



Protegrity Installation Guide 9.1.0.5

Created on: Nov 19, 2024

Copyright

Copyright © 2004-2024 Protegility Corporation. All rights reserved.

Protegility products are protected by and subject to patent protections;

Patent: <https://www.protegility.com/patents>.

Protegility logo is the trademark of Protegility Corporation.

NOTICE TO ALL PERSONS RECEIVING THIS DOCUMENT

Some of the product names mentioned herein are used for identification purposes only and may be trademarks and/or registered trademarks of their respective owners.

Windows, Azure, MS-SQL Server, Internet Explorer and Internet Explorer logo, Active Directory, and Hyper-V are registered trademarks of Microsoft Corporation in the United States and/or other countries.

Linux is a registered trademark of Linus Torvalds in the United States and other countries.

UNIX is a registered trademark of The Open Group in the United States and other countries.

SCO and SCO UnixWare are registered trademarks of The SCO Group.

Sun, Oracle, Java, and Solaris are the registered trademarks of Oracle Corporation and/or its affiliates in the United States and other countries.

Teradata and the Teradata logo are the trademarks or registered trademarks of Teradata Corporation or its affiliates in the United States and other countries.

Hadoop or Apache Hadoop, Hadoop elephant logo, Hive, and Pig are trademarks of Apache Software Foundation.

Cloudera and the Cloudera logo are trademarks of Cloudera and its suppliers or licensors.

Hortonworks and the Hortonworks logo are the trademarks of Hortonworks, Inc. in the United States and other countries.

Greenplum Database is the registered trademark of VMware Corporation in the U.S. and other countries.

Pivotal HD is the registered trademark of Pivotal, Inc. in the U.S. and other countries.

PostgreSQL or Postgres is the copyright of The PostgreSQL Global Development Group and The Regents of the University of California.

AIX, DB2, IBM and the IBM logo, and z/OS are registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide.

Utimaco Safeware AG is a member of the Sophos Group.

Xen, XenServer, and Xen Source are trademarks or registered trademarks of Citrix Systems, Inc. and/or one or more of its subsidiaries, and may be registered in the United States Patent and Trademark Office and in other countries.

VMware, the VMware “boxes” logo and design, Virtual SMP and VMotion are registered trademarks or trademarks of VMware, Inc. in the United States and/or other jurisdictions.

Amazon Web Services (AWS) and AWS Marks are the registered trademarks of Amazon.com, Inc. in the United States and other countries.

HP is a registered trademark of the Hewlett-Packard Company.

HPE Ezmeral Data Fabric is the trademark of Hewlett Packard Enterprise in the United States and other countries.

Dell is a registered trademark of Dell Inc.

Novell is a registered trademark of Novell, Inc. in the United States and other countries.

POSIX is a registered trademark of the Institute of Electrical and Electronics Engineers, Inc.

Mozilla and Firefox are registered trademarks of Mozilla foundation.

Chrome and Google Cloud Platform (GCP) are registered trademarks of Google Inc.

Table of Contents

| | |
|---|-----------|
| Copyright..... | 2 |
| | |
| Chapter 1 Introduction to this Guide..... | 12 |
| 1.1 Sections contained in this Guide..... | 12 |
| 1.2 Accessing the Protegility documentation suite..... | 13 |
| 1.2.1 Viewing product documentation..... | 13 |
| 1.2.2 Downloading product documentation..... | 13 |
| | |
| Chapter 2 Overview of Installation..... | 15 |
| 2.1 Audience..... | 15 |
| 2.2 Protegility Data Security Platform..... | 15 |
| 2.2.1 Architecture..... | 15 |
| | |
| Chapter 3 System Requirements..... | 17 |
| 3.1 System Requirements for the ESA..... | 17 |
| 3.2 Certificate Requirements..... | 17 |
| | |
| Chapter 4 Installing the ESA On-Premise..... | 19 |
| 4.1 Selecting Network Interface Cards (NICs)..... | 21 |
| 4.2 Configuring Network Settings..... | 22 |
| 4.3 Configuring Time Zone..... | 24 |
| 4.4 Configuring the Nearest Location..... | 25 |
| 4.5 Updating the Date and Time..... | 26 |
| 4.6 Updating the Keyboard Settings..... | 28 |
| 4.7 Configuring GRUB Settings..... | 29 |
| 4.8 Setting up Users and Passwords..... | 31 |
| 4.9 Licensing..... | 32 |
| 4.10 Installing Products..... | 32 |
| | |
| Chapter 5 Installing Appliances on Cloud Platforms..... | 34 |
| 5.1 Installing Protegility Appliances on Amazon Web Services (AWS)..... | 34 |
| 5.1.1 Verifying Prerequisites..... | 35 |
| 5.1.1.1 Prerequisites..... | 35 |
| 5.1.1.2 Hardware Requirements..... | 35 |
| 5.1.1.3 Network Requirements..... | 35 |
| 5.1.2 Obtaining the AMI..... | 36 |
| 5.1.3 Loading the Protegility Appliance from an Amazon Machine Image (AMI)..... | 40 |
| 5.1.3.1 Creating an Instance of the Protegility Appliance from the AMI..... | 40 |
| 5.1.3.2 Configuring the Virtual Private Cloud (VPC)..... | 42 |
| 5.1.3.3 Adding a Subnet to the Virtual Private Cloud (VPC)..... | 42 |
| 5.1.3.4 Finalizing the Installation of Protegility Appliance on the Instance..... | 43 |
| 5.1.3.5 Connecting to an ESA instance (for DSG deployment)..... | 45 |
| 5.1.4 Increasing Disk Space on the Appliance..... | 46 |
| 5.2 Installing Protegility Appliances on Azure..... | 47 |
| 5.2.1 Prerequisites..... | 47 |
| 5.2.1.1 Hardware Requirements..... | 47 |
| 5.2.1.2 Network Requirements..... | 48 |
| 5.2.2 Azure Cloud Utility..... | 48 |
| 5.2.3 Setting up Azure Virtual Network..... | 48 |
| 5.2.4 Creating a Resource Group..... | 48 |
| 5.2.5 Creating a Storage Account..... | 49 |
| 5.2.6 Creating a Container..... | 49 |
| 5.2.7 Obtaining the Azure BLOB..... | 49 |



| | |
|--|-----------|
| 5.2.8 Creating Image from the Azure BLOB..... | 52 |
| 5.2.9 Creating a VM from the Image..... | 52 |
| 5.2.10 Accessing the Appliance..... | 53 |
| 5.2.11 Finalizing the Installation of Protegility Appliance on the Instance..... | 53 |
| 5.2.11.1 Finalizing ESA Installation..... | 54 |
| 5.2.12 Connecting to an ESA Instance..... | 56 |
| 5.3 Installing Protegility Appliances on Google Cloud Platform (GCP)..... | 56 |
| 5.3.1 Verifying Prerequisites | 56 |
| 5.3.1.1 Prerequisites..... | 56 |
| 5.3.1.2 Hardware Requirements..... | 56 |
| 5.3.1.3 Network Requirements..... | 57 |
| 5.3.2 Configuring the Virtual Private Cloud (VPC)..... | 57 |
| 5.3.2.1 Adding a Subnet to the Virtual Private Cloud (VPC)..... | 58 |
| 5.3.3 Obtaining the GCP Image..... | 58 |
| 5.3.4 Converting the Raw Disk to a GCP Image..... | 60 |
| 5.3.5 Loading the Protegility Appliance from a GCP Image..... | 61 |
| 5.3.5.1 Creating a VM Instance from an Image..... | 61 |
| 5.3.5.2 Creating a VM Instance from a Disk..... | 62 |
| 5.3.5.3 Accessing the Appliance..... | 64 |
| 5.3.6 Finalizing the Installation of Protegility Appliance..... | 64 |
| 5.3.6.1 Finalizing ESA Installation..... | 65 |
| 5.3.7 Connecting to an ESA instance (for DSG deployment)..... | 67 |
| 5.3.8 Deploying the Instance of the Protegility Appliance with the Protectors..... | 67 |
| 5.3.9 Backing up and Restoring Data on GCP..... | 67 |
| 5.3.9.1 Creating a Snapshot of a Disk on GCP..... | 67 |
| 5.3.9.2 Restoring from a Snapshot on GCP..... | 68 |
| 5.3.10 Increasing Disk Space on the Appliance..... | 68 |
| Chapter 6 Configuring the ESA..... | 70 |
| 6.1 Configuring Authentication Settings..... | 70 |
| 6.2 Configuring Accounts and Passwords..... | 70 |
| 6.3 Configuring Syslog..... | 71 |
| 6.4 Configuring External Certificates..... | 71 |
| 6.5 Configuring SMTP..... | 71 |
| 6.6 Configuring SNMP..... | 72 |
| Chapter 7 Initializing the Policy Information Management (PIM) Module..... | 73 |
| Chapter 8 Configuring the ESA in a Trusted Appliances Cluster (TAC)..... | 74 |
| Chapter 9 Creating an Audit Store Cluster..... | 75 |
| 9.1 Completing the Prerequisites..... | 75 |
| 9.2 Initializing the Audit Store Cluster on the ESA..... | 75 |
| 9.3 Adding an ESA to the Audit Store Cluster..... | 76 |
| 9.4 Refreshing the Audit Store Cluster..... | 79 |
| 9.5 Configuring td-agent in the Audit Store Cluster..... | 79 |
| 9.6 Verifying the Audit Store Cluster..... | 80 |
| 9.7 Updating the Priority IP List for Signature Verification..... | 80 |
| Chapter 10 Verifying the ESA Installation from the Web UI..... | 82 |
| Chapter 11 Installing the Data Security Gateway (DSG)..... | 83 |
| 11.1 Installing the DSG..... | 83 |
| 11.2 Installing the DSG on Cloud Platforms..... | 83 |

| | |
|---|-----------|
| 11.2.1 Configuring Cloud Instances..... | 84 |
| Chapter 12 Protegility Data Protectors Installation..... | 85 |
| 12.1 General Architecture..... | 85 |
| 12.2 Installing the Log Forwarder on Protectors..... | 86 |
| 12.2.1 Installing the Log Forwarder..... | 87 |
| 12.2.1.1 Installing the Log Forwarder on Linux..... | 87 |
| 12.2.1.2 Installing the Log Forwarder on Windows..... | 88 |
| 12.2.1.3 Configuring the Log Forwarder for AIX platform..... | 90 |
| 12.2.2 Log Forwarder Configurations..... | 93 |
| 12.2.3 Running the Log Forwarder..... | 94 |
| 12.3 Installing and Uninstalling the PEP Server..... | 94 |
| 12.3.1 PEP Server Pre-Installation Preparation..... | 94 |
| 12.3.1.1 Disk Space Requirements..... | 95 |
| 12.3.1.2 Shared Memory Requirements..... | 95 |
| 12.3.1.3 Stack Memory Requirements..... | 95 |
| 12.3.1.4 Hostname Localhost Verification..... | 96 |
| 12.3.2 Installing the PEP Server..... | 96 |
| 12.3.2.1 PEP Server on Windows..... | 96 |
| 12.3.2.2 PEP Server on UNIX..... | 97 |
| 12.3.3 Configuring PEP Server..... | 98 |
| 12.3.3.1 Customization of pepserver.cfg File..... | 98 |
| 12.3.3.2 PEP Server Configuration Parameters..... | 99 |
| 12.3.3.3 Configuring the PEP Server to start automatically..... | 99 |
| 12.3.4 Configuring the PEP Server for AIX platform..... | 100 |
| 12.3.5 Synchronizing Certificate Files..... | 100 |
| 12.3.5.1 Retrieving Certificate Files..... | 101 |
| 12.3.5.2 Retrieving Certificate Files Examples..... | 101 |
| 12.3.6 Running the PEP Server..... | 102 |
| 12.3.7 Uninstalling the PEP Server from Unix or Linux Platforms..... | 103 |
| 12.3.8 Uninstalling the PEP Server from Windows Platform..... | 103 |
| 12.4 Protector Proxy Installation..... | 103 |
| 12.4.1 Introduction..... | 103 |
| 12.4.2 Installing and Uninstalling Protector Proxy..... | 104 |
| 12.4.2.1 Installing the Protector Proxy..... | 104 |
| 12.4.2.2 Running the Protector Proxy..... | 105 |
| 12.4.2.3 Configuring the Protector Proxy..... | 105 |
| 12.4.2.4 Uninstalling the Protector Proxy..... | 107 |
| 12.5 Installing and Uninstalling Application Protectors..... | 107 |
| 12.5.1 Installing Application Protectors C..... | 107 |
| 12.5.1.1 Setting up Application Protector C on Linux or Unix..... | 107 |
| 12.5.1.2 Setting up Application Protector C on Windows..... | 112 |
| 12.5.1.3 Uninstalling Application Protector C..... | 129 |
| 12.5.2 Installing Application Protector (AP) Java..... | 131 |
| 12.5.2.1 Setting up Application Protector Java on Linux or Unix..... | 131 |
| 12.5.2.2 Setting up Application Protectors Java on Windows..... | 136 |
| 12.5.2.3 Setting up Application Protector Java on AIX..... | 149 |
| 12.5.2.4 Configuring Application Protector Java..... | 152 |
| 12.5.2.5 Checking Installation Success..... | 152 |
| 12.5.2.6 Uninstalling Application Protector Java..... | 152 |
| 12.5.3 Installing Application Protector (AP) Python..... | 154 |
| 12.5.3.1 Setting up AP Python on Linux or Unix in a Production Environment..... | 155 |
| 12.5.3.2 Setting up AP Python on Windows in a Production Environment..... | 161 |
| 12.5.3.3 Setting Up AP Python in a Development Environment..... | 174 |
| 12.5.4 Installing Application Protector Go..... | 177 |
| 12.5.4.1 Setting up Application Protector Go on Linux or Unix..... | 177 |
| 12.5.4.2 Uninstalling Application Protector Go..... | 183 |
| 12.5.5 Installing Application Protector NodeJS..... | 184 |



| | |
|---|-----|
| 12.5.5.1 Setting up Application Protector NodeJS on Linux or Unix..... | 184 |
| 12.5.5.2 Installing Log Forwarder on Linux or Unix..... | 184 |
| 12.5.5.3 Installing PEP Server on Linux or Unix..... | 186 |
| 12.5.5.4 Installing Application Protector NodeJS on Linux or Unix..... | 188 |
| 12.5.5.5 Uninstalling Application Protector NodeJS..... | 188 |
| 12.5.6 Installing Application Protector (AP) .Net..... | 189 |
| 12.5.6.1 Setting up Application Protector .Net on Windows..... | 189 |
| 12.5.6.2 Uninstalling Application Protector .Net..... | 206 |
| 12.6 Installing and Uninstalling the Big Data Protector..... | 208 |
| 12.6.1 Installing the Big Data Protector Using the CDP Private Cloud Base (CDP-PVC-Base) Installer..... | 209 |
| 12.6.1.1 Understanding the Installation Methods..... | 209 |
| 12.6.1.2 Verifying the Prerequisites for Installing the Big Data Protector..... | 210 |
| 12.6.1.3 Extracting the Big Data Protector Package..... | 211 |
| 12.6.1.4 Running the Big Data Protector Configurator Script..... | 211 |
| 12.6.1.5 Setting up the Big Data Protector Parcels and CSDs..... | 214 |
| 12.6.1.6 Distributing the Big Data Protector Parcels to the Nodes..... | 216 |
| 12.6.1.7 Activating the Big Data Protector Parcels on the Nodes..... | 219 |
| 12.6.1.8 Starting the Big Data Protector Services..... | 224 |
| 12.6.1.9 Installing the Big Data Protector Parcels on a New Node..... | 237 |
| 12.6.1.10 Updating the Big Data Protector Configuration Parameters..... | 237 |
| 12.6.1.11 Setting the Big Data Protector Configuration for CDP-PVC-Base Distribution..... | 239 |
| 12.6.1.12 Working with the Certificate Parcels..... | 240 |
| 12.6.1.13 Updating the Fluent Bit Parcel..... | 248 |
| 12.6.1.14 Enabling the PEP Server Application Log File..... | 250 |
| 12.6.1.15 Installing the Big Data Protector after Upgrading to CDP-PVC-Base Distribution..... | 253 |
| 12.6.1.16 Performing an Upgrade of the CDP-PVC-Base Distribution..... | 256 |
| 12.6.1.17 Uninstalling the Big Data Protector..... | 256 |
| 12.6.2 Installing Big Data Protector using CDH Native Installer..... | 269 |
| 12.6.2.1 Verifying the Prerequisites for Installing the Big Data Protector..... | 270 |
| 12.6.2.2 Installing the Big Data Protector..... | 271 |
| 12.6.2.3 Downloading the Big Data Protector Package..... | 271 |
| 12.6.2.4 Extracting the Big Data Protector Package..... | 271 |
| 12.6.2.5 Running the Big Data Protector Configurator Script..... | 271 |
| 12.6.2.6 Setting up the Big Data Protector Parcels and CSDs..... | 274 |
| 12.6.2.7 Distributing the Big Data Protector Parcels to the Nodes..... | 275 |
| 12.6.2.8 Activating the Big Data Protector Parcel on the Nodes..... | 279 |
| 12.6.2.9 Verifying the Permissions and Ownerships for the Extracted Parcels..... | 282 |
| 12.6.2.10 Starting the Big Data Protector Services on the Nodes..... | 282 |
| 12.6.2.11 Managing the Big Data Protector on Cloudera Manager..... | 294 |
| 12.6.2.12 Updating the Certificates Parcel..... | 295 |
| 12.6.2.13 Updating the Fluent Bit Parcel..... | 297 |
| 12.6.2.14 Generating the <i>pepserver.log</i> File..... | 299 |
| 12.6.2.15 Performing an Upgrade of the CDH Distribution..... | 302 |
| 12.6.2.16 Uninstalling the Big Data Protector from all the Nodes..... | 302 |
| 12.6.3 Installing the Big Data Protector using Ambari Native Installer..... | 315 |
| 12.6.3.1 Verifying Prerequisites for Installing Big Data Protector..... | 316 |
| 12.6.3.2 Installing Big Data Protector..... | 316 |
| 12.6.3.3 Adding the Big Data Protector PEP Service..... | 328 |
| 12.6.3.4 Starting the Certificates Service..... | 334 |
| 12.6.3.5 Installing Big Data Protector on a New Node..... | 338 |
| 12.6.3.6 Installing Big Data Protector on a New Node using Cloudbreak..... | 338 |
| 12.6.3.7 Managing the Big Data Protector on HDP..... | 341 |
| 12.6.3.8 Updating the Certificates Management Pack..... | 341 |
| 12.6.3.9 Configuring Ambari to Set-up Multiple External Audit Stores..... | 342 |
| 12.6.3.10 Uninstalling the Big Data Protector Services from the Nodes..... | 346 |
| 12.6.4 Installing Big Data Protector using the Shell-based Installer..... | 355 |
| 12.6.4.1 Verifying Prerequisites for Installing Big Data Protector..... | 355 |
| 12.6.4.2 Extracting Files from the Installation Package..... | 357 |



| | |
|---|-----|
| 12.6.4.3 Updating the BDP.config File..... | 357 |
| 12.6.4.4 Setting up the Proxy..... | 358 |
| 12.6.4.5 Installing Big Data Protector..... | 359 |
| 12.6.4.6 Installing or Uninstalling Big Data Protector on Specific Nodes..... | 362 |
| 12.6.4.7 Uninstalling Big Data Protector from a Cluster..... | 363 |
| 12.6.5 Installing the Big Data Protector on an Amazon EMR Cluster..... | 364 |
| 12.6.5.1 Verifying the Prerequisites for Installing the Big Data Protector..... | 364 |
| 12.6.5.2 Creating an S3 Bucket on AWS..... | 365 |
| 12.6.5.3 Downloading the Big Data Protector Package..... | 365 |
| 12.6.5.4 Extracting the Big Data Protector Package..... | 365 |
| 12.6.5.5 Setting up the Proxy..... | 366 |
| 12.6.5.6 Running the Configurator Script..... | 368 |
| 12.6.5.7 Installing Big Data Protector on a New EMR Cluster..... | 373 |
| 12.6.5.8 Installing the Big Data Protector on an Existing EMR Cluster..... | 381 |
| 12.6.5.9 Using Hive UDFs with Amazon EMR..... | 396 |
| 12.6.5.10 Best Practices for Using Big Data Protector on EMR..... | 397 |
| 12.6.6 Installing the Big Data Protector on a Dataproc Cluster..... | 398 |
| 12.6.6.1 Verifying the Prerequisites for Installing the Big Data Protector..... | 398 |
| 12.6.6.2 Creating a Storage Bucket on the Google Cloud Platform..... | 399 |
| 12.6.6.3 Downloading the Big Data Protector Package..... | 400 |
| 12.6.6.4 Extracting the Big Data Protector Package..... | 400 |
| 12.6.6.5 Setting up the Proxy..... | 400 |
| 12.6.6.6 Running the Configurator Script..... | 402 |
| 12.6.6.7 Installing the Big Data Protector..... | 406 |
| 12.6.6.8 Uninstalling the Big Data Protector..... | 414 |
| 12.6.7 Installing Big Data Protector on an Azure HDInsight Cluster..... | 414 |
| 12.6.7.1 Verifying Prerequisites for Installing Big Data Protector..... | 414 |
| 12.6.7.2 Creating a Storage Location on the Azure Platform..... | 415 |
| 12.6.7.3 Downloading the Big Data Protector Package..... | 415 |
| 12.6.7.4 Extracting the Big Data Protector Package..... | 415 |
| 12.6.7.5 Setting up the Proxy..... | 416 |
| 12.6.7.6 Running the Configurator Script..... | 416 |
| 12.6.7.7 Installing Big Data Protector on a New HDInsight Cluster..... | 418 |
| 12.6.7.8 Installing Big Data Protector on an Existing HDInsight Cluster..... | 427 |
| 12.6.7.9 Scaling and Shrinking of the Nodes in the Cluster..... | 431 |
| 12.6.7.10 Best Practices for Using Big Data Protector on the HDInsights Cluster..... | 431 |
| 12.6.7.11 Uninstalling Big Data Protector..... | 432 |
| 12.6.8 Installing the Big Data Protector on an AWS Databricks Cluster using the Amazon S3 Bucket..... | 432 |
| 12.6.8.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using the S3 Bucket..... | 432 |
| 12.6.8.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks..... | 433 |
| 12.6.8.3 Creating an S3 Bucket on AWS..... | 435 |
| 12.6.8.4 Extracting the Big Data Protector Package..... | 435 |
| 12.6.8.5 Running the Configurator Script..... | 435 |
| 12.6.8.6 Modifying the <i>pepperserver.cfg</i> File..... | 438 |
| 12.6.8.7 Uploading the Installation Files to the Amazon S3 Bucket..... | 439 |
| 12.6.8.8 Installing the Big Data Hive and Spark Protector..... | 442 |
| 12.6.8.9 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 448 |
| 12.6.9 Installing the Big Data Protector on an AWS Databricks Cluster using the Workspace..... | 450 |
| 12.6.9.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using the Workspace..... | 450 |
| 12.6.9.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks..... | 451 |
| 12.6.9.3 Extracting the Big Data Protector Package..... | 453 |
| 12.6.9.4 Running the Configurator Script..... | 453 |
| 12.6.9.5 Modifying the <i>pepperserver.cfg</i> File..... | 456 |
| 12.6.9.6 Uploading the Installation Files to the Workspace Storage..... | 457 |
| 12.6.9.7 Installing the Big Data Hive and Spark Protector..... | 459 |
| 12.6.9.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 463 |



| | |
|--|-----|
| 12.6.10 Installing the Big Data Protector on an AWS Databricks Cluster using DBFS..... | 465 |
| 12.6.10.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using DBFS..... | 465 |
| 12.6.10.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks..... | 466 |
| 12.6.10.3 Extracting the Big Data Protector Package..... | 467 |
| 12.6.10.4 Running the Configurator Script..... | 467 |
| 12.6.10.5 Modifying the <i>pepperserver.cfg</i> File..... | 470 |
| 12.6.10.6 Uploading the Files using the Helper Script..... | 471 |
| 12.6.10.7 Installing the Big Data Hive and Spark Protector..... | 473 |
| 12.6.10.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 476 |
| 12.6.11 Installing the Big Data Protector on an Azure Databricks Cluster Using ABFSS..... | 478 |
| 12.6.11.1 Verifying the Prerequisites for Installing the Big Data Protector on an Azure Databricks Cluster using ABFSS..... | 478 |
| 12.6.11.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks with ABFSS..... | 479 |
| 12.6.11.3 Extracting the Big Data Protector Package..... | 481 |
| 12.6.11.4 Running the Configurator Script..... | 481 |
| 12.6.11.5 Modifying the <i>pepperserver.cfg</i> File..... | 485 |
| 12.6.11.6 Uploading the Installation Files to ADLS..... | 486 |
| 12.6.11.7 Installing the Big Data Hive and Spark Protector..... | 489 |
| 12.6.11.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 494 |
| 12.6.12 Installing the Big Data Protector on an Azure Databricks Cluster using the Workspace..... | 495 |
| 12.6.12.1 Verifying the Prerequisites for Installing the Big Data Protector on Azure Databricks Cluster using the Workspace..... | 495 |
| 12.6.12.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using the Workspace..... | 496 |
| 12.6.12.3 Extracting the Big Data Protector Package..... | 498 |
| 12.6.12.4 Running the Configurator Script..... | 498 |
| 12.6.12.5 Modifying the <i>pepperserver.cfg</i> File..... | 502 |
| 12.6.12.6 Uploading the Installation Files to the Workspace Storage..... | 503 |
| 12.6.12.7 Installing the Big Data Hive and Spark Protector..... | 504 |
| 12.6.12.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 508 |
| 12.6.13 Installing the Big Data Protector on an Azure Databricks Cluster using the Unity Catalog Volumes..... | 510 |
| 12.6.13.1 Verifying the Prerequisites for Installing the Big Data Protector on Azure Databricks Cluster using the Unity Catalog Volumes..... | 510 |
| 12.6.13.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using the Unity Catalog Volume..... | 511 |
| 12.6.13.3 Extracting the Big Data Protector Package..... | 512 |
| 12.6.13.4 Running the Configurator Script..... | 513 |
| 12.6.13.5 Modifying the <i>pepperserver.cfg</i> File..... | 516 |
| 12.6.13.6 Uploading the Installation Files to the Unity Catalog Volume..... | 517 |
| 12.6.13.7 Installing the Big Data Hive and Spark Protector..... | 520 |
| 12.6.13.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 523 |
| 12.6.14 Installing the Big Data Protector on an Azure Databricks Cluster Using DBFS..... | 525 |
| 12.6.14.1 Verifying the Prerequisites for Installing the Big Data Protector on an Azure Databricks Cluster using DBFS..... | 525 |
| 12.6.14.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using DBFS.. | 526 |
| 12.6.14.3 Extracting the Big Data Protector Package..... | 527 |
| 12.6.14.4 Running the Configurator Script..... | 527 |
| 12.6.14.5 Modifying the <i>pepperserver.cfg</i> File..... | 531 |
| 12.6.14.6 Uploading the Installation Files to DBFS..... | 532 |
| 12.6.14.7 Installing the Big Data Hive and Spark Protector..... | 533 |
| 12.6.14.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog..... | 537 |
| 12.6.15 Installing the Big Data Protector on the GCP Databricks Cluster..... | 538 |
| 12.6.15.1 Verifying the Prerequisites for Installing the Big Data Protector on a GCP Databricks Cluster..... | 538 |
| 12.6.15.2 Understanding the Installation Workflow of the Big Data Protector on GCP Databricks..... | 538 |
| 12.6.15.3 Extracting the Big Data Protector Package..... | 540 |
| 12.6.15.4 Running the Configurator Script..... | 540 |
| 12.6.15.5 Modifying the <i>pepperserver.cfg</i> File..... | 542 |



| | |
|---|-----|
| 12.6.15.6 Uploading the Files Manually to DBFS..... | 542 |
| 12.6.15.7 Installing the Big Data Hive and Spark Protector..... | 543 |
| 12.6.16 Installing the Big Data Protector on a CDP AWS Data Hub Platform..... | 545 |
| 12.6.16.1 Verifying the Prerequisites for Installing the Big Data Protector..... | 546 |
| 12.6.16.2 Extracting the Big Data Protector Package..... | 546 |
| 12.6.16.3 Running the Big Data Protector Configurator Script..... | 546 |
| 12.6.16.4 Registering the Recipe Scripts..... | 550 |
| 12.6.16.5 Creating and Registering the Custom Cluster Template..... | 551 |
| 12.6.16.6 Creating a Data Hub Cluster..... | 554 |
| 12.6.16.7 Setting the Big Data Protector Configuration for the CDP AWS Data Hub Platform..... | 557 |
| 12.6.16.8 Updating the Certificates Parcel on a AWS Data Hub Cluster..... | 560 |
| 12.6.16.9 Updating the Fluent Bit Configuration Parcel on a AWS Data Hub Cluster..... | 563 |
| 12.6.17 Installing the Big Data Protector on a CDP Azure Data Hub Platform..... | 566 |
| 12.6.17.1 Verifying the Prerequisites for Installing the Big Data Protector..... | 566 |
| 12.6.17.2 Extracting the Big Data Protector Package..... | 567 |
| 12.6.17.3 Running the Big Data Protector Configurator Script..... | 567 |
| 12.6.17.4 Registering the Recipe Scripts..... | 571 |
| 12.6.17.5 Creating and Registering the Custom Cluster Template..... | 572 |
| 12.6.17.6 Creating a Data Hub Cluster..... | 575 |
| 12.6.17.7 Setting the Big Data Protector Configuration for the CDP Azure Data Hub Platform..... | 578 |
| 12.6.17.8 Updating the Certificates Parcel on a Azure Data Hub Cluster..... | 581 |
| 12.6.17.9 Updating the Fluent Bit Configuration Parcel on a Azure Data Hub Cluster..... | 584 |
| 12.7 Installing and Uninstalling Database Protectors..... | 587 |
| 12.7.1 Installing and Uninstalling the MS SQL Database Protector..... | 587 |
| 12.7.1.1 Verifying the Prerequisites..... | 587 |
| 12.7.1.2 Installing the Log Forwarder..... | 588 |
| 12.7.1.3 Installing the PEP Server..... | 594 |
| 12.7.1.4 Installing the PEP for MS SQL Server..... | 598 |
| 12.7.1.5 Configuration of MS SQL Database Protector..... | 601 |
| 12.7.1.6 Installation of User Defined Functions (UDFs)..... | 601 |
| 12.7.1.7 Uninstalling the MS SQL Database Protector..... | 607 |
| 12.7.2 Installing and Uninstalling the Oracle Database Protector..... | 609 |
| 12.7.2.1 Installing the Oracle Database Protector..... | 609 |
| 12.7.2.2 Uninstalling the Oracle Database Protector..... | 616 |
| 12.7.3 Installing and Uninstalling the Teradata Database Protector..... | 617 |
| 12.7.3.1 Prerequisites..... | 617 |
| 12.7.3.2 Installing the Teradata Database Protector..... | 618 |
| 12.7.3.3 Installing the User Defined Functions (UDFs) for Teradata..... | 624 |
| 12.7.3.4 Uninstalling the Teradata Database Protector..... | 625 |
| 12.7.4 Installing and Uninstalling the Greenplum Database Protector..... | 626 |
| 12.7.4.1 Setup Overview for UNIX..... | 626 |
| 12.7.4.2 Prerequisites..... | 627 |
| 12.7.4.3 Installing the Greenplum Database Protector on a Single Node..... | 627 |
| 12.7.4.4 Installing and Uninstalling Greenplum Database Protector on Multiple Nodes..... | 628 |
| 12.7.4.5 Installing UDFs of Greenplum Database Protector..... | 630 |
| 12.7.4.6 Uninstalling the Greenplum Database Protector..... | 631 |
| 12.7.5 Installing and Uninstalling the DB2 Database Protector..... | 631 |
| 12.7.5.1 Verifying the Prerequisites for Installing the DB2 Database Protector..... | 631 |
| 12.7.5.2 Creating the Directory to Install the DB2 Database Protector..... | 632 |
| 12.7.5.3 Installing the Log Forwarder..... | 632 |
| 12.7.5.4 Installing the PEP Server..... | 635 |
| 12.7.5.5 Verifying the Status of the PEP Server..... | 636 |
| 12.7.5.6 Installing the PEP for DB2..... | 636 |
| 12.7.5.7 Installing the UDFs..... | 637 |
| 12.7.5.8 Verifying the Installation of UDFs..... | 638 |
| 12.7.5.9 Uninstalling the DB2 Database Protector..... | 638 |
| 12.7.6 Installing and Uninstalling Netezza Database Protector..... | 640 |
| 12.7.6.1 Prerequisites..... | 640 |



| | |
|---|------------|
| 12.7.6.2 Downloading and Extracting Protector Package..... | 641 |
| 12.7.6.3 Installing Netezza Database Protector on Netezza Host..... | 641 |
| 12.7.6.4 Configuring Netezza Database Protector..... | 643 |
| 12.7.6.5 Verifying Netezza Database Protector Installation and Configuration..... | 644 |
| 12.7.6.6 Using Netezza Database Protector..... | 644 |
| 12.7.6.7 Uninstallation..... | 646 |
| 12.7.7 Installing and Uninstalling the Trino Protector..... | 646 |
| 12.7.7.1 Verifying the Prerequisites for Installing the Trino Protector..... | 647 |
| 12.7.7.2 Extracting the Files from the Installation Package..... | 648 |
| 12.7.7.3 Running the Configurator Script..... | 648 |
| 12.7.7.4 Running the Installation Script on Every Node in the Trino Cluster..... | 651 |
| 12.7.7.5 Working with the Cluster Utilities..... | 655 |
| 12.7.7.6 Executing the Uninstallation Script..... | 665 |
| Chapter 13 Appendix A: PEP Server Configuration File..... | 669 |
| Chapter 14 Appendix: Configuring a Trusted Appliance Cluster (TAC) without Consul Integration..... | 675 |
| Chapter 15 Appendix: Verifying the Immutable Service on the ESA to Export Policy for Immutable Protectors..... | 676 |
| Chapter 16 Appendix B: Using Go Module with Private GitLab Repository..... | 677 |
| Chapter 17 Appendix: Configuring the IP address for the Docker Interface..... | 678 |
| Chapter Appendix: Audit Store Performance Analysis..... | 680 |

Chapter 1

Introduction to this Guide

1.1 Sections contained in this Guide

1.2 Accessing the Protegility documentation suite

This guide provides information that you need to install Protegility products.

1.1 Sections contained in this Guide

This guide is divided into the following sections:

- Section *Introduction to this Guide* is the Introduction to this Guide that defines the purpose and scope for this guide. In addition, it explains how information is organized in this guide.
- Section *Overview of Installation* provides an overview of the Protegility Data Security Platform and its general architecture.
- Section *System Requirements for the ESA* provides the system requirements.
- Section *Installing the ESA On-Premise* demonstrates the ESA Appliance Installation.
- Section *Installing Appliances on Cloud Platforms* describes installation on Cloud Platforms.
- Section *Configuring the ESA* provides information about configuring the ESA.
- Section *Initializing the Policy Information Management (PIM) Module* provides information about initializing the Policy Information Management (PIM) module.
- Section *Configuring the ESA in a Trusted Appliances Cluster (TAC)* provides information about configuring the Trusted Appliance Cluster.
- Section *Creating an Audit Store Cluster* provides information about configuring the Audit Store Cluster.
- Section *Verifying the ESA Installation from the Web UI* provides information about verifying the ESA installation.
- Section *Installing the Data Security Gateway (DSG)* explains the Protegility Gateway Technology. In addition, this section includes an overview of the Data Security Gateway (DSG).
- Section *Protegility Data Protectors Installation* explains the information about installing and uninstalling the protectors.
- Section *Appendix A: PEP Server Configuration File* contains the code for PEP server configuration file that demonstrates the usage of PEP server configuration file used for customization of the PEP servers.
- Section *Appendix: Configuring a Trusted Appliance Cluster (TAC) without Consul Integration* describes the steps to configure a Trusted Appliance Cluster (TAC) without Consul Integration.
- Section *Appendix: Verifying the Immutable Service on the ESA to Export Policy for Immutable Protectors* contains the steps to verify if the Immutable Service is installed on the ESA.
- Section *Appendix B: Using Go Module with Private GitLab Repository* describes the steps to use the Go module.
- Section *Appendix: Configuring the IP address for the Docker Interface* describes the steps to configure the IP address for the Docker Interface.

1.2 Accessing the Protegility documentation suite

This section describes the methods to access the *Protegility Documentation Suite* using the [My.Protegility](#) portal.

1.2.1 Viewing product documentation

The **Product Documentation** section under **Resources** is a repository for Protegility product documentation. The documentation for the latest product release is displayed first. The documentation is available in the HTML format and can be viewed using your browser. You can also view and download the *.pdf* files of the required product documentation.

1. Log in to the [My.Protegility](#) portal.
2. Click **Resources > Product Documentation**.
3. Click a product version.
The documentation appears.

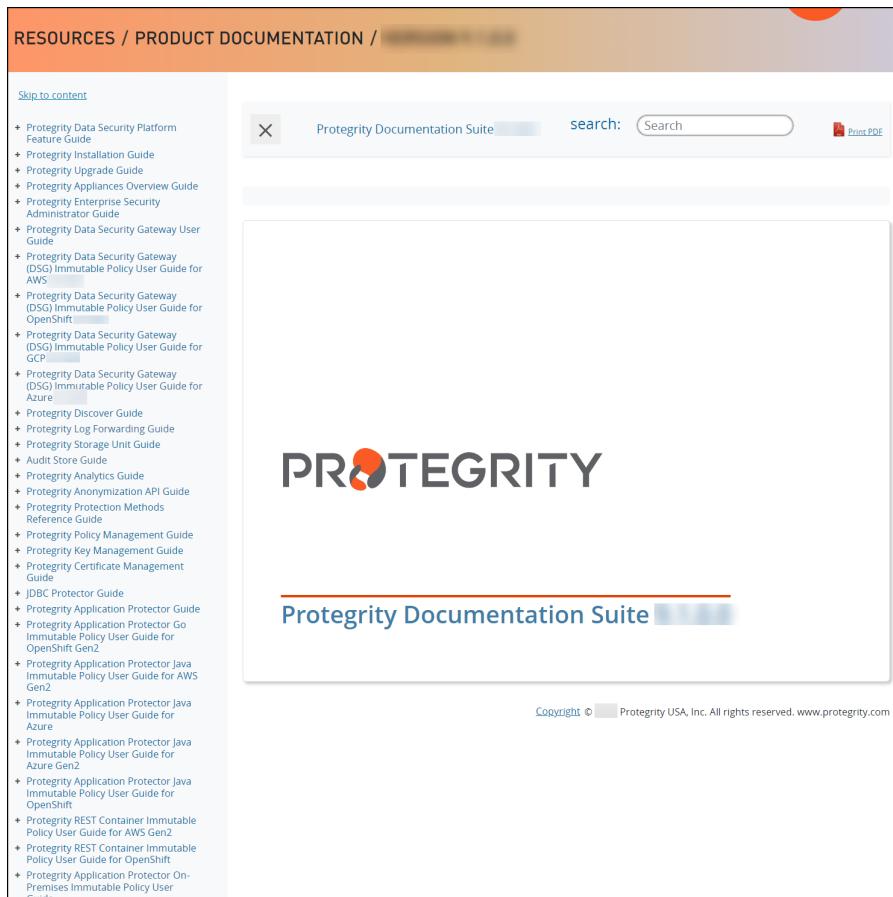


Figure 1-1: Documentation

4. Expand and click the link for the required documentation.
5. If required, then enter text in the **Search** field to search for keywords in the documentation.
The search is dynamic, and filters results while you type the text.
6. Click the **Print PDF** icon from the upper-right corner of the page.
The page with links for viewing and downloading the guides appears. You can view and print the guides that you require.

1.2.2 Downloading product documentation

This section explains the procedure to download the product documentation from the [My.Protegility](#) portal.

1. Click **Product Management > Explore Products**.
2. Select **Product Documentation**.
The **Explore Products** page is displayed. You can view the product documentation of various Protegility products as per their releases, containing an overview and other guidelines to use these products at ease.
3. Click **View Products** to advance to the product listing screen.
4. Click the **View** icon (⌚) from the **Action** column for the row marked **On-Prem** in the **Target Platform Details** column.
If you want to filter the list, then use the filters for: **OS**, **Target Platform**, and **Search** fields.
5. Click the icon for the action that you want to perform.

Chapter 2

Overview of Installation

2.1 Audience

2.2 Protegility Data Security Platform

This section provides a general overview of the Protegility Data Security Platform. In addition, this section contains the intended audience of this guide.

2.1 Audience

This Installation Guide is intended for the following stakeholders:

- Security professionals like security officers who are responsible for protecting business systems in organizations. They plan and ensure execution of security arrangement for their organization.
- System administrators and other technical personnel who are responsible for implementing data security solutions in their organization.
- System Architects who are responsible for providing expert guidance in designing, development and implementation of enterprise data security solution architecture for their business requirements.

2.2 Protegility Data Security Platform

The Protegility Data Security Platform is a comprehensive source of enterprise data protection solutions. Its design is based on a hub and spoke deployment architecture.

The Protegility Data Security Platform has following components:

Enterprise Security Administrator (ESA) – that handles the management of policies, keys, monitoring, auditing and reporting of protected systems in the enterprise.

Data Protectors – that protect sensitive data in the enterprise and deploy security policy for enforcement on each installed system. Policy is deployed from ESA to the Data Protectors. The Audit Logs of all activity on sensitive data are reported and stored in the Audit Store cluster on the ESA.

2.2.1 Architecture

The following diagram shows the general architecture of the Protegility Data Security Platform.

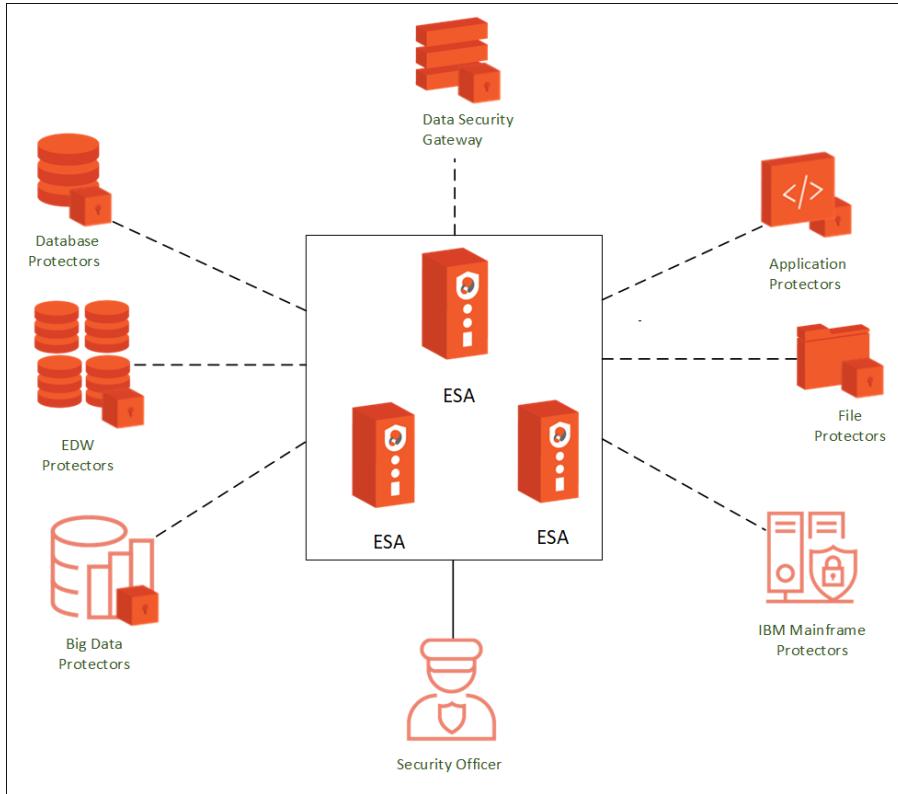


Figure 2-1: Protegility Data Security Platform

Chapter 3

System Requirements

3.1 System Requirements for the ESA

3.2 Certificate Requirements

This section describes the hardware requirements for the appliances components.

3.1 System Requirements for the ESA

The compatibility settings for your products to run smoothly are listed in the following table.

Table 3-1: Compatibility of Appliances Components

| Component | Compatibility |
|-----------------------|---|
| Application Protocols | HTTP 1.0, HTTP 1.1, SSL/TLS |
| WebServices | SOAP 1.1 and WSDL 1.1 |
| Web Browsers | Minimum supported Web Browser versions are as follows: <ul style="list-style-type: none">• Google Chrome version 123.0.6312.123 (64-bit)• Mozilla Firefox version 124.0.2 (64-bit)• Microsoft Edge version 123.0.2420.81 (64-bit) |

The minimum hardware configuration recommended is as follows:

| Hardware Components | Configuration |
|---------------------|--|
| CPU | Multicore Processor, with minimum 8 CPUs |
| RAM | 32 GB |
| Hard Disk | 320 GB |
| CPU Architecture | x86 |

3.2 Certificate Requirements

Certificates are used for secure communication between the ESA and protectors. The certificate-based communication and authentication involves a client certificate, server certificate, and a certifying authority that authenticates the client and server certificates.

The various components within the Protegility Data Security Platform that communicate with and authenticate each other through digital certificates are:

- ESA Web UI and ESA
- Audit Store



- ESA and Protectors
- Protegility Appliances and external REST clients

Note: Protegility client and server certificates are self-signed by Protegility. However, you can replace them by certificates signed by a trusted and commercial CA. These certificates are used for communication between various components in ESA.

For more information about the certificates, refer to the *Protegility Certificate Management Guide 9.1.0.5*.

Chapter 4

Installing the ESA On-Premise

4.1 Selecting Network Interface Cards (NICs)

4.2 Configuring Network Settings

4.3 Configuring Time Zone

4.4 Configuring the Nearest Location

4.5 Updating the Date and Time

4.6 Updating the Keyboard Settings

4.7 Configuring GRUB Settings

4.8 Setting up Users and Passwords

4.9 Licensing

4.10 Installing Products

The following steps explain the installation of the ESA v9.1.0.5 ISO image on-premise.

► To install the ESA appliance:

1. Insert the appliance installation media in the system disk drive.
2. Boot the system from the disk drive.
The following screen appears.



3. Press **ENTER** to start the installation.

The following screen appears.

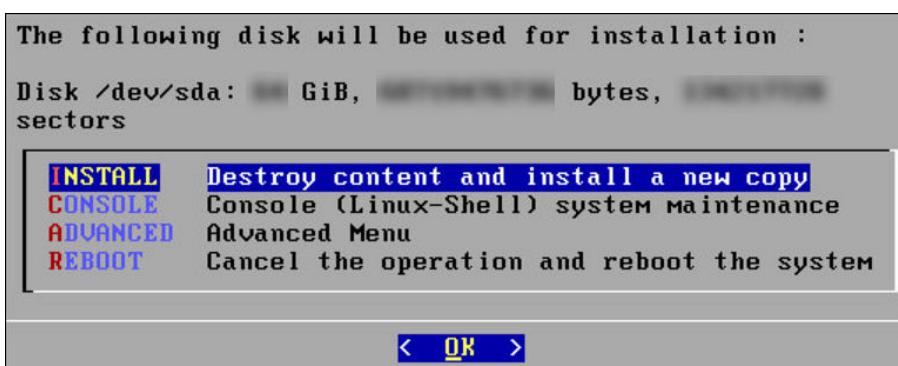


Figure 4-1: Select Installation Type Screen

The system will detect the number of hard drives that are present. If there are multiple hard drives, then it will allow you to choose the hard drive where you want to install the *OS* partition and the */opt* partition.

If there are multiple hard drives, then the following screen appears.

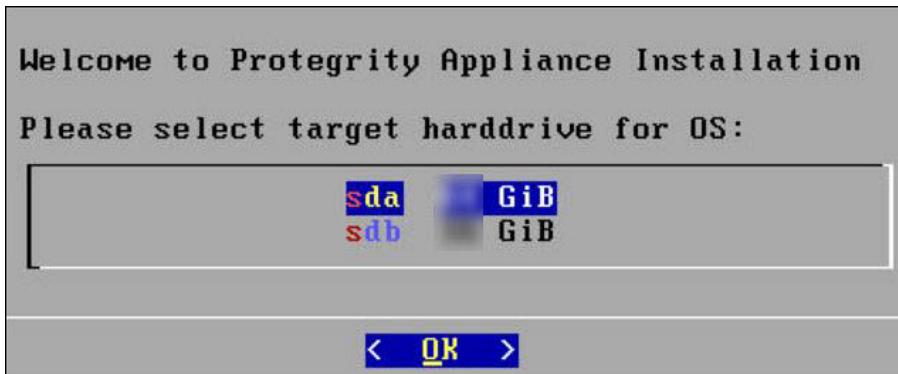


Figure 4-2: Select Target Hard Drive for OS partition

For storing the operating system-related data, select the hard drive where you want to install the OS partition and select **OK**.

The following screen appears.



Figure 4-3: Select Target Hard Drive for /opt partition

For storing the logs, configuration data, and so on select the hard drive where you want to install the */opt* partition and select **OK**.

4.1 Selecting Network Interface Cards (NICs)

Before you begin

The Network Interface Card (NIC) is a device through which appliances, such as, the ESA or the DSG, connect to each other on a network. You can configure multiple network interface cards (NICs) on the appliance.

The *ethMNG* interface is generally used for managing the appliance and *ethSRV* interface is used for binding the appliances for using other services.

For example, the appliance can use the *ethMNG* interface for the ESA Web UI and the *ethSRV* interface for enabling communication with different applications in an enterprise.

The following steps describe how to select management interfaces.

► To select multiple NICs:

- If there are multiple NICs, then the following screen appears.

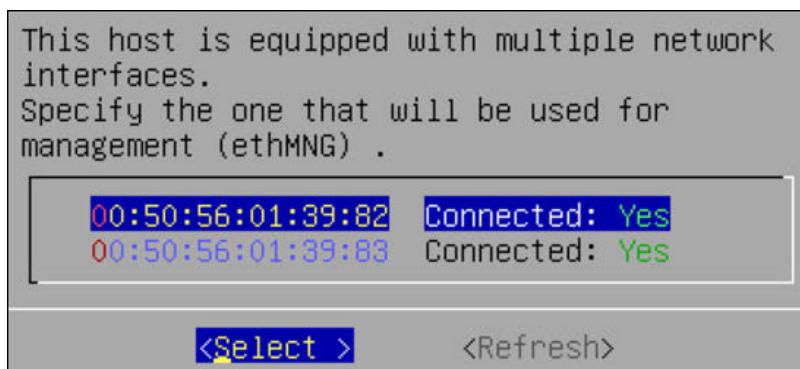


Figure 4-4: Selecting Multiple NICs

- Select the required NIC for management interface.

3. Choose **Select** and press **ENTER**.

4.2 Configuring Network Settings

After selecting the NIC for management, you configure the network for your appliance. During the network configuration, the system tries to connect to a DHCP server to obtain the hostname, default gateway, and IP addresses for the appliance. If the DHCP is not available, then you can configure the network information manually.

► To configure the network settings:

1. If the DHCP server is configured, then the following screen containing the network information appears.

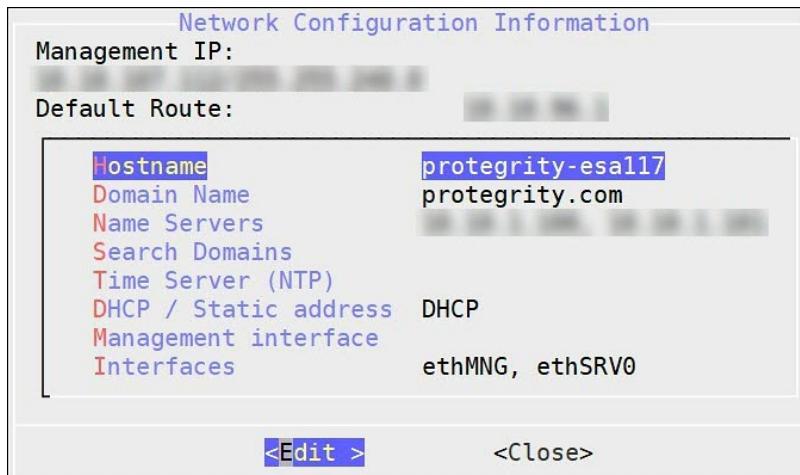


Figure 4-5: Network Configuration Information Screen

2. If the DHCP server is not available, then the following screen appears.

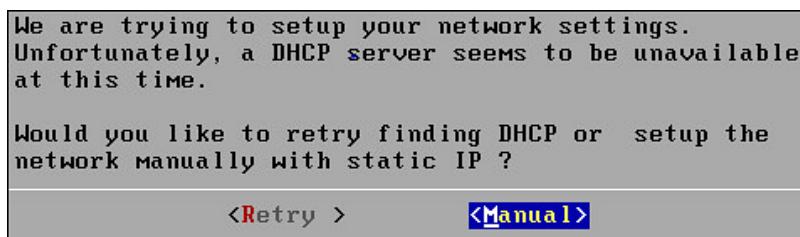


Figure 4-6: Network Entry Type Selection Screen

The Network Configuration Information screen appears.

3. Select **Manual** and press **ENTER**.

The following screen appears.

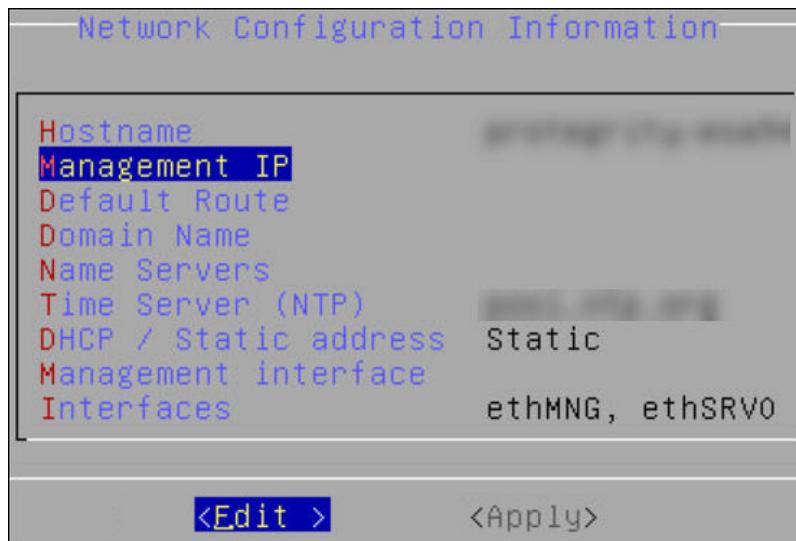


Figure 4-7: Network Configuration Information Screen

- Select **DHCP / Static address** to configure the DHCP / Static address for the appliance and choose **Edit**.
- Select **Static address** and choose **Update**.
- If you want to change the hostname of the appliance, then perform the following steps.
 - Select **Hostname** and select **Edit**.
 - Change the Hostname and select **OK**.
- Select **Management IP** to configure the management IP address for the appliance and choose **Edit**.

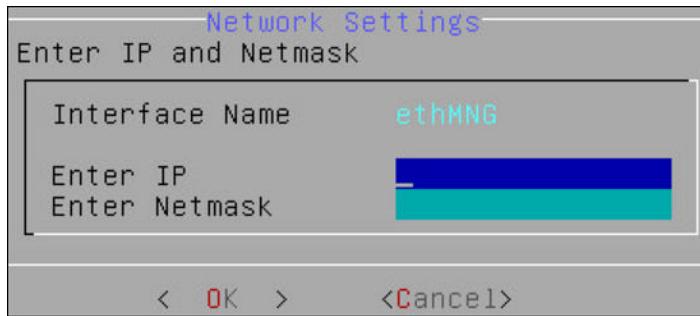


Figure 4-8: IP and Netmask Settings Screen

- Add the IP address assigned to the ethMNG interface. This IP address configures the ESA appliance to use the Web UI.
- Enter the Netmask. The ethMNG interface must be connected to the LAN with this Netmask value.
- Select **OK**.
- Select **Default Route** to configure the default route for the appliance and press **Edit**.



Figure 4-9: Default Route Setting Screen

- Enter the IP address for the default network traffic.
- Select **Apply**.

- f. Select **Domain Name** and press **Edit**.



Figure 4-10: Domain Name Setting Screen

- i. Enter the Appliance Domain Name. For example, *protegility.com*.
- ii. Press **Apply**.
- g. Select **Name Servers** and press **Edit**.



Figure 4-11: New Name Server Setting Screen

- i. Add the IP address of the name server.
- ii. Press **OK**.
- h. If you want to configure the NTP, then perform the following steps.
 - i. Select **Time Server (NTP)**, and press **Edit**.
 - ii. Add NTP time server on a TCP/IP network.
 - iii. Select **Apply**.
4. Select **Apply**.

The network settings are configured.

4.3 Configuring Time Zone

After you configure the network settings, the **Time Zone** screen appears. This section explains how to set the time zone.

► To set the Time Zone:

1. On the Time Zone screen, select the time zone.

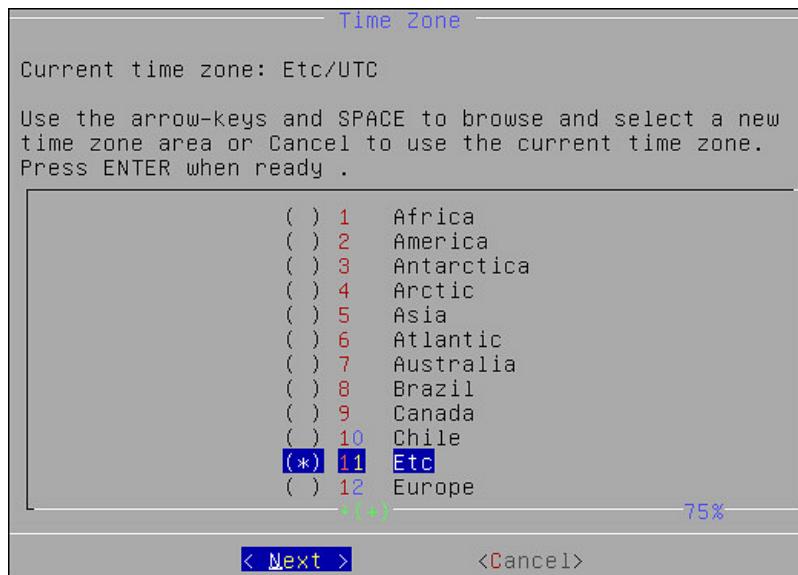


Figure 4-12: Time Zone Setting Screen

2. Press **Next**.
The time zone is configured.

4.4 Configuring the Nearest Location

After configuring the time zone, the **Nearest Location** screen appears.

► To Set the Nearest Location:

1. On the **Nearest Location** screen, enter the nearest location in GMT or UTC.

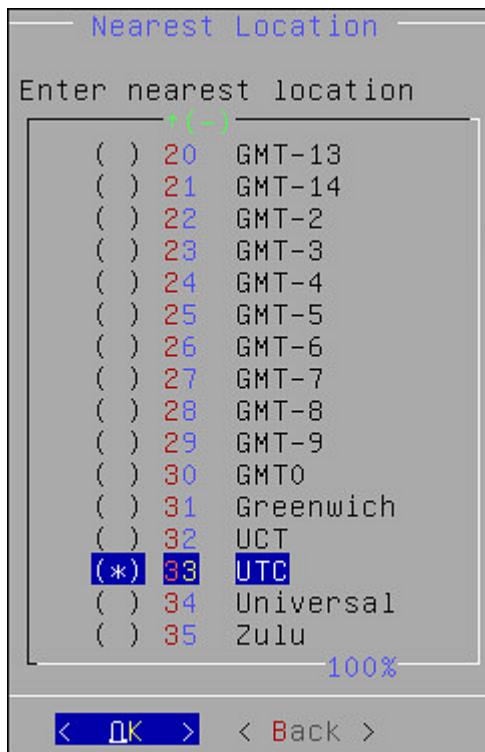


Figure 4-13: Nearest Location Setting Screen

2. Press **OK**.

The following screen appears.

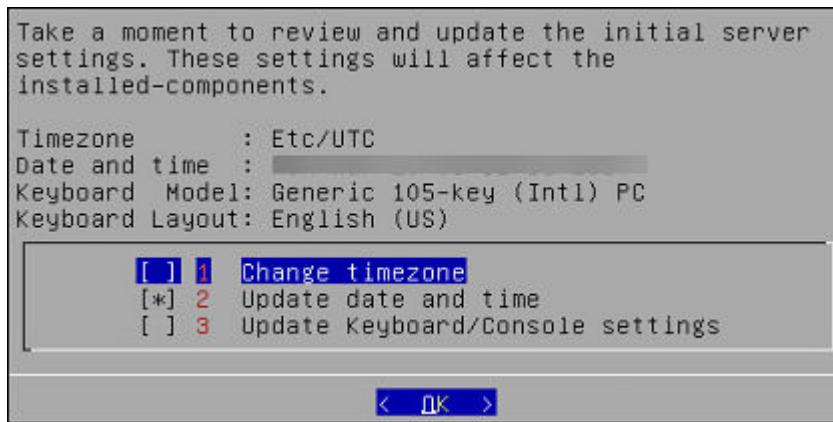


Figure 4-14: Initial Server Settings Screen

This screen also allows you to update the default settings of date and time, keyboard manufacturer, keyboard model, and keyboard layout.

4.5 Updating the Date and Time

► To Update the Date and Time:

1. Press SPACE and select **Update date and time**.
2. Press ENTER.

The following screen appears.

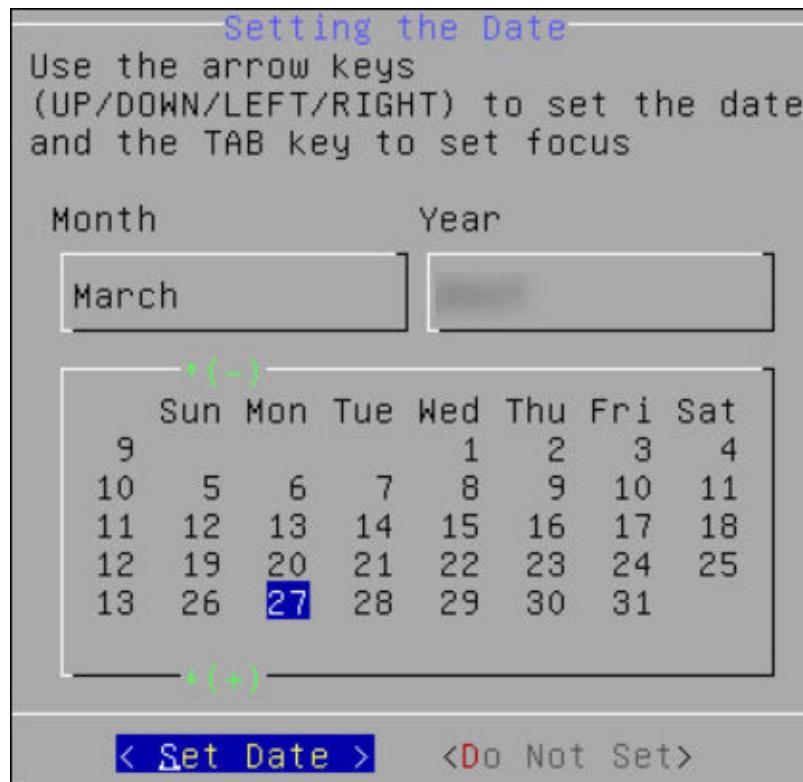


Figure 4-15: Setting the Date Screen

3. Select the date.
4. Select **Set Date** and press **ENTER**.

The next screen appears.

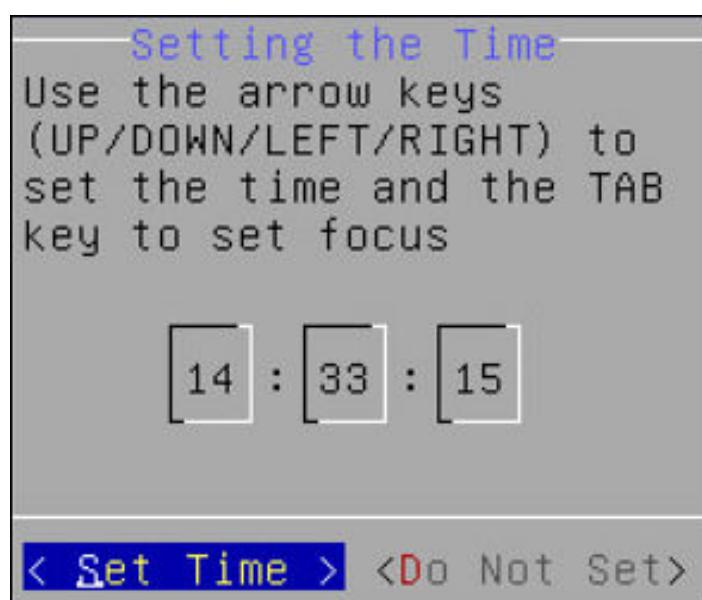


Figure 4-16: Setting the Time Screen

5. Set the time.
6. Click **Set Time** and press **ENTER**.

The date and time settings are configured.

4.6 Updating the Keyboard Settings

► To Update the Keyboard Settings:

1. Select **Update Keyboard or Console settings**.
2. Press **ENTER**.
3. Select the vendor and press the **SPACEBAR**.

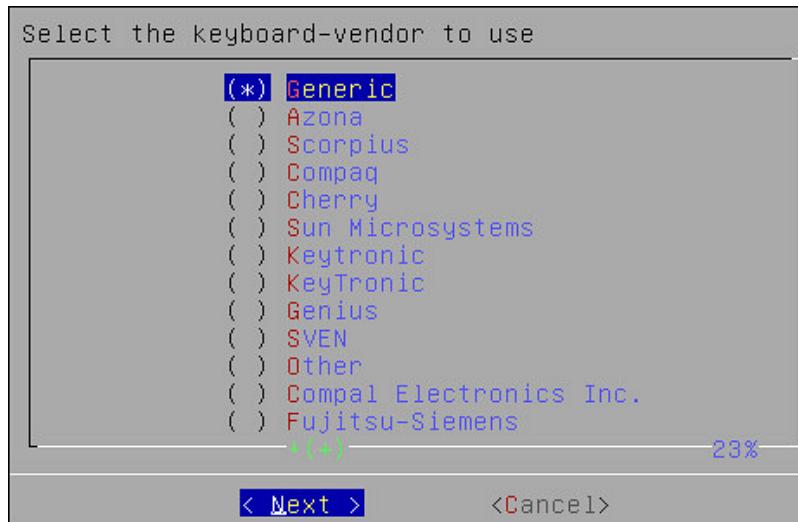


Figure 4-17: Set Keyboard Vendor Screen

4. Select **Next**.
If you select **Generic**, then a window with the list of generic keyboard models appears.
5. Select the model you use and press **Next**.

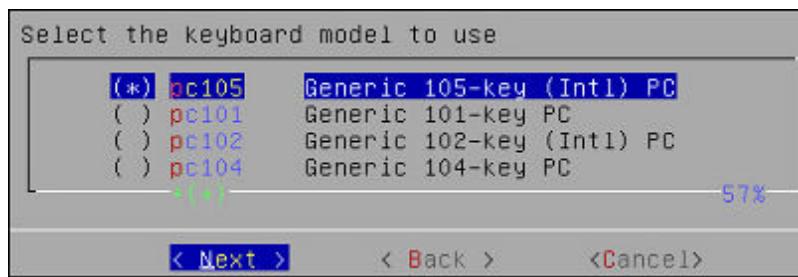


Figure 4-18: Set Keyboard Model Screen

6. On the next window, select the keyboard language. The default is **English (US)**.

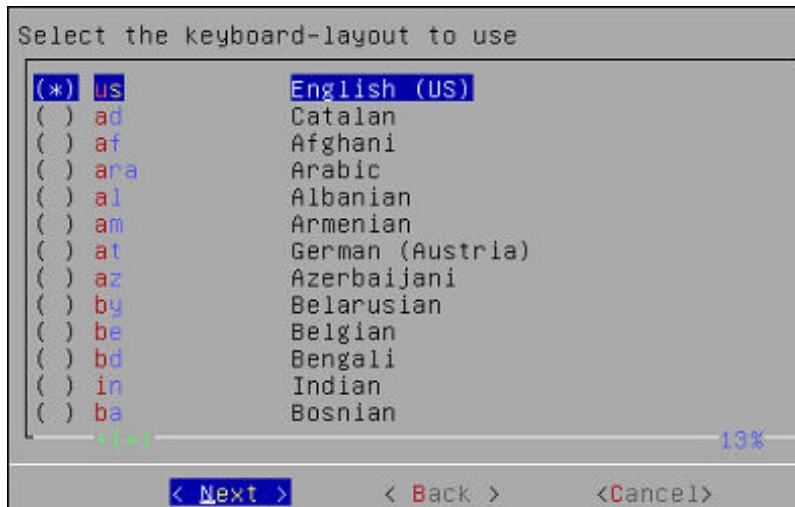


Figure 4-19: Set Keyboard Font Screen

7. Select **Next**.
8. On the next window, select the console font. The default is **Lat15-Fixed16**.

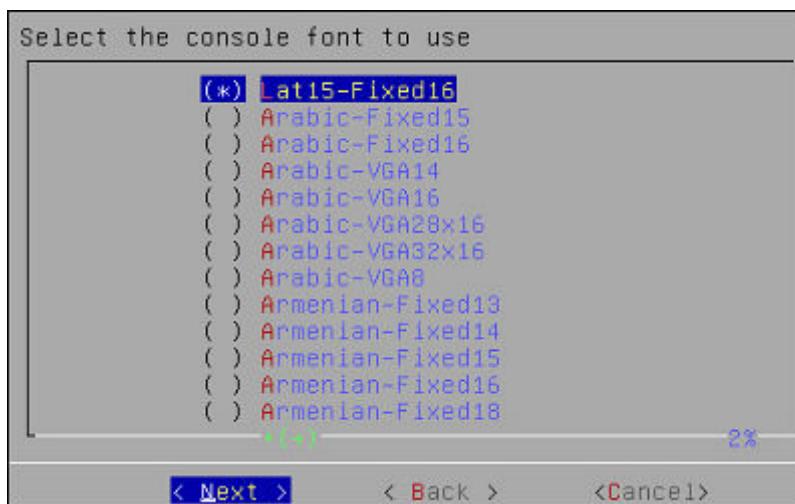


Figure 4-20: Set Console Font Screen

9. Press **Next**.
- A confirmation message appears.
10. Press **OK** to confirm.

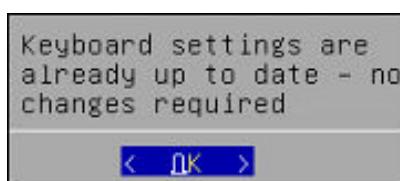


Figure 4-21: Keyboard Settings Confirmation Screen

4.7 Configuring GRUB Settings

Before you begin

On the Protegility appliances, GRUB version 2 (GRUB2) is used for loading the kernel. If you want to protect the boot configurations, then you can secure it by enforcing a username and password combination for the GRUB menu.

During installation for the ESA on-premise, a screen to configure GRUB credentials appears. If you want to protect the boot configurations, then you can secure it by enforcing a username and password combination for the GRUB menu. While installing the ESA, you can secure the GRUB menu by creating a username and setting password as described in the following task.

► To configure GRUB settings:

- From the GRUB Credentials page, press spacebar to select **Enable**.



Figure 4-22: Grub Settings

Note:

By default the **Disable** is selected. If you continue to choose **Disable**, then the security for the GRUB menu is disabled. It is recommended to enable GRUB to secure the appliance.

You can enable this feature from the CLI Manager after the installation is completed. On the CLI Manager, navigate to **Administration > GRUB Credential Settings** to enable the GRUB settings.

For more information about GRUB, refer to the section *Securing the GRand Unified Bootloader (GRUB)* in the *Protegility Appliances Overview Guide 9.1.0.5*.

- Select **OK**.

The following screen appears.

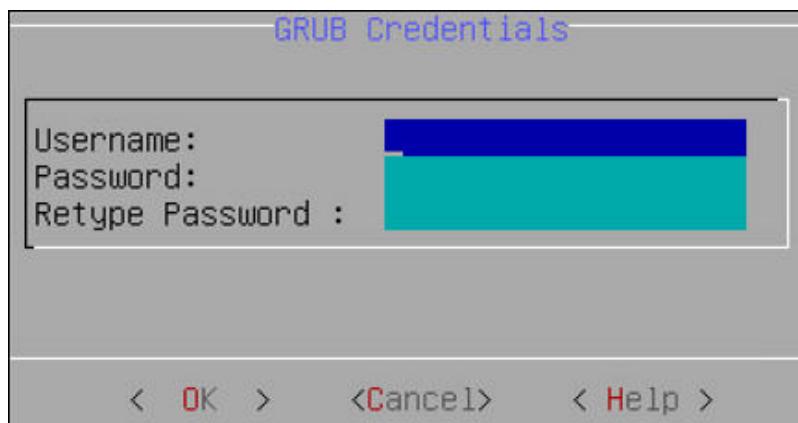


Figure 4-23: GRUB Credentials screen

- Enter a username in the **Username** text box.

Note:

The requirements for the **Username** are as follows:

- It should contain a minimum of three and maximum of 16 characters
- It should not contain numbers and special characters

4. Enter a password in the **Password** and **Re-type Password** text boxes.

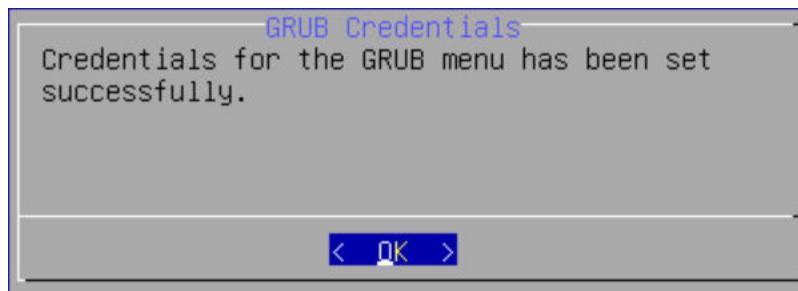
Note:

The requirements for the **Password** are as follows:

- It must contain at least eight characters
- It must contain a combination of alphabets, numbers, and printable characters

5. Select **OK** and press **ENTER**.

A message *Credentials for the GRUB menu has been set successfully* appears.



6. Select **OK**.

4.8 Setting up Users and Passwords

Only authorized users can access the appliance. The Protegility Data Security Platform defines a list of roles for each user who can access the appliance. These are system users and LDAP administrative users who have specific roles and permissions. When you install the appliance, the default users configured are as follows:

- **root**: Super user with access to all commands and files.
- **admin**: User with administrative privileges to perform all operations.
- **viewer**: User who can view, but does not have edit permissions.
- **local_admin**: Local administrator that can be used when the *admin* user is not accessible.

After completing the server settings, the **Users Passwords** screen appears that allows you set the passwords for the users.

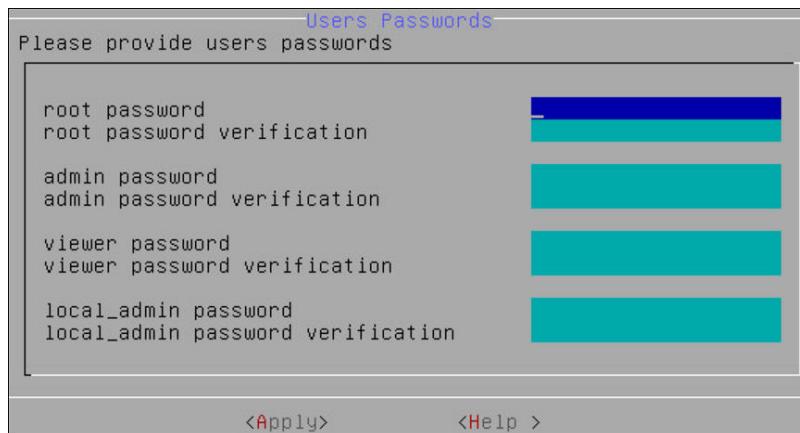


Figure 4-24: Users Passwords Screen

► To set the LDAP Users Passwords:

1. Add the passwords of the users.

Ensure that the passwords for the users comply with the password policies. For more information about the password policies, refer to the section *Password Policy Configuration* in the *Protegility Enterprise Security Administrator Guide 9.1.0.5*.

2. Select **Apply**.

The user passwords are set.

4.9 Licensing

After the appliance components are installed, the **Temporary License** screen appears. This system takes time. It is recommended to wait for few minutes before proceeding.

Note: After the ESA Appliance is installed, you must apply for a valid license within 30 days.

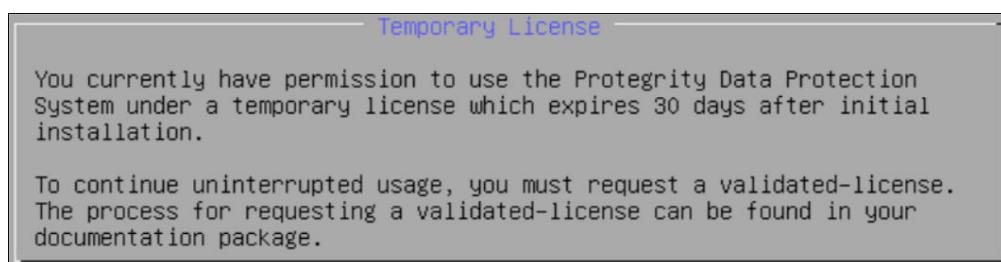


Figure 4-25: Temporary License Information Screen

For more information about licenses, refer to the *Protegility Data Security Platform Licensing Guide 9.0.0.0*.

4.10 Installing Products

In the final steps of installing the appliance, you are prompted to select the appliance components to install.

► To select products to install:

1. Press space and select the necessary products to install the following products.

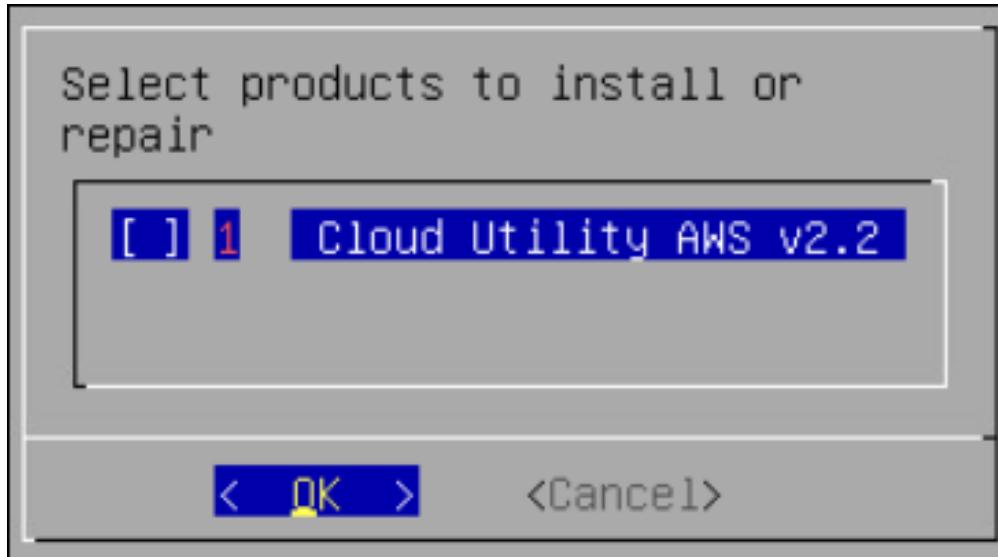


Figure 4-26: Install or Repair Products Screen

2. Click **OK**.
The selected products are installed.
3. After installation is completed, the following screen appears.

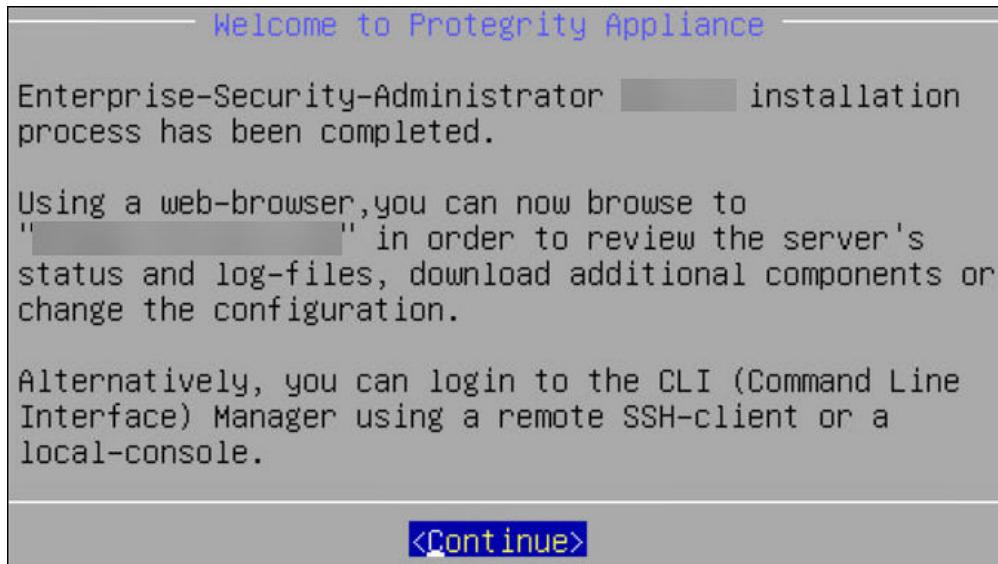


Figure 4-27: Welcome to Protegility Appliance

4. Select **Continue** to view the CLI Login screen.

Chapter 5

Installing Appliances on Cloud Platforms

[5.1 Installing Protegility Appliances on Amazon Web Services \(AWS\)](#)

[5.2 Installing Protegility Appliances on Azure](#)

[5.3 Installing Protegility Appliances on Google Cloud Platform \(GCP\)](#)

This section describes installing the appliances on Cloud platforms, such as, AWS, Azure, or GCP. For installing the appliances on cloud platforms, you must mount the image containing the Protegility appliance on a cloud instance or a virtual machine. After mounting the image, you must run the finalization procedure to install the appliance components.

The following steps must be completed to run an appliance on a cloud platform:

1. Configure the cloud instance
2. Finalize installation

5.1 Installing Protegility Appliances on Amazon Web Services (AWS)

Amazon Web Services (AWS) is a cloud-based computing service, which provides several services, such as computing power through Amazon Elastic Compute Cloud (EC2), storage through Amazon Simple Storage Service (S3), and so on.

The AWS stores Amazon Machine Images (AMIs), which are templates or virtual images containing an operating system, applications, and configuration settings.

Protegility appliances offer flexibility and can run in the following environments:

- **On-premise:** The appliance is installed and runs on dedicated hardware.
- **Virtualized:** The appliance is installed and runs on a virtual machine.
- **Cloud:** The appliance is installed and runs on or as part of a Cloud-based service.

Protegility provides AMIs that contain the appliance image, running on a customized and hardened Linux distribution.

This section describes the prerequisites and tasks for installing Protegility appliances on AWS. In addition, it describes some best practices for using the Protegility appliances on AWS effectively.

Caution:

The following features of the Protegility appliances are not available on the AWS platform:

- High Availability
- Full OS Backup/Restore

5.1.1 Verifying Prerequisites

This section describes the prerequisites including the hardware, software, and network requirements for installing and using Protegility appliances on AWS.

5.1.1.1 Prerequisites

Ensure that the following prerequisites are met:

- The Enterprise Security Administrator (ESA) appliances are installed, configured, and running.

For more information about the connection between the ESA and the Protector, refer the section [General Architecture](#).

- The IP address or host name of the ESA is noted.

- Ensure that Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to *section 4 Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.5](#).

5.1.1.2 Hardware Requirements

As the Protegility appliances are hosted and run on AWS, the hardware requirements are dependent on the configurations provided by Amazon. However, these requirements can autoscale as per customer requirements and budget.

The minimum recommendation for an appliance is 8 CPU cores and 32 GB memory. Based on this hardware requirement, select the *t3.2xlarge* configuration for creating an instance on AWS.

For more information about the hardware requirements of ESA, refer to section *System Requirements* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

Note:

The actual hardware configuration depends on the actual usage or amount of data and logs expected.

5.1.1.3 Network Requirements

The Protegility appliances on AWS are provided with an Amazon Virtual Private Cloud (VPC) networking environment. The Amazon VPC enables you to access other AWS resources, such as other instances of Protegility appliances on AWS.

You can configure the Amazon VPC by specifying its usable IP address range. You can also create and configure subnets, network gateways, and the security settings.

For more information about the Amazon VPC, refer to the Amazon VPC documentation at: http://docs.aws.amazon.com/AWSVPC/latest/UserGuide/VPC_Introduction.html.

If you are using the ESA or the DSG appliance with AWS, then ensure that the inbound and outbound ports of the appliances are configured in the Amazon Virtual Private Cloud (VPC), as described in this section, to ensure that they are able to interact with the other required components.

For more information about the list of inbound and outbound ports to be configured based on the appliance, refer to section *Open Listening Ports* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

5.1.1.3.1 Accessing the Internet

The following points list the ways in which you can provide or limit Internet access for an appliance instance in the VPC:

- If you need to connect the appliance to the Internet, then ensure that the appliance is on the default subnet so that it uses the Internet gateway that is included in the VPC.
- If you need to allow the appliance to initiate outbound connections to, and prevent inbound connections from the Internet, then ensure that you use a Network Address Translation (NAT) device.
- If you want to block the connection of the appliance to the Internet, then ensure that the appliance is on a private subnet.

5.1.1.3.2 Accessing a Corporate Network

If you need to connect the appliance to a corporate network, then ensure that you use an IPSec hardware VPN connection.

5.1.2 Obtaining the AMI

Before creating the instance on AWS, you must obtain the image from the [My.Protegility](#) portal. On the portal, you select the required ESA version and choose **AWS** as the target cloud platform. You then share the product to your cloud account. The following steps describe how to share the AMI to your cloud account.

► To obtain and share the AMI:

1. Log in to the [My.Protegility](#) portal with your user account.
2. Click **Product Management > Explore Products > Data Protection**.
3. Select the required ESA Platform Version.
The **Product Family** table will update based on the selected ESA Platform Version.

Note: The ESA Platform Versions listed in the drop-down menu reflect all versions that were either previously downloaded or shipped within the organization along with any newer versions available thereafter. You can check the list of products previously downloaded from **Product Management > My Product Inventory**.

4. Select the **Product Family**.
The description box will populate with the **Product Family** details.

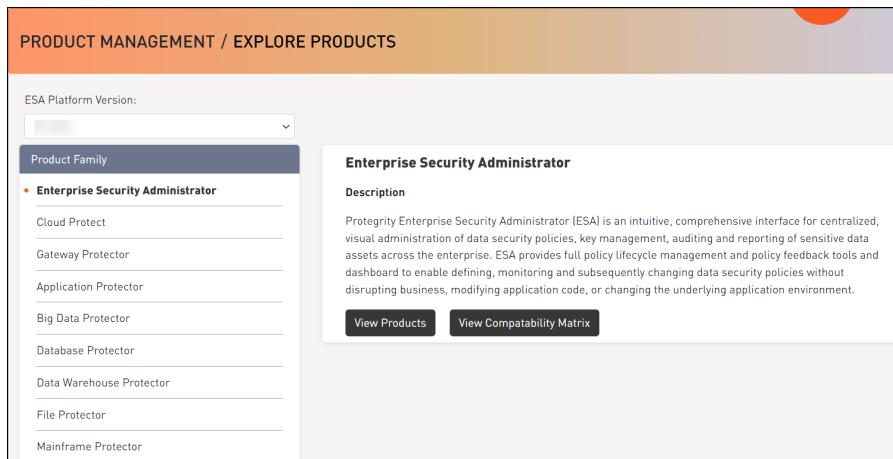


Figure 5-1: Product Family Screen

5. Click **View Products** to advance to the **Product List** screen.

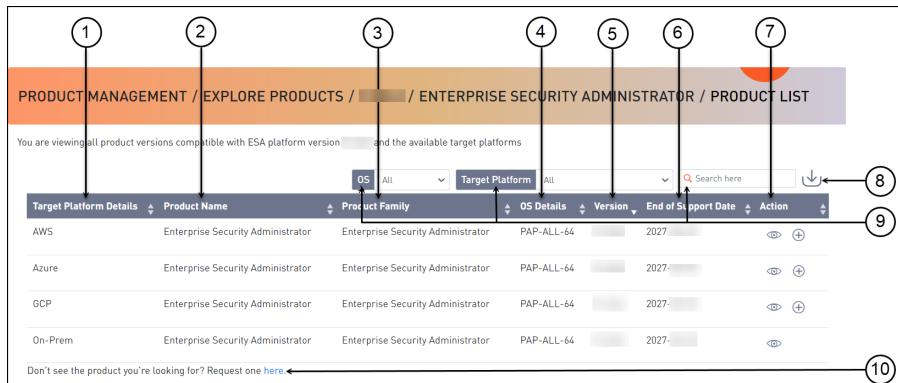


Figure 5-2: Product List Screen

Table 5-1: Product List Screen Description

| Callout | Element Name | Description |
|---------|--------------------------------|---|
| 1 | Target Platform Details | Shows details about the target platform. |
| 2 | Product Name | Shows the product name. |
| 3 | Product Family | Shows the product family name. |
| 4 | OS Details | Shows the operating system name. |
| 5 | Version | Shows the product version. |
| 6 | End of Support Date | Shows the final date that Protegility will provide support for the product. |
| 7 | Action | Click the View icon (eye icon) to open the Product Detail screen. |
| 8 | Export as CSV | Downloads a .csv file with the results displayed on the screen. |
| 9 | Search Criteria | Type text in the search field to specify the search filter criteria or filter the entries using the following options: <ul style="list-style-type: none">• OS• Target Platform |
| 10 | Request one here | Opens the Create Certification screen for a certification request. |

- Select the AWS cloud target platform you require and click the **View** icon (eye icon) from the **Action** column.
- The **Product Detail** screen appears.

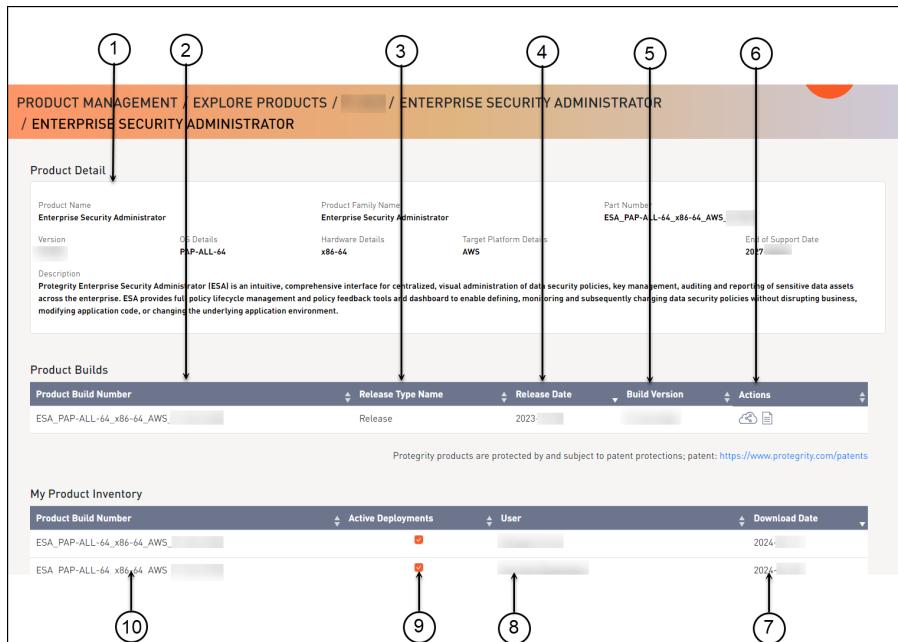


Figure 5-3: Product Detail Screen

Table 5-2: Product Details Screen Description

| Callout | Element Name | Description |
|---------|-----------------------------|---|
| 1 | Product Detail | Shows the following information about the product: <ul style="list-style-type: none"> • Product name • Family name • Part number • Version • OS details • Hardware details • Target platform details • End of support date • Description |
| 2 | Product Build Number | Shows the product build number. |
| 3 | Release Type Name | Shows the type of build, such as, release, hotfix, or patch. |
| 4 | Release Date | Shows the release date for the build. |
| 5 | Build Version | Shows the build version. |
| 6 | Actions | Shows the following options for download: <ul style="list-style-type: none"> • Click the Share Product icon () • Click the Download Signature icon () • Click the Download Readme icon () |
| 7 | Download Date | Shows the date when the file was downloaded. |
| 8 | User | Shows the user name who downloaded the build. |

| Callout | Element Name | Description |
|---------|-----------------------------|--|
| 9 | Active Deployment | Select the check box to mark the software as active. Clear the check box to mark the software as inactive. Note: This option is available only after you download a product. |
| 10 | Product Build Number | Shows the product build number. |

7. Click the **Share Product** icon () to share the desired cloud product.

Note:

If the user does not have access to cloud products or has access to cloud products and the Customer Cloud Account details are not available, then a message appears with the information that is required and the contact information for obtaining access to cloud share.

A dialog box appears and your available cloud accounts will be displayed.

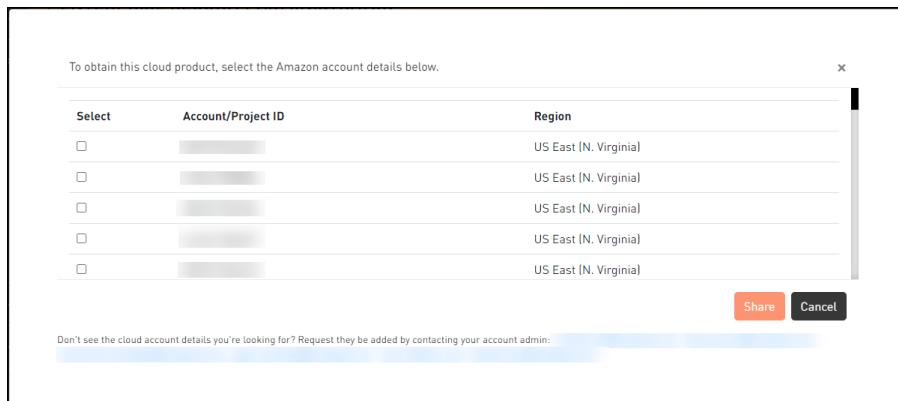


Figure 5-4: Account Selection Screen

8. Select your required cloud account in which to share the Protegility product.

9. Click **Share**.

A message box is displayed with the command line interface (CLI) instructions with the option to download a detailed PDF containing the cloud web interface instructions. Additionally, the instructions for sharing the cloud product are sent to your registered email address and to your notification inbox in [My.Protegility](#).

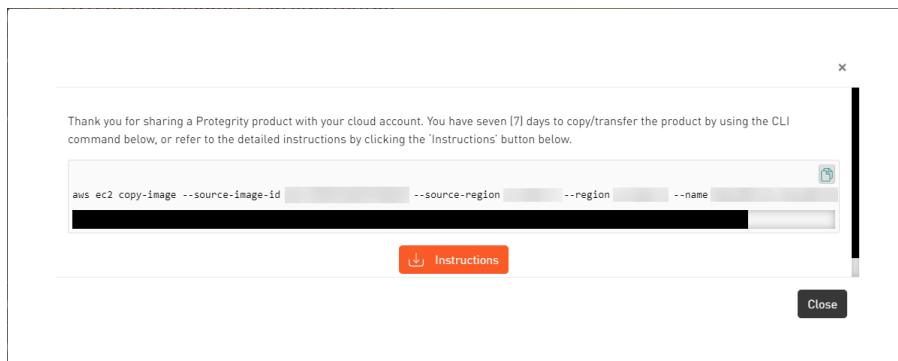


Figure 5-5: Sharing Command

10. Click the **Copy** icon () to copy the command for sharing the cloud product and run the command in CLI. Alternatively, click **Instructions** to download the detailed PDF instructions for cloud sharing using the CLI or the web interface.

Note:

The cloud sharing instruction file is saved in a `.pdf` format. You need a reader, such as, Acrobat Reader to view the file.

The Cloud Product will be shared with your cloud account for seven (7) days from the original share date in the [My.Protegility](#) portal.

After the seven (7) day time period, you need to request a new share of the cloud product through [My.Protegility](#).

5.1.3 Loading the Protegility Appliance from an Amazon Machine Image (AMI)

This section describes the tasks that need to be performed for loading the Protegility appliance from an AMI, which is provided by Protegility.

5.1.3.1 Creating an Instance of the Protegility Appliance from the AMI

Perform the following steps to create an instance of the Protegility appliance using an AMI.

► To create an instance of the Protegility appliance:

1. Access AWS at the following URL:

<https://aws.amazon.com/>

The AWS home screen appears.

2. Click the **Sign In to the Console** button.

The AWS login screen appears.

3. On the AWS login screen, enter the following details:

- Account Number
- User Name
- Password

4. Click the **Sign in** button.

After successful authentication, the AWS Management Console screen appears.

5. Click **Services**.

6. Navigate to **Compute > EC2**

The EC2 Dashboard screen appears.

7. Contact Protegility Support and provide your Amazon Account Number so that the required Protegility AMIs can be made accessible to the account.

8. Click on **AMIs** under the Images section.

The AMIs that are accessible to the user account appear in the right pane.

9. Select the AMI of the required Protegility appliance in the right pane.

10. Click the **Launch** button to launch the selected Protegility appliance.

The Choose an Instance screen appears.

11. Depending on the performance requirements, choose the required instance type.

For the ESA appliance, an instance with 32 GB RAM is recommended.

12. If you need to configure the details of the instance, then click the **Next: Configure Instance Details** button.

The Configure Instance Details screen appears.

13. Specify the following parameters on the Configure Instance Details screen:

- **Number of Instances:** The number of instances that you want to launch at a time.
- **Purchasing option:** The option to request Spot instances, which are unused EC2 instances. If you select this option, then you need to specify the maximum price that you are willing to pay for each instance on an hourly basis.

- **Network:** The VPC to launch the appliance in. If you need to create a VPC, then click the Create new VPC link. For more information about creating a VPC, refer to the section [Configuring VPC](#).
- **Subnet:** The Subnet to be used to launch the appliance. A subnet resides in one Availability zone. If you need to create a Subnet, then click the *Create new subnet* link.

For more information about creating a subnet, refer to the section [Adding a Subnet to the Virtual Private Cloud \(VPC\)](#).

- **Auto-assign Public IP:** The IP address from where your instance can be accessed over the Internet. You need to select Enable from the list.
- **Availability Zone:** A location within a region that is designed to be isolated from failures in other Availability Zones.
- **IAM role:** This option is disabled by default.
- **Shutdown behaviour:** The behaviour of the appliance when an OS-level shut down command is initiated.
- **Enable Termination Protection:** The option to prevent accidental termination of the appliance instance.
- **Monitoring:** The option to monitor, collate, and analyze the metrics for the instance of your appliance.

14. If you need to add additional storage to the instance of the appliance, then click the **Next: Add Storage** button. The Add Storage screen appears.
15. You can provision additional storage for the appliance by clicking the **Add New Volume** button. *Root* is the default volume for your instance. Alternatively, you can provision additional storage for the appliance later too.

For more information on configuring the additional storage on the instance of the appliance, refer to the section [Increasing Disk Space on the Appliance](#).

16. If you need to create a key-value pair, then click the **Next: Add Tags** button. The Add Tags screen appears. Depending on your instance requirements, add the key-value pairs.
17. If you need to configure the Security Group, then click the **Next: Configure Security Group** button. The Configure Security Group screen appears.
18. You can assign a security group from the available list. Alternatively, you can create security group with rules for the required inbound and outbound ports.
19. Click the **Review and Launch** button. The Review Instance Launch screen appears, listing all the details related to the instance of the appliance. You can review the required sections before you launch your instance.
20. Click the **Launch** button. The Key Pair dialog box appears.
21. Select the **Existing Key Pair** option and choose a key from the list of available key pairs. Alternatively, you can select the **Create a new Key Pair**, to create a new key pair.

Note:

If you proceed without a key pair, then the system will not be accessible.

22. Select the confirmation check box.
23. Click the **Launch Instances** button. The instance of the required Protegility appliance is launched and the *Launch Status* screen appears.
24. Click the **View Instances** button. The *Instances* screen appears listing the instance of the appliance.

-
25. If you need to use the instance of the appliance, then access the appliance CLI Manager using the IP address of the appliance.

5.1.3.2 Configuring the Virtual Private Cloud (VPC)

If you need to connect two Protegility appliances, or to the Internet, or a corporate network using a Private IP address, then you might need to configure the VPC.

For more information about the various inbound and outbound ports to be configured in the VPC, refer to section *Open Listening Ports* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

Perform the following steps to configure the VPC for the instance of the Protegility appliance.

► To configure the VPC for the Protegility appliance:

1. Ensure that you are logged in to AWS and at the AWS Management Console screen.
2. On the AWS Management Console, click **VPC** under the *Networking* section.
The VPC Dashboard screen appears.
3. Click on **Your VPCs** under the Virtual Private Cloud section.
The Create VPC screen appears listing all available VPCs in the right pane.
4. Click the **Create VPC** button.
The Create VPC dialog box appears.
5. Specify the following parameters on the Create VPC dialog box:
 - **Name tag:** The name of the VPC.
 - **CIDR block:** The range of the IP addresses for the VPC in $x.x.x.x/y$ form where $x.x.x.x$ is the IP address and y is the /16 and /28 netmask.
 - **Tenancy:** The tenancy parameter for the VPC, which can be set to *Default* or *Dedicated*. If the value of this parameter is set to *Default*, then it will select the tenancy attribute, which is specified at the launch of the instance of the appliance for the VPC.
6. Click the **Yes, Create** button.
The VPC is created.

5.1.3.3 Adding a Subnet to the Virtual Private Cloud (VPC)

You can add Subnets to your VPC. A subnet resides in an Availability zone. When you create a subnet, you can specify the CIDR block.

Perform the following steps to create the subnet for your VPC.

► To create a Subnet:

1. Ensure that you are logged in to AWS and at the AWS Management Console screen.
2. On the AWS Management Console, click VPC under the Networking section.
The VPC Dashboard screen appears.
3. Click **Subnets** under the Virtual Private Cloud section.
The create subnet screen appears listing all available subnets in the right pane.

- Click the **Create Subnet** button.

The Create Subnet dialog box appears.

- Specify the following parameters on the Create Subnet dialog box.

- Name tag:** The name for the Subnet.
- VPC:** The VPC for which you want to create a subnet.
- Availability Zone:** The Availability zone where the subnet resides.
- CIDR block:** The range of the IP addresses for the VPC in x.x.x.x/y form where x.x.x.x is the IP address and y is the /16 and /28 netmask.

- Click the **Yes, Create** button.

The Subnet is created.

5.1.3.4 Finalizing the Installation of Protegility Appliance on the Instance

When you install the appliance, it generates multiple security identifiers such as, keys, certificates, secrets, passwords, and so on. These identifiers ensure that sensitive data is unique between two appliances in a network. When you receive a Protegility appliance image, the identifiers are generated with certain values. If you use the security identifiers without changing their values, then security is compromised and the system might be vulnerable to attacks. Using the **Rotate Appliance OS Keys**, you can randomize the values of these security identifiers for an appliance. During the finalization process, you run the key rotation tool to secure your appliance.

Note:

If you do not complete the finalization process, then some features of the appliance may not be functional including the Web UI.

For example, if the OS keys are not rotated, then you might not be able to add appliances to a Trusted Appliances Cluster (TAC).

Note:

For information about the default passwords, refer to the section *Launching the ESA instance on Amazon Web Services* in the *Release Notes*.

5.1.3.4.1 Logging in and Finalising the AWS Instance using the SSH Client

After installing the Protegility Appliance on AWS, you must log in to the AWS instance using the SSH Client.

► To login to the AWS instance using the SSH Client:

- Start the local SSH Client.
- Perform the SSH operation on the AWS instance using the key pair utilizing the following command.

```
ssh -i <path of the private key pair> local_admin@<IP address of the AWS instance>
```

Note:

Ensure that you use the *local_admin* user to perform the SSH operation.

- Press **Enter**.



The following screen appears.

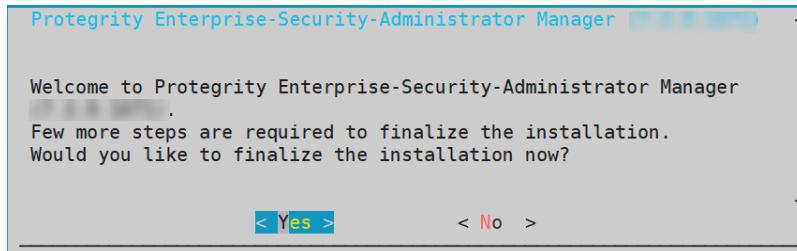


Figure 5-6: Finalizing Installation Confirmation screen

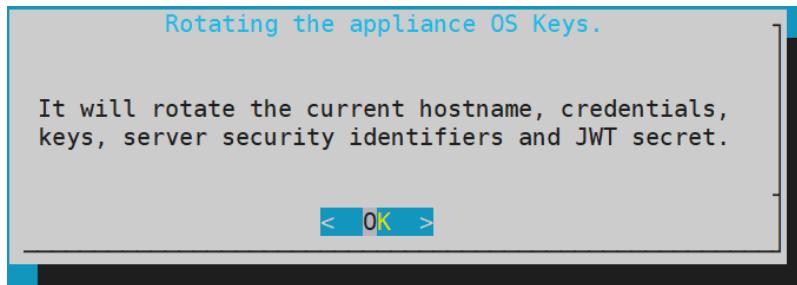
Caution:

Ensure that the finalization process is initiated from a single session only. If you start finalization simultaneously from a different session, then the "*Finalization is already in progress.*" message appears. You must wait until the finalization of the instance is successfully completed.

Additionally, ensure that the appliance session is not interrupted. If the session is interrupted, then the instance becomes unstable and the finalization process is not completed on that instance.

4. Select **Yes** to initiate the finalization process.

A confirmation screen to rotate the appliance OS keys appears.



Note:

If you select **No** in the **Finalizing Installation Confirmation** screen in Step 3, then the finalization process is not initiated.

To manually initiate the finalization process, navigate to **Tools > Finalize Installation** and press **ENTER**.

5. Select **OK** to rotate the appliance OS keys.

The following screen appears.

User's Passwords
Please provide user's passwords

| | |
|-----------------------------------|--|
| root password | |
| root password verification | |
| admin password | |
| admin password verification | |
| viewer password | |
| viewer password verification | |
| local_admin password | |
| local_admin password verification | |

<Apply> <Help >

- To update the user passwords, provide the credentials for the following users:
 - root
 - admin
 - viewer
 - local_admin
- Select **Apply**.

The user passwords are updated and the appliance OS keys are rotated.

The finalization process is completed.

Note: The *SSH Authentication Type* by default, is set to *Publickey*. Ensure that you use the *Password + Publickey* for accessing the CLI. You can change the authentication type from the ESA Web UI, once the finalization is completed.

Note:

The appliance comes with some products installed by default. If you want to verify the installed products or install additional products, then navigate to **Administration > -- Installations and Patches -- > Add/Remove Services**.

For more information about installing products, refer to section [Installing Products](#).

5.1.3.5 Connecting to an ESA instance (for DSG deployment)

If you are using an instance of the DSG appliance, then you need to provide the connectivity details related to an instance of the ESA appliance in the DSG appliance using the CLI Manager.

Note:

Ensure that you run the *Appliance-rotation-tool* on the ESA before you setup the communication of the DSG appliance with the ESA appliance.

For more information about running the Appliance-rotation-tool on the ESA, refer to the section *Running the Appliance-Rotation-Tool* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

For more information about connecting to an instance of the ESA appliance, refer to the section *Setting up ESA Communication* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

5.1.3.5.1 Deploying the Instance of the Protegility Appliance with the Protectors

You can configure the various protectors that are a part of the Protegility Data Security Platform with the instance of the ESA appliance running on AWS.

Depending on the Cloud-based environment which hosts the protectors, the protectors can be configured with the instance of the ESA appliance in one of the following ways:

- If the protectors are running on the same VPC as the instance of the ESA appliance, then the protectors need to be configured using the internal IP address of the appliance within the VPC.
- If the protectors are running on a different VPC than that of the instance of the ESA appliance, then the VPC of the instance of the ESA needs to be configured to connect to the VPC of the protectors.

5.1.4 Increasing Disk Space on the Appliance

After an instance of the appliance is created, you can increase the disk space on the appliance.

Perform the following steps to increase the disk space for the appliance on AWS.

► To increase disk space for the Appliance on AWS:

1. On the EC2 Dashboard screen, click **Volumes** under the *Elastic Block Store* section.
The Create Volume screen appears.
2. Click the **Create Volume** button.
The **Create Volume** dialog box appears.
3. Enter the required size of the additional disk space in the **Size (GiB)** text box.
4. Enter the snapshot ID of the instance, for which the additional disk space is required in the **Snapshot ID** text box.
5. Click the **Create** button.
The required additional disk space is created as a volume.
6. Right-click on the additional disk, which is created.
The pop-up menu appears.
7. Select **Attach Volume**.
The **Attach Volume** dialog box appears.
8. Enter the Instance ID or name tag of the appliance to add the disk space in the **Instance** text box.
9. Click the **Attach** button to add the disk space to the required appliance instance.
The disk space is added to the appliance instance.
10. After the disk space on the appliance instance is added, navigate to Instances under the *Instances* section.
11. Right-click on the appliance instance in which the disk space was added.
The pop-up menu appears.
12. Select **Instance State > Stop**.
The appliance instance is stopped.
13. Select **Instance State > Start**.
The appliance instance is started.



14. After the appliance instance is started, configure the additional storage on the appliance using the CLI Manager on the appliance.
- For more information on configuring the additional storage on the appliance, refer to section *Installation of Additional Hard Disks* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

5.2 Installing Protegility Appliances on Azure

Azure is a cloud computing service offered by Microsoft, which provides services for compute, storage, and networking. It also provides software, platform, and infrastructure services along with support for different programming languages, tools, and frameworks.

The Azure cloud platform includes the following components:

- **Resource groups:** Resource groups in Azure are a collection of multiple Azure resources, such as virtual machines, storage accounts, virtual networks, and so on. The resource groups enable you to manage and maintain the resources as a single entity.
- **Storage accounts:** Azure storage accounts contain all the Azure storage data objects, such as disks, blobs, files, queues, and tables. The data in the storage accounts are scalable, secure, and highly available.
- **BLOB storage:** The BLOB storage is used to store unstructured data on the Azure cloud platform.

5.2.1 Prerequisites

This section describes the prerequisites, including the hardware and network requirements, for installing and using Protegility appliances on Azure.

The following prerequisites are essential to install the Protegility appliances on Azure:

- Sign in URL for the Azure account
- Authentication credentials for the Azure account
- Working knowledge of Azure
- Access to the [My.Protegility](#) portal

Before you begin:

Ensure that you use the following order to create a virtual machine on Azure:

| Order | Description |
|-------|---|
| 1 | Create a Resource Group |
| 2 | Create a Storage Account |
| 3 | Create a Container |
| 4 | Obtain the Azure BLOB |
| 5 | Create an image from the BLOB |
| 6 | Create a VM from the image |

5.2.1.1 Hardware Requirements

As the Protegility appliances are hosted and run on Azure, the hardware requirements are dependent on the configurations provided by Microsoft. However, these requirements can change based on the customer requirements and budget.

The minimum recommendation for an appliance is 8 CPU cores and 32 GB memory. Based on this hardware requirement, select the required configuration for creating an instance on Azure.

For more information about the hardware requirements of ESA, refer to section *System Requirements* in the *Protegility Appliances Overview Guide 9.1.0.5*.

Note:

The actual hardware configuration depends on the actual usage or amount of data and logs expected.

5.2.1.2 Network Requirements

The Protegility appliances on Azure are provided with an Azure virtual networking environment. The virtual network enables you to access other instances of Protegility resources in your project.

For more information about configuring Azure virtual network, refer to section [Setting up Azure Virtual Network](#).

5.2.2 Azure Cloud Utility

The Azure Cloud Utility is an appliance component that is available for supporting features specific to Azure Cloud Platform. For Protegility appliances, this component must be installed to utilize the services of Azure Accelerated Networking and Azure Linux VM agent. When you upgrade or install the appliance from an Azure v9.0.0 blob, the **Azure Cloud Utility** is installed automatically in the appliance.

Note:

If you are utilizing the Azure Accelerated Networking or Azure Linux VM agent, then it is recommended to not uninstall this component.

5.2.3 Setting up Azure Virtual Network

The Azure virtual network is a service that provides connectivity to the virtual machine and services on Azure. You can configure the Azure virtual network by specifying the usable IP addresses. You can also create and configure subnets, network gateways, and security settings.

For more information about setting up Azure virtual network, refer to the Azure virtual network documentation at:

<https://docs.microsoft.com/en-us/azure/virtual-network/virtual-networks-overview>

If you are using the ESA or the DSG appliance with Azure, ensure that the inbound and outbound ports of the appliances are configured in the virtual network.

For more information about the list of inbound and outbound ports, refer to section *Open Listening Ports* in the *Protegility Appliances Overview Guide 9.1.0.5*.

5.2.4 Creating a Resource Group

Resource Groups in Azure are a collection of multiple Azure resources, such as virtual machines, storage accounts, virtual networks, and so on. The resource groups enable to manage and maintain the resources as a single entity.

For more information about creating resource groups, refer to the Azure resource group documentation at,

<https://docs.microsoft.com/en-us/azure/azure-resource-manager/resource-group-portal>



5.2.5 Creating a Storage Account

Azure storage accounts contain all the Azure storage data objects, such as disks, blobs, files, queues, and tables. The data in the storage accounts are scalable, secure, and highly available.

For more information about creating storage accounts, refer to the Azure storage accounts documentation at,

<https://docs.microsoft.com/en-us/azure/storage/common/storage-quickstart-create-account>

5.2.6 Creating a Container

The data storage objects in a storage account are stored in a container. Similar to directories in a file system, the container in Azure contain BLOBS. You add a container in Azure to store the ESA BLOB.

For more information about creating a container, refer to the following link.

<https://docs.microsoft.com/en-us/azure/storage/blobs/storage-quickstart-blobs-portal>

5.2.7 Obtaining the Azure BLOB

In Azure, you can share files across different storage accounts. The ESA that is packaged as a BLOB, is shared across storage accounts on Azure. A BLOB is a data type that is used to store unstructured file formats. Azure supports BLOB storage to store unstructured data, such as audio, text, images, and so on. The BLOB of the appliance is shared by Protegility to the client's storage account.

Before creating the instance on Azure, you must obtain the BLOB from the [My.Protegility](#) portal. On the portal, you select the required ESA version and choose **Azure** as the target cloud platform. You then share the product to your cloud account. The following steps describe how to share the BLOB to your cloud account.

► To obtain and share the BLOB:

1. Log in to the [My.Protegility](#) portal with your user account.
2. Click **Product Management > Explore Products > Data Protection**.
3. Select the required ESA Platform Version.
The **Product Family** table will update based on the selected ESA Platform Version.

Note: The ESA Platform Versions listed in the drop-down menu reflect all versions that were either previously downloaded or shipped within the organization along with any newer versions available thereafter. You can check the list of products previously downloaded from **Product Management > My Product Inventory**.

4. Select the **Product Family**.
The description box will populate with the **Product Family** details.

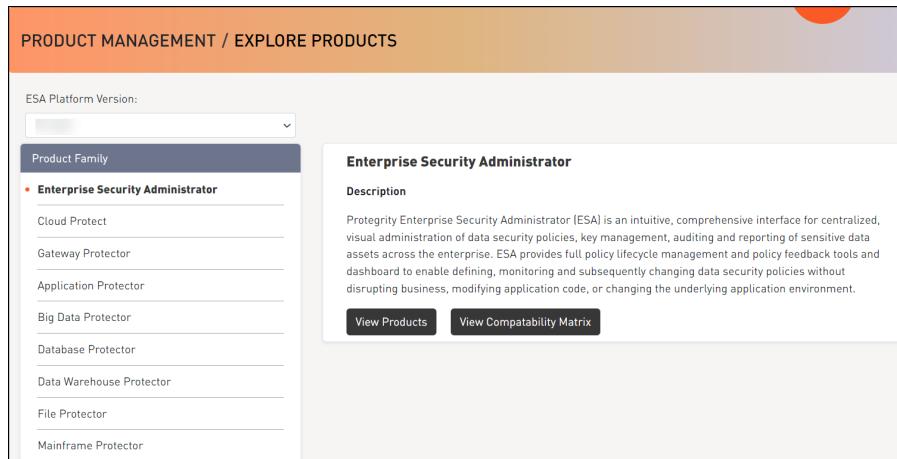


Figure 5-7: Product Family Screen

- Click **View Products** to advance to the **Product List** screen.

Figure 5-8: Product List Screen

Table 5-3: Product List Screen Description

| Callout | Element Name | Description |
|---------|--------------------------------|---|
| 1 | Target Platform Details | Shows details about the target platform. |
| 2 | Product Name | Shows the product name. |
| 3 | Product Family | Shows the product family name. |
| 4 | OS Details | Shows the operating system name. |
| 5 | Version | Shows the product version. |
| 6 | End of Support Date | Shows the final date that Protegility will provide support for the product. |
| 7 | Action | Click the View icon (ⓘ) to open the Product Detail screen. |
| 8 | Export as CSV | Downloads a .csv file with the results displayed on the screen. |
| 9 | Search Criteria | Type text in the search field to specify the search filter criteria or filter the entries using the following options: • OS • Target Platform |
| 10 | Request one here | Opens the Create Certification screen for a certification request. |

- Select the **Azure** cloud target platform you require and click the **View** icon (ⓘ) from the **Action** column. The **Product Detail** screen appears.

- Click the **Share Product** icon () to share the Azure cloud product.

Note:

If the user does not have access to cloud products or has access to cloud products and the Customer Cloud Account details are not available, then a message appears with the information that is required and the contact information for obtaining access to cloud share.

A dialog box appears and your available cloud accounts will be displayed.

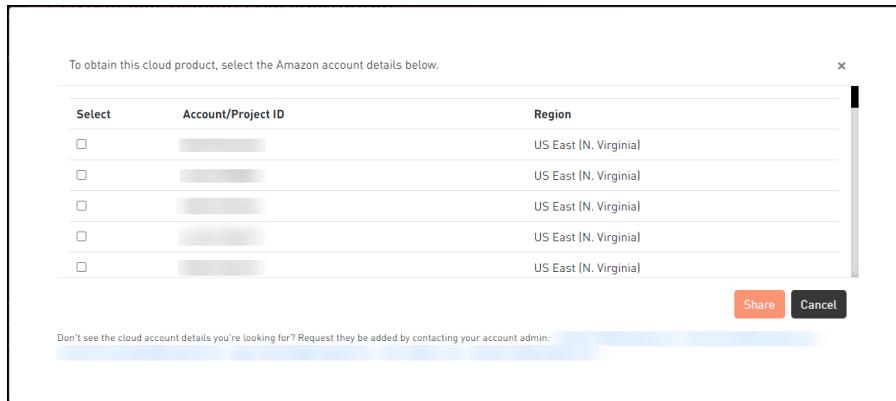


Figure 5-9: Account Selection Screen

- Select your required cloud account in which to share the Protegility product.

- Click **Share**.

A message box is displayed with the command line interface (CLI) instructions with the option to download a detailed PDF containing cloud web interface instructions. Additionally, the instructions for sharing the cloud product are sent to your registered email address and to your notification inbox in [My.Protegility](#).

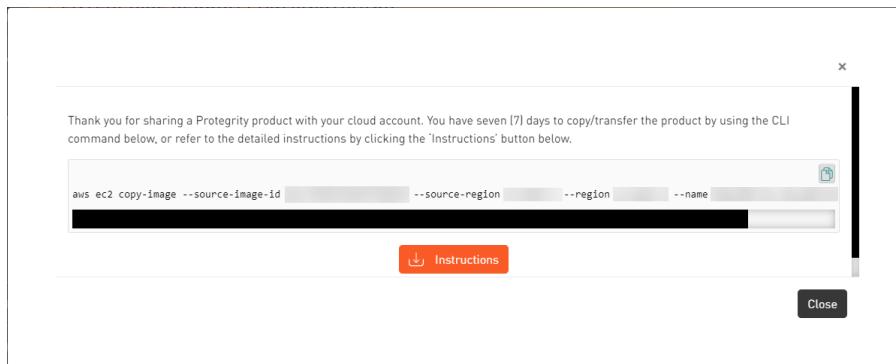


Figure 5-10: Sharing Command

- Click the **Copy** icon () to copy the command for sharing the cloud product and run the command in CLI. Alternatively, click **Instructions** to download the detailed PDF instructions for cloud sharing using the CLI or the web interface.

Note:

The cloud sharing instruction file is saved in a `.pdf` format. You need a reader, such as, Acrobat Reader to view the file.

The Cloud Product will be shared with your cloud account for seven (7) days from the original share date in the [My.Protegility](#) portal.

After the seven (7) day time period, you need to request a new share of the cloud product through [My.Protegility](#).

5.2.8 Creating Image from the Azure BLOB

After you obtain the BLOB from Protegility, you must create an image from the BLOB. The following steps describe the parameters that must be selected to create an image.

► To create an image from the BLOB:

1. Log in to the Azure portal.
2. Select **Images** and click **Create**.
3. Enter the details in the **Resource Group**, **Name**, and **Region** text boxes.
4. In the **OS disk** option, select **Linux**.
5. In the **VM generation** option, select **Gen 1**.
6. In the **Storage blob** drop-down list, select the Protegility Azure BLOB.
7. Enter the appropriate information in the required fields and click **Review + create**.
The image is created from the BLOB.

5.2.9 Creating a VM from the Image

After obtaining the image, you can create a VM from it. For more information about creating a VM from the image, refer to the following link.

<https://docs.microsoft.com/en-us/azure/virtual-machines/linux/quick-create-portal#create-virtual-machine>

► To create a VM:

1. Login in to the Azure homepage.
2. Click **Images**.
The list of all the images appear.
3. Select the required image.
4. Click **Create VM**.
5. Enter details in the required fields.
6. Select **SSH public key** in the **Authentication type** option.

Note:

As a security measure for the appliances, it is recommended to not use the **Password based mechanism** as an authentication type.

7. In the **Username** text box, enter the name of a user.

Note:

This user is added as an OS level user in the appliance. Ensure that the following usernames are not provided in the **Username** text box:

- Appliance OS users
- Appliance LDAP users

For more information about Appliance OS users and Appliance LDAP users, refer to sections *OS Users in Appliances* and *Managing Users* in the [Protegility Appliances Overview Guide 9.1.0.5](#)

8. Select the required SSH public key source.
9. Enter the required information in the *Disks, Networking, Management*, and *Tags* sections.
10. Click **Review + Create**.
The VM is created from the image.
11. After the VM is created, you can access the appliance from the CLI Manager or Web UI.

Note:

The OS user that is created in step 7 does not have SSH access to the appliance. If you want to provide SSH access to this user, login to the appliance as another administrative user and toggle SSH access. In addition, update the user to permit Linux shell access (`/bin/sh`).

For more information about toggle SSH access and Linux shell access, refer the section *Managing Local OS Users* in the [Protegility Appliances Overview Guide 9.1.0.5](#)

5.2.10 Accessing the Appliance

After setting up the virtual machine, you can access the appliance through the IP address that is assigned to the virtual machine. It is recommended to access the appliance with the administrative credentials.

Note:

If the number of unsuccessful password attempts exceed the defined value in the password policy, then the account gets locked.

For more information on the password policy for the admin and viewer users, refer the section *Password Policy for All Appliance Services*, and for the *root* and *local_admin* OS users, refer the section *Managing Local OS Users* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

5.2.11 Finalizing the Installation of Protegility Appliance on the Instance

When you install the appliance, it generates multiple security identifiers such as, keys, certificates, secrets, passwords, and so on. These identifiers ensure that sensitive data is unique between two appliances in a network. When you receive a Protegility appliance image, the identifiers are generated with certain values. If you use the security identifiers without changing their values, then security is compromised and the system might be vulnerable to attacks. Using the **Rotate Appliance OS Keys**, you can randomize the values of these security identifiers for an appliance. During the finalization process, you run the key rotation tool to secure your appliance.

Note:

If you do not complete the finalization process, then some features of the appliance may not be functional including the Web UI.

For example, if the OS keys are not rotated, then you might not be able to add appliances to a Trusted Appliances Cluster (TAC).

Note:



For information about the default passwords, refer to the section *Launching the ESA instance on Amazon Web Services* in the [Release Notes](#).

5.2.11.1 Finalizing ESA Installation

You can finalize the installation of the ESA after signing in to the CLI Manager.

Caution:

Ensure that the finalization process is initiated from a single session only. If you start finalization simultaneously from a different session, then the "*Finalization is already in progress.*" message appears. You must wait until the finalization of the instance is successfully completed.

Additionally, ensure that the appliance session is not interrupted. If the session is interrupted, then the instance becomes unstable and the finalization process is not completed on that instance.

► To finalize ESA installation:

1. Sign in to the ESA CLI Manager of the instance created using the default administrator credentials.
The following screen appears.

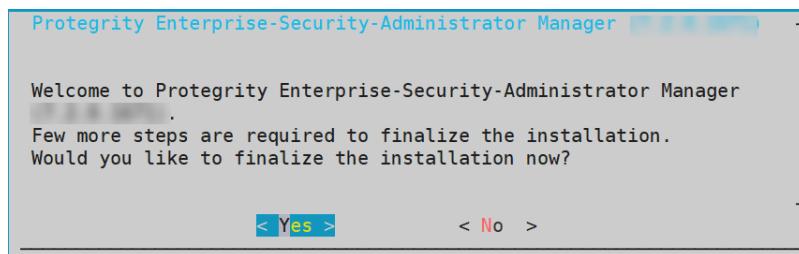


Figure 5-11: Finalizing Installation Confirmation screen

For more information about default credentials, refer the section *Creating an Instance of the Protegility Appliance from the AMI* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

2. Select **Yes** to initiate the finalization process.

The screen to enter the administrative credentials appears.

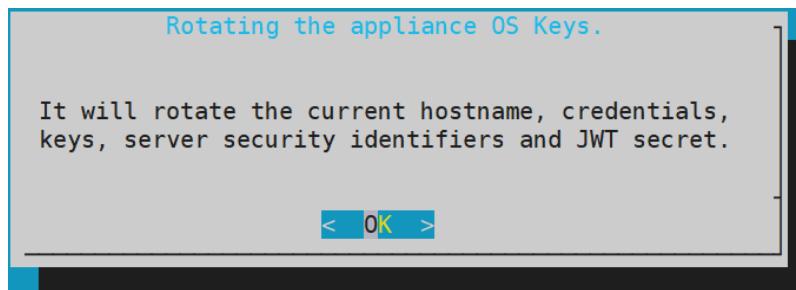
Note:

If you select **No**, then the finalization process is not initiated.

To manually initiate the finalization process, navigate to **Tools > Finalize Installation** and press **ENTER**.

3. Enter the credentials for the *admin* user and select **OK**.

A confirmation screen to rotate the appliance OS keys appears.



4. Select **OK** to rotate the appliance OS keys.

The following screen appears.

User's Passwords
Please provide user's passwords

| | |
|-----------------------------------|------------|
| root password | [Redacted] |
| root password verification | [Redacted] |
| admin password | [Redacted] |
| admin password verification | [Redacted] |
| viewer password | [Redacted] |
| viewer password verification | [Redacted] |
| local_admin password | [Redacted] |
| local_admin password verification | [Redacted] |

<Apply> <Help >

- To update the user passwords, provide the credentials for the following users:
 - root
 - admin
 - viewer
 - local_admin
- Select **Apply**.

The user passwords are updated and the appliance OS keys are rotated.

The finalization process is completed.

Note:

If you want to verify the installed products or install additional products, then navigate to **Administration > -- Installations and Patches -- > Add/Remove Services**.

For more information about installing products, refer to the section [Installing Products](#).

For more information about rotating the OS keys, refer to section *Rotate Appliance OS Keys* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

After rotating appliance keys, rotate the Audit Store certificates using the steps from *Rotating Audit Store Certificates* in the [Audit Store Guide 9.1.0.5](#).

5.2.12 Connecting to an ESA Instance

If you are using an instance of the DSG appliance on Azure, you must connect it to an instance of the ESA appliance. Using the CLI manager, you must provide the connectivity details of the ESA appliance in the DSG appliance.

For more information about connecting a DSG instance with ESA, refer to the section *Setting up ESA Communication* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

5.3 Installing Protegility Appliances on Google Cloud Platform (GCP)

The Google Cloud Platform (GCP) is a cloud computing service offered by Google, which provides services for compute, storage, networking, cloud management, security, and so on. The following products are available on GCP:

- **Google Compute Engine** provides virtual machines for instances
- **Google App Engine** provides a Software Developer Kit (SDK) to develop products
- **Google Cloud Storage** is a storage platform to store large data sets
- **Google Container Engine** is a cluster-oriented container to develop and manage Docker containers

Protegility provides the images for GCP that contain either the Enterprise Security Administrator (ESA), or the Data Security Gateway (DSG).

This section describes the prerequisites and tasks for installing Protegility appliances on GCP. In addition, it describes some best practices for using the Protegility appliances on GCP effectively.

5.3.1 Verifying Prerequisites

This section describes the prerequisites including the hardware, software, and network requirements for installing and using Protegility appliances on GCP.

5.3.1.1 Prerequisites

The following prerequisite is essential to install the Protegility appliances on GCP:

- A GCP account and the following information:
 - Login URL for the GCP account
 - Authentication credentials for the GCP account
 - Access to the [My.Protegility](#) portal

5.3.1.2 Hardware Requirements

As the Protegility appliances are hosted and run on GCP, the hardware requirements are dependent on the configurations provided by GCP. However, these requirements can autoscale as per customer requirements and budget.

The minimum recommendation for an appliance is 8 CPU cores and 32 GB memory. Based on this hardware requirement, select the *e2-standard-8* configuration for creating an instance on GCP.

For more information about the hardware requirements of ESA, refer the section *System Requirements* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

Note:

The actual hardware configuration depends on the actual usage or amount of data and logs expected.

5.3.1.3 Network Requirements

The Protegility appliances on GCP are provided with a Google Virtual Private Cloud (VPC) networking environment. The Google VPC enables you to access other instances of Protegility resources in your project.

You can configure the Google VPC by specifying the IP address range. You can also create and configure subnets, network gateways, and the security settings.

For more information about the Google VPC, refer to the VPC documentation at: <https://cloud.google.com/vpc/docs/vpc>

If you are using the ESA or the DSG appliance with GCP, then ensure that the inbound and outbound ports of the appliances are configured in the VPC.

For more information about the list of inbound and outbound ports, refer to section *Open Listening Ports* in the *Protegility Appliances Overview Guide 9.1.0.5*.

5.3.2 Configuring the Virtual Private Cloud (VPC)

You must configure your Virtual Private Cloud (VPC) to connect to different Protegility appliances.

► To configure a VPC:

1. Ensure that you are logged in to the GCP Console.
2. Navigate to the **Home** screen.
3. Click the navigation menu on the Home screen.
4. Under Networking, navigate to **VPC network > VPC networks**.
The *VPC networks* screen appears.
5. Click **CREATE VPC NETWORK**.
The *Create a VPC network* screen appears.
6. Enter the name and description of the VPC network in the **Name** and **Description** text boxes.
7. Under the Subnets area, Click **Custom** to add a subnet.
 - a. Enter the name of the subnet in the **Name** text box.
 - b. Click **Add a Description** to enter a description for the subnet.
 - c. Select the region where the subnet is placed from the **Region** drop-down menu.
 - d. Enter the IP address range for the subnet in the **IP address range** text box.
For example, *10.0.0.0/99*.
- e. Select **On** or **Off** from the Private Google Access options to set access for VMs on the subnet to access Google services without assigning external IP addresses.
- f. Click **Done**.

Note:

Click **Add Subnet** to add another subnet.

8. Select **Regional** from the Dynamic routing mode option.
9. Click **Create** to create the VPC.
The VPC is added to the network.

5.3.2.1 Adding a Subnet to the Virtual Private Cloud (VPC)

You can add a subnet to your VPC.

► To add a subnet:

1. Ensure that you are logged in to the GCP Console.
2. Under Networking, navigate to **VPC network > VPC networks**.
The *VPC networks* screen appears.
3. Select the VPC.
The *VPC network details* screen appears.
4. Click **EDIT**.
5. Under Subnets area, click **Add Subnet**.
The **Add a subnet** screen appears.
6. Enter the subnet details.
7. Click **ADD**.
8. Click **Save**.

The subnet is added to the VPC.

5.3.3 Obtaining the GCP Image

Before creating the instance on GCP, you must obtain the image from the [My.Protegility](#) portal. On the portal, you select the required ESA version and choose **GCP** as the target cloud platform. You then share the product to your cloud account. The following steps describe how to share the image to your cloud account.

► To obtain and share the image:

1. Log in to the [My.Protegility](#) portal with your user account.
2. Click **Product Management > Explore Products > Data Protection**.
3. Select the required ESA Platform Version.
The **Product Family** table will update based on the selected ESA Platform Version.

Note: The ESA Platform Versions listed in the drop-down menu reflect all versions that were either previously downloaded or shipped within the organization along with any newer versions available thereafter. You can check the list of products previously downloaded from **Product Management > My Product Inventory**.

4. Select the **Product Family**.
The description box will populate with the **Product Family** details.

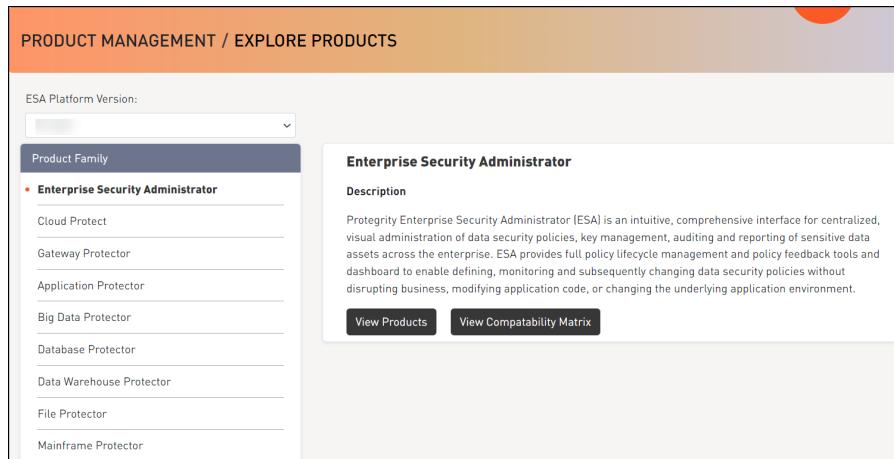


Figure 5-12: Product Family Screen

- Click **View Products** to advance to the **Product List** screen.

The screenshot shows the 'Enterprise Security Administrator / PRODUCT LIST' screen. It displays a table of compatible products across different target platforms. Callouts numbered 1 through 10 point to specific elements: 1. Target Platform Details; 2. Product Name; 3. Product Family; 4. OS Details; 5. Version; 6. End of Support Date; 7. Action; 8. Export as CSV; 9. Search Criteria; 10. Request one here.

Figure 5-13: Product List Screen

Table 5-4: Product List Screen Description

| Callout | Element Name | Description |
|---------|--------------------------------|---|
| 1 | Target Platform Details | Shows details about the target platform. |
| 2 | Product Name | Shows the product name. |
| 3 | Product Family | Shows the product family name. |
| 4 | OS Details | Shows the operating system name. |
| 5 | Version | Shows the product version. |
| 6 | End of Support Date | Shows the final date that Protegility will provide support for the product. |
| 7 | Action | Click the View icon (ocular icon) to open the Product Detail screen. |
| 8 | Export as CSV | Downloads a .csv file with the results displayed on the screen. |
| 9 | Search Criteria | Type text in the search field to specify the search filter criteria or filter the entries using the following options: <ul style="list-style-type: none">• OS• Target Platform |
| 10 | Request one here | Opens the Create Certification screen for a certification request. |

- Select the **GCP** cloud target platform you require and click the **View** icon (ocular icon) from the **Action** column. The **Product Detail** screen appears.
- Click the **Share Product** icon (cloud icon) to share the GCP cloud product.

Note:

If the user does not have access to cloud products or has access to cloud products and the Customer Cloud Account details are not available, then a message appears with the information that is required and the contact information for obtaining access to cloud share.

A dialog box appears and your available cloud accounts will be displayed.

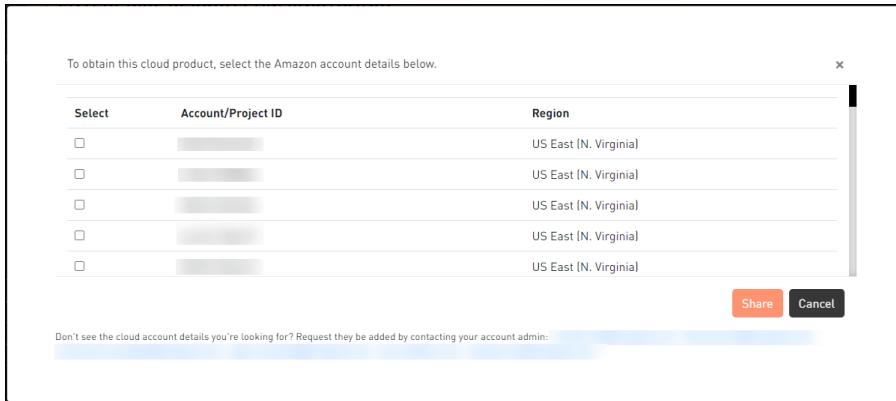


Figure 5-14: Account Selection Screen

8. Select your required cloud account in which to share the Protegility product.

9. Click **Share**.

A message box is displayed with the command line interface (CLI) instructions with the option to download a detailed PDF containing cloud web interface instructions. Additionally, the instructions for sharing the cloud product are sent to your registered email address and to your notification inbox in *My.Protegility*.

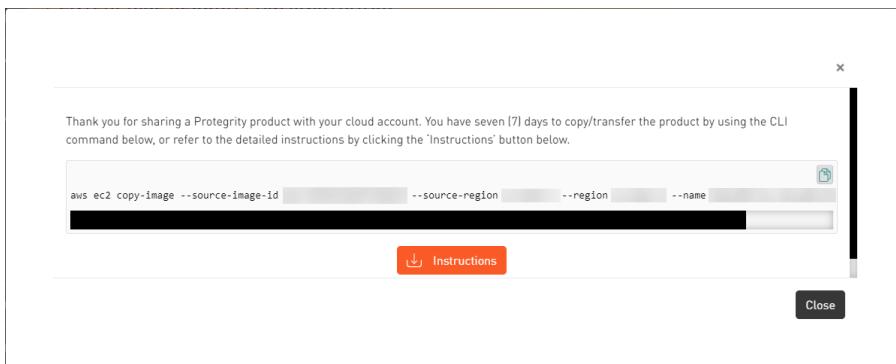


Figure 5-15: Sharing Command

10. Click the **Copy** icon () to copy the command for sharing the cloud product and run the command in CLI. Alternatively, click **Instructions** to download the detailed PDF instructions for cloud sharing using the CLI or the web interface.

Note:

The cloud sharing instruction file is saved in a *.pdf* format. You need a reader, such as, Acrobat Reader to view the file.

The Cloud Product will be shared with your cloud account for seven (7) days from the original share date in the *My.Protegility* portal.

After the seven (7) day time period, you need to request a new share of the cloud product through *My.Protegility*.

5.3.4 Converting the Raw Disk to a GCP Image

After obtaining the image from Protegility, you can proceed to create a virtual image. However, the image provided is available as disk in a raw format. This must be converted to a GCP specific image before you create an instance. The following steps provide the details of converting the image in a raw format to a GCP-specific image.

► To convert the image:

1. Login to the GCP Console.
2. Run the following command.

```
gcloud compute images create <Name for the new GCP Image> --source-uri gs://<Name of the storage location where the raw image is obtained>/<Name of the GCP image>
```

For example,

```
gcloud compute images create esa80 --source-uri gs://stglocation80/esa-pap-all-64-x86-64-gcp-8-0-0-0-1924.tar.gz
```

The raw image is converted to a GCP-specific image. You can now create an instance using this image

5.3.5 Loading the Protegility Appliance from a GCP Image

This section describes the tasks that you must perform to load the Protegility appliance from an image that is provided by Protegility. You must create a VM instance using the image provided in the following two methods:

- Creating a VM instance from the Protegility appliance image provided
- Creating a VM instance from a disk that is created with an image of the Protegility appliance

5.3.5.1 Creating a VM Instance from an Image

This section describes how to create a VM instance from an appliance image provided to you.

► To create a VM instance from an image:

1. Ensure that you are logged in to the GCP.
2. Click **VM instances**.
The *VM instances* screen appears.
3. Click **CREATE INSTANCE**.
The *Create an instance* screen appears.
4. Enter the following information:
 - **Name:** Name of the instance
 - **Description:** Description for the instance
5. Select the region and zone from the **Region** and **Zone** drop-down menus respectively.
6. Under the **Machine Type** area, select the processor and memory configurations based on the requirements.

Note:

Click **Customize** to customize the memory, processor, and core configuration.

7. Under the Boot disk area, click **Change** to configure the boot disk.
The *Boot disk screen* appears.
 - a. Click **Custom Images**.
 - b. Under the **Show images from** drop-down menu, select the project where the image of the appliance is provided.



- c. Select the image for the root partition.
 - d. Select the required disk type from the **Boot disk type** drop-down list.
 - e. Enter the size of the disk in the **Size (GB)** text box.
 - f. Click **Select**.
- The disk is configured.
8. Under the *Identity and API access* area, select the account from the **Service Account** drop-down menu to access the Cloud APIs.
- Depending on the selection, select the access scope from the **Access Scope** option.
9. Under the *Firewall* area, select the **Allow HTTP traffic** or **Allow HTTPS traffic** checkboxes to permit HTTP or HTTPS requests.
10. Click **Networking** to set the networking options.
- a. Enter data in the **Network tags** text box.
 - b. Click **Add network interface** to add a network interface.
- If you want to edit a network interface, then click the edit icon ().
11. Click **Create** to create and start the instance.

5.3.5.2 Creating a VM Instance from a Disk

You can create disks using the image provided for your account. You must create a boot disk using the OS image. After creating the disk, you can attach it to an instance.

This section describes how to create a disk using an image. Using this disk, you then create a VM instance.

5.3.5.2.1 Creating a Disk from the GCP Image

Perform the following steps to create a disk using an image.

 To create a disk of the Protegility appliance:

1. Access the GCP domain at the following URL:
<https://cloud.google.com/>
The GCP home screen appears.
2. Click **Console**.
The GCP login screen appears.
3. On the GCP login screen, enter the following details:
 - User Name
 - Password
4. Click **Sign in**.
After successful authentication, the GCP management console screen appears.
5. Click **Go to the Compute Engine dashboard** under the *Compute Engine* area.
The *Dashboard* screen appears
6. Click **Disk** on the left pane.
The *Disk* screen appears.

7. Click **CREATE DISK** to create a new disk.
The *Create a disk* screen appears.
8. Enter the following details:
 - **Name:** Name of the disk
 - **Description:** Description for the disk
9. Select one of the following options from the **Type** drop-down menu:
 - **Standard persistent disk**
 - **SSD persistent disk**
10. Select the region and zone from the **Region** and **Zone** drop-down menus respectively.
11. Select one of the following options from the **Source Type** option:
 - **Image:** The image of the Protegility appliance that is provided.
Select the image from the Source Image drop-down menu.

Note:

Ensure that you have access to the Protegility appliance images.

- **Snapshot:** The snapshot of a disk
 - **Blank:** Create a blank disk
12. Enter the size of the disk in the **Size (GB)** text box.
 13. Select **Google-managed** key from the **Encryption** option.
 14. Click **Create**.

The disk is created.

5.3.5.2.2 Creating a VM Instance from a Disk

This section describes how to create a VM instance from a disk that is created from an image.

For more information about creating a disk, refer to section [Creating a Disk from the GCP Image](#).

► To create a VM instance from a disk:

1. Ensure that you are logged in to the GCP Console.
2. Click **VM instances**.
The *VM instances* screen appears.
3. Click **CREATE INSTANCE**.
The *Create an instance* screen appears.
4. Enter information in the following text boxes:
 - **Name**
 - **Description**
5. Select the region and zone from the **Region** and **Zone** drop-down menus respectively.
6. Under the **Machine Type** section, select the processor and memory configuration based on the requirements.
Click **Customize** to customize your memory, processor and core configuration.
7. Under *Boot disk* area, click **Change** to configure the boot disk.

The *Boot disk* screen appears.

- Click **Existing Disks**.
 - Select the required disk created with the Protegility appliance image.
 - Click **Select**.
8. Under Firewall area, select the **Allow HTTP traffic** or **Allow HTTPS traffic** checkboxes to permit HTTP or HTTPS requests.
9. Click **Create** to create and start the instance.

5.3.5.3 Accessing the Appliance

After setting up the virtual machine, you can access the appliance through the IP address that is assigned to the virtual machine. It is recommended to access the appliance with the administrative credentials.

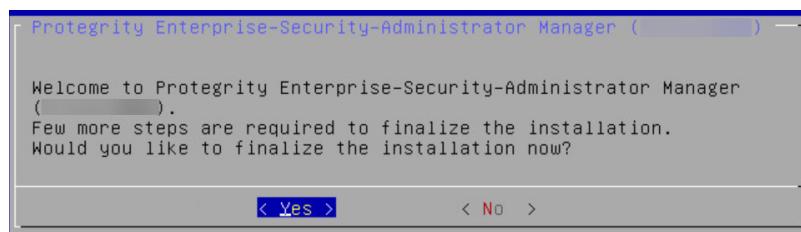
Note:

If the number of unsuccessful password attempts exceed the defined value in the password policy, the account gets locked.

For more information on the password policy for the admin and viewer users, refer the section *Password Policy for All Appliance Services*, and for the *root* and *local_admin* OS users, refer the section *Managing Local OS Users* in the *Protegility Appliances Overview Guide 9.1.0.5*.

5.3.6 Finalizing the Installation of Protegility Appliance

After installing AWS, the appliance framework is installed without the Protegility-specific components and products. You must finalize the installation to complete the installation. When you login to the CLI Manager, the following screen appears.



Select **Yes** to finalize the installation.

Note:

You must have *administrator* privileges to finalize the installation.

Alternatively, you can also finalize the installation later.

During the finalizing of installation, the following sensitive data is configured:

- Hostname
- Appliance certificates
- Service credentials
- SSH Keys
- Appliance security identifiers

All the other components, such as logging, reporting, and policy management are also installed.

Note:

After the execution of the rotation tool, you may install the latest updates/patches/service-packs. You may contact Protegility support for any information related to installation of the latest updates/patches/service-packs.

For more information about running the Appliance-rotation-tool on the Protegility appliance, refer to section *Running the Appliance-Rotation-Tool* in the *Protegility Enterprise Security Administrator Guide 9.1.0.5*.

5.3.6.1 Finalizing ESA Installation

You can finalize the installation of the ESA after signing in to the CLI Manager.

Caution:

Ensure that the finalization process is initiated from a single session only. If you start finalization simultaneously from a different session, then the "*Finalization is already in progress.*" message appears. You must wait until the finalization of the instance is successfully completed.

Additionally, ensure that the appliance session is not interrupted. If the session is interrupted, then the instance becomes unstable and the finalization process is not completed on that instance.

► To finalize ESA installation:

1. Sign in to the ESA CLI Manager of the instance created using the default administrator credentials.
The following screen appears.



Figure 5-16: Finalizing Installation Confirmation screen

For more information about default credentials, refer the section *Creating an Instance of the Protegility Appliance from the AMI* in the *Protegility Appliances Overview Guide 9.1.0.5*.

2. Select **Yes** to initiate the finalization process.

The screen to enter the administrative credentials appears.

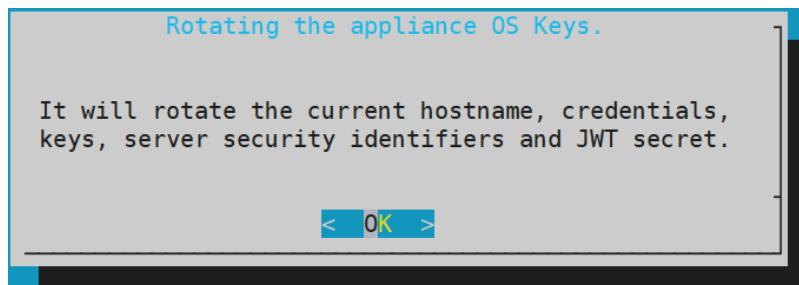
Note:

If you select **No**, then the finalization process is not initiated.

To manually initiate the finalization process, navigate to **Tools > Finalize Installation** and press **ENTER**.

3. Enter the credentials for the *admin* user and select **OK**.

A confirmation screen to rotate the appliance OS keys appears.



4. Select **OK** to rotate the appliance OS keys.

The following screen appears.

A screenshot of a "User's Passwords" configuration screen. It asks for "Please provide user's passwords". There are four pairs of password fields: "root password" and "root password verification", "admin password" and "admin password verification", "viewer password" and "viewer password verification", and "local_admin password" and "local_admin password verification". Each pair consists of a grey input field and a red password strength bar. At the bottom are two buttons: "<Apply>" and "<Help >".

- a. To update the user passwords, provide the credentials for the following users:

- root
- admin
- viewer
- local_admin

- b. Select **Apply**.

The user passwords are updated and the appliance OS keys are rotated.

The finalization process is completed.

Note:

If you want to verify the installed products or install additional products, then navigate to **Administration > -- Installations and Patches -- > Add/Remove Services**.

For more information about installing products, refer to the section [Installing Products](#).

For more information about rotating the OS keys, refer to section *Rotate Appliance OS Keys* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

After rotating appliance keys, rotate the Audit Store certificates using the steps from *Rotating Audit Store Certificates* in the [Audit Store Guide 9.1.0.5](#).

5.3.7 Connecting to an ESA instance (for DSG deployment)

If you are using an instance of the DSG appliance on GCP, you must connect it to the instance of the ESA appliance. Using the CLI manager, you must provide the connectivity details of the ESA appliance in the DSG appliance.

For more information about connecting to an instance of the ESA appliance, refer to the section *Setting up ESA Communication* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

5.3.8 Deploying the Instance of the Protegility Appliance with the Protectors

You can configure the various protectors that are a part of the Protegility Data Security Platform with the instance of the ESA appliance running on GCP.

Depending on the Cloud-based environment which hosts the protectors, the protectors can be configured with the instance of the ESA appliance in one of the following ways:

- If the protectors are running on the same VPC as the instance of the ESA appliance, then the protectors need to be configured using the internal IP address of the appliance within the VPC.
- If the protectors are running on a different VPC than that of the instance of the ESA appliance, then the VPC of the instance of the ESA needs to be configured to connect to the VPC of the protectors.

5.3.9 Backing up and Restoring Data on GCP

A snapshot represents a state of an instance or disk at a point in time. You can use a snapshot of an instance or a disk to backup or restore information in case of failures.

5.3.9.1 Creating a Snapshot of a Disk on GCP

This section describes the steps to create a snapshot of a disk.

► To create a snapshot on GCP:

1. On the **Compute Engine** dashboard, click **Snapshots**.
The *Snapshots* screen appears.
2. Click **Create Snapshot**.
The *Create a snapshot* screen appears.
3. Enter information in the following text boxes.
 - Name - Name of the snapshot
 - Description – Description for the snapshot
4. Select the required disk for which the snapshot is to be created from the **Source Disk** drop-down list.
5. Click **Add Label** to add a label to the snapshot.
6. Enter the label in the **Key** and **Value** text boxes.
7. Click **Add Label** to add additional tags.

8. Click **Create**.

Note: Ensure that the status of the snapshot is set to **completed**.

Note: Ensure that you note the snapshot id.

5.3.9.2 Restoring from a Snapshot on GCP

This section describes the steps to restore data using a snapshot.

Note: Ensure that the snapshot of the disk is created.

► To restore data using a snapshot on GCP:

1. Navigate to **Compute Engine > VM instances**.
The *VM instances* screen appears.
2. Select the required instance.
The screen with instance details appears.
3. Stop the instance.
4. After the instance is stopped, click **EDIT**.
5. Under the **Boot Disk** area, remove the existing disk.
6. Click **Add Item**.
7. Select the **Name** drop-down list and click **Create a disk**.
The **Create a disk** screen appears.
8. Under **Source Type** area, select the required snapshot.
9. Enter the other details, such as, **Name**, **Description**, **Type**, and **Size (GB)**.
10. Click **Create**.
The snapshot of the disk is added in the **Boot Disk** area.
11. Click **Save**.
The instance is updated with the new snapshot.

5.3.10 Increasing Disk Space on the Appliance

After creating an instance on GCP, you can add a disk to your appliance.

► To add a disk to a VM instance:

1. Ensure that you are logged in to the GCP Console.
2. Click **VM instances**.
The *VM instances* screen appears.
3. Select the instance.
The *VM instance* details screen appears.
4. Click **EDIT**.

5. Under **Additional disks**, click **Add new disk**.
6. Enter the disk name in the **Name** field box.
7. Select the disk permissions from the **Mode** option.
8. If you want to delete the disk or keep the disk after the instance is created, select the required option from the **Deletion rule** option.
9. Enter the disk size in GB in the **Size (GB)** field box.
10. Click **Done**.
11. Click **Save**.

The disk is added to the VM instance.

Chapter 6

Configuring the ESA

- [6.1 Configuring Authentication Settings](#)
 - [6.2 Configuring Accounts and Passwords](#)
 - [6.3 Configuring Syslog](#)
 - [6.4 Configuring External Certificates](#)
 - [6.5 Configuring SMTP](#)
 - [6.6 Configuring SNMP](#)
-

This section contains configurations that can be performed after the installation of the ESA is completed.

6.1 Configuring Authentication Settings

User authentication is the process of identifying someone who wants to gain access to a resource. A server contains protected resources that are only accessible to authorized users. When you want to access any resource on the server, the server uses different authentication mechanism to confirm your identity.

You can configure the authentication using for the following methods.

- Basic Authentication
- Client Certificates
- JSON Web Token (JWT)
- Proxy Authentication
- Single Sign-On (SSO)

For more information about configuring the authentication settings, refer to the [Protegility Appliances Overview Guide 9.1.0.5](#).

6.2 Configuring Accounts and Passwords

You can change your current password from the CLI Manager. The CLI Manager includes options to change passwords and permissions for multiple users.

For more information on configuring accounts and passwords, refer to section *Accounts and Passwords Management* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

6.3 Configuring Syslog

The Appliance Logs tool can be differentiated into appliance common logs and appliance-specific logs. Syslog is a log type that is common for all appliances.

For more information about configuring syslog, refer the section Working with Logs in *Protegility Appliances Overview Guide 9.1.0.5* and *Protegility Enterprise Security Administrator Guide 9.1.0.5*.

6.4 Configuring External Certificates

External certificates or digital certificates are used to encrypt online communications securely between two entities over the Internet. It is a digitally signed statement that is used to assert the online identities of individuals, computers, and other entities on the network, utilizing the security applications of Public Key Infrastructure (PKI). Public Key Infrastructure (PKI) is the standard cryptographic system that is used to facilitate the secure exchange of information between entities.

For more information on configuring certificates, refer the section Certificate Management in ESA in *Protegility Certificate Management Guide 9.1.0.5*.

6.5 Configuring SMTP

The Simple Mail Transfer Protocol (SMTP) setting allows the system to send emails. You can set up an email server that supports the notification features in Protegility Reports.

► To configure SMTP from Web UI:

1. Login to the ESA.
2. Navigate to **Settings > Network**.
3. Click the **SMTP Settings** tab.

The following screen appears.

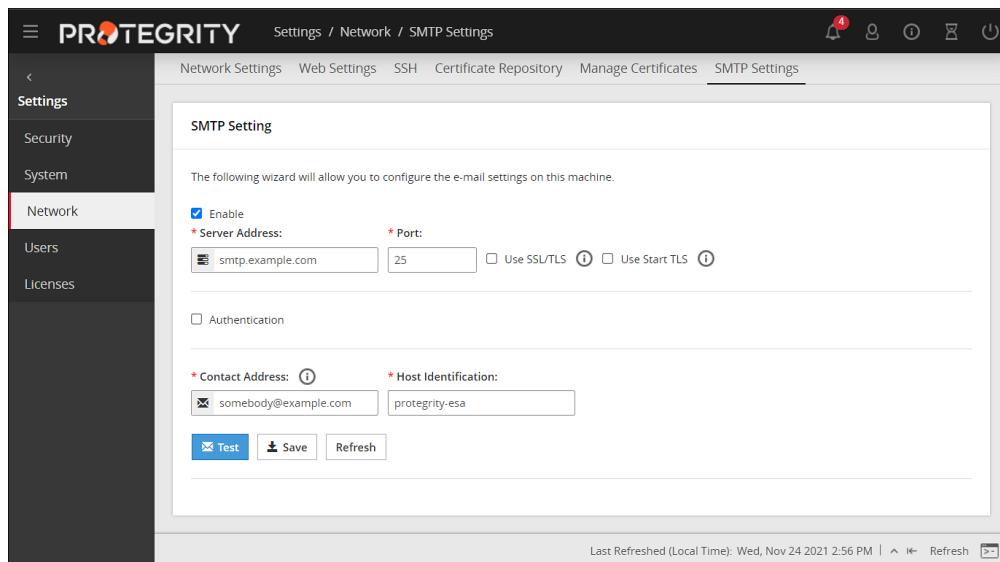


Figure 6-1: Email Settings screen

For more information about configuring SMTP, refer to the section *Email Setup* in the *Protegility Appliances Overview Guide 9.1.0.5*.

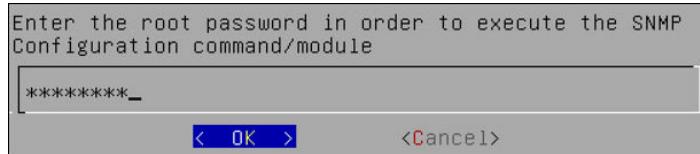
6.6 Configuring SNMP

Using Simple Network Management Protocol (SNMP), you can query the appliance performance data.

By default, due to security reasons, the SNMP service is disabled. To enable the service and provide its basic configuration (listening address, community string) you can use the SNMP tool available in the CLI Manager.

► To initialize SNMP configuration:

1. Login to the CLI Manager.
2. Navigate to **Networking > SNMP Configuration**.
3. Enter the root password to execute the SNMP configuration and click **OK**.



The following screen appears.

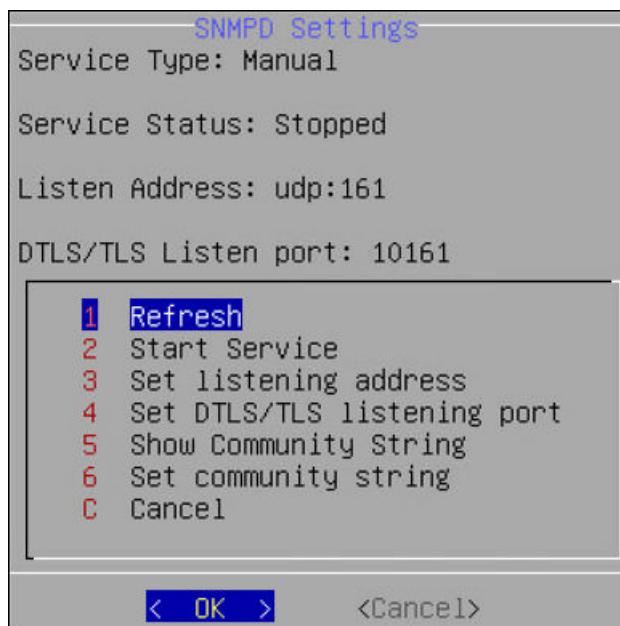


Figure 6-2: SNMP Configuration Tool

You can also start the SNMP Service from the Web UI. Navigate to **System > Services** to start the SNMP service.

For more information about configuring SNMP, refer to the section *Configure SNMP* in the *Protegility Appliances Overview Guide 9.1.0.5*.

Chapter 7

Initializing the Policy Information Management (PIM) Module

After completing the installation of the ESA, you must initialize the Policy Information Management (PIM) module, which creates the keys-related data and the policy repository.

► To initialize the PIM module:

1. In a web browser, enter the ESA IP address in the window task bar.
2. Enter the **Username** and **Password**.
3. Click **Sign in**.
The ESA dashboard appears.
4. Navigate to **Policy Management > Dashboard**.
The following screen to initialize PIM appears.

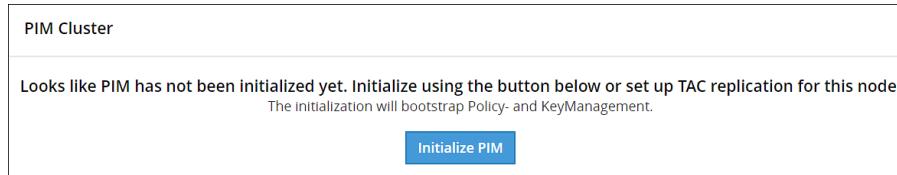


Figure 7-1: PIM Initialization Screen

5. Click **Initialize PIM**.
A confirmation message appears.
6. Click **OK**.
The Policy management screen appears.

Chapter 8

Configuring the ESA in a Trusted Appliances Cluster (TAC)

In a scenario where the ESAs are configured in a TAC setup, you must add at least three Appliance nodes for the Audit Store cluster.

The following figure illustrates the TAC setup.

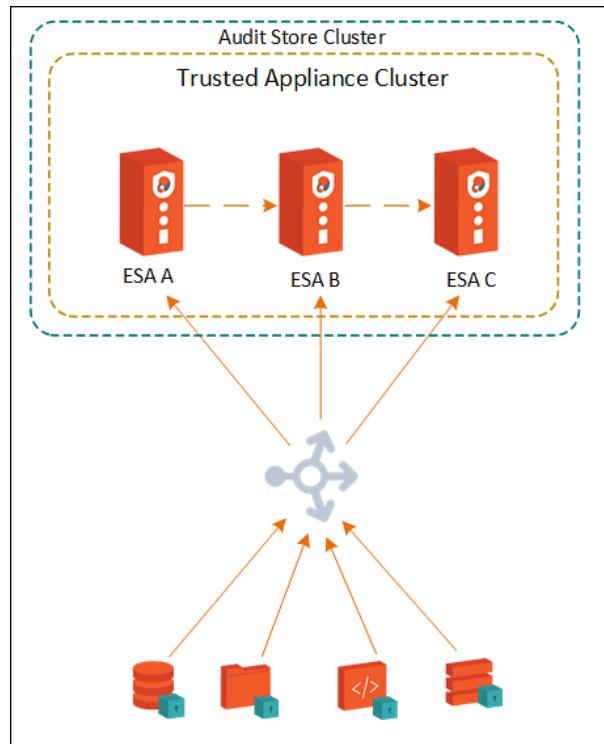


Figure 8-1: TAC on ESA 9.1.0.3

1. A TAC is established between the primary appliance ESA A and the Secondary ESAs, ESA B and ESA C.

For more information about TAC, refer to the section *Trusted Appliances Cluster (TAC)* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

2. Data replication for policies, forensics, or DSG configuration takes place between all the ESAs.

For more information about replication tasks, refer to the section *Working with Backup and Restore* in the [Protegility Appliances Overview Guide 9.1.0.5](#).

3. The Audit Store cluster is enabled for the ESAs.

For more information about enabling Audit Store Cluster, refer to the section [Using Protegility Analytics](#).

4. All the ESAs are added as a part of the Audit Store Cluster.

For more information about adding an ESA to the Audit Store Cluster, refer to the section [Creating an Audit Store Cluster](#).

Chapter 9

Creating an Audit Store Cluster

[9.1 Completing the Prerequisites](#)

[9.2 Initializing the Audit Store Cluster on the ESA](#)

[9.3 Adding an ESA to the Audit Store Cluster](#)

[9.4 Refreshing the Audit Store Cluster](#)

[9.5 Configuring *td-agent* in the Audit Store Cluster](#)

[9.6 Verifying the Audit Store Cluster](#)

[9.7 Updating the Priority IP List for Signature Verification](#)

The following sections describes the steps to create an Audit Store Cluster.

9.1 Completing the Prerequisites

Ensure that the following prerequisites are met before configuring the Audit Store Cluster. Protegility recommends that the Audit Store Cluster has a minimum of 3 ESAs for creating a highly-available multi-node Audit Store cluster.

1. Prepare and set up three ESAs v9.1.0.5.
2. Add the first ESA to the TAC. This will be the Primary ESA.

For more information about installing the ESA, refer to the section [Installing the ESA On-Premise](#) or [Installing Appliances on Cloud Platforms](#).

3. Add the remaining ESAs to the TAC. These will be the Secondary ESAs in the TAC.

For more information about installing the ESA, refer to the section [Installing the ESA On-Premise](#) or [Installing Appliances on Cloud Platforms](#).

9.2 Initializing the Audit Store Cluster on the ESA

Complete the steps provided in this section on the first ESA or the Primary ESA in the TAC. When you select this option, Protegility Analytics is configured to retrieve data from the local Audit Store. Additionally, the required processes, such as *td-agent*, is started and Protegility Analytics is initialized. The Audit Store cluster is initialized on the local machine so that other nodes can join this Audit Store cluster.

Perform the following steps to configure the Audit Store.

1. Login to the ESA Web UI.
2. Verify that the Audit Store services are running by navigating to **System > Services > Audit Store**.
3. Navigate to **Analytics**.

The following screen appears.



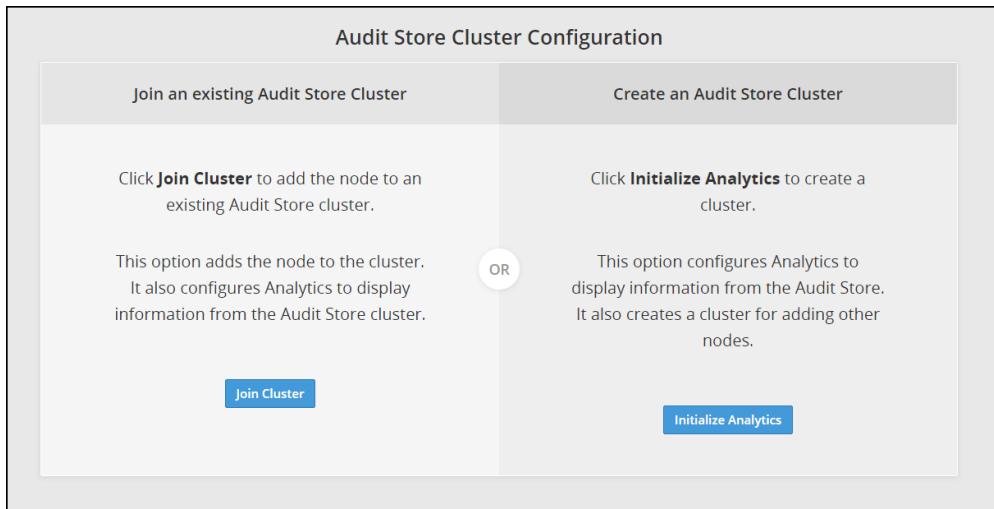


Figure 9-1: Analytics Screen

4. Click **Initialize Analytics**.

Protegility Analytics is initialized, the internal configuration is updated for creating the local Audit Store cluster, the *td-agent* service is started, and logs are read from the Audit Store. Other Audit Store nodes can now join this Audit Store cluster.

Protegility Analytics is now configured and retrieves data for the reports from the Audit Store. The data is available on the **Analytics > Forensics** tab on the ESA Web UI as shown in the following figure.

The screenshot shows the Protegility ESA Web UI with the 'Forensics' tab selected. The main pane displays audit log data with columns for additional_info.description, additional_info.module, additional_info.procedure, client.ip, client.username, cnt, correlationid, endtime, filetype, index_node, and inde. A message at the top of the list states 'Analytics successfully initialized! Logs will be loaded soon.' The bottom of the page shows a status bar with 'Last Refreshed (Local Time):' and a refresh button.

Figure 9-2: Forensics

9.3 Adding an ESA to the Audit Store Cluster

If multiple ESAs need to be added to the Audit Store cluster, such as multiple ESAs in a TAC, then the steps in this section need to be performed. In this case, the current ESA that you are adding will be a node in the Audit Store cluster. After the configurations are completed, the required processes are started and the logs are read from the Audit Store cluster. Complete the steps in this section to join an existing Audit Store cluster.

Caution:

The Audit Store cluster information is updated when a node joins the Audit Store cluster. This information is updated across the Audit Store cluster. Hence, nodes must be added to an Audit Store cluster one at a time. Adding multiple nodes to the Audit Store at the same

time using the ESA Web UI would make the cluster information inconsistent, make the Audit Store cluster unstable, and would lead to errors.

Ensure that the following prerequisites are met:

- The health status of the Audit Store node that you are connecting to is green or yellow.
- The health status of the Audit Store node that you are adding to the cluster is green or yellow.

Note: To check the health status of a node, login to ESA Web UI of the node, click **Audit Store Management**, and view the **Cluster Status** from the upper-right corner of the screen.

Perform the following steps to add a node to the Audit Store cluster.

Note: Ensure that the Audit Store cluster is created on the node that you want to join. You need to perform this step only if you need multiple ESAs or are implementing a TAC.

For more information about creating an Audit Store cluster, refer to the section *Initializing the Audit Store Cluster on the ESA*.

Important: Before joining an Audit Store cluster, ensure that the SSH Authentication type is set to **Public key** or **Password + PublicKey**.

For more information about setting the authentication, refer to the section *Working with Secure Shell (SSH) Keys* in the *Protegility Appliances Overview Guide 9.1.0.5*.

1. Login to the Web UI of the second ESA.
2. Verify that the Audit Store services are running by navigating to **System > Services > Audit Store**.
3. Navigate to **Analytics**.

The following screen appears.

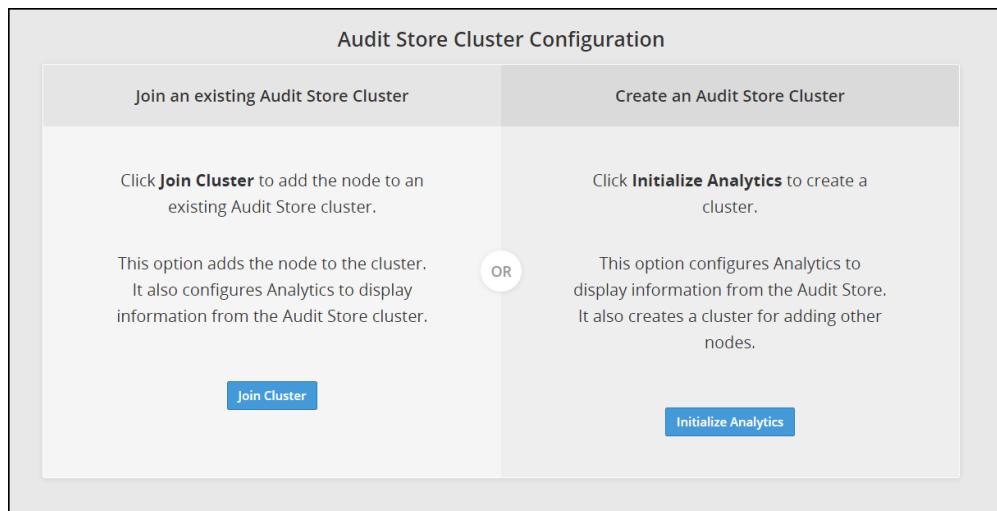


Figure 9-3: Analytics Screen

4. Click **Join Cluster**.

The following screen appears.

Join an existing Audit Store Cluster

Target node IP/Hostname*

Node IP/Hostname

Username*

Username

Password*

Password

Clear cluster data! This operation will clear data from the node. I have backed up the data and want to continue with this operation.

Join Cluster **Cancel**

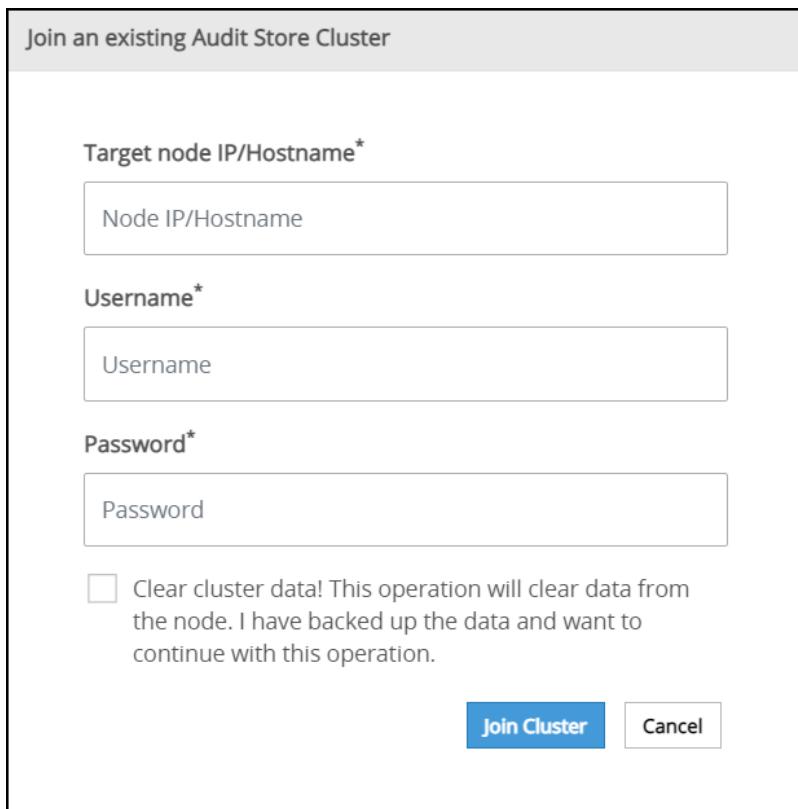


Figure 9-4: Joining an Audit Store Cluster

- Specify the IP address or the hostname of the Audit Store cluster to join.

Note: Only use hostname if the hostname is resolved between the nodes.

Ensure that Protegility Analytics is initialized and the Audit Store cluster is already created on the target node. A node cannot join the cluster if Protegility Analytics is not initialized on the target node.

For more information about initializing the Audit Store, refer to the section [Initializing the Audit Store Cluster on the ESA](#).

- Specify the admin username and password for the Audit Store cluster.

Note: If required, then select the **Clear cluster data** check box to clear the Audit Store data from the current node before joining the Audit Store cluster. The check box will only be enabled if the node has data, that is, if Analytics is installed and initialized on the node. Else, this check box is disabled.

- Click **Join Cluster**.

The internal configuration is updated for the Audit Store cluster, the *td-agent* service is started, and the node is added to the Audit Store cluster.

Protegility Analytics is now configured and retrieves data for the reports from the Audit Store cluster. The data is available on the **Analytics** tab on the ESA Web UI as shown in the following figure.

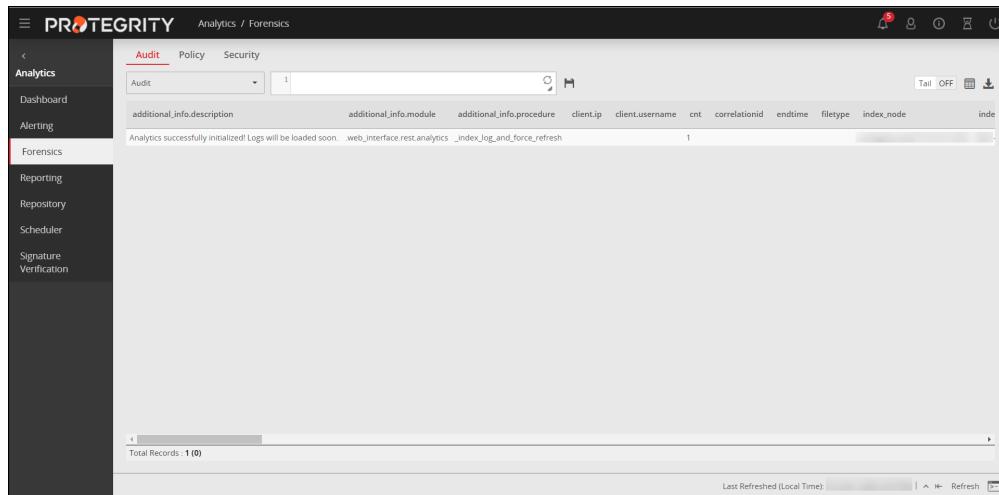


Figure 9-5: Protegity Analytics

9.4 Refreshing the Audit Store Cluster

Complete the steps in this section to refresh the ESA for the Audit Store Cluster.

1. Login to the ESA Web UI of the ESA node
2. Navigate to **System > Task Scheduler**.
3. Click the **Audit Store Management Update Unicast Hosts** task.
4. Click **Run now** and then click **OK** in the confirmation box.
5. If you are using a TAC, then perform the steps provided in this section on the other ESAs in the Audit Store Cluster.

9.5 Configuring td-agent in the Audit Store Cluster

Complete the following steps after adding the ESA node to the Audit Store cluster. This configuration is required for processing and storing the logs received by the Audit Store.

Note: This step must be performed on all the ESAs in the Audit Store cluster.

Before performing the steps provided here, verify that the Audit Store cluster health status is green on the **Audit Store Management** screen of the ESA Web UI.

1. Login to the CLI Manager of the *ESA* node.
2. Navigate to **Tools > PLUG - Forward logs to Audit Store**.
3. Enter the *root* password and select **OK**.
4. Enter the *username* and *password* for the administrative user, such as, *admin*.
5. Select **OK**.
6. In the *Setting ESA Communication* screen, select **OK**.
7. Specify the IP addresses of all the ESA machines in the cluster, separated by commas.

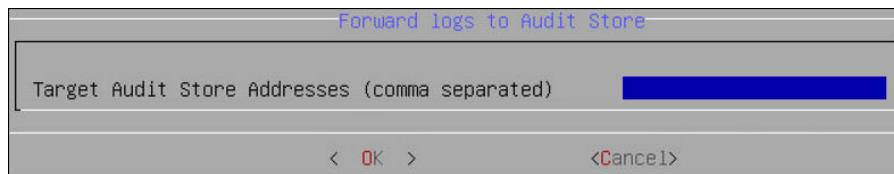


Figure 9-6: Forward Logs

8. Select **OK**.
9. Type *y* to fetch certificates for communicating with the ESA and select **OK**.

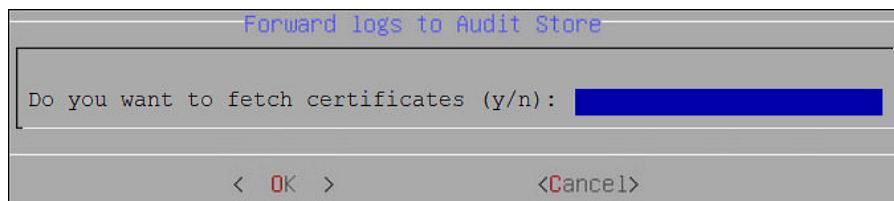


Figure 9-7: Fetch Certificates

10. Enter the *admin* username and password and select **OK**.
- Repeat the steps provided in this section on all the ESAs in the Audit Store Cluster.

9.6 Verifying the Audit Store Cluster

View the Audit Store Management page to verify that the configurations that you performed were completed successfully using the steps provided here.

1. Login to the ESA Web UI.
2. Navigate to the **Audit Store Management** page.
3. Verify that the nodes are added to the cluster. The health of the nodes must be either green or yellow.
4. If you added additional ESAs for creating a TAC, then verify that the ESA has only the master role.

| Node IP | Master | Data | Ingest | Action | Name | Up Time | Disk Total (Bytes) | Disk Used (Bytes) | Disk Avail (Bytes) | RAM |
|---------------|--------|------|--------|------------|------|---------|--------------------|-------------------|--------------------|------|
| 192.168.1.101 | ✓ | ✓ | ✓ | Edit Roles | | 19s | 39,502,524,416 | 6,955,266,048 | 32,547,258,368 | 6.44 |
| 192.168.1.102 | ✓ | ✓ | ✓ | Edit Roles | | 2.5h | 39,502,524,416 | 7,021,211,648 | 32,481,312,768 | 16.6 |
| 192.168.1.103 | ✓ | ✓ | ✓ | Edit Roles | | 4.9m | 45,709,819,904 | 5,415,325,696 | 40,294,494,208 | 28.3 |

Figure 9-8: Nodes Added to Cluster

9.7 Updating the Priority IP List for Signature Verification

Signature verification jobs run on the ESA and use the ESA's processing time. Ensure that you update the priority IP list for the default signature verification jobs after you set up the system. By default, the Primary ESA will be used for the priority IP. If you have multiple ESAs in the priority list, then additional ESAs are available to process the signature verifications jobs that must be processed. This frees up the Primary ESA's processor to handle other important tasks.

For example, if the maximum jobs to run on an ESA is set to 4 and 10 jobs are queued to run on 2 ESAs, then 4 jobs are started on the first ESA, 4 jobs are started on the second ESA, and 2 jobs will be queued to run till an ESA job slot is free to accept and run the queued job.

For more information about scheduling jobs, refer to the section *Using the Scheduler* in the [Protegrity Analytics Guide 9.1.0.5](#).

For more information about signature verification jobs, refer to the section *Verifying Signatures* in the [Protegrity Analytics Guide 9.1.0.5](#).

Use the steps provided in this section to update the priority IP list.

1. Login to the ESA Web UI.
2. Navigate to **Analytics > Scheduler**.
3. From the **Action** column, click the **Edit** icon () for the **Signature Verification** task.
4. Update the **Priority IPs** field with the list of the ESAs available separating the IPs using commas.
5. Click **Save**.

Chapter 10

Verifying the ESA Installation from the Web UI

After you install the ESA v9.1.0.5, you can perform the following steps to verify the installation.

- To verify the ESA installation from the Web UI:

1. Login to the ESA Web UI.

The ESA dashboard appears.

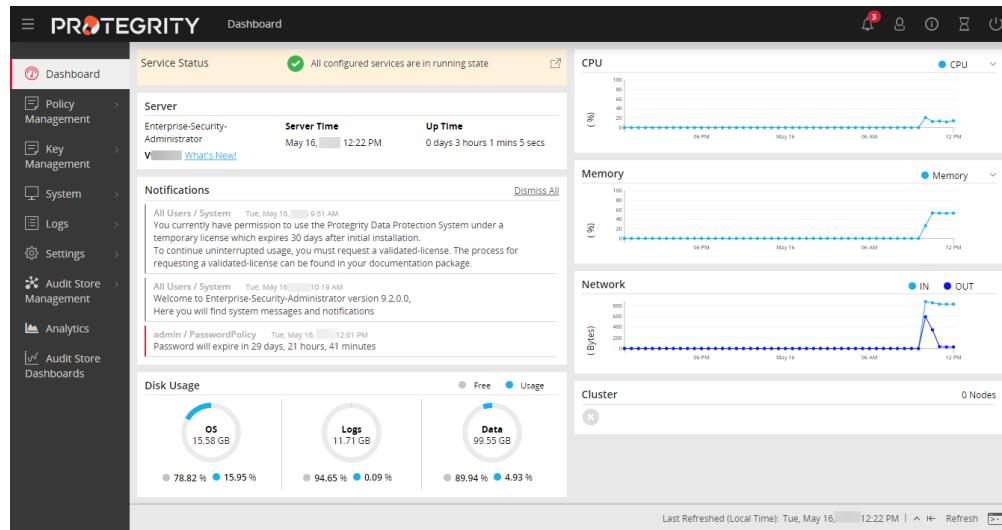


Figure 10-1: ESA Dashboard

2. Navigate to **System > Information**.

The screen displaying the information of your system appears.

3. Under the **Installed Patches** area, the **ESA_9.1.0.5** entry appears.

4. Navigate to **System > Services** and ensure that all the required services are running.

Chapter 11

Installing the Data Security Gateway (DSG)

[11.1 Installing the DSG](#)

[11.2 Installing the DSG on Cloud Platforms](#)

The Protegility Data Security Gateway Technology represents security operations on the network.

11.1 Installing the DSG

The DSG install requires an existing ESA, which serves as a single point of management for the data security policy, rules configuration, and on-going monitoring of the system.

For information about the ESA version supported by this release of the DSG, refer to the section *Products Compatibility Matrix* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

The DSG installation includes the following two installation paths:

- Applying the DSG patch on the compatible ESA

Note: For example, if you want to install DSG v3.1.0.5, then you must apply the *ESA_PAP-ALL-64_x86-64_9.1.0.5.xxxx.DSGUP.pty* patch on the ESA v9.1.0.5. This patch is applied on the ESA to extend the ESA with the DSG Web UI.

For more information about applying the DSG patch on the compatible ESA, refer to the section *Installing the DSG* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

- Installing the DSG, which is often referred as a DSG node

For more information about the installation order for installing the DSG on the ESA and the DSG nodes, refer to the section *Installing the DSG* in the [Protegility Data Security Gateway User Guide 3.1.0.5](#).

11.2 Installing the DSG on Cloud Platforms

For installing the Data Security Gateway (DSG) on cloud platforms, you must mount the image containing the Protegility appliance on a cloud instance or a virtual machine. The supported cloud platforms are Amazon Web Services (AWS), Microsoft Azure, and Google Cloud Platform (GCP).

You must configure the cloud instance to run an appliance on a cloud platform.

11.2.1 Configuring Cloud Instances

Before mounting the appliance images on the cloud platforms, you must create virtual machines or instances on them. This procedure involves configuring network settings on the cloud platforms, creating accounts, containers, and so on.

The process to create instances varies between the different platforms, such as, AWS, Azure, and GCP.

Table 11-1: Configuring Cloud Instances

| To | Refer |
|------------------------------------|--|
| Install the DSG appliance on AWS | Section <i>Installing Data Security Gateway (DSG) on Amazon Web Services (AWS)</i> in the Protegility Data Security Gateway User Guide 3.1.0.5 |
| Install the DSG appliance on Azure | Section <i>Installing the Data Security Gateway (DSG) on Microsoft Azure</i> in the Protegility Data Security Gateway User Guide 3.1.0.5 |
| Install the DSG appliance on GCP | Section <i>Installing the Data Security Gateway (DSG) on Google Cloud Platform (GCP)</i> in the Protegility Data Security Gateway User Guide 3.1.0.5 |



Chapter 12

Protegility Data Protectors Installation

[12.1 General Architecture](#)

[12.2 Installing the Log Forwarder on Protectors](#)

[12.3 Installing and Uninstalling the PEP Server](#)

[12.4 Protector Proxy Installation](#)

[12.5 Installing and Uninstalling Application Protectors](#)

[12.6 Installing and Uninstalling the Big Data Protector](#)

[12.7 Installing and Uninstalling Database Protectors](#)

The Enterprise Security Administrator (ESA) delivers centralized control over data security policy, keys, and reporting. A rich set of protectors enforce data security policy at protection points distributed throughout the enterprise. The ESA works with the following protectors: Cloud Protect, Gateway Protector, Application Protector, Big Data Protector, File Protector, and Data Warehouse Protector. They are deployed to any server in the enterprise containing sensitive data to be protected. Protection is based on a data security policy, which is deployed to every protection point from the ESA.

The protectors are implemented in different technologies and handle sensitive data in different forms, whether structured or unstructured.

12.1 General Architecture

This section explains the general architecture how ESA delivers security policy information to all protectors while the protectors deliver audit logs back to ESA. These tasks are facilitated by two agents:

- Communication Agent (referred to as PEP Server)
- Protection Enforcement Agent

Note: From v9.1.0.1 onwards, the architecture consists of 3 ESAs. During the Protector configuration, if you do not have PSUs as part of your architecture, then use the Primary ESA instead of the PSU for the configuration, such as, forwarding logs to the Audit Store. However, if you do have PSUs as a part of your architecture, then ensure that you keep the PSUs updated with all the available patches and fixes.

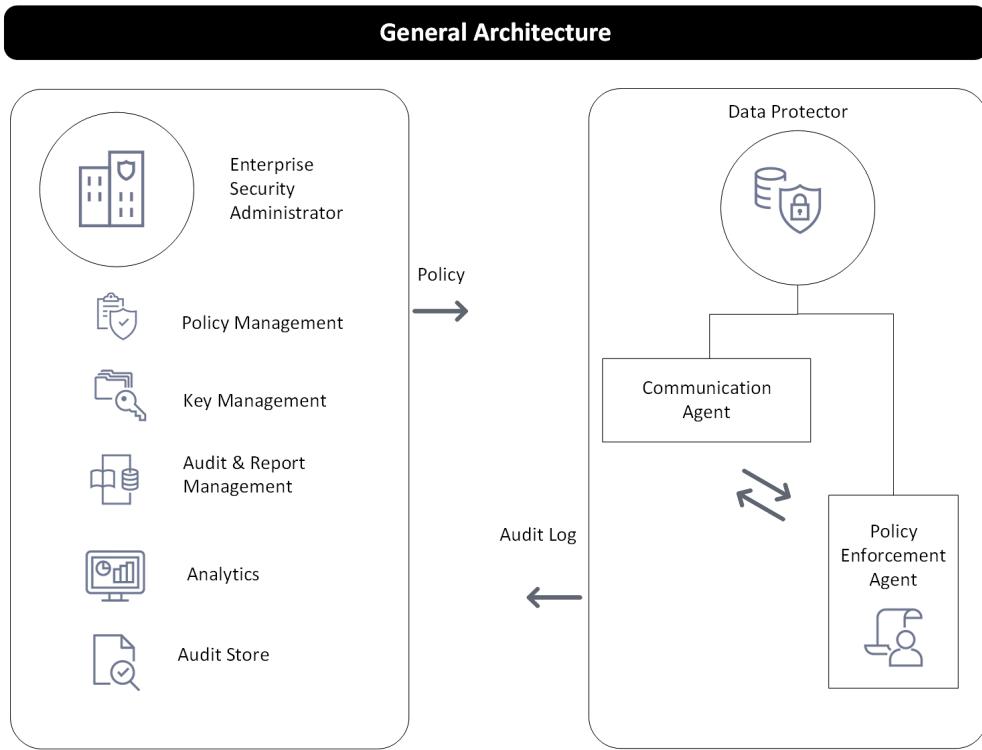


Figure 12-1: General Architecture of ESA working with Protegility Data Protectors

The **Communication Agent** is responsible for accepting the policy that is deployed from the Hub Controller. It prepares the policy in one of several forms so that it can be used by the actual Protection Enforcement Agent. It also collects, protects, and sends the audit logs to the main Audit Store repository in the ESA. This is the connection between the ESA and a Data Protector, also known as the PEP Server. The movement of Policy and Audit Logs between ESA and the Data Protectors is done through this secure channel.

The **Policy Enforcement Agent** is responsible for the actual enforcement of the policy and the protection of sensitive data. Its implementation differs based on the host technology and what the protector protects: database, file, or application. ESA allows security officers to easily specify data security requirements and distribute them across the enterprise to be executed locally. Once the policy is determined and set in ESA, it is deployed to Protegility Data Protectors for enforcement on their installed systems, ensuring consistent security enterprise-wide.

12.2 Installing the Log Forwarder on Protectors

The Log Forwarder collects logs generated by the system and forwards it to the Audit Store that is located on the ESA. This section explains the procedure to install the Log Forwarder on protectors.

For more information about the Log Forwarder, refer to the section *Introduction to Logging* in the [Protegility Log Forwarding Guide 9.1.0.5](#).

Caution:

After the Log Forwarder service is restarted, any data security operation, such as, protect or unprotect, may result in loss of the first data security operation audit. Subsequent data security operation counts are displayed accurately in the *Forensics*.

12.2.1 Installing the Log Forwarder

This section explains the specifics of installing the Log Forwarder on different platforms, namely Linux and Windows.

Note: AIX does not support the Log Forwarder. You can install the Log Forwarder either on the Linux machine or on the Windows machine.

12.2.1.1 Installing the Log Forwarder on Linux

The Log Forwarder can be installed through the Linux installer or through a silent installation.

Note:

There is no upper limit to the number of Audit Store endpoints that can be added to Log Forwarder on Linux through the Linux installer or through a silent installation without editing the *upstream_es.cfg* file.

When you install the Log Forwarder, the system automatically sets up a directory structure with the required files in the */opt/protegility* directory.

12.2.1.1 Installation using the Linux Installer

Caution:

If your Linux distribution uses *dash* as the shell, then you must run the installer with *bash* instead. For example,

```
>bash ./LogforwarderSetup_Linux_x64_<version>.sh
```

► To install the Log Forwarder on Linux:

1. Run the *LogforwarderSetup_<os>_<version>.sh* file from the Terminal or double-click the file and select **Run in Terminal**.
2. Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store Port number where the Log forwarder listens for logs.

The added Audit Store endpoint is displayed on the command prompt.

Note: The default port number is *9200*.

3. If you want to add more than one Audit Store endpoints, then press the **Y** key when **Do you want to add another audit store endpoint?** message appears on the command prompt.

The added Audit Store endpoints are displayed on the command prompt.

Note: If you need to add *n* Audit store endpoints, then repeat the Step 2 and Step 3 *n* times.

4. Press the **Y** key to install into the destination directory.

The directories are created under */opt/protegility*, and the required installation files are installed in these directories.

Silent Mode of Installation

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.



Table 12-1: Parameter List for Silent Installation

| Parameter | Description |
|------------------|---|
| --endpoint or -e | IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is 9200. |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the **--dir** parameter to the command to specify the Log Forwarder installation directory and **--pemdir** parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]  
--dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.2.1.2 Installing the Log Forwarder on Windows

The Log Forwarder can be installed through the Windows wizard or through silent installation.

When you install the Log Forwarder, the system automatically sets up a directory structure with the required files in the ..\Protegility\fluent-bit directory.

► To install the Log Forwarder on Windows:

12.2.1.2 Installation using the Windows Wizard

1. Double-click or run the *LogforwarderSetup_<OS>_<version>.exe* file.
2. From the Setup Wizard, click **Next**.
The *ESA Connectivity Information* screen appears.
3. Enter the Audit Store endpoint (IP address:port number).

Note: The default port number is 9200.

Note: If you are using more than one Audit Store, then after the installation completes, you must edit the *upstream_es.cfg* file .

To add ESAs, navigate to the `..\Protegility\fluent-bit\data\config.d` directory, and edit the `upstream_es.cfg` file as follows. The `/NODE` block must be added for each new Audit Store.

```
[NODE]
Name node-1
Host 10.37.4.150
Port 9200
tls on
tls.verify off
Pipeline logs_pipeline
[NODE]
Name node-2
Host 10.37.4.158
Port 9200
tls on
tls.verify off
Pipeline logs_pipeline
```

Table 12-2: Parameters to add a new node

| Parameter | Description |
|------------|---|
| Name | Set a name for the Audit Store |
| Host | IP address or hostname of the Audit Store |
| Port | Set the port number Note: The default port number is <i>9200</i> . |
| tls | Enable or disable TLS support. Set this parameter to <i>on</i> to enable TLS support and to <i>off</i> to disable TLS support. Note: The default tls setting is <i>on</i> . |
| tls.verify | Force certification validation. Set this parameter to <i>on</i> to enforce certificate validation and to <i>off</i> to disable certificate verification. Note: The default tls.verify setting is <i>off</i> . |
| Pipeline | Set a filter for the Audit Store. Note: The default Pipeline setting is <i>logs_pipeline</i> . |

Restart the Log Forwarder service after editing the file.

4. Enter the directory where the PEP server is installed.
5. Click **Next**.
The *Select Destination Location* screen appears.
6. Browse to the directory in which you want to install the Log Forwarder, or retain the default location (recommended).
7. Click **Next**.
The *Ready to Install* screen appears.
8. Click **Install**.
9. From the **Completing the Logforwarder Setup Wizard** screen, click **Finish** to complete the installation and exit. The directories are created under the installation directory that was defined and the installation files are installed in these directories.

Silent Mode of Installation

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.



Table 12-3: Parameter List for Silent Installation

| Parameter | Description |
|------------------------------------|--|
| -endpoint1, -endpoint2, -endpoint3 | <p>Audit Store IP address and the Port number where the Log forwarder listens for logs</p> <p>Note: The default port number is 9200.</p> <p>Note: The parameters <i>-endpoint2</i> and <i>-endpoint3</i> are optional.</p> |
| -dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the ..\Protegility\fluent-bit directory. |
| -pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the ..\Protegility directory. |

At the command prompt, type the following command from the installer directory.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the Log Forwarder installation directory and the *-pemdir* parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>] -dir <Log Forwarder installation directory> -pemdir
<PEP server installation directory>
```

12.2.1.3 Configuring the Log Forwarder for AIX platform

This section explains the specifics of configuring the Log Forwarder for the AIX platform.

The data protectors installed on the AIX platform communicates with the Log Forwarder on the following operating systems:

- Linux machine
- Windows machine

12.2.1.3.1 Configuring the Log Forwarder on the Linux machine for the AIX platform

This section describes the specifics to configure the standalone Log Forwarder on the Linux machine to communicate with the data protectors being installed and configured on the AIX platform.



Before you begin

Ensure that the Log Forwarder installed on the Linux machine and the data protector installed on the AIX platform are able to communicate with each other. If they are not able to communicate, then create a security rule in the firewall setting to allow traffic from the data protector installed on the AIX platform.

Note: For more information about Installing the Log Forwarder on the Linux platform, refer to the section [Installing the Log Forwarder on Linux](#).

► To configure the Log forwarder on the Linux machine:

1. Login and open a CLI on the machine where the Log Forwarder is installed.

2. Navigate to the `config.d` directory using the following command.

```
cd /opt/protegility/fluent-bit/data/config.d
```

3. Backup the existing `in_tcp.conf` file using the following command.

```
cp in_tcp.conf in_tcp.conf_backup
```

4. Update the value of the `Listen` setting in the input block to the `0.0.0.0` as shown in the following code snippet.

Important:

The `0.0.0.0` interface accepts connections from any network interface. To enhance the security, you can update the value of the `Listen` setting to a IP address of a specific network interface.

Note:

The default value of the `Listen` setting is `127.0.0.1`.

```
[INPUT]
  Name      tcp
  Listen    0.0.0.0
  Tag       logdata
  Port      15780
  Format    json
  storage.type filesystem
  Chunk_Size 2048
  Buffer_Size 8192
  Buffer_Max_Size 12M
  Mem_Buf_Limit 64M
  Routes    forward
```

5. Run the following command to start the Log Forwarder service.

```
<INSTALL_DIR>/fluent-bit/bin/logforwarderctrl start
```

6. Login and open a CLI on the Protector machine.
7. Navigate to the `/opt/protegility/defiance_dps/bin/data` directory.
8. Backup the existing `pepserver.conf` file.
9. Navigate to the `Logging Configuration` section in the `pepserver.cfg` file.



10. Update the value of the *host* setting to the *IP address of Linux Machine* where the Log Forwarder is installed and running.

Note:

The default value of the *host* IP address is 127.0.0.1.

Note:

The default value of the *port* number is 15780. Ensure that the value of the *port* setting in the *in_tcp.conf* and the *pepserver.conf* files are the same.

The code snippet is shown here.

```
# Fluentbit host and port values (mostly localhost) where logs will be
# forwarded from the protector.
host = 127.0.0.1
port = 15780
```

11. Start the PEP server installed on the AIX machine.

12.2.1.3.2 Configuring the Log Forwarder on the Windows machine for the AIX platform

This section describes the specifics to configure the standalone Log Forwarder on the Windows machine to communicate with the data protectors being installed and configured on the AIX platform.

Before you begin

Ensure that the Log Forwarder installed on the Windows machine and the data protector installed on the AIX platform are able to communicate with each other. If they are not able to communicate, then create a security rule in the firewall setting to allow traffic from the data protector installed on the AIX platform.

Note: For more information about Installing the Log Forwarder on the Windows platform, refer to the section [Installing the Log Forwarder on Windows](#).

► To configure the Log forwarder on the Windows machine:

1. Login and open a CLI on the machine where the Log Forwarder is installed.
2. Navigate to the *config.d* directory using the following command.

```
cd <LOG_FORWARDER_INSTALL_DIR>\Protegility\fluent-bit\data\config.d
```

3. Backup the existing *in_tcp.conf* file using the following command.

```
xcopy in_tcp.conf in_tcp.conf_backup
```

4. Update the value of the *Listen* setting in the input block to the *0.0.0.0* as shown in the following code snippet.

Important:

The *0.0.0.0* interface accepts connections from any network interface. To enhance the security, you can update the value of the *Listen* setting to a IP address of a specific network interface.

Note:

The default value of the *Listen* setting is 127.0.0.1.

```
[INPUT]
  Name      tcp
  Listen    0.0.0.0
  Tag       logdata
  Port      15780
  storage.type memory
  Chunk_Size 2048
  Buffer_Size 8192
  Buffer_Max_Size 12M
  Mem_Buf_Limit 64M
  Routes    forward
```

- Run the following command to start the Log Forwarder service.

```
net start logforwarder
```

- Login and open a CLI on the Protector machine.
- Navigate to the */opt/protegility/defiance_dps/bin/data* directory.
- Backup the existing *pepserver.conf* file.
- Navigate to the *Logging Configuration* section in the *pepserver.cfg* file.
- Update the value of *host* setting to the *IP address of Windows Machine* where the Log Forwarder is installed and running.

Note:

The default value of the *host* IP address is 127.0.0.1.

The default value of the *port* number is 15780. Ensure that the value of the *port* setting in the *in_tcp.conf* and the *pepserver.conf* files are the same.

The code snippet is shown here.

```
# Fluentbit host and port values (mostly localhost) where logs will be
# forwarded from the protector.

host = 127.0.0.1
port = 15780
```

- Start the PEP server installed on the AIX machine.

12.2.2 Log Forwarder Configurations

After the Log Forwarder is installed, you can configure how logs must be handled in a situation where the connection to the Log Forwarder in the protector is lost.

For more information about configuring the Log Forwarder, refer to the section [Appendix A: PEP Server Configuration File](#).

12.2.3 Running the Log Forwarder

After the Log Forwarder is installed and configuration settings are done, you are ready to start the Log Forwarder service.

Caution:

Ensure that the *logforwarder* service is started before starting the *PEP server* service.

Table 12-4: Steps for running Log Forwarder in Windows

| To... | Perform the following ... |
|--|--|
| Start or stop the Log Forwarder as a Windows service | On the command line, run: <code>net start logforwarder</code> <code>net stop logforwarder</code> |
| Start/stop the Log Forwarder using Windows Control Panel | Navigate to Windows Control Panel > Administrative Tools > Services > logforwarder . |

Table 12-5: Steps for running Log Forwarder in Linux

| To... | Perform the following ... |
|--|---|
| Start/stop the Log Forwarder as a daemon | On the command line, run: <code><INSTALL_DIR>/fluent-bit/bin/ logforwarderctrl start</code> <code><INSTALL_DIR>/fluent-bit/bin/ logforwarderctrl stop</code> You can add a script that calls the daemon. It will be called automatically at system start-up. |
| View status of the Log Forwarder daemon | On the console window, run: <code><INSTALL_DIR>/fluent-bit/bin/ logforwarderctrl status</code> |

12.3 Installing and Uninstalling the PEP Server

This section explains the generic procedures how to install, configure and uninstall PEP Server.

12.3.1 PEP Server Pre-Installation Preparation

Before setting up the PEP server, ensure that your environment meets the minimum requirements.



12.3.1.1 Disk Space Requirements

Table 12-6: Disk Space Requirements

| Configuration | Free Disk Space on Windows | Free Disk Space on Unix | Notes |
|--|----------------------------|-------------------------|---|
| PEP Server | 20MB per node | 10MB per node | For systems under heavy load with auditing turned on, the space required for the PEP Server (20 MB/node) can be higher. |
| User defined functions (UDFs) and procedures | 10MB | 10MB | |
| Total | Min. 140MB | Min. 125MB | |

12.3.1.2 Shared Memory Requirements

Some operating systems have a maximum limit on the size that a software program can create, which is commonly referred to as the *shmmmax* or *kernel.shmmmax* limit.

The maximum shared memory (shmax) is recommended to be 512 MB, which is greater than the space that the largest token table (402653184 bytes) will occupy when stored in the shared memory. If the default shared memory of your OS is less than the largest token table size, then you should manually change it.

The commands to view and set the limits differ between operating systems. Thus, for example, on Solaris, the system kernel parameters can be read/modified using the *sysctl* program.

You should change the *kernel.shmmmax* parameter in the *sysctl.conf* file available in the */etc* directory, and then restart the system. Alternatively, use the following commands:

sysctl -w kernel.shmmmax=<value> to write in *sysctl.conf*

sysctl -p /etc/sysctl.conf to read or reload the values from *sysctl.conf*

12.3.1.3 Stack Memory Requirements

Some operating systems limit the amount of stack and data space that an application can allocate when being executed.

Thus, by default, AIX installations have limited values on the stack and the data memory resource limits even for the root account login on a console. Such default limitations of the system may lead to problems with deployment of big token tables. To avoid problems while deploying token tables, you need to manually increase the stack memory limit.

You can increase the stack memory by creating a separate system account, for example *protegility*, for the execution of the PEP server.

Note: The users whose security settings are affected, must log off and log in before the settings are applied.

Then you will need to modify the *Limits* file in the */etc/security/* directory with the following entries:

```
protegility:
data = -1
stack = -1
```

12.3.1.4 Hostname Localhost Verification

The PEP server requires the localhost name value to be set to start correctly. The IP address of the localhost could either be *127.0.0.1* or the system IP. If the IP address is neither of these, then the value must be updated.

The localhost value can be checked on most distributions of Unix with the host command:

```
host localhost
```

If you receive an error message, such as Hostname lookup failure or Unknown, instead of *127.0.0.1* or the system IP address as the localhost value, then perform the following steps.

1. Remove the IP address value from the DNS.
2. Configure the localhost by adding the correct IP address to */etc/hosts* directory.

If the DNS record of a faulty localhost value cannot be removed or changed, then modify the order of resolving hostnames. For example, for AIX, modify the order in the file */etc/netsvc.conf* to ensure that the local definition in */etc/hosts* is checked first.

Note:

In AIX platforms, export the *LDR_CNTRL* variable to the *.profile* directory. The following syntax explains the command to export the *LDR_CNTRL* variable.

For P5 versions:

```
export LDR_CNTRL=MAXDATA=0x5000000
```

For P6 versions:

```
export LDR_CNTRL=LARGE_PAGE_TEXT=Y@LARGE_PAGE_DATA=M
```

12.3.2 Installing the PEP Server

This section explains the specifics of installation of PEP server on different platforms, such as, Windows and UNIX.

12.3.2.1 PEP Server on Windows

When you install the PEP server, the system automatically sets up a directory structure with the required files.

The following table describes the default directory structure.

Table 12-7: Default Installation Directory Structure

| Directory | Description |
|----------------------------|---|
| .. \Protegity\Defiance DPS | The default installation directory which includes the <i>\bin</i> and <i>\data</i> directories. |
| \bin | Includes the executable files of the PEP server. |
| \data | Includes the PEP server configuration file, log file, and certificates. |

► To install the PEP server on Windows:

1. Double-click or run the *PepServerSetup_<OS>_<Version>.exe* file.

2. From the Setup Wizard, click **Next**.
The *ESA Connectivity Information* screen appears.
3. Enter the ESA IP Address, ESA username, and ESA Password to download certificates.
4. Click **Next**.
The *Select Destination Location* screen appears.
5. Browse to the directory to which you want to install the Protegility PEP server, or leave the default location (recommended).
6. Click **Next**.
The *Ready to Install* screen appears.
7. Click **Install**.
8. From the **Completing the PEP Server Wizard** screen, click **Finish** to complete the installation and exit. Directories are created under the installation directory you specified and the installation files are installed in these directories.

Note: If you require the PEP server service to start automatically after every restart of the system, then define the PEP server service in the startup with the required run levels.

Silent Mode of Installation

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.

Table 12-8: Parameter List for Silent Installation

| Parameter | Description |
|-----------|--|
| -esa | Specifies the ESA IP address. |
| -esaport | Specifies the ESA port, which is optional. The default value is 8443. |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the.. Protegility Defiance DPS directory. |

At the command prompt, type the following command from the installer directory.

```
PepServerSetup_Windows_x64.exe -certuser <username> -esa <esaIP> -certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
PepServerSetup_Windows_x64.exe -certuser <username> -esa <esaIP> -certpw <password>
-dir <installation-directory-path>
```

12.3.2.2 PEP Server on UNIX

When you install the PEP Server, the system automatically creates the necessary directories in the directory you specify during installation.

The following table describes these default directories.

Table 12-9: Default Installation Directory Structure

| Directory | Description |
|-------------------------------|---|
| /opt/protegility/defiance_dps | The default installation directory which includes the \bin and \data directories. |



| Directory | Description |
|-----------|---|
| /bin | Includes the executable files of the Protegity PEP server. |
| /data | Includes the PEP server configuration file, log file, and certificates. |

Caution:

If your Linux distribution uses *dash* as shell, then you must run the installer with bash instead. For example,

```
>bash ./PepServerSetup_Linux_x64_<version>.sh
```

► To install the PEP Server on UNIX:

1. Run the *pepper_server_setup_<os>_<version>.sh* file from the Terminal or double-click the file and select Run in Terminal.
2. Enter the ESA Host Name or IP Address.
The ESA host will be written to the *pepper_server.cfg* file.
3. Press ENTER to install into the destination directory.
Directories are created under */opt/protegity/defiance_dps*, and the required installation files are installed in these directories.

Caution: Ensure that ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

4. Enter the user name for downloading the certificates.
5. Press ENTER.
6. Enter the required password for downloading the certificates.
7. Press ENTER.
8. Open the *pepper_server.cfg* file located in the *opt/protegity/defiance_dps/data* directory and configure the communication ID.

If you are installing a Proxy node for a multi-node Teradata environment, then you need to enable proxies. This option is available only for the PEP server installation on Linux x64.

For more information, refer to *Protegity Database Protector Guide 9.1.0.0*. The following scenario explains the installation specifics for installing these Proxies.

12.3.3 Configuring PEP Server

After the PEP server is installed, it should be configured for your specific installation. This section explains main PEP server configuration parameters and their recommended settings.

12.3.3.1 Customization of *pepper_server.cfg* File

The *pepper_server.cfg* file is the configuration file for the PEP server. It is installed in the specified *defiance_dps/data* directory. If required, then customize this file after installation and before the PEP Server starts. The following table describes sections that are available in the *pepper_server.cfg* file.

The format of the file is a standard UNIX configuration format. A number sign (#) starts a comment, so typing this sign as the first character of a line makes it a comment line.

For more information on the *pepsvcfg.cfg* configuration file, refer to section [Appendix: PEP Server Configuration File](#).

12.3.3.2 PEP Server Configuration Parameters

The following table describes sections that are available in the *pepsvcfg.cfg* file.

Table 12-10: Configuration Parameters

| Section | Description |
|-----------------------|---|
| Application | Specifies the location where PEP server saves the temporary files. Defines whether to add the PEP server's IP address to request headers which is required when the PEP server is communicating with the ESA via a proxy. Specifies the communication ID (Teradata, SQL Server or Oracle, DB2) to use and specify a path to the post-deploy script. |
| Logging | Defines the level of logging and the file to write logs. Defines whether to append logs to existing log file or recreate log file after restarting the PEP server. |
| Policy Management | Defines where the private keys for the certificates should be located. Specifies the interval (in seconds) to refresh the policies. Defines what value to return, if data to protect is an empty string. |
| Application Protector | Specifies the listener port for Application Protector Client/Server. |
| Audit | Defines where the private keys for the certificates should be located. Specifies if logs should be sent to a log server. If yes, then logs will be sent to the specified host and port. If no, then logs will be stored locally and will not be sent to any log server. It informs the PEP to log every callout instead of one per SQL statement. |
| Administration | Specifies the listener port for the administration interface. |
| Member | Specifies how policy users are checked against policy. Configure the case-sensitivity of the policy users. |

12.3.3.3 Configuring the PEP Server to start automatically

If you require the PEP server service to start automatically after every restart of the system, then define the PEP server service in the start-up.

► To configure the PEP server to Start Automatically after a System restart:

1. Login to the *root* user account.
2. Copy the PEP Server control script file (*pepsrvctrl.sh*) from the <PROTEGILITY_DIR>/defiance_dps/bin/*pepsrvctrl* directory to /etc/init.d/ directory with the name *pepsvcfg* using the following command.
`opt/protegility/defiance_dps/bin/pepsrvctrl /etc/init.d/pepsvcfg`
3. Navigate to the /etc/init.d directory.
4. Add the execute permissions for the PEP Server control script file using the following command.
`chmod +x pepsvcfg`
5. Add the PEP Server control script file in the *chkconfig* utility using the following command.
`sudo chkconfig --add pepsvcfg`
6. If required, then verify whether the PEP Server control script file is included in the *chkconfig* utility using the following command.
`sudo chkconfig --list pepsvcfg`
7. Set the PEP Server to start automatically after a system reboot using the following command.
`sudo chkconfig pepsvcfg on`
8. If you are uninstalling the protector, then remove the PEP Server control script from the /etc/init.d/ directory.

12.3.4 Configuring the PEP Server for AIX platform

Perform the following steps to install and configure the standalone PEP server to communicate with the data protectors being installed and configured on the AIX platform.

► To configure the PEP Server for AIX platform:

1. Login and open a CLI on the AIX protector machine where the PEP server is installed.
2. To navigate to the `/opt/protegility/defiance_dps/data` directory, run the following command.

```
cd /opt/protegility/defiance_dps/data
```

3. To backup the existing `pepserver.conf` file, run the following command.

```
cp pepsolver.conf pepsolver.conf_backup
```

4. To edit the `pepsolver.cfg` file, and update the value of the `host` parameter.

```
vim pepsolver.cfg
```

5. Navigate to the *Logging Configuration* section in the `pepsolver.cfg` file.

Note:

- The default value of the `host` IP address is 127.0.0.1.
- For installing the protector on AIX platform, set the value of the `host` parameter with the IP address of the linux or windows machine where the Log Forwarder is installed.
- The default value of the `port` number is 15780. Ensure that the value of the `port` setting in the `in_tcp.conf` and the `pepsolver.conf` files are the same.

The code snippet is shown here.

```
# Fluentbit host and port values (mostly localhost) where logs will be
forwarded from the protector.
host = 127.0.0.1
port = 15780
```

6. To start the PEP server installed on the AIX machine, run the following command.

```
./pepsrvctrl start
```

12.3.5 Synchronizing Certificate Files

As a post-installation task, you need to synchronize ESA certificates with the installed and configured PEP Server. This section explains how to do that on different platforms.

PEP Server installer on all platforms, by default includes a Curl tool. Curl is used to download the certificate files from ESA to the PEP Server. With Curl, the files are stored centrally on the appliance in their own `.tgz` archives, since the appliance is the master manager and storehouse for the files.

The Curl script that facilitate synchronization is available for Windows and UNIX operating systems. The following table describes the script.

Table 12-11: Curl Script Files

| File | Description |
|---|--|
| <i>GetCertificates.bat</i> (Windows) | Compares certificates between the Appliance and PEP Server and synchronizes them if necessary. |
| <i>GetCertificates.sh</i> (UNIX) | |

The following sections describe how to use the Curl tool with UNIX and Windows.

Note: When you login to ESA for the first time, ensure that you change the password before running the *GetCertificates* command.

12.3.5.1 Retrieving Certificate Files

After installing the protector, you run the script to retrieve the certificates. This ensures the certificates are synchronized between the PEP Server and ESA. The command syntax is the same for Windows and UNIX.

The following example shows how to retrieve the ESA certificates as user admin on Windows:

GetCertificates.bat -u admin

From the *bin* directory of your Protegility Data Security Platform installation, run

GetCertificates.ext -u username where:

| | |
|-----------------|--|
| <i>Ext</i> | The file extension, either <i>bat</i> (for Windows) or <i>sh</i> (for UNIX). |
| <i>Username</i> | The user identifier for logon. |

Note: If you type the command with incorrect arguments or no arguments, then the system displays Help for the command.

12.3.5.2 Retrieving Certificate Files Examples

This section shows some examples to retrieve certificate files. On a PEP Server, you typically download certificates from ESA. The following example shows the screen display from the sequence of commands.

```
C:\Program Files\Protegility\Defiance DPS\bin>GetCertificates.bat -u admin
Downloading certificates from 10.10.97.15:8443...
Enter host password for user 'admin':
% Total    % Received % Xferd  Average Speed   Time      Time     Current
          Dload  Upload Total   Spent    Left   Speed
100 20480  100 20480    0      0 26086      0 --::-- --::-- --::-- 27489
Extracting certificates...
Certificates successfully downloaded and stored in /opt/protegility/defiance_dps/data.
```



The following example shows a sample of an error message if a problem is encountered while running the *GetCertificates.bat* command.

```
-----
C:\Program Files\Protegility\Defiance DPS\bin>GetCertificates.bat -u admin
Enter host password for user 'admin':
 % Total    % Received % Xferd  Average Speed   Time     Time      Time  Current
               Dload  Upload   Total   Spent    Left  Speed
0       0     0      0       0        0      0 --:--:-- --:--:-- --:--:--     0
Failed to download certificates
HTTP response code: 401
```

12.3.6 Running the PEP Server

After the PEP Server is installed and configuration settings are done, you are ready to run the PEP Server.

Note: If you have entered incorrect information in the *pepserver.cfg* file, then you will encounter errors when you try to start the PEP Server. The errors will be logged in the *pepserver.log* file.

You can run the PEP Server as a Windows service, in a console window, or as a UNIX or Linux daemon. The table below explains how to do that.

Table 12-12: Steps for running PEP Server

| To... | Perform the following ... |
|--|--|
| Start or stop the PEP Server as a Windows service. | <p><i>In the command line, run:</i></p> <pre>net start <Defiance PEP Server> net stop <Defiance PEP Server></pre> <p>(where “Defiance PEP Server” is the PEP Server default name)</p> |
| Start/stop the PEP Server using Windows Control Panel | Navigate to Windows Control Panel > Administrative Tools > Services > Defiance PEP Server, start/stop. |
| Start the PEP Server in a command line on UNIX/Linux/Windows | <p>Put in the PEP Server setup file into <<i>INSTALL_DIR</i>>/</p> <p><i>In the command line, run:</i></p> <pre>pepserver -verbose -dir <INSTALL_DIR>/data</pre> |
| Start/stop the PEP Server as a daemon | <p><i>In the command line, run:</i></p> <pre><INSTALL_DIR>/defiance_dps/bin/pepsrvctrl start <INSTALL_DIR>/defiance_dps/bin/pepsrvctrl stop</pre> <p>You can add a script that calls the daemon. It will be called automatically at system start-up.</p> |
| View status of the PEP Server daemon | In the console window, run: |



| To... | Perform the following ... |
|-------|--|
| | <code><INSTALL_DIR>/defiance_dps/pepserver/ pepsrvctrl status</code> |

Note: If the PEP Server is stopped incorrectly, or if it goes down, then the policy that exists in the shared memory is removed. If you terminate PEP Server process, then the PEP Server shared memory needs to be removed manually (use `ipcrm -m id` command).

12.3.7 Uninstalling the PEP Server from Unix or Linux Platforms

The uninstallation is provided for both *nix and Windows separately.

► To uninstall from Unix/Linux/*nix platforms

If the PEP Server is stopped, then run the following command to delete the directory.

`rm -rf /opt/protegrity/defiance_dps`

12.3.8 Uninstalling the PEP Server from Windows Platform

► To uninstall the PEP server on the Windows platform:

Depending on the directory where the PEP server is installed, run one of the following executable commands:

1. If the PEP server is installed in *C:\Program Files\Protegrity\Defiance DPS*, then run the following executable file.
C:\Program Files\Protegrity\Defiance DPS\unins000.exe
2. If the PEP server is installed in *C:\Program Files(x86)\Protegrity\Defiance DPS*, then run the following executable file.
C:\Program Files(x86)\Protegrity\Defiance DPS\unins000.exe

This uninstalls the PEP server.

12.4 Protector Proxy Installation

This section explains the Protector Proxy, generic procedures how to install, run, configure and uninstall the Protector Proxy.

12.4.1 Introduction

ESA 6.6.x provided the use of a proxy (called DPS Proxy) for deployments where protector nodes could not communicate with ESA directly due to network segmentation and thus had to communicate with ESA via the DPS proxy.

Starting from ESA 7.0, the DPS proxy has been enhanced to improve scalability to support more nodes communicating with ESA via the proxy that is now renamed to Protector Proxy.

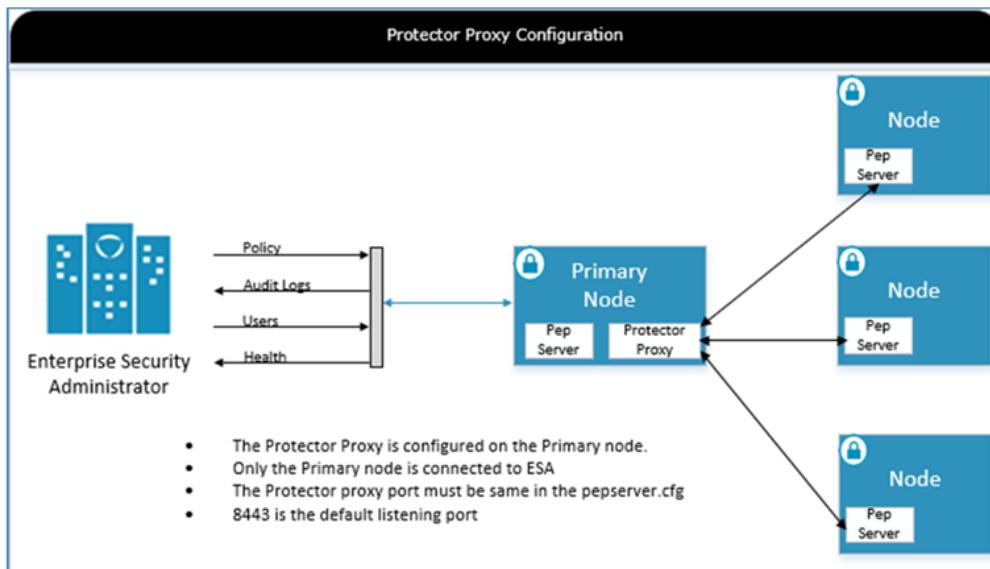


Figure 12-2: Protector Proxy Configuration

Deployed between ESA and the protectors, the Protector proxy supports all the communication between ESA and a protector for downloading policies, users, sending audit logs and health information.

The Protector proxy is a standalone component, packaged along with the protector archive. When the protector archive is extracted it will have a separate installable for the protector proxy.

For ESA 7.1, the protector proxy is bundled with Big Data, Teradata, and Netezza protectors, only for Linux platforms.

12.4.2 Installing and Uninstalling Protector Proxy

This section explains installing the Protector proxy when ESA is initially installed and is being set up.

12.4.2.1 Installing the Protector Proxy

► To Install the Protector Proxy:

Execute the following script.

```
./ProxySetup_Linux_x64_<version>.sh -esa <ESA IP>
```

For example, `./ProxySetup_Linux_x64_7.2.0.x.sh -esa <ESA IP>`

The Proxy is installed successfully in the `/opt/protegility` directory where the default port is 8443.

Note:

You can use the following command to install the Protector proxy using a different port and in a different directory.

```
For example, ./ProxySetup_Linux_x64_7.2.0.x.sh -esa <ESA IP> -port <PORT> -dir <DIRECTORY>
```

12.4.2.2 Running the Protector Proxy

► To run the Protector Proxy:

1. Go to `/opt/protegriity/proxy/bin` to start the Proxy.
2. Run the following command to run the Protector proxy in background.

`./proxyctrl start`

In this case, the value inside the `/opt/protegriity/proxy/bin/daemon.cfg` is ON.

The Protector Proxy starts.

3. Run the following command to run the Protector proxy in foreground.

`./proxyctrl start -fg -`

In this case, the value inside the `/opt/protegriity/proxy/conf/daemon.cfg` is OFF.

Note: Running the Protector Proxy in foreground is applicable in case of Cloudera because cloudera daemons usually run in foreground.

The Proxy starts.

Note: Ensure that the port that you are using for proxy is not used by other services. If there is any software which is already using port 8443 on the proxy machine, then reconfigure the proxy to use some other port.

12.4.2.3 Configuring the Protector Proxy

The following snippet shows the default configuration of the `proxy.cfg` file.

```
include daemon.cfg;

.....
.....
.....

stream {
    server {
        listen                8443;
        proxy_pass            esa_backend;
        proxy_timeout         10s;
        proxy_connect_timeout 30s;

        # proxy_bind <some_ip>:8443;
        # proxy_buffer_size 16k;
    }

    upstream esa_backend {
        server <esa_host>:8443 max_fails=3 fail_timeout=30s;
    }
}
```

The following table lists some important configurable parameters in the `proxy.cfg` file. Users can change these values, as required.

| Parameter | Default value | Details |
|-------------------------------------|----------------------|--|
| <i>server/listen</i> | 8443 | Specifies the listening port of the proxy server. If you want to use another port as the listening port, then you can modify the <i>proxy.cfg</i> file in the <i>/opt/protegilityproxy/conf</i> directory. |
| <i>server/proxy_pass</i> | <i>esa_backend</i> | <p>Specifies the address of the proxied server, where the Protector Proxy forwards the request.</p> <p>You can also choose to specify a variable as the parameter value. For example, you can specify the value of the parameter as <i>esa_backend</i>. You can then specify the same variable as the name of the <i>upstream</i> block. This enables you to specify a list of servers as the proxied servers, within the <i>upstream</i> block.</p> <p>In the default configuration of the <i>proxy.cfg</i> file, only one ESA is specified under the <i>upstream</i> block. However, you can also specify multiple ESAs.</p> |
| <i>server/proxy_timeout</i> | 10s | Specifies the time period (in seconds) between two successive read or write operations between the Protector Proxy and the client (Protector) or between the Protector Proxy and the proxied server (ESA), after which the connection times out. |
| <i>server/proxy_connect_timeout</i> | 30s | Specifies the time period (in seconds) to establish a connection between the Protector Proxy and the proxied server, after which the connection times out. |
| <i>upstream/server</i> | <i>esa_host:8443</i> | <p>Specifies the hostname or IP address of the ESA.</p> <p>The <i>esa_host</i> value is replaced with the hostname or the IP address of the ESA provided by the user during the installation of the Protector Proxy.</p> <p>You can modify the <i>proxy.config</i> file to specify multiple server values.</p> |
| <i>upstream/maxfails</i> | 3 | Specifies the maximum number of attempts allowed to establish a communication between the Protector Proxy and the ESA. The consecutive attempts must happen within the time period specified by the <i>fail_timeout</i> parameter. |
| <i>upstream/fail_timeout</i> | 30s | Specifies the time period (in seconds) within which consecutive attempts must be made. The Protector Proxy will attempt to establish a communication with the ESA if fails to establish a connection on the first attempt. |



12.4.2.4 Uninstalling the Protector Proxy

► To uninstall the Protector Proxy:

1. Go to `/opt/protegility/proxy/bin`.
2. To stop the Protector proxy, run the following command.
`./proxyctrl stop`
The Protector Proxy stops.
3. Go to `/opt/protegility/proxy` directory and delete the directory.

12.5 Installing and Uninstalling Application Protectors

This section describes the procedures to install and uninstall the Application Protectors.

12.5.1 Installing Application Protectors C

12.5.1.1 Setting up Application Protector C on Linux or Unix

This section describes how to install the Application Protector (AP) C on a Linux or Unix platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP C:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

► To setup the AP C on the Linux or Unix platform:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_GCC-3.2.3_<version>.tgz* file to any location on the machine where you want to install the protector.
2. Extract the AP C installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_GCC-3.2.3_<version>.tgz
```

The following setup files are extracted:

- *XCDevSetup_Linux_x64_<version>.sh*
- *XCSamplesSetup_Linux_x64_<version>.sh*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*



12.5.1.1.1 Installing Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.1.1.1.1 Installing Log Forwarder on Linux or Unix Using Linux Installer

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer.

► To install the Log Forwarder on a Linux or Unix platform using the Linux installer:

1. Run the Log Forwarder installer using the following command.

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store port number where the Log Forwarder listens for logs.

Note: The default port number is *15780*.

3. Press ENTER.

The added Audit Store endpoint appears on the screen.

The prompt to enter an additional Audit Store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

4. If you want to add more than one Audit Store endpoint, then type *y* else type *n*.

Note: If you need to add *n* Audit Store endpoints, then repeat the *Step 2* and *Step 3* *n* times.

5. Type the *y* key to install into the destination directory.

The Log Forwarder is installed in the */opt/protegility/fluent-bit/* directory.

6. Start the Log Forwarder component by using the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl start
```

The Log Forwarder is successfully installed.

12.5.1.1.1.2 Silent Mode of Installation of Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-13: Parameter List for Silent Installation

| Parameter | Description |
|------------------|---|
| --endpoint or -e | The IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is 15780. |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the **--dir** parameter to the command to specify the Log Forwarder installation directory and **--pemdir** parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>] --dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.5.1.1.2 Installing PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.1.1.2.1 Installing PEP Server on Linux or Unix using Linux Installer

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer.

► To install the PEP server on a Linux or Unix platform:

- Run the PEP server installer using the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

- Enter the ESA Host Name or IP Address.
- Press ENTER.



The prompt to enter the username for downloading certificates appears.

```
Please enter the user name for downloading certificates:
```

4. Enter the username for downloading the certificates.
5. Press ENTER.

The prompt to enter the password for downloading the certificates appears.

```
Please enter the password for downloading certificates:
```

6. Press ENTER to install into the destination directory.

Directories are created under `/opt/protegility/defiance_dps` by default, and the required installation files are installed in these directories.

Caution: Ensure that the ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

To manually install the certificates to the `/opt/protegility/defiance_dps/data` directory of the PEP server, navigate to the `/opt/protegility/defiance_dps/bin` directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the *administrator* user.

Verify that the following files have been copied to the `/opt/protegility/defiance_dps/data` directory:

- `CA.pem`
- `keyinternal.plm`
- `pepperver.cfg`
- `pepperver.pid`
- `authesa.plm`
- `cert.key`
- `cert.pem`
- `certkeyup.bin`

7. Start the PEP server by using the following command.

```
/opt/protegility/defiance_dps/bin/pepsrvctrl start
```

The PEP server is successfully installed.

12.5.1.1.2.2 Silent Mode of Installation of PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform through the Silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-14: Parameter List for Silent Installation

| Parameter | Description |
|-----------------------|---|
| <code>-esa</code> | Specifies the ESA IP address. |
| <code>-esaport</code> | Specifies the ESA port, which is optional. The default value is <code>8443</code> . |

| Parameter | Description |
|-----------|--|
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |

At the command prompt, type the following command from the installer directory.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password> -dir <installation-directory-path>
```

12.5.1.1.3 Installing Application Protector C on Linux or Unix using Linux Installer

This section describes how to install the AP C on a Linux or Unix platform using the Linux installer.

► To install the AP C on the Linux or Unix platform using the Linux installer:

1. Run the AP C installer using the following command.

```
./XCDevSetup_Linux_x64_<version>.sh
```

The prompt to continue the installation appears.

This will install XCDev on your computer.

Do you want to continue? [yes or no]

2. If you want to continue with the installation of the XCDev, then type *yes* else type *no*.

If you type *yes*, then the prompt to enter the installation directory appears.

```
Please enter installation directory
[/opt/protegility]:
```

The XCDev is installed in /opt/protegility/defiance_xc by default.

If you type *no*, then the installation of the AP C aborts.

Note:

The *xcpep.plm*, *xcclient.plm*, and *xclite.plm* files are installed in the /opt/protegility/defiance_xc/bin directory.

The *.plm* files are shared libraries that are required for compiling the applications which use AP C APIs.

Link your application to the required *.plm* file, depending on the Application Protector you want to use.

For users working with *xcclient*, Protegility recommends using the AP REST container.

For users working with *xelite*, Protegility recommends using the Immutable AP C.

3. Run the AP C sample installer using the following command.

```
./XCSamplesSetup_Linux_x64_<version>.sh
```

Note:

C Sharp is not a part of Protegility's Application Protector Suite.

A C Sharp sample code, *BulkProtect*, is provided as part of the Application Protector (AP) C installable package as a reference.

For more information about configuring and running the sample code, refer to the section *BulkProtect Function in C Sharp* in the [Application Protector Guide 9.1.0.0](#).

4. If you want to continue with the installation of the XCSamples, then type *yes* else type *no*.

If you type *yes*, then the prompt to enter the installation directory appears.

```
Please enter installation directory  
[/opt/protegility]:
```

The XCSamples are installed in */opt/protegility/defiance_xc* by default.

If you type *no*, then the installation of the XCSamples aborts.

The following samples are installed in the */opt/protegility/defiance_xc/samples/* directory:

- *xcapcsample/*
- *xccsharpsample/*

For more information about the samples, refer to the *readme.txt* file located in the respective sample directories.

The AP C is successfully installed.

12.5.1.2 Setting up Application Protector C on Windows

This section describes how to install the AP C on a Windows platform.

Before you begin

Ensure that the following prerequisites are met before installing Application Protector:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.2](#).

► To setup the AP C on the Windows platform:

1. Download the *ApplicationProtector_WIN-ALL-64_x86-64_VS-2K16_<version>.zip* installation package to any location on the machine where you want to install the protector.
2. Extract the files from the *ApplicationProtector_WIN-ALL-64_x86-64_VS-2K16_<version>.zip* installation package.

The following setup files are extracted:

- *LogforwarderSetup_Windows_x64_<version>.exe*
- *PepServerSetup_Windows_x64_<version>.exe*
- *XCDevSetup_Windows_x64_<version>.exe*
- *XCSamplesSetup_Windows_x64_<version>.exe*

12.5.1.2.1 Installing Log Forwarder on Windows

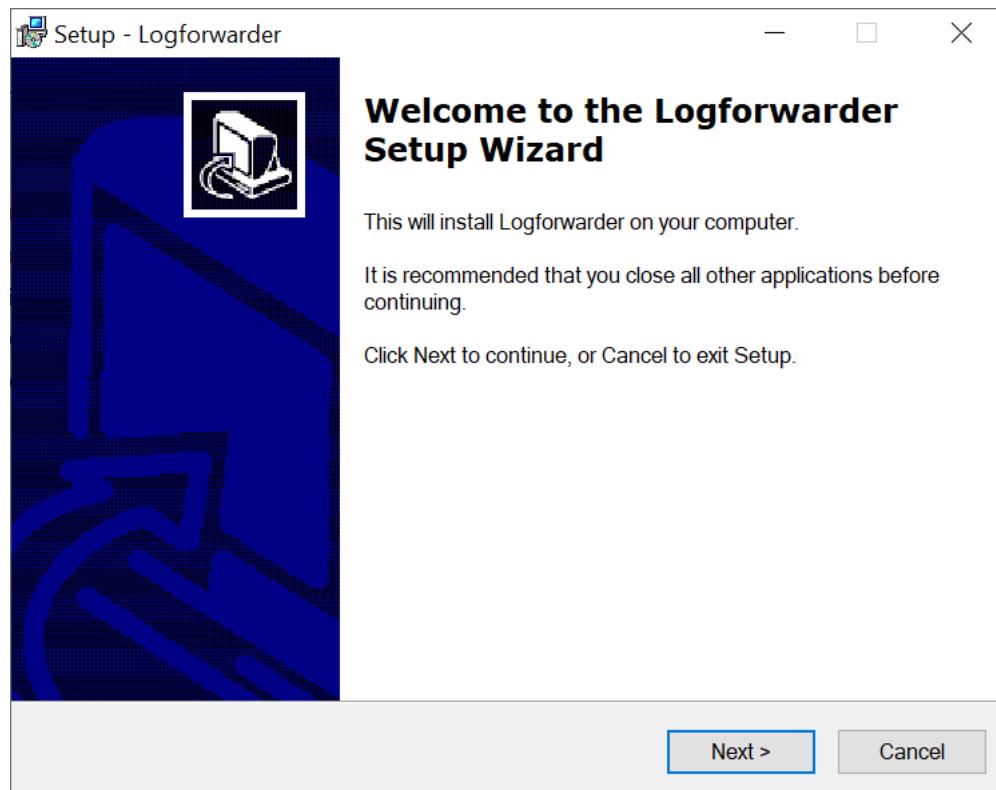
This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard and through the Silent mode of installation.

12.5.1.2.1.1 Using Windows Wizard

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard.

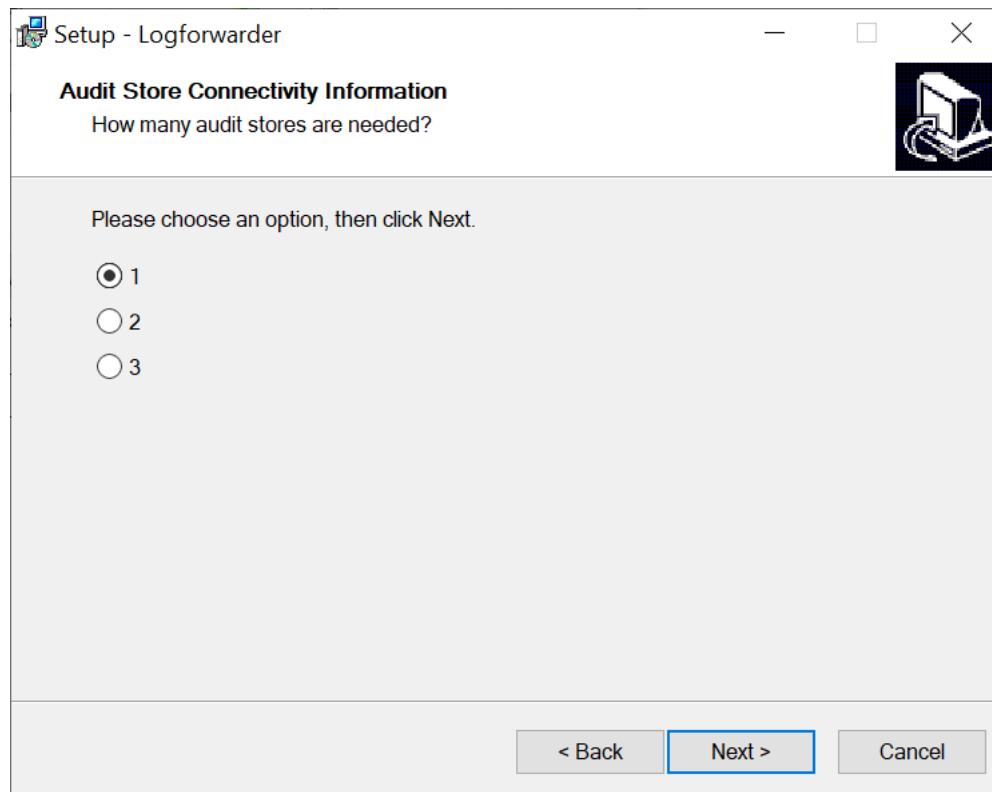
► To install the Log Forwarder on the Windows platform through the Windows Wizard:

1. Run the *LogforwarderSetup_Windows_x64_<version>.exe* file from the created directory.
The **Welcome to the Log Forwarder Setup Wizard** screen appears.



2. Click **Next**.

The **Audit Store Connectivity Information** screen appears to select the number of audit stores that are needed.

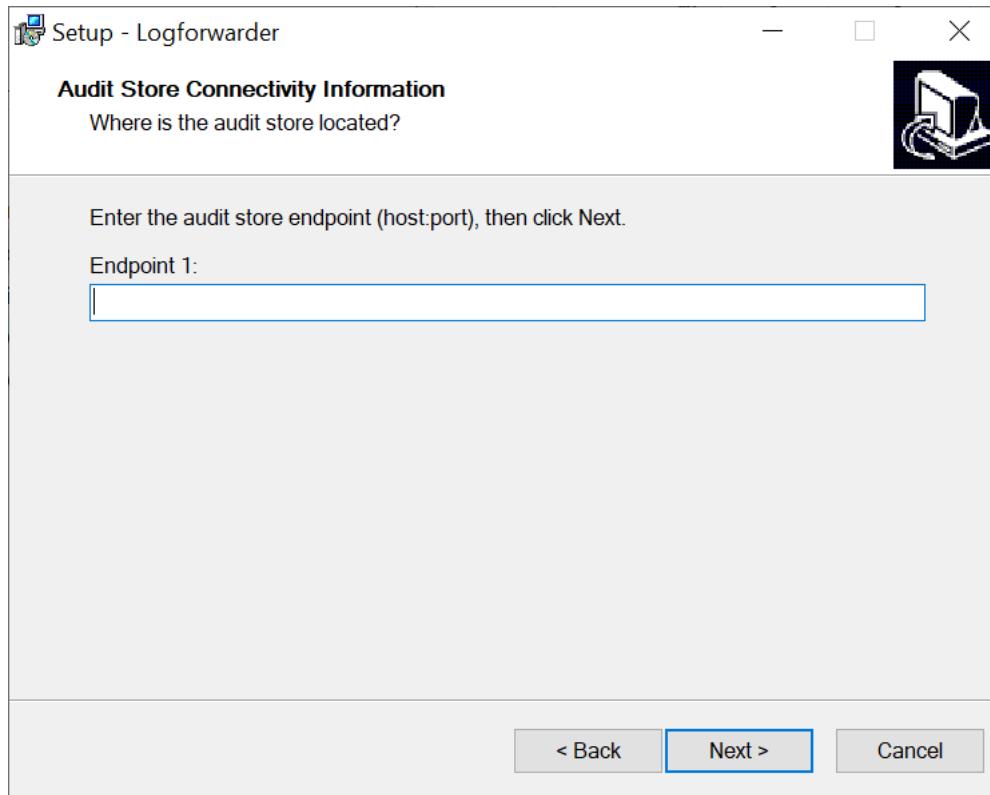


3. Select the number of Audit Stores as required.

The maximum number of Audit Stores that can be configured is 3.

4. Click **Next**.

The **Audit Connectivity Information** screen appears to enter the location of the audit store.



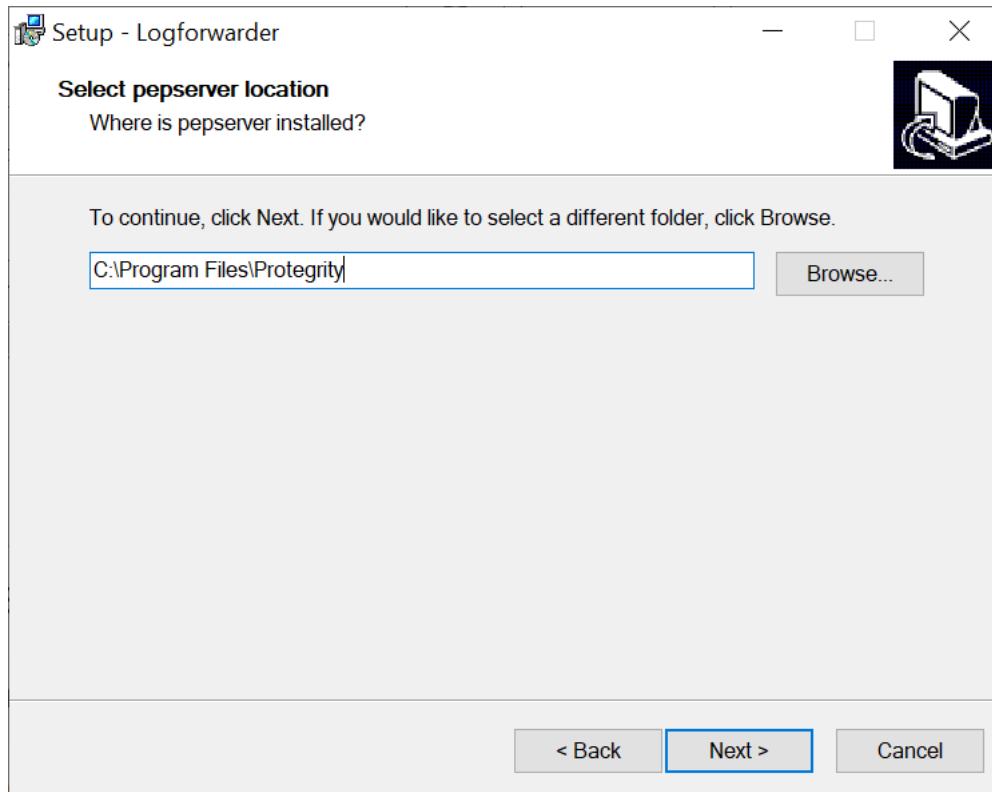
5. Enter the Hostname/IPAddress:port where the Audit Store is configured.

Note:

The default port number for the Audit Store is *9200*.

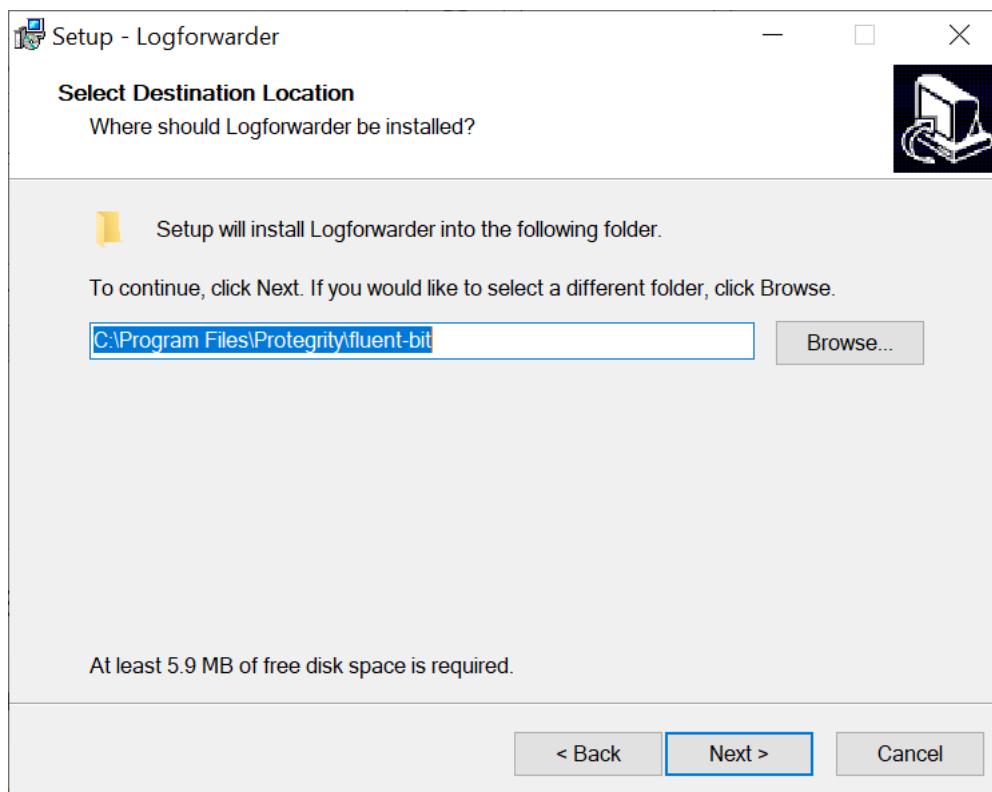
6. Click **Next**.

The **Select pepserver location** screen appears.



7. Click **Next**.

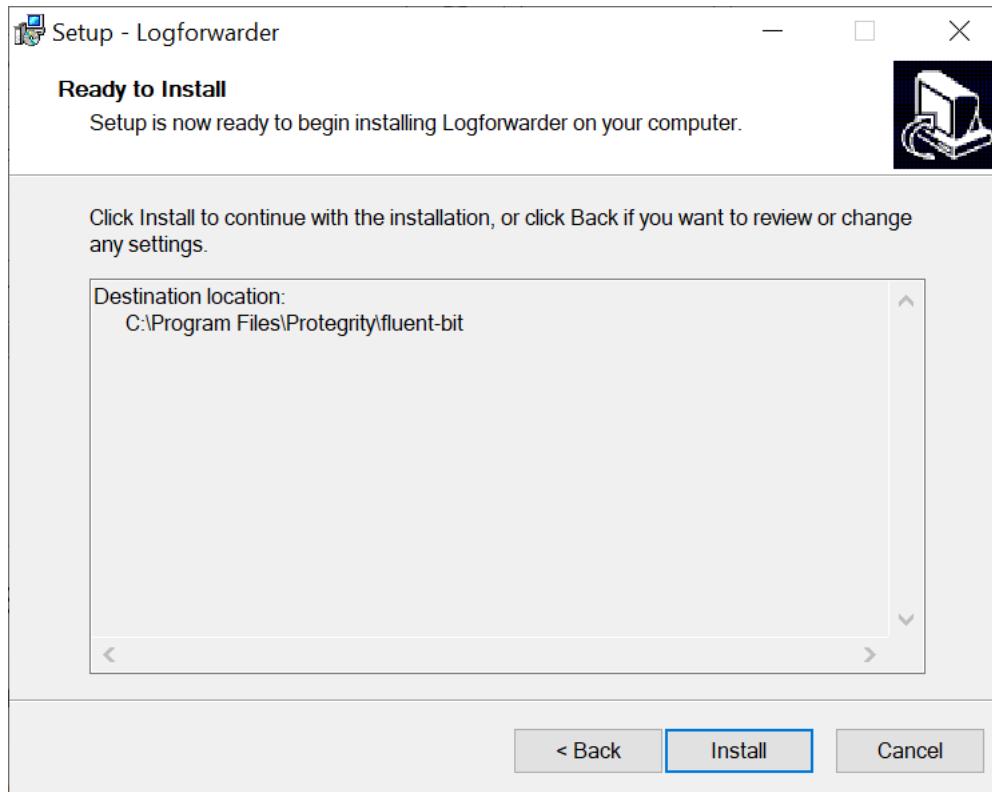
The **Select Destination Location** screen appears.



8. Set the installation directory for the Log Forwarder to *C:\Program Files\Protegility\fluent-bit*.

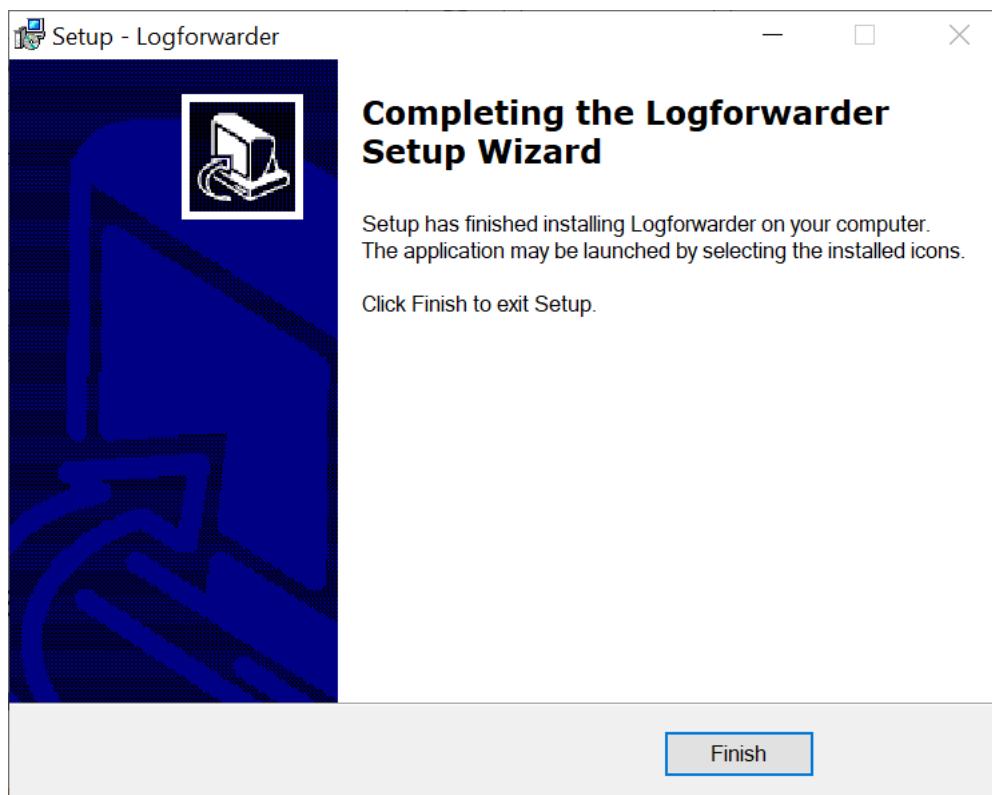
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Logforwarder Setup Wizard** screen appears.



11. From the **Completing the Logforwarder Setup Wizard** screen, click **Finish** to complete the installation and exit.

Post successful installation, you can find the installation files in the directories that are created under the defined installation directory.

The Log Forwarder is installed successfully.

12.5.1.2.1.2 Using Silent Mode of Installation

This section describes how to install the Log Forwarder on a Windows platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-15: Parameter List for Silent Installation

| Parameter | Description |
|------------------------------------|--|
| -endpoint1, -endpoint2, -endpoint3 | <p>Audit Store IP address and the port number where the Log Forwarder listens for logs.</p> <p>Note: The default port number is <i>9200</i>.</p> <p>Note: The parameters <i>-endpoint2</i> and <i>-endpoint3</i> are optional.</p> |
| -dir | <p>Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility\fluent-bit</i>.</p> |
| -pemdir | <p>Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility</i>.</p> |

At the command prompt, type the following command from the installer directory.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the Log Forwarder installation directory and the *-pemdir* parameter to the command to specify the PEP server installation directory.

The following snippet displays a sample command.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>] -dir <Log Forwarder installation directory> -pemdir
<PEP server installation directory>
```

12.5.1.2.2 Installing PEP Server on Windows

This section describes how to install the PEP server on a Windows platform through the Windows Wizard and through the Silent mode of installation.

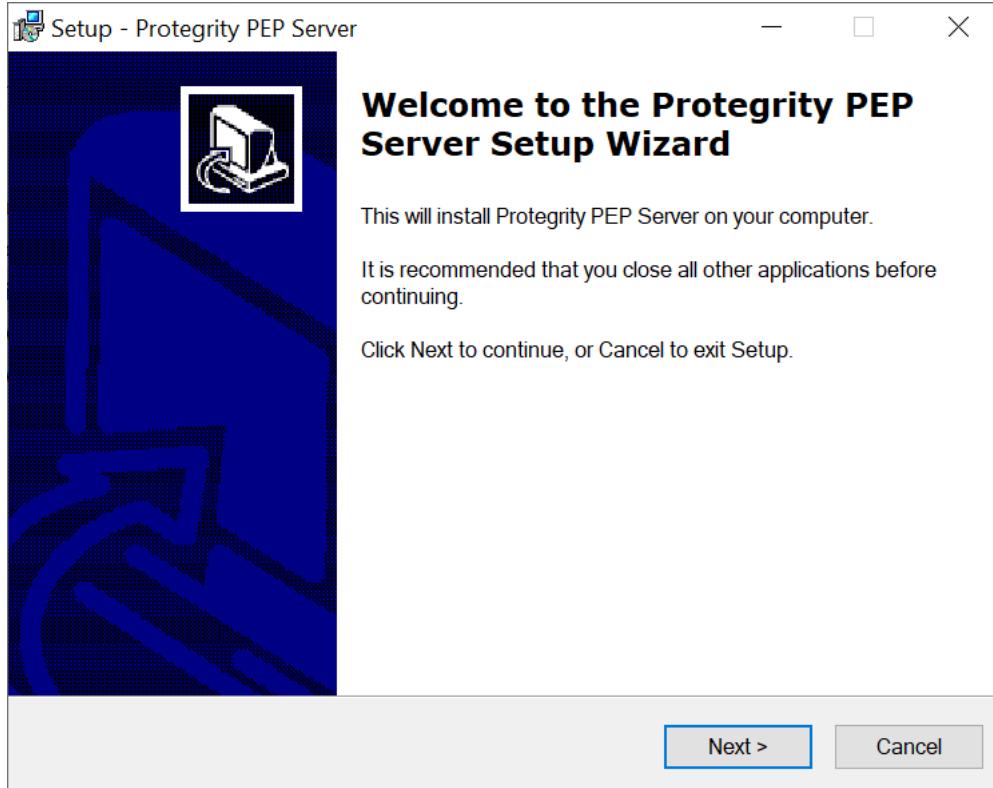
12.5.1.2.2.1 Using Windows Wizard

This section describes how to install the PEP server on a Windows platform through the Windows Wizard.

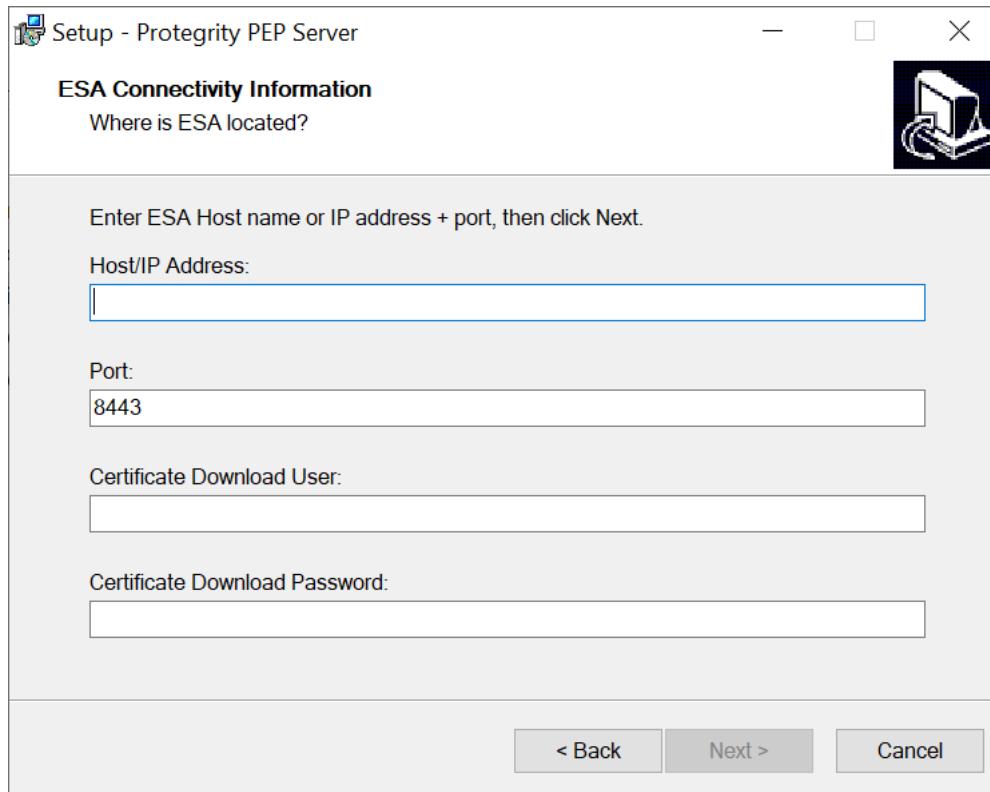


► To install the PEP server on the Windows platform through the Windows Wizard:

1. Run the *PepServerSetup_Windows_<bit-version>_<version>.exe* file from its installed directory.
The **Welcome to the Protegility PEP Server Setup Wizard** appears.



2. Click **Next**.
The **ESA Connectivity Information** screen appears.



3. Enter the ESA Host name or IP Address in the **Host/IP Address** field.
4. Enter the ESA host listening port details in the **Port** field.

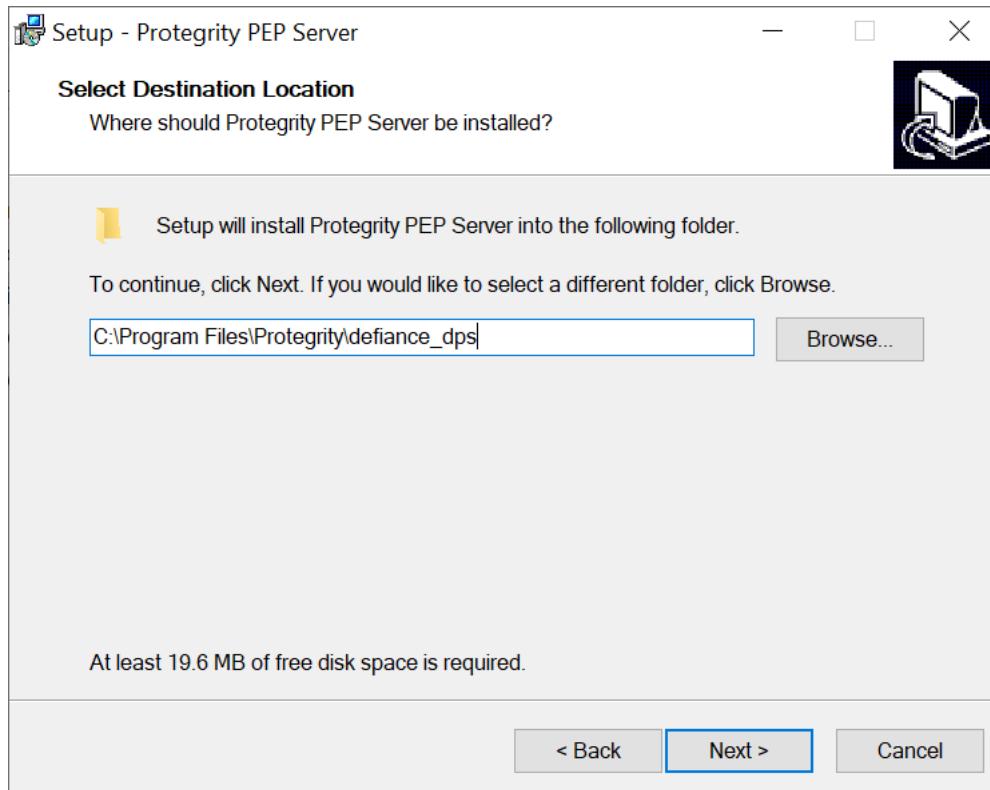
Note: Ensure that the ESA is up and running with the HubController service in running status to enable the automatic downloading of certificates.

5. Enter the **Certificate Download User**.

Note: It is recommended to use *admin* as the user.

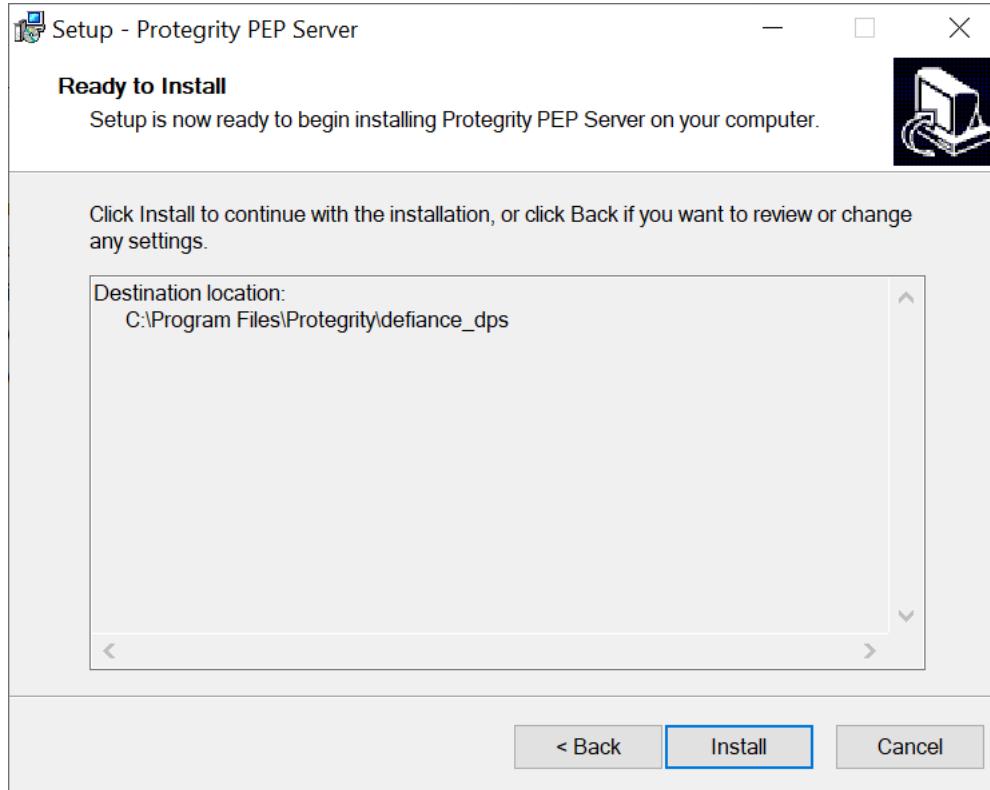
6. Enter the **Certificate Download Password**.
7. Click **Next**.

The **Select Destination Location** screen appears.



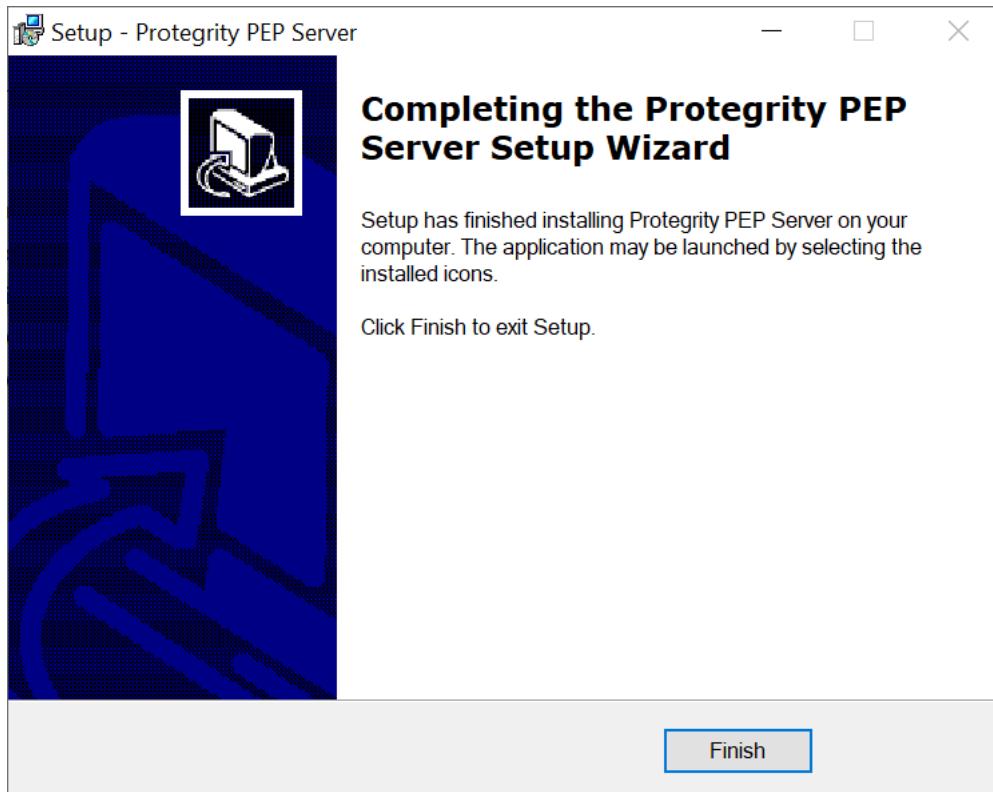
8. Set the installation directory for the PEP server to *C:\Program Files\Protegility\defiance_dps*.
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Protegility PEP Server Setup Wizard** screen appears.



11. From the **Completing the Protegility PEP Server Setup Wizard** screen, click **Finish** to complete the installation and exit.

The directories are created under the installation directory that was defined and the installation files are installed in these directories.

Note:

If you want to manually install the certificates in the *<path to where pepserver is installed>\defiance_dps\data* directory of the PEP server, then navigate to the *<path to where pepserver is installed>\defiance_dps\bin* directory, and run the following command:

```
GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the user *admin*.

The PEP server is installed successfully.

12.5.1.2.2 Using Silent Mode of Installation

This section describes how to install the PEP server on a Windows platform through the silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.

Table 12-16: Parameter List for Silent Installation

| Parameter | Description |
|-----------|-------------------------------|
| -esa | Specifies the ESA IP address. |

| Parameter | Description |
|-----------|---|
| -esaport | Specifies the ESA port, which is optional. The default value is <i>8443</i> . |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i><path to where pepserver is installed>\defiance_dps</i> . |

At the command prompt, type the following command from the installer directory.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
-dir <installation-directory-path>
```

12.5.1.2.3 Installing Application Protector C on Windows

This section describes how to install the AP C on a Windows platform.

► To install the AP C on the Windows platform:

1. Run the *XCDevSetup_Windows_x64_<version>.exe* installer from the created directory.

The **Welcome to the Defiance XCDev Setup Wizard** screen appears.

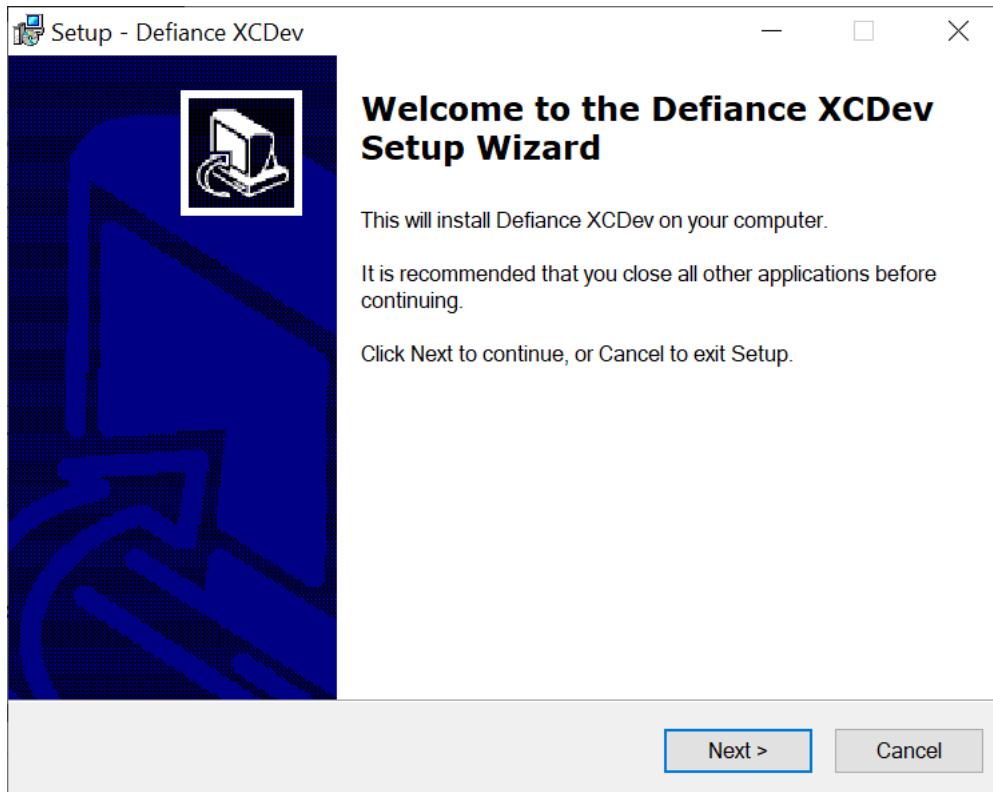


Figure 12-3: Welcome to the Defiance XCDev Setup Wizard

2. Click **Next**.
- The **Select Destination Location** screen appears.

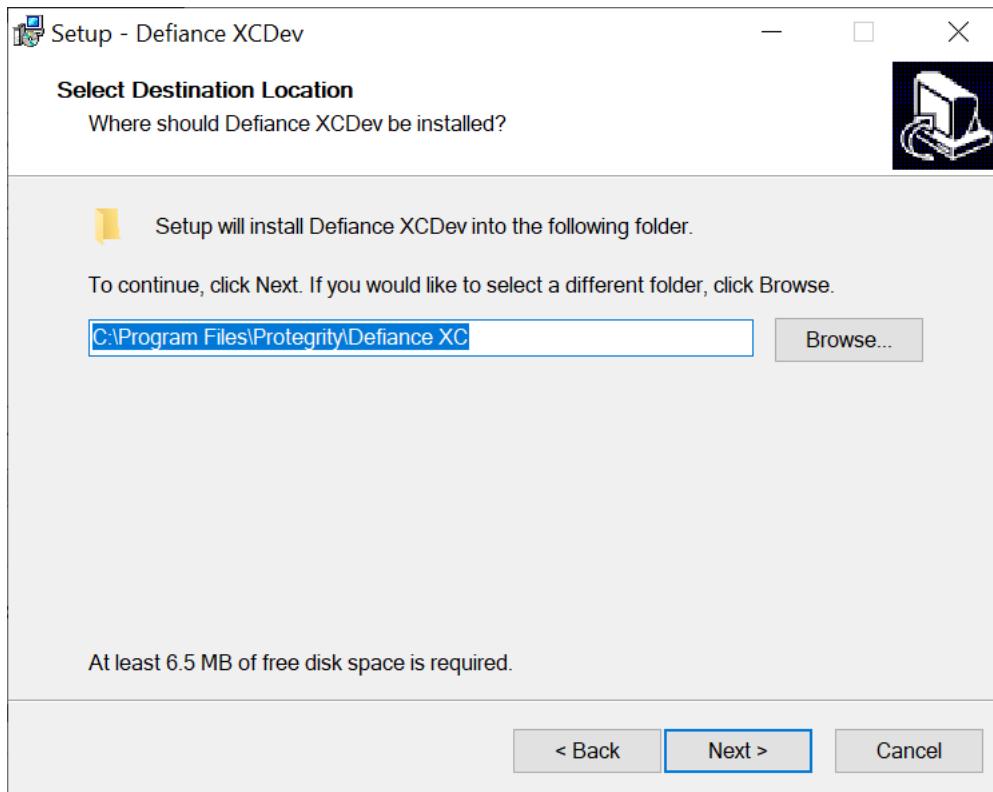


Figure 12-4: Select Destination Location

3. Set the installation directory to `C:\Program Files\Protegility\Defiance XC`.
4. Click **Next**.

The **Ready to Install** screen appears.

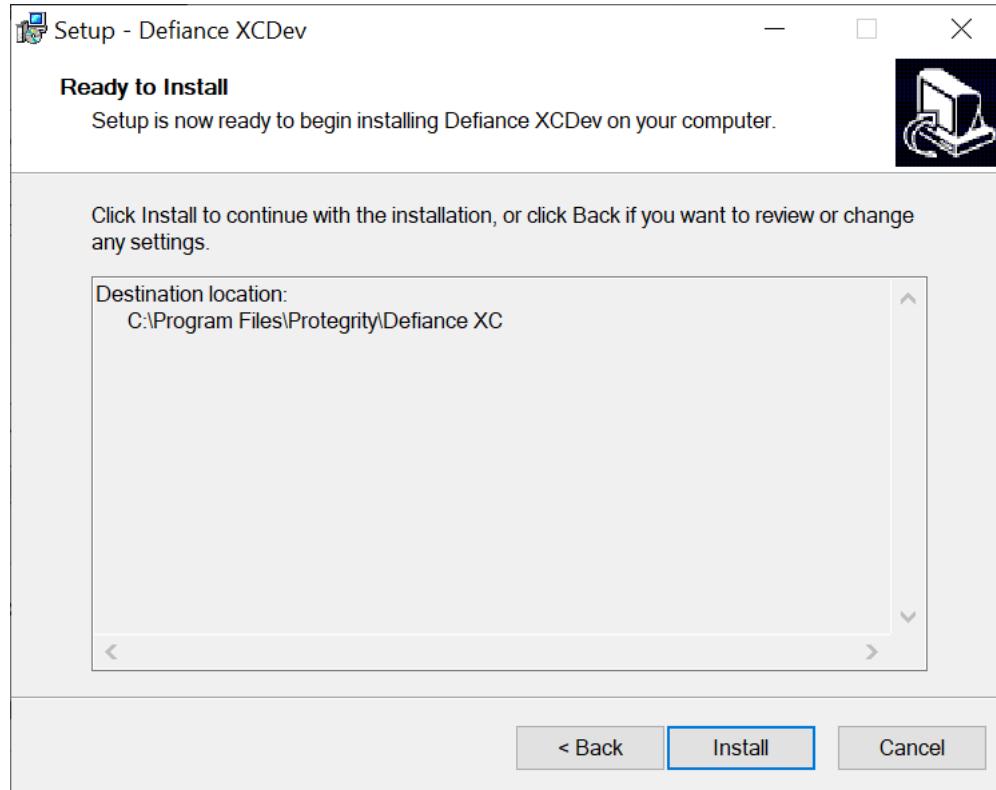


Figure 12-5: Ready to Install

5. Click **Install**.

The **Completing the Defiance XCDev Setup Wizard** screen appears.

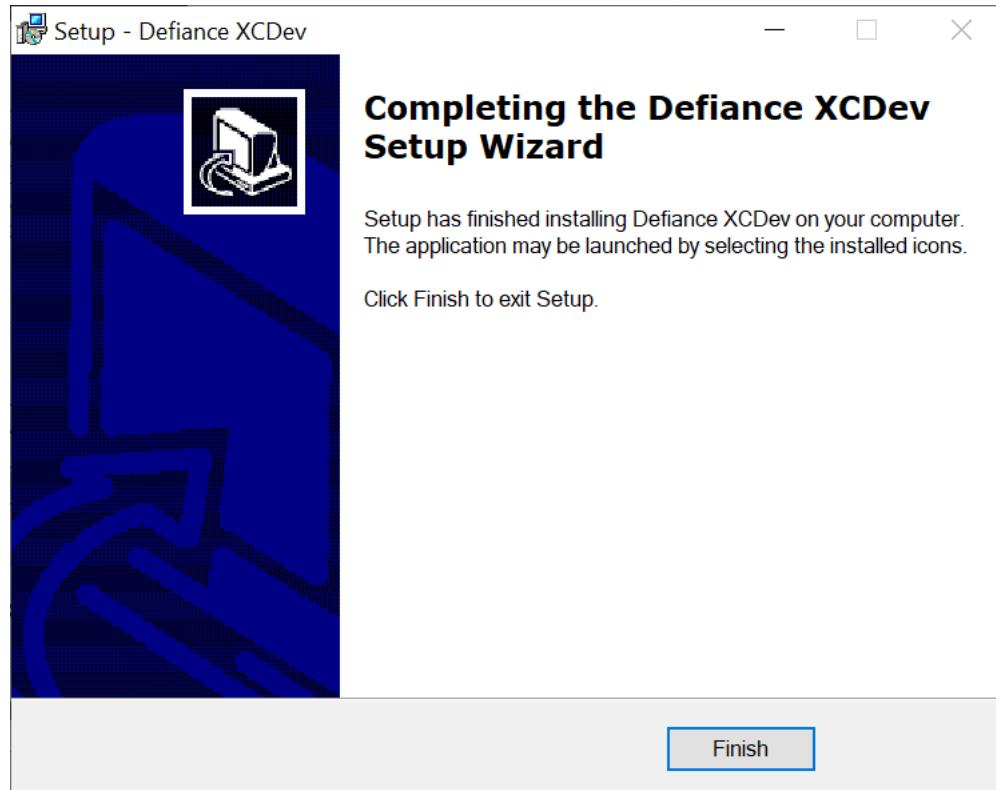


Figure 12-6: Completing the Defiance XCDev Setup Wizard

- From the **Completing the Defiance XCDev Setup Wizard** screen, click **Finish** to complete the installation and exit.

The AP C is installed successfully.

Note:

The *xcclient.plm*, *xclite.plm*, and *xcpep.plm* files are installed in the <path to where Defiance XCDev is installed>\Defiance XC\bin directory. The header files are installed in the <path to where Defiance XCDev is installed>\Defiance XC\include directory.

The *.plm* files are shared libraries that are required for compiling the applications which use AP C APIs.

Link your application to the required *.plm* file, depending on the Application Protector that you want to use.

For users working with *xcclient*, Protegility recommends using the AP REST container.

For users working with *xclite*, Protegility recommends using the Immutable AP C.

12.5.1.2.4 Installing the Application Protector C Samples on Windows

This section describes how to install the AP C samples on a Windows platform.

► To install the AP C Samples on the Windows platform:

- Run the *XCSamplesSetup_Windows_x64_<version>.exe* installer from the created directory.

The **Welcome to the Defiance XCSamples Setup Wizard** screen appears.

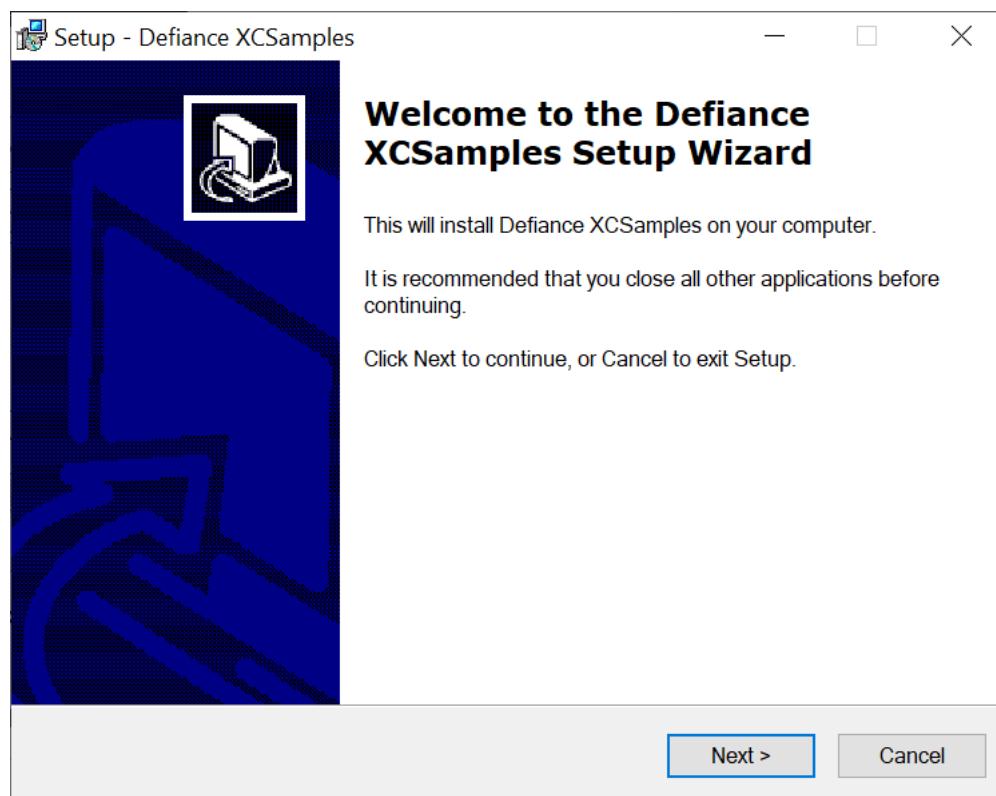


Figure 12-7: Welcome to the Defiance XCSamples Setup Wizard

2. Click **Next**.

The **Select Destination Location** screen appears.

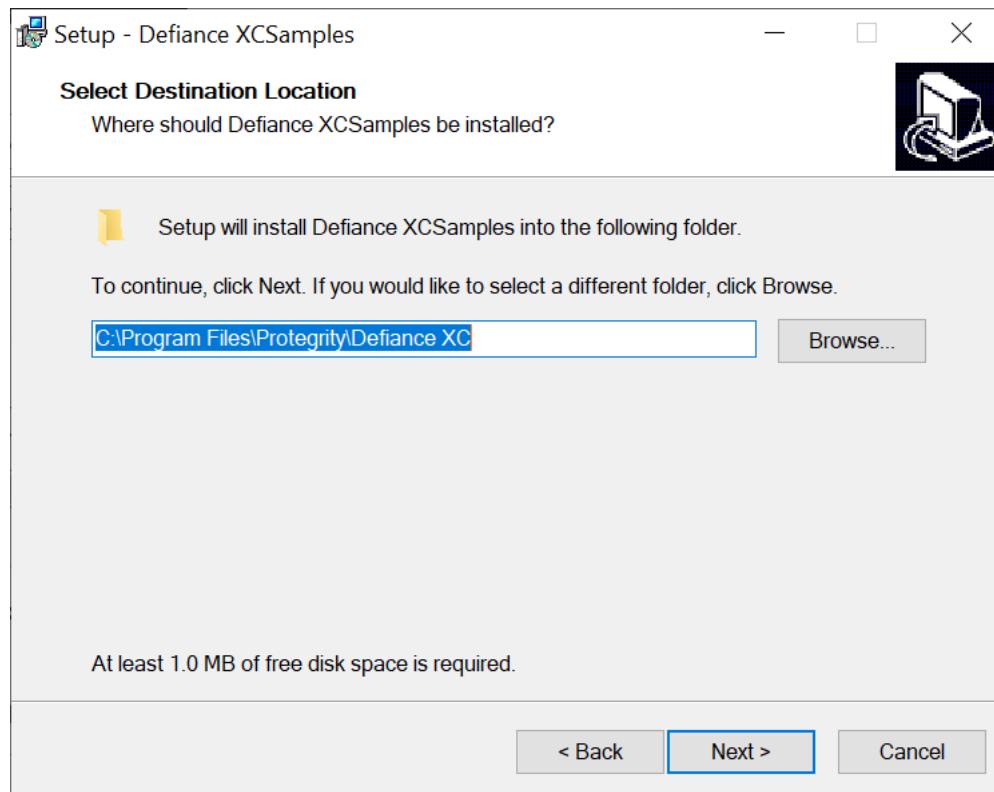


Figure 12-8: Select Destination Location

3. Set the installation directory to `C:\Program Files\Protegility\DefianceXC`.
4. Click **Next**.

The **Ready to Install** screen appears.

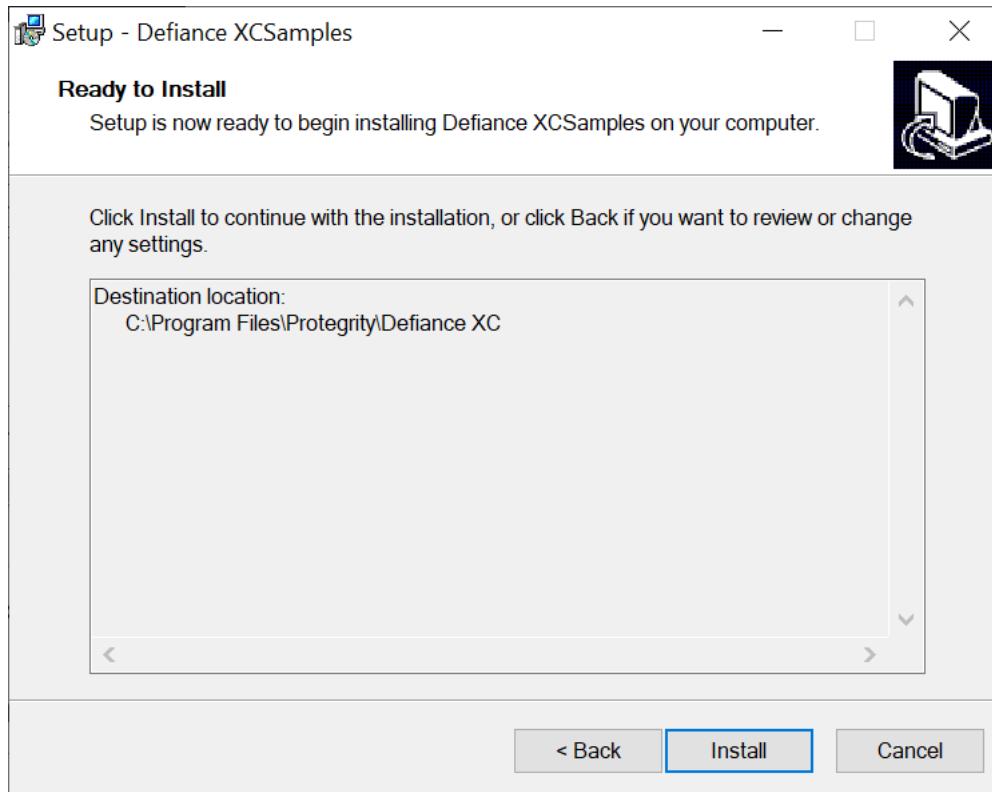


Figure 12-9: Ready to Install

5. Click **Install**.
The **Completing the Defiance XCSamples Setup Wizard** screen appears.

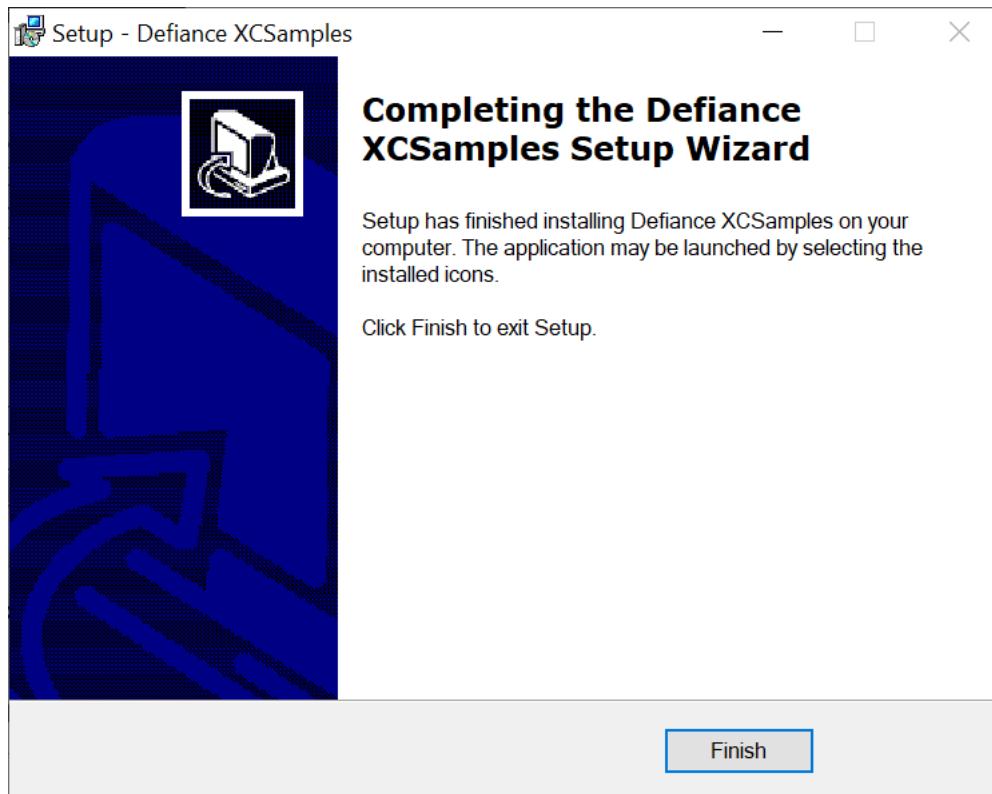


Figure 12-10: Completing the Defiance XCSamples Setup Wizard

6. From the **Completing the Defiance XCSamples Setup Wizard** screen, click **Finish** to complete the installation and exit.

Note: The sample files will be installed in the *<path to where Defiance XCDev is installed>\Defiance XC\sample* directory.

The AP C samples are installed successfully.

12.5.1.3 Uninstalling Application Protector C

This section describes the steps to uninstall the Application Protector C on the different platforms.

12.5.1.3.1 Uninstalling the Log Forwarder from AIX, Linux, or Unix

This section describes how to uninstall the Log Forwarder from an AIX, Linux, or Unix platform.

► To uninstall the Log Forwarder from the AIX, Linux, or Unix platform:

1. Navigate to the */opt/protegility/fluent-bit/bin* directory.
2. Stop the *fluentbit* component using the following command.
`./logforwarderctrl stop`
3. Delete the *fluentbit* directory.

The Log Forwarder is uninstalled.

12.5.1.3.2 Uninstalling the Log Forwarder from the Windows Platform

This section describes how to uninstall the Log Forwarder from the Windows platform.

► To uninstall the Log Forwarder on Windows:

1. Perform the following steps to stop the Log Forwarder.
 - a. From the Windows Start Menu, search and select **Services**.
 - b. Navigate to the *Logforwarder* directory.
 - c. Right-click the **Logforwarder** service and click **Stop**.
2. Run the **logforwarder** uninstall utility located in the *C:\Program Files\Protegility\fluent-bit* directory.
3. After the Log Forwarder is uninstalled, delete the following directory.

`fluent-bit`

The Log Forwarder is uninstalled.

12.5.1.3.3 Uninstalling the PEP Server from AIX, Linux or Unix

This section describes how to uninstall the PEP server from an AIX, Linux or Unix platform.

► To uninstall the PEP server from the AIX, Linux or Unix platform:

1. Navigate to the `/opt/protegility/defiance_dps/bin` directory.
2. Stop the PEP server using the following command.
`./pepsrvctrl stop`
3. Delete the `/opt/protegility/defiance_dps` directory.

The PEP server is uninstalled.

12.5.1.3.4 Uninstalling the PEP Server from the Windows Platform

This section describes how to uninstall the PEP server from the Windows platform.

► To uninstall the PEP server from the Windows platform:

1. Perform the following steps to stop the PEP server.
 - a. From the Windows Start Menu, search and select *Services*.
 - b. Navigate to the *Protegility PEP Server* directory.
 - c. Right-click the service and click **Stop**.
2. Run the uninstall utility located in the `..|Defiance DPS` directory.
3. After the PEP server is uninstalled, delete the following directories:
 - *Defiance DPS*
 - *Database Protector*
 - *Protegility*

The PEP server is uninstalled.

12.5.1.3.5 Uninstalling the Application Protector C from the Linux or Unix Platform

This section describes how to uninstall the AP C from a Linux or Unix platform.

► To uninstall the AP C from the Linux or Unix platform:

1. Navigate to the installation directory `/opt/protegility/defiance_xc/bin`.
2. Remove the `bin` directory from the `defiance_xc` directory.

The AP C is uninstalled.

12.5.1.3.6 Uninstalling the Application Protector C from the Windows Platform

This section describes how to uninstall the AP C from a Windows platform.

► To uninstall the AP C from the Windows platform:

1. Navigate to the *C:\Program Files\Protegrity\DefianceXC* installation directory.
2. Run the uninstall utility located in the *C:\Program Files\Protegrity\DefianceXC* directory.
The AP C is uninstalled.
3. Delete the *C:\Program Files\Protegrity\DefianceXC* directory.

12.5.1.3.7 Uninstalling the Application Protector C Samples from the Linux or Unix Platform

This section describes how to uninstall the AP C Samples from a Windows platform.

► To uninstall the AP C Samples from the Linux or Unix platform:

1. Navigate to the installation directory */opt/protegrity/defiance_xc/samples*.
2. Remove the *samples* directory from the *defiance_xc* directory.

The AP C Samples are uninstalled.

12.5.1.3.8 Uninstalling the Application Protector C Samples from the Windows Platform

This section describes how to uninstall the AP C samples from a Windows platform.

► To uninstall the AP C Samples from the Windows platform:

1. Navigate to the *C:\Program Files\Protegrity\DefianceXC* installation directory.
2. Run the uninstall utility located in the *C:\Program Files\Protegrity\DefianceXC* directory.
The AP C samples are uninstalled.
3. Delete the *C:\Program Files\Protegrity\DefianceXC* directory.

12.5.2 Installing Application Protector (AP) Java

Protegrity Application Protector (AP) Java provides APIs that integrate with the customer application to protect, unprotect, and reprotect sensitive data. The AP Java can be used with any customer application that is developed using the Java programming language.

This section describes a standard setup of installing and uninstalling the Application Protector Java on different platforms.

12.5.2.1 Setting up Application Protector Java on Linux or Unix

This section describes how to install the Application Protector (AP) Java on a Linux or Unix platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP Java:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.



- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.2](#).

► To setup the AP Java on the Linux or Unix platform:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_JRE-1.8-64_<version>.tgz* file to any location on the machine where you want to install the protector.
2. Extract the AP Java installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_JRE-1.8-64_<version>.tgz
```

The following setup files are extracted:

- *APJavaSetup_Linux_x64_<version>.sh*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*

12.5.2.1.1 Installing Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.2.1.1.1 Installing Log Forwarder on Linux or Unix Using Linux Installer

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer.

► To install the Log Forwarder on a Linux or Unix platform using the Linux installer:

1. Run the Log Forwarder installer using the following command.

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store port number where the Log Forwarder listens for logs.

Note: The default port number is *15780*.

3. Press ENTER.

The added Audit Store endpoint appears on the screen.

The prompt to enter an additional Audit Store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

4. If you want to add more than one Audit Store endpoint, then type *y* else type *n*.

Note: If you need to add *n* Audit Store endpoints, then repeat the *Step 2* and *Step 3* *n* times.

5. Type the *y* key to install into the destination directory.

The Log Forwarder is installed in the */opt/protegility/fluent-bit/* directory.

6. Start the Log Forwarder component by using the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl start
```

The Log Forwarder is successfully installed.

12.5.2.1.1.2 Silent Mode of Installation of Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-17: Parameter List for Silent Installation

| Parameter | Description |
|------------------|--|
| --endpoint or -e | The IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is 15780. |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *--dir* parameter to the command to specify the Log Forwarder installation directory and *--pemdir* parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>] --dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.5.2.1.2 Installing PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.2.1.2.1 Installing PEP Server on Linux or Unix using Linux Installer

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer.

► To install the PEP server on a Linux or Unix platform:

1. Run the PEP server installer using the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the ESA Host Name or IP Address.

3. Press ENTER.

The prompt to enter the username for downloading certificates appears.

```
Please enter the user name for downloading certificates:
```

4. Enter the username for downloading the certificates.

5. Press ENTER.

The prompt to enter the password for downloading the certificates appears.

```
Please enter the password for downloading certificates:
```

6. Press ENTER to install into the destination directory.

Directories are created under `/opt/protegility/defiance_dps` by default, and the required installation files are installed in these directories.

Caution: Ensure that the ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

To manually install the certificates to the `/opt/protegility/defiance_dps/data` directory of the PEP server, navigate to the `/opt/protegility/defiance_dps/bin` directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the `administrator` user.

Verify that the following files have been copied to the `/opt/protegility/defiance_dps/data` directory:

- `CA.pem`
- `keyinternal.plm`
- `pepperserver.cfg`
- `pepperserver.pid`

- *authesa.plm*
- *cert.key*
- *cert.pem*
- *certkeyup.bin*

7. Start the PEP server by using the following command.

```
/opt/protegility/defiance_dps/bin/pepsrvctrl start
```

The PEP server is successfully installed.

12.5.2.1.2.2 Silent Mode of Installation of PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform through the Silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-18: Parameter List for Silent Installation

| Parameter | Description |
|-----------|---|
| -esa | Specifies the ESA IP address. |
| -esaport | Specifies the ESA port, which is optional. The default value is <i>8443</i> . |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

At the command prompt, type the following command from the installer directory.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password> -dir <installation-directory-path>
```

12.5.2.1.3 Installing Application Protector Java on Linux or Unix using Linux Installer

This section describes how to install the AP Java on a Linux or Unix platform using the Linux installer.

► To install the AP Java on the Linux or Unix platform using the Linux installer:

1. Run the AP Java installer using the following command.

```
./APJavaSetup_Linux_x64_<version>.sh
```

The prompt to continue the installation appears.

This will install AP Java SDK on your computer.

Do you want to continue? [yes or no]

2. If you want to continue with the installation of the AP Java SDK, then type *yes* else type *no*.

If you type *yes*, then the prompt to enter the installation directory appears.

Please enter installation directory
[/opt/protegility]:

If you type *no*, then the installation of the AP Java aborts.

The AP Java is installed successfully.

The default installation directories for the Linux platform are listed in the following table.

Table 12-19: AP Java Installation Directory

| Platform | Directory |
|------------|--|
| Linux/Unix | /opt/protegility/applicationprotector/java |

12.5.2.2 Setting up Application Protectors Java on Windows

This section describes how to install the AP Java on a Windows platform.

Before you begin

Ensure that the following prerequisites are met before installing Application Protector:

- The ESA appliance is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

► To setup the AP Java on the Windows platform:

1. Download the *ApplicationProtector_WIN-ALL-64_x86-64_JRE-1.8-64_<version>.zip* installation package to any location on the machine where you want to install the protector.
2. Extract the files from the *ApplicationProtector_WIN-ALL-64_x86-64_JRE-1.8-64_<version>.zip* installation package.

The following setup files are extracted:

- *LogforwarderSetup_Windows_x64_<version>.exe*
- *PepServerSetup_Windows_x64_<version>.exe*



- *APJavaSetup_Windows_x64_<version>.exe*

12.5.2.2.1 Installing Log Forwarder on Windows

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard and through the Silent mode of installation.

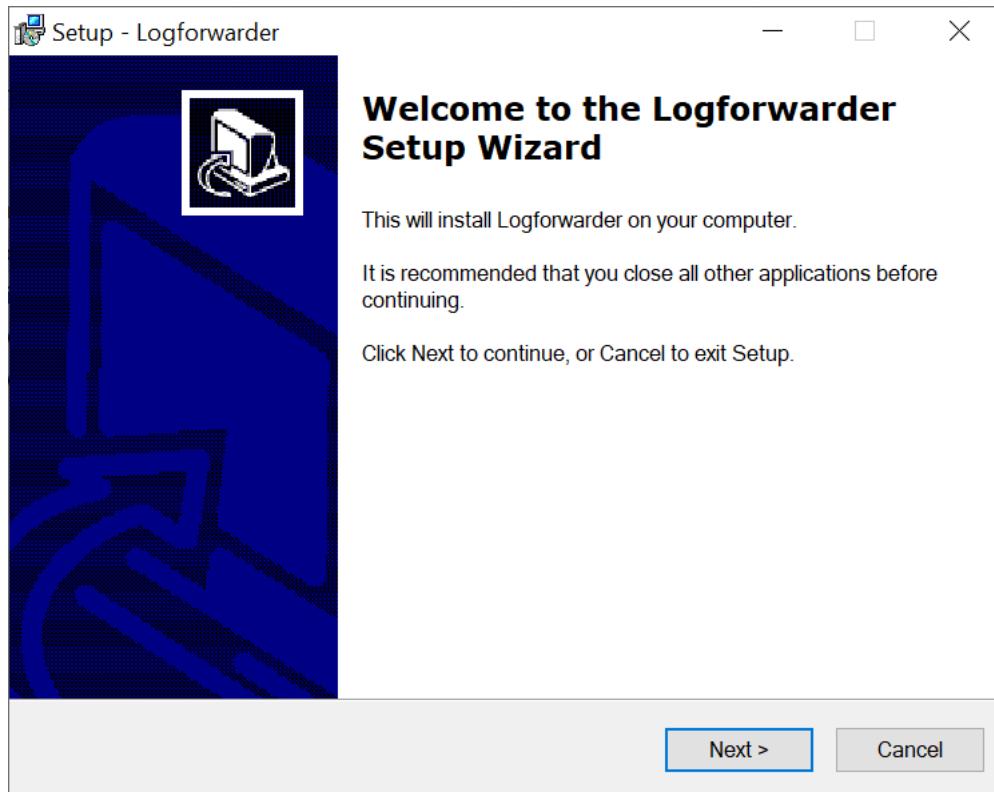
12.5.2.2.1.1 Using Windows Wizard

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard.

► To install the Log Forwarder on the Windows platform through the Windows Wizard:

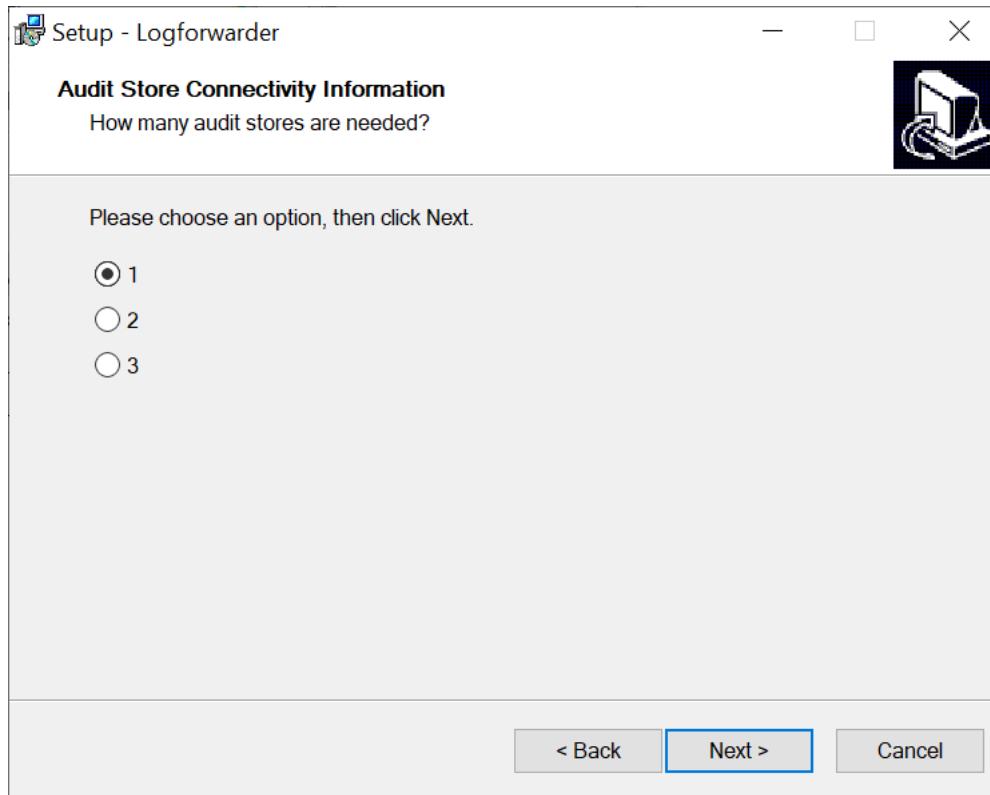
1. Run the *LogforwarderSetup_Windows_x64_<version>.exe* file from the created directory.

The **Welcome to the Log Forwarder Setup Wizard** screen appears.



2. Click **Next**.

The **Audit Store Connectivity Information** screen appears to select the number of audit stores that are needed.

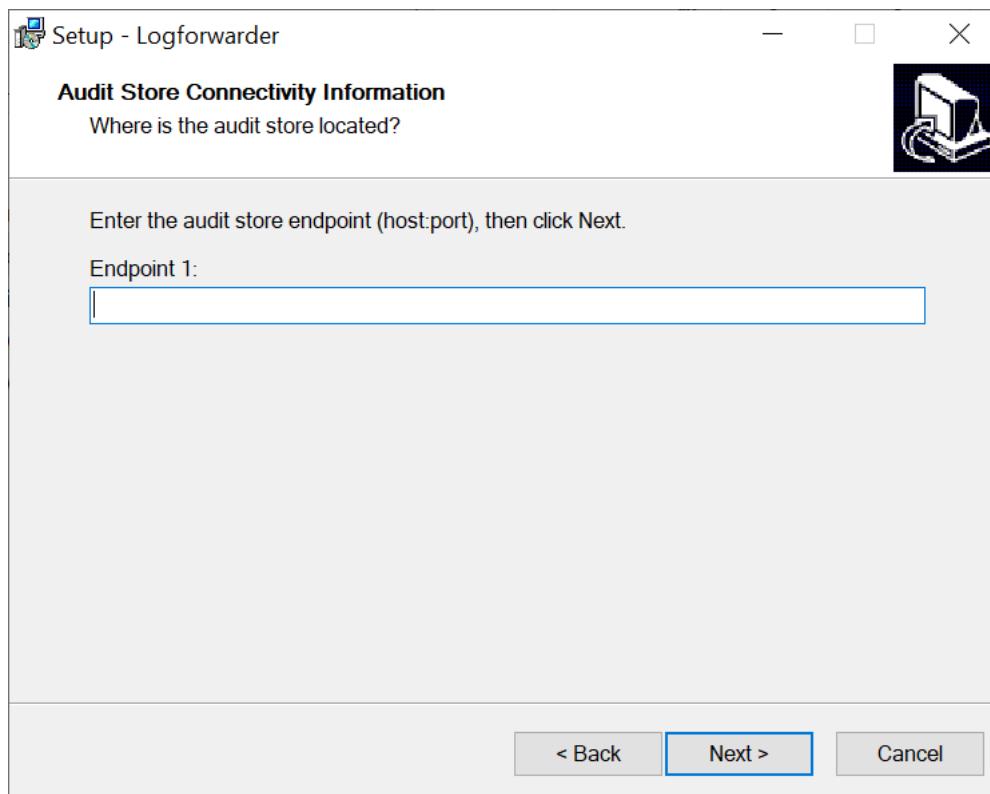


3. Select the number of Audit Stores as required.

The maximum number of Audit Stores that can be configured is 3.

4. Click **Next**.

The **Audit Connectivity Information** screen appears to enter the location of the audit store.



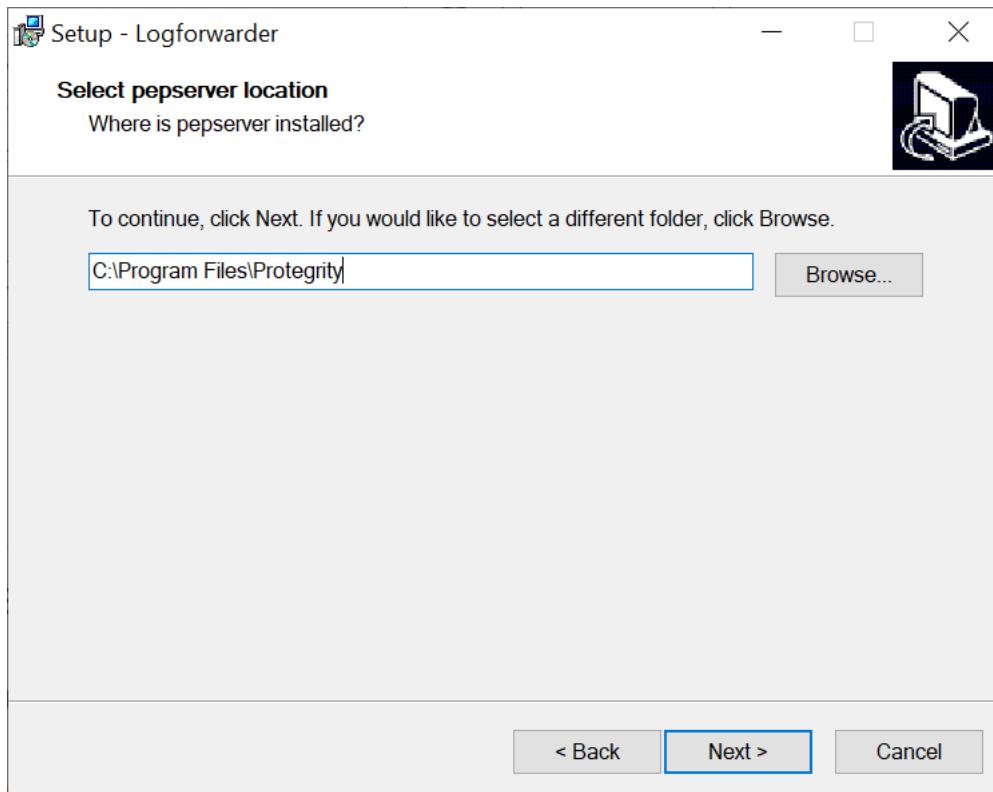
5. Enter the Hostname/IPAddress:port where the Audit Store is configured.

Note:

The default port number for the Audit Store is **9200**.

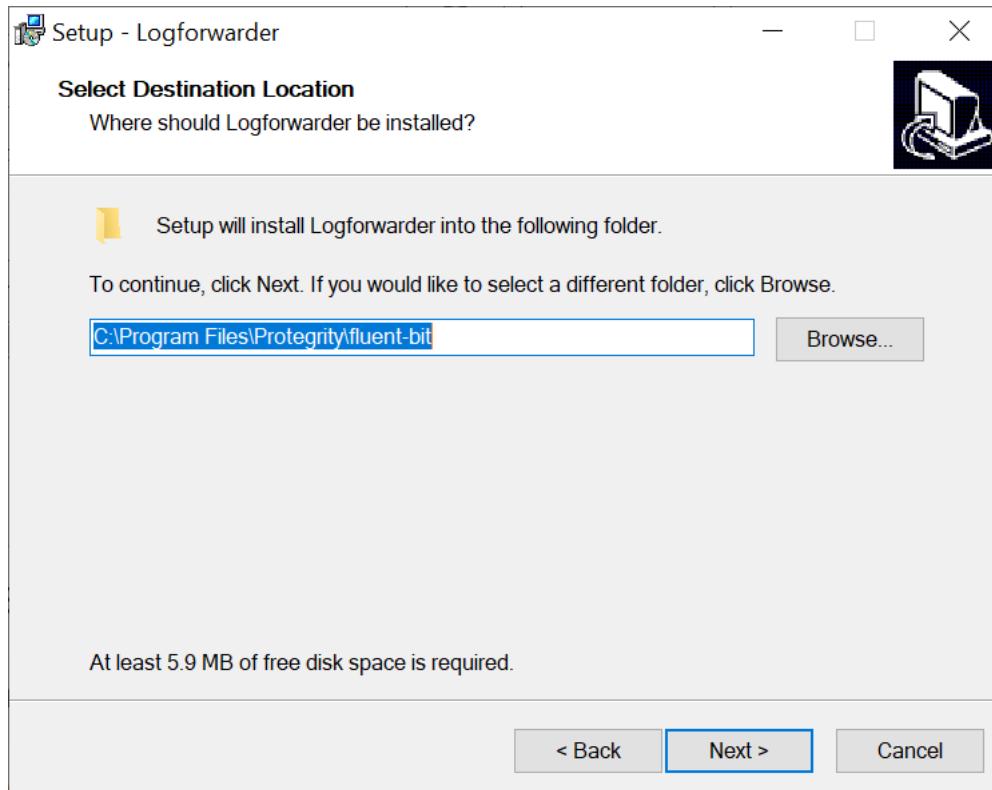
6. Click **Next**.

The **Select pepserver location** screen appears.



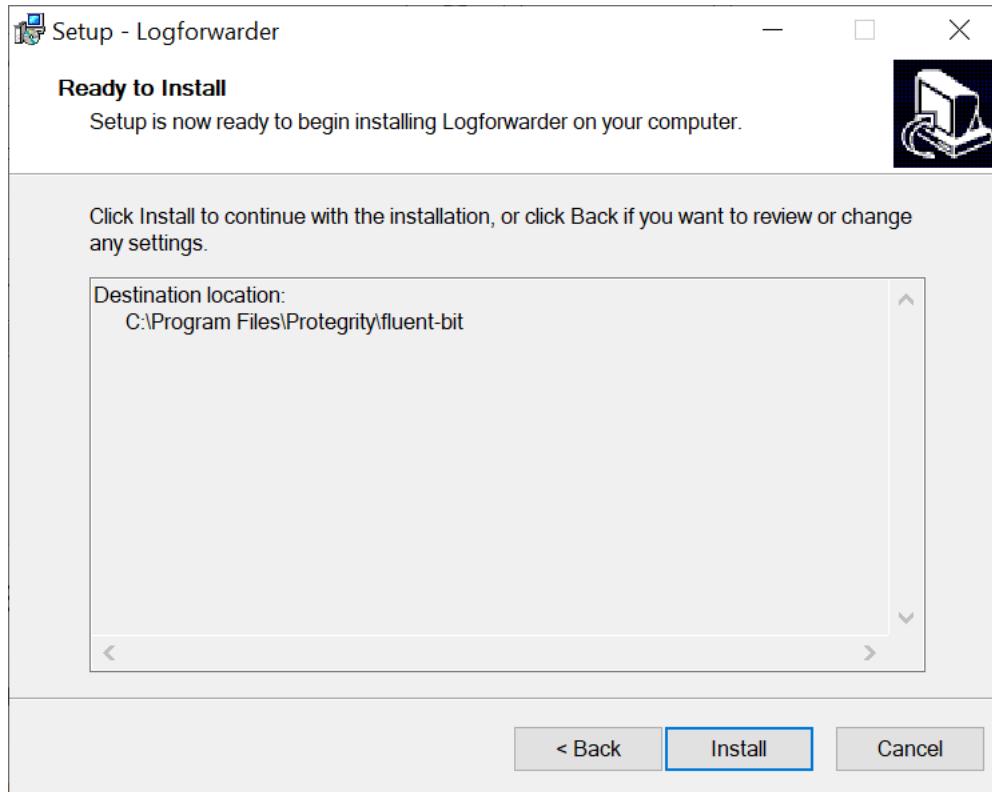
7. Click **Next**.

The **Select Destination Location** screen appears.



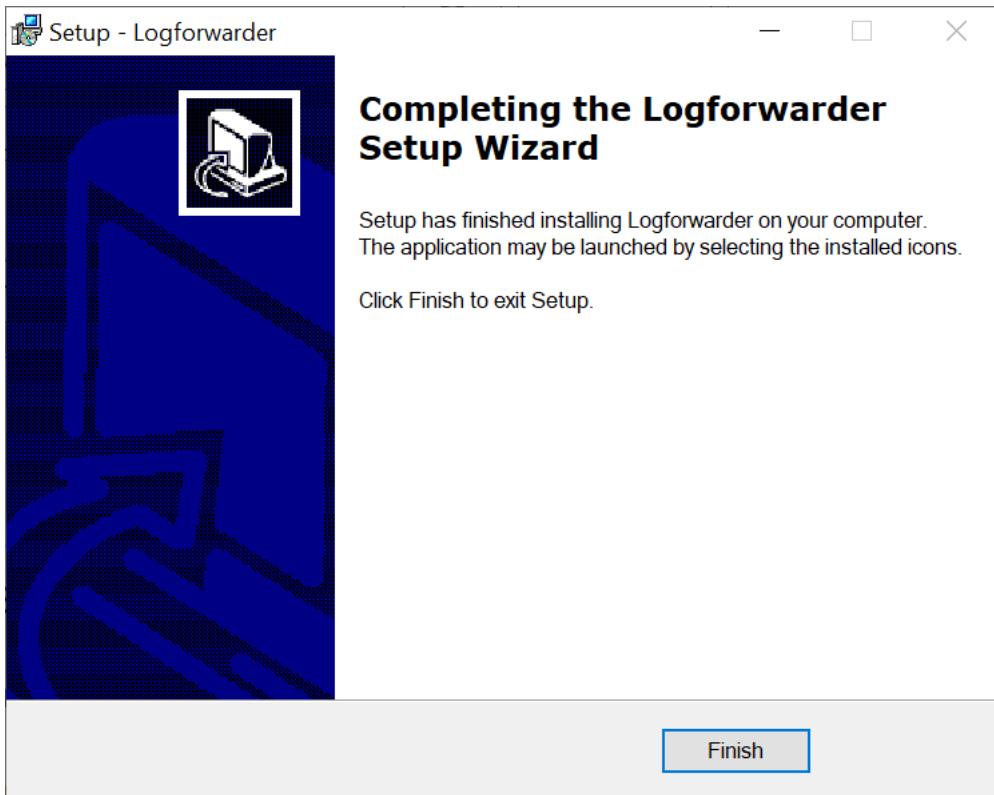
8. Set the installation directory for the Log Forwarder to *C:\Program Files\Protegity\fluent-bit*.
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Logforwarder Setup Wizard** screen appears.



11. From the **Completing the Logforwarder Setup Wizard** screen, click **Finish** to complete the installation and exit.

Post successful installation, you can find the installation files in the directories that are created under the defined installation directory.

The Log Forwarder is installed successfully.

12.5.2.2.1.2 Using Silent Mode of Installation

This section describes how to install the Log Forwarder on a Windows platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-20: Parameter List for Silent Installation

| Parameter | Description |
|------------------------------------|--|
| -endpoint1, -endpoint2, -endpoint3 | <p>Audit Store IP address and the port number where the Log Forwarder listens for logs.</p> <p>Note: The default port number is <i>9200</i>.</p> <p>Note: The parameters <i>-endpoint2</i> and <i>-endpoint3</i> are optional.</p> |
| -dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility\fluent-bit</i> . |

| Parameter | Description |
|-----------|--|
| -pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C : \Program Files\Protegility</i> . |

At the command prompt, type the following command from the installer directory.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the Log Forwarder installation directory and the *-pemdir* parameter to the command to specify the PEP server installation directory.

The following snippet displays a sample command.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>] -dir <Log Forwarder installation directory> -pemdir <PEP server installation directory>
```

12.5.2.2.2 Installing PEP Server on Windows

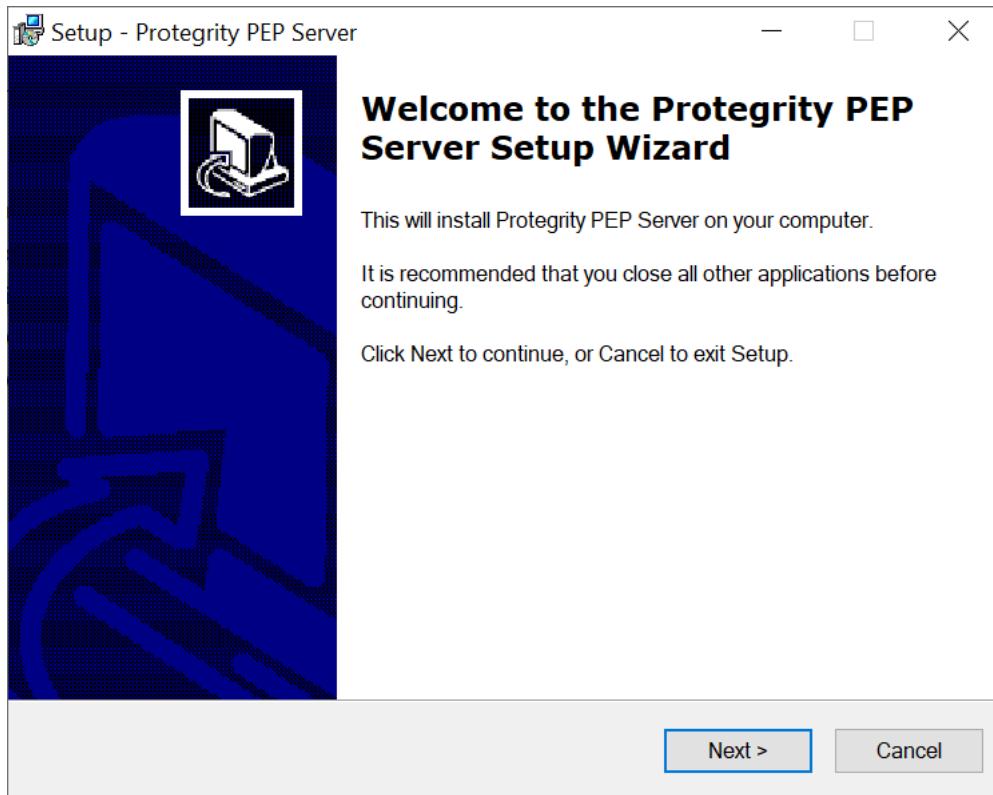
This section describes how to install the PEP server on a Windows platform through the Windows Wizard and through the Silent mode of installation.

12.5.2.2.1 Using Windows Wizard

This section describes how to install the PEP server on a Windows platform through the Windows Wizard.

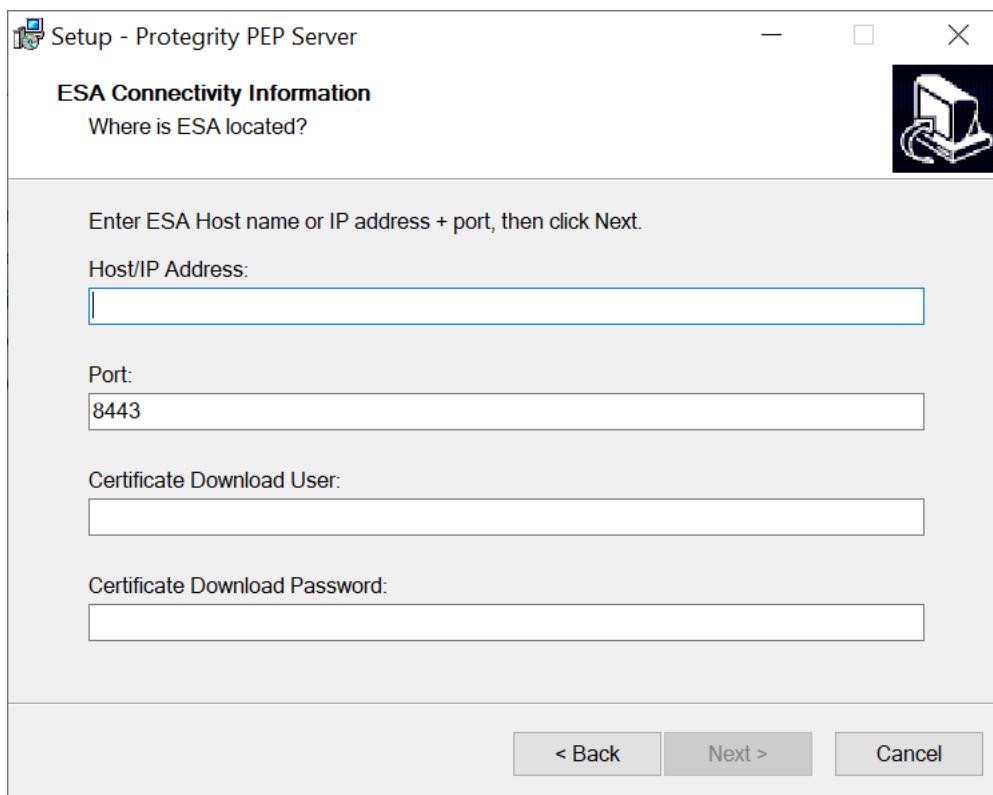
► To install the PEP server on the Windows platform through the Windows Wizard:

1. Run the *PepServerSetup_Windows_<bit-version>_<version>.exe* file from its installed directory.
The **Welcome to the Protegility PEP Server Setup Wizard** appears.



2. Click **Next**.

The **ESA Connectivity Information** screen appears.



3. Enter the ESA Host name or IP Address in the **Host/IP Address** field.

- Enter the ESA host listening port details in the **Port** field.

Note: Ensure that the ESA is up and running with the HubController service in running status to enable the automatic downloading of certificates.

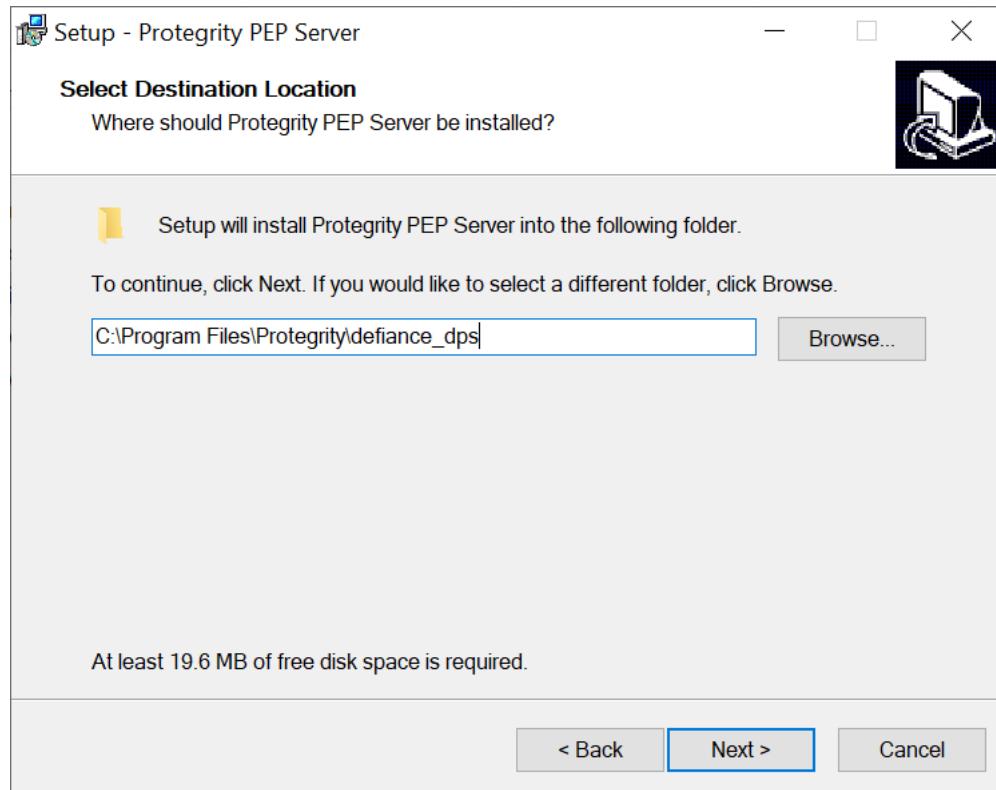
- Enter the **Certificate Download User**.

Note: It is recommended to use *admin* as the user.

- Enter the **Certificate Download Password**.

- Click **Next**.

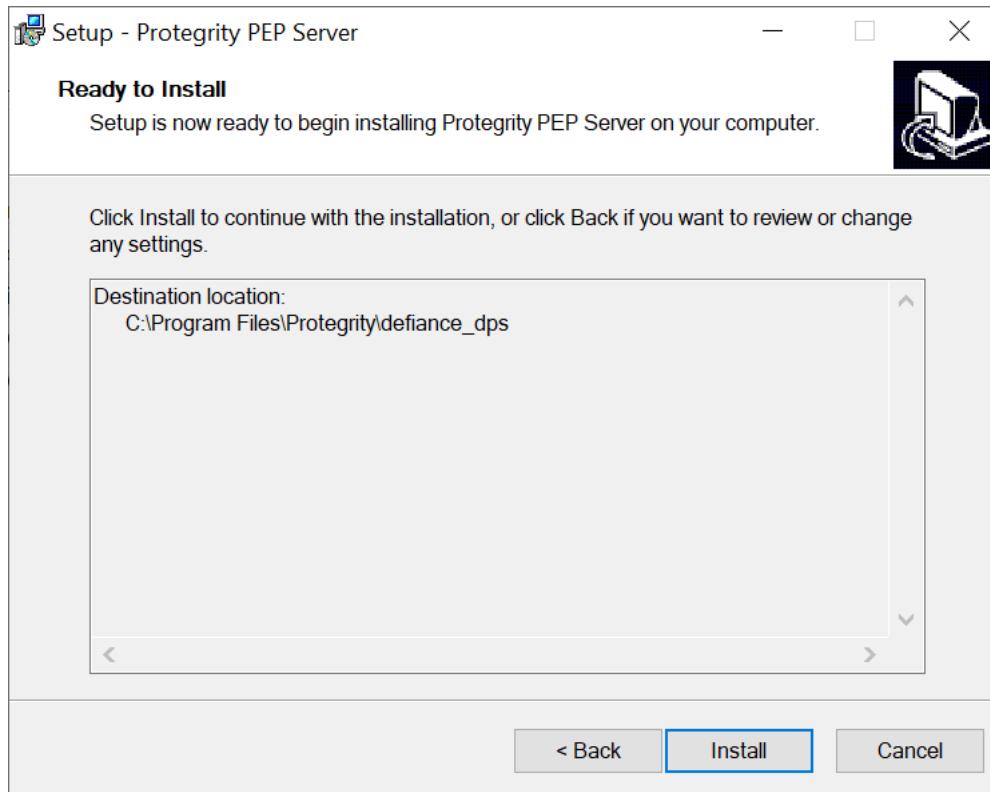
The **Select Destination Location** screen appears.



- Set the installation directory for the PEP server to *C :\Program Files\Protegility\defiance_dps*.

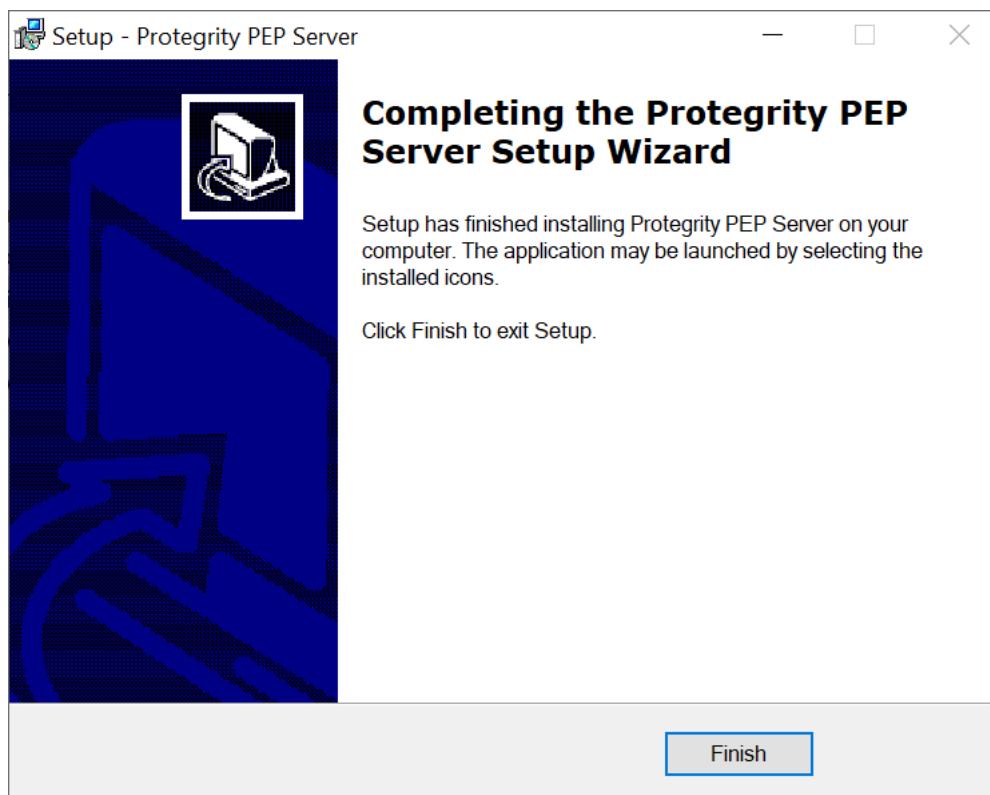
- Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Protegility PEP Server Setup Wizard** screen appears.



11. From the **Completing the Protegility PEP Server Setup Wizard** screen, click **Finish** to complete the installation and exit.

The directories are created under the installation directory that was defined and the installation files are installed in these directories.

Note:

If you want to manually install the certificates in the *<path to where pepserver is installed>\defiance_dps\data* directory of the PEP server, then navigate to the *<path to where pepserver is installed>\defiance_dps\bin* directory, and run the following command:

```
GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the user *admin*.

The PEP server is installed successfully.

12.5.2.2.2 Using Silent Mode of Installation

This section describes how to install the PEP server on a Windows platform through the silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.

Table 12-21: Parameter List for Silent Installation

| Parameter | Description |
|-----------|---|
| -esa | Specifies the ESA IP address. |
| -esaport | Specifies the ESA port, which is optional. The default value is <i>8443</i> . |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i><path to where pepserver is installed>\defiance_dps</i> . |

At the command prompt, type the following command from the installer directory.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
-dir <installation-directory-path>
```

12.5.2.2.3 Installing Application Protector Java on Windows

This section describes how to install the AP Java on a Windows platform.

- To install the AP Java on the Windows platform:



- Run the *APJavaSetup_Windows_x64_<version>.exe* installer from the created directory.
The **Welcome to Defiance DPS AP Java API Setup Wizard** screen appears.

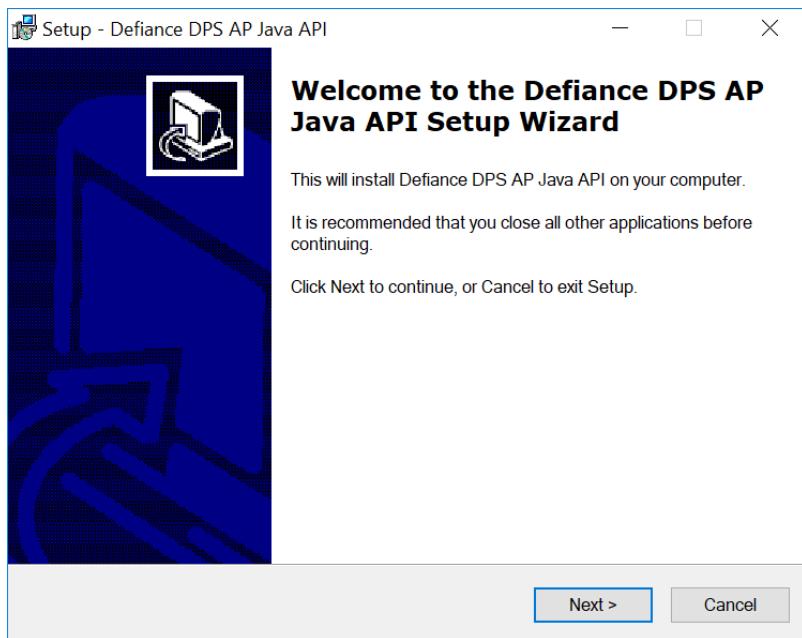


Figure 12-11: Welcome to Defiance DPS AP Java API Setup Wizard

- Click **Next**.
The **Select Destination Location** screen appears.

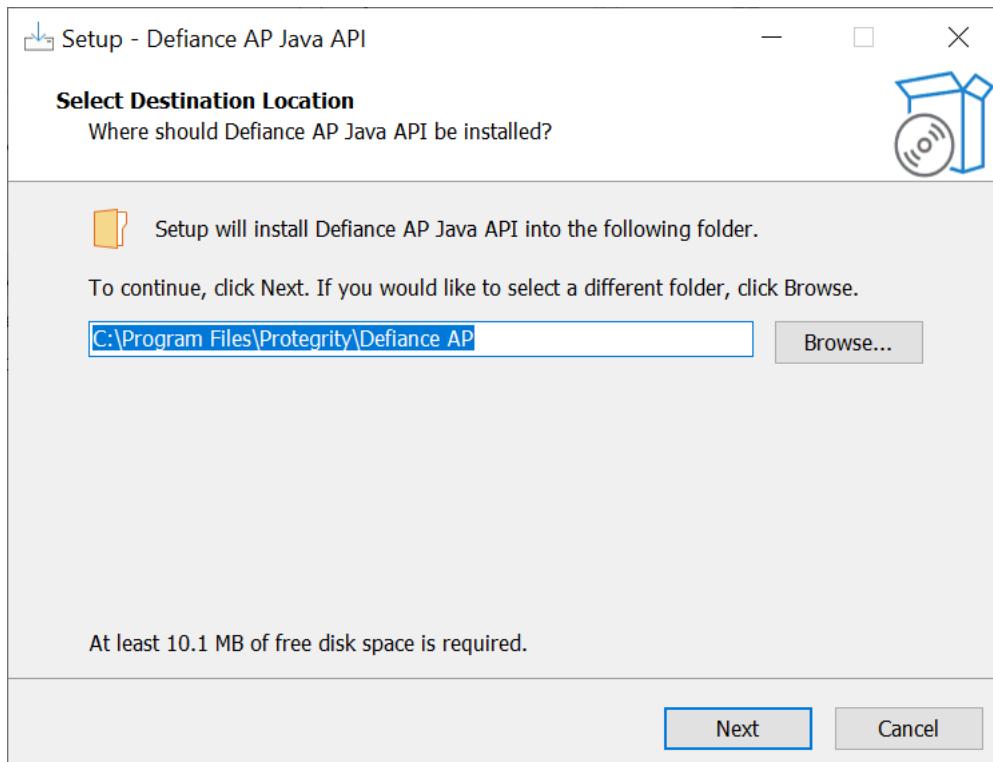


Figure 12-12: Select Destination Location

- Set the installation directory to *C:\Program Files\Protegility\Defiance AP*.
- Click **Next**.
The **Ready to Install** screen appears.

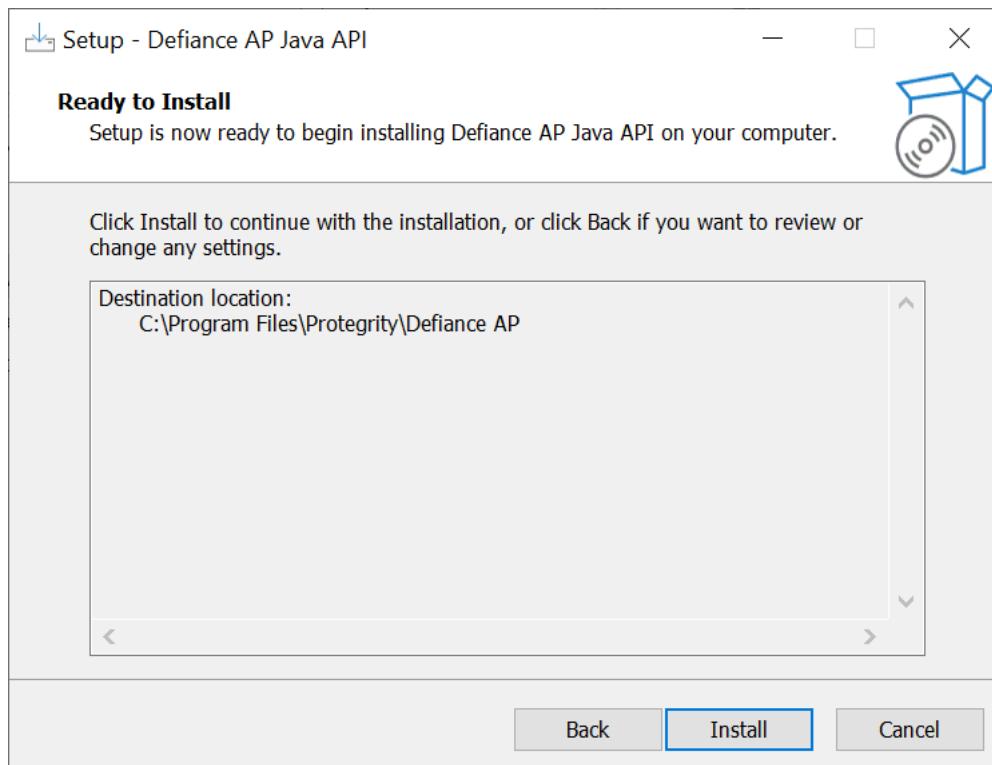


Figure 12-13: Ready to Install

5. Click **Install**.

The **Completing the Defiance AP Java API Setup Wizard** screen appears.

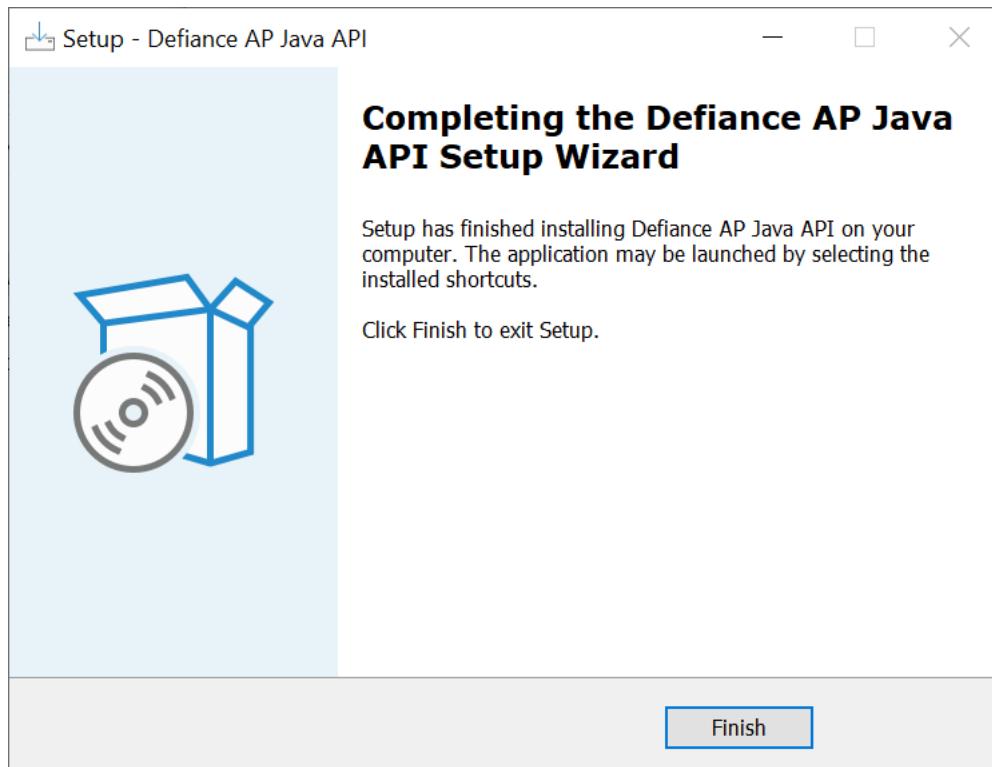


Figure 12-14: Completing the Defiance AP Java API Setup Wizard

6. From the **Completing the Defiance AP Java API Setup Wizard** screen, click **Finish** to complete the installation and exit.
The AP Java is installed successfully.

The default installation directories for different platforms are given in the following table.

Table 12-22: AP Java Installation Directory

| Platform | Directory |
|----------|---|
| win64 | <p><i>C:\Program Files\Protegility\Defiance AP\java\lib</i></p> <p>For API documentation:</p> <p><i>C:\Program Files\Protegility\Defiance AP\java\doc</i></p> |

7. Check that the following libraries are installed in the AP Java installation directory:
 - *ApplicationProtectorJava.jar*
 - *ApplicationProtectorJava.properties*
 - *jna-5.8.0.jar*
 - *jna-platform-5.8.0.jar*
 - *jpeplite.plm*
8. Link your application to the installed *.jar* and *ApplicationProtectorJava.properties* files.

The AP Java is installed successfully.

12.5.2.3 Setting up Application Protector Java on AIX

This section describes how to set up the AP Java on an AIX platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP Java:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of the PIM ensures that the cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

► To setup the AP Java on the AIX platform:

1. Download the *ApplicationProtector_AIX-ALL-64_PPC-64_JRE-1.8-64_<version>.tgz* installation package to any location on the machine where you want to install the protector.
2. Extract the AP Java installation package using the following command.

```
gunzip ApplicationProtector_AIX-ALL-64_PPC-64_JRE-1.8-64_<version>.tgz
```

The *ApplicationProtector_AIX-ALL-64_PPC-64_JRE-1.8-64_<version>.tar* file is extracted.

3. Extract the *ApplicationProtector_AIX-ALL-64_PPC-64_JRE-1.8-64_<version>.tar* file using the following command.

```
tar -xvf ApplicationProtector_AIX-ALL-64_PPC-64_JRE-1.8-64_<version>.tar
```



The following setup files are extracted:

- *LogforwarderSetup_Windows_x64_<version>.exe*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_AIX_ppc64_<version>.sh*
- *APJavaSetup_AIX_ppc64_<version>.sh*

Note:

It is mandatory to install the Log Forwarder before installing the PEP server to ensure that the Application Protector is configured correctly.

4. To install the Log Forwarder either on a Linux machine or a Windows machine, use the respective script provided in the build package for installation:

- *LogforwarderSetup_Linux_x64_<version>.sh*
- *LogforwarderSetup_Windows_x64_<version>.exe*

Note:

The Fluent Bit is incompatible with AIX.

For more information about the supported platforms, refer to <https://docs.fluentbit.io/manual/installation/supported-platforms>.

For more information about configuring the Log Forwarder for the AIX platform, refer to the section [Configuring the Log Forwarder for AIX platform](#)

5. Install the PEP server by running the following command.

```
./PepServerSetup_AIX_ppc64_<version>.sh
```

Caution: Ensure that the ESA is up and running with the *HubController* service in running status to enable automatic downloading of certificates.

To manually install the certificates to the */opt/protegility/defiance_dps/data* directory of the PEP server, navigate to the */opt/protegility/defiance_dps/bin* directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the *administrator* user.

Verify that the following files have been copied to the */opt/protegility/defiance_dps/data* directory:

- *CA.pem*
- *keyinternal.plm*
- *pepperver.cfg*
- *pepperver.pid*
- *authesa.plm*
- *cert.key*
- *cert.pem*
- *certkeyup.bin*



- Navigate to the *Logging Configuration* section in the *pepserver.cfg* file.

The following code snippet appears on the screen.

```
# Fluentbit host and port values (mostly localhost) where logs will be
# forwarded from the protector.
host = 127.0.0.1
port = 15780
```

The default value of the *host* IP address is *127.0.0.1*.

- For installing the protector on the AIX platform, set the value of the *host* parameter with the IP address of the Linux or Windows machine where the Log Forwarder is installed.
- The default value of the *port* number is *15780*. Ensure that the value of the *port* parameter in the *in_tcp.conf* file and the *pepserver.conf* file matches.

For more information about configuring the PEP server for the AIX platform, refer to the section [Configuring the PEP Server for AIX platform](#).

- Start the PEP server by running the following commands.

```
export LDR_CNTRL=MAXDATA=0x50000000
```

```
/opt/protegility/defiance_dps/bin/pepsrvctrl start
```

```
unset LDR_CNTRL
```

- Run the AP Java installer using the following command.

```
./APJavaSetup_AIX_ppc64_<version>.sh
```

The following table lists the default installation directory for the AP Java on the AIX platform.

Table 12-23: AP Java Installation Directory

| Platform | Directory |
|----------|---|
| AIX | <i>/opt/protegility/applicationprotector/java/lib</i> |

Verify that the following libraries are installed in the */opt/protegility/applicationprotector/java/lib* directory:

- ApplicationProtectorJava.jar*
- ApplicationProtectorJava.properties*
- jna-5.8.0.jar*
- jna-platform-5.8.0.jar*
- jpeplite.plm*

- Link your application to the installed *ApplicationProtectorJava.properties* and *.jar* files.

For more information about configuring the Application Protector Java and verifying whether the protector has been successfully installed, refer to the sections [Configuring Application Protector Java](#) and [Checking Installation Success](#).

The AP Java is installed successfully.

12.5.2.4 Configuring Application Protector Java

This section describes how to configure the AP Java on the AIX, Linux or Unix, and Windows platform.

► To configure the AP Java on the AIX, Linux or Unix, and Windows platform:

1. Setup the Java classpath.

Table 12-24: Java Classpath

| Operating System | Classpath |
|------------------|---|
| AIX/Linux/Unix | /opt/protegity/applicationprotector/java/lib |
| Windows | C:\program files\protegity\defiance AP\Java\lib |

2. Before the trusted application that would be using the AP Java, can successfully load the *ApplicationProtectorJava.jar* file, ensure that the Java classpath and *jpeelite.plm* path are accurately setup and configured.
3. Deploy a policy to test the application.

Note: For more information about deploying a policy, refer to the section *Creating and Deploying Policies* in the *Protegity Policy Management Guide 9.1.0.2*.

12.5.2.5 Checking Installation Success

This section describes how to verify that the AP Java is successfully installed.

► To verify that the AP Java is successfully installed:

1. Initialize the AP Java.

For more information about the AP Java initialization API, refer to the section *getProtector* in the *Protegity APIs, UDFs, and Commands Reference Guide 9.1.0.0*.

2. Run the *getVersion* method.

The AP Java version number appears on the screen.

3. To verify that the entire installation is ready to protect data, test some sample data with the *protect* method.

For more information on the *getVersion* and *protect* methods, refer to section *Application Protector (AP) Java APIs* in the *Protegity APIs, UDFs, Commands Reference Guide 9.1.0.0*.

For more information about using the AP Java sample, refer to the section *Running AP Java APIs* in the *Protegity Application Protector Guide 9.1.0.0*.

12.5.2.6 Uninstalling Application Protector Java

This section describes how to uninstall the different components of the AP Java on different platforms.

12.5.2.6.1 Uninstalling the Log Forwarder from AIX, Linux, or Unix

This section describes how to uninstall the Log Forwarder from an AIX, Linux, or Unix platform.

► To uninstall the Log Forwarder from the AIX, Linux, or Unix platform:

1. Navigate to the `/opt/protegility/fluent-bit/bin` directory.
2. Stop the `fluentbit` component using the following command.
`./logforwarderctrl stop`
3. Delete the `fluentbit` directory.

The Log Forwarder is uninstalled.

12.5.2.6.2 Uninstalling the Log Forwarder from the Windows Platform

This section describes how to uninstall the Log Forwarder from the Windows platform.

► To uninstall the Log Forwarder on Windows:

1. Perform the following steps to stop the Log Forwarder.
 - a. From the Windows Start Menu, search and select **Services**.
 - b. Navigate to the `Logforwarder` directory.
 - c. Right-click the **Logforwarder** service and click **Stop**.
2. Run the `logforwarder` uninstall utility located in the `C:\Program Files\Protegility\fluent-bit` directory.
3. After the Log Forwarder is uninstalled, delete the following directory.

`fluent-bit`

The Log Forwarder is uninstalled.

12.5.2.6.3 Uninstalling the PEP Server from AIX, Linux or Unix

This section describes how to uninstall the PEP server from an AIX, Linux or Unix platform.

► To uninstall the PEP server from the AIX, Linux or Unix platform:

1. Navigate to the `/opt/protegility/defiance_dps/bin` directory.
2. Stop the PEP server using the following command.
`./pepsrvctrl stop`
3. Delete the `/opt/protegility/defiance_dps` directory.

The PEP server is uninstalled.

12.5.2.6.4 Uninstalling PEP Server from Windows

This section describes how to uninstall the PEP server from the Windows platform.

► To uninstall the PEP Server from the Windows platform:

1. Stop the PEP server by performing the following steps.
 - a. From the Windows Start Menu, search and select *Services*.
 - b. Navigate to the *Protegility PEP Server* service.
 - c. Right-click the service and click **Stop**.
2. Navigate to the *C:\Program Files\Protegility\defiance_dps* directory.
3. Run the following file to uninstall the PEP server.
unins000.exe
4. Delete the following directory.
C:\Program Files\Protegility\defiance_dps

The PEP server is uninstalled.

12.5.2.6.5 Uninstalling Application Protector Java from AIX, Linux or Unix

This section describes how to uninstall the AP Java from an AIX, Linux or Unix platform.

► To uninstall the AP Java from the AIX, Linux or Unix platform:

1. Navigate to the */opt/protegility/applicationprotector* directory.
2. Delete the *applicationprotector* directory.

The AP Java is uninstalled.

12.5.2.6.6 Uninstalling the Application Protector Java from the Windows Platform

This section describes how to uninstall the AP Java from a Windows platform.

► To uninstall the AP Java from the Windows platform:

1. Navigate to the *C:\Program Files\Protegility\Defiance AP* installation directory.
2. Run the uninstall utility located in the *C:\Program Files\Protegility\Defiance AP* directory.
3. Delete the *C:\Program Files\Protegility\Defiance AP* directory.

The AP Java is uninstalled.

12.5.3 Installing Application Protector (AP) Python

The Application Protector (AP) Python provides APIs for data protection and only Trusted Applications or users of the APIs can use AP Python. The **Policies and Trusted Applications** option under **Policy Management** in the ESA allows the user to add, edit, and rename the Trusted Applications.

You can choose to install AP Python in the following modes:

- ***AP Python in a production environment*** - In this mode, you can use the AP Python APIs to protect, unprotect, and reprotect the data using the data elements created in the ESA. To use AP Python in this mode, the ESA must be already available and the PEP server must be running.
- ***AP Python in a development environment*** - In this mode, a set of sample users and data elements have been provided that can be used to simulate the behavior of the AP Python APIs. You do not require the ESA or the PEP server to run AP Python in this mode. You can use this mode to test the integration of your application with the AP Python APIs. This mode is also known as the AP Python mock implementation.

Warning: When you use the AP Python APIs with the sample users and data elements, the data is *not* tokenized or encrypted. **Use this mode only for testing purposes.**

Caution: Ensure that only one of the two AP Python modes are installed in your environment at any one time. That means do not install both the production mode and development mode of AP Python in your environment at the same time.

Note: You can switch from the development mode to production mode without any changes to the code.

12.5.3.1 Setting up AP Python on Linux or Unix in a Production Environment

This section describes how to install the AP Python on a Linux or Unix platform in a production environment.

Before you begin

Ensure that the following prerequisites are met before installing the AP Python:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

- Python 3, any version from 3.7 to 3.11 must be installed on the machine where you are installing the AP Python.
- Latest version of *pip*, which is a package manager for Python modules, must be installed on the machine where you are installing the AP Python.

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

► To set up the AP Python on the Linux or Unix platform in the production environment:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_PY-3.11-<version>.tgz* file to any location on the machine where you want to install the protector.
2. Extract the AP Python installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_PY-3.11-<version>.tgz
```

The following setup files are extracted:

- *APPythonSetup_Linux_x64-<version>.tar*
- *APPythonDevSetup_Linux_x64-<version>.tar*



- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*

12.5.3.1.1 Installing Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.3.1.1.1 Installing Log Forwarder on Linux or Unix Using Linux Installer

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer.

► To install the Log Forwarder on a Linux or Unix platform using the Linux installer:

1. Run the Log Forwarder installer using the following command.

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store port number where the Log Forwarder listens for logs.

Note: The default port number is *15780*.

3. Press ENTER.

The added Audit Store endpoint appears on the screen.

The prompt to enter an additional Audit Store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

4. If you want to add more than one Audit Store endpoint, then type *y* else type *n*.

Note: If you need to add *n* Audit Store endpoints, then repeat the *Step 2* and *Step 3* *n* times.

5. Type the *y* key to install into the destination directory.

The Log Forwarder is installed in the */opt/protegility/fluent-bit/* directory.

6. Start the Log Forwarder component by using the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl start
```

The Log Forwarder is successfully installed.

12.5.3.1.1.2 Silent Mode of Installation of Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-25: Parameter List for Silent Installation

| Parameter | Description |
|------------------|---|
| --endpoint or -e | The IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is 15780. |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *--dir* parameter to the command to specify the Log Forwarder installation directory and *--pemdir* parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>] --dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.5.3.1.2 Installing PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.3.1.2.1 Installing PEP Server on Linux or Unix using Linux Installer

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer.

► To install the PEP server on a Linux or Unix platform:

1. Run the PEP server installer using the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the ESA Host Name or IP Address.
3. Press ENTER.



The prompt to enter the username for downloading certificates appears.

```
Please enter the user name for downloading certificates:
```

4. Enter the username for downloading the certificates.
5. Press ENTER.

The prompt to enter the password for downloading the certificates appears.

```
Please enter the password for downloading certificates:
```

6. Press ENTER to install into the destination directory.

Directories are created under `/opt/protegility/defiance_dps` by default, and the required installation files are installed in these directories.

Caution: Ensure that the ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

To manually install the certificates to the `/opt/protegility/defiance_dps/data` directory of the PEP server, navigate to the `/opt/protegility/defiance_dps/bin` directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>]
[-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the *administrator* user.

Verify that the following files have been copied to the `/opt/protegility/defiance_dps/data` directory:

- `CA.pem`
- `keyinternal.plm`
- `pepperver.cfg`
- `pepperver.pid`
- `authesa.plm`
- `cert.key`
- `cert.pem`
- `certkeyup.bin`

7. Start the PEP server by using the following command.

```
/opt/protegility/defiance_dps/bin/pepsrvctrl start
```

The PEP server is successfully installed.

12.5.3.1.2.2 Silent Mode of Installation of PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform through the Silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-26: Parameter List for Silent Installation

| Parameter | Description |
|-----------------------|---|
| <code>-esa</code> | Specifies the ESA IP address. |
| <code>-esaport</code> | Specifies the ESA port, which is optional. The default value is <code>8443</code> . |

| Parameter | Description |
|-----------|--|
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the /opt/protegility directory. |

At the command prompt, type the following command from the installer directory.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password> -dir <installation-directory-path>
```

12.5.3.1.3 Installing Application Protector Python on Linux or Unix

This section describes how to install the AP Python in a production environment on a Linux or Unix platform.

► To install the AP Python in a production environment on a Linux or Unix platform:

- Run the following command to install the AP Python in a production environment.

```
pip install APPythonSetup_Linux_x64_<version>.tar
```

This installs the AP Python in the production environment on the Linux or Unix platform.

The default installation directory for the Linux or Unix platform is /usr/local/lib/python<version>/site-packages.

- Verify that the following directories are created in the AP Python installation directory:
 - appython
 - mocks
 - pypepprovider
- Perform the following steps to access the AP Python Pydoc, which contains the API documentation.
 - Run the following command to extract the AP Python setup file:


```
tar -xvf APPythonSetup_Linux_x64_<version>.tar
```

 The appython-<version> directory is extracted.
 - Navigate to the appython-<version>\docs\ directory.
 - Open the index.html file in a browser to access the AP Python Pydoc.

Note:

If you are setting up the AP Python in a virtual Linux or Unix environment, then convert the appython-<version>\docs directory to a zip file, download it locally, and then open the index.html file in a browser to access the AP Python Pydoc.



12.5.3.1.4 Verifying the Installation of AP Python

This section describes how to verify if the AP Python is installed successfully.

► To verify that the AP Python has been successfully installed:

1. Login to the machine where the AP Python is installed.
2. To verify the version of the AP Python, run the following command:

```
pip list
```

The name and version of the installed AP Python package are displayed on the console.

```
appython      9.1.0.0.4
```

3. Alternatively, run the *get_version* API.

This method does not validate the Trusted Application.

To verify that the AP Python is ready to protect data, test some sample data with the *protect* method.

For more information about the *get_version* and *protect* APIs, refer to the section *Application Protector (AP) Python APIs* in the [Protegility APIs, UDFs, and Commands Reference Guide 9.1.0.0](#).

12.5.3.1.5 Uninstalling Application Protector Python from the Production Environment

This section describes how to uninstall the Log Forwarder, PEP server and AP Python from the production environment.

12.5.3.1.5.1 Uninstalling the Log Forwarder from AIX, Linux, or Unix

This section describes how to uninstall the Log Forwarder from an AIX, Linux, or Unix platform.

► To uninstall the Log Forwarder from the AIX, Linux, or Unix platform:

1. Navigate to the */opt/protegility/fluent-bit/bin* directory.
2. Stop the *fluentbit* component using the following command.
`./logforwarderctrl stop`
3. Delete the *fluentbit* directory.

The Log Forwarder is uninstalled.

12.5.3.1.5.2 Uninstalling the PEP Server from AIX, Linux or Unix

This section describes how to uninstall the PEP server from an AIX, Linux or Unix platform.

► To uninstall the PEP server from the AIX, Linux or Unix platform:

1. Navigate to the */opt/protegility/defiance_dps/bin* directory.

2. Stop the PEP server using the following command.
`./pepsrvctrl stop`
3. Delete the `/opt/protegility/defiance_dps` directory.

The PEP server is uninstalled.

12.5.3.1.5.3 Uninstalling Application Protector Python

This section describes how to uninstall AP Python from the production environment on the Windows, Linux, or Unix platforms.

► To uninstall the AP Python from the production environment:

1. Login to the machine from where you want to uninstall the AP Python.
2. Uninstall the AP Python by running the following command.

```
pip uninstall appython
```

12.5.3.2 Setting up AP Python on Windows in a Production Environment

This section describes how to install the AP Python on a Windows platform in a production environment.

Before you begin

Ensure that the following prerequisites are met before installing the AP Python:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

- Python 3, version 3.7 must be installed on the machine where you are installing the AP Python.
- Latest version of *pip*, which is a package manager for Python modules, must be installed on the machine where you are installing the AP Python.

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

► To setup the AP Python on the Windows platform in the production environment:

1. Download the *ApplicationProtector_WIN-ALL-64_x86-64_PY-3.7-<version>.zip* file to any location on the machine where you want to install the protector.
2. Extract the files from the *ApplicationProtector_WIN-ALL-64_x86-64_PY-3.7-<version>.zip* installation package.

The following setup files are extracted:

- *LogforwarderSetup_Windows_x64-<version>.exe*
- *PepServerSetup_Windows_x64-<version>.exe*
- *APPythonSetup_Windows_x64-<version>.tar*



- APPythonDevSetup_Windows_x64_<version>.tar

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

12.5.3.2.1 Installing Log Forwarder on Windows

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard and through the Silent mode of installation.

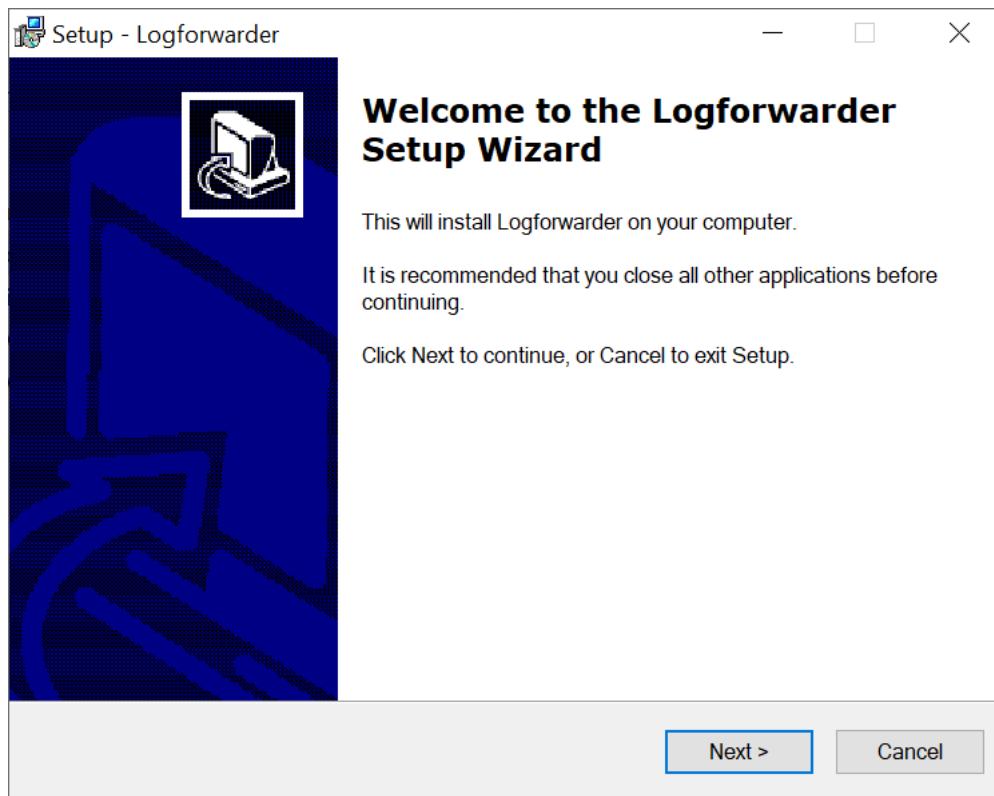
12.5.3.2.1.1 Using Windows Wizard

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard.

► To install the Log Forwarder on the Windows platform through the Windows Wizard:

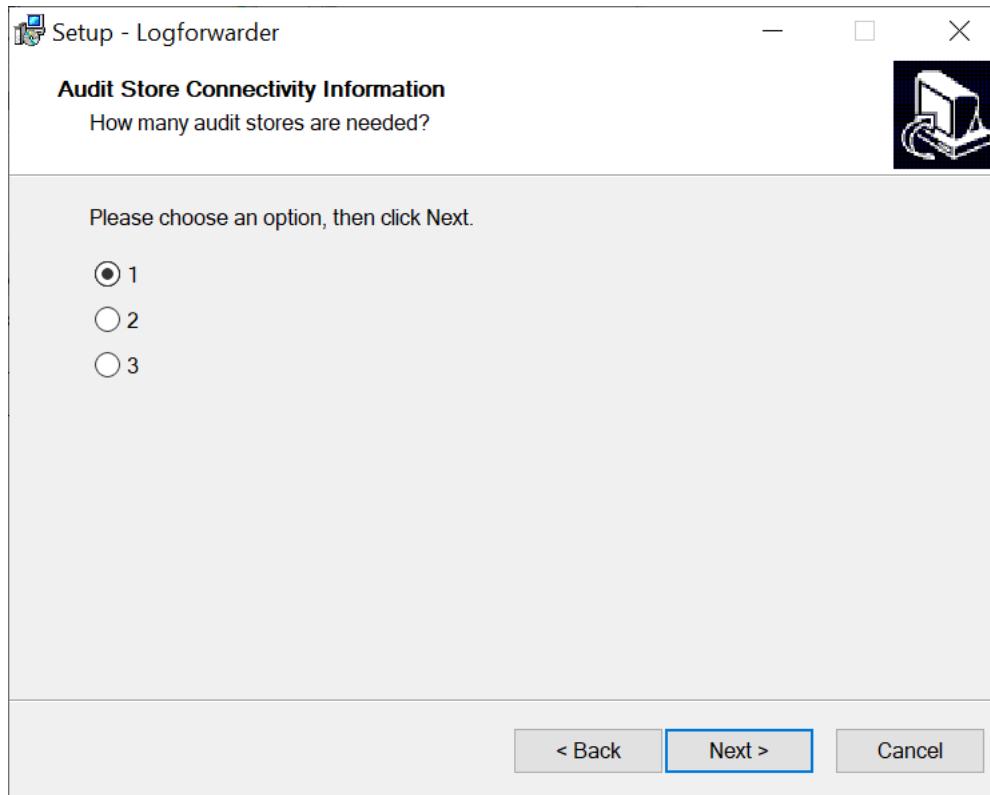
1. Run the *LogforwarderSetup_Windows_x64_<version>.exe* file from the created directory.

The **Welcome to the Log Forwarder Setup Wizard** screen appears.



2. Click **Next**.

The **Audit Store Connectivity Information** screen appears to select the number of audit stores that are needed.

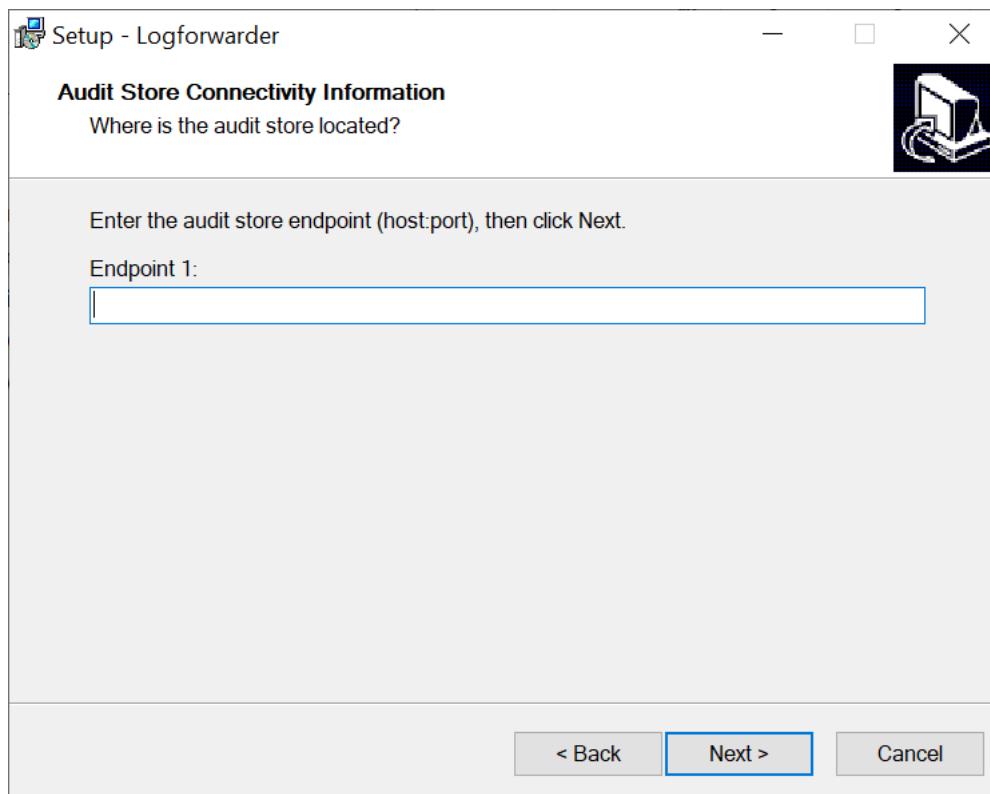


3. Select the number of Audit Stores as required.

The maximum number of Audit Stores that can be configured is 3.

4. Click **Next**.

The **Audit Connectivity Information** screen appears to enter the location of the audit store.



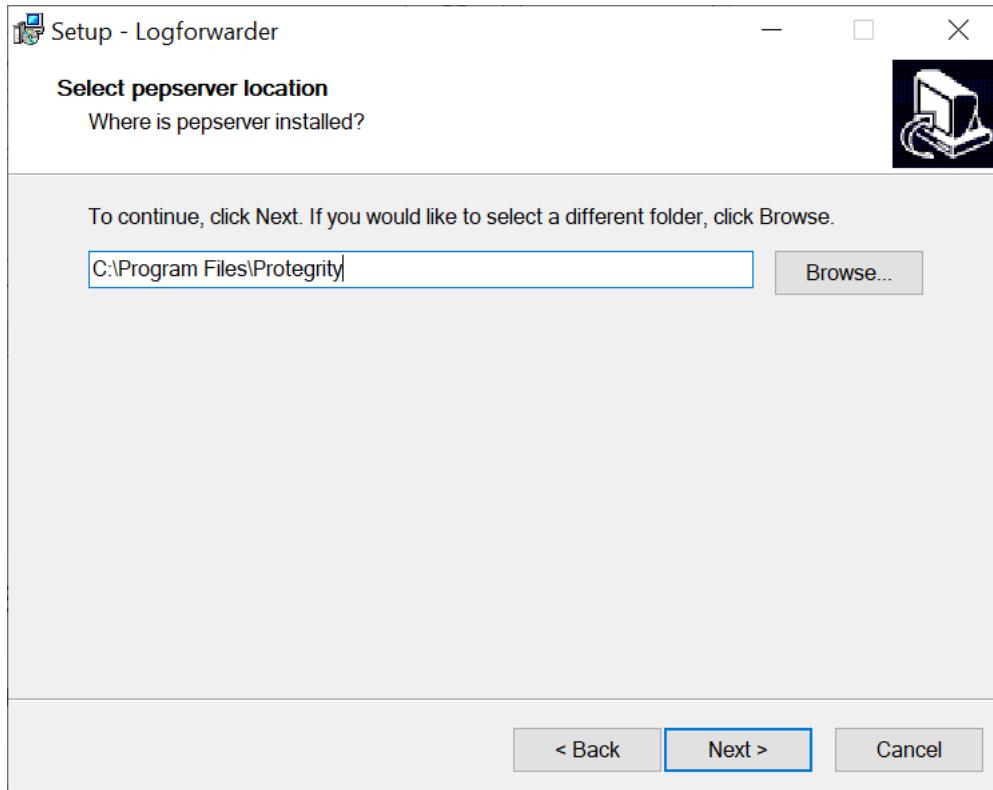
5. Enter the Hostname/IPAddress:port where the Audit Store is configured.

Note:

The default port number for the Audit Store is **9200**.

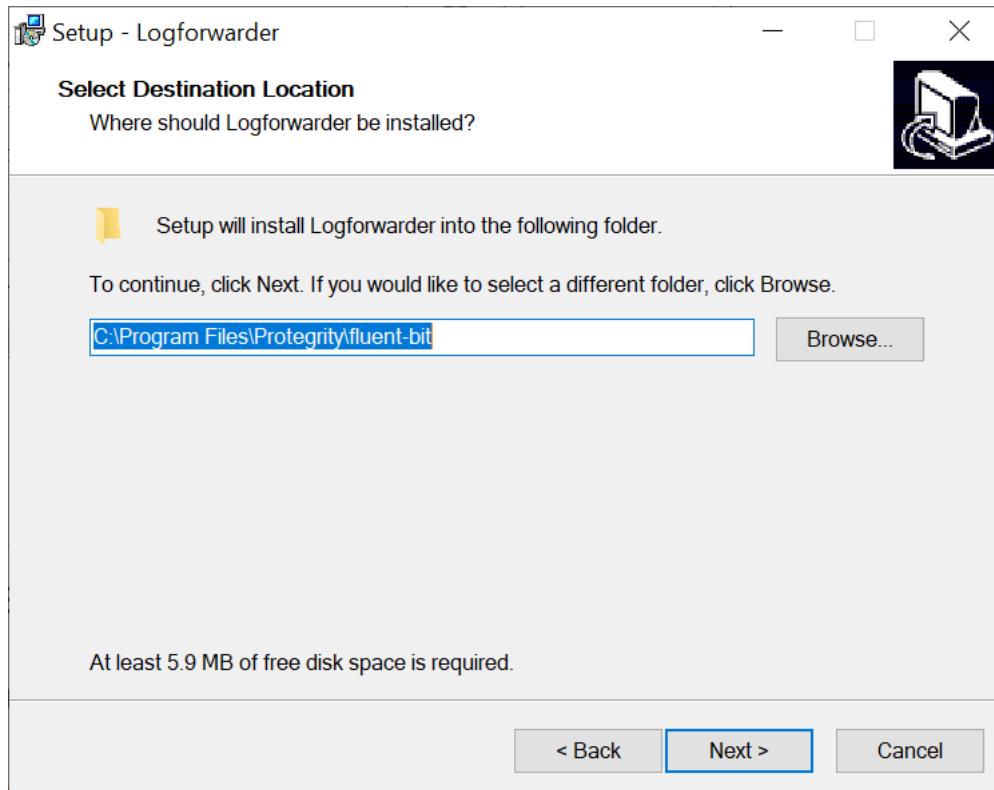
6. Click **Next**.

The **Select pepserver location** screen appears.



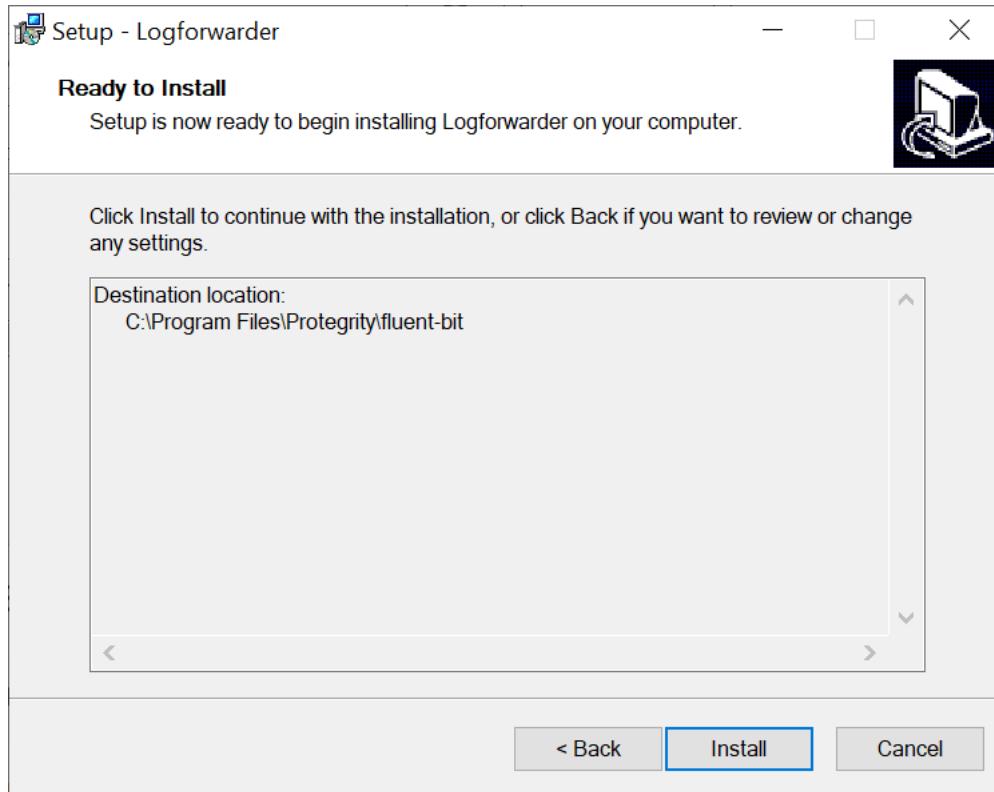
7. Click **Next**.

The **Select Destination Location** screen appears.



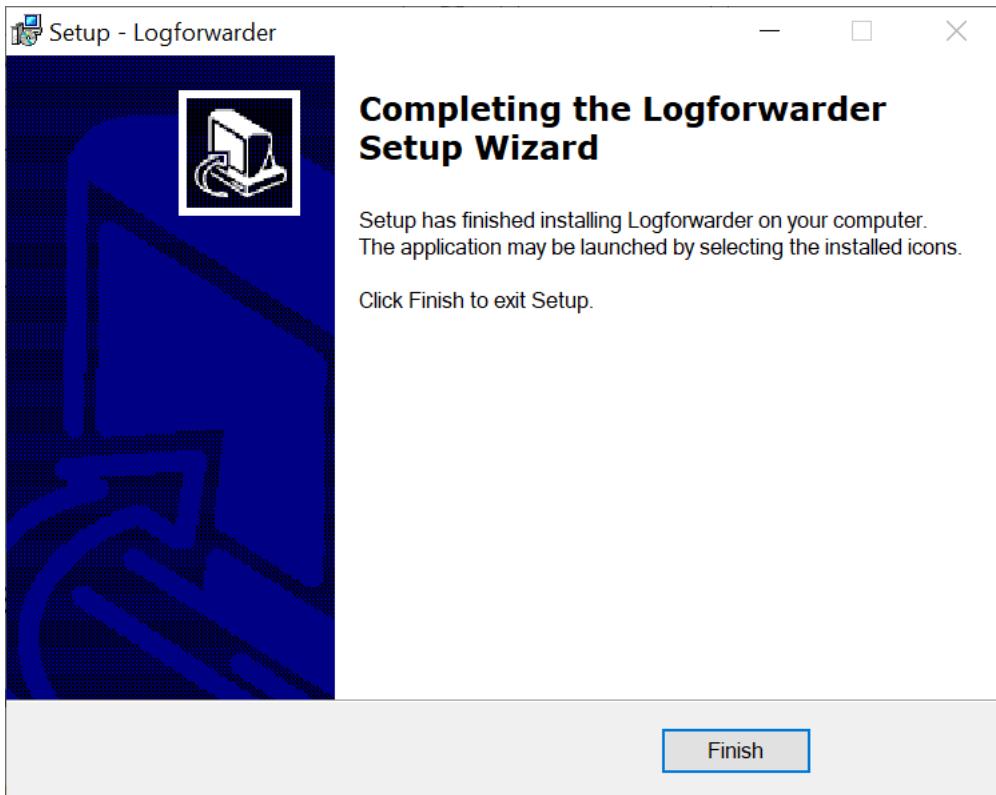
8. Set the installation directory for the Log Forwarder to *C:\Program Files\Protegity\fluent-bit*.
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Logforwarder Setup Wizard** screen appears.



11. From the **Completing the Logforwarder Setup Wizard** screen, click **Finish** to complete the installation and exit.

Post successful installation, you can find the installation files in the directories that are created under the defined installation directory.

The Log Forwarder is installed successfully.

12.5.3.2.1.2 Using Silent Mode of Installation

This section describes how to install the Log Forwarder on a Windows platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-27: Parameter List for Silent Installation

| Parameter | Description |
|------------------------------------|--|
| -endpoint1, -endpoint2, -endpoint3 | <p>Audit Store IP address and the port number where the Log Forwarder listens for logs.</p> <p>Note: The default port number is <i>9200</i>.</p> <p>Note: The parameters <i>-endpoint2</i> and <i>-endpoint3</i> are optional.</p> |
| -dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility\fluent-bit</i> . |

| Parameter | Description |
|-----------|--|
| -pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C : \Program Files\Protegility</i> . |

At the command prompt, type the following command from the installer directory.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the Log Forwarder installation directory and the *-pemdir* parameter to the command to specify the PEP server installation directory.

The following snippet displays a sample command.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>] -dir <Log Forwarder installation directory> -pemdir <PEP server installation directory>
```

12.5.3.2.2 Installing PEP Server on Windows

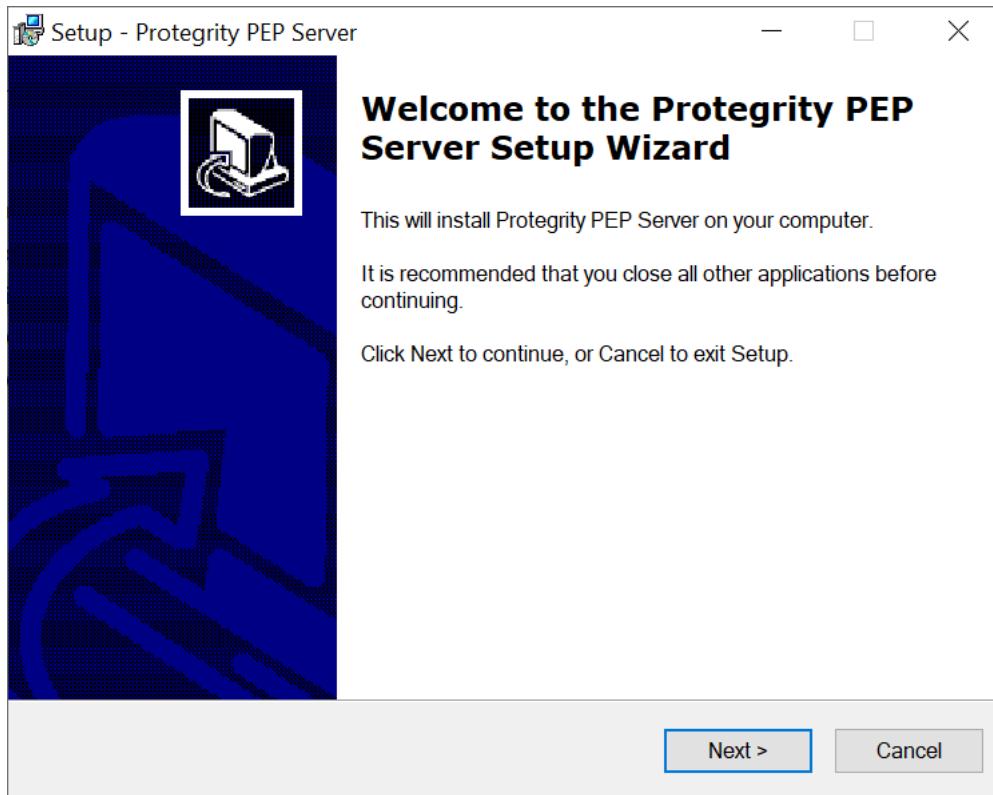
This section describes how to install the PEP server on a Windows platform through the Windows Wizard and through the Silent mode of installation.

12.5.3.2.2.1 Using Windows Wizard

This section describes how to install the PEP server on a Windows platform through the Windows Wizard.

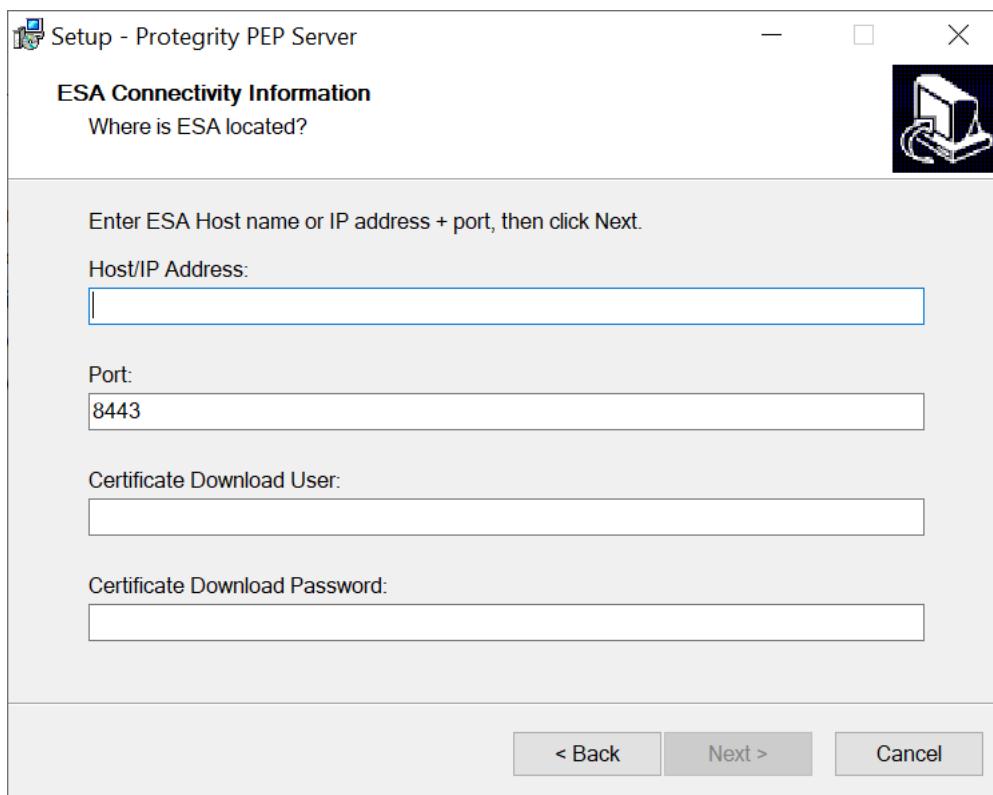
► To install the PEP server on the Windows platform through the Windows Wizard:

1. Run the *PepServerSetup_Windows_<bit-version>_<version>.exe* file from its installed directory.
The **Welcome to the Protegility PEP Server Setup Wizard** appears.



2. Click **Next**.

The **ESA Connectivity Information** screen appears.



3. Enter the ESA Host name or IP Address in the **Host/IP Address** field.

- Enter the ESA host listening port details in the **Port** field.

Note: Ensure that the ESA is up and running with the HubController service in running status to enable the automatic downloading of certificates.

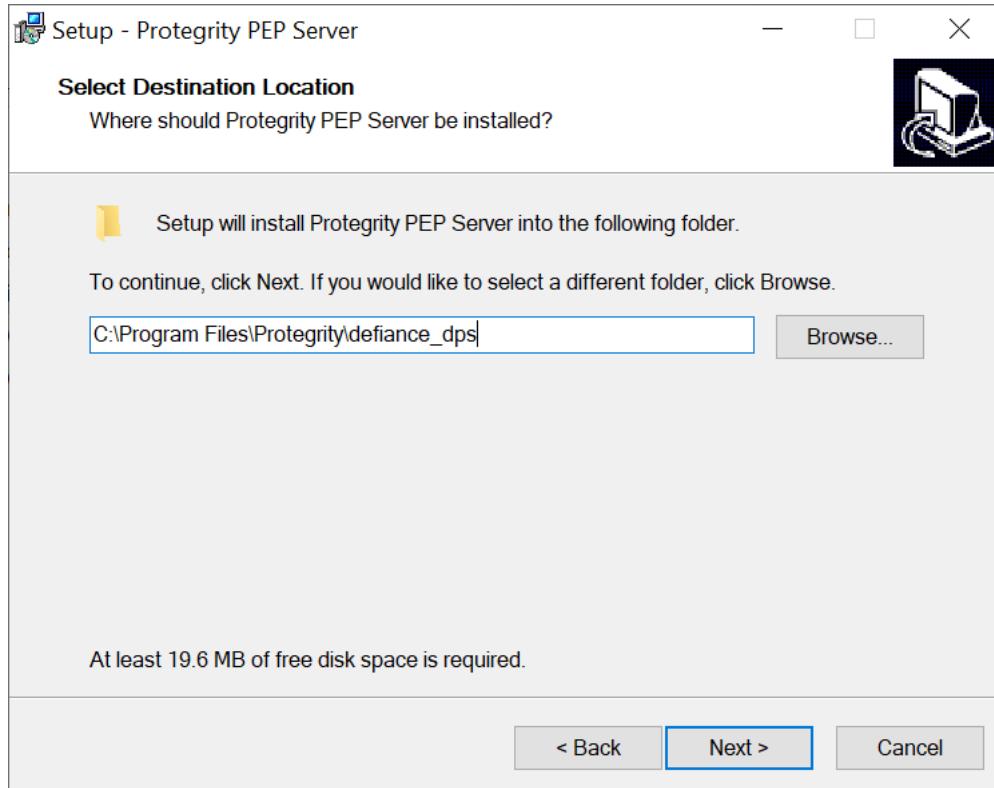
- Enter the **Certificate Download User**.

Note: It is recommended to use *admin* as the user.

- Enter the **Certificate Download Password**.

- Click **Next**.

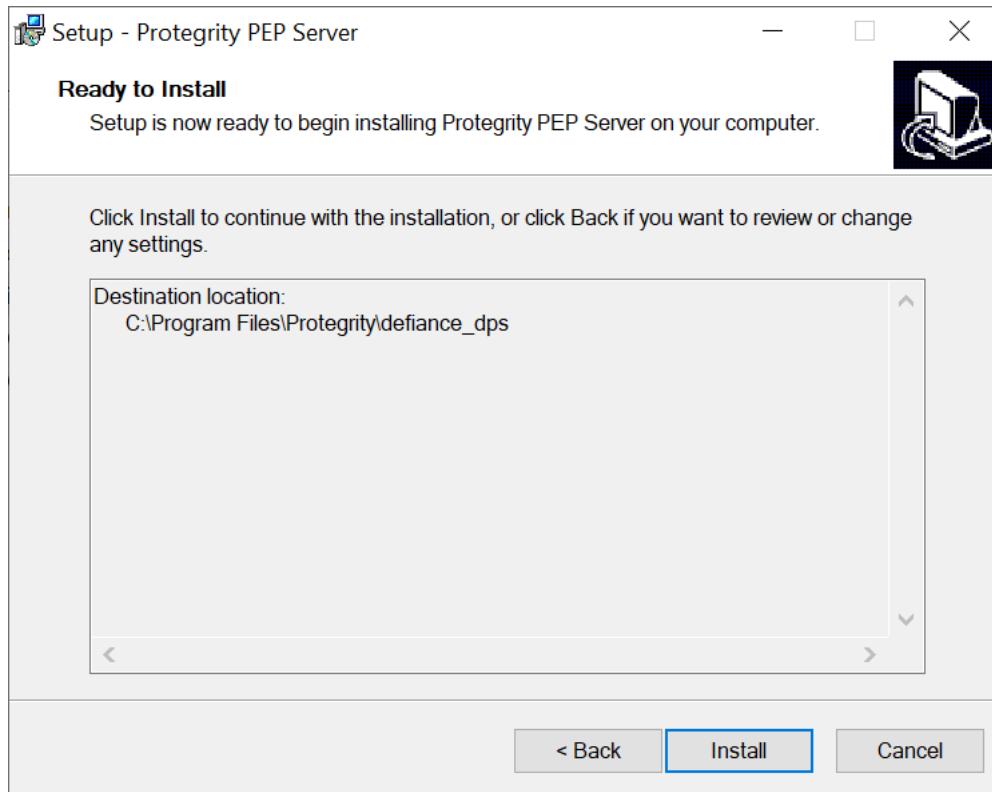
The **Select Destination Location** screen appears.



- Set the installation directory for the PEP server to *C:\Program Files\Protegility\defiance_dps*.

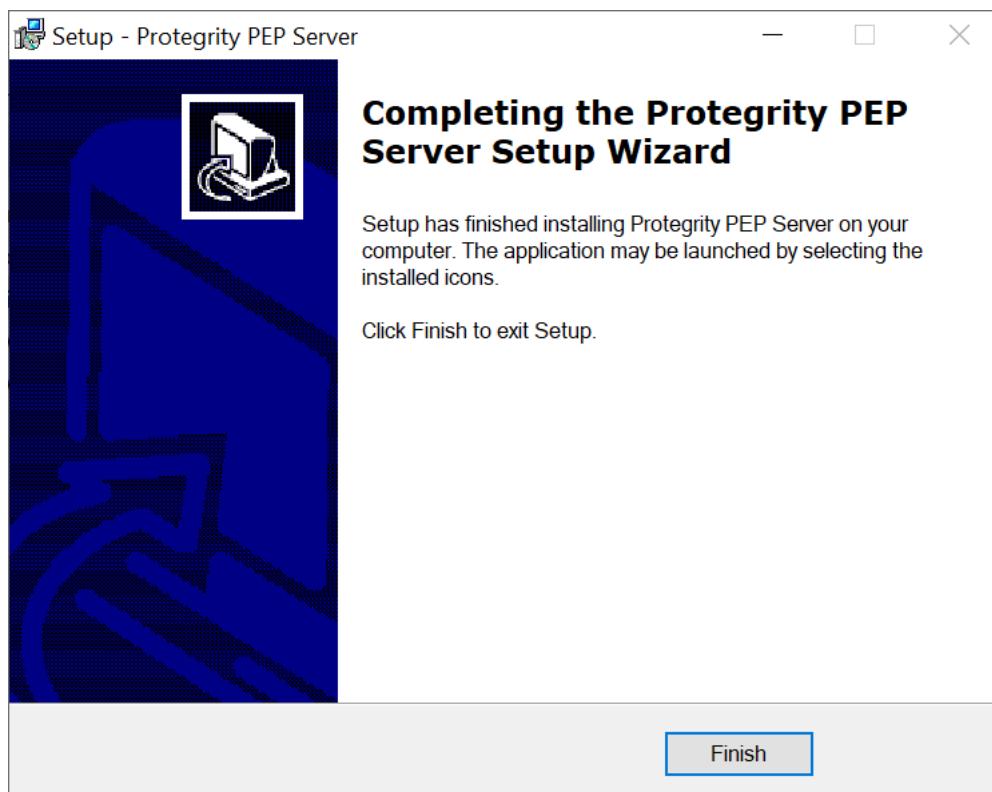
- Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Protegility PEP Server Setup Wizard** screen appears.



11. From the **Completing the Protegility PEP Server Setup Wizard** screen, click **Finish** to complete the installation and exit.

The directories are created under the installation directory that was defined and the installation files are installed in these directories.

Note:

If you want to manually install the certificates in the *<path to where pepserver is installed>\defiance_dps\data* directory of the PEP server, then navigate to the *<path to where pepserver is installed>\defiance_dps\bin* directory, and run the following command:

```
GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the user *admin*.

The PEP server is installed successfully.

12.5.3.2.2 Using Silent Mode of Installation

This section describes how to install the PEP server on a Windows platform through the silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.

Table 12-28: Parameter List for Silent Installation

| Parameter | Description |
|-----------|---|
| -esa | Specifies the ESA IP address. |
| -esaport | Specifies the ESA port, which is optional. The default value is 8443. |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i><path to where pepserver is installed>\defiance_dps</i> . |

At the command prompt, type the following command from the installer directory.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
-dir <installation-directory-path>
```

12.5.3.2.3 Installing Application Protector Python on Windows

This section describes how to install the AP Python on a Windows platform in a production environment.

► To install the AP Python on the Windows platform in the production environment:



- Run the following executable file to install the AP Python in a production environment.

```
pip install APPythonDevSetup_Windows_x64_<version>.tar
```

This installs the AP Python.

The default installation directory for the Windows platform is

C:\Users\<User>\AppData\Local\Programs\Python\Python37\Lib\site-packages.

- Verify that the following directories are created in the AP Python installation directory:

- appython*
- pypepprovider*

- Perform the following steps to access the AP Python Pydoc, which contains the API documentation.

- Extract the following AP Python setup file.

```
APPythonDevSetup_Windows_x64_<version>.tar
```

The *APPythonDevSetup_Windows_x64_<version>* directory is extracted.

- Navigate to the *APPythonDevSetup_Windows_x64_<version>\dist\appython-<version>.tar* directory.

- Extract the *appython-<version>.tar* in the *\dist* directory.

The *appython-<version>* directory is extracted.

- Navigate to the *appython-<version>\docs\index* directory.

- Open the *index.html* file in a browser to access the AP Python Pydoc.

The AP Python is successfully installed in the production environment.

12.5.3.2.4 Verifying the Installation of AP Python

This section describes how to verify if the AP Python is installed successfully.

► To verify that the AP Python has been successfully installed:

- Login to the machine where the AP Python is installed.
- To verify the version of the AP Python, run the following command:

```
pip list
```

The name and version of the installed AP Python package are displayed on the console.

```
appython          9.1.0.0.4
```

- Alternatively, run the *get_version* API.

This method does not validate the Trusted Application.

To verify that the AP Python is ready to protect data, test some sample data with the *protect* method.

For more information about the *get_version* and *protect* APIs, refer to the section *Application Protector (AP) Python APIs* in the *Protegility APIs, UDFs, and Commands Reference Guide 9.1.0.0*.

12.5.3.2.5 Uninstalling Application Protector Python from the Production Environment

This section describes how to uninstall the Log Forwarder, PEP server and AP Python from the production environment.

12.5.3.2.5.1 Uninstalling the Log Forwarder from the Windows Platform

This section describes how to uninstall the Log Forwarder from the Windows platform.

► To uninstall the Log Forwarder on Windows:

1. Perform the following steps to stop the Log Forwarder.
 - a. From the Windows Start Menu, search and select **Services**.
 - b. Navigate to the *Logforwarder* directory.
 - c. Right-click the **Logforwarder** service and click **Stop**.
2. Run the **logforwarder** uninstall utility located in the *C:\Program Files\Protegility\fluent-bit* directory.
3. After the Log Forwarder is uninstalled, delete the following directory.

fluent-bit

The Log Forwarder is uninstalled.

12.5.3.2.5.2 Uninstalling PEP Server from Windows

This section describes how to uninstall the PEP server from the Windows platform.

► To uninstall the PEP Server from the Windows platform:

1. Stop the PEP server by performing the following steps.
 - a. From the Windows Start Menu, search and select *Services*.
 - b. Navigate to the *Protegility PEP Server* service.
 - c. Right-click the service and click **Stop**.
2. Navigate to the *C:\Program Files\Protegility\defiance_dps* directory.
3. Run the following file to uninstall the PEP server.
unins000.exe
4. Delete the following directory.
C:\Program Files\Protegility\defiance_dps

The PEP server is uninstalled.

12.5.3.2.5.3 Uninstalling Application Protector Python

This section describes how to uninstall AP Python from the production environment on the Windows, Linux, or Unix platforms.

► To uninstall the AP Python from the production environment:

1. Login to the machine from where you want to uninstall the AP Python.
2. Uninstall the AP Python by running the following command.

```
pip uninstall appython
```

12.5.3.3 Setting Up AP Python in a Development Environment

This section describes how to install the AP Python in a development environment.

12.5.3.1 Setting up AP Python on Linux or Unix in a Development Environment

This section describes how to install the AP Python in a development environment on a Linux or Unix platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP Python:

- Python 3, any version from 3.7 to 3.11, must be installed on the machine where you are installing the AP Python.
- Latest version of *pip*, which is a package manager for Python modules, must be installed on the machine where you are installing the AP Python.

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

► To install the AP Python in a development environment on a Linux or Unix platform:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_PY-3.11_<version>.tgz* file to any location on the machine where you want to install the protector.
2. Extract the AP Java installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_PY-3.11_<version>.tgz
```

The following setup files are extracted:

- *APPythonSetup_Linux_x64_<version>.tar*
- *APPythonDevSetup_Linux_x64_<version>.tar*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*

3. Run the following script to install AP Python in a development environment.

```
pip install APPythonDevSetup_Linux_x64_<version>.tar
```

This installs the AP Python in the development environment on the Linux or Unix platform.

The default installation directory for the Linux or Unix platform is */usr/local/lib/python<version>/site-packages*.

4. Verify that the following directories are created in the AP Python installation directory:

- *appython*
- *mocks*
- *pypepprovider*

- To verify the version of the AP Python, run the following command:

```
pip list
```

The name and version of the installed AP Python package are displayed on the console.

```
appython-dev           9.1.0.0.4
```

Note: For more information on how to use the AP Python APIs in a development environment, refer to the section *Using AP Python in Development Environment* in the *Protegility APIs, UDFs, and Commands Reference Guide 9.1.0.0*.

- Perform the following steps to access the AP Python Pydoc, which contains the API documentation.

- Run the following command to extract the AP Python setup file.

```
tar -xvf APPythonDevSetup_Linux_x64_<version>.tar
```

The *appython_dev-<version>* directory is extracted.

- Navigate to the *appython_dev-<version>\docs* directory.
- Open the *index.html* file in a browser to access the AP Python Pydoc.

Note:

If you are setting up the AP Python in a virtual Linux or Unix environment, then convert the *appython-<version>\docs* directory to a zip file, download it locally, and then open the *index.html* file in a browser to access the AP Python Pydoc.

12.5.3.3.2 Setting up AP Python on Windows in a Development Environment

This section describes how to install the AP Python on a Windows platform in a development environment.

Before you begin

Ensure that the following prerequisites are met before installing the AP Python:

- Python 3, version 3.7, must be installed on the machine where you are installing the AP Python.
- Latest version of *pip*, which is a package manager for Python modules, must be installed on the machine where you are installing the AP Python.

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

► To setup the AP Python on the Windows platform in the development environment:

- On the target machine, extract the AP Python installation package.

ApplicationProtector_WIN-ALL-64_x86-64_PY-3.7_<version>.zip

The following setup files are extracted:

- LogforwarderSetup_Windows_x64_<version>.exe*
- PepServerSetup_Windows_x64_<version>.exe*
- APPythonSetup_Windows_x64_<version>.tar*



- *APPythonDevSetup_Windows_x64_<version>.tar*

Caution: Ensure that you install either one of the two AP Python modes, such as, production or development, in your environment at a time.

2. Extract the following installable package to install the AP Python in a Development environment.

```
pip install APPythonDevSetup_Windows_x64_<version>.tar
```

The *APPythonDevSetup_Windows_x64_<version>* directory is extracted.

This installs the AP Python in the development environment.

The default installation directory for the Windows platform is

C:\Users\<User>\AppData\Local\Programs\Python\Python37\Lib\site-packages.

3. Verify that the following directories are created in the AP Python installation directory:

- *appython*
- *mocks*
- *pypepprovider*

4. To verify the version of the AP Python, run the following command.

```
pip list
```

The name and version of the installed AP Python package are displayed on the console.

```
appython-dev (9.1.0.0.4)
```

For information about using the AP Python APIs in a Development environment, refer to the section *Using AP Python in a Development Environment* in the *APIs, UDFs, and Commands Reference Guide 9.1.0.0*.

5. Perform the following steps to access the AP Python Pydoc, which contains the API documentation.

- a. Extract the AP Python setup file.

```
APPythonDevSetup_Windows_x64_<version>.tar
```

The *APPythonDevSetup_Windows_x64_<version>* directory is extracted.

- b. Navigate to the *APPythonDevSetup_Windows_x64_<version>\dist\appython_dev-<version>.tar* directory.

- c. Extract the *appython_dev-<version>.tar* in the *\dist* directory.

The *appython_dev-<version>* directory is extracted.

- d. Navigate to the *appython_dev-<version>\docs\index* directory.

- e. Open the *index.html* file in a browser to access the AP Python Pydoc.

The AP Python is successfully installed in the development environment.

12.5.3.3 Uninstalling AP Python from the Development Environment

This section describes how to uninstall AP Python from the development environment on the Windows, Linux, or Unix platforms.

- To uninstall the AP Python from the development environment:

1. Login to the machine where AP Python is installed.
2. Uninstall the AP Python by running the following command.

```
pip uninstall appython-dev
```

12.5.4 Installing Application Protector Go

This section explains a standard setup of installing and uninstalling the Application Protector Go on the Linux platform.

12.5.4.1 Setting up Application Protector Go on Linux or Unix

This section describes how to install the AP Go on a Linux or Unix platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP Go:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

► To setup the AP Go on a Linux or Unix platform:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_GO-1-64_<version>.tgz* file to any location on the machine where you want to install the protector.
2. Extract the AP Go installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_GO-1-64_<version>.tgz
```

The following setup files are extracted:

- *APGoSetup_Linux_x64_<version>.sh*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*

12.5.4.1.1 Installing Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.4.1.1.1 Installing Log Forwarder on Linux or Unix Using Linux Installer

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer.

► To install the Log Forwarder on a Linux or Unix platform using the Linux installer:

- Run the Log Forwarder installer using the following command.

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

- Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store port number where the Log Forwarder listens for logs.

Note: The default port number is *15780*.

- Press ENTER.

The added Audit Store endpoint appears on the screen.

The prompt to enter an additional Audit Store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

- If you want to add more than one Audit Store endpoint, then type *y* else type *n*.

Note: If you need to add *n* Audit Store endpoints, then repeat the *Step 2* and *Step 3* *n* times.

- Type the *y* key to install into the destination directory.

The Log Forwarder is installed in the */opt/protegility/fluent-bit/* directory.

- Start the Log Forwarder component by using the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl start
```

The Log Forwarder is successfully installed.

12.5.4.1.1.2 Silent Mode of Installation of Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-29: Parameter List for Silent Installation

| Parameter | Description |
|------------------|--|
| --endpoint or -e | The IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is <i>15780</i> . |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

| Parameter | Description |
|-----------|---|
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *--dir* parameter to the command to specify the Log Forwarder installation directory and *--pemdir* parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>] --dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.5.4.1.2 Installing PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.4.1.2.1 Installing PEP Server on Linux or Unix using Linux Installer

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer.

► To install the PEP server on a Linux or Unix platform:

1. Run the PEP server installer using the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the ESA Host Name or IP Address.
3. Press ENTER.

The prompt to enter the username for downloading certificates appears.

```
Please enter the user name for downloading certificates:
```

4. Enter the username for downloading the certificates.
5. Press ENTER.
6. The prompt to enter the password for downloading the certificates appears.

```
Please enter the password for downloading certificates:
```

6. Press ENTER to install into the destination directory.



Directories are created under `/opt/protegrity/defiance_dps` by default, and the required installation files are installed in these directories.

Caution: Ensure that the ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

To manually install the certificates to the `/opt/protegrity/defiance_dps/data` directory of the PEP server, navigate to the `/opt/protegrity/defiance_dps/bin` directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>]
[-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the *administrator* user.

Verify that the following files have been copied to the `/opt/protegrity/defiance_dps/data` directory:

- `CA.pem`
- `keyinternal.plm`
- `pepserver.cfg`
- `pepserver.pid`
- `authesa.plm`
- `cert.key`
- `cert.pem`
- `certkeyup.bin`

7. Start the PEP server by using the following command.

```
/opt/protegrity/defiance_dps/bin/pepsrvctrl start
```

The PEP server is successfully installed.

12.5.4.1.2.2 Silent Mode of Installation of PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform through the Silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-30: Parameter List for Silent Installation

| Parameter | Description |
|------------------------|--|
| <code>-esa</code> | Specifies the ESA IP address. |
| <code>-esaport</code> | Specifies the ESA port, which is optional. The default value is <code>8443</code> . |
| <code>-certuser</code> | Specifies the ESA user to download certificates. |
| <code>-certpw</code> | Specifies the ESA user password to download certificates. |
| <code>-dir</code> | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <code>/opt/protegrity</code> directory. |



At the command prompt, type the following command from the installer directory.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>
-certpw <password> -dir <installation-directory-path>
```

12.5.4.1.3 Installing Application Protector Go on Linux or Unix

This section describes how to install the AP Go on a Linux or Unix platform.

► To install the AP Go on the Linux or Unix platform:

You can install the AP Go using either the GOPATH approach or the Go Module approach.

Using the Go Module approach:

- a. Run the AP Go installer using the following command.

```
./APGoSetup_Linux_x64_<version>.sh
```

The prompt to continue the installation appears.

```
*****
Welcome to the Defiance DPS AP Go API Setup Wizard
*****
This will install AP Go API on your computer.
Do you want to continue? [yes or no]
```

- b. If you want to continue with the installation of the AP Go SDK, then type *yes* else type *no*.

If you type *yes*, then the prompt to enter the installation directory appears.

```
Please enter installation directory
[/opt/protegility]:
```

If you type *no*, then the installation of the AP Go aborts.

- c. After you enter the installation directory, the following prompt appears.

```
Would you like to install AP Go as a Go Module? [yes or no]
```

- d. Type *yes* to install the AP Go using the Go Module approach.

Note: Ensure that you have installed the Go binaries on your local machine to install the AP Go using the Go Module approach.

- e. Provide the path to the *private* Version Control System repository (VCS). For example, *privaterrepo.example.com/app* where *app* is the name of project. The private VCS path will be used by the installer as the module path for the *go mod init* command.

The code path to be published in the *private* repository is pointed out by the AP Go installer after successful installation. For example, you can publish the code under `/opt/protegility/applicationprotector/go/src/privaterepo.example.com/app` directory.

Caution: For the AP Go module to work accurately, the Protegility Native Library (`gopepprovider.plm`) is included in the `/opt/protegility/applicationprotector/go/src/privaterepo.example.com/app` directory. This library will also be a part of the published code.

- f. To use the AP Go module in your application, you need to add the import path (the repository path where the AP Go module is published) as per your repository path. For example,

```
import "privaterepo.example.com/app/apgo"
```

For more information about how to configure Go to use the private repository module, refer to the section *Appendix B: Using the Go Module with a Private GitLab Repository* in the [Protegility Application Protector Guide 9.1.0.0](#).

Using the GOPATH approach:

- a. Run the AP Go installer using the following command.

```
./APGoSetup_Linux_x64_<version>.sh
```

The prompt to continue the installation appears.

```
*****
Welcome to the Defiance DPS AP Go API Setup Wizard
*****
This will install AP Go API on your computer.
Do you want to continue? [yes or no]
```

- b. If you want to continue with the installation of the AP Go SDK, then type *yes* else type *no*.

If you type *yes*, then the prompt to enter the installation directory appears.

```
Please enter installation directory
[/opt/protegility]:
```

If you type *no*, then the installation of the AP Go aborts.

- c. After you enter the installation directory, the following prompt appears.

```
Would you like to install AP Go as a Go Module? [yes or no]
```

- d. Type *no* to install the AP Go using the GOPATH approach.

- e. Set the following environment variables to run the AP Go application:

```
export GOPATH=$GOPATH:/opt/protegility/applicationprotector/go/
export CGO_CFLAGS="-I/opt/protegility/applicationprotector/go/include"
export CGO_LDFLAGS="-L/opt/protegility/applicationprotector/go/lib"
export LD_LIBRARY_PATH=$LD_LIBRARY_PATH:/opt/protegility/applicationprotector/go/lib
```

The AP Go is installed in the `/opt/protegility` directory using the GOPATH approach.

- f. To use the AP Go package in your application, you need to add the import path. For example,

```
import "protegility.com/apgo"
```

The default installation directory for the AP Go on the Linux or Unix platform is listed in the following table.

Table 12-31: AP Go Installation Directory

| Platform | Directory |
|------------|--|
| Linux/Unix | /opt/protegility/applicationprotector/go |

12.5.4.2 Uninstalling Application Protector Go

As the policies are saved in the ESA and the encrypted or clear data are saved separately, it is not mandatory to take a backup of all the policies, LDAP, and configuration files. However, it is recommended that you take a backup of the files before uninstalling the Application Protector Go.

12.5.4.2.1 Uninstalling the Log Forwarder from AIX, Linux, or Unix

This section describes how to uninstall the Log Forwarder from an AIX, Linux, or Unix platform.

► To uninstall the Log Forwarder from the AIX, Linux, or Unix platform:

1. Navigate to the `/opt/protegility/fluent-bit/bin` directory.
2. Stop the `fluentbit` component using the following command.
`./logforwarderctrl stop`
3. Delete the `fluentbit` directory.

The Log Forwarder is uninstalled.

12.5.4.2.2 Uninstalling the PEP Server from the Linux Platform

This section describes how to uninstall the PEP server from the Linux platform.

► To uninstall the PEP Server from Linux:

1. Navigate to the `/opt/protegility/defiance_dps/bin` directory.
2. Stop the PEP server using the following command:
`/pepsrvctrl stop all`
3. Delete the `/opt/protegility/defiance_dps` directory.

12.5.4.2.3 Uninstalling the Application Protector Go

This section describes how to uninstall the AP Go from the Linux or Unix platform.

► To uninstall the AP Go from the Linux or Unix platform:

1. Navigate to the `/opt/protegility/applicationprotector/go` directory.
2. Delete the `/go` directory.

The AP Go in uninstalled.

12.5.5 Installing Application Protector NodeJS

The Protegility Application Protector (AP) NodeJS provides APIs that integrate with the customer application to protect, unprotect, and reprotect sensitive data. The AP NodeJS can be used with any customer application that is developed for NodeJS.

This section describes a standard setup for installing and uninstalling the AP NodeJS on a Linux platform.

12.5.5.1 Setting up Application Protector NodeJS on Linux or Unix

This section describes how to install the AP NodeJS on a Linux or Unix platform.

Before you begin

Ensure that the following prerequisites are met before installing the AP NodeJS:

- The ESA is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that the Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that the cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.2*.

- NodeJS, version 12 or higher, must be installed on the machine where you are installing the AP NodeJS.
- Ensure one of the following packages is installed on the machine where you are installing the AP NodeJS.
 - *npm*: If you want to install the AP NodeJS using the *npm* approach, then the *npm* package manager must be installed.
 - *Yarn*: If you want to install the AP NodeJS using the *Yarn* approach, then the *Yarn* package manager must be installed.

► To setup the AP NodeJS on the Linux or Unix platform:

1. Download the *ApplicationProtector_Linux-ALL-64_x86-64_NodeJS-12-64_<version>.tgz* package to any location on the machine where you want to install the protector.
2. Extract the AP NodeJS installation package using the following command.

```
tar -xvf ApplicationProtector_Linux-ALL-64_x86-64_NodeJS-12-64_<version>.tgz
```

The following setup files are extracted:

- *apnode-<version>.tgz*
- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*

12.5.5.2 Installing Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.5.2.1 Installing Log Forwarder on Linux or Unix Using Linux Installer

This section describes how to install the Log Forwarder on a Linux or Unix platform using the Linux installer.

► To install the Log Forwarder on a Linux or Unix platform using the Linux installer:

- Run the Log Forwarder installer using the following command.

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

- Enter the Audit Store endpoint that is the Audit Store IP address and the Audit Store port number where the Log Forwarder listens for logs.

Note: The default port number is *15780*.

- Press ENTER.

The added Audit Store endpoint appears on the screen.

The prompt to enter an additional Audit Store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

- If you want to add more than one Audit Store endpoint, then type *y* else type *n*.

Note: If you need to add *n* Audit Store endpoints, then repeat the *Step 2* and *Step 3* *n* times.

- Type the *y* key to install into the destination directory.

The Log Forwarder is installed in the */opt/protegility/fluent-bit/* directory.

- Start the Log Forwarder component by using the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl start
```

The Log Forwarder is successfully installed.

12.5.5.2.2 Silent Mode of Installation of Log Forwarder on Linux or Unix

This section describes how to install the Log Forwarder on a Linux or Unix platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-32: Parameter List for Silent Installation

| Parameter | Description |
|------------------|--|
| --endpoint or -e | The IP address and port number of the Audit Store instance. You can add multiple Audit Store endpoints. Note: The default port number is <i>15780</i> . |
| --dir | Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

| Parameter | Description |
|-----------|---|
| --pemdir | Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <i>/opt/protegility</i> directory. |

At the command prompt, type the following command from the installer directory.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *--dir* parameter to the command to specify the Log Forwarder installation directory and *--pemdir* parameter to the command to specify the PEP server installation directory. The following snippet displays a sample command.

```
./LogforwarderSetup_Linux_x64_<version>.sh --endpoint <ip address:port number> [--endpoint <ip address:port number>] --dir <Log Forwarder installation directory> --pemdir <PEP server installation directory>
```

12.5.5.3 Installing PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer or through the Silent mode of installation.

12.5.5.3.1 Installing PEP Server on Linux or Unix using Linux Installer

This section describes how to install the PEP server on a Linux or Unix platform using the Linux installer.

► To install the PEP server on a Linux or Unix platform:

1. Run the PEP server installer using the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the Audit Store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

2. Enter the ESA Host Name or IP Address.

3. Press ENTER.

The prompt to enter the username for downloading certificates appears.

```
Please enter the user name for downloading certificates:
```

4. Enter the username for downloading the certificates.

5. Press ENTER.

The prompt to enter the password for downloading the certificates appears.

```
Please enter the password for downloading certificates:
```

6. Press ENTER to install into the destination directory.

Directories are created under `/opt/protegrity/defiance_dps` by default, and the required installation files are installed in these directories.

Caution: Ensure that the ESA is up and running with the HubController service in running status to enable automatic downloading of certificates.

To manually install the certificates to the `/opt/protegrity/defiance_dps/data` directory of the PEP server, navigate to the `/opt/protegrity/defiance_dps/bin` directory and run the following command:

```
./GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>]
[-p portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the *administrator* user.

Verify that the following files have been copied to the `/opt/protegrity/defiance_dps/data` directory:

- `CA.pem`
- `keyinternal.plm`
- `pepperver.cfg`
- `pepperver.pid`
- `authesa.plm`
- `cert.key`
- `cert.pem`
- `certkeyup.bin`

7. Start the PEP server by using the following command.

```
/opt/protegrity/defiance_dps/bin/pepsrvctrl start
```

The PEP server is successfully installed.

12.5.5.3.2 Silent Mode of Installation of PEP Server on Linux or Unix

This section describes how to install the PEP server on a Linux or Unix platform through the Silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-33: Parameter List for Silent Installation

| Parameter | Description |
|------------------------|--|
| <code>-esa</code> | Specifies the ESA IP address. |
| <code>-esaport</code> | Specifies the ESA port, which is optional. The default value is <code>8443</code> . |
| <code>-certuser</code> | Specifies the ESA user to download certificates. |
| <code>-certpw</code> | Specifies the ESA user password to download certificates. |
| <code>-dir</code> | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is the <code>/opt/protegrity</code> directory. |



At the command prompt, type the following command from the installer directory.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>  
-certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
./PepServerSetup_Linux_x64_<version>.sh -esa <esaIP> -esaport <esaPort> -certuser <username>  
-certpw <password> -dir <installation-directory-path>
```

12.5.4 Installing Application Protector NodeJS on Linux or Unix

This section describes how to install the AP NodeJS on a Linux or Unix platform.

► To install the AP NodeJS on the Linux or Unix platform:

You can install the AP NodeJS using either the *npm* approach or the *Yarn* approach.

Using the *npm* approach:

- a. Run the following command to initialize the project and create a *package.json* file.

```
npm init
```

After answering the questions to create the initial structure of the project, the *package.json* file is created.

- b. Run the following command to install the AP NodeJS.

```
npm install apnode-<version>.tgz
```

Using the *Yarn* approach:

- a. Run the following command to initialize the project and create a *package.json* file.

```
yarn init
```

After answering the questions to create the initial structure of the project, the *package.json* file is created.

- b. Run the following command to install the AP NodeJS.

```
yarn add file:apnode-<version>.tgz
```

The AP NodeJS is installed successfully.

12.5.5 Uninstalling Application Protector NodeJS

This section describes how to uninstall the AP NodeJS from a Linux or Unix platform.

12.5.5.1 Uninstalling the Log Forwarder from AIX, Linux, or Unix

This section describes how to uninstall the Log Forwarder from an AIX, Linux, or Unix platform.

► To uninstall the Log Forwarder from the AIX, Linux, or Unix platform:

1. Navigate to the `/opt/protegility/fluent-bit/bin` directory.
2. Stop the `fluentbit` component using the following command.
`./logforwarderctrl stop`
3. Delete the `fluentbit` directory.

The Log Forwarder is uninstalled.

12.5.5.2 Uninstalling the PEP Server from AIX, Linux or Unix

This section describes how to uninstall the PEP server from an AIX, Linux or Unix platform.

- To uninstall the PEP server from the AIX, Linux or Unix platform:

1. Navigate to the `/opt/protegility/defiance_dps/bin` directory.
2. Stop the PEP server using the following command.
`./pepsrvctrl stop`
3. Delete the `/opt/protegility/defiance_dps` directory.

The PEP server is uninstalled.

12.5.5.3 Uninstalling AP NodeJS from Linux or Unix

This section describes how to uninstall the AP NodeJS from a Linux or Unix platform.

- To uninstall the AP NodeJS from the Linux or Unix platform:

1. If you installed the AP NodeJS using the `npm` approach, then use the following command to uninstall the AP Node from the Linux or Unix platform.

```
npm uninstall apnode
```

2. If you installed the AP NodeJS using the `Yarn` approach, then use the following command to uninstall the AP NodeJS from the Linux or Unix platform.

```
yarn remove apnode
```

12.5.6 Installing Application Protector (AP) .Net

The Protegility Application Protector (AP) .NET provides APIs that integrate with the customer application to protect and unprotect sensitive data. The AP .Net can be used with any customer application that is developed using the .Net Core and C# programming language.

This section describes a standard setup of installing and uninstalling the Application Protector .NET on the Windows platform.

12.5.6.1 Setting up Application Protector .Net on Windows

This section describes how to install the AP .Net on a Windows platform.

Before you begin

Ensure that the following prerequisites are met before installing Application Protector:

- The ESA appliance is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- Ensure that Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.4](#).

- Any supported version from .Net Standard 2.0 must be installed on the machine where you are installing the AP .Net.

► To setup the AP .Net on the Windows platform:

1. Download the *ApplicationProtector_WIN-ALL-64_x86-64_NET-STD-2.0-64_<version>.zip* installation package to any location on the machine where you want to install the protector.
2. Extract the files from the *ApplicationProtector_WIN-ALL-64_x86-64_NET-STD-2.0-64_<version>.zip* installation package.
The following setup files are extracted:
 - *LogforwarderSetup_Windows_x64_<version>.exe*
 - *PepServerSetup_Windows_x64_<version>.exe*
 - *APDotNetSetup_Windows_x64_<version>.exe*

12.5.6.1.1 Installing Log Forwarder on Windows

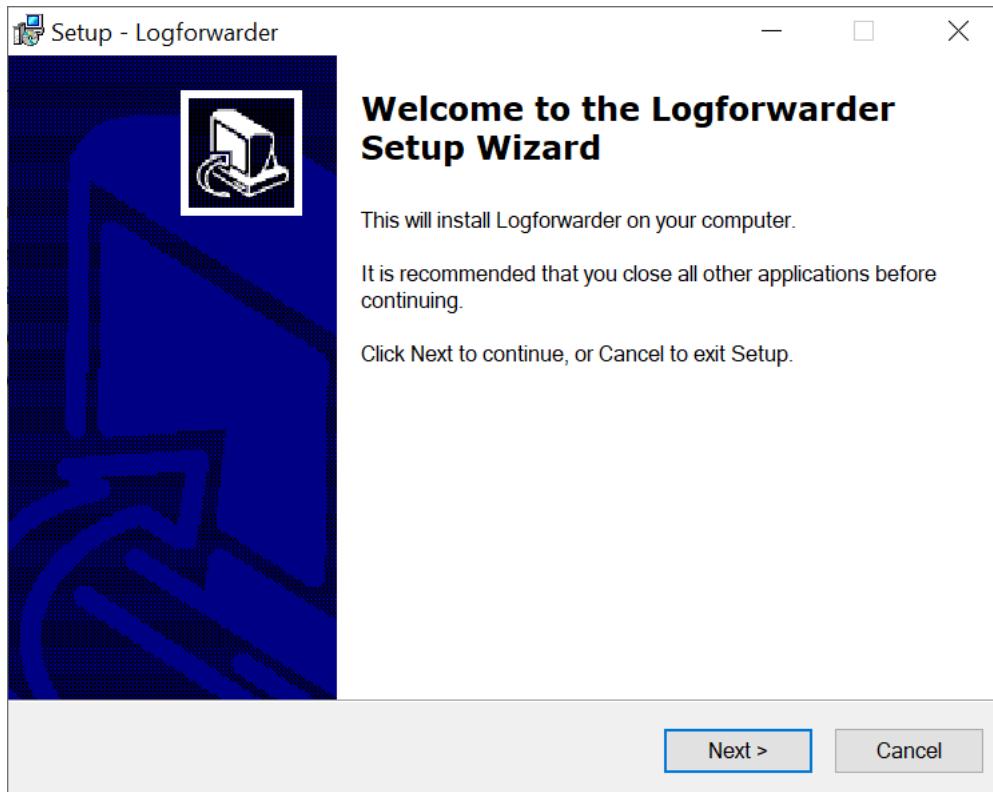
This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard and through the Silent mode of installation.

12.5.6.1.1.1 Using Windows Wizard

This section describes how to install the Log Forwarder on a Windows platform through the Windows Wizard.

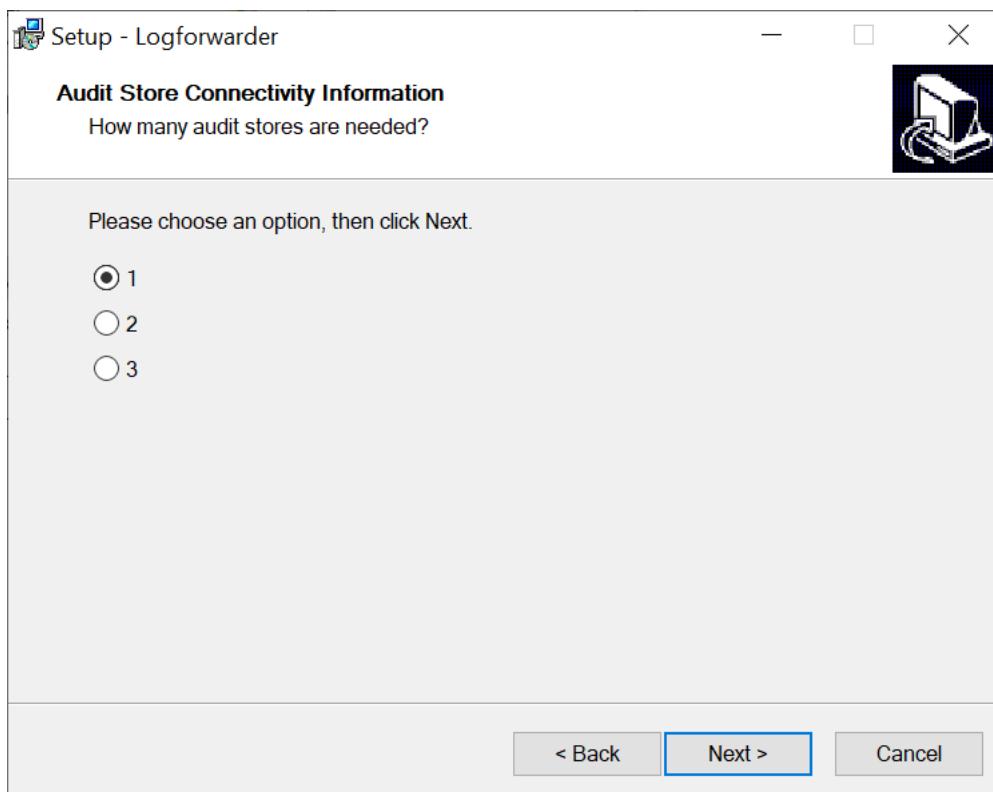
► To install the Log Forwarder on the Windows platform through the Windows Wizard:

1. Run the *LogforwarderSetup_Windows_x64_<version>.exe* file from the created directory.
The **Welcome to the Log Forwarder Setup Wizard** screen appears.



2. Click **Next**.

The **Audit Store Connectivity Information** screen appears to select the number of audit stores that are needed.

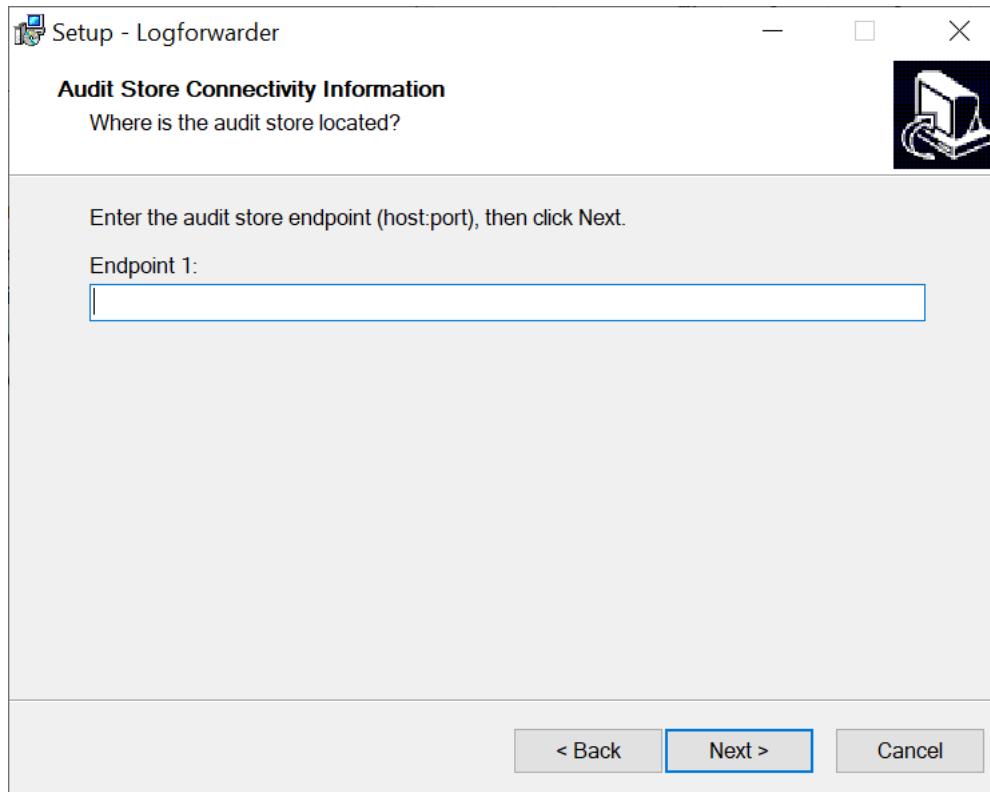


3. Select the number of Audit Stores as required.

The maximum number of Audit Stores that can be configured is 3.

4. Click **Next**.

The **Audit Connectivity Information** screen appears to enter the location of the audit store.



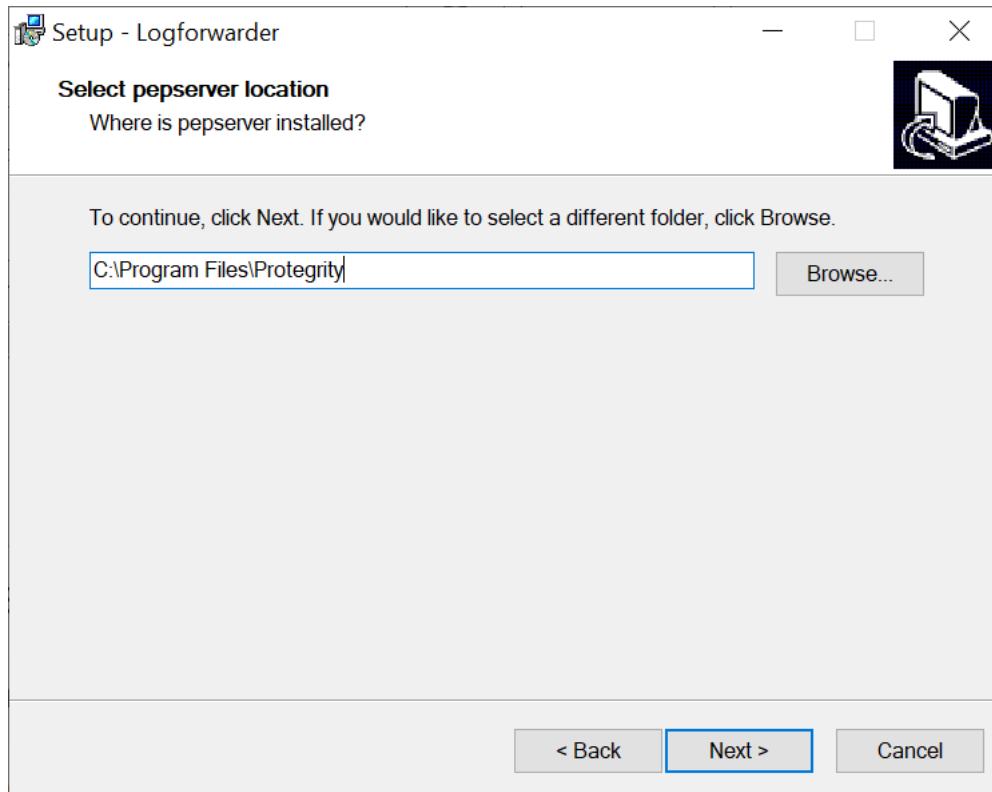
5. Enter the Hostname/IPAddress:port where the Audit Store is configured.

Note:

The default port number for the Audit Store is *9200*.

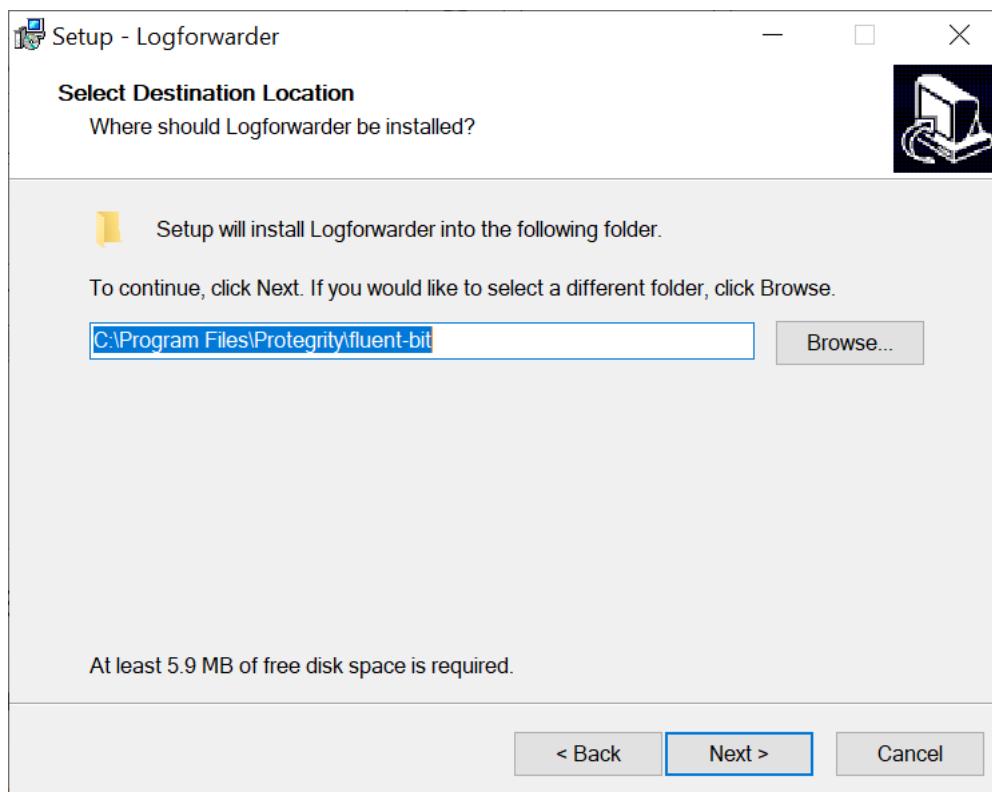
6. Click **Next**.

The **Select pepserver location** screen appears.



7. Click **Next**.

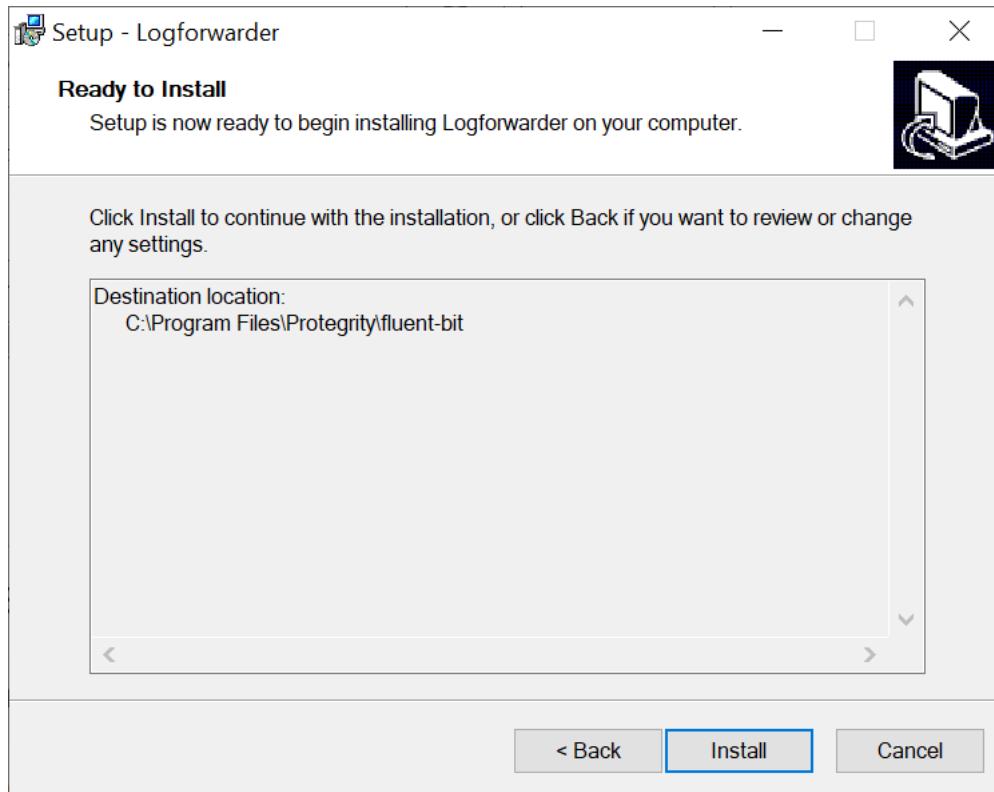
The **Select Destination Location** screen appears.



8. Set the installation directory for the Log Forwarder to *C:\Program Files\Protegility\fluent-bit*.

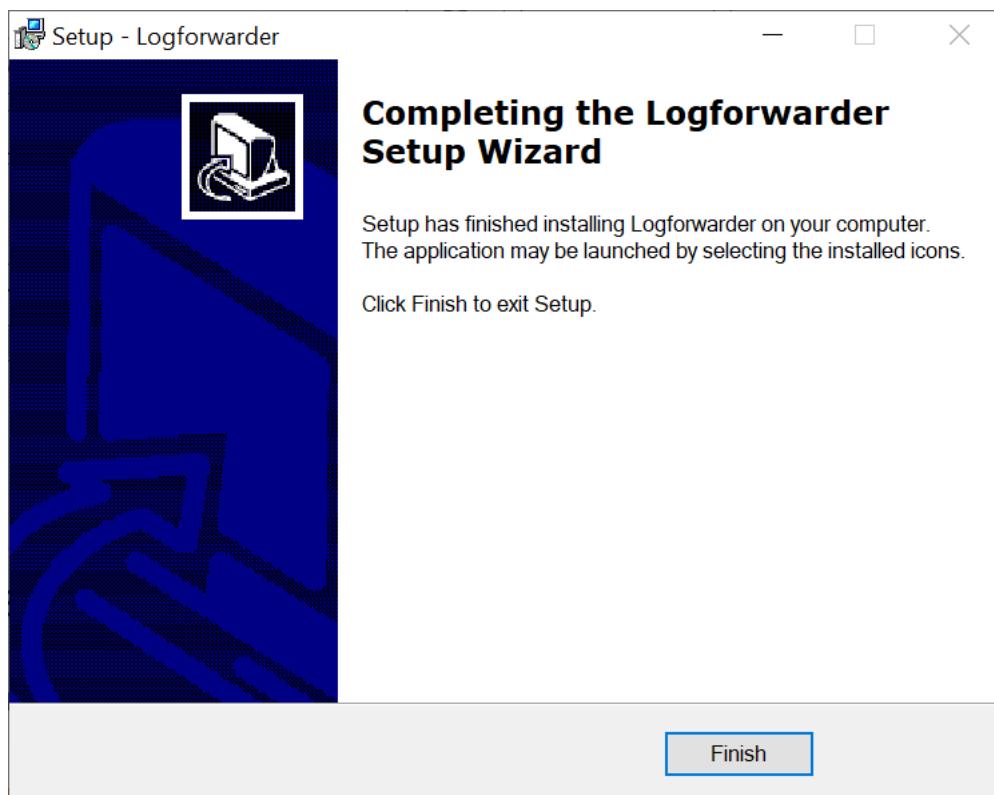
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Logforwarder Setup Wizard** screen appears.



11. From the **Completing the Logforwarder Setup Wizard** screen, click **Finish** to complete the installation and exit.

Post successful installation, you can find the installation files in the directories that are created under the defined installation directory.

The Log Forwarder is installed successfully.

12.5.6.1.1.2 Using Silent Mode of Installation

This section describes how to install the Log Forwarder on a Windows platform through the Silent mode of installation.

You can also execute the Log Forwarder installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in the Silent mode.

Table 12-34: Parameter List for Silent Installation

| Parameter | Description |
|------------------------------------|--|
| -endpoint1, -endpoint2, -endpoint3 | <p>Audit Store IP address and the port number where the Log Forwarder listens for logs.</p> <p>Note: The default port number is <i>9200</i>.</p> <p>Note: The parameters <i>-endpoint2</i> and <i>-endpoint3</i> are optional.</p> |
| -dir | <p>Installation directory of the Log Forwarder, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility\fluent-bit</i>.</p> |
| -pemdir | <p>Installation directory of the PEP server, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i>C:\Program Files\Protegility</i>.</p> |

At the command prompt, type the following command from the installer directory.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>]
```

If you want to install the Log Forwarder and the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the Log Forwarder installation directory and the *-pemdir* parameter to the command to specify the PEP server installation directory.

The following snippet displays a sample command.

```
.\LogforwarderSetup_Windows_x64.exe -endpoint1 <ip address:port number> [-endpoint2 <ip address:port number>]
[-endpoint3 <ip address and port number>] -dir <Log Forwarder installation directory> -pemdir
<PEP server installation directory>
```

12.5.6.1.2 Installing PEP Server on Windows

This section describes how to install the PEP server on a Windows platform through the Windows Wizard and through the Silent mode of installation.

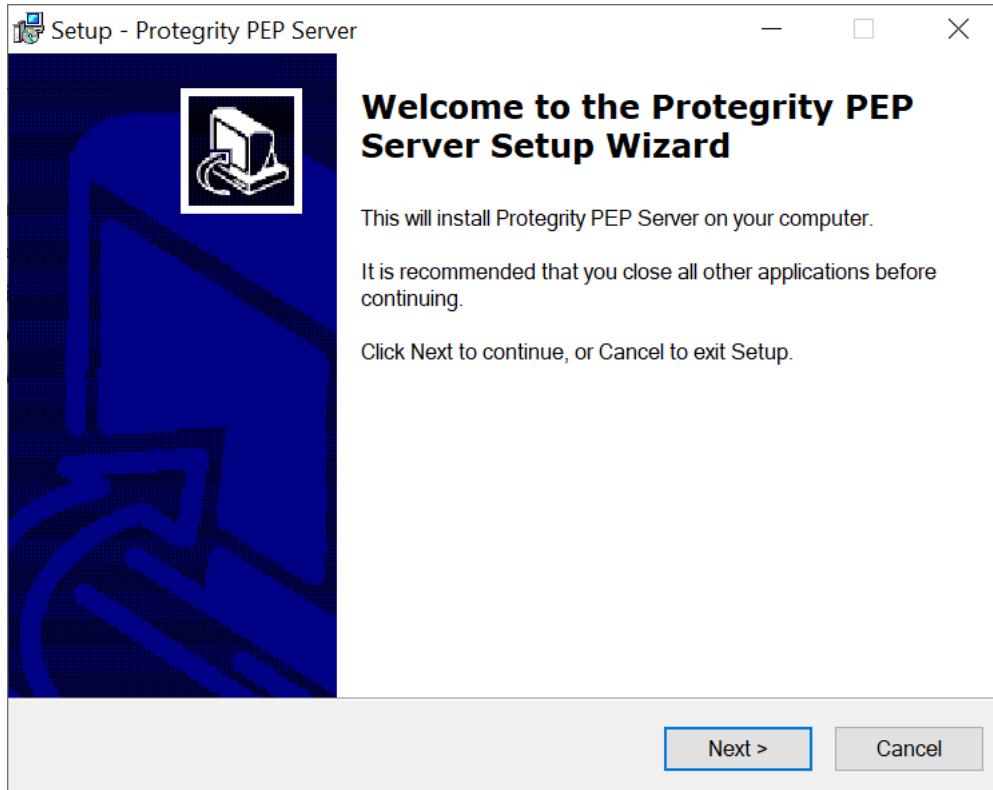
12.5.6.1.2.1 Using Windows Wizard

This section describes how to install the PEP server on a Windows platform through the Windows Wizard.

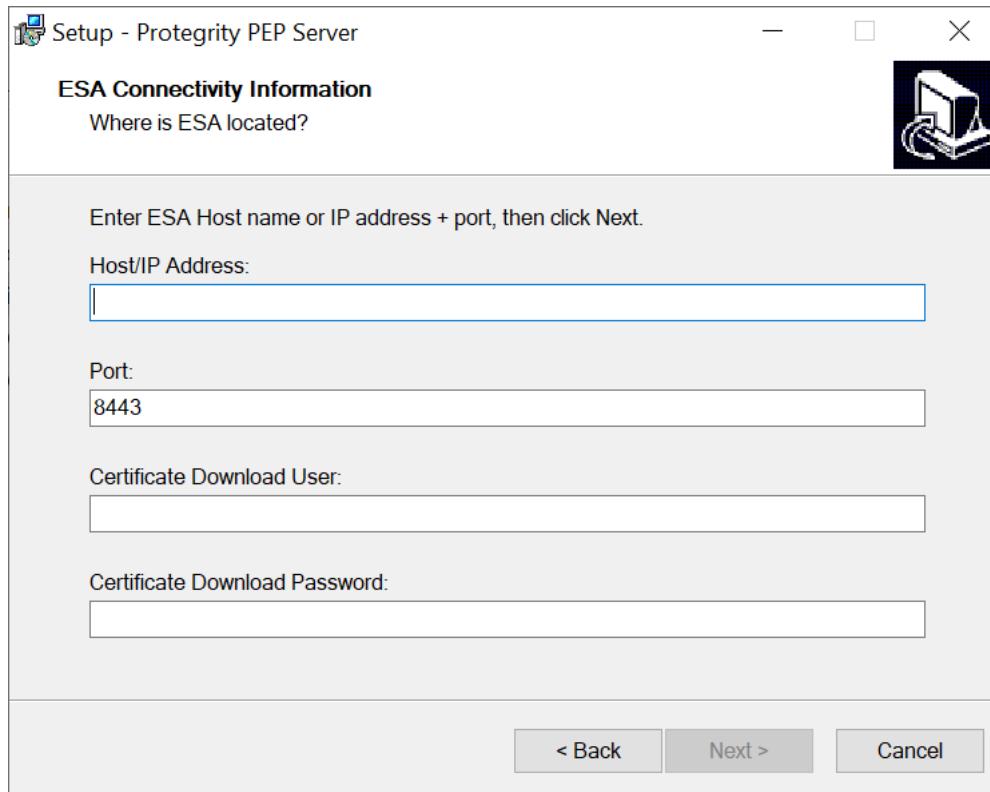


► To install the PEP server on the Windows platform through the Windows Wizard:

1. Run the *PepServerSetup_Windows_<bit-version>_<version>.exe* file from its installed directory.
The **Welcome to the Protegility PEP Server Setup Wizard** appears.



2. Click **Next**.
The **ESA Connectivity Information** screen appears.



3. Enter the ESA Host name or IP Address in the **Host/IP Address** field.
4. Enter the ESA host listening port details in the **Port** field.

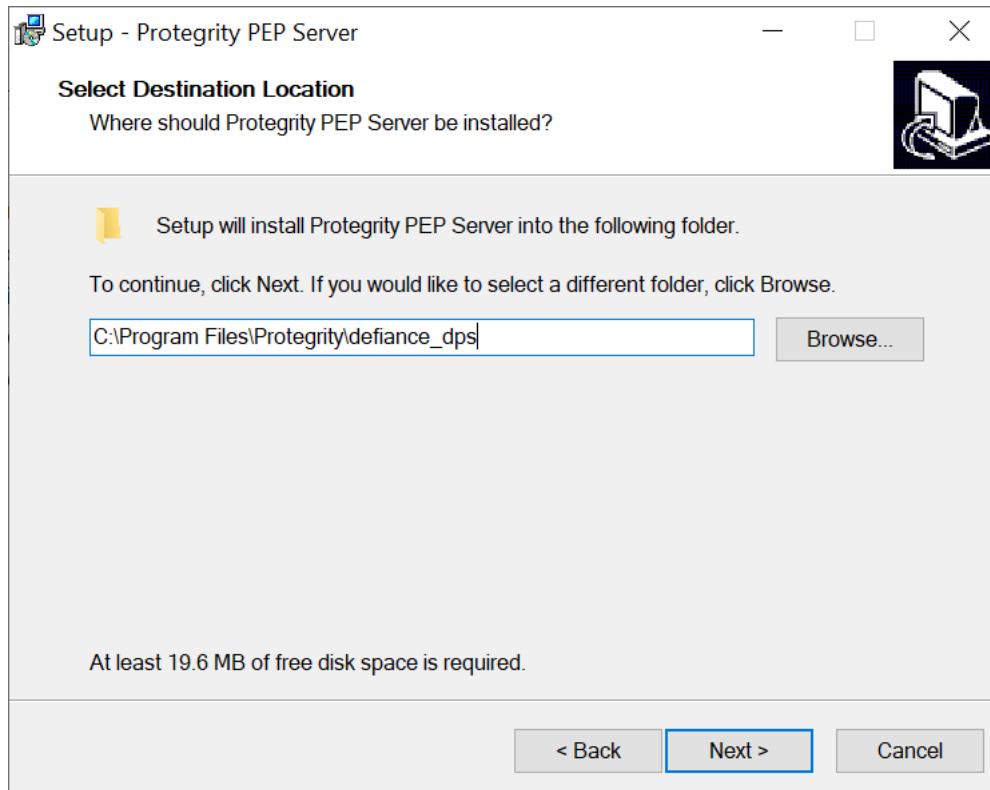
Note: Ensure that the ESA is up and running with the HubController service in running status to enable the automatic downloading of certificates.

5. Enter the **Certificate Download User**.

Note: It is recommended to use *admin* as the user.

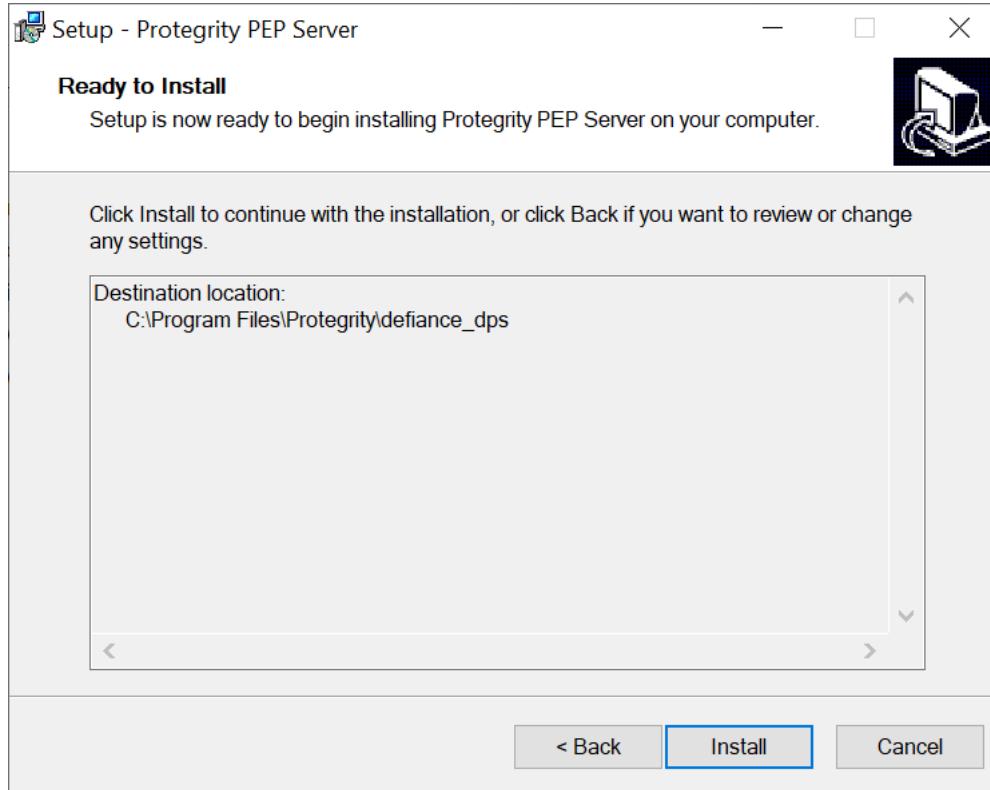
6. Enter the **Certificate Download Password**.
7. Click **Next**.

The **Select Destination Location** screen appears.



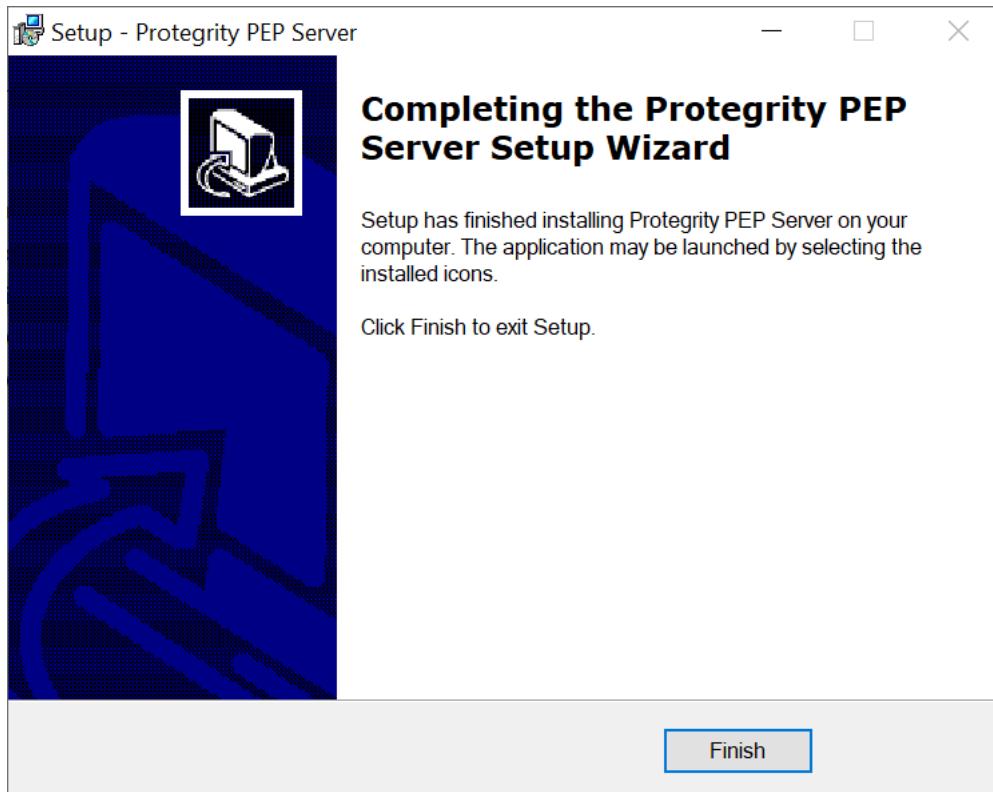
8. Set the installation directory for the PEP server to *C:\Program Files\Protegility\defiance_dps*.
9. Click **Next**.

The **Ready to Install** screen appears.



10. Click **Install**.

The **Completing the Protegility PEP Server Setup Wizard** screen appears.



11. From the **Completing the Protegility PEP Server Setup Wizard** screen, click **Finish** to complete the installation and exit.

The directories are created under the installation directory that was defined and the installation files are installed in these directories.

Note:

If you want to manually install the certificates in the *<path to where pepserver is installed>\defiance_dps\data* directory of the PEP server, then navigate to the *<path to where pepserver is installed>\defiance_dps\bin* directory, and run the following command:

```
GetCertificates -u admin <Admin Username> [-h <ESA host name or IP address>] [-portno] [-d directory]
```

This initiates secure communication between the PEP server and the ESA.

Enter the password for the user *admin*.

The PEP server is installed successfully.

12.5.6.1.2.2 Using Silent Mode of Installation

This section describes how to install the PEP server on a Windows platform through the silent mode of installation.

You can also execute the PEP server installer without any manual intervention, which is also known as the Silent mode of installation. The following parameters must be provided to execute the installer in Silent mode.

Table 12-35: Parameter List for Silent Installation

| Parameter | Description |
|-----------|-------------------------------|
| -esa | Specifies the ESA IP address. |

| Parameter | Description |
|-----------|---|
| -esaport | Specifies the ESA port, which is optional. The default value is <i>8443</i> . |
| -certuser | Specifies the ESA user to download certificates. |
| -certpw | Specifies the ESA user password to download certificates. |
| -dir | Specifies the installation directory, which is optional. If the installation directory is not specified, then the installation path is the default directory, which is <i><path to where pepserver is installed>\defiance_dps</i> . |

At the command prompt, type the following command from the installer directory.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
```

If you want to install the PEP server in a directory other than the default directory, then you can add the *-dir* parameter to the command to specify the directory. The following command displays a sample snippet.

```
PepServerSetup_Windows_<bit-version>.exe -certuser <username> -esa <esaIP> -certpw <password>
-dir <installation-directory-path>
```

12.5.6.1.3 Installing Application Protector .Net on Windows

This section describes how to install the AP .Net on a Windows platform.

Important:

Ensure that the *C:\Users\Administrator\.nuget\packages* directory does not contain the *dotnetprotector* folder before installing the AP .Net.

► To install the AP .Net on the Windows platform:

1. Run the *APDotNetSetup_Windows_x64_<version>.exe* installer from the created directory.

The **Select Destination Location** screen appears.

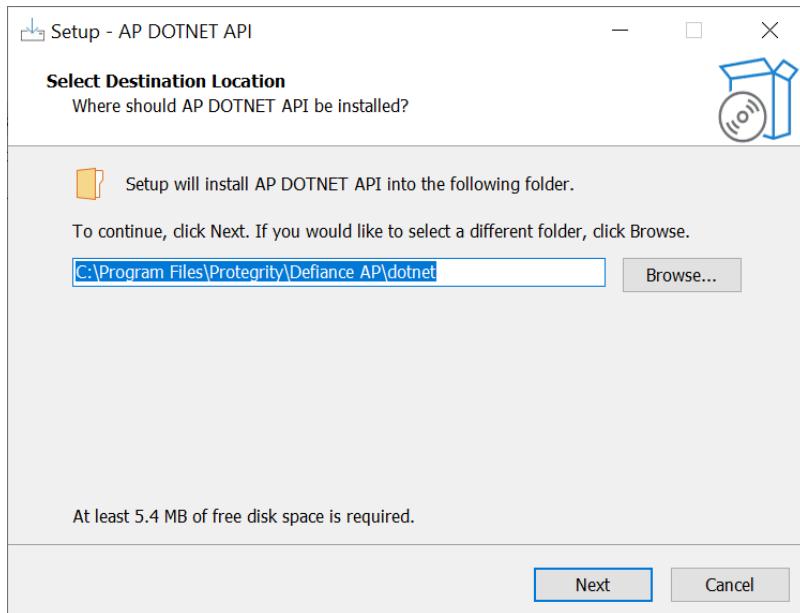


Figure 12-15: Select Destination Location

2. Set the installation directory to `C:\Program Files\Protegility\Defiance AP\dotnet`.
3. Click **Next**.

The **Ready to Install** screen appears.

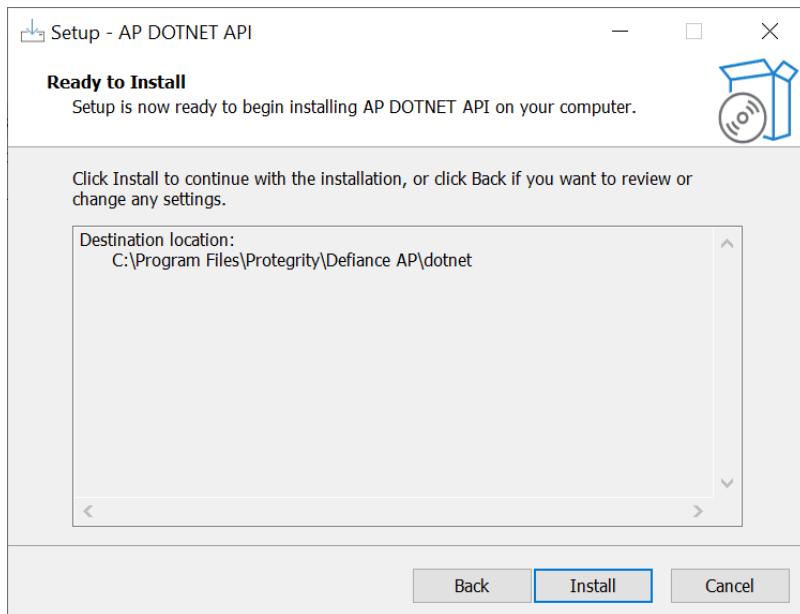


Figure 12-16: Ready to Install

4. Click **Install**.

The **Completing the Defiance AP DOTNET API Setup Wizard** screen appears.

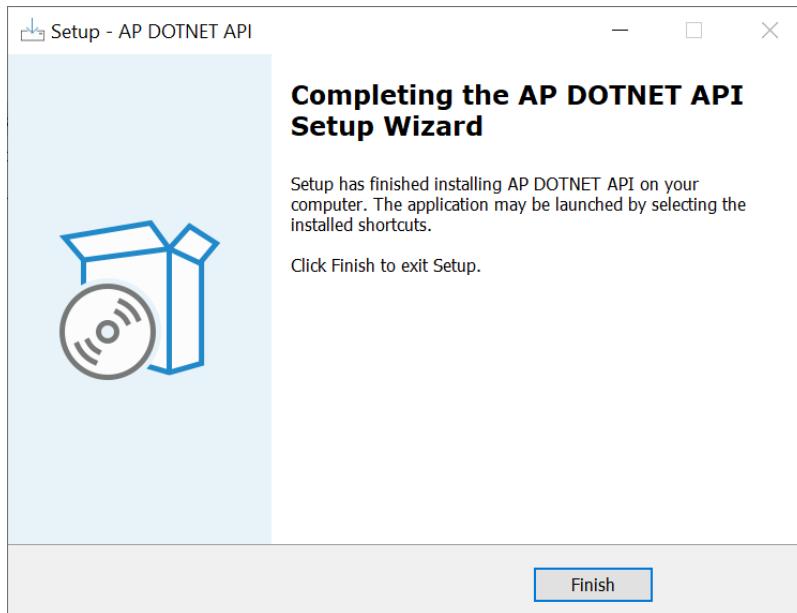


Figure 12-17: Completing the Defiance AP DOTNET API Setup Wizard

5. From the **Completing the Defiance AP DOTNET API Setup Wizard** screen, click **Finish** to complete the installation and exit.
- The AP .Net is installed successfully.

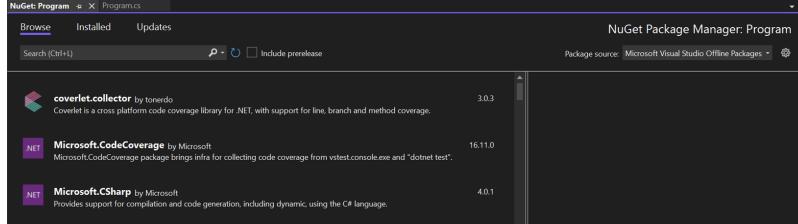
The default installation directories for different platforms are given in the following table.

Table 12-36: AP .Net Installation Directory

| Platform | Directory |
|----------------|---|
| Windows 64-bit | <p><i>C:\Program Files\Protegility\Defiance AP\dotnet\lib</i></p> <p>For API documentation:</p> <p><i>C:\Program Files\Protegility\Defiance AP\dotnet\doc</i></p> |

6. Verify that the following files are installed in the AP .NET installation directory:
 - *DotNetProtector.<version>.nupkg* - NuGet package for the AP .NET
 - *dotnetprovider.plm* - Dynamically loadable module of the AP .NET for Windows
7. Install the *DotNetProtector.<version>.nupkg* NuGet package.
 - a. In **Solution Explorer**, right-click on **Dependencies**, and select **Manage NuGet Packages**.

The **NuGet Package Manager** screen appears.



- b. Click the **Settings** icon next to the **Package source** list.

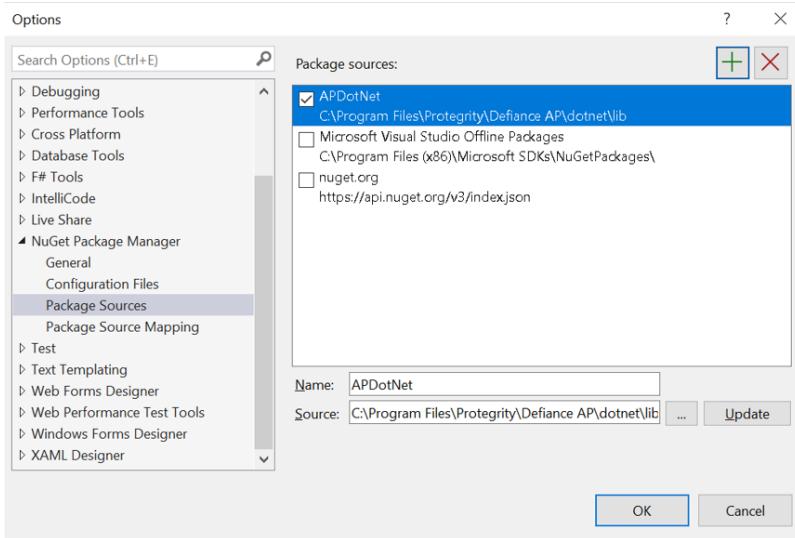
The **Options** dialog box appears.

- Click on the ellipsis icon next to the **Source** field to browse for and select the *lib* directory that contains the *DotNetProtector.<version>.nupkg* package.

For example, select the *C:\Program Files\Protegility\Defiance AP\dotnet\lib* directory.

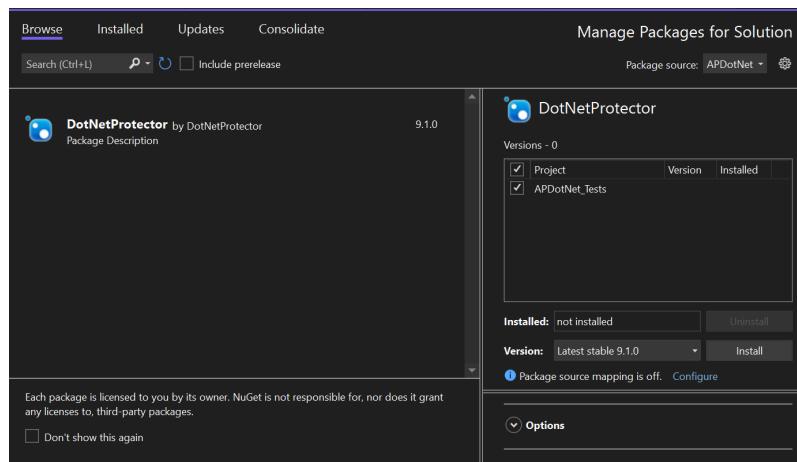
- Specify a name for the package source in the **Name** field, and then click **Update**.

The *lib* directory path appears in the **Package sources** list.



- Click **OK** to close the **Options** dialog box.
- In the **Package source** list on the **NuGet Package Manager** screen, select the package source that you have created in *step d*, and then click **Browse**.

The **DotNetProtector** package appears in the list of packages.

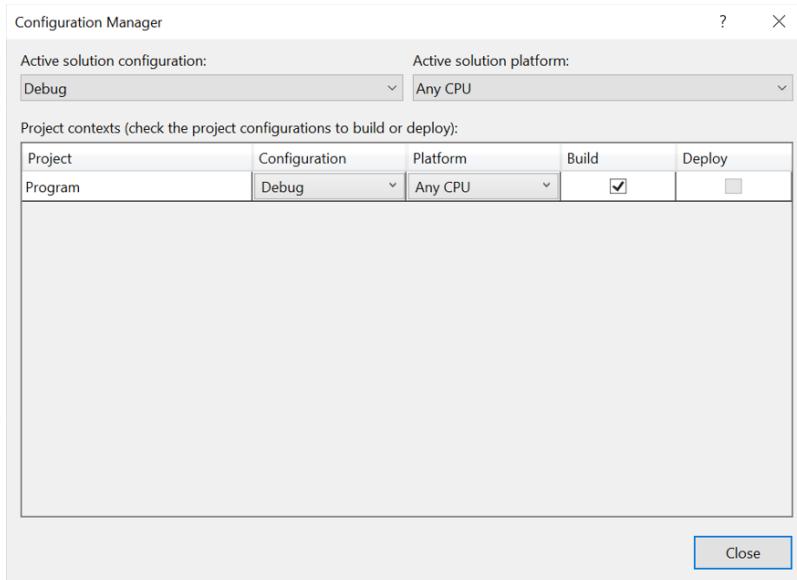


- Click **Install**.

The **DotNetProtector** package is installed. The **Error List** section displays a warning message because the CPU platform specified in the **Configuration Manager** does not match the CPU platform that was used to build the sample application. By default, the platform value is set to **Any CPU**.

- Perform the following steps to remove the warning message.
- Navigate to **Build > Configuration Manager**.

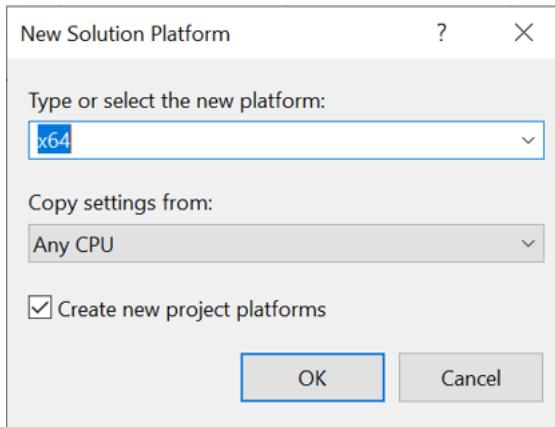
The **Configuration Manager** dialog box appears.



ii. In the **Active solution configuration** list, select **Release**.

iii. Click the **Platform** list, and then click **New**.

The **New Solution Platform** dialog box appears.



iv. In the **New** platform list, select **x64**.

You need to select **x64** because the Sample Application has been built using the x64 platform.

v. Select the **Create new solution platforms** checkbox.

vi. Click **OK**.

The Warning message disappears.

i. Run the program.

For more information about installing a NuGet package in Visual Studio, refer to the [Microsoft documentation](#).

8. Add the path where the *dotnetprovider.plm* file is located in the *Path* system variable.

For example, add the file path *C:\Program Files\Protegility\Defiance AP\dotnet\lib* in the *Path* system variable.

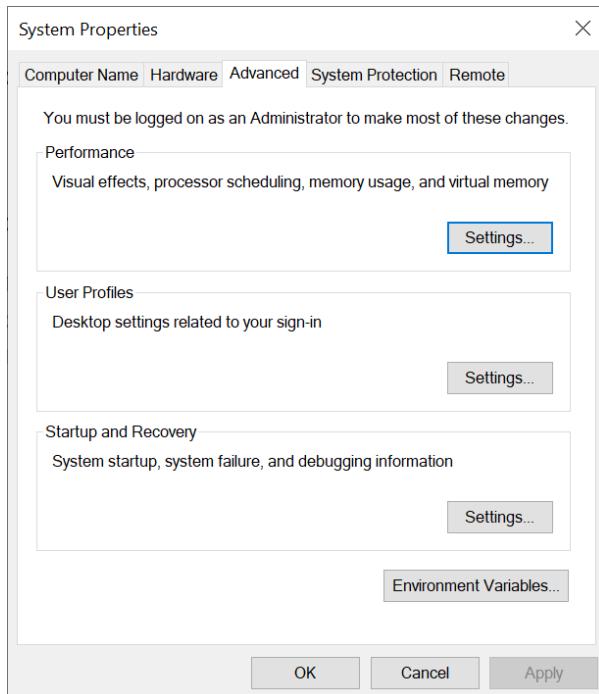
The AP .Net is installed successfully.

12.5.6.1.4 Setting Up the Environment Variables

This section describes the steps to setup the environment variables for running the AP .NET application on a Windows platform.

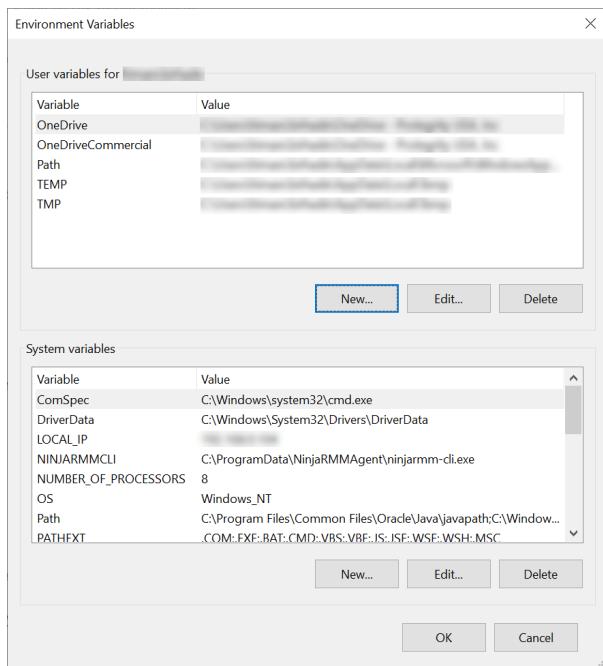
► To setup the environment variables on the Windows platform:

1. On the machine where you have installed the AP .NET, navigate to **Control Panel > System > Advanced system settings**.
The **System Properties** dialog box appears.



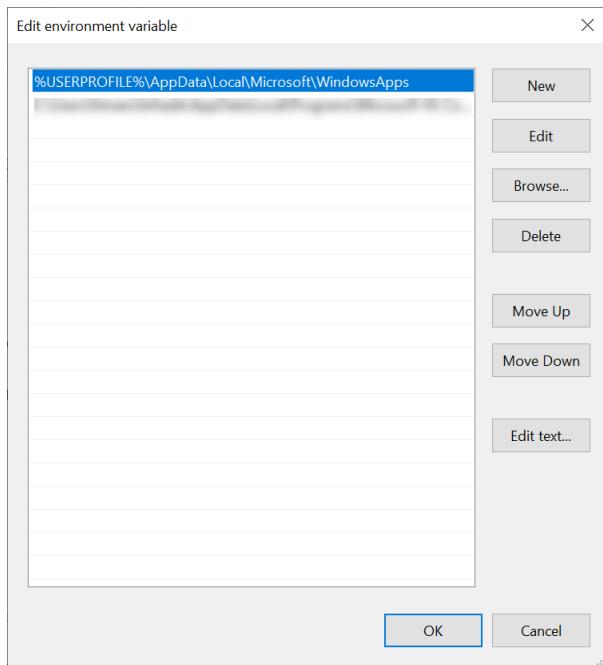
2. Click **Environment Variables**.

The **Environment Variables** dialog box appears.



3. Select the **Path** variable, and then click **Edit**.

The **Edit environment variable** dialog box appears.



4. Click **New**.
5. Add the path where the *dotnetprovider.plm* file is located.
For example, specify the path as *C:\Program Files\Protegility\Defiance AP\dotnet\lib*.
6. Click **OK** to save the changes.

12.5.6.2 Uninstalling Application Protector .Net

This section describes how to uninstall the AP .Net from a Windows platform.

12.5.6.2.1 Uninstalling the Log Forwarder from the Windows Platform

This section describes how to uninstall the Log Forwarder from the Windows platform.

► To uninstall the Log Forwarder on Windows:

1. Perform the following steps to stop the Log Forwarder.
 - a. From the Windows Start Menu, search and select **Services**.
 - b. Navigate to the *Logforwarder* directory.
 - c. Right-click the **Logforwarder** service and click **Stop**.
2. Run the **logforwarder** uninstall utility located in the *C:\Program Files\Protegility\fluent-bit* directory.
3. After the Log Forwarder is uninstalled, delete the following directory.

fluent-bit

The Log Forwarder is uninstalled.

12.5.6.2.2 Uninstalling PEP Server from Windows

This section describes how to uninstall the PEP server from the Windows platform.

► To uninstall the PEP Server from the Windows platform:

1. Stop the PEP server by performing the following steps.
 - a. From the Windows Start Menu, search and select *Services*.
 - b. Navigate to the *Protegility PEP Server* service.
 - c. Right-click the service and click **Stop**.
2. Navigate to the *C:\Program Files\Protegility\defiance_dps* directory.
3. Run the following file to uninstall the PEP server.
unins000.exe
4. Delete the following directory.
C:\Program Files\Protegility\defiance_dps

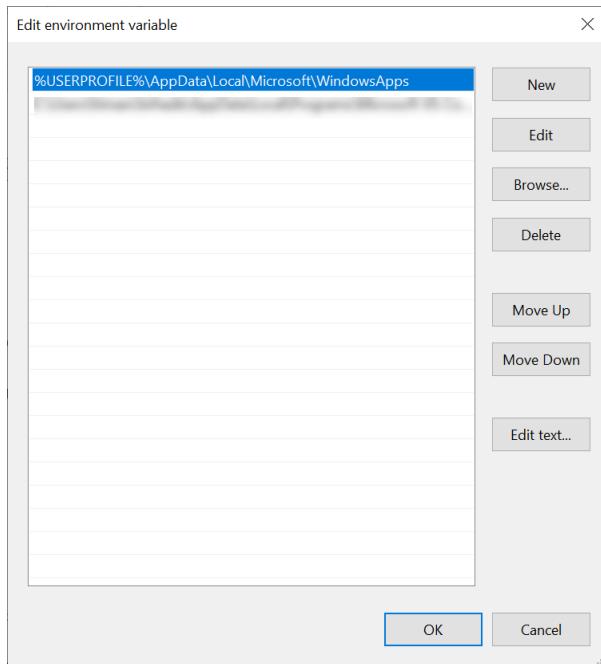
The PEP server is uninstalled.

12.5.6.2.3 Uninstalling the Application Protector .Net from the Windows Platform

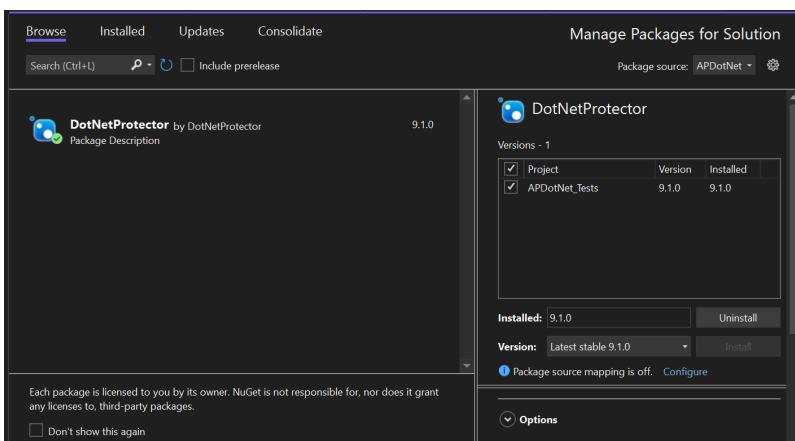
This section describes how to uninstall the AP .Net from a Windows platform.

► To uninstall the AP .Net from the Windows platform:

1. Navigate to the *C:\Program Files\Protegility\Defiance_AP\dotnet* installation directory.
2. Run the uninstall utility located in the *C:\Program Files\Protegility\Defiance_AP\dotnet* directory.
3. Delete the *C:\Program Files\Protegility\Defiance_AP* directory.
4. Delete the path of the *dotnetprovider.plm* file that was added in the environment variables.



5. Uninstall the AP .Net NuGet package using the following steps:
 - a. Navigate to **Manage NuGet Packages for Solution**.
 - b. Browse and select **APDotNet**.



- c. Click **Uninstall**.
- d. Navigate to `C:\Users\Administrator\.nuget\packages` directory.
- e. Delete the `dotnetprotector` folder.

The AP .Net is uninstalled.

12.6 Installing and Uninstalling the Big Data Protector

This section describes the procedure to install and uninstall the Big Data Protector.

Note:

Starting from the Big Data Protector 7.2.0 release, the HDFS File Protector (HDFSFP) is deprecated.

The version numbers in the screenshots, in the following sub-sections, are used as an example to explain the process.

12.6.1 Installing the Big Data Protector Using the CDP Private Cloud Base (CDP-PVC-Base) Installer

This section describes the procedures to install and uninstall the Protegity Big Data Protector using Cloudera Manager.

Caution: The CDP Data Center (CDP-DC) is renamed as CDP Private Cloud Base (CDP-PVC-Base). The CDP Data Center (CDP-DC) term is henceforth referred to as CDP Private Cloud Base (CDP-PVC-Base) in Protegility documentation.

Note: The build number in the installation files and the Cloudera Manager Web interface will reflect the version number of the build that you download from the [My.Cloudera](#) portal.

12.6.1.1 Understanding the Installation Methods

Depending on the requirements of the cluster configuration, you can install the Big Data Protector in the following ways:

- Installing the Big Data Protector without a proxy: Use this method if the cluster is configured in an open network system where all the nodes have connectivity with the ESA. The following diagram represents the Big Data Protector installation without a proxy.

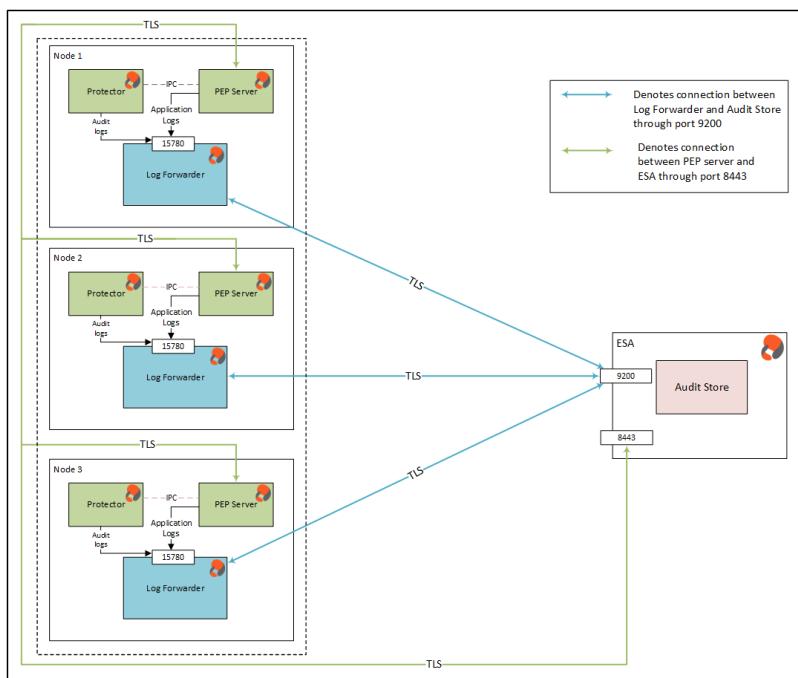


Figure 12-18: Big Data Protector Installation without a Proxy

- Installing the Big Data Protector with a proxy: Use this method if the cluster is configured in a network system where the nodes do not have connectivity with the ESA. Typically, you must configure the Proxy service on the *Lead (Edge)* node because it has connectivity to the ESA. The following diagram represents the Big Data Protector installation with a proxy.

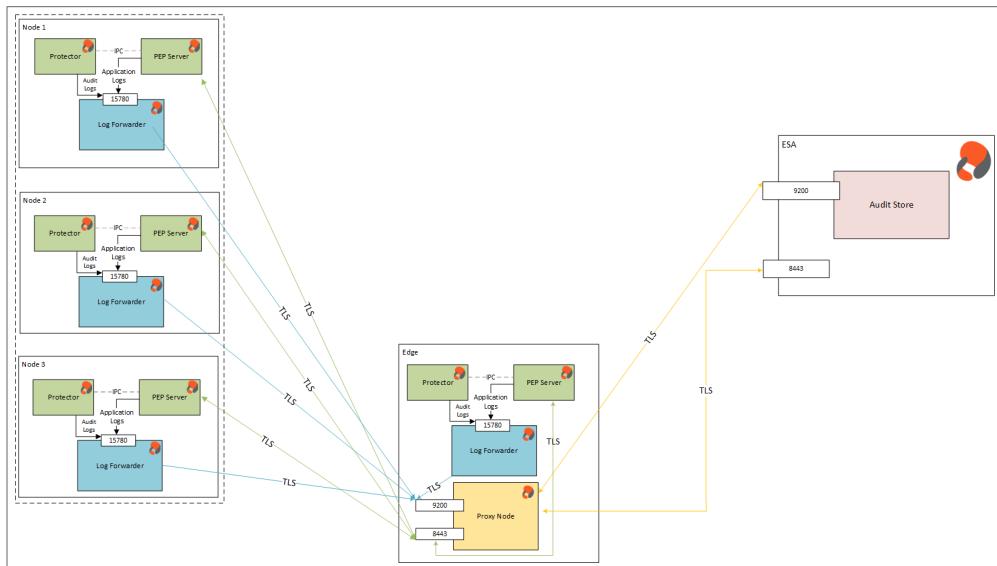


Figure 12-19: Big Data Protector Installation with a Proxy

Note:

For the Big Data Protector 9.1.0.0 release, the CDP Private Cloud Base (CDP-PVC-Base), which includes Cloudera Runtime and Cloudera Manager, version 7.1, is used for reference.

For more information about CDP-PVC-Base, version 7.1.x, refer to <https://docs.cloudera.com/cdp-private-cloud-base/latest/index.html>.

12.6.1.2 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before installing the Big Data Protector from the Cloudera Manager:

- The Hadoop cluster is installed, configured, and running CDP-PVC-Base and Cloudera Manager, version 7.1, or later.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running.
- The ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector, are listed in the following table:

Table 12-37: List of Ports for the Big Data Protector

| Destination Port | Protocols | Sources | Destinations | Descriptions |
|------------------|-----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Protector on the Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to <i>localhost</i> through port 15780. The PEP server Application Logs are also written to <i>localhost</i> through port 15780. The Log Forwarder reads the logs from that socket. |



| Destination Port | Protocols | Sources | Destinations | Descriptions |
|--|-----------|--|---|--|
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses <i>localhost</i> port 16700. |
| Port Requirement for Proxy Node (Optional) | | | | |
| 8443 | TCP | PEP server on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The PEP server communicates with the Proxy node through port 8443, which substitutes the stream to the ESA. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The Log Forwarder communicates with the Proxy node through port 9200, which substitutes the stream to the Protegility Audit Store Appliance. |

- The user, installing the Big Data Protector, has the requisite permissions to perform the following tasks:
 - Copy the Big Data Protector parcels and CSDs to the Cloudera Manager repository directories
 - Restart the Cloudera SCM Server
- If you are installing the Big Data Protector on a cluster, then ensure that it is installed on all the nodes in the cluster.
- The group *ptyitusr* and the user *ptyitusr*, which are responsible to manage the Big Data Protector-related services, are not present on the cluster nodes, which are managed by Cloudera Manager.

12.6.1.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script that will generate the Big Data Protector parcels and CSDs required to install the Big Data Protector on all the nodes in the cluster, which are managed by Cloudera Manager.

► To extract the files from the installation package:

1. Login to the CLI on the Master node that has connectivity to the ESA.
2. Copy the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_CDP-PVC-Base-7-64_9.1.0.0.x.tgz* to any directory.
For example, */opt/bigdata*.
3. To extract the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_CDP-PVC-Base-7-64_9.1.0.0.x.tgz
```

4. Press ENTER.
The command extracts the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* file from the Big Data Protector installation package.

12.6.1.4 Running the Big Data Protector Configurator Script

You must run the Big Data Protector configurator script to download certificates from the ESA, and create the parcels and CSDs for Big Data Protector.

► To run the configurator script and generate the Big Data Protector Parcels and CSDs:

- Run the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* script from the directory where it is extracted. The prompt to continue the configuration of Big Data Protector appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP PVC Base

Do you want to continue? [yes or no]:
```

- To start the configuration of Big Data Protector, type *yes*.

- Press ENTER.

The prompt to select the type of installation files appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Select the type of Installation files you want to generate.
[ 1: Create All ] : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

- To create the Big Data Protector parcels and CSDs, type *1*.
- To update the *PTY_CERT* parcels with an incremented patch version, type *2*.

Note: For more information about updating the *PTY_CERT* parcel, refer to section [Updating the Certificates Parcel](#).

- To update the *PTY_FLUENTBIT_CONF* parcel with an incremented patch version, type *3*.

Note: For more information about updating the *PTY_FLUENTBIT_CONF* parcel, refer to section [Updating the Fluent Bit Parcel](#).

- Press ENTER.

The prompt to select the operating system for the Cloudera Manager parcel appears.

```
Select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el7 ]   : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 2: el8 ]   : RHEL 8 and clones (CentOS, Scientific Linux, etc)
[ 3: el9 ]   : RHEL 9 and clones (CentOS, Scientific Linux, etc)
[ 4: sles12 ] : SuSE Linux Enterprise Server 12.x

Enter the no.:
```

- Depending on the requirements, type *1*, *2*, *3*, or *4* to select the operating system version for the Big Data Protector parcels.

- Press ENTER.

The prompt to enter the ESA hostname or IP address appears.

```
Enter the ESA Hostname or IP Address:
```

- Enter the ESA hostname or IP address.



11. Press ENTER.

The prompt to enter the ESA host listening port appears.

```
Enter ESA host listening port [8443]:
```

12. If you want to use the default value of the ESA host listening port, which is *8443*, then press ENTER.

13. If you have configured an external proxy having connectivity with the ESA to download the certificates from the ESA, then enter the external Proxy listening port.

14. Press ENTER.

The prompt to enter the ESA user name appears.

```
Enter ESA Username:
```

15. Enter the ESA user name.

16. Press ENTER.

The prompt to enter the password for the ESA appears.

```
Enter host password for user '<user_name>':
```

17. Enter the ESA administrator password.

18. Press ENTER.

The certificates are downloaded from the ESA and the prompt to create the *PTY_FLUENTBIT_CONF* parcel containing the custom Fluent Bit configuration file(s) for an external audit store appears.

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current |
|---------|------------|---------|---------------|---------|---------|--------|---------|
| Dload | Upload | Total | | Spent | Left | Speed | |
| 100 | 20480 | 100 | 20480 | 0 | 0 | 14401 | 0 |
| | | | | 0:00:01 | 0:00:01 | --::-- | 14402 |

Do you want to package any custom Fluent-Bit configuration files for External Audit Store?
 [yes] : Create a *PTY_FLUENTBIT_CONF* parcel containing configuration files to be used with External Audit Store.
 [no] : Skip this step.

```
[ yes or no ]:
```

19. To package the Fluent Bit configuration file(s) for an external audit store, type *yes*.

20. Press ENTER.

The prompt to enter the local directory path containing the Fluent Bit configuration files appears.

Do you want to package any custom Fluent-Bit configuration files for External Audit Store?
 [yes] : Create a *PTY_FLUENTBIT_CONF* parcel containing configuration files to be used with External Audit Store.
 [no] : Skip this step.
 [yes or no] : yes
 Creation of *PTY_FLUENTBIT_CONF* parcel is enabled.
 Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:

Note: The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across the CDP nodes through the Cloudera Manager. Ensure that you name the custom Fluent Bit configuration files for the external audit store with the **.conf* extension.

21. Enter the local directory path that contains the Fluent Bit configuration files.

22. Press ENTER.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration
files for External Audit Store: /root/fluentbit_file_output
```

```
Generating Installation files...
```

```
Big Data Protector parcels & CSDs are generated in ./Installation_Files/ directory.
```

NOTE:

```
Copy Big Data Protector CSDs (jars) to Cloudera Manager local csd repository.
Copy Big Data Protector parcels (*.parcel and *.sha files) to Cloudera Manager local
parcel repository.
```

The configurator script generates the following Big Data Protector parcels and CSDs in the *Installation_Files* directory:

- *BDP_PEP-9.1.0.0.x.jar*
- *PTY_BDP-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_BDP-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*
- *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*
- *PTY_PROXY-9.1.0.0.x.jar*

If you type *no* at the prompt to create the *PTY_FLUENTBIT_CONF* parcel, then the installer will skip the creation of the Fluent Bit parcel and proceed to generate the installation files.

```
Do you want to package any custom Fluent-Bit configuration files for External Audit
Store?
```

```
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used
with External Audit Store.
[ no ] : Skip this step.
```

```
[ yes or no ] : no
```

```
Creation of PTY_FLUENTBIT_CONF parcel is skipped.
```

```
Generating Installation files...
```

```
Big Data Protector parcels & CSDs are generated in ./Installation_Files/ directory.
```

NOTE:

```
Copy Big Data Protector CSDs (jars) to Cloudera Manager local csd repository.
Copy Big Data Protector parcels (*.parcel and *.sha files) to Cloudera Manager local
parcel repository.
```

12.6.1.5 Setting up the Big Data Protector Parcels and CSDs

After the Big Data Protector parcels and CSDs are copied to the local Cloudera repository directories, you must restart the Cloudera SCM server to ensure that Cloudera Manager identifies the new CSD and parcel files and displays the Big Data Protector services in the **Add Services** section in Cloudera Manager.

► To set up the Big Data Protector Parcels and CSDs:

1. Login to the Master node.

Caution: Ensure that you delete the older versions of the Big Data Protector parcels and *.jar* files before you install the new parcels and *.jar* files to the local repository of the Cloudera Manager.

2. Copy the following Big Data Protector parcels with the *.parcel* extension and their corresponding checksum files with the *.sha* extension to the local parcel repository of Cloudera Manager:

- *PTY_BDP-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_BDP-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*
- *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel.sha*

Note: The local parcels for the Cloudera Manager are stored in the */opt/cloudera/parcel-repo/* directory.

3. Copy the following *.jar* files file to the local CSD repository:

- *BDP_PEP-9.1.0.0.x.jar*
- *PTY_PROXY-9.1.0.0.x.jar*

Note:

- Copy the *PTY_PROXY-9.1.0.0.x.jar* file only if you want to configure a proxy service.
- The local CSD or *.jar* files for Cloudera Manager are stored in the */opt/cloudera/csd/* directory.

4. Navigate to the local parcel repository directory.

Note: The local parcel files are available in the */opt/cloudera/parcel-repo/* directory.

5. To assign the ownership permissions for the *Cloudera SCM* user to the Protegility Big Data Protector parcels and checksum files, run the following command:

```
chown cloudera-scm:cloudera-scm PTY_*
```

6. Press ENTER.

7. To assign *640* permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```

8. Press ENTER.

The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.

9. Navigate to the local CSD repository directory.

Note: The local CSD or *.jar* files are available in the */opt/cloudera/csd* directory.

10. To assign the ownership permissions for the Cloudera SCM user to the Big Data Protector CSD or *.jar* files, run the following command:

```
chown cloudera-scm:cloudera-scm *
```

11. Press ENTER.



12. To assign *640* permissions to the CSD or *.jar* files, run the following command.

```
chmod 640 *
```

13. Press ENTER.

The command assigns read and write permissions to the owner, read permissions to the group, and restricts access for all other users.

14. To restart the Cloudera SCM server and load the Big Data Protector CSDs in the Cloudera Manager, run the following command:

```
service cloudera-scm-server restart
```

15. Press ENTER.

The Cloudera Manager detects the new parcels in the local parcel repository.

Note: You must restart the Cloudera SCM server to ensure that the Big Data Protector services are listed on the **Add Services** page in Cloudera Manager.

12.6.1.6 Distributing the Big Data Protector Parcels to the Nodes

You must distribute the following Big Data Protector parcels to the nodes in the cluster before installing or activating them on the nodes:

- Big Data Protector parcel: *PTY_BDP*
- Certificates parcel: *PTY_CERT*
- Fluent Bit configuration parcel: *PTY_FLUENTBIT_CONF*

Note: To distribute the Big Data Protector parcels to the nodes, *Cluster Administrator* privileges are required.

For more information about the required role, refer to <https://docs.cloudera.com/cloudera-manager/7.1.1/managing-clusters/topics/cm-parcels.html>.

Note: The build number in the screen shots for the Cloudera Manager user interface will reflect version number of the Big Data Protector build that you downloaded from the [My.Protegility](#) portal.

► To distribute the Big Data Protector Parcels to the Nodes in the Cluster:

1. Using a browser, navigate to the Cloudera Manager page.

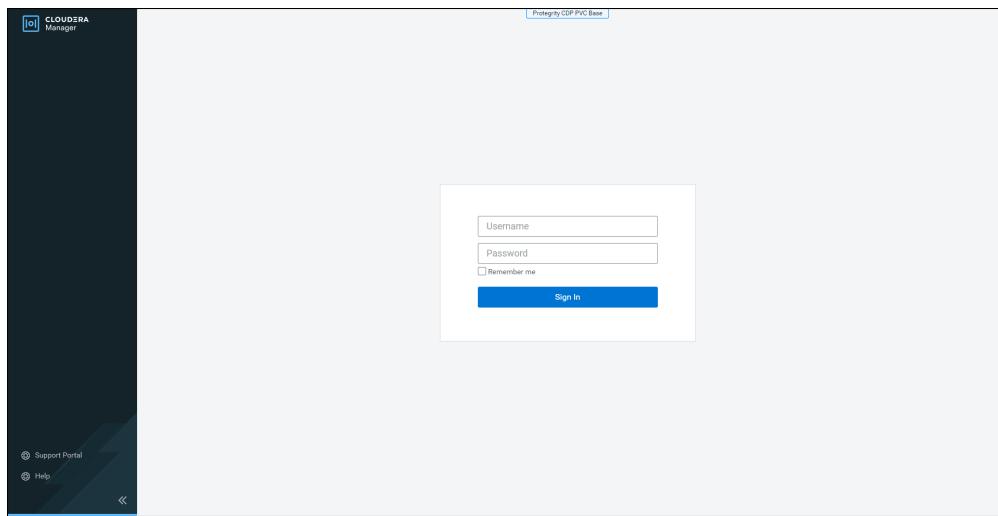


Figure 12-20: Cloudera Manager screen

2. Enter the **Username**.
3. Enter the **Password**.
4. Click **Sign In**.

The Cloudera Manager Home page appears.

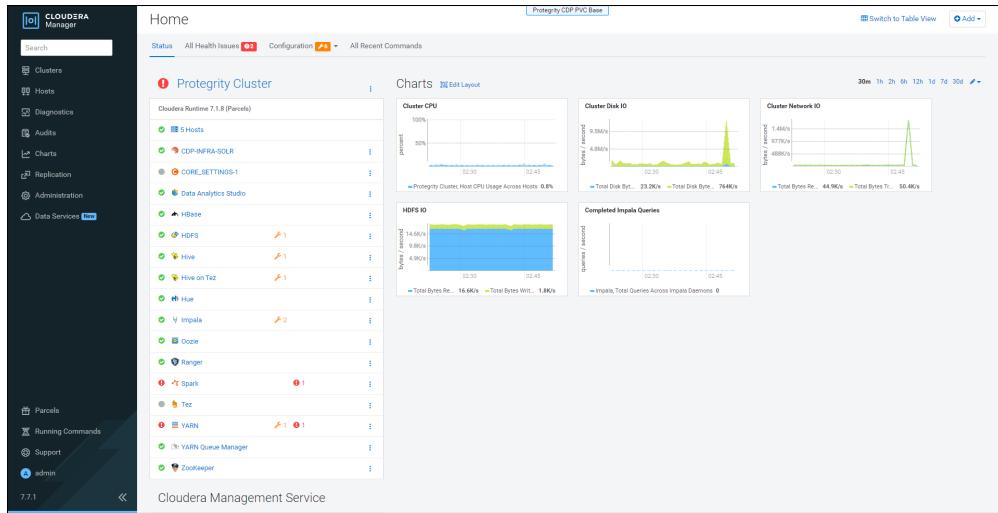


Figure 12-21: Cloudera Manager Home Page

5. Navigate to **Administration > Settings**.
The **Settings** page appears.
6. To view the settings related to parcels, from the **Filters** pane, under **CATEGORY**, click **Parcels**.
The options related to the parcels appear.
7. Ensure that you select the **Create Users and Groups**, and **apply File Permissions for Parcels** checkbox.

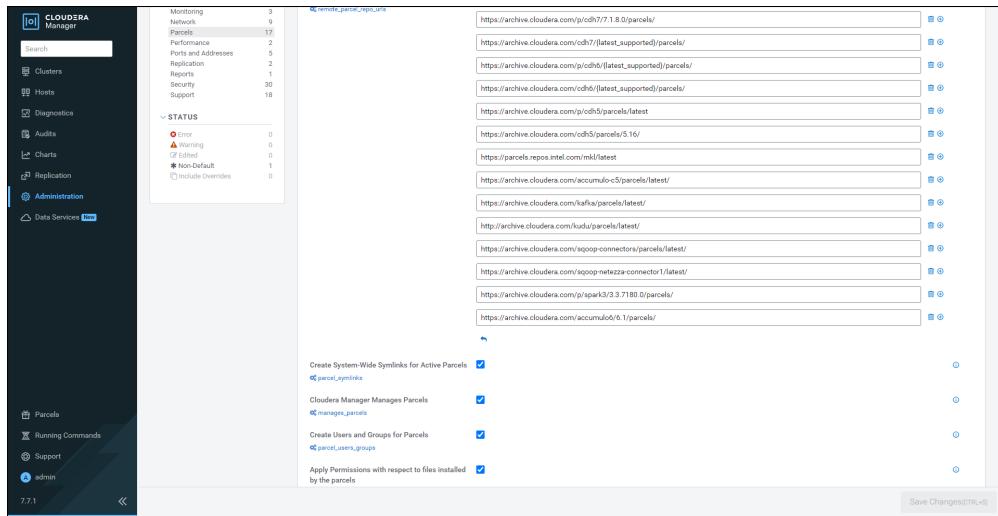


Figure 12-22: Options on the Settings screen

- From the left pane, click **Parcels**.

The **Cloudera Manager Parcels** page appears.

The screenshot shows the 'Parcels' page in Cloudera Manager. The sidebar includes links for Clusters, Hosts, Diagnostics, Audits, Charts, Replication, Administration, and Data Services. The main area is titled 'Protegility Cluster' and lists various parcels with their versions and statuses. The 'PTY_BDP' parcel is highlighted with a red border. Other visible parcels include ACCUMULO, CLOUDERA_RUNTIME, CDH_6, KAFKA, KEYTRUSTEE_SERVER, KUDU, KUDU_BDP, PTY_CERT, PTY_FLUENTBIT_CONF, SOOOP_NETEZZA_CONNECTOR, SOOOP_TERADATA_CONNECTOR, and mlib. Each parcel row has a 'Download' and 'Distribute' button.

Figure 12-23: Cloudera Manager Parcels Page

Note: The *PTY_FLUENTBIT_CONF* parcel will be visible only if you choose to add the location of the Fluent Bit configuration files while generating the installation files.

- Ensure that the following Protegility parcels appear on the **Parcels** page:

- PTY_BDP*: Big Data Protector parcel
- PTY_CERT*: Certificates parcel
- PTY_FLUENTBIT_CONF*: Fluent Bit configuration parcel

| | | | |
|--------------------|--------------------|------------|------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |

Figure 12-24: Protegility Parcels

- To distribute the Big Data Protector parcel, besides the *PTY_BDP* parcel, click **Distribute**. The distribution of the Big Data Protector parcel starts.
- To distribute the Certificates parcel, besides the *PTY_CERT* parcel, click **Distribute**. The distribution of the Certificates parcel starts.

12. To distribute the Fluent Bit configuration parcel, besides the *PTY_FLUENTBIT_CONF* parcel, click **Distribute**.
The distribution of the Fluent Bit configuration parcel starts.



Figure 12-25: Distribution of Parcels Starts

After the Protegility parcels are distributed to the nodes, their status on the **Parcels** page is updated to **Distributed**, and the **Activate** button appears.

| Parcel Name | Version | Status | Action |
|---------------------------|---|------------------------|-----------------|
| ACCUMULO | 1.7.3.6.5.0.ACCUMULO5.0.p0.8 | Available Remotely | Download |
| Cloudera Runtime | 7.1.8-1.cdh7.1.8.p0.30990532 | Distributed, Activated | Deactivate |
| CDH 6 | 6.3.4.1.cdh6.3.4.p0.6751008 | Available Remotely | Download |
| CDH 5 | 5.16.2.1.cdh5.16.2.p0.8 | Available Remotely | Download |
| KAFKA | 4.1.0-1.4.1.0.p0.4 | Available Remotely | Download |
| KEYTRUSTEE_SERVER | 7.1.8-0.1.keytrustee7.1.8.0.p0.30990532 | Available Remotely | Download |
| KUDU | 1.4.0.1.kudu5.1.2.p0.8 | Available Remotely | Download |
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| SPARK3 | 3.3.0.3.3.7180.0.274-1.p0.31212967 | Available Remotely | Download |
| SOOOP_NETEZZA_CONNECTOR | 1.5.1/p | Available Remotely | Download |
| SOOOP_TERADATA_CONNECTOR | 1.5/p | Available Remotely | Download |
| m4i | 2023.1.0.46342 | Available Remotely | Download |

Figure 12-26: Protegility Parcels Distributed

12.6.1.7 Activating the Big Data Protector Parcels on the Nodes

After distributing the Big Data Protector parcels on the cluster nodes, you must activate the parcels so that the Big Data Protector-related services can be added and started on the nodes in the cluster.

► To activate the Big Data Protector Parcels on the Nodes:

1. Using a browser, navigate to the Cloudera Manager screen.

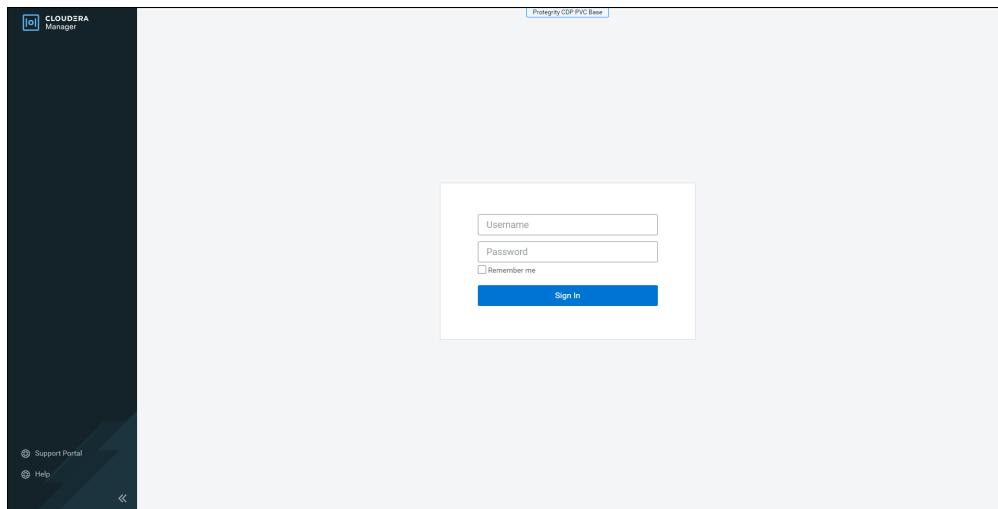


Figure 12-27: Cloudera Manager Login Page

2. Enter the **Username**.
3. Enter the **Password**.
4. Click **Sign In**.

The Cloudera Manager Home page appears.

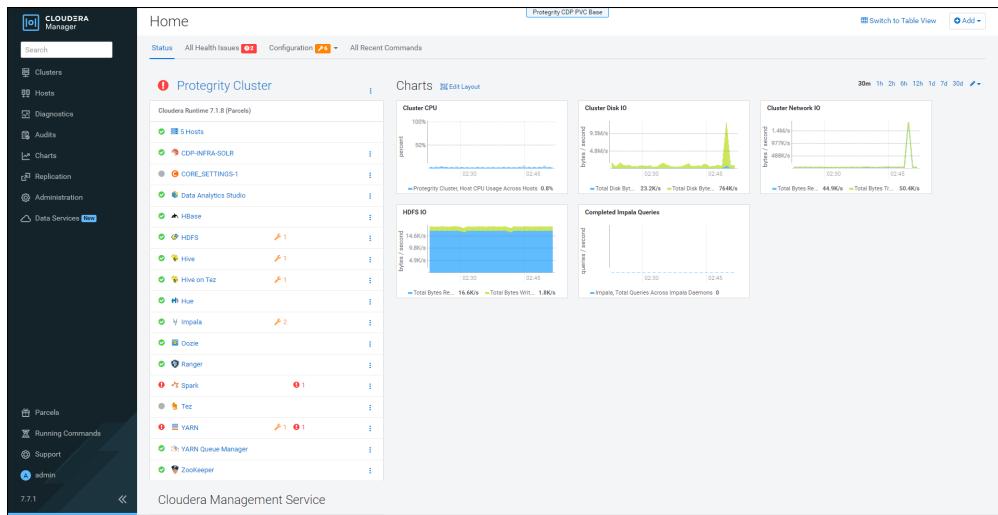


Figure 12-28: Cloudera Manager Home Page

5. From the left pane, click **Parcels**.

The **Cloudera Manager Parcels** page appears.

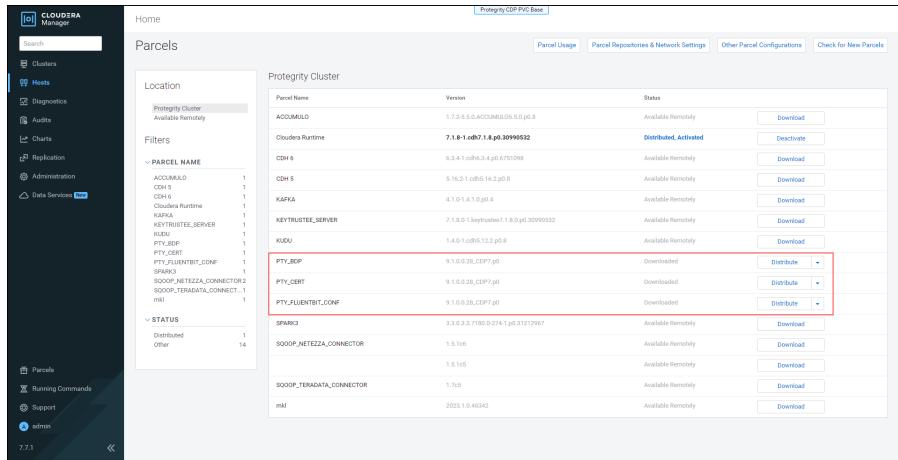


Figure 12-29: Cloudera Manager Parcels Page

Note: The *PTY_FLUENTBIT_CONF* parcel will be visible only if you choose to add the location of the Fluent Bit configuration files while generating the installation files.

6. To activate the Big Data Protector parcel, besides the *PTY_BDP* parcel, click **Activate**. A prompt to confirm the activation of the parcel appears.





Figure 12-30: Prompt to Activate the Big Data Protector Protegility Parcel

7. To activate the Big Data Protector parcel, click **OK**.
The Big Data Protector parcel is activated.
8. To activate the Certificates parcel, besides the *PTY_CERT* parcel, click **Activate**.
A prompt to confirm the activation of the parcel appears.



Figure 12-31: Prompt to Activate the Certificates Parcel

9. To activate the Certificates parcel, click **OK**.
The Certificates parcel is activated.
10. To activate the Fluent Bit configuration parcel, besides the *PTY_FLUENTBIT_CONF* parcel, click **Activate**.
A prompt to confirm the activation of the parcel appears.



Figure 12-32: Prompt to Activate the Fluent Bit Configuration Parcel

11. To activate the *PTY_FLUENTBIT_CONF* parcel, click **OK**.
After the Protegility parcels are activated on the nodes, their status on the **Parcels** page is updated to **Distributed, Activated**.
The **Deactivate** button appears.

| Parcel Name | Version | Status | Action |
|--------------------------|---|------------------------|----------------------------|
| Accumulo | 1.7.2-5.5.0-ACUMULO-0.0.0-8 | Available Remotely | Download |
| CDH 6 | 6.3.4-1.0.0-0.1-p0.0751098 | Available Remotely | Download |
| CDH 5 | 5.16.2.1-p0.0.16.2-p0.8 | Available Remotely | Download |
| KAFKA | 4.1.0-1.1.0-p0.4 | Available Remotely | Download |
| KEYTRUSTEE_SERVER | 7.1.0-0.1-keytrustee-7.1.0.0-p0.3099532 | Available Remotely | Download |
| KUDU | 1.4.0-0.1-p0.0.10.2-p0.8 | Available Remotely | Download |
| PTY_BDP | 9.1.0-0.2_kCDP7_p0 | Distributed, Activated | Deactivate |
| PTY_CERT | 9.1.0-0.2_kCDP7_p0 | Distributed, Activated | Deactivate |
| PTY_FUENTBIT_CONF | 9.1.0-0.2_kCDP7_p0 | Distributed, Activated | Deactivate |
| SPARK | 3.3.0-0.3.7100-0.214-1-p0.31212967 | Available Remotely | Download |
| SOOOP_NETEZZA_CONNECTOR | 1.5.16 | Available Remotely | Download |
| SOOOP_TERADATA_CONNECTOR | 1.7c5 | Available Remotely | Download |
| mail | 2023.1.0.46342 | Available Remotely | Download |

Figure 12-33: Protegility Parcels Activated

12. Navigate to the Cloudera Manager Home page.

The Cloudera Manager Home page appears and the required services are displayed with their configuration states in the Cluster area.

Figure 12-34: Hadoop Services

Note: If the configuration state of any Hadoop services is stale, then  appears beside the service.

13. To redeploy the service configuration and restart the service, besides the required Hadoop service, click .

The screenshot shows the Cloudera Manager interface with the 'Clusters' section selected. In the center, there's a table titled 'Stale Configurations' with two columns: 'SERVICE' and 'ROLE TYPE'. The 'SERVICE' column lists HDFS, Hive, Oozie, Spark, YARN, and ZooKeeper, each with a count of 1. The 'ROLE TYPE' column lists DataNode, HBase Thrift Server, HBase Master, Hive Metastore Server, Hue Server, JobHistory Server, Kerberos Ticket Renewer, Load Balancer, Master, NameNode, MapReduce, Oozie Server, ResourceManager, SecondaryNameNode, and Server, also each with a count of 1. A blue button at the bottom right says 'Restart Stale Services'.

Figure 12-35: Configuration State of Hadoop Services

- To restart the configuration, click **Restart Stale Services**.

The **Review Changes** page appears.

The screenshot shows the 'Review Changes' page. It has a sidebar with 'Clusters' selected. The main content area has a heading 'Review Changes' with the sub-instruction 'All services running with outdated configurations in the cluster and their dependencies will be restarted.' Below this is a checkbox labeled 'Re-deploy client configuration'. At the bottom right are 'Back' and 'Restart Now' buttons.

Figure 12-36: Review Changes Page

- Click **Restart Now**.

Cloudera Manager restarts the cluster to re-deploy the updated configuration on all the nodes in the cluster. After the configuration deployment is complete, Cloudera Manager enables the **Continue** button.

Review Changes

Command Details

Restart Awaiting Staleness Computation Command

Status: Finished Context: Protegility Cluster Jun 2, 3:54:36 AM 6.9m

All requested services successfully restarted.

Completed 2 of 2 step(s).

Show All Steps Show Only Failed Steps Show Only Running Steps

| Action | Target | Time | Duration |
|--|---------------------|-------------------|----------|
| Execute global command Wait for configuration staleness computation | | Jun 2, 3:54:36 AM | 39ms |
| Execute command Restart on cluster Protegility Cluster | Protegility Cluster | Jun 2, 3:54:37 AM | 6.9m |
| Execute command Stop on cluster Protegility Cluster | Protegility Cluster | Jun 2, 3:54:38 AM | 2.5m |
| Execute command Stop concurrently on 2 services | | Jun 2, 3:54:38 AM | 1.35s |
| Execute command Stop concurrently on 3 services | | Jun 2, 3:54:40 AM | 2.1m |
| Execute command Stop concurrently on 2 services | | Jun 2, 3:56:46 AM | 1.38s |
| Execute command Stop concurrently on 2 services | | Jun 2, 3:57:47 AM | 10.44s |
| Execute command Stop on service CDP-INFRA-GOLR | CDP-INFRA-GOLR | Jun 2, 3:56:58 AM | 7.68s |
| Execute command Stop on service HDFS | HDFS | Jun 2, 3:57:05 AM | 2.24s |
| Execute command Stop on service YARN Queue Manager | YARN Queue Manager | Jun 2, 3:57:08 AM | 1.34s |
| Execute command Stop concurrently on 2 services | | Jun 2, 3:57:09 AM | 1.89s |
| Execute command Start on cluster Protegility Cluster | Protegility Cluster | Jun 2, 3:57:12 AM | 3.9m |
| Execute command Deploy Client Configuration on cluster Protegility Cluster | Protegility Cluster | Jun 2, 4:01:05 AM | 25.76s |

Back Continue

Figure 12-37: Restart Stale Services Page

16. Click **Continue**.

The **Cloudera Manager Home** page appears.

17. Restart the Cloudera Management Service to re-deploy the service configuration for the stale configurations.

12.6.1.8 Starting the Big Data Protector Services

After you have distributed and activated the parcels on the nodes in the cluster, you must start the services to begin using the Big Data Protector. Depending on your connectivity options, you will have to start the following services:

- Big Data Protector PEP Service
- Proxy Service

Note: You require the proxy service only if your network configuration requires a proxy server that acts as a connection between the nodes and the ESA.

12.6.1.8.1 Starting the Big Data Protector PEP Service

If you want to use the Big Data Protector, then you must start the Big Data Protector PEP service on all the nodes in the cluster.

Before you begin

Before starting the Big Data Protector PEP service, ensure that the following Big Data Protector-related parcels are in the **Activated** state:

- Big Data Protector parcel: *PTY_BDP*
- Certificates parcel: *PTY_CERT*
- Fluent Bit configuration parcel: *PTY_FLUENTBIT_CONF*

► To start the Big Data Protector PEP Service on the Nodes:

1. Login to the Cloudera Manager web interface.
2. Besides the cluster name, click the kebab menu . The cluster drop-down list appears.



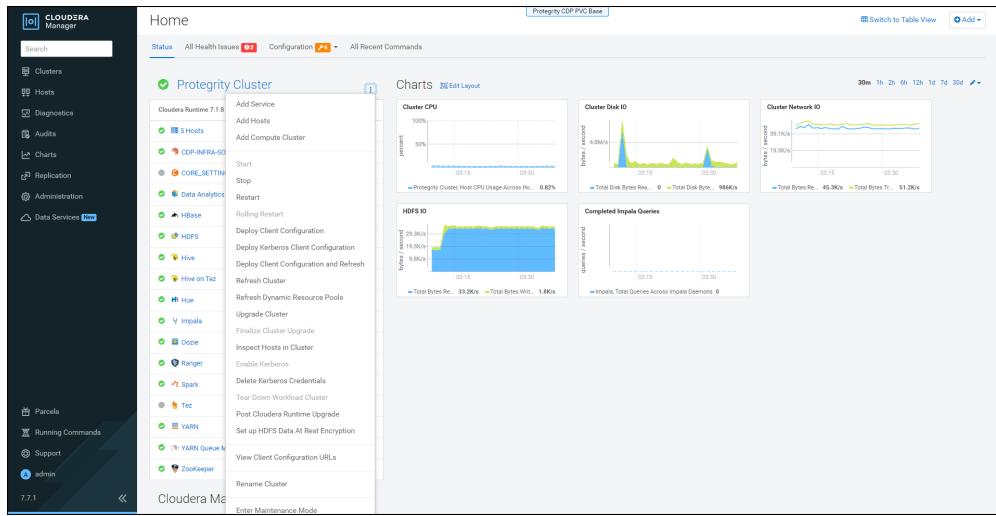


Figure 12-38: Cluster Drop Down

3. Select Add Service.

The cluster services wizard page appears.

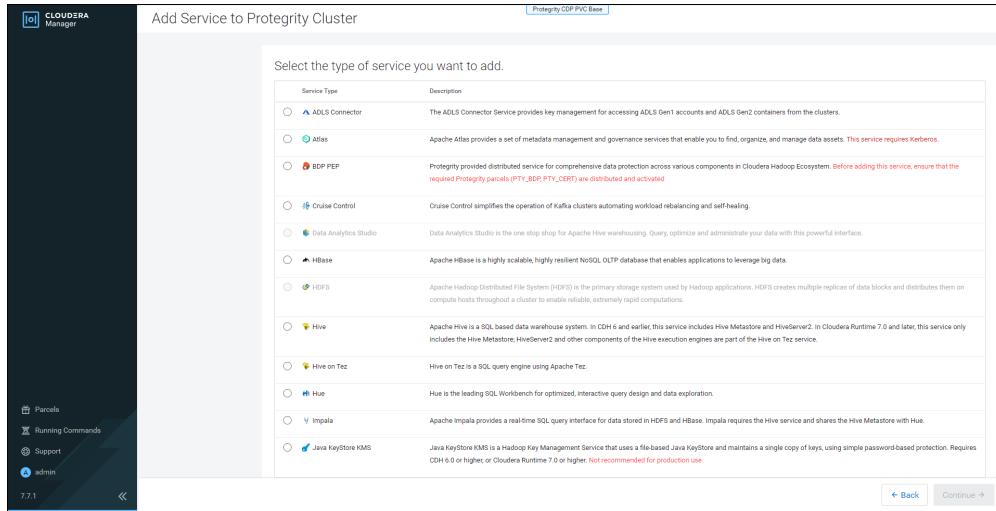


Figure 12-39: Cluster Services Page

4. From the Service Type list, select BDP PEP.

When you select the service, Cloudera enables the Continue button.

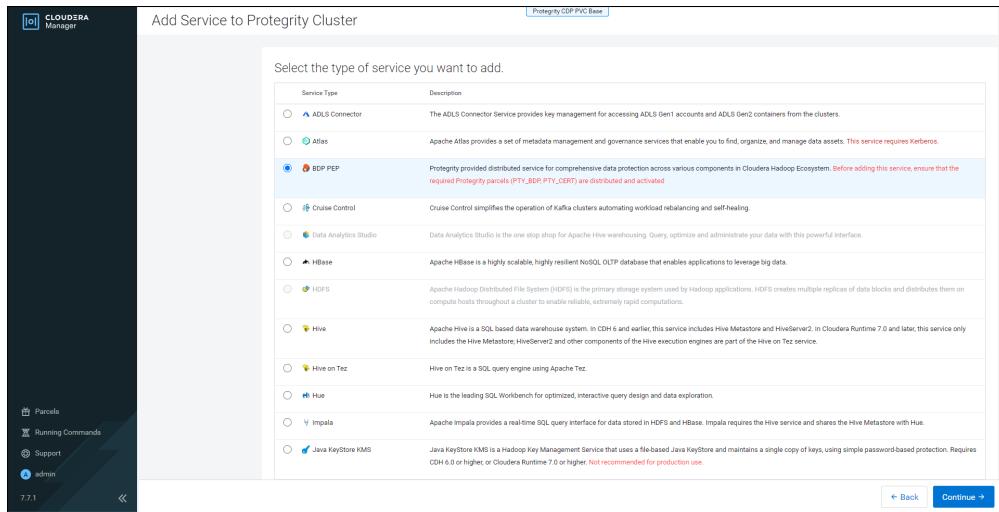


Figure 12-40: Cluster Services Page

5. Click Continue.

The **Assign Roles** page appears.

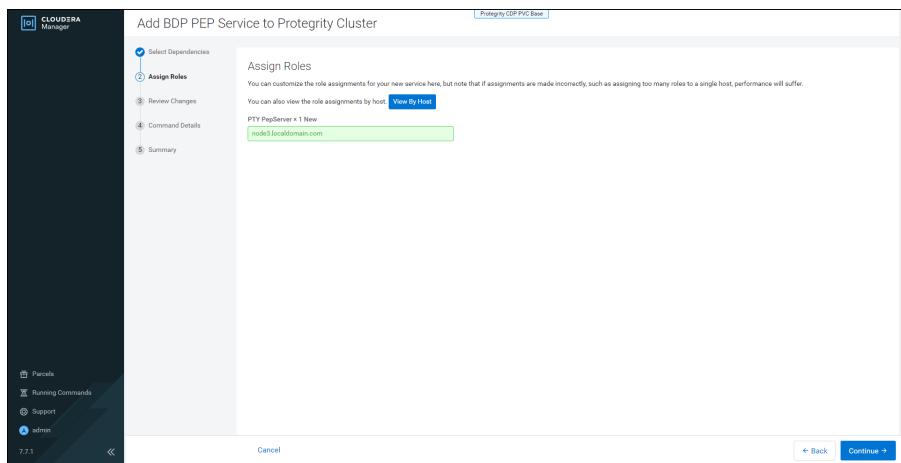


Figure 12-41: Assign Roles Page

6. Click the highlighted text box.

The list of nodes in the cluster appear.

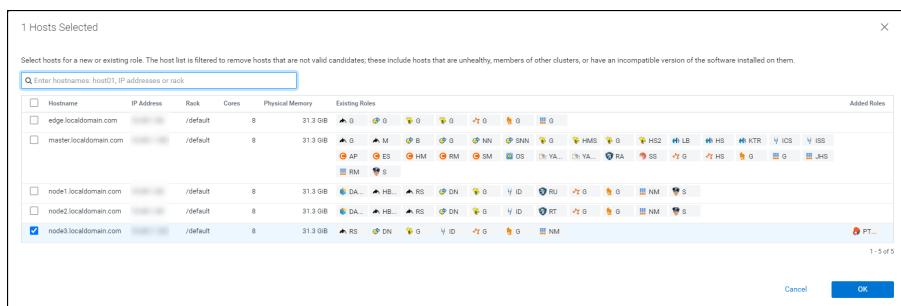


Figure 12-42: List of Available Nodes to Install the BDP PEP Service

7. To select all the host nodes in the list to install the BDP PEP service, select the **Hostname checkbox.**

When you select the **Hostname** checkbox, Cloudera enables the **OK** button.

Note: The *PTY PepServer* and the *PTY Log Forwarder* roles are installed on the selected node.

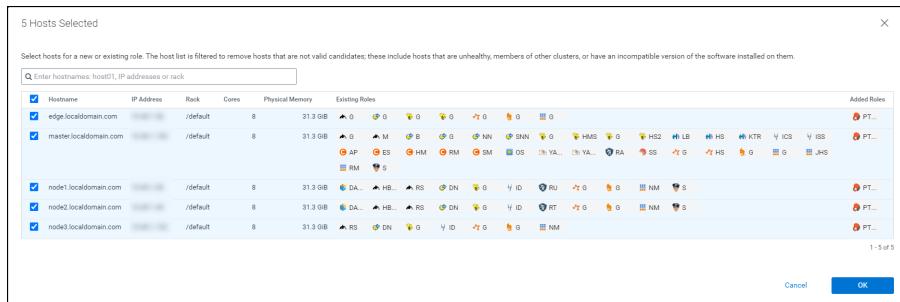


Figure 12-43: Nodes Selected to Install the BDP PEP Service

8. Click OK.

The **Assign Roles** page appears with the nodes in the cluster, which are selected for installing the *BDP PEP* service.

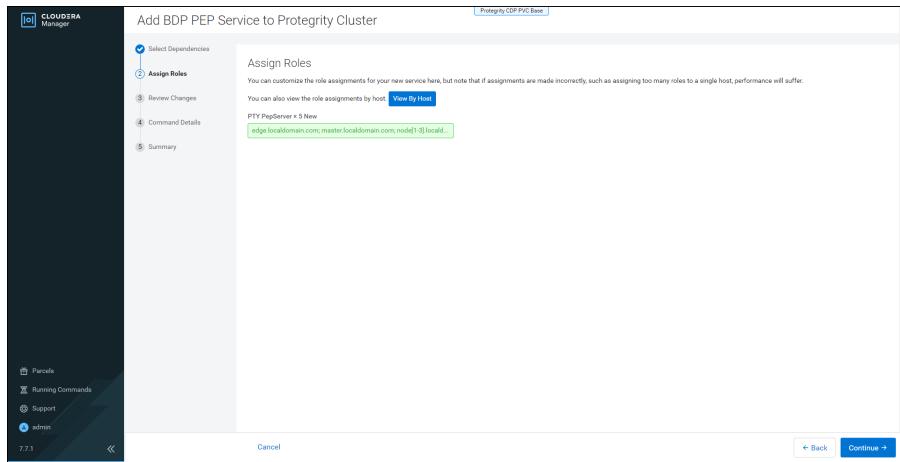


Figure 12-44: Nodes Selected to Install the BDP PEP Service

9. Click Continue.

The **Review Changes** page appears.

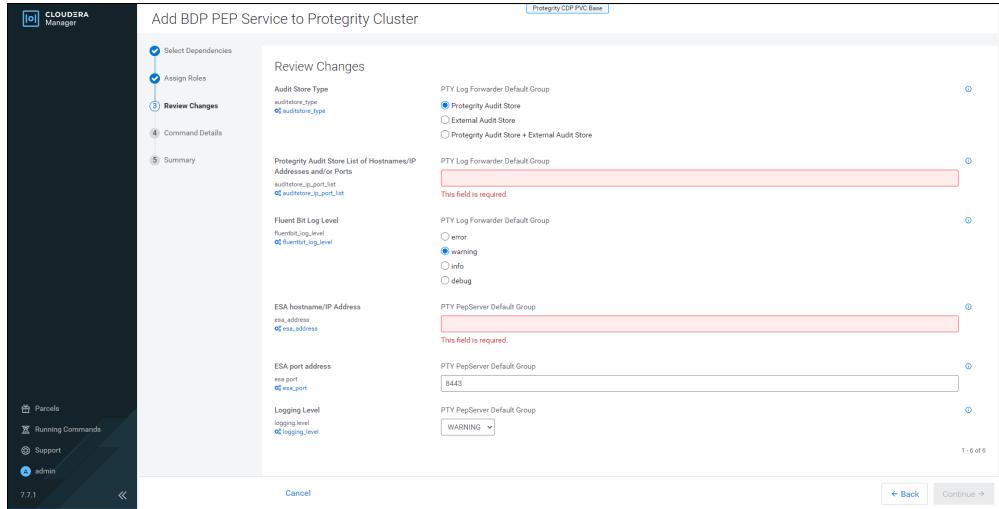


Figure 12-45: Review Changes for BDP PEP Service Page

10. Depending on the Audit Store type, select any one of the following options:

Table 12-38: Options to select the Audit Store Type

| Option | Description |
|--------------------------------|---|
| Protegility Audit Store | To use the default setting select the Protegility Audit Store option. If you select Protegility Audit Store , then the default Fluent Bit |

| Option | Description |
|---|--|
| | configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store. |
| External Audit Store | Enter the comma-separated IP/ports using the accurate syntax in the External Audit Store box. If you select External Audit Store , then enter NA in the Protegility Audit Store List of Hostnames/IP Address and/or Ports box. Ensure that the <i>PTY_FLUENTBIT_CONF</i> parcel is distributed and activated. If you select External Audit Store , then the default Fluent Bit configuration files used for Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| Protegility Audit Store + External Audit Store | To use a combination of the default setting with an external audit store, select Protegility Audit Store + External Audit Store . If you select Protegility Audit Store + External Audit Store , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

11. Enter the IP address of the ESA in the **ESA hostname/IP Address** box.

Cloudera Manager enables the **Continue** button.

12. Click **Continue**.

The **Summary** page appears.

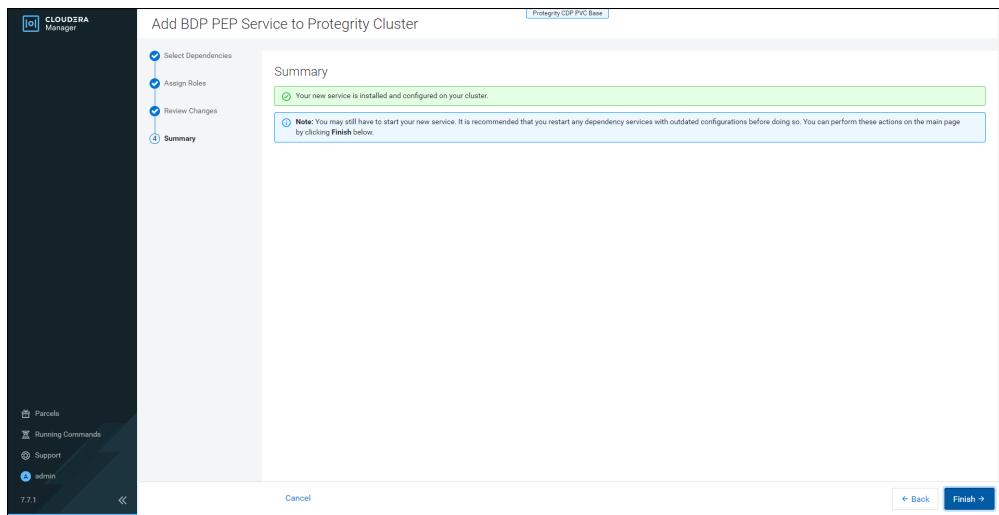


Figure 12-46: Summary Page to add the BDP PEP Service to the Protegility Cluster

13. Click **Finish**.

The Cloudera Manager Home page appears and the BDP PEP service is added on all the nodes in the cluster.

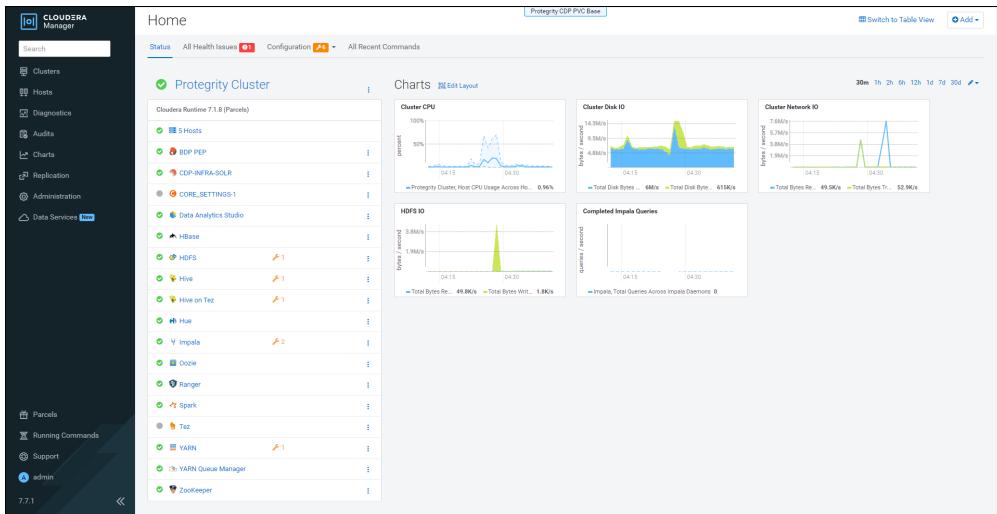


Figure 12-47: BDP PEP Service in Cluster

Note: In the Cloudera Manager native installer, there is a caveat in the *BDP PEP* service that the *PTY Log Forwarder* and the *PTY PepServer* roles will be started at the same time on a cluster node. Therefore, some of the initial PEP server application logs would not be sent to the Log Forwarder and in turn, the logs would not be forwarded to the Audit Store. After the Log Forwarder starts up, it will start forwarding the PEP server application logs.

By default, the *BDP PEP* service is stopped.

14. To start the *BDP PEP* service, besides *BDP PEP*, click the kebab menu icon . The **BDP PEP Actions** sub-menu appears.

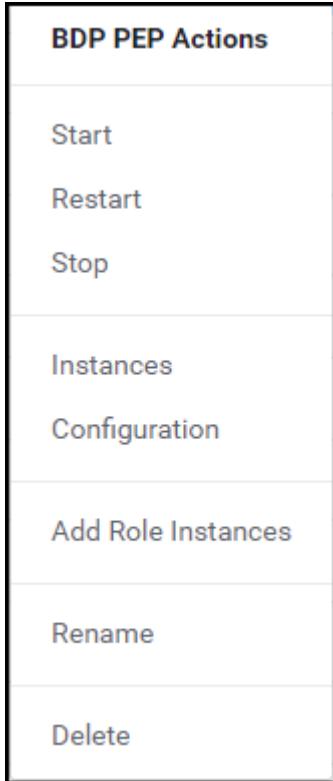


Figure 12-48: BDP PEP Actions sub-menu

15. From the sub-menu, select **Start**. The prompt to confirm the action appears.

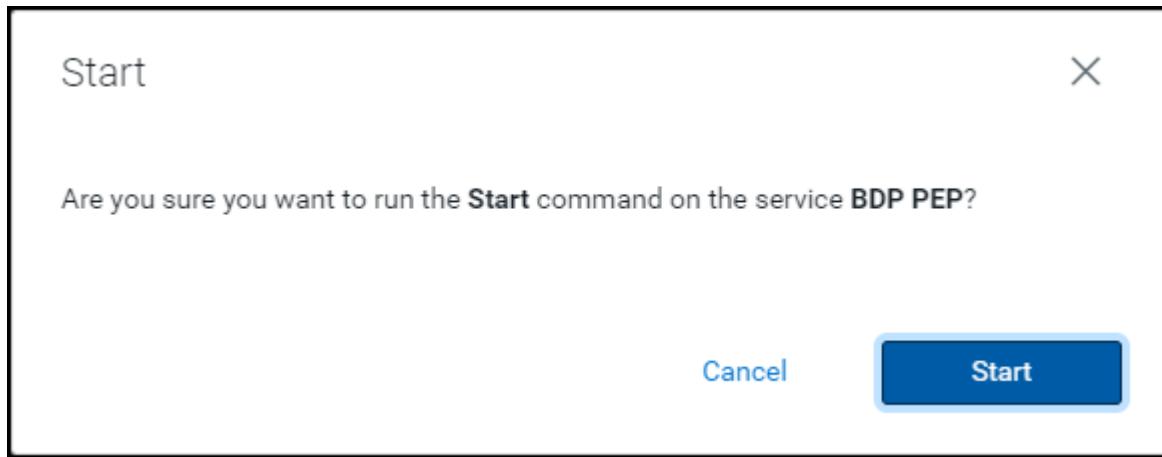


Figure 12-49: Prompt to Start the BDP PEP service

16. Click Start.

Cloudera Manager starts the BDP PEP service on all the nodes in the cluster.

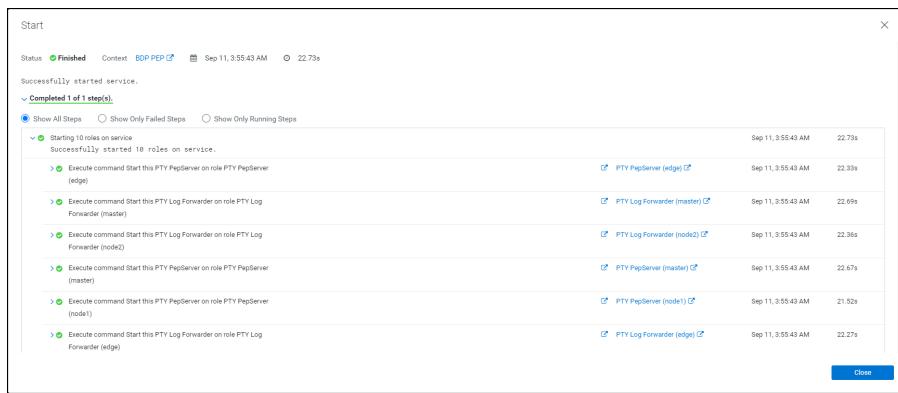


Figure 12-50: BDP PEP Service Started on all the Nodes in the Cluster

12.6.1.8.2 Starting the Proxy Service

If you want to utilize the *PTY_Proxy* service, then you must add and start the *PTY_Proxy* service on the required nodes in the cluster. Typically, the Lead (Edge) node is used for installing the proxy service.

Note: The *PTY_PROXY* service is required if the cluster is configured in a closed network system, where the nodes do not have connectivity with the ESA. Typically, you must configure the Proxy service on the *Lead (Edge)* node because it has connectivity to the ESA.

You can configure the Proxy service either when Big Data Protector is being installed or at a later stage after the Big Data Protector is already installed.

Warning: The *PTY_Proxy* service can receive and forward the TCP traffic coming from Log forwarder(s), only to a single Protegility Audit Store appliance. Multiple Audit Store endpoints are not supported.

Before you begin

Before starting the *PTY_Proxy* service, ensure that the following Big Data Protector-related parcels are in the **Activated** state:

- Big Data Protector parcel: *PTY_BDP*
- Certificates parcel: *PTY_CERT*

► To start the *PTY_Proxy* Service on the Nodes:

- Besides the cluster name, click the kebab menu . The cluster drop-down list appears.

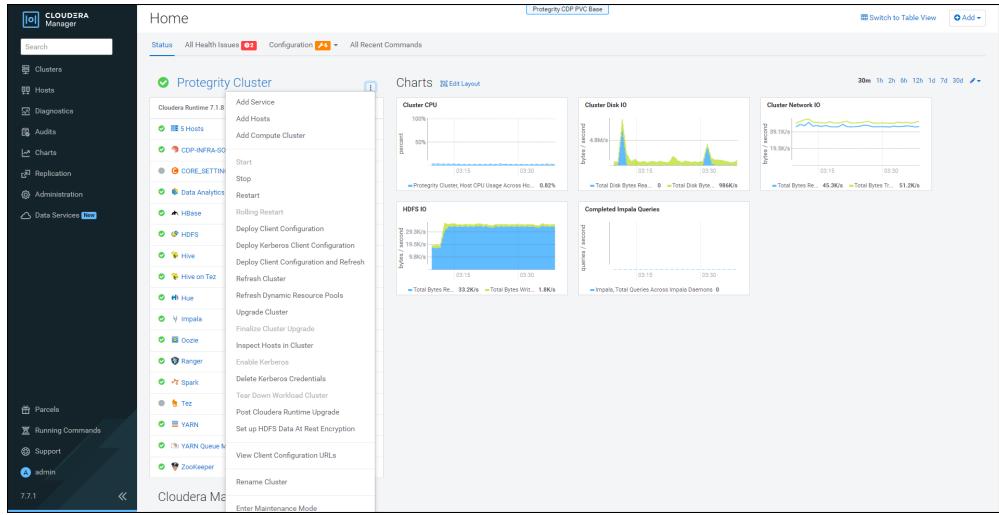


Figure 12-51: Cluster Drop-Down List

- Select **Add Service**.

The cluster services wizard page appears.

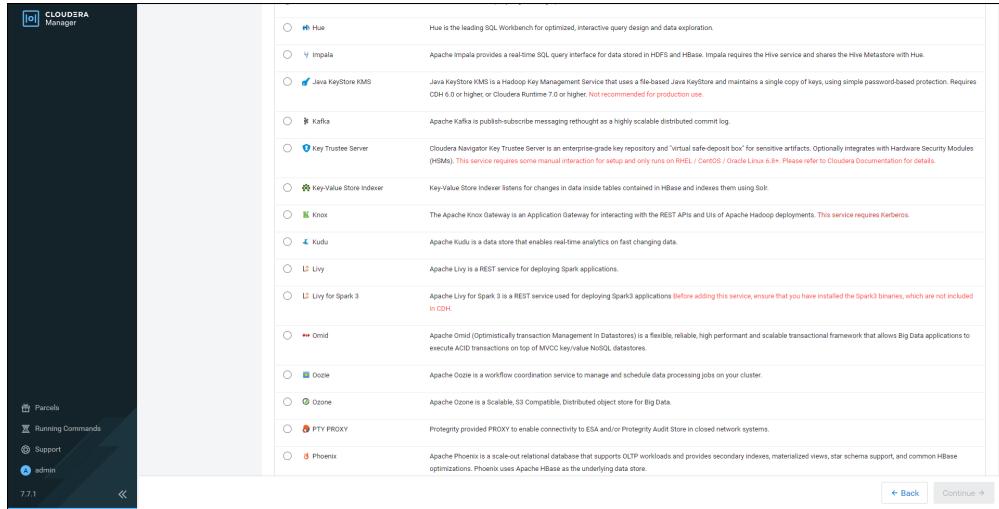


Figure 12-52: Cluster Services Page

- From the **Service Type** list, select **PTY PROXY**.

When you select the service, Cloudera enables the **Continue** button.

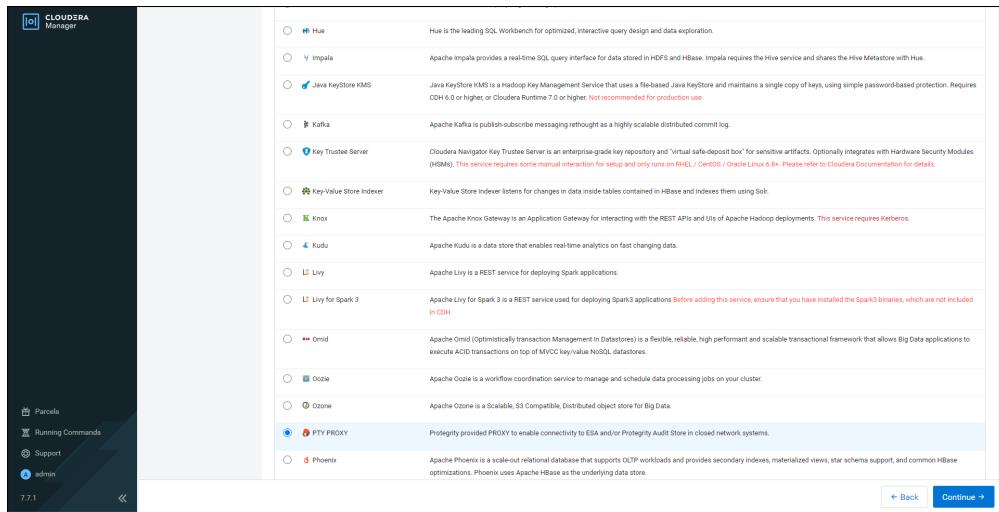


Figure 12-53: Selecting the PTY Proxy Service

- #### **4. Click Continue.**

The **Assign Roles** page appears.

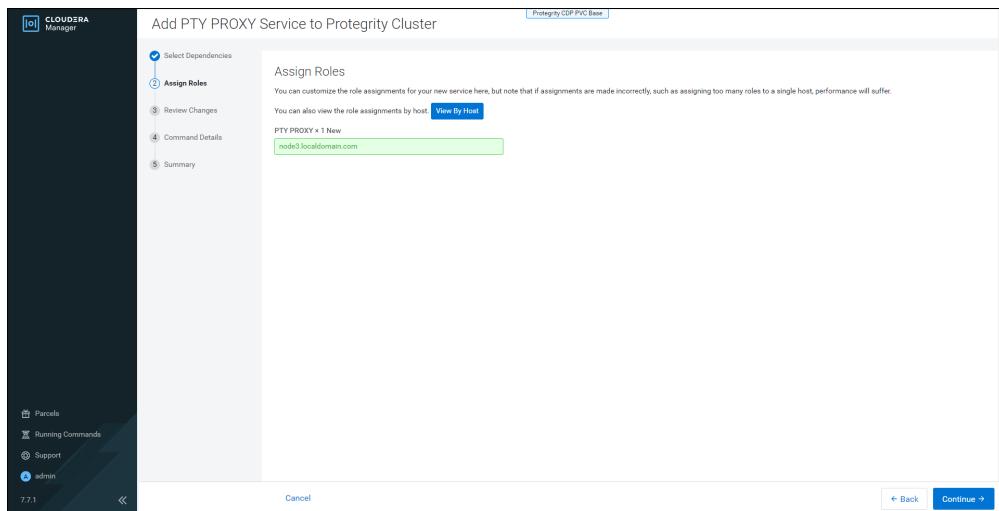


Figure 12-54: Assign Roles Page

- Click the box that contains the name of the node where you want to configure the *PTY_PROXY* service. The list of nodes in the cluster appear.

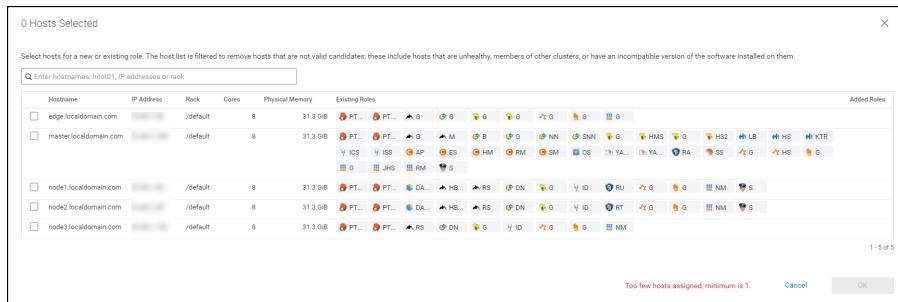


Figure 12-55: Selecting the Node to Install the PTY PROXY Service

- To select the required host node in the list to install the *PTY PROXY* service, select the checkbox against the node. Cloudera enables the **OK** button.

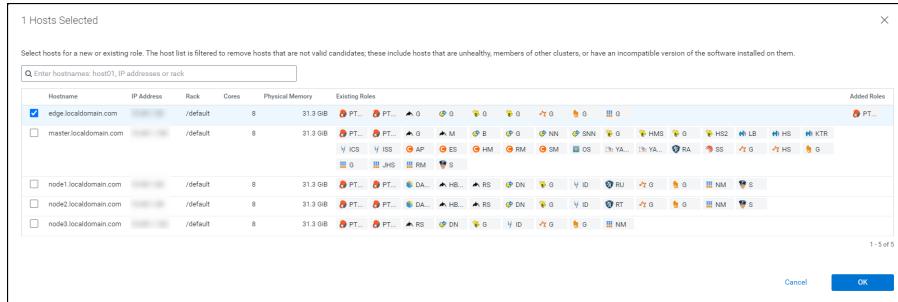


Figure 12-56: Node Selected to Install the PTY_Proxy Service

7. Click OK.

The *PTY PROXY* page is updated with the required node in the cluster, which is selected for installing the *PTY PROXY* service.

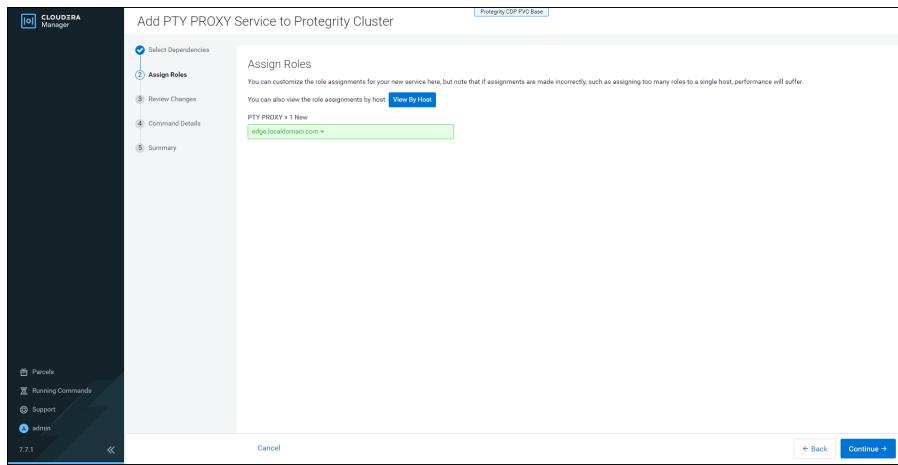


Figure 12-57: PTY PROXY Host

8. Click Continue.

The configuration **Review Changes** page appears.

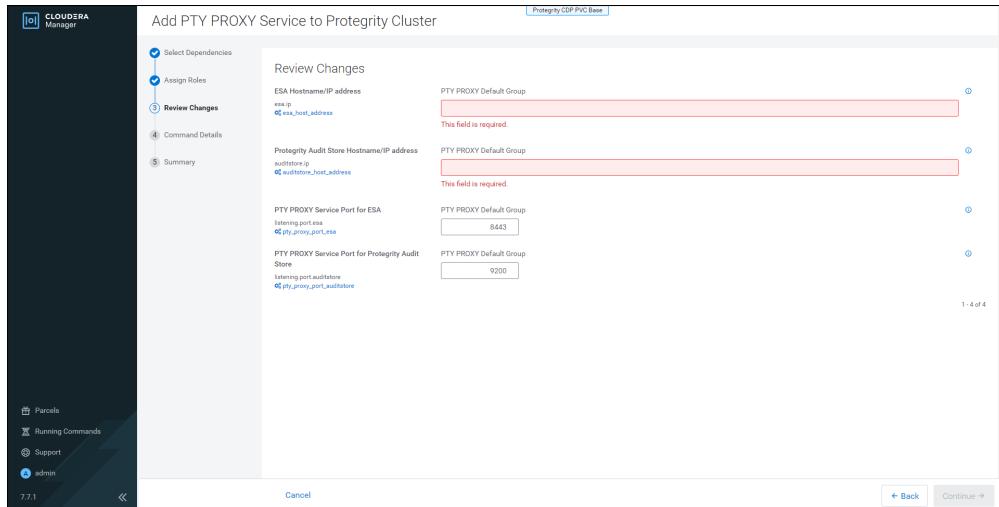


Figure 12-58: Review Changes for PTY PROXY Service Page

9. Enter the IP address of the Audit Store in the **Audit Store Hostname/IP address** box.

10. Enter the IP address of the ESA in the **ESA Hostname/IP address** box.

Cloudera enables the **Continue** button.

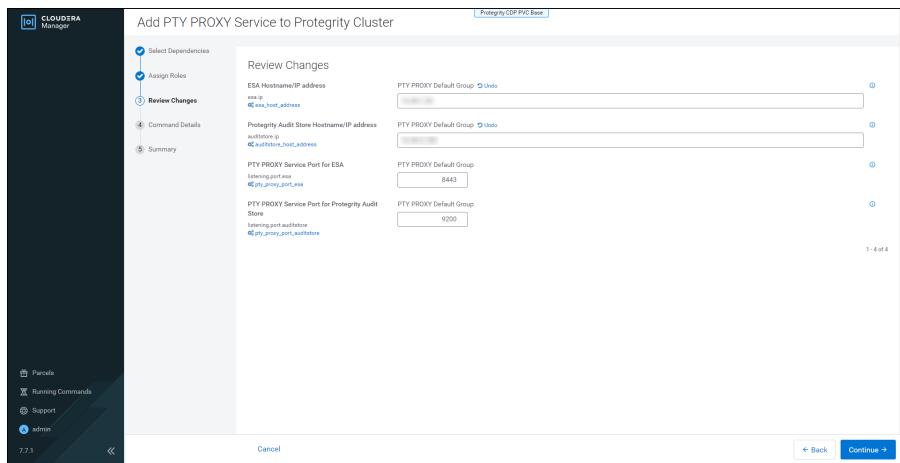


Figure 12-59: Review Changes Page to add the PTY_Proxy Service

11. Click Continue.

The Summary page appears.

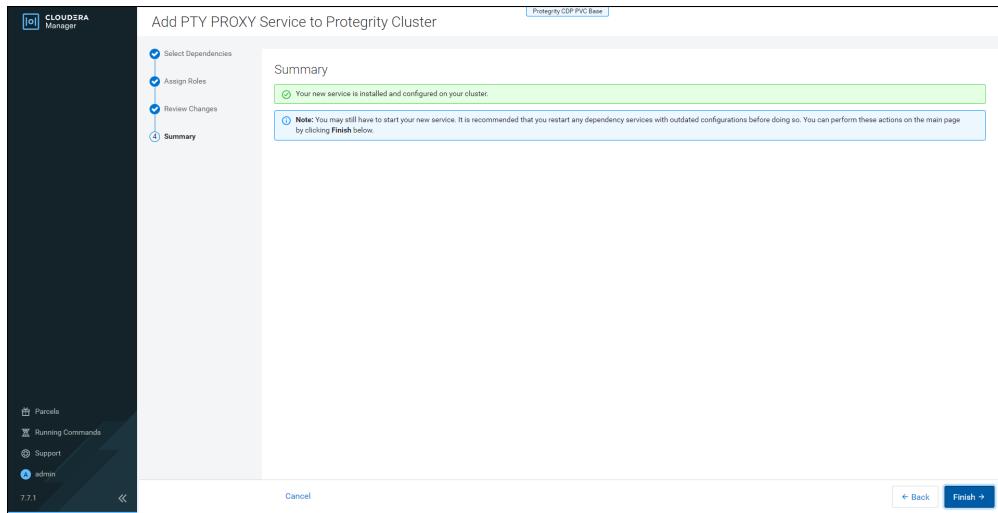


Figure 12-60: Summary Page

12. Click Finish.

The Cloudera Manager Home page appears and the *PTY PROXY* service is added on all the nodes in the cluster.

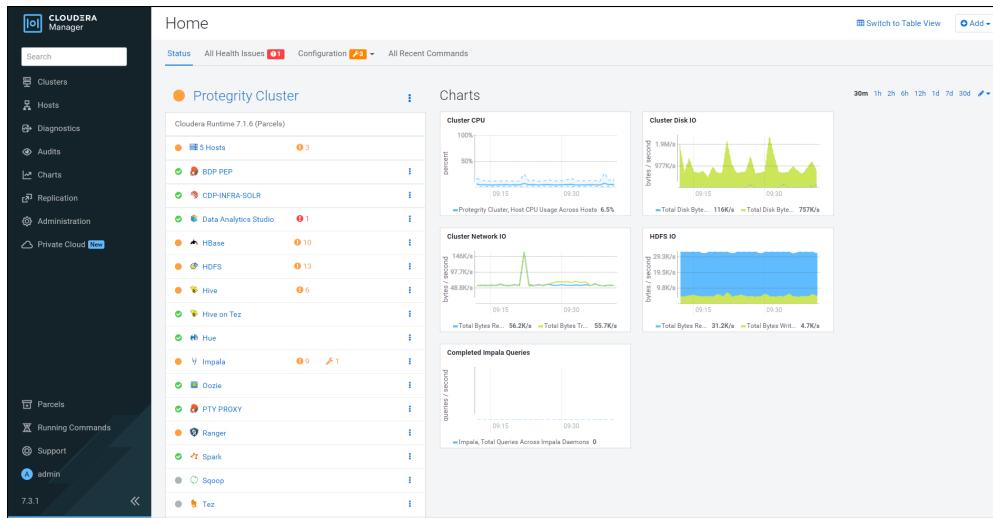


Figure 12-61: PTY PROXY Service in Cluster

Note: The *PTY Proxy* service only supports forwarding the log traffic to a single Protegility Audit Store Appliance IP address and port.

13. To start the *PTY Proxy* service, besides *PTY Proxy*, click the kebab menu . The **PTY Proxy Actions** sub-menu appears.

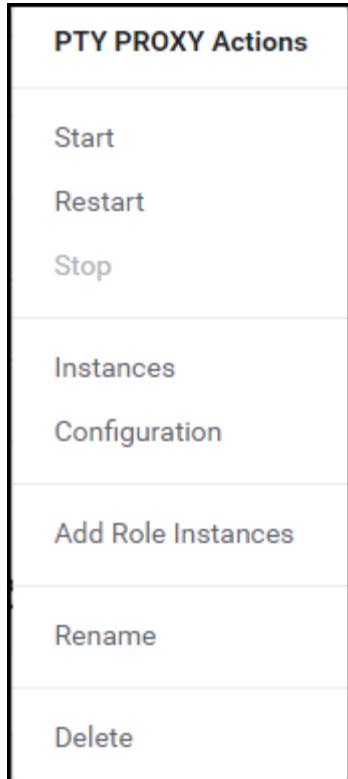


Figure 12-62: PTY Proxy Actions sub-menu

14. From the sub-menu, select **Start**. The prompt to confirm the action appears.

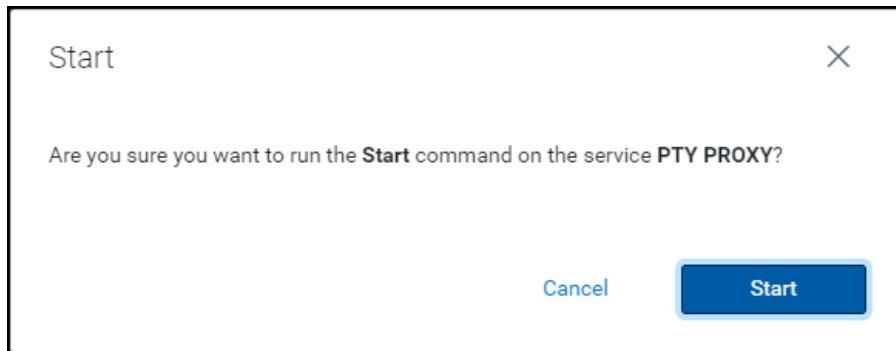


Figure 12-63: Prompt to Start the PTY Proxy service

15. Click **Start**.

Cloudera Manager starts the *PTY Proxy* service on the node where you have installed it.



Figure 12-64: Starting the PTY Proxy Service

16. Click **Close**.

The Cloudera Manager redirects you to the home page with the *PTY Proxy* service installed and started on the cluster.

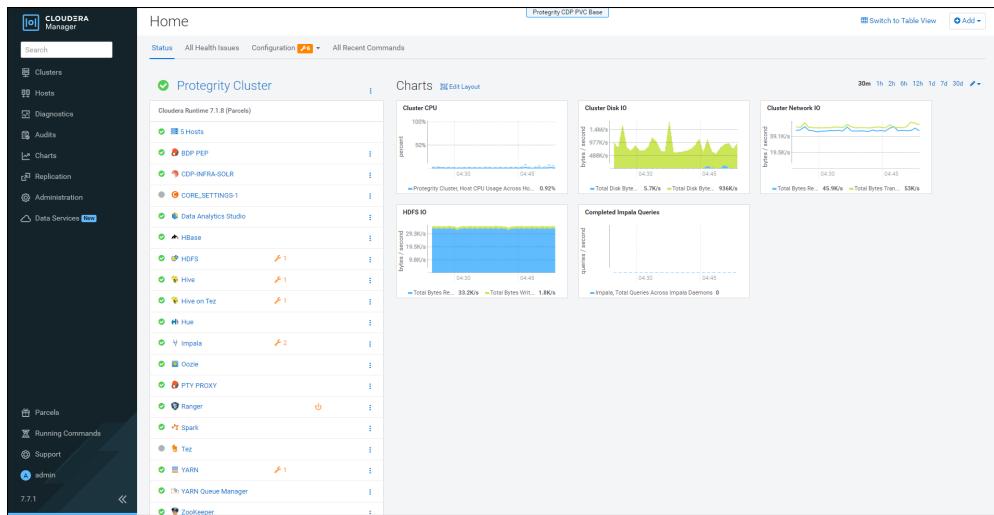


Figure 12-65: PTY Proxy Service Installed

12.6.1.8.3 Configuring the Proxy Service

This section provides information on how to configure the *PTY_Proxy* service in the following scenarios:

- You have not added and started the Big Data Protector PEP service
- You have added and started the Big Data Protector PEP service

- To configure the *PTY_Proxy* service when you have not added and started the Big Data Protector PEP service, perform the following steps.

Note: For more information about starting the BDP PEP service, refer to section [Starting the Big Data Protector PEP Service](#).

- To add the *PTY_Proxy* service and start it, perform the steps as mentioned in section [Starting the Proxy Service](#).
- To enter the IP address or the hostname of the node running the *PTY PROXY* service in the *ESA IP Address* and the *Audit store IP address* prompts instead of the *ESA IP address* or hostname and the *Audit store IP address* or hostname respectively, that appears when you add the *BDP PEP* service to the nodes in the cluster.
- To configure the *PTY_Proxy* service when you have added and started the Big Data Protector PEP service, perform the following steps.
 - Navigate to the **Configurations** page of the BDP PEP service.
 - Replace the ESA IP address or hostname with the IP address or hostname of the host running the *PTY PROXY* service.
 - Replace the Audit Store address or hostname with the IP address or hostname of the host running the *PTY PROXY* service.
 - Restart the *BDP PEP* service.

12.6.1.9 Installing the Big Data Protector Parcels on a New Node

If you want to install the Big Data Protector on a new node in an existing cluster, then you must distribute and activate the following parcels on the new node:

- PTY_BDP*
- PTY_CERT*
- PTY_FLUENTBIT_CONF*

Note: The Cloudera Manager handles the distribution and activation of the Big Data Protector parcels.

Ensure that the *PTY_Pepserver* and the *PTY Log Forwarder* roles, that are part of the *BDP PEP* service, are added to the new node.

Note: For more information about starting the BDP PEP service, refer to section [Starting the Big Data Protector PEP Service](#).

12.6.1.10 Updating the Big Data Protector Configuration Parameters

This section provides information on updating the configuration parameters for the following Big Data Protector services:

- BDP PEP* service
- PTY_Proxy* service

12.6.1.10.1 Updating the Configuration Parameters for the *BDP PEP* Service

You can modify the following PEP server configuration parameters for the *BDP PEP* service:

Table 12-39: Configuration Parameters for the *BDP PEP* Service

| Parameter | Description | Possible Values |
|---|--|---|
| Protegility Audit Store List of Hostnames/IP Addresses and/or Ports | Comma-delimited list of Protegility Audit Store appliance Hostnames/IP addresses and/or Ports where Fluent Bit sends the logs. | <ul style="list-style-type: none"> <code>hostname[:port]</code> <code>[,hostname[:port],hostname[:port]...]</code> - Enter the list of the IP address (in the specified syntax) of the Protegility Audit |



| Parameter | Description | Possible Values |
|-------------------------|---|---|
| | | <p>Store appliance if the Audit Store type is Protegility Audit Store or Protegility Audit Store + External Audit Store. By default, 9200 is set for empty ports.</p> <ul style="list-style-type: none"> NA - Enter this value if you select External Audit Store as the Audit Store type. |
| Fluent Bit Log Level | The Log Forwarder (Fluent Bit) logging verbosity level | <ul style="list-style-type: none"> Error Warning Info Debug |
| ESA hostname/IP Address | Specify the ESA hostname or IP address where the Log Forwarder sends the logs. If the PTY_Proxy service is used, then it should be the host name or the IP address of the PTY_Proxy node. | Host name or IP address |
| Logging Level | Specify the level of details for the PEP server logs. | <ul style="list-style-type: none"> OFF - No logging Severe Warning Info Config ALL (Default) |
| Logging Mode | If the connection to the Fluent Bit is lost, then set how the logs must be handled. This setting is available only for the protector logs and not for the application logs. | <ul style="list-style-type: none"> drop error |

For more information about the PEP server configuration file, refer to section [PEP Server Configuration File](#).

Note: The Shared Memory is world-readable by default. To change the permissions of the Shared memory, add the following Shared Memory management parameters in the *PTY PepServer Advanced Configuration Snippet (Safety Valve)* for *pepper.cfg* parameter:

- sharedmemory.groupname=ptyituser*
- sharedmemory.worldreadable=no*

Ensure that all the service and system users that are trying to access the shared memory are added to the *ptyitusr* group on all the nodes. Else, it may cause an operation failure.

The *ptyitusr* is a sample group used for representational purposes. The user can use their own group name instead.

12.6.1.10.2 Updating the Configuration Parameters for the *PTY_Proxy* Service

You can modify the following configuration parameters for the *PTY_Proxy* service:

Table 12-40: Configuration Parameters for the PTY Proxy Service

| Parameter | Description | Possible Values |
|---------------------------------|--|-------------------------|
| Audit Store hostname/IP Address | Specify the host name or IP address where the Log Forwarder sends logs. If the <i>PTY_Proxy</i> service is used, then it should be the host name or the IP address of the <i>PTY_Proxy</i> node. | Host name or IP address |

| Parameter | Description | Possible Values |
|-------------------------|--|-------------------------|
| ESA hostname/IP Address | Specify the ESA host name or IP address where the Log Forwarder sends logs. If the <i>PTY Proxy</i> service is used, then it should be the host name or the IP address of the <i>PTY Proxy</i> node. | Host name or IP address |

12.6.1.11 Setting the Big Data Protector Configuration for CDP-PVC-Base Distribution

After you install the Big Data Protector, depending on the CDP-PVC-Base services that you will use, you must set the configuration parameters for the Big Data Protector as per the recommendations in the following table:

Table 12-41: Recommended Configuration for Big Data Protector

| Service | Protector Jar Configuration | Protegility Native Library Configuration |
|--|--|--|
| Hive on Tez | <p>In the <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: HIVE_CLASSPATH Value: /opt/cloudera/ parcels/PTY_BDP/pephive/lib/ pephive-*.*jar:/opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:/opt/ cloudera/parcels/PTY_BDP/ bdp_version/:\${HIVE_CLASSPATH}</pre> <p>In the <i>Hive Service Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-site.xml</i>:</p> <pre>Name: hive.exec.pre.hooks Value: com.protegility.hive.PtyHiveUserPreHook</pre> | <p>In the <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: JAVA_LIBRARY_PATH Value: /opt/cloudera/ parcels/PTY_BDP/jpeplite/ lib:\${JAVA_LIBRARY_PATH}</pre> |
| Tez | <pre>Name: tez.cluster.additional.class path.prefix Value: /opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:/opt/ cloudera/parcels/PTY_BDP/ pephive/lib/*</pre> | <pre>Name: tez.am.launch.env = Value: LD_LIBRARY_PATH=/opt/ cloudera/parcels/CDH/lib/ hadoop/lib/native:/opt/ cloudera/parcels/PTY_BDP/ jpeplite/lib</pre> |
| HBase | <pre>Name: hbase.coprocessor.region.clas ses Value: com.protegility.hbase.PTYRegi onObserver</pre> | |
| Spark on Yarn (Spark version >= 2.4.0 and < 3.0.0) | <p>In 'Spark Client Advanced Configuration Snippet (Safety Valve)' for <i>spark-conf/spark-defaults.conf</i>:</p> <pre>spark.executor.plugins=com.p</pre> | |



| Service | Protector Jar Configuration | Protegility Native Library Configuration |
|---------|---|--|
| | rotegility.spark.PtyExecSpark Plugin | |

Important: If you want to uninstall the Big Data Protector, then ensure that you roll back the configuration parameters, to their previous values, that you set after installing the Big Data Protector.

► To configure Impala using the CDP-PVC-Base Native Installer:

1. Ensure that the cluster is installed, configured, and running.
2. To create the `/opt/protegility/impala/udfs` directory in HDFS, run the following command:

```
# sudo -u hdfs hadoop fs -mkdir /opt/protegility/impala/udfs
```

3. To assign Impala supergroup permissions to the `/opt/protegility/impala/udfs` path, run the following command:

```
# sudo -u hdfs hadoop fs -chown -R impala:supergroup /opt/protegility/impala/udfs
```

4. Navigate to the `/opt/cloudera/parcels/PTY_BDP/pepimpala/` directory.
5. To load the `pepimpala.so` file to the `/opt/Protegility/impala/udfs` directory, run the following command:

```
sudo -u hdfs hadoop fs -put pepimpala.so /opt/protegility/impala/udfs
```

In this case, the name of the shared objects file considered as `pepimpala.so`. Typically, the name of the shared objects file is `pepimpala<xx>_RHEL.so`, where <xx> is the version of the file, which needs to be considered.

6. Navigate to the `/opt/cloudera/parcels/PTY_BDP/pepimpala/sqlscripts/` directory.

This directory contains the SQL scripts for defining the Protegility UDFs for the Impala protector.

7. If you are not using a Kerberos-enabled Hadoop cluster, then execute the `createobjects.sql` script to load the Protegility UDFs for the Impala protector.

```
impala-shell -i <IP address of any Impala slave node> -f /opt/cloudera/parcels/PTY_BDP/  
pepimpala/sqlscripts/createobjects.sql
```

8. If you are using a Kerberos-enabled Hadoop cluster, then execute the `createobjects.sql` script to load the Protegility UDFs for the Impala protector.

```
impala-shell -i <IP address of any Impala slave node> -f /opt/cloudera/parcels/PTY_BDP/  
pepimpala/sqlscripts/createobjects.sql -k
```

12.6.1.12 Working with the Certificate Parcels

This section provides information on the following tasks:

- Updating the Certificates Parcel
- Updating the Certificates Parcel without Restarting the PEP server

12.6.1.12.1 Updating the Certificates Parcel

If customers have updated the certificates in the ESA, with which the Big Data Protector is configured, then the Certificates parcel must be updated with the new certificates. The updated Certificates parcel must be utilized by all the nodes in the cluster.

► To utilize the updated certificates:

1. Login to the node, which contains the Big Data Protector configurator script.
2. Run the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* script.
The prompt to continue the configuration of the Big Data Protector appears.

```
*****
        Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP PVC Base

Do you want to continue? [yes or no]:
```

3. To start configuration of the Big Data Protector, type *yes*.
4. Press ENTER.
The prompt to select the type of installation file appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Select the type of Installation files you want to generate.
[ 1: Create All ] : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ] :
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. To update the ESA certificates in the *PTY_CERT* parcel, type *2*.
6. Press ENTER.
The prompt to select the operating system for the parcel appears.

```
Select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el7 ] : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 2: el8 ] : RHEL 8 and clones (CentOS, Scientific Linux, etc)
[ 3: el9 ] : RHEL 9 and clones (CentOS, Scientific Linux, etc)
[ 4: sles12 ] : SuSE Linux Enterprise Server 12.x

Enter the no.:
```

7. Depending on the requirements, type *1*, *2*, *3*, or *4* to select the operating system version for the Big Data Protector parcels.
8. Press ENTER.
The prompt to enter the ESA hostname or IP address appears.

```
Enter the ESA Hostname or IP Address:
```

9. Enter the ESA hostname or IP address.
10. Press ENTER.
The prompt to enter the ESA host listening port appears.

```
Enter ESA host listening port [8443]:
```

11. If you want to use the default value of the ESA host listening port, which is *8443*, then press ENTER.

12. If you have configured an external proxy having connectivity with the ESA to download the certificates and password binaries from the ESA, then enter the external Proxy listening port.
13. Press ENTER.
The prompt to enter the ESA user name appears.

```
Enter ESA Username:
```
14. Enter the ESA user name.
15. Press ENTER.
The prompt to enter the password for the ESA appears.

```
Fetching Certificates from ESA....
```

```
Enter host password for user '<user_name>':
```
16. Enter the ESA administrator password.
17. Press ENTER.
The prompt querying the version of the activated *PTY_CERT* parcel appears.
You can verify the version of the activated *PTY_CERT* parcel from the parcel name, such as *PTY_CERT-x.x.x.x_CDPx.x.p<version>-<OS_Version>.parcel*, where the <version> parameter denotes the patch version of the *PTY_CERT* parcel.
For instance, if the name of the current activated *PTY_CERT* parcel is *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*, then the patch version of the *PTY_CERT* parcel is 0. When specifying the patch version of the parcel, do not include the character *p*.
18. Enter the current activated patch version of the *PTY_CERT* parcel.
19. Press ENTER.
The updated *PTY_CERT* parcel *PTY_CERT-9.1.0.0.x_CDP7.p<updated version>-<OS_Version>.parcel* is generated in *./Installation_Files*/directory.
20. Copy the new Certificate parcel to the local parcel repository of Cloudera Manager.
The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.
21. Navigate to the local parcel repository directory.
In this case, the local parcel repository is stored in the */opt/cloudera/parcel-repo/* directory.
22. To assign the ownership permissions for Cloudera SCM to the new Certificate parcel and checksum file, run the following command:

```
chown cloudera-scm:cloudera-scm PTY_*
```
23. Press ENTER.
24. To set *640* permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```
25. Press ENTER.
The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.
26. Login to the Cloudera Manager web interface.
27. Navigate to the **Parcels** page.
The **Parcels** page appears.
28. To fetch the updated parcels, click **Check for New Parcels**.
Cloudera Manager fetches the updated *PTY_CERT* parcel.

29. Distribute the new Certificate parcel to the nodes.

Note: For more information about distributing the new Certificate parcel, refer to section [Distributing the Big Data Protector Parcels to the Nodes](#).

30. Activate the new Certificate parcel on the nodes.

Note: For more information about activating the new Certificate parcel, refer to section [Activating the Big Data Protector Parcels on the Nodes](#).

12.6.1.12.2 Updating the Certificates Parcel without a Restart

After you update the certificate parcel and distribute them to the nodes, you must restart the dependant services. This restart enables Cloudera Manager to fetch the updated certificate parcel. However, restarting the dependant services results in a loss of production hours. Therefore, Protegility has introduced a feature wherein you can update the certificate parcel without restarting the cluster or the dependant services.

► To update the certificates parcel without restarting the PEP server:

1. Login to the host machine, which contains the Big Data Protector configurator script.
2. Run the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* script.
The prompt to continue the configuration of Big Data Protector appears.
3. To start configuration of Big Data Protector, type *yes*.
4. Press ENTER.
The following prompt appears.

```
Select the type of Installation files you want to generate?
[ 1: Create All ]      : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. To update the ESA certificates in the *PTY_CERT* parcel, type *2*.
6. Press ENTER.
The prompt to select the version of the operating system appears.

```
Select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 2: el8 ]      : RHEL 8 and clones (CentOS, Scientific Linux, etc)
[ 3: el9 ]      : RHEL 9 and clones (CentOS, Scientific Linux, etc)
[ 4: sles12 ]    : SuSE Linux Enterprise Server 12.x

Enter the no.:
```

7. Depending on the requirements, type *1*, *2*, *3*, or *4* to select the OS version for the Big Data Protector parcels.
8. Press ENTER.
The prompt to enter the hostname or the IP address of the ESA appears.

9. Enter the hostname or the IP address of the ESA.
10. Press ENTER.
The prompt to enter the ESA host listening port appears.
11. If you want to use the default value of the ESA host listening port, which is *8443*, then press ENTER.
12. If you have configured an external proxy having connectivity with the ESA to download the certificates and password binaries from the ESA, then enter the external Proxy listening port.
13. Press ENTER.
The prompt to enter the user name for the ESA appears.
14. Enter the ESA user name.
15. Press ENTER.
The prompt to enter the password for the ESA appears.
16. Enter the ESA administrator password.
17. Press ENTER.
A prompt querying the version of the activated *PTY_CERT* parcel appears.
You can verify the version of the activated *PTY_CERT* parcel from the parcel name, such as *PTY_CERT-x.x.x.x_CDPx.x.p<version>-<OS_Version>.parcel*, where the <version> parameter denotes the patch version of the *PTY_CERT* parcel.

For instance, if the name of the current activated *PTY_CERT* parcel is *PTY_CERT-9.1.0.0.x_CDP7.p0-<OS_Version>.parcel*, then the patch version of the *PTY_CERT* parcel is *0*. When specifying the patch version of the parcel, do not include the character *p*.

18. Enter the current activated patch version of the *PTY_CERT* parcel.
19. Press ENTER.
The updated *PTY_CERT* parcel *PTY_CERT-9.1.0.0.x_CDP7.p<updated version>-<OS_Version>.parcel* is generated in the *./Installation_Files*/directory.
20. Copy the new Certificate parcel to the local parcel repository of Cloudera Manager.
The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.
21. Navigate to the local parcel repository directory.
In this case, the local parcel repository is stored in the */opt/cloudera/parcel-repo/* directory.
22. To assign the ownership permissions for Cloudera SCM to the new Certificate parcel and checksum file, run the following command.

```
chown cloudera-scm:cloudera-scm PTY_*
```
23. Press ENTER.
24. To set *640* permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```
25. Press ENTER.
The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.
26. Login to the Cloudera Manager user interface.
27. From the left pane, click **Parcels**.
The **Cloudera Manager Parcels** page appears.

The screenshot shows the Cloudera Manager interface with the 'Parcels' page selected. In the 'Protégility Cluster' section, the 'PTY_CERT' parcel is highlighted with a red border. Other visible parcels include ACCUMULO, CDH 6, COH 5, KAFKA, KEYTRUSTEE_SERVER, KUDU, PTY_BDP, PTY_FLUENTBIT_CONF, SPARKS, SOOOP_NETEZZA_CONNECTOR_2, SOOOP_TERADATA_CONNECT_1, and mlib.

Figure 12-66: Cloudera Manager Parcels Page

Note: The *PTY_FLUENTBIT_CONF* parcel will be visible only if you choose to add the location of the Fluent Bit configuration files while generating the installation files.

- To distribute the Certificates parcel, besides the *PTY_CERT* parcel, click **Distribute**.
Cloudera Manager distributes the Certificates parcel to all the nodes and enables the **Activate** button.

The screenshot shows the Cloudera Manager interface with the 'Parcels' page selected. In the 'Protégility Cluster' section, the 'PTY_CERT' parcel is highlighted with a red border. Other visible parcels include ACCUMULO, CDH 6, COH 5, KAFKA, KEYTRUSTEE_SERVER, KUDU, PTY_BDP, PTY_FLUENTBIT_CONF, SPARKS, SOOOP_NETEZZA_CONNECTOR_2, SOOOP_TERADATA_CONNECT_1, and mlib.

Figure 12-67: Distribution of the Updated Certificates Parcel

- To activate the certificates parcel without a restart, besides the *PTY_CERT* parcel, click **Activate**.
The prompt to activate the certificates parcel appears.

The dialog box contains the following text:
Activate PTY_CERT 9.1.0.0.28_CDP7.p1 on Protégility Cluster

These services need to be restarted.
BDP PEP

 Restart
 Activate Only

Cancel OK

Figure 12-68: Prompt to Activate the Certificates Parcel

- Select **Activate Only**.

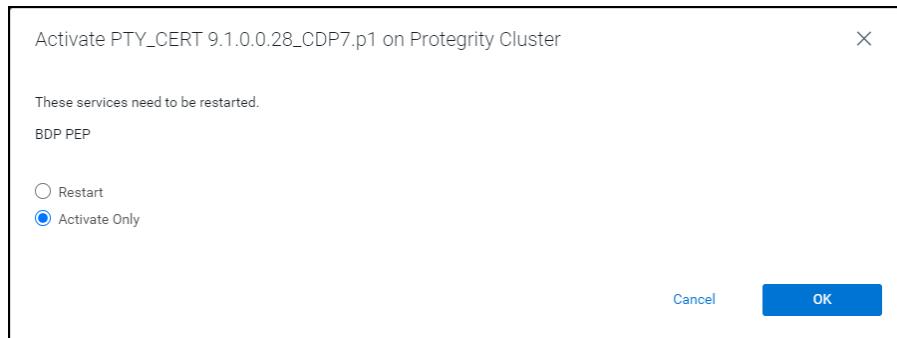


Figure 12-69: Prompt to Only Activate the Certificates Parcel

31. Click OK.

Cloudera Manager deactivates the existing certificates parcel from all the nodes and activates the updated certificates parcel on all the nodes.

After the activation is complete, Cloudera Manager enables the **Deactivate** option for the *PTY_CERT* parcel.

| Parcel Name | Version | Status | Action |
|---------------------------|--|------------------------|----------------------------|
| ACCUMULO | 1.7.0.6.5.0-CDH-0.5.0-p0.8 | Available Remotely | Download |
| Cloudera Runtime | 7.1.0-1.0#7.1.0.p0.30990532 | Distributed, Activated | Deactivate |
| CDH 6 | 4.3.1.0#6.3.4.p0.6751058 | Available Remotely | Download |
| CDH 5 | 5.16.2.1.0#5.16.2.p0.8 | Available Remotely | Download |
| KAFKA | 4.1.0.1.4.1.0#0.4 | Available Remotely | Download |
| KEYTRUSTEE_SERVER | 7.1.0.0.0#keytrustee/7.1.0.p0.30990532 | Available Remotely | Download |
| KUDU | 1.4.0.1.0#5.1.3.2.p0.8 | Available Remotely | Download |
| PTY_BDP | 1 | Available Remotely | Download |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| SPARKS | 1 | Available Remotely | Download |
| SOODOP_NETEZZA_CONNECTOR | 3.0.0.3.3.7180.0.274.1.p0.31212967 | Available Remotely | Download |
| SOODOP_TERADATA_CONNECTOR | 1.5.16 | Available Remotely | Download |
| SOODOP_TERADATA_CONNECTOR | 1.7.0.5 | Available Remotely | Download |
| mail | 2023.1.0.46342 | Available Remotely | Download |

Figure 12-70: Updated Certificates Parcel Activated

32. Navigate to the Cloudera Manager home page.

The Cloudera Manager home page indicates a stale configuration in the BDP PEP service because we activated the updated certificates parcel without a restart.

Figure 12-71: Stale Configuration in the BDP PEP Service

Note: In this method, you can safely ignore the stale configuration alert because the update certificate feature does not require a restart of the BDP PEP service.

33. To view the service page for PEP server, click **BDP PEP**.
The **BPD PEP** page appears.

The screenshot shows the Protegility Cluster interface with the BDP PEP service selected. The left sidebar includes options like Clusters, Hosts, Audits, Charts, Replication, Administration, and Data Services. The main area displays a 'Status Summary' section with three items: PTY PepServer (Good Health), PTY Log Forwarder (Good Health), and Hosts (Good Health). Below this is a 'Health History' section stating 'No results found.' To the right, there are two charts: 'Informational Events' and 'Important Events and Alerts', both showing data from 04:00 to 04:15. A top navigation bar includes 'Actions' with a power icon, 'Status', 'Instances', 'Configuration', 'Commands', 'Charts Library', 'Audits', and 'Quick Links'. A timestamp at the top right indicates '30 minutes preceding Sep 12, 4:17 AM EDT'.

Figure 12-72: BDP PEP Page

34. To update the certificates parcel on all the nodes, select **Actions > Rotate certificates for all PepServers**.

The screenshot shows the same Protegility Cluster interface as Figure 12-72, but the 'Actions' dropdown menu is open. The menu items include Start, Restart, Rotate certificates for all PepServers (which is highlighted with a blue border), Stop, Add Role Instances, Rename, Delete, and Enter Maintenance Mode. The rest of the page remains consistent with Figure 12-72, showing the 'Status Summary' and 'Health History' sections.

Figure 12-73: Rotate Certificates for all PepServers

The prompt to confirm the action appears.

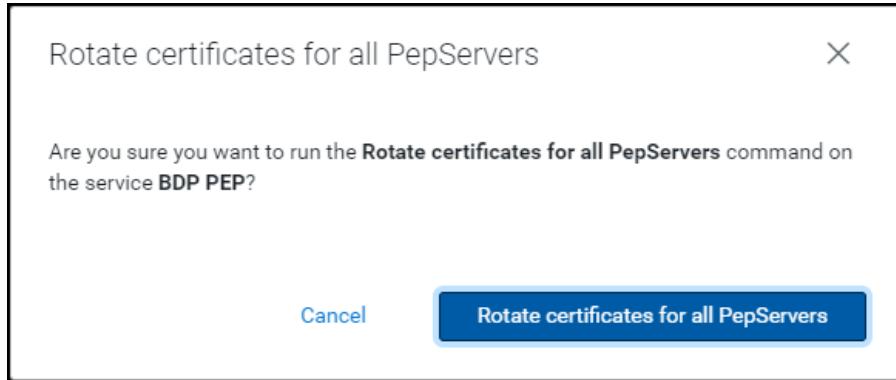


Figure 12-74: Rotate Certificates for all PEP Servers

35. Click **Rotate certificates for all PepServers**.

Cloudera Manager executes the rotate certificate command and updates the certificates parcel on all the nodes on the cluster.

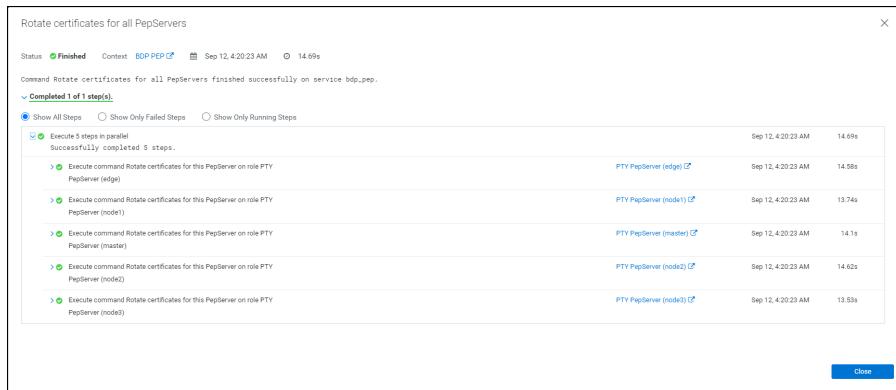


Figure 12-75: Rotate Certificate Command to Update the Certificates Parcel

36. Click **Close**.

- The command extracts the certificates from the latest activated *PTY_CERT* parcel directory (*/opt/cloudera/parcels/PTY_CERT/data/esacerts.tar*) to the default PEP server directory (*/opt/cloudera/parcels/PTY_BDP/defiance_data/data/*) on each node.
- The PEP server will establish a TLS connection, download the policy, and fetch the certificates from the */defiance_dps/data/* directory every time it polls the ESA. This eliminates the need to restart the PEP server to fetch the updated certificates.

Note: The *BDP PEP* service in Cloudera Manager will fetch the updated certificates (*PTY_CERT*) parcel on the new node whenever you add a new node to the existing cluster.

12.6.1.13 Updating the Fluent Bit Parcel

If you want to use a newer set of custom Fluent Bit configuration files to send the logs to an External Audit Store, then you must update, distribute, and activate the Fluent Bit parcel on all the nodes in the cluster.

► To update the Fluent Bit parcel:

- Login to the host machine, which contains the Big Data Protector configurator script.
- Run the *BDPConfigurator_CDP-PVC-Base-7_9.1.0.0.x.sh* script.

The prompt to continue the configuration of Big Data Protector appears.

```
*****
        Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP PVC Base

Do you want to continue? [yes or no]:
```

3. To start configuration of the Big Data Protector, type *yes*.
4. Press ENTER.

The prompt to select the type of installation file appears.

```
Select the type of Installation files you want to generate.
[ 1: Create All ]      : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                         Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                         : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                         Use this if you want to set Custom Fluent-Bit configuration
files to
                         forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. To update the Fluent Bit parcel, type *3*.
6. Press ENTER.

The prompt to select the operating system version appears.

```
Select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 2: el8 ]      : RHEL 8 and clones (CentOS, Scientific Linux, etc)
[ 3: el9 ]      : RHEL 9 and clones (CentOS, Scientific Linux, etc)
[ 4: sles12 ]    : SuSE Linux Enterprise Server 12.x

Enter the no.:
```

7. Depending on the requirements, type *1*, *2*, *3*, or *4* to select the operating system version for the Big Data Protector parcels.
8. Press ENTER.

The prompt to enter the local directory path that stores the Fluent Bit configuration files appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration
files for External Audit Store:
```

9. Type the local directory path that stores the Fluent Bit configuration files.
10. Press ENTER.

The prompt to enter the current version of the Fluent Bit configuration parcel appears.

```
Generating Installation files...
```

NOTE:

You can verify the version of the activated PTY_FLUENTBIT_CONF parcel from the parcel name, such as PTY_FLUENTBIT_CONF-x.x.x.x.p<version>-<os>.parcel, where the <version> parameter denotes the patch version of the PTY_FLUENTBIT_CONF parcel.

For Example: If the current activated PTY_FLUENTBIT_CONF parcel is PTY_FLUENTBIT_CONF-x.x.x.x.CDPx.x.p0-<os>.parcel, the patch version of the PTY_FLUENTBIT_CONF parcel will be 0. Do NOT include 'p' while specifying the version.

Enter the <version> of the current PTY_FLUENTBIT_CONF Parcel as specified in the parcel name [0]:

11. Type the version of the Fluent Bit configuration parcel.

12. Press ENTER.

The installer generates the *PTY_FLUENTBIT_CONF* parcel in the *./Installation_Files/* directory.

The updated PTY_FLUENTBIT_CONF parcel 'PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p1-<OS_Version>.parcel' is generated in *./Installation_Files/* directory.

NOTE:

Copy PTY_FLUENTBIT_CONF-9.1.0.0.x_CDP7.p1-<OS_Version>.parcel and .sha files to Cloudera Manager local parcel repository.

13. Copy the new *PTY_FLUENTBIT_CONF* parcel to the local parcel repository of Cloudera Manager.

The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.

14. Navigate to the local parcel repository directory.

In this case, the local parcel repository is stored in the */opt/cloudera/parcel-repo/* directory.

15. To assign the ownership permissions for the Cloudera SCM to the new Fluent Bit configuration parcel and checksum file, run the following command:

```
chown cloudera-scm:cloudera-scm PTY_*
```

16. Press ENTER.

17. To assign 640 permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```

18. Press ENTER.

The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.

19. Login to the Cloudera Manager web interface.

20. Navigate to the **Parcels** page.

The **Parcels** page appears.

21. To fetch the updated parcels, click **Check for New Parcels**.

The Cloudera Manager will fetch the updated *PTY_FLUENTBIT_CONF* parcel.

22. Distribute the new *PTY_FLUENTBIT_CONF* parcel to the nodes.

Note: For more information about distributing the new *PTY_FLUENTBIT_CONF* parcel, refer to section [Distributing the Big Data Protector Parcels to the Nodes](#).

23. Activate the new *PTY_FLUENTBIT_CONF* parcel on the nodes.

Note: For more information about activating the new *PTY_FLUENTBIT_CONF* parcel, refer to section [Activating the Big Data Protector Parcels on the Nodes](#).

12.6.1.14 Enabling the PEP Server Application Log File

A new feature is introduced to enable the application log file for the PEP server. You can access this feature using the BDP PEP configuration page in Cloudera Manager. In addition to enabling the application log file for the PEP server, you can also set the following parameters:

- **PepServer Application Log File Directory Path** - specifies the location to create the PEP server application log file
- **PepServer Application Log File Name** - specifies the name for the PEP server application log file

► To enable the PEP Server Application Log file:



1. Using a browser, navigate to the Cloudera Manager page.

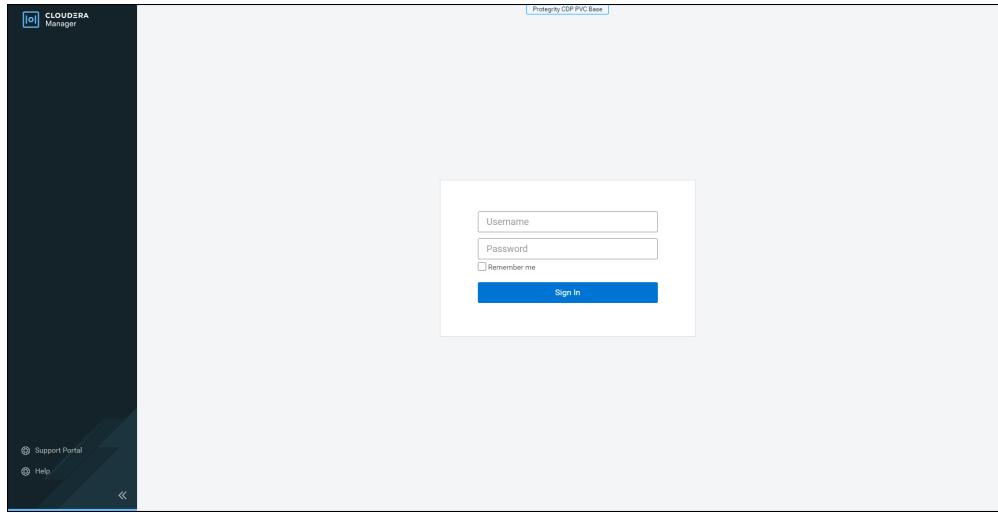


Figure 12-76: Cloudera Manager Login Page

2. Enter the **Username**.
3. Enter the **Password**.
4. Click **Sign In**.

The Cloudera Manager **Home** page appears.

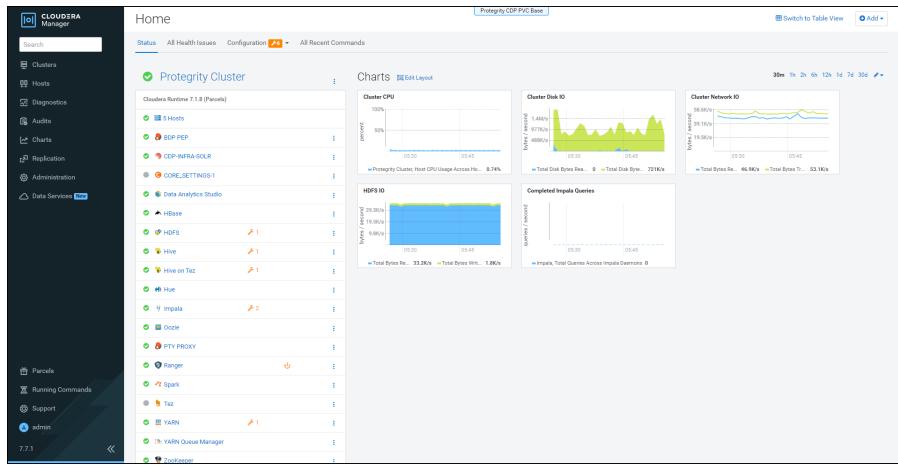


Figure 12-77: Cloudera Manager Home Page

5. Under the cluster name, click **BDP PEP**.
The **BDP PEP** page appears.

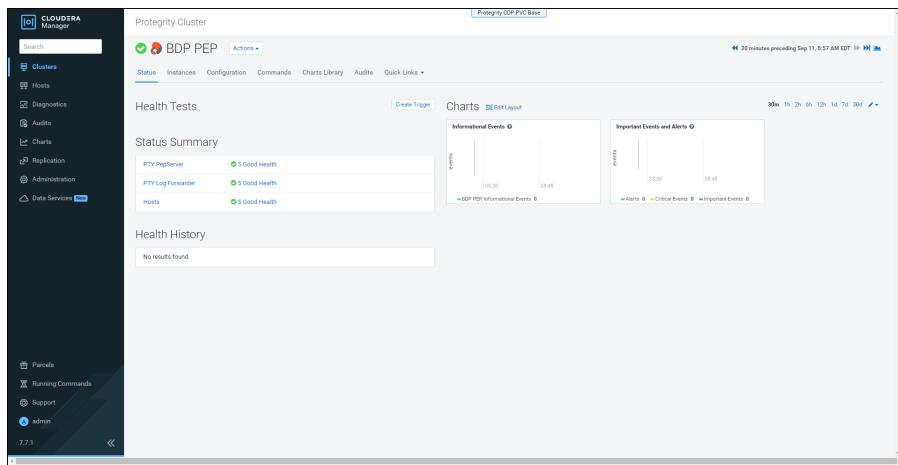


Figure 12-78: BDP PEP Page

- To view and modify the parameters, click the **Configuration** tab.
The **Configuration** tab appears.

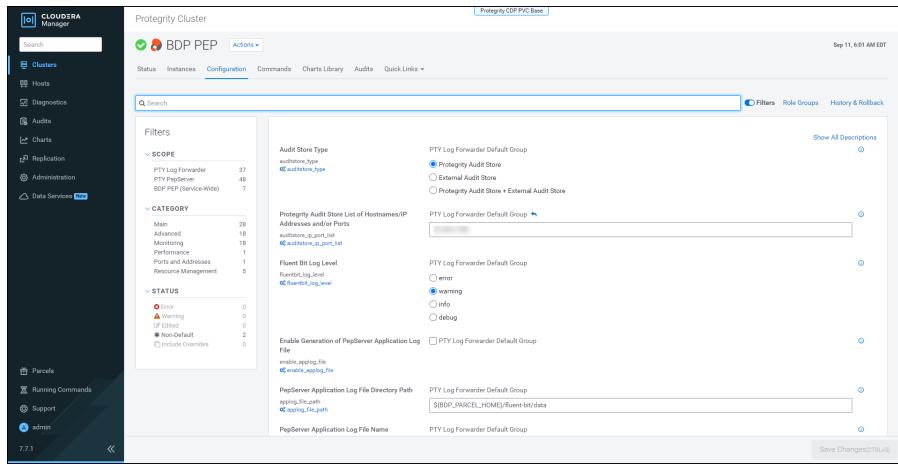


Figure 12-79: Configuration Tab for the BDP PEP Service

- To enable the PEP server application log file, besides **Enable Generation of PepServer Application Log File**, select the checkbox.
 - To specify the location where you want to generate the application log file, under **PepServer Application Log File Directory Path**, set the required location.
 - To specify a name for the PEP server application log file, under **PepServer Application Log File Name**, enter the required name.
- After you modify the value for any parameter, Cloudera enables the **Save Changes(CTRL+S)** button.

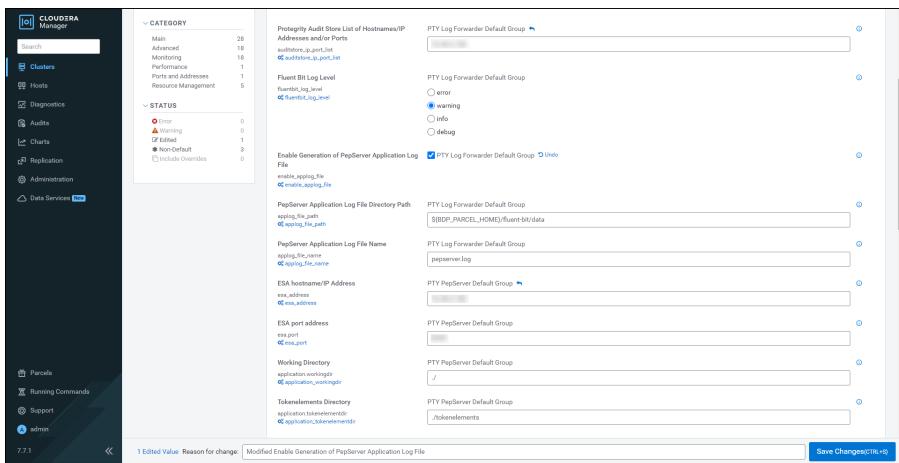


Figure 12-80: Enabling the PEP Server Application Log File

- To persist the changes, click **Save Changes(CTRL+S)**.

Cloudera saves the configuration changes and a confirmation appears.

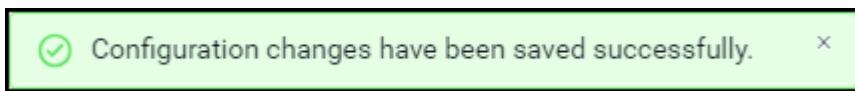


Figure 12-81: Confirmation for Configuration Updates

- To persist the new changes, restart the *BDP PEP* service.

Caution:

- The log entries written to *pepper.log* file can fill up the disk over a period of time.
- It is recommended to monitor the *pepper.log* file and rotate the log entries depending on your system configuration. Protegility does not support the rotation of log entries in the *pepper.log* file.

12.6.1.15 Installing the Big Data Protector after Upgrading to CDP-PVC-Base Distribution

Cloudera enables the following upgrade paths:

- Upgrading from CDH to CDP-PVC-Base
- Upgrading from HDP to CDP-PVC-Base

Caution: If you are upgrading to CDP-PVC-Base distribution from CDH or HDP, then you must first uninstall the Big Data Protector before starting the upgrade. After the upgrade of the CDP-PVC-Base distribution is complete, you must install the Big Data Protector version that is compatible with the updated version of the CDP-PVC-Base distribution.

12.6.1.15.1 Installing the Big Data Protector after Upgrading from CDH to CDP-PVCBase

Cloudera supports upgrading from CDH version 5.13.x and higher to CDP-PVC-Base version 7.1.x.

Caution: If you are upgrading to CDP-PVC-Base distribution from CDH or HDP, then you must first uninstall the Big Data Protector before starting the upgrade. After the upgrade of the CDP-PVC-Base distribution is complete, you must install the Big Data Protector version that is compatible with the updated version of the CDP-PVC-Base distribution.

► To upgrade from CDH Versions 5.13.x onwards to CDP-PVC-Base, version 7.1.x:

1. Note all the custom *BDP PEP* configurations because the customizations will be lost after you uninstall the Big Data Protector.
 2. Stop the following services:
 - *BDP PEP*
 - *PTY_PROXY*

Note: For more information about stopping the services, refer to the section [Stopping and Removing the Big Data Protector Services from all the Nodes](#).

3. Remove the following services:
 - *BDP PEP*
 - *PTY_PROXY*

Note: For more information about removing the services, refer to the section [Stopping and Removing the Big Data Protector Services from all the Nodes](#).

4. Deactivate the following parcels:
 - *PTY_BDP*
 - *PTY_CERT*
 - *PTY_FLUENTBIT_CONF*

Note: For more information about deactivating the parcels, refer to the section [Deactivating the Big Data Protector Parcels from all the Nodes](#).

5. Remove the following parcels:
 - *PTY_BDP*
 - *PTY_CERT*
 - *PTY_FLUENTBIT_CONF*

Note: For more information about removing the parcels, refer to the section [Removing the Big Data Protector Parcels from all the Nodes](#).

6. Upgrade the Cloudera Manager from version 5.13.x to Cloudera Manager version 7.x in CDP-PVC-Base.

Note: For more information about upgrading the Cloudera Manager, refer to https://docs.cloudera.com/cdp/latest/upgrade-cdh/topics/ug_cm_upgrade_before.html.

7. Upgrade the CDH clusters from versions 5.13.x onwards to Cloudera Runtime version 7.1.x in CDP-PVC-Base.

Note: For more information about upgrading Cloudera Runtime, refer to https://docs.cloudera.com/cdp/latest/upgrade-cdh/topics/ug_cdh_upgrading_top.html.

8. Upgrade the Big Data Protector parcels to the supported CDP-PVC-Base distribution.

Note: For more information about upgrading the Big Data Protector parcels, refer to section [Installing the Big Data Protector using CDP Private Cloud Base \(CDP-PVC-Base\) Native Installer](#).

9. Distribute the following parcels to all the nodes in the cluster:
 - *PTY_BDP*
 - *PTY_CERT*



- *PTY_FLUENTBIT_CONF*

Note: For more information about distributing the Big Data Protector parcels to the nodes, refer to the section [Distributing the Big Data Protector Parcels to the Nodes](#).

10. Activate the following parcels to all the nodes in the cluster:

- *PTY_BDP*
- *PTY_CERT*
- *PTY_FLUENTBIT_CONF*

Note: For more information about activating the Big Data Protector parcels on all the nodes, refer to the section [Activating the Big Data Protector Parcels on the Nodes](#).

11. Start the *BDP PEP* and *PTY_PROXY* services.

Note: For more information about starting the Big Data Protector services on the nodes, refer to the section [Starting the Big Data Protector Services](#).

12. Restore custom PEP service configurations manually, if any.

12.6.1.15.2 Installing the Big Data Protector after Upgrading from HDP to CDP-PVC-Base

You can upgrade from HDP version 2.6.5 and higher and Ambari version 2.6.2.2 and higher to CDP-PVC-Base version 7.1.x.

Caution: If you are upgrading to CDP-PVC-Base distribution from CDH or HDP, then you must first uninstall the Big Data Protector before starting the upgrade. After the upgrade of the CDP-PVC-Base distribution is complete, you must install the Big Data Protector version that is compatible with the updated version of the CDP-PVC-Base distribution.

► To upgrade from HDP version 2.6.5 and higher and Ambari version 2.6.2.2 and higher to CDP-PVC-Base version 7.1.x:

1. Note all the custom *BDP PEP* configurations as customization could be lost after uninstalling the Big Data Protector.
2. Uninstall the Big Data Protector from HDP version 2.6.5.

Note: For more information about uninstalling the Big Data Protector from HDP, refer to section [Uninstalling the Big Data Protector Services from the Nodes](#).

3. Upgrade HDP to CDP-PVC-Base.

Note: For more information about upgrading HDP to CDP-PVC-Base, refer to <https://docs.cloudera.com/cdp/latest/upgrade/topics/cdpdc-hdp-overview.html>.

4. Install Big Data Protector on CDP-PVC-Base.

Note: For more information about installing Big Data Protector on CDP-PVC-Base, refer to section [Installing the Big Data Protector using CDP Private Cloud Base \(CDP-PVC-Base\) Native Installer](#).

5. Restore custom PEP service configurations manually, if any.

12.6.1.16 Performing an Upgrade of the CDP-PVC-Base Distribution

If you are performing a rolling upgrade of the CDP-PVC-Base distribution, then you must uninstall the Big Data Protector before starting the rolling upgrade. After the rolling upgrade of the CDP-PVC-Base distribution is complete, you must install the Big Data Protector version that is compatible with the updated version of the CDP-PVC-Base distribution. If you are using CDP-PVC-Base versions 7.1 and lower and need to upgrade the version of CDP-PVC-Base, then perform the following steps.

► To perform a rolling upgrade of minor version of CDP-PVC-Base:

1. Ensure that you do not invoke the Protegity APIs or UDFs or the Protegity HBase coprocessor will fail during the rolling upgrade to a minor version of CDP-PVC-Base.
2. Perform the rolling restart of the stale services to deploy the recommended service configurations.
3. Perform the rolling upgrade of the required CDP-PVC-Base minor version.
4. Perform the rolling restart of the stale services to deploy the recommended service configurations.

12.6.1.17 Uninstalling the Big Data Protector

The process of uninstalling the Big Data Protector involves the following steps:

1. Removing the Big Data Protector-related services from all the nodes in the cluster
2. Deactivating the Big Data Protector parcels from all the nodes in the cluster.
3. Removing the Big Data Protector parcels from all the nodes in the cluster.
4. Deleting the Big Data Protector parcels from the local repository of Cloudera Manager.
5. Deleting the *.jar* files from the local repository of Cloudera Manager.

12.6.1.17.1 Stopping and Removing the Big Data Protector Services from all the Nodes

Before you deactivate the Big Data Protector parcels from all the nodes in the cluster, you must stop and remove the Big Data Protector-related services from all the nodes.

► To stop and remove the all the Big Data Protector related services from all the nodes in the cluster:

1. To stop and remove the *BDP PEP* service, perform the following steps.
 - a. On the Cloudera Manager Home page, besides the BDP PEP service, click the kebab menu . The **BDP PEP Actions** drop-down menu appears.

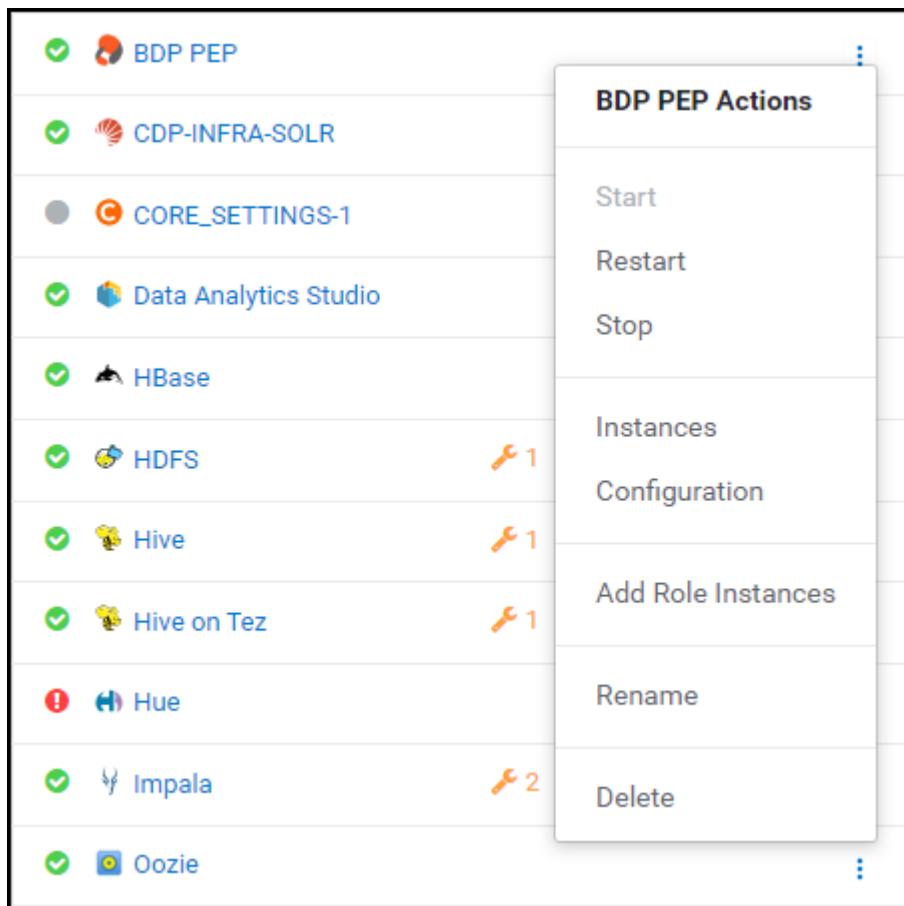


Figure 12-82: BDP PEP Actions drop-down Menu

- b. Select **Stop**.

The prompt to confirm the termination of the *BDP PEP* service appears.

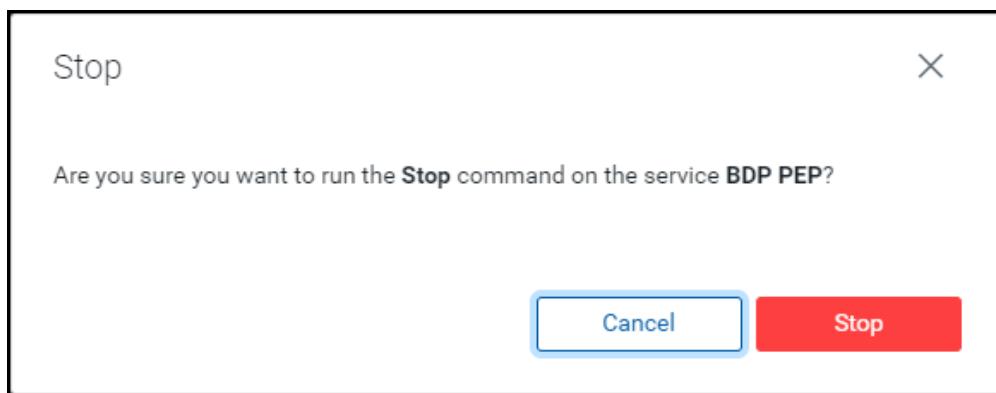


Figure 12-83: Prompt to Stop the BDP PEP Service

- c. Click **Stop**.

The BDP PEP service is terminated and the following page appears.

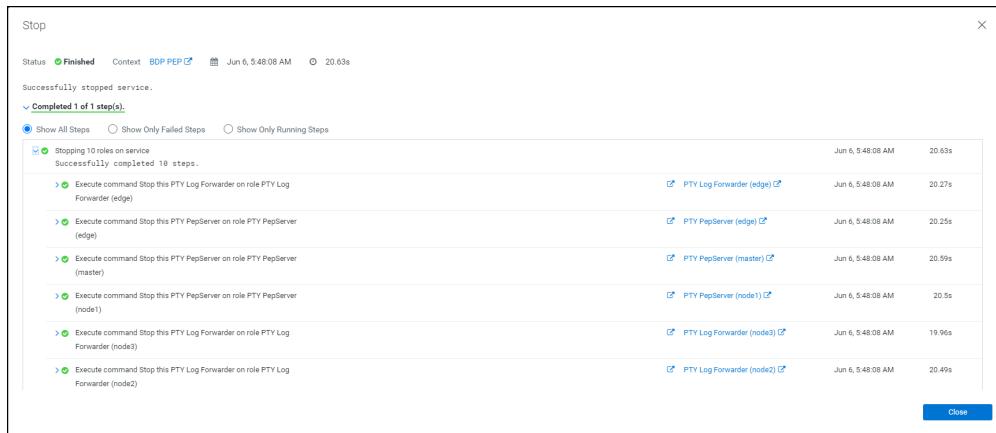


Figure 12-84: BDP PEP Service Stopped

d. Click **Close**.

The BDP PEP service is stopped and the status is updated on the Home page of the Cloudera Manager.

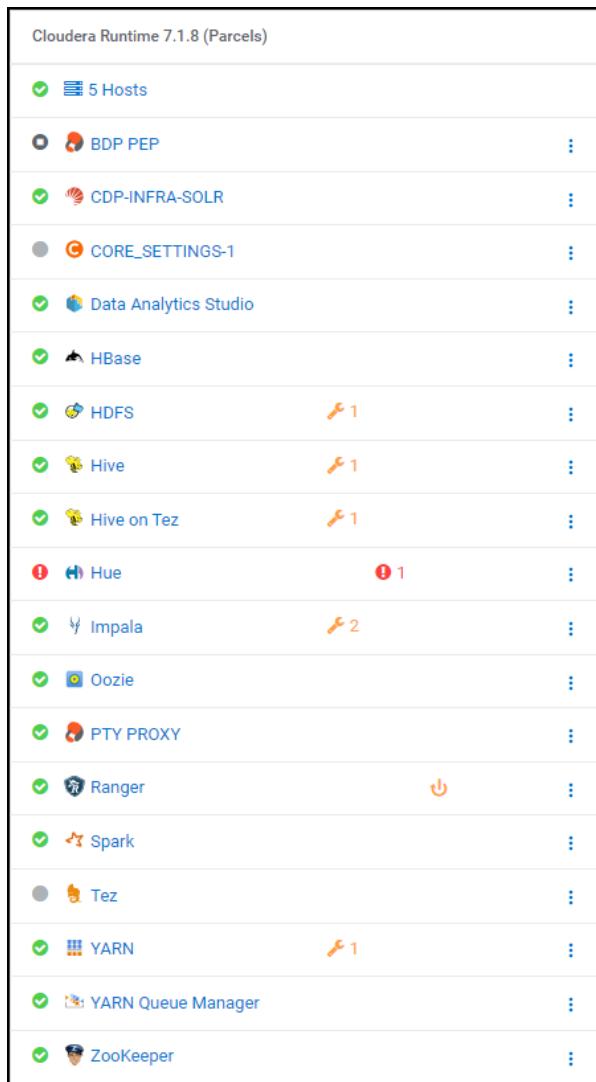


Figure 12-85: BDP PEP Service Stopped in the Cluster

- e. Besides the BDP PEP service, click the kebab menu . The **BDP PEP Actions** drop-down list appears.

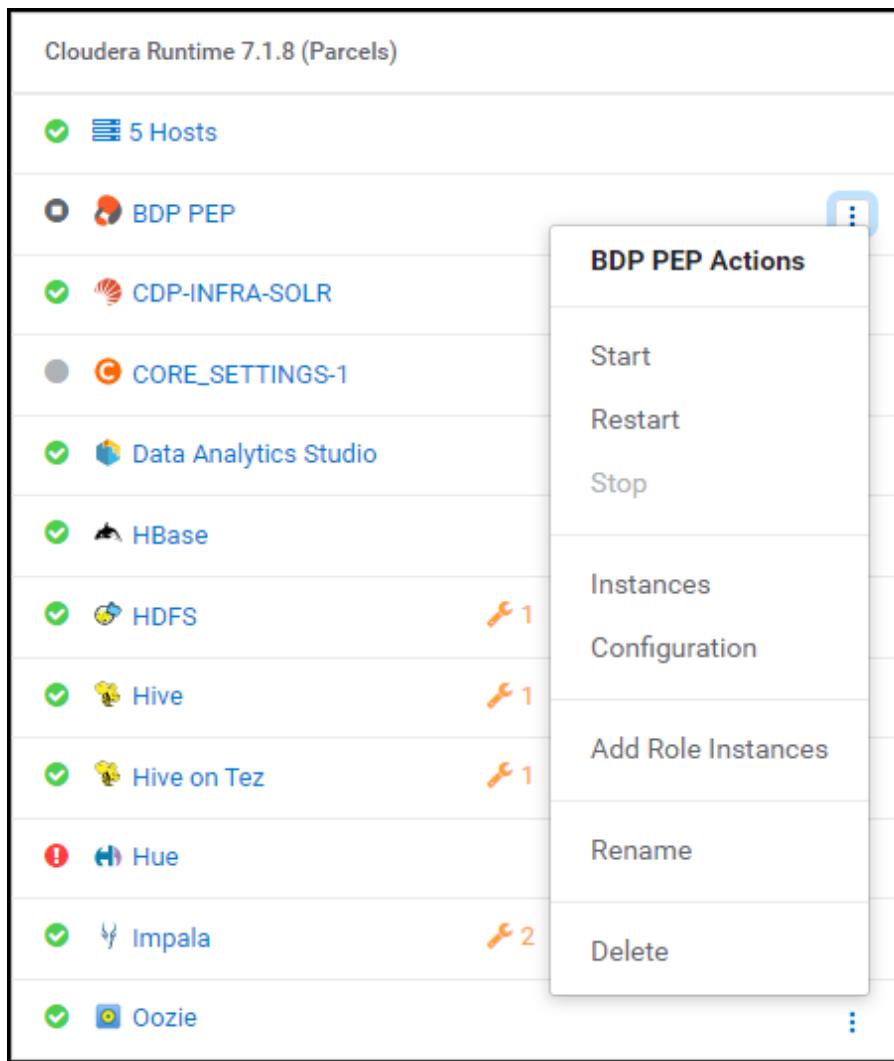


Figure 12-86: BDP PEP Actions drop-down list

f. Select **Delete**.

The prompt to confirm the deletion of the BDP PEP service appears.

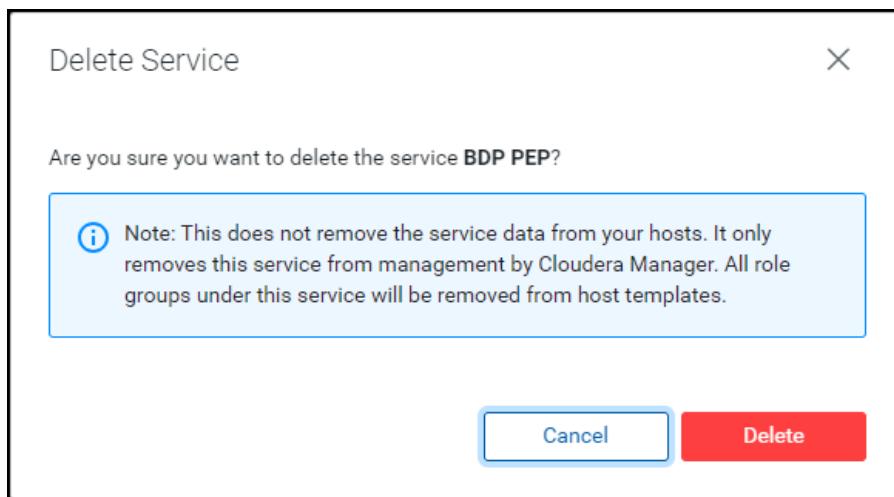


Figure 12-87: Prompt to Delete the BDP PEP Service

g. Click **Delete**.

The BDP PEP service is removed from all the nodes in the cluster.

2. To stop and remove the *PTY_PROXY* service, perform the following steps.

- a. Besides the **PTY PROXY** service, click the kebab menu . The **PTY PROXY Actions** drop-down menu appears.

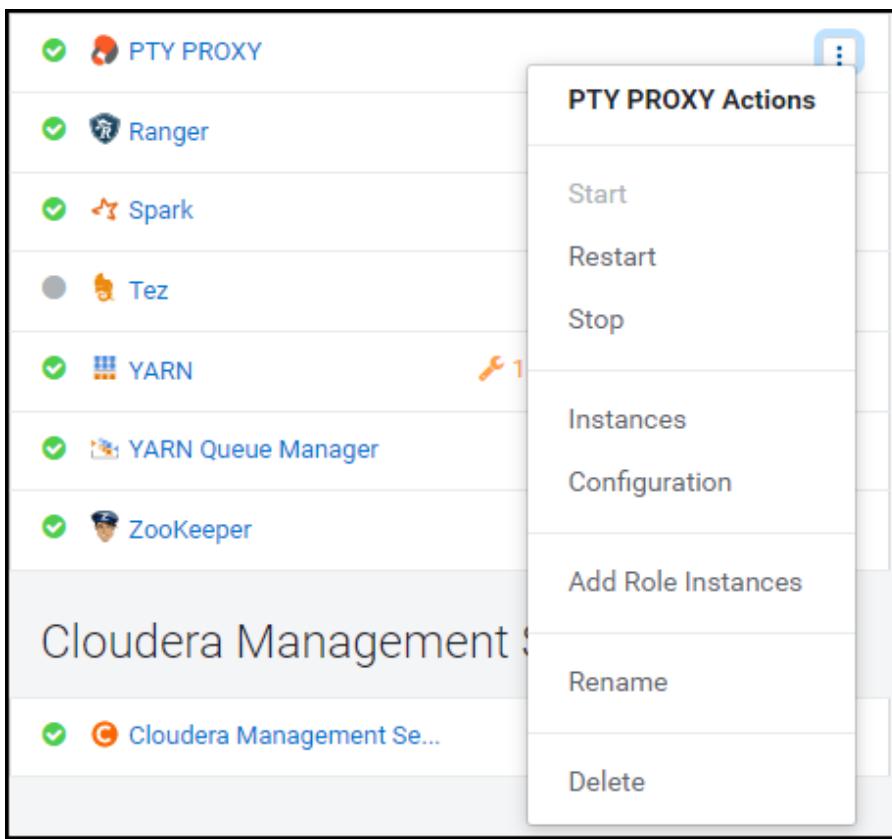


Figure 12-88: PTY PROXY Actions drop-down Menu

- b. Select **Stop**.
The prompt to confirm the termination of the **PTY PROXY** service appears.

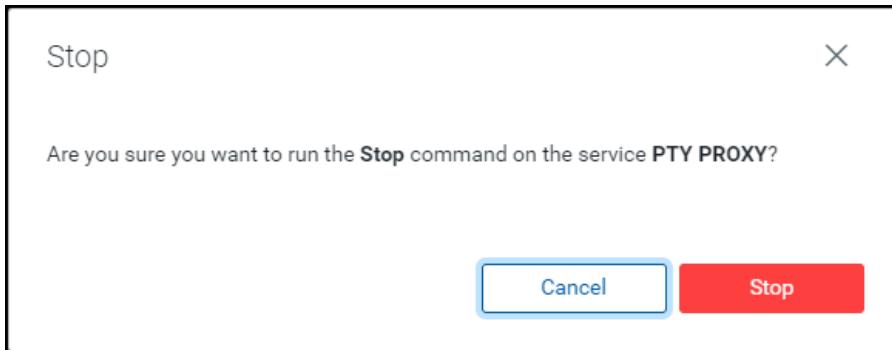


Figure 12-89: Prompt to Stop the PTY PROXY Service

- c. Click **Stop**.
The **PTY PROXY** service is terminated.
d. Click **Close**.
The **PTY PROXY** service is stopped and the status is updated.
e. Besides the **PTY PROXY** service, click . The **PTY PROXY Actions** drop-down menu appears.

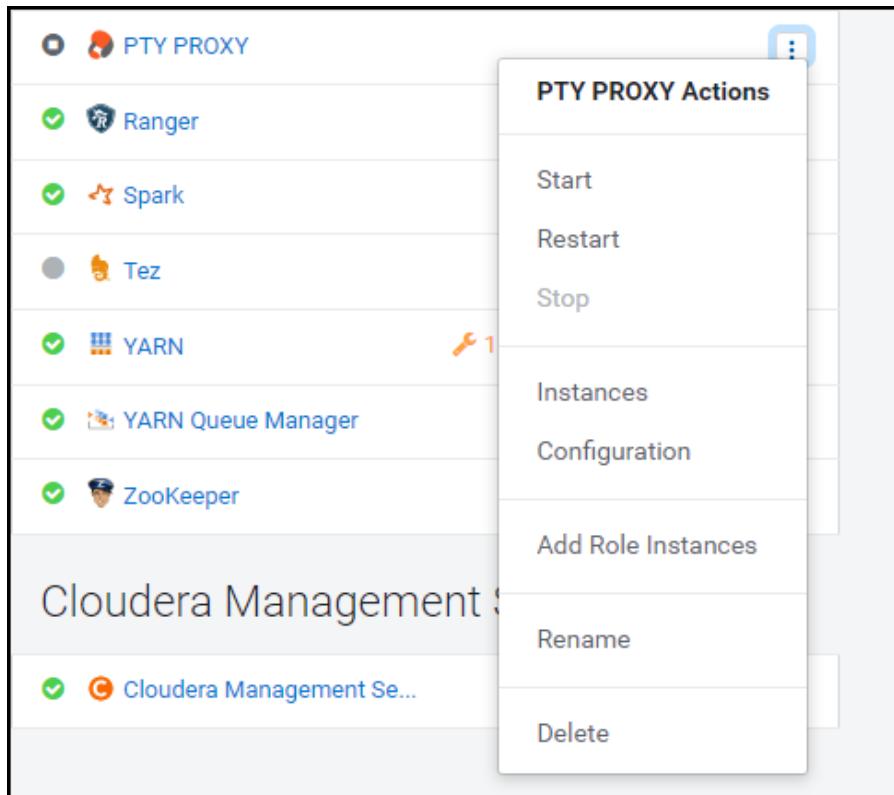


Figure 12-90: PTY PROXY Actions drop-down Menu

- f. Select **Delete**.

The prompt to confirm the deletion of the *PTY PROXY* service appears.

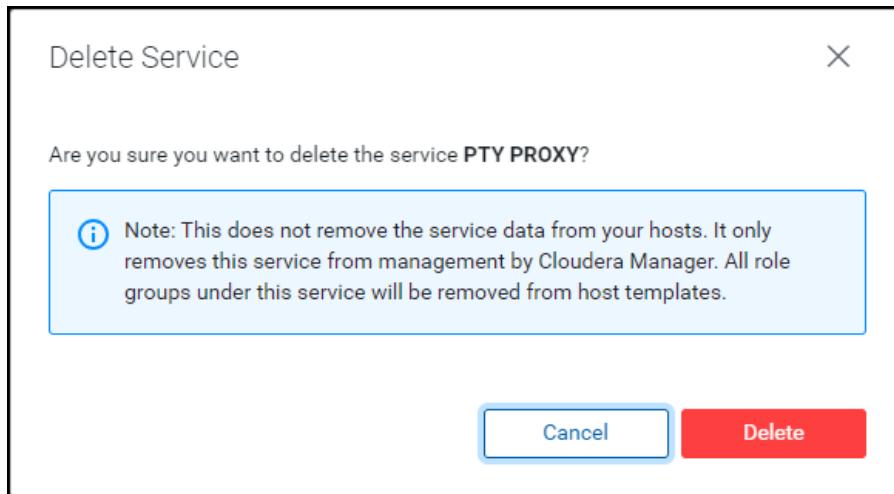


Figure 12-91: Prompt to Delete the PTY_PROXY Service

3. Click **Delete**.

The *PTY PROXY* service is removed from all the nodes in the cluster.

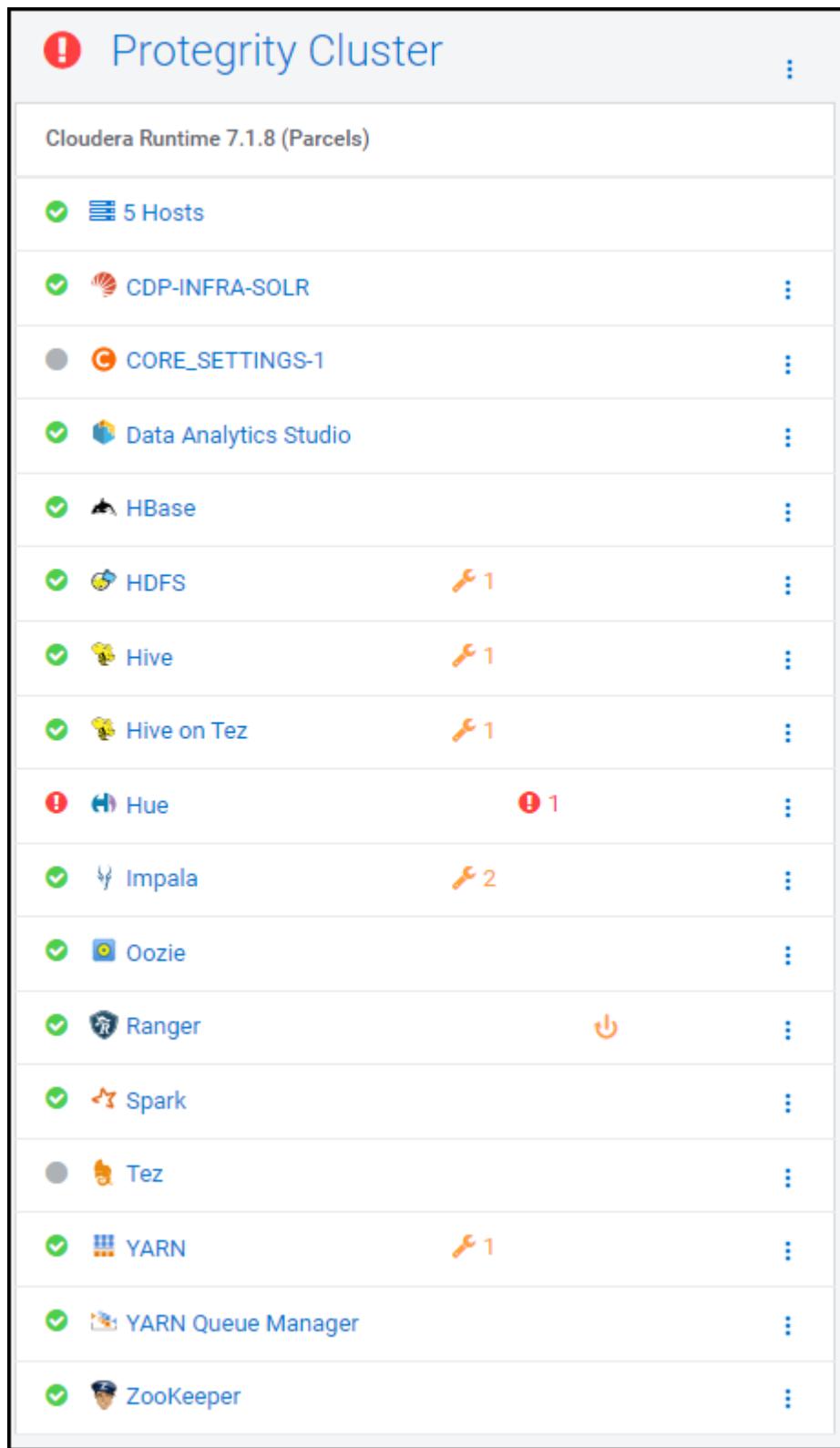


Figure 12-92: Big Data Protector Services Removed from the Nodes in the Cluster

12.6.1.17.2 Deactivating the Big Data Protector Parcels from all the Nodes

After you remove the Big Data Protector-related services from all the nodes in the cluster, you must deactivate the Big Data Protector parcels from all the nodes.

► To deactivate the Big Data Protector Parcels from all Nodes in the Cluster:

1. On the Cloudera Manager home page, click **Parcels**.
The **Parcels** page appears.

| Parcel Name | Version | Status | Action |
|---------------------------|---------------------------------------|-------------------------------|-------------------|
| ACUMULO | 1.7.2.5.6.0.ACUMULO.05.0.p0.8 | Available Remotely | Download |
| Cloudera Runtime | 7.1.8-1.cdp7.1.8.p0.30990532 | Distributed, Activated | Deactivate |
| CDH 6 | 6.3.4-1.cdh6.3.4.p0.6751098 | Available Remotely | Download |
| CDH 5 | 5.16.1-1.cdh5.5.2.p0.8 | Available Remotely | Download |
| KAFKA | 4.1.0-1.4.1.0.p0.4 | Available Remotely | Download |
| KEYTRUSTEE_SERVER | 7.1.8-0.1.keytrustee7.1.8.p0.30990532 | Available Remotely | Download |
| KUDU | 1.4.0-1.cdh5.12.2.p0.8 | Available Remotely | Download |
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| SPARK3 | 3.3.0-3.7180.0.274-1.p0.31212467 | Available Remotely | Download |
| SOOOP_NETEZZA_CONNECTOR | 1.5.1c6 | Available Remotely | Download |
| SOOOP_TERADATA_CONNECTOR | 1.7c6 | Available Remotely | Download |
| mkl | 2023.1.46342 | Available Remotely | Download |

Figure 12-93: Cloudera Manager Parcels Page

The following Protegility parcels appear on the Parcels page:

- **PTY_BDP**: Big Data Protector parcel
- **PTY_CERT**: Certificates parcel
- **PTY_FLUENTBIT_CONF**: Fluent Bit configuration parcel

Note: The **PTY_FLUENTBIT_CONF** Fluent Bit configuration parcel will be visible only if you have selected it during installation.

2. To deactivate the Fluent Bit configuration parcel, besides the **PTY_FLUENTBIT_CONF** parcel, click **Deactivate**.

| | | | |
|--------------------|--------------------|------------------------|------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | Deactivate |

Figure 12-94: Protegility Parcels Activated

The prompt to confirm the deactivation of the parcel appears.

Deactivate PTY_FLUENTBIT_CONF 9.1.0.0.28_CDP7.p0 on Protegility Cluster

Are you sure?

Cancel OK

Figure 12-95: Prompt to Deactivate the Fluent Bit Configuration Parcel

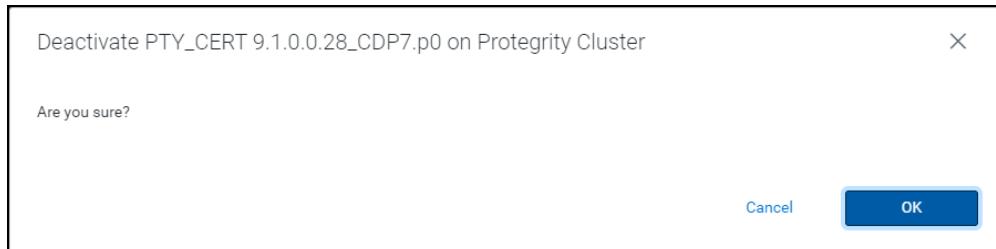
3. To deactivate the **PTY_FLUENTBIT_CONF** parcel, click **OK**.
4. To deactivate the certificates parcel, besides the **PTY_CERT** parcel, click **Deactivate**.



| | | | |
|--------------------|--------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | <button>Activate</button> |

Figure 12-96: Protegity Parcels Activated

The prompt to confirm the deactivation of the parcel appears.

*Figure 12-97: Prompt to Deactivate the Certificates Parcel*

- To deactivate the *PTY_CERT* parcel, click **OK**.

| | | | |
|--------------------|--------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | <button>Activate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | <button>Activate</button> |

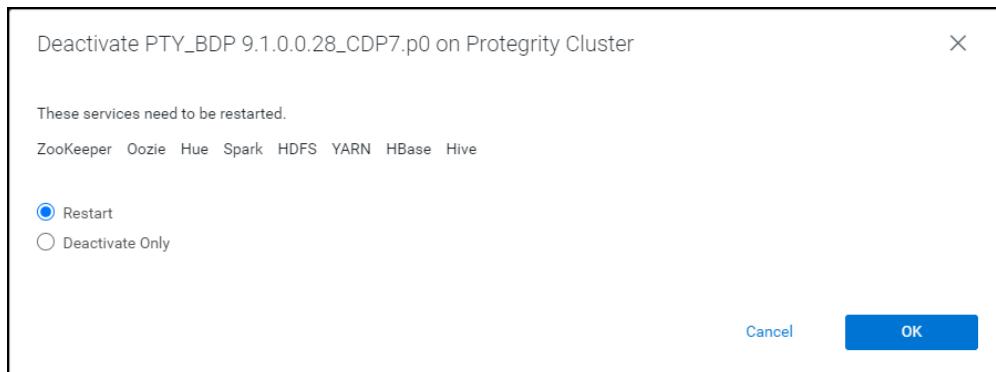
Figure 12-98: Protegity Parcels Activated

- To deactivate the Big Data Protector parcel, besides the *PTY_BDP* parcel, click **Deactivate**.

| | | | |
|--------------------|--------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | <button>Activate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | <button>Activate</button> |

Figure 12-99: Protegity Parcels Activated

The prompt to confirm the deactivation of the parcel and restart of the dependent services appears.

*Figure 12-100: Prompt to Deactivate the Big Data Protector Parcel*

- To restart the services, which are dependent on the parcel that needs to be deactivated, select **Restart**. Alternatively, to just deactivate the parcel, select **Deactivate Only**.

Note: You can restart the dependent services later also. However, it is recommended to restart the dependent services immediately. This will ensure that the dependent services do not utilize the parcel that is being deactivated.

- To deactivate the Big Data Protector parcel, click **OK**.

Note: Alternatively, to terminate the deactivation, click **Abort**.

The deactivation of the Big Data Protector parcel starts.

- To complete the deactivation of the Big Data Protector parcel, click **Close**.

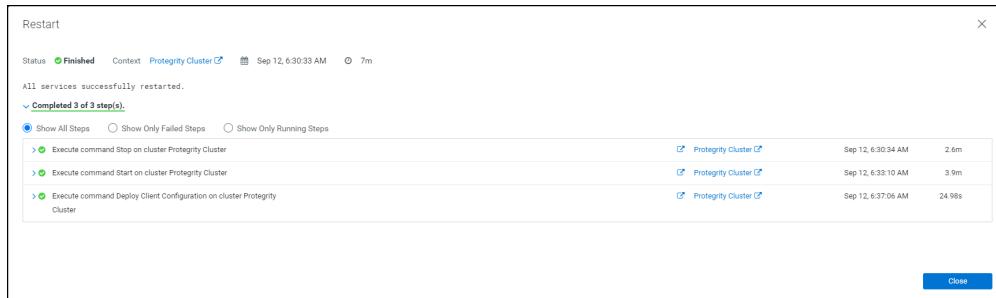


Figure 12-101: Big Data Protector Parcel Deactivated

After you deactivate the **PTY_FLUENTBIT_CONF**, **PTY_CERT**, and **PTY_BDP** parcels, their status on the **Parcels** changes to **Distributed**, and the **Activate** button appears.

| Parcel Name | Version | Status | Action |
|--------------------------|--|------------------------|------------|
| ACCUMULO | 1.7.3.5.0.0.ACCUMULO.05.5.0.p0.8 | Available Remotely | Download |
| CDH 5 | 5.16.1.0.cdh5.16.2.p0.8 | Available Remotely | Download |
| CDH 6 | 6.3.4.1.cdh6.3.4.p0.6791098 | Available Remotely | Download |
| CDH Runtime | 7.1.8-1.cdh7.1.8.p0.3099052 | Distributed, Activated | Deactivate |
| CDRuntime | 4.1.0-1.4.1.0.p0.4 | Available Remotely | Download |
| KAFKA | 4.1.0-1.4.1.0.p0.4 | Available Remotely | Download |
| KEYTRUSTEE_SERVER | 7.1.8.0.1.keytrustee7.1.8.0.p0.3099052 | Available Remotely | Download |
| Kudu | 1.4.0.1.kudu1.12.2.p0.8 | Available Remotely | Download |
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| SPARK3 | 3.3.0.3.3.7180.0.274.1.p0.31212967 | Available Remotely | Download |
| SOOOP_NETEZZA_CONNECTOR | 1.5.1c5 | Available Remotely | Download |
| SOOOP_TERADATA_CONNECTOR | 1.7c5 | Available Remotely | Download |
| mkl | 2023.1.0.46342 | Available Remotely | Download |

Figure 12-102: Big Data Protector Parcels Deactivated

12.6.1.17.3 Removing the Big Data Protector Parcels from all the Nodes

After deactivating the Big Data Protector parcels from the Cloudera Manager, you must remove the following Big Data Protector parcels from all the nodes:

- PTY_BDP**: Big Data Protector parcel
- PTY_CERT**: Certificates parcel
- PTY_FLUENTBIT_CONF**: Fluent Bit configuration parcel

► To remove the Big Data Protector Parcels from all the Nodes in the Cluster:

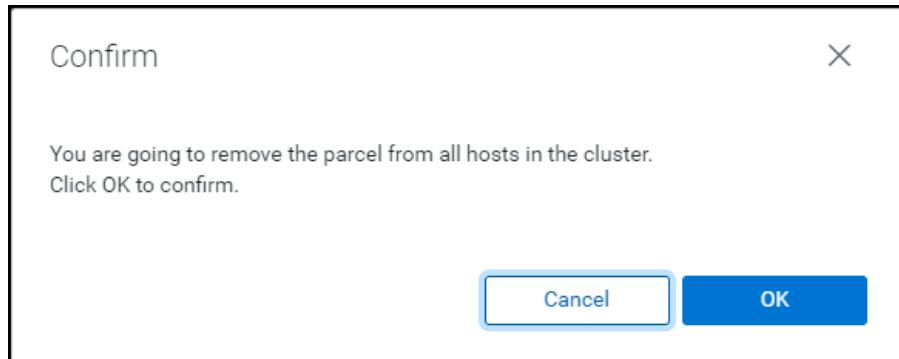
- On the Cloudera Manager **Parcels** page, besides the Big Data Protector parcel, click . The drop-down menu appears.

| | | | |
|--------------------|--------------------|-------------|--|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Distributed | Activate  |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | Remove From Hosts |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Activate  |

Figure 12-103: Drop-down menu

2. Select **Remove From Hosts**.

The prompt to confirm the removal of the Big Data Protector parcel appears.

*Figure 12-104: Prompt to Remove the Big Data Protector Parcel*

3. Click **OK**.

The Big Data Protector parcel is removed from all the nodes in the cluster.

| | | | |
|--------------------|--------------------|-------------|--|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute  |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | Activate  |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Activate  |

Figure 12-105: Big Data Protector Parcel Removed

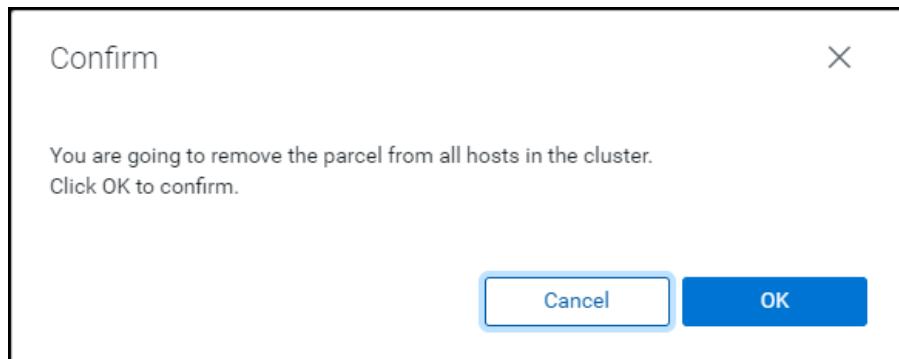
4. Besides the *PTY_CERT* parcel, click  .
The drop-down menu appears.

| | | | |
|--------------------|--------------------|-------------|--|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute  |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Distributed | Activate  |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Remove From Hosts |

Figure 12-106: Drop-down menu

5. Select **Remove From Hosts**.

The prompt to confirm the removal of the Certificates parcel appears.

*Figure 12-107: Prompt to Remove the Certificates Parcel*

6. Click **OK**.

The Certificate parcel is removed from all the nodes in the cluster.

7. Besides the *PTY_FLUENTBIT_CONF* parcel, click  .

The drop-down menu appears.

| | | | |
|--------------------|----------------------------------|--------------------|-------------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Distributed | Activate |
| SPARK3 | 3.3.0.3.7180.0.274-1.p0.31212967 | Available Remotely | Remove From Hosts |

Figure 12-108: Drop-down menu

8. Select **Remove From Hosts**.

The prompt to confirm the removal of the Fluent Bit configuration parcel appears.

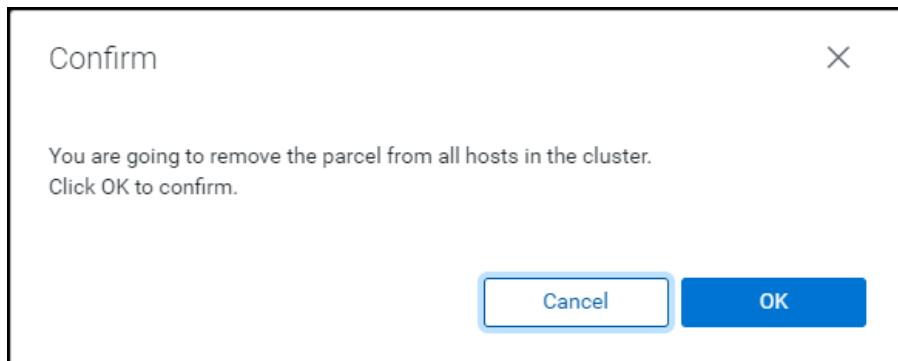


Figure 12-109: Prompt to Remove the Fluent Bit Configuration Parcel

9. Click **OK**.

The Fluent Bit configuration parcel is removed from all the nodes in the cluster.

| | | | |
|--------------------|--------------------|------------|------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |

Figure 12-110: Fluent Bit Configuration Parcel Removed

12.6.1.17.4 Deleting the Big Data Protector Parcels from the Repository

After removing the Big Data Protector parcel from the nodes, you must delete the following Big Data Protector parcels from the local Cloudera Manager repository:

- *PTY_BDP*: Big Data Protector parcel
- *PTY_CERT*: Certificates parcel
- *PTY_FLUENTBIT_CONF*: Fluent Bit configuration parcel

► To delete the Big Data Protector Parcels from the Local Repository:

1. On the Cloudera Manager web interface, navigate to the **Parcels** page. The **Parcels** page appears.

2. Besides the *PTY_BDP* parcel, click . The drop-down menu appears.

| | | | |
|--------------------|--------------------|------------|------------|
| PTY_BDP | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Downloaded | Delete |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute |

Figure 12-111: Drop-down menu

3. Select **Delete**.

The prompt to confirm the deletion of the Big Data Protector parcel appears.

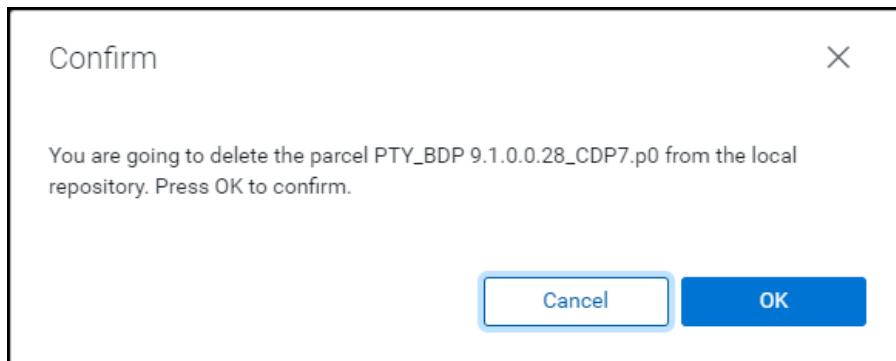


Figure 12-112: Prompt to Delete the Big Data Protector Parcel

4. Click **OK**.

The Big Data Protector parcel is deleted from the local repository.

5. Besides the *PTY_CERT* parcel, click  .
The drop-down menu appears.

| | | | |
|--------------------|------------------------------------|--------------------|--|
| PTY_CERT | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute  |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Downloaded | Delete  |
| SPARK3 | 3.3.0.3.3.7180.0.274-1.p0.31212967 | Available Remotely | Download |

Figure 12-113: Drop-down menu

6. Select **Delete**.

The prompt to confirm the deletion of the Certificates parcel appears.

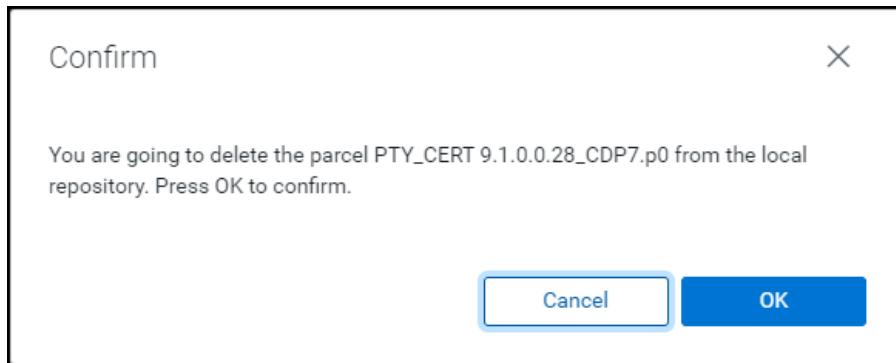


Figure 12-114: Prompt to Delete the Certificates Parcel

7. Click **OK**.

The Certificates parcel is deleted from the local repository.

8. Besides the *PTY_FLUENTBIT_CONF* parcel, click  .
The drop-down menu appears.

| | | | |
|--------------------|---|--------------------|--|
| KEYTRUSTEE_SERVER | 7.1.8.0-1.keytrustee7.1.8.0.p0.30990532 | Available Remotely | Download |
| KUDU | 1.4.0-1.cdh5.12.2.p0.8 | Available Remotely | Download |
| PTY_FLUENTBIT_CONF | 9.1.0.0.28_CDP7.p0 | Downloaded | Distribute  |
| SPARK3 | 3.3.0.3.3.7180.0.274-1.p0.31212967 | Available Remotely | Delete |

Figure 12-115: Drop-down menu

9. Select **Delete**.

The prompt to confirm the deletion of the Fluent Bit configuration parcel appears.

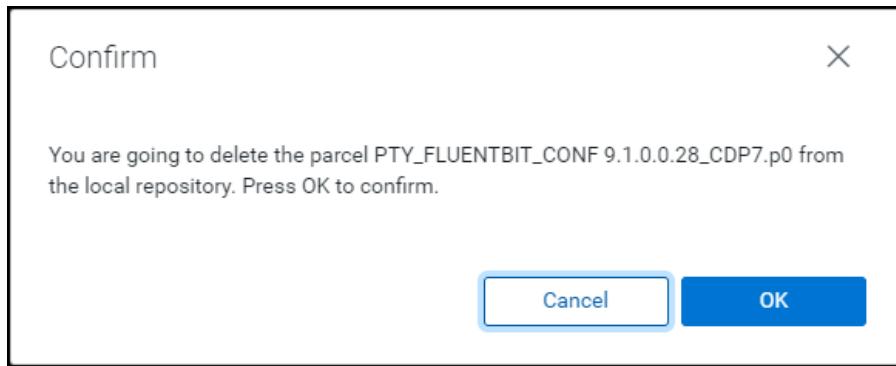


Figure 12-116: Prompt to Delete the Fluent Bit Configuration Parcel

10. Click **OK**.

The Fluent Bit configuration parcel is deleted from the local repository.

11. After all the Big Data Protector parcels are deleted from the repository, remove the Big Data Protector related configuration updates from the cluster.

Note: For more information about removing the Big Data Protector configuration updates from the cluster, refer to section [Setting the Big Data Protector Configuration for CDP-PVC-Base Distribution](#).

12.6.1.17.5 Deleting the CSD Jar Files

The last step in the uninstall process is to delete the following *.jar* files from the local repository of the Cloudera Manager:

- *BDP_PEP.jar*
- *PTY_PROXY.jar*

► To delete the *.jar* files from the local repository of the Cloudera Manager:

1. Login to the Master node.
2. Navigate to the */opt/cloudera/csd/* directory.
3. Delete the following files having the *.jar* extension:
 - *BDP_PEP.jar*
 - *PTY_PROXY.jar*
4. Restart the Cloudera Manager server.
5. After the Cloudera Manager server starts up, restart the Cloudera Management services on the Cloudera Manager web interface.

12.6.2 Installing Big Data Protector using CDH Native Installer

This section describes the procedures to install and uninstall Protegity Big Data Protector 9.1.0.0 release using Cloudera Manager.

Note: The HDFSFP-related components in the CDH Native installer (such as the *PTY_HDFSFP* parcel, *BDP_HDFSFP* CSD, and the *BDP_HDFSFP* service) will not be available starting from the Big Data Protector 7.2.0 release, as the HDFS File Protector (HDFSFP) is deprecated.

For the Big Data Protector 9.1.0.0 release, the CDH and Cloudera Manager, version 6.2, is used for reference.

For more information about CDH and Cloudera Manager, version 6.2, refer to https://docs.cloudera.com/documentation/enterprise/6/6.2/topics/cdh_intro.html.

12.6.2.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before installing Big Data Protector from Cloudera Manager:

- The CDH, version 6.2, cluster is running.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-42: List of Ports for the Big Data Protector

| Destination Ports No. | Protocols | Sources | Destinations | Descriptions |
|--|-----------|--|---|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a Policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all logs to Protegility Audit Store appliance (Elastic Search) through port 9200. |
| 15780 | | Protector on the Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses localhost port 16700. |
| Port Requirement for Proxy Node (Optional) | | | | |
| 8443 | TCP | PEP server on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The PEP server communicates with the Proxy node through port 8443 which substitutes the stream to ESA. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The Log Forwarder communicates with the Proxy node through port 9200 which substitutes the stream to Protegility Audit Store Appliance. |

- The user who is installing Big Data Protector has the required permissions to perform the following tasks:
 - Copy the Big Data Protector parcels and CSDs to the Cloudera Manager repository directories
 - Restart the Cloudera SCM Server
- If you are installing the Big Data Protector on a cluster, then ensure that it is installed on all the nodes in the cluster.
- The group *ptytusr* and the user *ptytusr*, which are responsible to manage the Big Data Protector-related services, are not present in the Hadoop cluster nodes, which are managed by Cloudera Manager.

12.6.2.2 Installing the Big Data Protector

The following sections describes the tasks for installing Big Data Protector on all the nodes in a Hadoop cluster, which are managed by Cloudera Manager.

Caution:

- Ensure that you are logged in as the required user with the relevant privileges for installing Big Data Protector and managing the Cloudera SCM Server.
- The *cloudera-scm* user is the default user for running the Cloudera management services.
- This section considers the *cloudera-scm* user as the default user for installing Big Data Protector and running the Cloudera management services.

12.6.2.3 Downloading the Big Data Protector Package

► To download the Big Data Protector package:

1. After receiving the Big Data Protector installation package from Protegity, login to the Master node in the cluster.
2. Copy the Big Data Protector installation package *BigDataProtector_Linux-ALL-64_x86-64_CDH-6.2-64_9.1.0.0.x.tgz* to the required directory.

Note: Ensure that the Master node is able to connect to the ESA.

12.6.2.4 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script to generate the Big Data Protector parcels and CSDs required for the installation of Big Data Protector on the nodes in the Hadoop cluster, which are managed by Cloudera Manager.

► To extract the files from the installation package:

1. Login to the CLI on the Master node node that has connectivity to the ESA.
2. Copy the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_CDH-6.2-64_9.1.0.0.x.tgz* to a directory, such as the */opt/bigdata* directory.
3. To extract the *BDPConfigurator_CDH-6.2_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_CDH-6.2-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the *BDPConfigurator_CDH-6.2_9.1.0.0.x.sh* file from the Big Data Protector installation package.

12.6.2.5 Running the Big Data Protector Configurator Script

You must run the Big Data Protector configurator script to download certificates from the ESA, and create the parcels and CSDs for Big Data Protector.

► To generate the Big Data Protector Parcels and CSDs:

- Run the *BDPConfigurator_CDH-6.2_9.1.0.0.x.sh* script from the directory where it is extracted.

A prompt to continue the configuration of Big Data Protector appears.

```
./BDPConfigurator_CDH-6.2_9.1.0.0.x.sh
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDH.

Do you want to continue? [yes or no]
```

- To start the configuration of Big Data Protector, type *yes*.

- Press ENTER.

The prompt to select the type of the installation appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDH.

Do you want to continue? [yes or no]
yes

Big Data Protector Configurator started...
Unpacking...
Extracting files...
```

```
Please select the type of Installation files you want to generate?
[ 1: Create All ] : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

- To create the Big Data Protector parcels and CSDs, type *1*.

- To update the *PTY_CERT* parcels with an incremented patch version, type *2*.

For more information about updating the *PTY_CERT* parcel, refer to section [Updating the Certificates Parcel](#).

- To update the *PTY_FLUENTBIT_CONF* parcel with an incremented patch version, type *3*.

For more information about updating the *PTY_FLUENTBIT_CONF* parcel, refer to section [Updating the Fluent Bit Parcel](#).

- Press ENTER.

The prompt to select the operating system appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.
```

```
[ 1: el6 ] : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ] : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ] : SuSE Linux Enterprise Server 12.x
```

```
Please enter the no.:
```

- Depending on the requirements, type *1*, *2*, or *3* to select the operating system version for the Big Data Protector parcels.

9. Press ENTER.

The prompt to enter the ESA Hostname or IP address appears.

```
Please enter the ESA Host or IP Address[]:
```

10. Enter the ESA Hostname or IP address.

11. Press ENTER.

The prompt to enter the listening port for the ESA appears.

```
Enter ESA host listening port [8443]:
```

12. If you want to use the default value of the ESA host listening port, which is *8443*, then press ENTER.

13. If you have configured an external proxy having connectivity with the ESA to download the certificates from the ESA, then enter the external Proxy listening port.

14. Press ENTER.

The prompt to enter the ESA user name appears.

```
Enter ESA User name :
```

15. Enter the ESA user name.

16. Press ENTER.

The prompt to enter the password for the ESA appears.

```
Fetching Certificates from ESA....
```

```
Enter host password for user '<username>':
```

17. Enter the ESA administrator password.

18. Press ENTER.

The certificates are downloaded from the ESA and the prompt to create the *PTY_FLUENTBIT_CONF* parcel containing the custom Fluent Bit configuration file(s) for an external audit store.

```
Do you want to package any custom Fluent-Bit configuration files for External Audit Store?  
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used  
with External Audit Store.  
[ no ] : Skip this step.
```

```
[ yes or no ] :
```

19. To include the Fluent Bit configuration file(s) for an external audit store, type *yes*20. Press *ENTER*.

The prompt to enter the directory to store the configuration files for Fluent Bit appears.

```
Do you want to package any custom Fluent-Bit configuration files for External Audit  
Store?  
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used  
with External Audit Store.  
[ no ] : Skip this step.  
[ yes or no ] : yes  
Creation of PTY_FLUENTBIT_CONF parcel is enabled.  
Enter the local directory path on this machine that stores the Fluent-Bit configuration  
files for External Audit Store:
```

Note: The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across CDP nodes through Cloudera Manager. Ensure that you name the custom Fluent Bit configuration file(s) for the external audit store with the **.conf* extension.

If you type *no* at the prompt to create the *PTY_FLUENTBIT_CONF* parcel, then the installer will skip the creation of the Fluent Bit parcel and proceed to generate the installation files.

```
Do you want to package any custom Fluent-Bit configuration files for External Audit  
Store?
```

```
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used
with External Audit Store.
[ no ] : Skip this step.

[ yes or no ] : no

Creation of PTY_FLUENTBIT_CONF parcel is skipped.

Generating Installation files...

Big Data Protector parcels & CSDs are generated in ./Installation_Files/ directory.

NOTE:
    Copy Big Data Protector CSDs (jars) to Cloudera Manager local csd repository.
    Copy Big Data Protector parcels (*.parcel and *.sha files) to Cloudera Manager local
parcel repository.
```

21. Enter the local directory path that contains the Fluent Bit configuration files.

22. Press *ENTER*.

Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store: /root/fluentbit_file_output

Generating Installation files...

Big Data Protector parcels & CSDs are generated in ./Installation_Files/ directory.

NOTE:

Copy Big Data Protector CSDs (jars) to Cloudera Manager local csd repository.
 Copy Big Data Protector parcels (*.parcel and *.sha files) to Cloudera Manager local parcel repository.

The following Big Data Protector parcels and CSDs are generated in the *Installation_Files/* directory:

- *BDP_PEP-9.1.0.0.x.jar*
- *PTY_BDP-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel*
- *PTY_BDP-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel.sha*
- *PTY_CERT-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel*
- *PTY_CERT-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel.sha*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel*
- *PTY_FLUENTBIT_CONF-9.1.0.0.x_CDH6.2.p0-<OS_Version>.parcel.sha*
- *PTY_PROXY-9.1.0.0.x.jar*

12.6.2.6 Setting up the Big Data Protector Parcels and CSDs

After the Big Data Protector parcels and CSDs are copied to the respective local Cloudera repository directories, you need to restart the Cloudera SCM server to ensure that Cloudera Manager identifies the new CSD files and display the Big Data Protector services in the Add Services section in Cloudera Manager.

► To set up the Big Data Protector Parcels and CSDs:

1. Copy the Big Data Protector parcels with the *.parcel* extension and respective checksum files with the *.sha* extension to the local parcel repository of Cloudera Manager.

The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.

2. Copy the Big Data Protector CSD files with the *.jar* extension to the local CSD repository.

The default local CSD repository for Cloudera Manager is located in the */opt/cloudera/csd/* directory.

3. Navigate to the local parcel repository directory.

Note: The local parcel repository is stored in the `/opt/cloudera/parcel-repo/` directory.

4. To assign the ownership permissions for the *Cloudera SCM* user to the Protegity Big Data Protector parcels and checksum files, run the following command.

```
chown cloudera-scm:cloudera-scm *
```

5. Press *ENTER*.

6. To set 640 permissions to the parcel files, run the following command.

```
chmod 640 *
```

7. Navigate to the local CSD repository directory.

Note: The local CSD repository is located in the `/opt/cloudera/csd` directory.

8. To assign the ownership permissions for the Cloudera SCM user to the Big Data Protector CSD files, run the following command.

```
chown cloudera-scm:cloudera-scm *
```

9. Press *ENTER*.

10. To set 640 permissions to the CSD files, run the following command.

```
chmod 640 *
```

11. To restart the Cloudera SCM server and load the Big Data Protector CSD files in the Cloudera Manager, run the following command.

```
service cloudera-scm-server restart
```

12. Press *ENTER*.

The new parcels in the local parcel repository are detected by Cloudera Manager.

Note: You must restart the Cloudera SCM server to ensure that the Big Data Protector services are listed on the **Add Services** screen in Cloudera Manager.

12.6.2.7 Distributing the Big Data Protector Parcels to the Nodes

You must distribute the following Big Data Protector parcels to the cluster nodes before installing or activating them on the nodes:

- Big Data Protector parcel: `PTY_BDP`
- Certificates parcel: `PTY_CERT`
- Fluent Bit configuration parcel: `PTY_FLUENTBIT_CONF`

To distribute the Big Data Protector parcels to the nodes, *Cluster Administrator* privileges are required.

Note: For more information about the required role, refer to <https://docs.cloudera.com/cloudera-manager/7.1.1/managing-clusters/topics/cm-parcels.html>.

- To distribute the Big Data Protector Parcels to the Nodes in the Cluster:

- Using a browser, navigate to the Cloudera Manager screen.

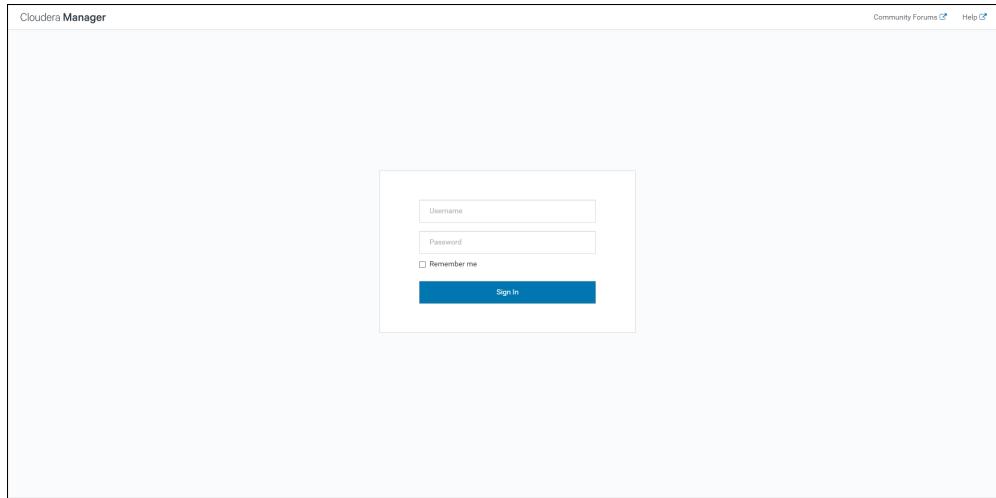


Figure 12-117: Cloudera Manager screen

- Enter the required user name for logging in to Cloudera Manager.
- Enter the required password for logging in to Cloudera Manager.
- Click **Sign In**.

The Cloudera Manager Home screen appears.

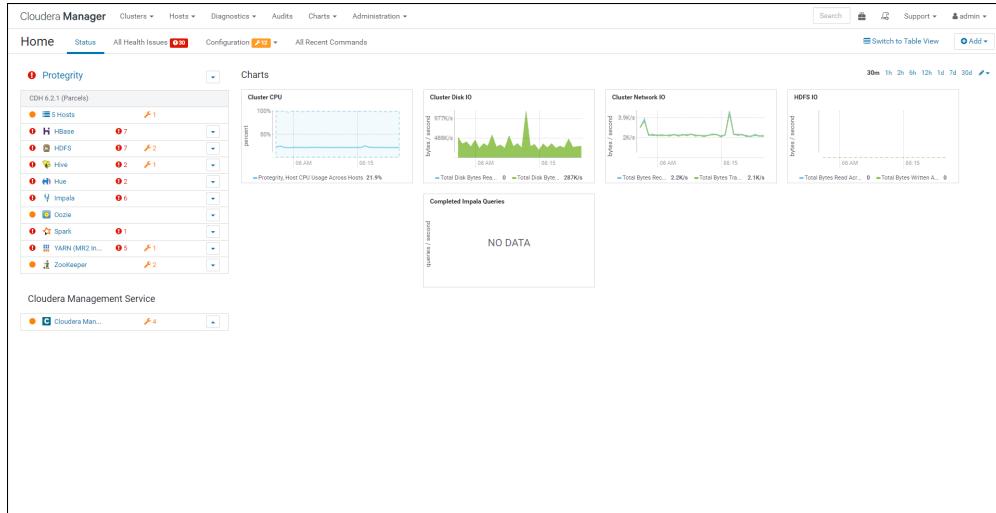


Figure 12-118: Cloudera Manager Home screen

- Navigate to **Administration > Settings**.
The **Settings** page appears.

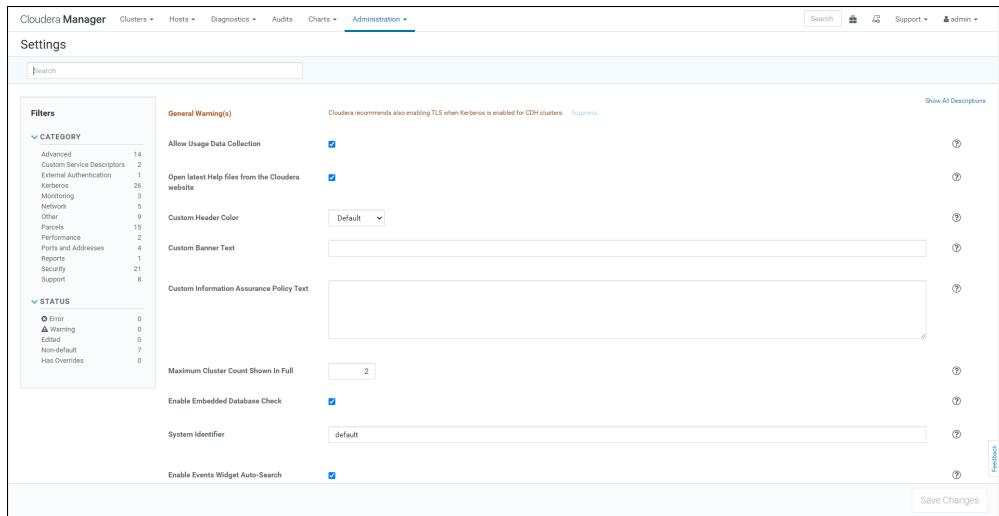


Figure 12-119: Settings Page

6. To view the settings related to parcels, from the **Filters** pane, under **CATEGORY**, click **Parcels**. The options related to parcels appear.
7. Ensure that the check box beside the **Create Users and Groups, and apply File Permissions for Parcels** option is selected.

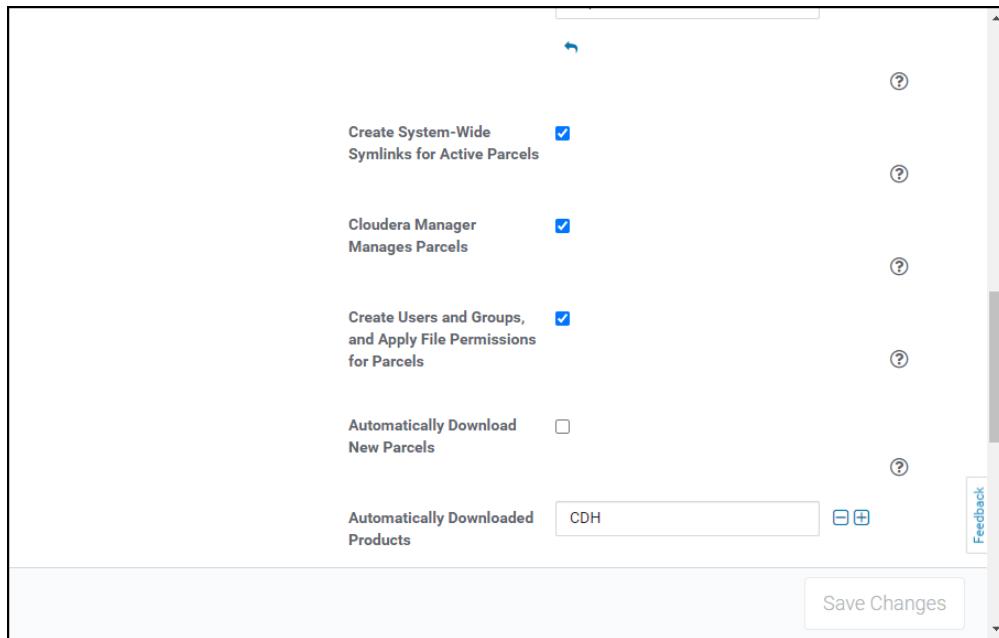


Figure 12-120: Options on the Settings screen

8. From the top navigation pane, navigate to **Clusters > Parcels**. The Cloudera Manager Parcels page appears.

The screenshot shows the Cloudera Manager interface with the 'Parcels' tab selected. In the 'Protegility' section, several parcels are listed with their names, versions, and statuses. The 'PTY_FLUENTBIT_CONF' parcel is highlighted with a red box. It has a version of '9.1.0.0.8,CDH6.2.par' and is currently 'Downloaded'. There are 'Distribute' and 'Deactivate' buttons next to it.

| Location | Parcel Name | Version | Status | Action Buttons |
|-------------|---------------------------|-------------------------------------|-------------------------------|-----------------------------|
| Protegility | ACUMULO | 1.7.2-8.5.0.ACUMULO.05.5.0.par.8 | Available Remotely | <button>Download</button> |
| Filters | CDH 6 | 6.3.2.1.cdh6.3.2.par.160554 | Downloaded | <button>Distribute</button> |
| | PTY_FLUENTBIT_CONF | 6.2.1-1.cdh6.2.1.par.1425774 | Distributed, Activated | <button>Deactivate</button> |
| | KAFKA | 5.16.2.1.cdh5.16.2.par.8 | Available Remotely | <button>Distribute</button> |
| | PTY_BDP | 4.1.0.1.4.1.0.par.4 | Available Remotely | <button>Download</button> |
| | PTY_CERT | 9.1.0.0.8,CDH6.2.par | Available Remotely | <button>Distribute</button> |
| | KUDU | 1.4.0.1.cdh5.12.2.par.8 | Available Remotely | <button>Download</button> |
| | PTY_BDP | 9.1.0.0.8,CDH6.2.par | Downloaded | <button>Distribute</button> |
| | PTY_CERT | 9.1.0.0.8,CDH6.2.par | Downloaded | <button>Distribute</button> |
| | SOOOP_NETEZZA_CONNECTOR | 1.5.1cb | Available Remotely | <button>Download</button> |
| | | 1.5.1cd | Available Remotely | <button>Download</button> |
| | SOOOP_TERADATA_CONNECTOR | 1.7.6S | Available Remotely | <button>Download</button> |
| | mail | 2022.0.2.136 | Available Remotely | <button>Download</button> |

Figure 12-121: Cloudera Manager Parcels Page

Note: The *PTY_FLUENTBIT_CONF* parcel will be visible only if you choose to add the location of the Fluent Bit configuration files while generating the installation files.

9. Ensure that the following Protegility parcels appear on the **Parcels** page:
 - *PTY_BDP*: Big Data Protector parcel
 - *PTY_CERT*: Certificates parcel
 - *PTY_FLUENTBIT_CONF*: Fluent Bit configuration parcel
10. To distribute the Big Data Protector parcel, besides the *PTY_BDP* parcel, click **Distribute**. The distribution of the Big Data Protector parcel starts.
11. To distribute the Certificates parcel, besides the *PTY_CERT* parcel, click **Distribute**. The distribution of the Certificates parcel starts.
12. To distribute the Fluent Bit configuration parcel, besides the *PTY_FLUENTBIT_CONF* parcel, click **Distribute**. The distribution of the Fluent Bit configuration parcel starts.

The screenshot shows the Cloudera Manager interface with the 'Parcels' tab selected. In the 'Protegility' section, the same list of parcels is shown. The 'PTY_FLUENTBIT_CONF' parcel is highlighted with a red box and has a status of 'Distributing % Details'. There are 'Cancel' buttons next to the distribution progress bars.

| Location | Parcel Name | Version | Status | Action Buttons |
|-------------|---------------------------|-------------------------------------|-------------------------------|-----------------------------|
| Protegility | ACUMULO | 1.7.2-8.5.0.ACUMULO.05.5.0.par.8 | Available Remotely | <button>Download</button> |
| Filters | CDH 6 | 6.3.2.1.cdh6.3.2.par.160554 | Downloaded | <button>Distribute</button> |
| | PTY_FLUENTBIT_CONF | 6.2.1-1.cdh6.2.1.par.1425774 | Distributing % Details | <button>Cancel</button> |
| | KAFKA | 5.16.2.1.cdh5.16.2.par.8 | Available Remotely | <button>Distribute</button> |
| | PTY_BDP | 4.1.0.1.4.1.0.par.4 | Available Remotely | <button>Download</button> |
| | PTY_CERT | 9.1.0.0.8,CDH6.2.par | Distributing % Details | <button>Cancel</button> |
| | PTY_FLUENTBIT_CONF | 9.1.0.0.8,CDH6.2.par | Distributing % Details | <button>Cancel</button> |
| | SOOOP_NETEZZA_CONNECTOR | 1.5.1cb | Available Remotely | <button>Download</button> |
| | | 1.5.1cd | Available Remotely | <button>Download</button> |
| | SOOOP_TERADATA_CONNECTOR | 1.7.6S | Available Remotely | <button>Download</button> |
| | mail | 2022.0.2.136 | Available Remotely | <button>Download</button> |

Figure 12-122: Distribution of Parcels Starts

After the Protegility parcels are distributed to the nodes, their status on the Parcels screen is updated to **Distributed**, and the **Activate** button appears.

| Location | Parcels | Protegility | |
|--------------------------------|---------|--------------------------|--|
| Protegility Available Remotely | | | |
| Filters | | | |
| PARCEL NAME | | | |
| ACUMULO | 1 | ACUMULO | Version 1.7.2.5.0.0.ACUMULO.05.5.0.p0.8 Status Available Remotely Download |
| CDH 5 | 1 | CDH 6 | Version 6.3.2.1.cdh6.3.2.p0.1400054 Status Downloaded Distribute |
| CDH 6 | 3 | | Version 6.3.2.1.cdh6.2.1.p0.1425774 Status Distributed, Activated Deactivate |
| KAFKA | 1 | | Version 6.2.0.1.cdh6.2.0.p0.907373 Status Downloaded Distribute |
| KUDU | 1 | | Version 5.16.2.1.cdh6.16.2.p0.8 Status Available Remotely Download |
| PTY_BDP | 1 | KAFKA | Version 4.1.0.1.4.1.p0.4 Status Available Remotely Download |
| PTY_CERT | 1 | KUDU | Version 1.4.0.1.cdh6.12.2.p0.8 Status Available Remotely Download |
| PTY_FLUENTBIT_CONF | 1 | | |
| SQOOP_TERADATA_CONNECTOR | 1 | PTY_BDP | Version 9.1.0.0.8.CDH6.2.p0 Status Distributed Activate |
| sqoop_teradata_connector | 1 | PTY_CERT | Version 9.1.0.0.8.CDH6.2.p0 Status Distributed Activate |
| mil | 1 | PTY_FLUENTBIT_CONF | Version 9.1.0.0.8.CDH6.2.p0 Status Distributed Activate |
| | | SQOOP_NETEZZA_CONNECTOR | Version 1.5.1cS Status Available Remotely Download |
| | | | Version 1.5.1cS Status Available Remotely Download |
| | | SQOOP_TERADATA_CONNECTOR | Version 1.7cS Status Available Remotely Download |
| | | mil | Version 2022.0.2.135 Status Available Remotely Download |

Figure 12-123: Protegility Parcels Distributed

12.6.2.8 Activating the Big Data Protector Parcel on the Nodes

Before you begin

After distributing the Big Data Protector parcels on the cluster nodes, you need to activate the parcels so that the Big Data Protector-related services can be added and then started on the nodes in the cluster.

To activate the Big Data Protector Parcels on the Nodes:

1. To activate the Big Data Protector parcel, besides the *PTY_BDP* parcel, click **Activate**. The prompt to confirm the activation of the parcel appears.

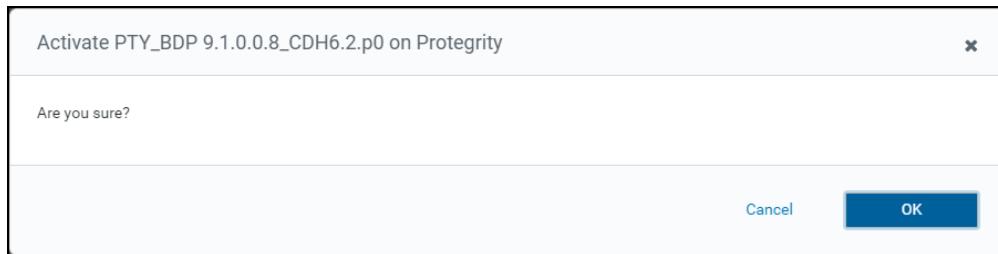


Figure 12-124: Prompt to Activate the *PTY_BDP* Parcel

2. To activate the Big Data Protector parcel, click **OK**.

Tip: To terminate activation, click **Cancel**.

3. To activate the Certificates parcel, besides the *PTY_CERT* parcel, click **Activate**. The prompt to confirm the activation of the parcel appears.

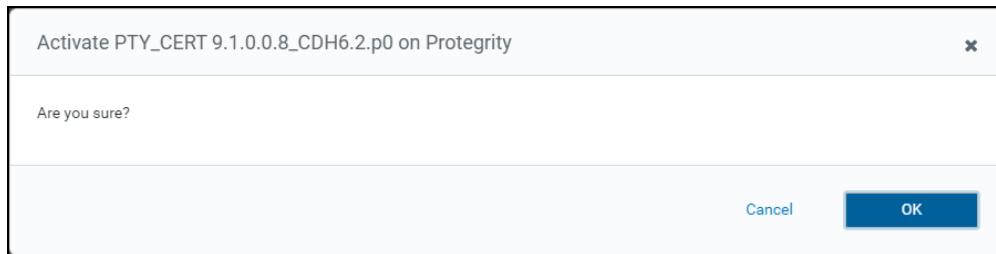


Figure 12-125: Prompt to Activate the *PTY_CERT* Parcel

4. To activate the *PTY_CERT* parcel, click **OK**.
5. To activate the Fluent Bit configuration parcel, besides the *PTY_FLUENTBIT_CONF* parcel, click **Activate**. A prompt to confirm the activation of the parcel appears.

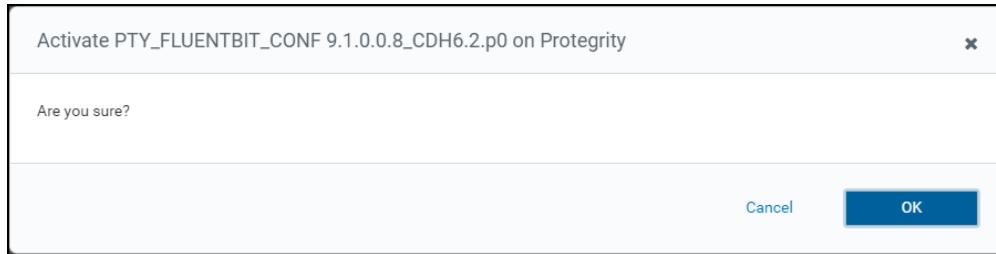


Figure 12-126: Prompt to Activate the Fluent Bit Configuration Parcel

6. To activate the *PTY_FLUENTBIT_CONF* parcel, click **OK**.

After the Protegility parcels are activated on the nodes, their status on the Parcels screen is updated to *Distributed*, *Activated*, and the **Deactivate** button appears.

| Parcels | | |
|-----------------|--------------------------|-------------------------------|
| Location | Parcel Name | Status |
| Protegility | ACUMULO | Available Remotely |
| | CDH 5 | Download |
| | CDH 6 | Downloaded |
| | PTY_BOP | Distributed, Activated |
| | KAFKA | Download |
| | KUDU | Download |
| | PTY_CERT | Download |
| | PTY_FLUENTBIT_CONF | Download |
| | SOOOP_NETEZZA_CONNECTOR | Download |
| | SOOOP_TERADATA_CONNECTOR | Download |
| | mkl | Download |
| ✓ STATUS | | |
| Distributed | 4 | |
| Other | 10 | |

Figure 12-127: Protegility Parcels Activated

7. Navigate to the **Cloudera Manager** home page.

The **Cloudera Manager** home page appears and the required Hadoop services are displayed with their configuration states in the *Cluster* area.

Note: If the configuration state of any Hadoop services is stale, then  appears beside the service.

8. To redeploy the service configuration and restart the service, besides the required service, click .

The screenshot shows the 'Stale Configurations' section of the Cloudera Manager interface. On the left, there are filters for 'Parcels', 'FILE', 'Services', and 'Role Type'. Under 'Services', several services are listed with their counts: HBase (1), HDFS (1), Hive (1), Hue (1), Oozie (1), Spark (1), and Zookeeper (1). Under 'Role Type', various roles are listed with their counts, such as DataNode (1), HBase Master Server (1), History Server (1), and so on. At the top right, a message indicates '8 services affected' with a 'Show' button. A large blue button at the bottom right says 'Restart Stale Services'.

Figure 12-128: Configuration State of Services

- To restart the configuration, click **Restart Stale Services**.

The **Review Changes** page appears.

The screenshot shows the 'Review Changes' page. It has two main sections: 'Review Changes' on the left and 'Command Details' on the right. The 'Review Changes' section contains a message: 'All services running with outdated configurations in the cluster and their dependencies will be restarted.' Below this is a checkbox labeled 'Re-deploy client configuration'. At the bottom right, there are 'Back' and 'Restart Now' buttons.

Figure 12-129: Review Changes Page

- To deploy the updated configuration, click **Restart Now**.

The updated service configuration is deployed to the nodes and the service is restarted.

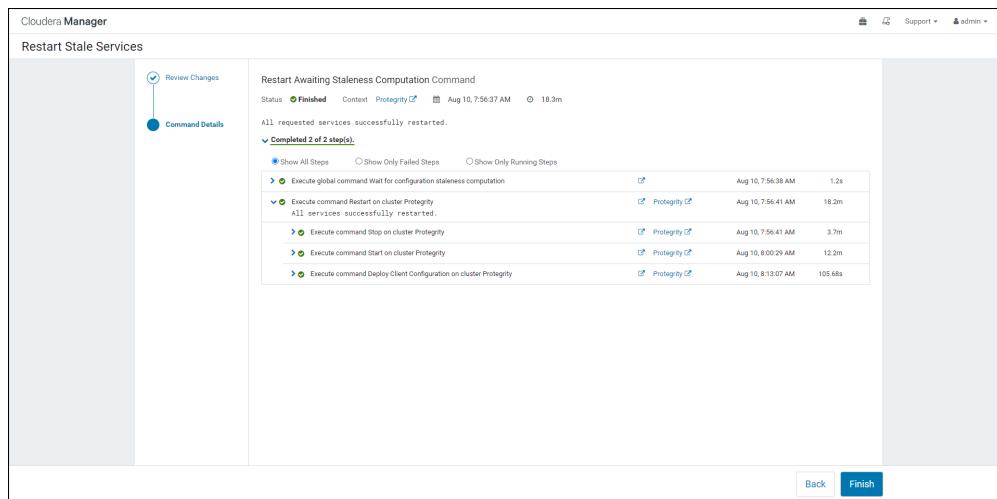


Figure 12-130: Service Restart screens

11. Click **Finish**.

The Protegity Big Data Protector parcels are installed on all the required nodes in the cluster.

12. Restart the Cloudera Management Service to redeploy the service configuration in for the stale configurations.

12.6.2.9 Verifying the Permissions and Ownerships for the Extracted Parcels

After distributing and activating the parcels, you must ensure that the parcels extracted on all the nodes have the correct file permissions and ownerships.

Attention: You must perform these steps on all the nodes in the cluster.

► To verify and ensure the permissions and ownership:

1. Navigate to the `/opt/cloudera/parcels/` directory.
2. To view the permissions of the `BDP_PEP` and `PTY_CERT` parcels, execute the following command.

```
ls -ld PTY*
```

3. Press ENTER.

The command displays the permissions for the parcels.

```
[root@master parcels]# ls -ld PTY*
lrwxrwxrwx 1 root root 24 Jul 22 12:32 PTY_BDP -> PTY_BDP-9.1.0.0.x_CDH6.2.p0
drwxr-xr-x 12 root root 165 Jul 19 08:32 PTY_BDP-9.1.0.0.x_CDH6.2.p0
lrwxrwxrwx 1 root root 25 Jul 22 12:33 PTY_CERT -> PTY_CERT-9.1.0.0.x_CDH6.2.p0
drwxr-xr-x 4 root root 30 Jul 19 08:32 PTY_CERT-9.1.0.0.x_CDH6.2.p0
```

Note: Every sub-directory and file within these parcel directories must have the ownership set to `ptyitusr:ptyitusr` and the permissions must be set according to the permission metadata set in the `/opt/cloudera/parcels/PTY_BDP/meta/permissions.json` and the `/opt/cloudera/parcels/PTY_CERT/meta/permissions.json` files.

12.6.2.10 Starting the Big Data Protector Services on the Nodes

After activating the Big Data Protector parcels on the nodes, you must add and start the following Big Data Protector-related services on the cluster:

- BDP PEP service: Add and start the *BDP_PEP* service on all nodes in the cluster
- PTY Proxy service: Add and start the *PTY_PROXY* service on the required nodes in the cluster. If the cluster is configured in a way where the nodes do not have connectivity with the ESA, then the node is typically the *Lead (Edge)* node.

12.6.2.10.1 Starting the Big Data Protector PEP Service

If you need to use Big Data Protector, then you need to start the Big Data Protector PEP service on all the nodes in the cluster.

Before you begin

Before starting the Big Data Protector PEP service, ensure that the following Big Data Protector-related parcels are in Activated state:

- Big Data Protector parcel: *PTY_BDP*
- Certificates parcel: *PTY_CERT*
- Fluent Bit configuration parcel: *PTY_FLUENTBIT_CONF*

► To start the Big Data Protector PEP Service on the Nodes:

1. Besides the cluster name, click the kebab menu icon . The cluster drop down appears.

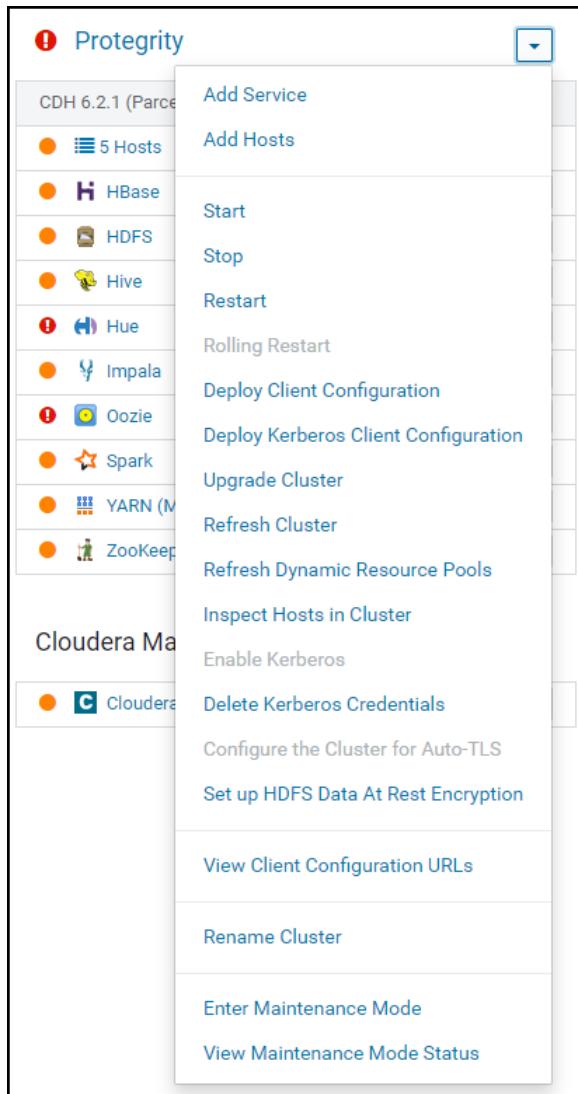


Figure 12-131: Cluster Drop Down

2. Select Add Service.

The cluster services wizard page appears.

| Service Type | Description |
|--|--|
| <input type="radio"/> ADLS Connector | The ADLS Connector service provides key management for accessing ADLS Gen1 accounts and ADLS Gen2 containers from CDH services. |
| <input type="radio"/> Accumulo | The Apache Accumulo sorted, distributed key/value store is a robust, scalable, high performance data storage and retrieval system. This service only works with releases meant to run on top of CDH. |
| <input type="radio"/> BDP PEP | Provides a distributed service for comprehensive data protection across various components in the Hadoop ecosystem. |
| <input type="radio"/> Flume | Flume collects and aggregates data from almost any source into a persisted store such as HDFS. |
| <input type="radio"/> HBase | Apache HBase provides random, real-time, read/write access to large data sets (requires HDFS and ZooKeeper). |
| <input type="radio"/> HDFS | Apache Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications. HDFS creates multiple replicas of data blocks and distributes them on compute hosts throughout a cluster to enable reliable, extremely rapid computations. |
| <input type="radio"/> Hive | Hive is a data warehouse system that offers a SQL-like language called HiveQL. |
| <input type="radio"/> Hue | Hue is a graphical user interface to work with the Cloudera Distribution including Apache Hadoop (requires HDFS, MapReduce, and Hive). |
| <input type="radio"/> Impala | Impala provides a real-time SQL query interface for data stored in HDFS and Hbase. Impala requires the Hive service and shares the Hive Metastore with Hive. |
| <input type="radio"/> Isilon | EMC Isilon is a distributed file system. |
| <input checked="" type="radio"/> Java KeyStore KMS | The Hadoop Key Management Service with file-based Java KeyStore. Maintains a single copy of keys, using simple password-based protection. Requires CDH 6.0+. Not recommended for production use. |
| <input type="radio"/> Kafka | Apache Kafka is publish-subscribe messaging rethought as a distributed commit log. |
| <input type="radio"/> Key-Value Store Indexer | Key-Value Store Indexer listens for changes in data inside tables contained in HBase and indexes them using Solr. |
| <input type="radio"/> Kudu | Kudu is a true column store for the Hadoop ecosystem. |
| <input type="radio"/> Oozie | Oozie is a workflow coordination service to manage data processing jobs on your cluster. |

Figure 12-132: Cluster Services Page

- From the **Service Type** list, select **BDP PEP**.
Cloudera enables the **Continue** button.

4. Click **Continue**.
The **Assign Roles** page appears.

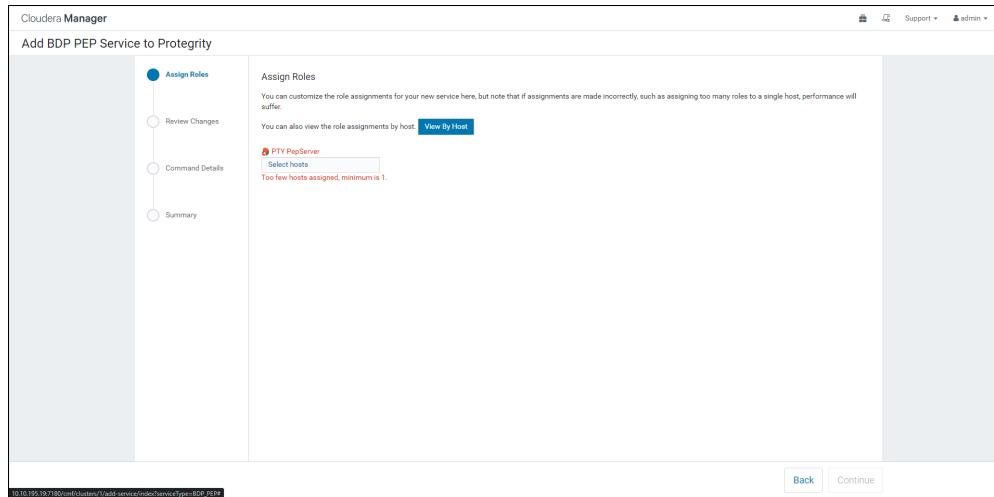


Figure 12-133: Assign Roles Page

5. Click the highlighted text box.
The list of nodes in the cluster appear.
 6. To select all the host nodes in the list to install the BDP PEP service, select the **Hostname** checkbox.

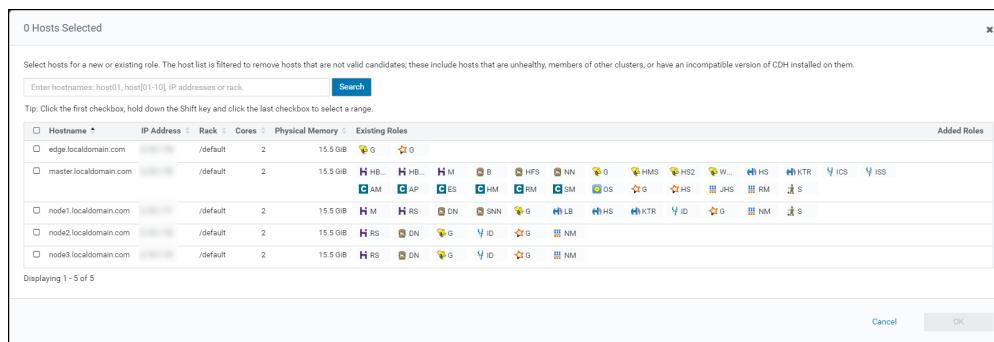


Figure 12-134: List of Nodes for Installing the BDP PEP Service

Cloudera enables the **OK** button.

Note: The *PTY PepServer* and the *PTY Log Forwarder* roles are installed on the selected node.

- The BDP PEP screen appears with the nodes in the cluster, which are selected for installing the *BDP PEP* service.

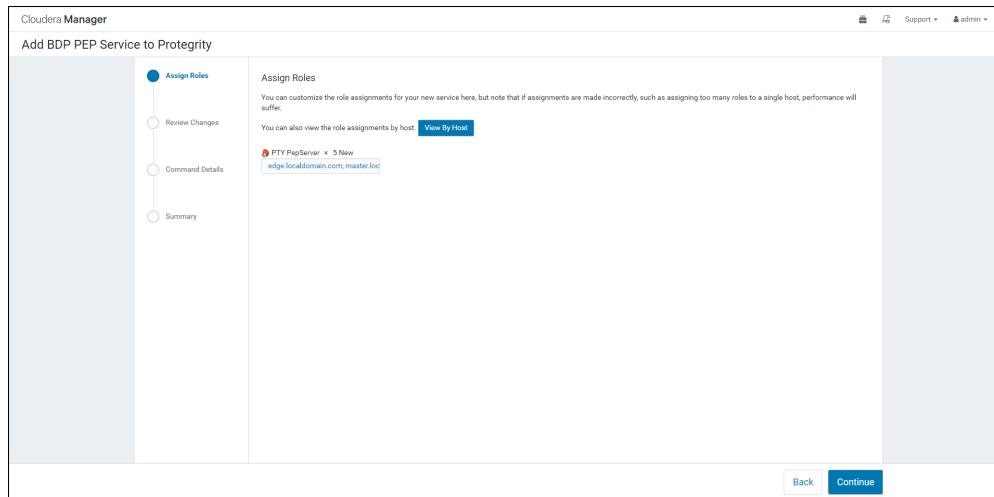


Figure 12-135: List of Nodes for Installing the BDP PEP Service

8. Click **Continue**.

The configuration **Review Changes** page appears.

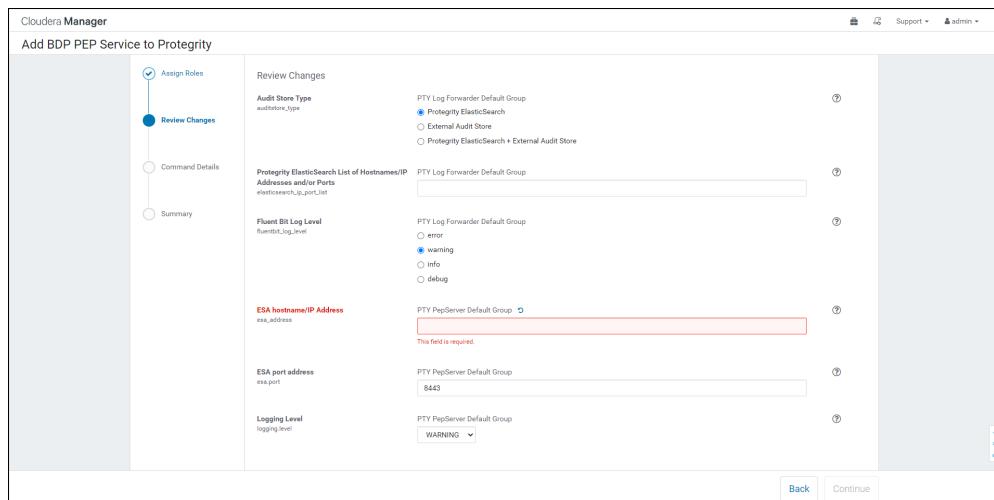


Figure 12-136: Review Changes for BDP PEP Service Page

9. Depending on the Audit Store type, select any one of the following options:

| Option | Description |
|---|--|
| Protegility Elasticsearch | Enter the comma-separated IP/ports using the accurate syntax in the Protegility Elasticsearch List of Hostnames/IP Addresses and/or Ports box |
| External Audit Store | Enter the comma-separated IP/ports using the accurate syntax in the Protegility Elasticsearch List of Hostnames/IP Addresses and/or Ports box. Ensure that the <i>PTY_FLUENTBIT_CONF</i> parcel is distributed and activated. |
| Protegility Elasticsearch + External Audit Store | Enter the comma-separated IP/ports using the correct syntax in the Protegility Elasticsearch List of Hostnames/IP Addresses and/or Ports box. |

Note: When you select **External Audit Store** and **Protegility Elasticsearch + External Audit Store**, the custom configuration files are copied from the currently activated *PTY_FLUENTBIT_CONF* parcel to the */opt/cloudera/parcels/PTY_BDP/fluent-bit/data/config.d* directory when the *BDP_PEP* service is started or restarted.

10. Enter the IP address of the ESA in the **ESA hostname/IP Address** box.

11. Click **Continue**.

The Big Data Protector PEP server service is started on the required nodes in the cluster.

Figure 12-137: BDP PEP Server Service Page

12. Click Continue.

The **Summary** page appears.

Figure 12-138: Summary Page to add BDP PEP Service to the Protegility Cluster

13. Click **Finish**.

The Cloudera Manager Home screen appears and the BDP PEP service is added and started on all the nodes in the cluster.

Note: In the Cloudera Manager native installer, there is a caveat in the *BDP PEP* service that the *PTY Log Forwarder* and the *PTY PepServer* roles will be started at the same time on a cluster node. Therefore, some of the initial PEP server application logs would not be sent to the Log Forwarder and in turn, the logs would not be forwarded to the Audit Store. After the Log Forwarder starts up, it will start forwarding the PEP server application logs.

12.6.2.10.2 Starting the Proxy Service

If you want to utilize the *PTY_Proxy* service, then you must add and start the *PTY_Proxy* service on the required nodes in the cluster. Typically, the Lead (Edge) node is used for installing the proxy service.

Note: The *PTY_PROXY* service is required if the cluster is configured in a way, where the all the nodes do not have connectivity with the ESA. Typically, you must configure the Proxy service on the *Lead (Edge)* node because it has connectivity to the ESA.

You can configure the Proxy service either when Big Data Protector is being installed or at a later stage after the Big Data Protector is already installed.

Warning: The *PTY_Proxy* service can receive and forward the TCP traffic coming from Log forwarder(s), only to a single Protegility ElasticSearch appliance. Multiple ElasticSearch endpoints are not supported.

Before you begin

Before starting the *PTY_Proxy* service, ensure that the following Big Data Protector-related parcels are in the *Activated* state:

- Big Data Protector parcel: *PTY_BDP*
- Certificates parcel: *PTY_CERT*
- Fluent Bit configuration parcel: *PTY_FLUENTBIT_CONF*

► To start the *PTY_Proxy* Service on the Nodes:

1. Besides the cluster name, click the kebab menu icon . The cluster drop down appears.

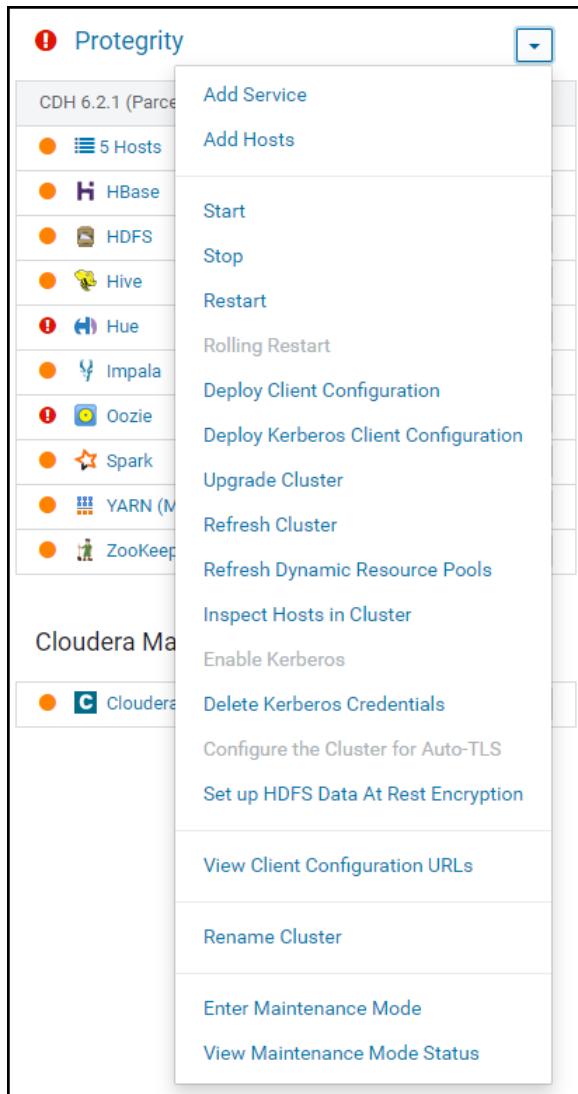


Figure 12-139: Cluster Drop Down

2. Select Add Service.

The cluster services wizard page appears.

| Service Type | Description |
|--|--|
| <input type="radio"/> ADLS Connector | The ADLS Connector service provides key management for accessing ADLS Gen1 accounts and ADLS Gen2 containers from CDH services. |
| <input type="radio"/> Accumulo | The Apache Accumulo sorted, distributed key/value store is a robust, scalable, high performance data storage and retrieval system. This service only works with releases meant to run on top of CDH. |
| <input type="radio"/> BDP PEP | Provides a distributed service for comprehensive data protection across various components in Hadoop Ecosystem. |
| <input type="radio"/> Flume | Flume collects and aggregates data from almost any source into a persisted store such as HDFS. |
| <input type="radio"/> HBase | Apache HBase provides random, real-time, read/write access to large data sets (requires HDFS and ZooKeeper). |
| <input type="radio"/> HDFS | Apache Hadoop Distributed File System (HDFS) is the primary storage system used by Hadoop applications; HDFS creates multiple replicas of data blocks and distributes them on compute hosts throughout a cluster to enable reliable, extremely rapid computations. |
| <input type="radio"/> Hive | Hive is a data warehouse system that offers a SQL-like language called HiveQL. |
| <input type="radio"/> Hue | Hue is a graphical user interface to work with the Cloudera Distribution including Apache Hadoop (requires HDFS, MapReduce, and Hive). |
| <input type="radio"/> Impala | Impala provides a real-time SQL query interface for data stored in HDFS and Hbase. Impala requires the Hive service and shares the Hive Metastore with Hive. |
| <input type="radio"/> Isilon | EMC Isilon is a distributed filesystem. |
| <input checked="" type="radio"/> Java KeyStore KMS | The Hadoop Key Management Service with file-based Java KeyStore. Maintains a single copy of keys, using simple password-based protection. Requires CDH 6.0+. <small>Not recommended for production use.</small> |
| <input type="radio"/> Kafka | Apache Kafka is publish-subscribe messaging rethought as a distributed commit log. |
| <input type="radio"/> Key-Value Store Indexer | Key-Value Store Indexer listens for changes in data inside tables contained in HBase and indexes them using Solr. |
| <input type="radio"/> Kudu | Kudu is a true column store for the Hadoop ecosystem. |
| <input type="radio"/> Oozie | Oozie is a workflow coordination service to manage data processing jobs on your cluster. |
| <input type="radio"/> PTY PROXY | Provides PROXY to enable connectivity to ESA and/or Audit Store PSU (Protegility ElasticSearch) in closed network systems. |

Figure 12-140: Cluster Services Page

3. From the Service Type list, select PTY PROXY.

Cloudera enables the **Continue** button.

4. Click **Continue**.
The **Assign Roles** page appears.

Figure 12-141: Assign Roles Page

5. Click the above highlighted text box.
The list of nodes in the cluster appear.
6. To select the required host node in the list to install the *PTY PROXY* service, select the check box against the node.

| Hostname | IP Address | Rack | Cores | Physical Memory | Existing Roles | Added Roles |
|--|------------|------|---------|-----------------|----------------|-------------|
| <input checked="" type="checkbox"/> edge.localdomain.com | /default | 2 | 15.5 GB | | | |
| <input type="checkbox"/> master.localdomain.com | /default | 2 | 15.5 GB | | | |
| <input type="checkbox"/> node1.localdomain.com | /default | 2 | 15.5 GB | | | |
| <input type="checkbox"/> node2.localdomain.com | /default | 2 | 15.5 GB | | | |
| <input type="checkbox"/> node3.localdomain.com | /default | 2 | 15.5 GB | | | |

Figure 12-142: Selection of Node for Installing the *PTY PROXY* Service

7. Click **OK**.
The *PTY PROXY* screen is updated with the required node in the cluster, which is selected for installing the *PTY PROXY* service.

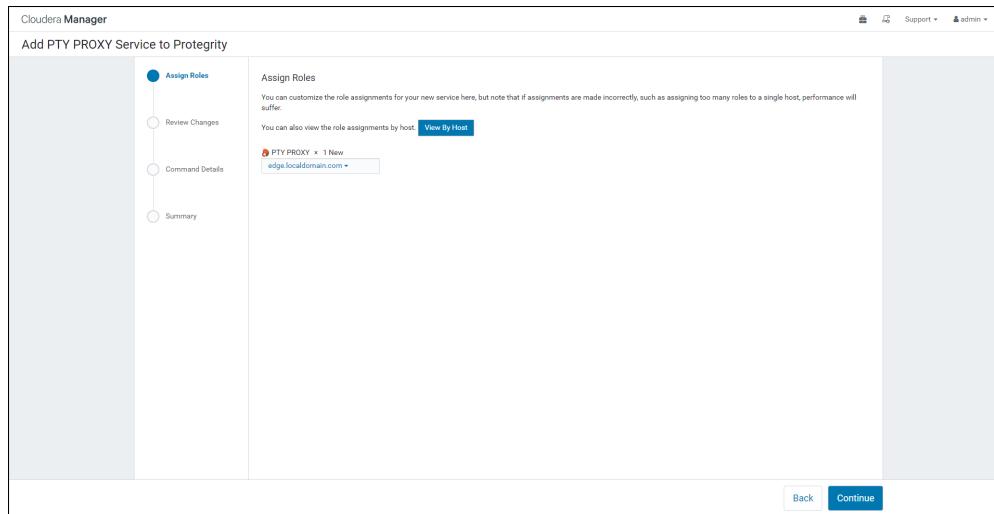


Figure 12-143: PTY PROXY Host

8. Click **Continue**.

The configuration **Review Changes** screen appears.

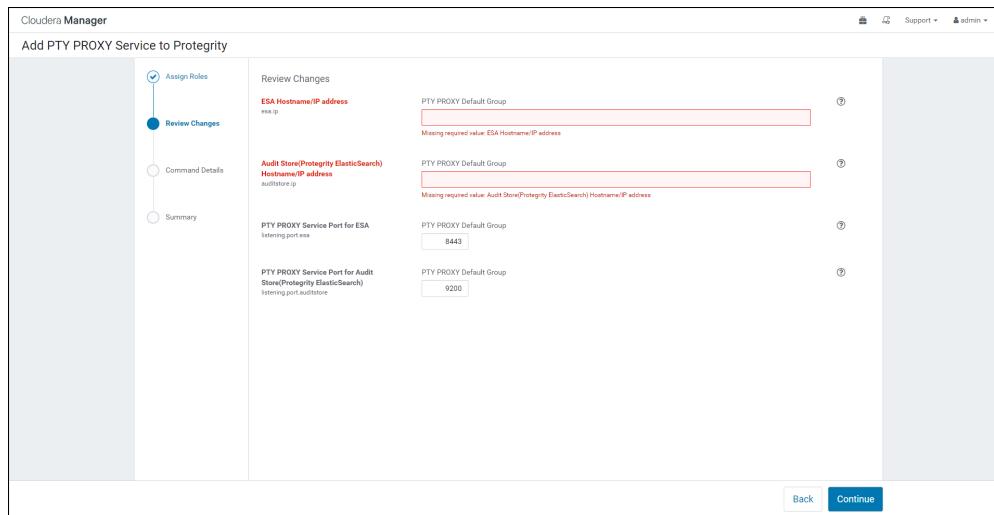


Figure 12-144: Review Changes for PTY PROXY Service Page

9. Enter the IP address of the Audit Store in the **Audit Store Hostname/IP address** box.

Note: Ensure that you provide the IP address or hostname of the host running the *PTY PROXY* service in the *ESA IP Address* prompt instead of the *ESA IP address* or *hostname*, that appears when the *BDP PEP* service is added to the nodes in the cluster.

10. Enter the IP address of the ESA in the **ESA Hostname/IP address** box.

Note: Replace the *ESA IP address* or *hostname* with the *IP address* or *hostname* of the host running the *PTY PROXY* service in the *Configurations* section

11. Click **Continue**.

The *PTY PROXY* service is started on the required node in the cluster.

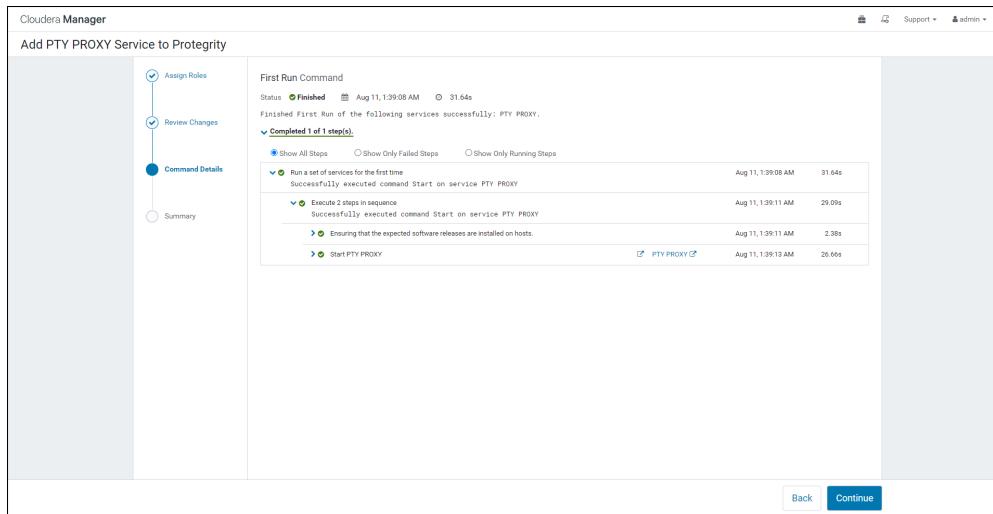


Figure 12-145: PTY PROXY Service screen

12. Click Continue.

The **Summary** page appears.

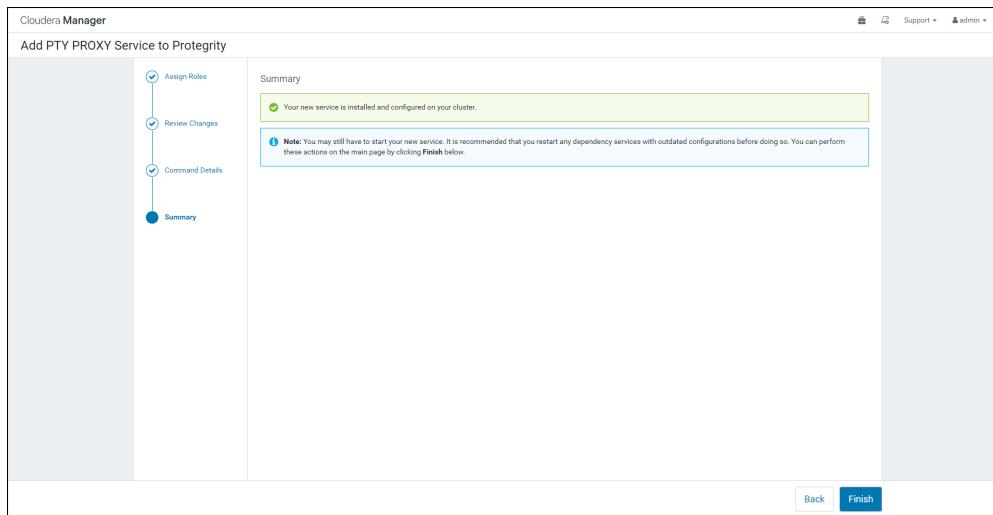


Figure 12-146: Summary Page

13. Click Finish.

The Cloudera Manager Home screen appears and the *PTY PROXY* service is added and started on all the nodes in the cluster.

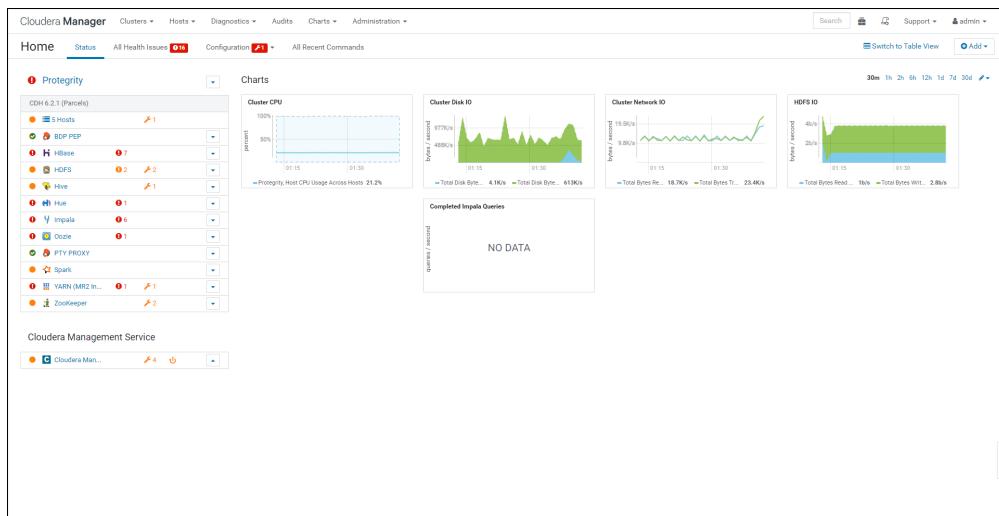


Figure 12-147: PTY PROXY Service in Cluster

Note: The *PTY Proxy* only supports forwarding log traffic to a single Protegity ElasticSearch Appliance IP address and port.

12.6.2.10.3 Configuring the Proxy Service when the Big Data Protector PEP Service is not Added and Started

► To configure the Proxy service when the Big Data Protector PEP service is not added and started:

1. After the *PTY Proxy* service is added and started as described in the previous task, add and start the Big Data Protector PEP service.

Note: For more information about adding and starting the Big Data Protector PEP service, refer to section [Starting the Big Data Protector PEP Service](#).

2. Ensure that you provide the IP address or hostname of the host running the *PTY PROXY* service in the ESA IP Address prompt instead of the ESA IP address or hostname, that appears when the BDP PEP service is added to the nodes in the cluster.

12.6.2.10.4 Configuring the Proxy Service when the Big Data Protector PEP Service is Added and Running

► To configure the Proxy service when the Big Data Protector PEP service is added and running:

1. After the *PTY Proxy* service is added and started as described in the previous task, navigate to the *Configurations* section of the BDP PEP service.
2. Replace the ESA IP address or hostname with the IP address or hostname of the host running the *PTY PROXY* service in the *Configurations* section.
3. Restart the BDP PEP service.

12.6.2.10.5 Installing Big Data Protector Parcel on a New Node

If you need to install Big Data Protector on a new node in an existing Hadoop cluster, then it requires distribution and activation of the following parcels on the new node:

- PTY_BDP
- PTY_CERT
- PTY_FLUENTBIT_CONF

The distribution and activation of the Big Data Protector parcels is done by the Cloudera framework.

In addition, ensure that the *PTY_Pepserver* and the *PTY Log Forwarder* roles, that are part of the BDP PEP service, are added to the new node.

12.6.2.11 Managing the Big Data Protector on Cloudera Manager

You can manage the Big Data Protector using the Cloudera Manager by modifying the PEP server and log forwarder configuration parameters in the *BDP PEP* service.

12.6.2.11.1 Updating the Configuration Parameters for the BDP PEP Service

You can modify the following PEP server configuration parameters for the *BDP PEP* service:

Table 12-43: Configuration Parameters for the BDP PEP Service

| Parameter | Description | Possible Values |
|---|--|--|
| Protegility ElasticSearch List of Hostnames/IP Addresses and/or Ports | Comma-delimited List of Protegility ElasticSearch appliance Hostnames/IP addresses and/or Ports where Fluent Bit sends the logs. | Allowed Syntax: hostname[:port] [,hostname[:port],hostname[:port]...] (By default 9200 is set for empty ports) |
| Fluent Bit Log Level | The Log Forwarder (Fluent Bit) logging verbosity level | <ul style="list-style-type: none"> • error • Warning • Info • debug |
| ESA hostname/IP Address | Specify the ESA hostname or IP address where the Log Forwarder sends the logs. If the <i>PTY Proxy</i> service is used, then it should be the hostname or the IP address of the <i>PTY Proxy</i> node. | Host name or IP address |
| Logging Level | Specify the level of details for the PEP server logs. | <ul style="list-style-type: none"> • OFF - No logging • Severe • Warning • Info • Config • ALL (Default) |
| Logging Mode | If the connection to the Fluent Bit is lost, then set how the logs must be handled. This setting is applicable only for the protector logs and not for the application logs. | <ul style="list-style-type: none"> • drop • error |

Note:

- For more information about the PEP Server Configuration file, refer to the section [PEP Server Configuration File](#).
- The Shared Memory is world-readable by default. To change the permissions of the Shared memory, add the following Shared Memory management parameters in the *PTY PepServer Advanced Configuration Snippet (Safety Valve)* for *pepper:cfg* parameter:
 - *sharedmemory.groupname=ptyituser*
 - *sharedmemory.worldreadable=no*
- Ensure that all the service and system users that are trying to access the shared memory are added to the *ptyitusr* group in all the nodes. Otherwise, it may cause operation failure.



- The *ptyitusr* is a sample group used only for representational purposes. The user can use their own group name instead.

12.6.2.11.2 Updating the Configuration Parameters for the PTY Proxy Service

You can modify the following configuration parameters for the *PTY Proxy* service:

Table 12-44: Configuration Parameters for the PTY Proxy Service

| Parameter | Description | Possible Values |
|---|--|-------------------------|
| Audit Store (ElasticSearch) hostname/IP Address | Specify the host name or IP address where the Log Forwarder sends logs. If the <i>PTY Proxy</i> service is used, then it should be the host name or the IP address of the <i>PTY Proxy</i> node. | Host name or IP address |
| ESA hostname/IP Address | Specify the ESA host name or IP address where the Log Forwarder sends logs. If the <i>PTY Proxy</i> service is used, then it should be the host name or the IP address of the <i>PTY Proxy</i> node. | Host name or IP address |

12.6.2.12 Updating the Certificates Parcel

If customers have updated the certificates in the ESA, with which the Big Data Protector is configured, then the Certificates parcel would need to be updated with the new certificates. The updated Certificates parcel needs to be utilized by the nodes in the cluster.

► To utilize updated Certificates:

1. Login to the host machine, which contains the Big Data Protector configurator script.
2. Run the *BDPConfigurator_CDH-6.2_9.1.0.0.x.sh* script.
A prompt to continue the configuration of Big Data Protector appears.

```
./BDPConfigurator_CDH-6.2_9.1.0.0.x.sh
*****
***** Welcome to the Big Data Protector Configurator Wizard *****
***** This will setup the Big Data Protector Installation Files for CDH.

Do you want to continue? [yes or no]
```

3. To start configuration of Big Data Protector, type *yes*.
4. Press ENTER.

The following prompt appears.

```
Please select the type of Installation files you want to generate?
[ 1: Create All ] : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. To update the ESA certificates in the *PTY_CERT* parcel, type *2*.



6. Press ENTER.

The prompt to select the version of the operating system appears.

```
Please select the OS version for Cloudera Manager Parcel.  
This will be used as the OS Distro suffix in the Parcel name.
```

```
[ 1: el6 ] : RHEL 6 and clones (CentOS, Scientific Linux, etc)  
[ 2: el7 ] : RHEL 7 and clones (CentOS, Scientific Linux, etc)  
[ 3: sles12 ] : SuSE Linux Enterprise Server 12.x
```

```
Please enter the no.:
```

7. Depending on the requirements, type *1*, *2*, or *3* to select the OS version for the Big Data Protector parcels.

8. Press ENTER.

The prompt to enter the ESA Hostname or IP address appears.

```
Please enter the ESA Host or IP Address[]:
```

9. Enter the ESA Hostname or IP address.

10. Press ENTER.

The prompt to enter the ESA host listening port appears.

```
Enter ESA host listening port [8443]:
```

11. If you want to use the default value of the ESA host listening port, which is *8443*, then press ENTER.

12. If you have configured an external proxy having connectivity with the ESA to download the certificates and password binaries from the ESA, then enter the external Proxy listening port.

13. Press ENTER.

The prompt to enter the ESA user name appears.

```
Enter ESA User name :
```

14. Enter the ESA user name.

15. Press *ENTER*.

The prompt to enter the password for the ESA appears.

```
Enter host password for user '<username>':
```

16. Enter the ESA administrator password.

17. Press ENTER.

The prompt to enter the version of the activated *PTY_CERT* parcel appears.

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current |
|-----------|------------|---------|---------------|------|--------|--------|---------|
| Dload | Upload | Total | Spent | Left | Speed | | |
| 100 20480 | 100 20480 | 0 0 | 23346 | 0 | --::-- | --::-- | 23325 |

```
-----  
Generating Installation files...
```

NOTE:

You can verify the version of the activated *PTY_CERT* parcel from the parcel name, such as *PTY_CERT-x.x.x.x_CDHx.x.p<version>-<os>.parcel*, where the <version> parameter denotes the patch version of the *PTY_CERT* parcel.

For Example: If the current activated *PTY_CERT* parcel is *PTY_CERT-x.x.x.x_CDHx.x.p0-<os>.parcel*, the patch version of the *PTY_CERT* parcel will be 0. Please do NOT include 'p' while specifying the version.

```
Enter the <version> of the current PTY_CERT Parcel as specified in the parcel name [0]:
```

You can verify the version of the activated *PTY_CERT* parcel from the parcel name, such as *PTY_CERT-x.x.x.x_CDHx.x.p<version>-<OS_Version>.parcel*, where the <version> parameter denotes the patch version of the *PTY_CERT* parcel.

For instance, if the name of the current activated *PTY_CERT* parcel is *PTY_CERT-9.1.0.0.x_CDHx.x.p0-<OS_Version>.parcel*, then the patch version of the *PTY_CERT* parcel is *0*. When specifying the patch version of the parcel, do not include the character *p*.

18. Enter the current activated patch version of the *PTY_CERT* parcel.

19. Press ENTER.

The updated *PTY_CERT* parcel *PTY_CERT-9.1.0.0.x_CDHx.x.p<updated version>-<OS_Version>.parcel* is generated in the *./Installation_Files*/directory.

```
The updated PTY_CERT parcel 'PTY_CERT-9.1.0.0.x_CDH6.2.p1-el7.parcel' is generated in ./  
Installation_Files/ directory.
```

NOTE:

```
Copy PTY_CERT-9.1.0.0.x_CDH6.2.p1-el7.parcel and .sha files to Cloudera Manager local  
parcel repository.
```

20. Copy the new Certificate parcel to the local parcel repository of the Cloudera Manager.

Note: The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.

21. Navigate to the local parcel repository directory.

22. To assign ownership permissions for the Cloudera SCM to the new Certificate parcel and the checksum file, run the following command.

```
chown cloudera-scm:cloudera-scm PTY_*
```

23. Press ENTER.

24. To set 640 permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```

25. Press ENTER.

The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.

26. Login to the Cloudera Manager web interface.

27. Navigate to the **Parcels** page.

The **Parcels** page appears.

28. To fetch the updated parcels, click **Check for New Parcels**.

The Cloudera Manager will fetch the updated *PTY_CERT* parcel.

29. Distribute the new Certificate parcel to the nodes.

Note: For more information about distributing the new Certificate parcel, refer to section [Distributing the Big Data Protector Parcels to the Nodes](#).

30. Activate the new Certificate parcel on the nodes.

Note: For more information about activating the new Certificate parcel, refer to section [Activating the Big Data Protector Parcel on the Nodes](#).

12.6.2.13 Updating the Fluent Bit Parcel

If the user wants to use a newer set of custom Fluent Bit configuration files for sending logs to an External Audit Store, then the Fluent Bit parcel must be updated, distributed, and activated across the cluster nodes through the Cloudera Manager.

- To update the Fluent Bit parcel:

1. Login to the host machine, which contains the Big Data Protector configurator script.
2. Run the *BDPConfigurator_CDH-6.2_9.1.0.0.x.sh* script.

A prompt to continue the configuration of Big Data Protector appears.

```
./BDPConfigurator_CDH-6.2_9.1.0.0.8.sh
*****
        Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDH.

Do you want to continue? [yes or no]
```

3. To start the configuration of the Big Data Protector, type *yes*.
4. Press ENTER.

The following prompt appears.

```
Please select the type of Installation files you want to generate?
[ 1: Create All ]      : Creates entire Big Data Protector CSDs and Parcels.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. To update the Fluent Bit parcel, type *3*.
6. Press ENTER.

The following prompt appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.
```

```
[ 1: el6 ]      : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ]   : SuSE Linux Enterprise Server 12.x
```

```
Please enter the no.:
```

7. Depending on the requirements, type *1*, *2*, or *3* to select the OS version for the Big Data Protector parcels.
8. Press ENTER.

A prompt to enter the local directory path that stores the Fluent Bit configuration files appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration
files for External Audit Store:
```

9. Type the local directory path that stores the Fluent Bit configuration files.
10. Press ENTER.

A prompt to enter the current version of the Fluent Bit configuration parcel appears.

```
Generating Installation files...
```

NOTE:

You can verify the version of the activated PTY_FLUENTBIT_CONF parcel from the parcel name, such as PTY_FLUENTBIT_CONF-x.x.x.x.p<version>-<os>.parcel, where the <version> parameter denotes the patch version of the PTY_FLUENTBIT_CONF parcel.

For Example: If the current activated PTY_FLUENTBIT_CONF parcel is PTY_FLUENTBIT_CONF-x.x.x.x.CDHx.x.p0-<os>.parcel, the patch version of the PTY_FLUENTBIT_CONF parcel will be 0. Please do NOT include 'p' while specifying the version.

Enter the <version> of the current PTY_FLUENTBIT_CONF Parcel as specified in the parcel name [0]:

11. Type the version of the Fluent Bit configuration parcel.

12. Press ENTER.

The installer generates the *PTY_FLUENTBIT_CONF* parcel in the *./Installation_Files/* directory.

The updated PTY_FLUENTBIT_CONF parcel 'PTY_FLUENTBIT_CONF-9.1.0.0.x_CDH6.2.p1-el7.parcel' is generated in *./Installation_Files/* directory.

NOTE:

Copy PTY_FLUENTBIT_CONF-9.1.0.0.x_CDH6.2.p1-el7.parcel and .sha files to Cloudera Manager local parcel repository.

13. Copy the new *PTY_FLUENTBIT_CONF* parcel to the local parcel repository of Cloudera Manager.

The default local parcel repository for Cloudera Manager is located in the */opt/cloudera/parcel-repo/* directory.

14. Navigate to the local parcel repository directory.

15. To assign ownership permissions for the Cloudera SCM to the new Certificate parcel and the checksum file, run the following command.

```
chown cloudera-scm:cloudera-scm PTY_*
```

16. Press ENTER.

17. To set 640 permissions to the parcel files, run the following command.

```
chmod 640 PTY_*
```

18. Press ENTER.

The command assigns read and write permissions to the owner, read permissions to the group, and restricts access to all other users.

19. Login to the Cloudera Manager web interface.

20. Navigate to the **Parcels** page.

The **Parcels** page appears.

21. To fetch the updated parcels, click **Check for New Parcels**.

The Cloudera Manager will fetch the updated *PTY_FLUENTBIT_CONF* parcel.

22. Distribute the new *PTY_FLUENTBIT_CONF* parcel to the nodes.

Note: For more information about distributing the new *PTY_FLUENTBIT_CONF* parcel, refer to section [Distributing the Big Data Protector Parcels to the Nodes](#).

23. Activate the new *PTY_FLUENTBIT_CONF* parcel on the nodes.

Note: For more information about activating the new *PTY_FLUENTBIT_CONF* parcel, refer to section [Activating the Big Data Protector Parcel on the Nodes](#).

12.6.2.14 Generating the *pepserver.log* File

By default, the *pepserver.log* file has been deprecated from the architecture starting from the Big Data Protector version 8.1.0.0. However, you can configure the log forwarder to generate the *pepserver.log* file.

Note:

- The process of generating the *pepserver.log* file is an optional step.
- If you do not want to generate the *pepserver.log* file, then all the log entries will be populated in the Protegility Storage Unit.

► To generate the *pepserver.log* file:

1. To install the Big Data Protector, perform the steps mentioned in the *Protegility Installation Guide 9.1.0.2*.
The solution mentioned in this section is compatible with the Big Data Protector, build version 9.1.0.0.8 and later.
2. Login to the Master node.
3. On the Master node, create a directory.
For example, *~/pepserver_log_confs/*
4. Navigate to the *~/pepserver_log_confs/* directory.
5. Create the *out_file_pepserver_log.conf*file.
6. Add the following content in the *out_file_pepserver_log.conf*file.

```
[FILTER]
Name rewrite_tag
Match logdata
Rule $logtype ^(Application)$ applog true
Emitter_Name re_emitted

[FILTER]
Name lua
Match applog
script config.d/restructure.lua
call restructure

[OUTPUT]
Name file
Match applog
Path /opt/cloudera/parcels/PTY_BDP/defiance_dps/data
File pepserver.log
Format template
Template {current_time} ({level}) {description}
storage.total_limit_size 5M
```

Note:

- When you add the content in the *out_file_pepserver_log.conf*file, ensure to maintain the spaces, indentation, and the blank lines as per the sample.
- If you want to generate the *pepserver.log* file in any other directory, then set the required location in the *Path* parameter in the *[OUTPUT]* section.
- Set the file size limit for the *pepserver.log* file depending on your requirements.

7. Save the changes to the *out_file_pepserver_log.conf*file.
8. In the *~/pepserver_log_confs/* directory, create the *restructure.lua* file.
9. Add the following content in the *restructure.lua* file.

```
-- A lua script that restructures the json to non-nested json structure
-- and converts the time.

function restructure(tag, timestamp, record)
new_record = {}

current_time = os.date('%Y-%m-%d %H:%M:%S', record["origin"]["time_utc"])

new_record["level"] = record["level"]
new_record["current_time"] = current_time
new_record["description"] = record["additional_info"]["description"]
```



```

        return 2, timestamp, new_record
end

```

Note: When you add the content in the *restructure.lua* file, ensure to maintain the spaces, indentation, and the blank lines as per the sample.

- On the Master node, run the configurator script to generate the *PTY_FLUENTBIT_CONF* parcel.

Note: Perform the steps from 1 to 7 as mentioned in the section [Updating the Fluent Bit Parcel](#).

- While generating the *PTY_FLUENTBIT_CONF* parcel, the configurator script will prompt for the directory path that contains the Fluent-Bit configuration files for the external audit store.

Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:

Note: If you have already configured an external audit store to save the logs, then ensure that you save the *out_file_pepserver_log.conf* and the *restructure.lua* files in the same directory that contains the Fluent-Bit configuration files.

- Enter the path of the directory that contains the *out_file_pepserver_log.conf* and the *restructure.lua* files.
For example, *~/pepserver_log_confs/*.
- After you generate the *PTY_FLUENTBIT_CONF* parcel, perform the steps from 13 onwards as mentioned in the section [Updating the Fluent Bit Parcel](#).
- Navigate to the *BDP PEP* service configuration page.
- In the *BDP PEP* service configuration page, under **Audit Store Type**, select any one of the following options:

| Option | Description |
|--|--|
| External Audit Store | Enter the comma-separated IP/ports using the accurate syntax in the Protegility Elasticsearch List of Hostnames/IP Addresses and/or Ports box. Ensure that the <i>PTY_FLUENTBIT_CONF</i> parcel is distributed and activated. |
| Protegility Elastic Search + External Audit Store | Enter the comma-separated IP/ports using the correct syntax in the Protegility Elasticsearch List of Hostnames/IP Addresses and/or Ports box. |

- Save the changes to the *BDP PEP* service configuration.

- Restart the *BDP PEP* service.

The *BDP PEP* service will extract the *out_file_pepserver_log.conf* and *restructure.lua* files from the *PTY_FLUENTBIT_CONF* parcel to the */fluent-bit/data/config.d/* directory on all the nodes in the cluster.

```

[root@ip-      config.d]# pwd
/opt/cloudera/parcels/PTY_BDP/fluent-bit/data/config.d
[root@ip-      config.d]# ll
total 28
-rw-r----- 1 ptyitusr ptyitusrgroup 585 Dec  9 06:52 in_error_tail.conf
-rw-r----- 1 ptyitusr ptyitusrgroup 1517 Dec  9 06:52 in_tail.conf
-rw-r----- 1 ptyitusr ptyitusr     1570 Dec  9 06:17 in_tcp.conf
-rw-r----- 1 ptyitusr ptyitusr     2055 Dec  9 06:52 out_elastic.conf
-rw-r----- 1 ptyitusr ptyitusrgroup  474 Dec  9 06:31 out_file_pepserver_log.conf
-rw-r----- 1 ptyitusr ptyitusrgroup  375 Dec  9 06:31 restructure.lua
-rw-r----- 1 ptyitusr ptyitusr     841 Dec  9 06:52 upstream_es.cfg
[root@ip-      config.d]#

```

Figure 12-148: *out_file_pepserver_log.conf* and *restructure.lua* Files in the */config.d/* Directory

The logforwarder will generate the *pepserver.log* file in the */opt/cloudera/parcels/PTY_BDP/defiance_dps/data/* directory on all the nodes in the cluster.

```
[root@ip-... data]# ll
total 10044
-rw-r-----. 1 ptyitusr ptyitusr      5530850 May  2 2022 authesa.plm
-rw-r-----. 1 ptyitusr ptyitusrgroup    2000 Dec  9 05:43 CA.pem
-rw-r-----. 1 ptyitusr ptyitusrgroup    3434 Dec  9 05:43 cert.key
-rw-r-----. 1 ptyitusr ptyitusrgroup     28 Dec  9 05:43 certkeyup.bin
-rw-r-----. 1 ptyitusr ptyitusrgroup    1834 Dec  9 05:43 cert.pem
-rw-r--r--. 1 ptyitusr ptyitusrgroup    8192 Dec  9 06:17 error.pos
-rw-r--r--. 1 ptyitusr ptyitusrgroup   32768 Dec  9 06:45 error.pos-shm
-rw-r--r--. 1 ptyitusr ptyitusrgroup      0 Dec  9 06:45 error.pos-wal
-rw-r-----. 1 ptyitusr ptyitusr       4678325 May  2 2022 keyinternal.plm
-rw-r-----. 1 ptyitusr ptyitusrgroup     477 Dec  9 06:45 pepserver.cfg
-rw-r--r--. 1 ptyitusr ptyitusrgroup     45 Dec  9 06:45 pepserver.log
-rw-r-----. 1 ptyitusr ptyitusrgroup      6 Dec  9 06:45 pepserver.pid
[root@ip-... data]#
```

Figure 12-149: The *pepserver.log* File in the */opt/cloudera/parcels/PTY_BDP/defiance_dps/data/* Directory

Important:

- Log entries, over a period of time, will fill up the *pepserver.log* file.
- The logforwarder process will terminate when the *pepserver.log* file exceeds the allocated file size.
- It is recommended to monitor the *pepserver.log* file and rotate the log entries depending on your system configuration. Protegility does not support the rotation of log entries in the *pepserver.log* file.

12.6.2.15 Performing an Upgrade of the CDH Distribution

If you are performing a rolling upgrade of the CDH distribution, then you need to uninstall Big Data Protector before starting the rolling upgrade. After the rolling upgrade of the CDH distribution is completed, you need to install the Big Data Protector version that is compatible with the updated version of the CDH distribution.

If you are using CDH versions 6.3 and lower and need to upgrade the version of CDH, then perform the following steps.

12.6.2.15.1 Performing Rolling Upgrade of Minor Version of CDH

► To perform rolling upgrade of minor version of CDH:

1. Ensure that you do not invoke the Protegility APIs or UDFs or the Protegility HBase coprocessor will fail during the rolling upgrade to a minor version of CDH.
2. Perform the rolling restart of the stale services to deploy the recommended service configurations.
3. Perform the rolling upgrade of the required CDH minor version.
4. Perform the rolling restart of the stale services to deploy the recommended service configurations.

12.6.2.16 Uninstalling the Big Data Protector from all the Nodes

► To uninstall Big Data Protector from all the nodes:

1. Remove the following Big Data Protector-related services from all the nodes in the cluster:
 - a. BDP PEP service
 - b. PTY Proxy service, only if it is enabled
2. Deactivate the Big Data Protector parcels from all the nodes in the cluster.

3. Remove the Big Data Protector parcels from all the nodes in the cluster.
4. Delete the Big Data Protector parcels from the Cloudera Manager local repository.

12.6.2.16.1 Removing the Big Data Protector Services from all Nodes

Before you deactivate the Big Data Protector parcels from all the nodes in the cluster, you must remove the Big Data Protector-related services from all the nodes.

► To remove the Big Data Protector Services from all Nodes in the Cluster:

1. On the Cloudera Manager Home screen, besides the BDP PEP service, click . The **BDP PEP Actions** menu appears.

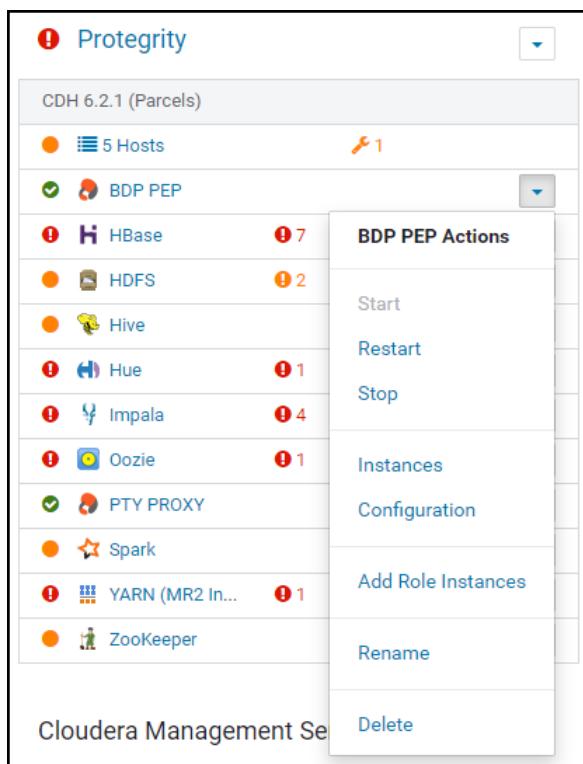


Figure 12-150: BDP PEP Actions Menu

2. To stop the BDP PEP service, select **Stop**. The prompt to confirm the termination of the *BDP PEP* service appears.

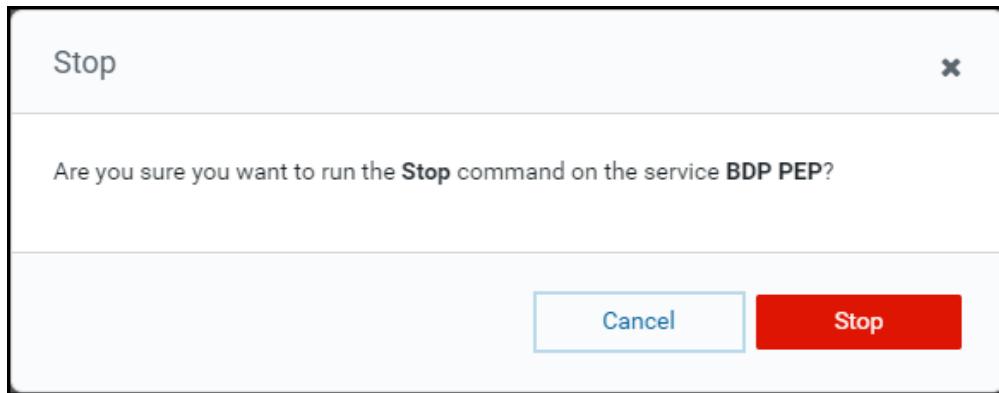


Figure 12-151: Confirmation Dialog Box for Stopping the BDP PEP Service

3. Click **Stop**.

The BDP PEP service is terminated and the following page appears.

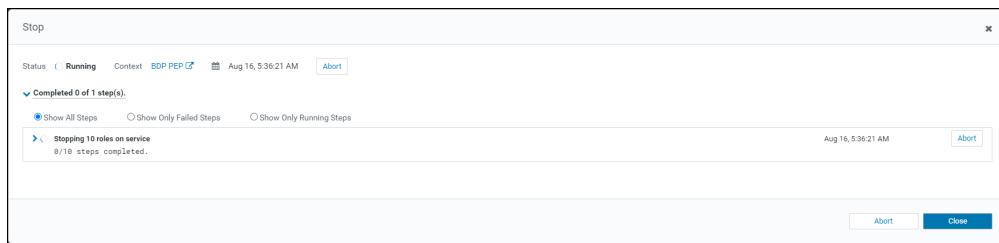


Figure 12-152: BDP PEP Service Stopped

4. Click **Close**.

The BDP PEP service is stopped and the status is updated in the Cloudera Manager.

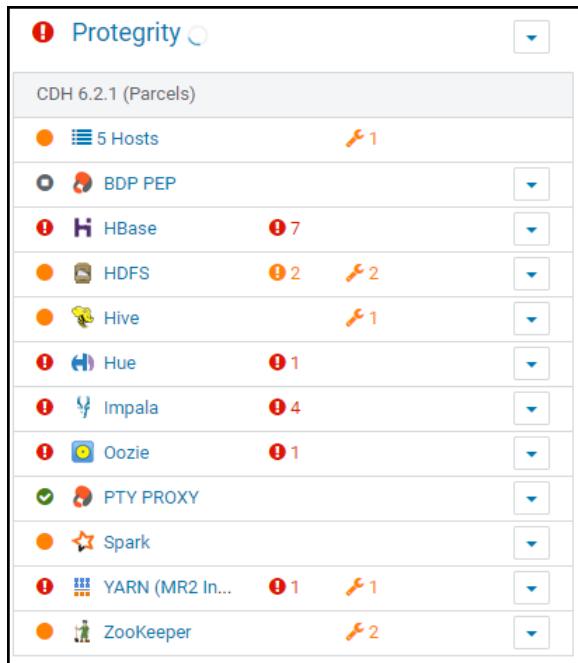


Figure 12-153: BDP PEP Service Stopped in the Cluster

5. On the Cloudera Manager Home screen, besides the BDP PEP service, click . The **BDP PEP Actions** menu appears.

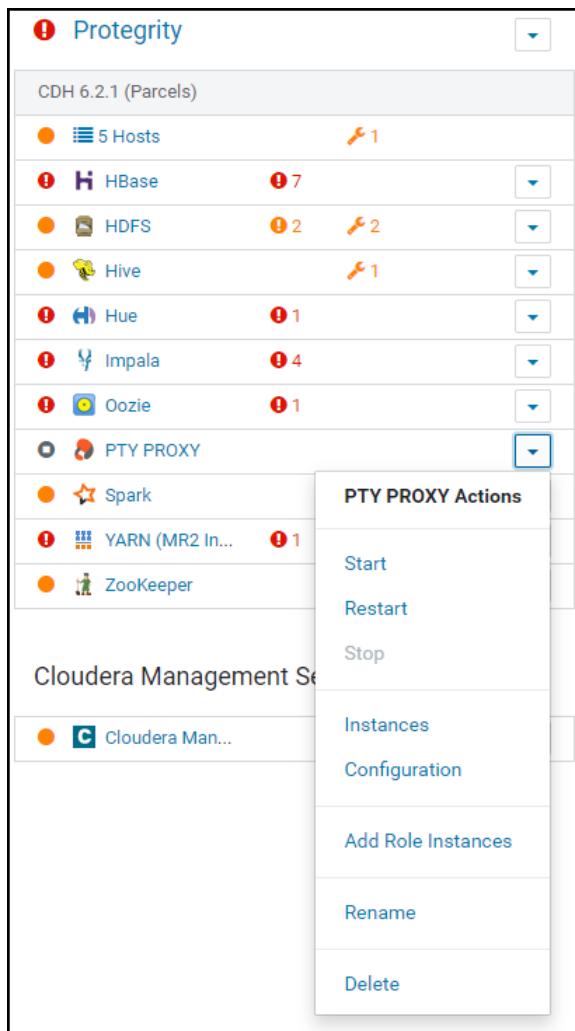


Figure 12-154: BDP PEP Actions Menu

6. Select **Delete**.
A prompt to confirm the deletion of the BDP PEP service appears.

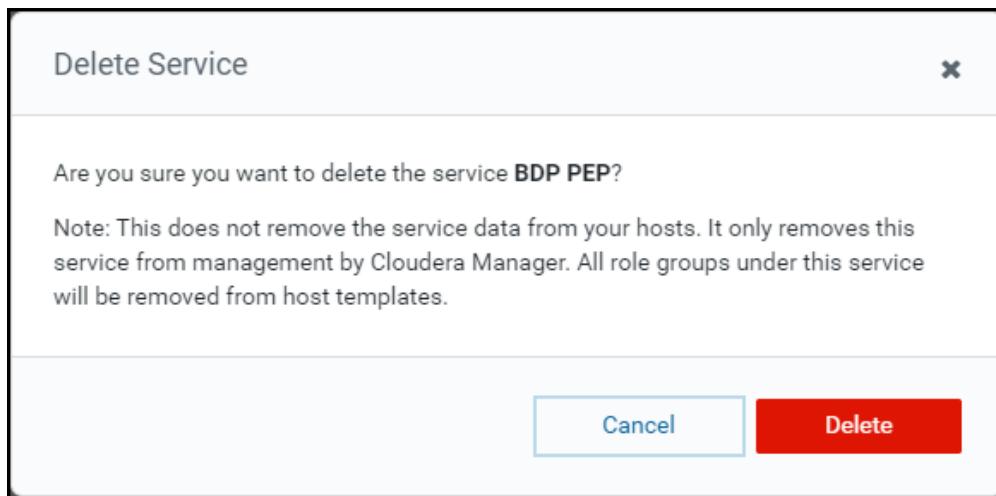


Figure 12-155: Prompt to Delete the BDP PEP Service

7. Click **Delete**.
The BDP PEP service is removed from all the nodes in the cluster.
8. If the *PTY PROXY* service is installed and running, the besides the service, click ▾.

The **PTY PROXY Actions** menu appears.

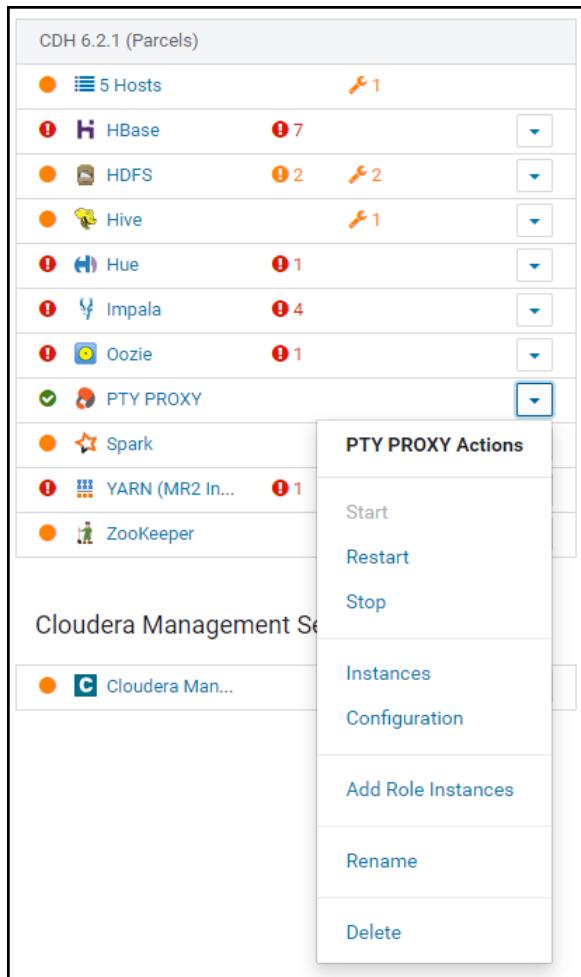


Figure 12-156: PTY PROXY Actions Menu

9. Select **Stop**.
A prompt to confirm the termination of the *PTY PROXY* service appears.

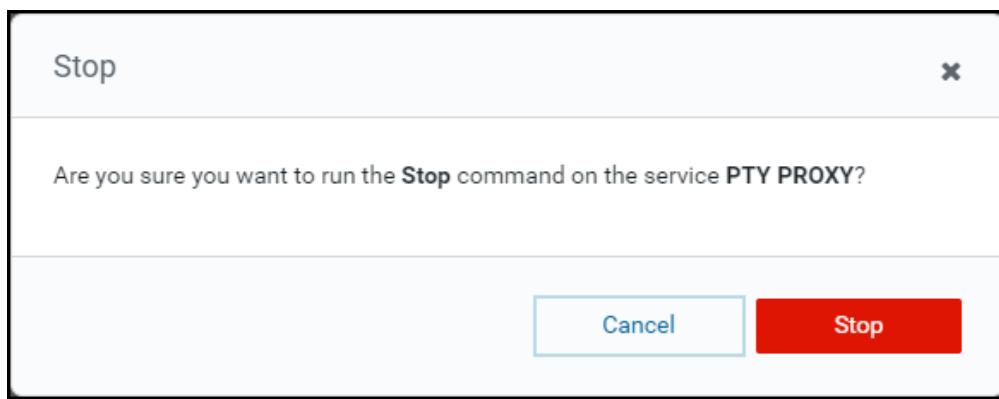


Figure 12-157: Confirmation Dialog Box for Terminating the PTY PROXY Service

10. Click **Stop**.
The *PTY PROXY* service is terminated.
11. Click **Close**.
The *PTY PROXY* service is stopped and the status is updated.

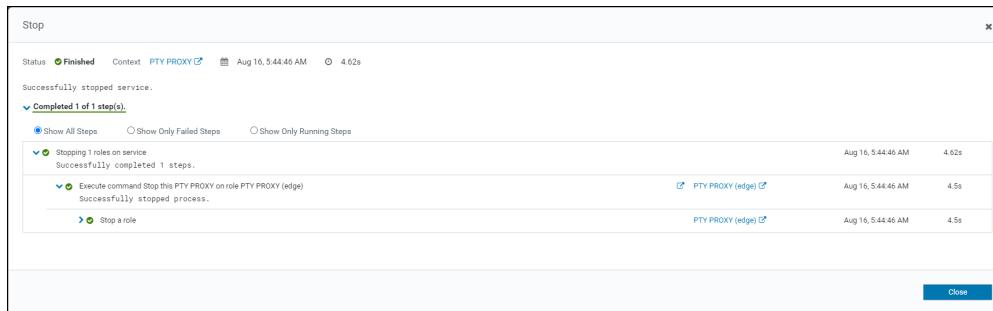


Figure 12-158: PTY PROXY Service Stopped

- On the Cloudera Manager screen, besides the *PTY PROXY* service, click . The **PTY PROXY Actions** menu appears.

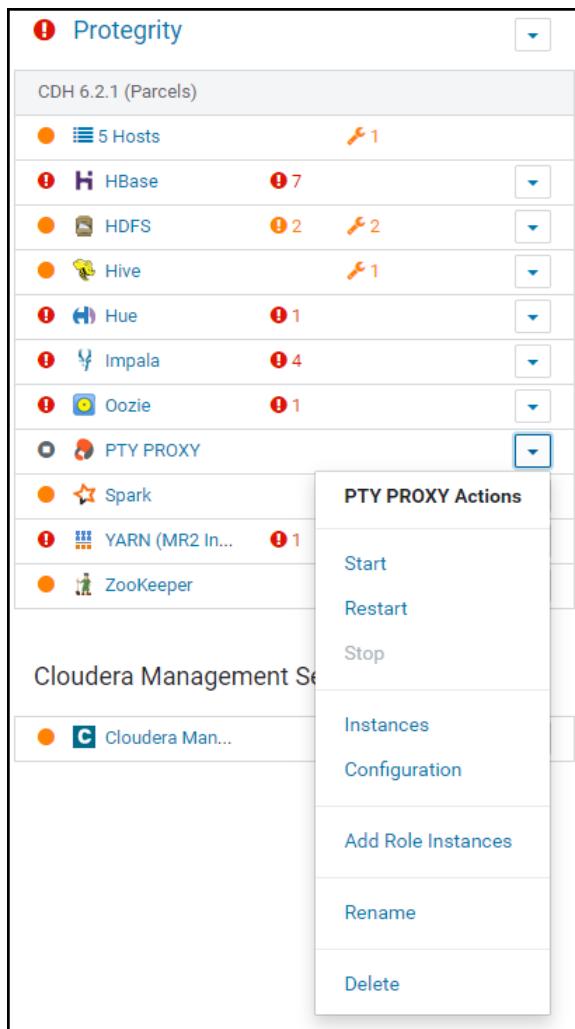


Figure 12-159: PTY PROXY Actions Menu

- Select **Delete**. A prompt to confirm the deletion of the *PTY PROXY* service appears.

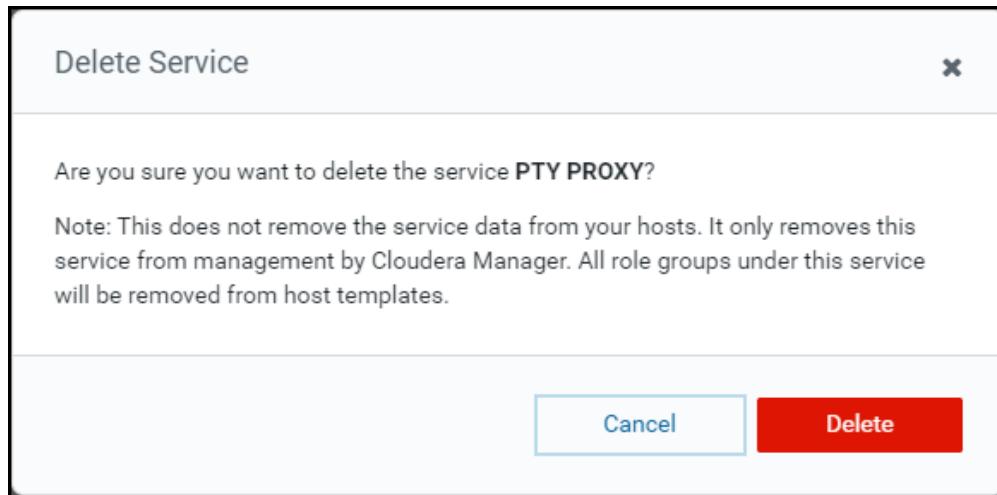


Figure 12-160: Prompt to Delete the PTY PROXY Service

- Click **Delete**.
The **PTY PROXY** service is removed from all the nodes in the cluster.

A screenshot of the Protegility interface showing a list of services and their status. The services listed are:

| Service | Status | Issues | Actions |
|---------------------|---------|--------|---------|
| CDH 6.2.1 (Parcels) | Normal | 0 | ▼ |
| 5 Hosts | Normal | 1 | ▼ |
| HBase | Warning | 7 | ▼ |
| HDFS | Warning | 2 | ▼ |
| Hive | Normal | 1 | ▼ |
| Hue | Warning | 1 | ▼ |
| Impala | Warning | 4 | ▼ |
| Oozie | Warning | 1 | ▼ |
| Spark | Normal | 0 | ▼ |
| YARN (MR2 In... | Warning | 1 | ▼ |
| ZooKeeper | Normal | 2 | ▼ |

Figure 12-161: Big Data Protector Services Removed from the Nodes in the Cluster

12.6.2.16.2 Deactivating the Big Data Protector Parcels from all the Nodes

After you remove the Big Data Protector-related services from all the nodes in the cluster, you must de-activate the Big Data Protector parcels from all the nodes.

- To de-activate the Big Data Protector Parcels from all Nodes in the Cluster:

- From the top navigation pane, navigate to **Clusters > Parcels**.

| Location | Parcel Name | Version | Status | Action |
|--------------------------------|-----------------------------|---------------------------------|------------------------|-----------------------------|
| Protegility Available Remotely | ACUMULO | 1.7.2.6.5.0.ACUMULO105.5.0.p0.8 | Available Remotely | <button>Download</button> |
| Filters | CDH 6 | 6.3.2.1.cdh6.3.2.p0.1603554 | Downloaded | <button>Distribute</button> |
| ▼ PARCEL NAME | 6.2.1-1.cdh6.2.1.p0.1425774 | | Distributed, Activated | <button>Deactivate</button> |
| | KAFKA | 5.16.2-1.cdh5.16.2.p0.8 | Available Remotely | <button>Download</button> |
| | KUDU | 4.1.0-1.4.1.p0.4 | Available Remotely | <button>Download</button> |
| | PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Downloaded | <button>Distribute</button> |
| | PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Downloaded | <button>Distribute</button> |
| | PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Downloaded | <button>Distribute</button> |
| SQOOP_NETEZZA_CONNECTOR | SQOOP_NETEZZA_CONNECTOR | 1.5.1cb | Available Remotely | <button>Download</button> |
| mkl | | 1.5.1cb | Available Remotely | <button>Download</button> |
| ▼ STATUS | SQOOP_TERADATA_CONNECTOR | 1.7.0cb | Available Remotely | <button>Download</button> |
| Distributed | mkl | 2022.0.2.136 | Available Remotely | <button>Download</button> |
| Other | | | | |

Figure 12-162: Cloudera Manager Parcels Page

The following Protegility parcels appear on the Parcels screen:

- PTY_BDP**: Big Data Protector parcel
- PTY_CERT**: Certificates parcel
- PTY_FLUENTBIT_CONF**: Fluent Bit configuration parcel

Note: The **PTY_FLUENTBIT_CONF** Fluent Bit configuration parcel will be visible only if you have selected it during installation.

- To deactivate the Fluent Bit configuration parcel, besides the **PTY_FLUENTBIT_CONF** parcel, click **Deactivate**.

| | | | |
|--------------------|---------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |

Figure 12-163: Protegility Parcels Activated

A prompt to confirm the deactivation of the parcel appears.

Deactivate PTY_FLUENTBIT_CONF 9.1.0.0.8_CDH6.2.p0 on Protegility

Are you sure?

Cancel
OK

Figure 12-164: Protegility Parcel Deactivation Confirmation

- To deactivate the **PTY_FLUENTBIT_CONF** parcel, click **OK**.
- To deactivate the certificates parcel, besides the **PTY_CERT** parcel, click **Deactivate**.

| | | | |
|--------------------|---------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Distributed | <button>Activate</button> |

Figure 12-165: Protegility Parcels Activated



A prompt to confirm the deactivation of the parcel appears.

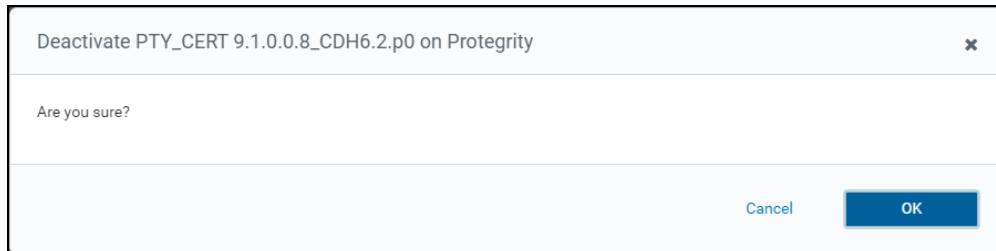


Figure 12-166: Prompt to Deactivate the Certificates Parcel

- To deactivate the **PTY_CERT** parcel, click **OK**.

After the Protegity parcels are deactivated from the nodes in the cluster, their status on the Parcels screen is updated to **Distributed** and the **Activate** button appears.

| | | | |
|--------------------|---------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Distributed | <button>Activate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Distributed | <button>Activate</button> |

Figure 12-167: Protegity Parcels Deactivated

- To deactivate the Big Data Protector parcel, besides the **PTY_BDP** parcel, click **Deactivate**.

| | | | |
|--------------------|---------------------|------------------------|-----------------------------|
| PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Distributed, Activated | <button>Deactivate</button> |
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Distributed | <button>Activate</button> |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Distributed | <button>Activate</button> |

Figure 12-168: Protegity Parcels Activated

A prompt to confirm the deactivation of the parcel and restart of the dependent services appears.

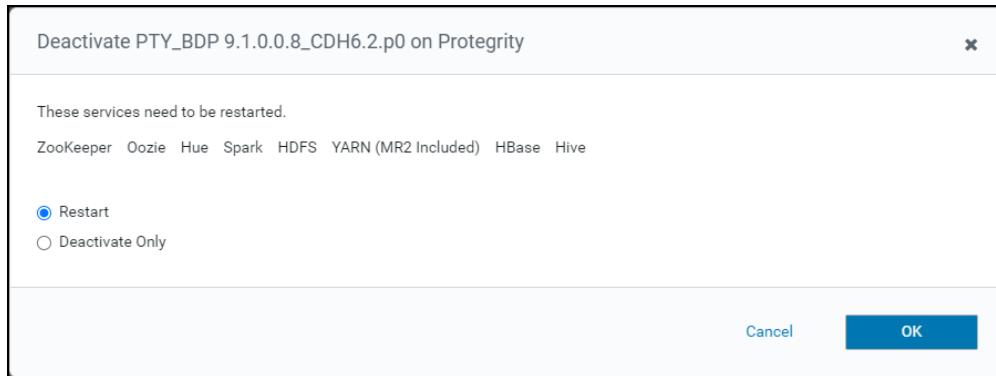


Figure 12-169: Protegity Parcel Deactivation Confirmation

- To restart the services, which are dependent on the parcel that needs to be deactivated, select **Restart**. Alternatively, to just deactivate the parcel, select **Deactivate Only**.

Note: You can restart the dependent services later also. However, it is recommended to restart the dependent services immediately. This will ensure that the dependent services do not utilize the parcel that is being deactivated.

- To deactivate the Big Data Protector parcel, click **OK**.

Tip: Alternatively, to terminate the deactivation, click **Abort**.

The deactivation of the Big Data Protector parcel starts.



Figure 12-170: Deactivation of the Big Data Protector Parcel

- To complete the deactivation of the Big Data Protector parcel, click **Close**.



Figure 12-171: Big Data Protector Parcel Deactivated

After the *PTY_FLUENTBIT_CONF*, *PTY_CERT*, and *PTY_BDP* parcels are deactivated, their status on the **Parcels** page changes to **Distributed**, and the **Activate** option appears.

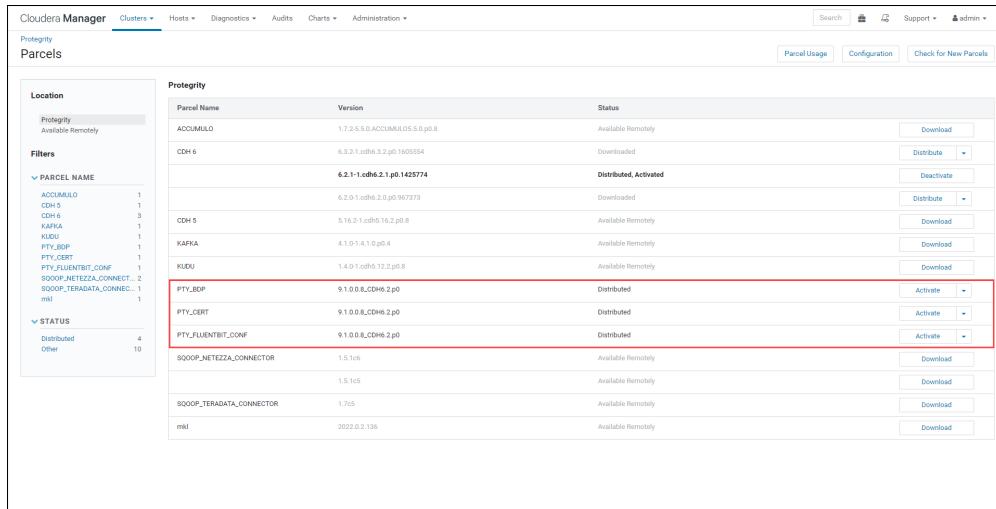


Figure 12-172: Big Data Protector Parcels Deactivated

12.6.2.16.3 Removing the Big Data Protector Parcels from all Nodes

After deactivating the Big Data Protector parcels from Cloudera Manager, you must remove the following Big Data Protector parcels from all the nodes:

- PTY_BDP*: Big Data Protector parcel
- PTY_CERT*: Certificates parcel
- PTY_FLUENTBIT_CONF*: Fluent Bit configuration parcel

► To remove the Big Data Protector Parcels from all the nodes in the Cluster:

- On the Cloudera Manager **Parcels** page, besides the Big Data Protector parcel, click . The drop-down menu appears.



| | | | |
|--------------------|-------------------|-------------|-------------------|
| PTY_BDP | 9.1.0.8_CDH6.2.p0 | Distributed | Activate |
| PTY_CERT | 9.1.0.8_CDH6.2.p0 | Distributed | Remove From Hosts |
| PTY_FLUENTBIT_CONF | 9.1.0.8_CDH6.2.p0 | Distributed | Activate |

Figure 12-173: Drop-down menu

- Select **Remove From Hosts**. The prompt to confirm the removal of the Big Data Protector parcel appears.

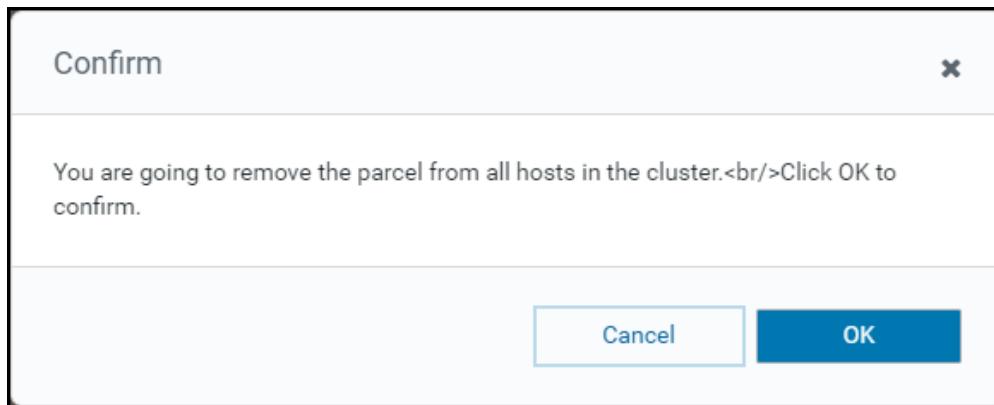


Figure 12-174: Confirmation Dialog Box for Removing the Protegility Parcel

- Click **OK**. The Big Data Protector parcel is removed from all the nodes in the cluster.
- Besides the Certificates parcel, click . The drop-down menu appears.



| | | | |
|--------------------|-------------------|-------------|-------------------|
| PTY_CERT | 9.1.0.8_CDH6.2.p0 | Distributed | Activate |
| PTY_FLUENTBIT_CONF | 9.1.0.8_CDH6.2.p0 | Distributed | Remove From Hosts |

Figure 12-175: Drop-down menu

- Select **Remove From Hosts**. The prompt to confirm the removal of the Certificates parcel appears.

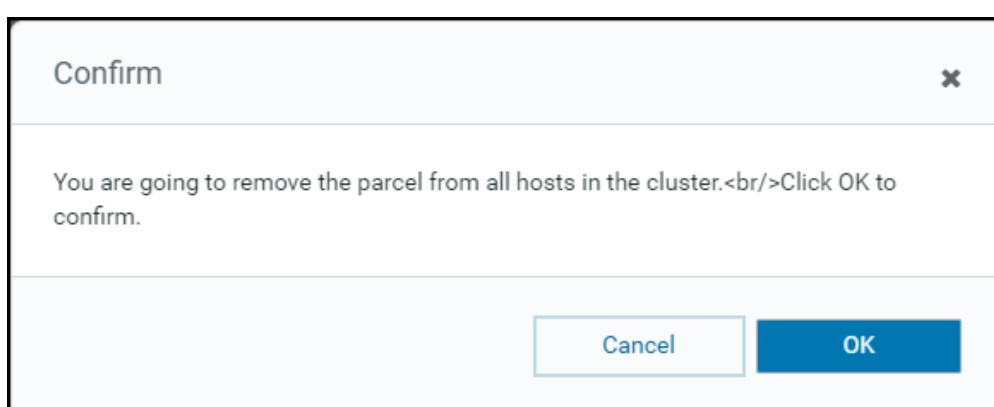


Figure 12-176: Confirmation Dialog Box for Removing the Protegility Parcel

- Click **OK**. The Certificates parcel is removed from all the nodes in the cluster.

7. Besides the Fluent Bit Configuration parcel, click . The drop-down menu appears.

| | | | |
|-------------------------|---------------------|--------------------|-------------------|
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Downloaded | Distribute |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Distributed | Activate |
| SQOOP_NETEZZA_CONNECTOR | 1.5.1c6 | Available Remotely | Remove From Hosts |

Figure 12-177: Drop-down menu

8. Select **Remove From Hosts**.

The prompt to confirm the removal of the Fluent Bit configuration parcel appears.

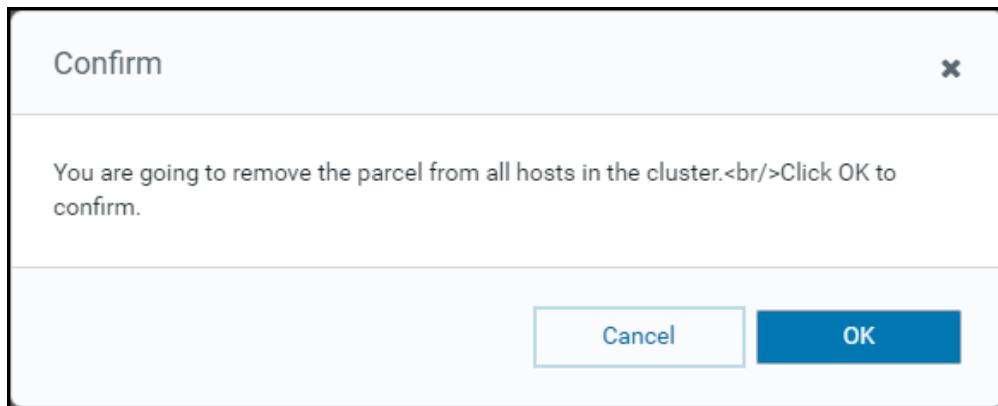


Figure 12-178: Prompt to Remove the Fluent Bit Configuration Parcel

9. Click **OK**.

The Fluent Bit configuration parcel is removed from all the nodes in the cluster.

12.6.2.16.4 Deleting the Big Data Protector Parcels from the Repository

After removing the Big Data Protector parcel from the nodes, you need to delete the Big Data Protector parcels from the local Cloudera Manager repository.

► To delete the Big Data Protector Parcels from the Local Repository:

1. On the Cloudera Manager **Parcels** page, besides the Big Data Protector parcel, click . The drop-down menu appears.

| | | | |
|---------|---------------------|------------|------------|
| PTY_BDP | 9.1.0.0.8_CDH6.2.p0 | Downloaded | Distribute |
| | | | Delete |

Figure 12-179: Drop-down Menu

2. Select **Delete**.

A dialog box to confirm the deletion of the Big Data Protector parcel appears.

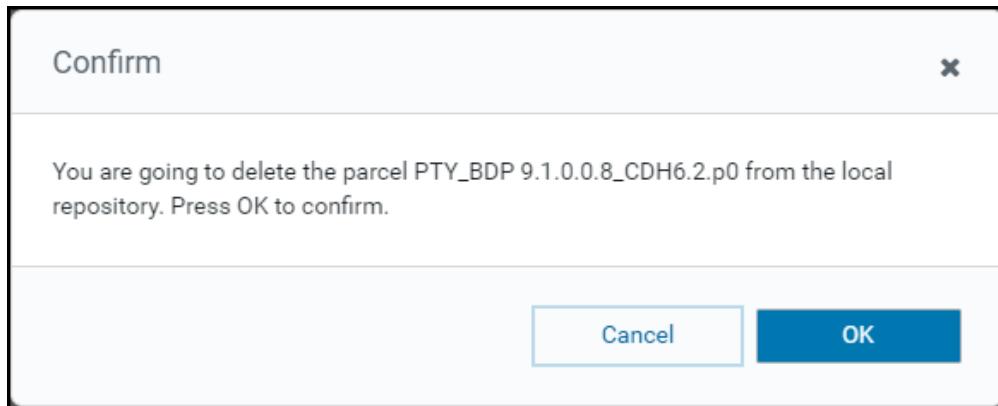


Figure 12-180: Confirmation Dialog Box for Deleting the Protegity Parcel

3. Click **OK**.
The Big Data Protector parcel is deleted from the local repository.

4. On the Cloudera Manager **Parcels** page, besides the Certificates parcel, click . The drop-down menu appears.

| | | | |
|--------------------|---------------------|------------|------------|
| PTY_CERT | 9.1.0.0.8_CDH6.2.p0 | Downloaded | Distribute |
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Downloaded | Delete |

Figure 12-181: Drop-down Menu

5. Select **Delete**.
The prompt to confirm the deletion of the Certificates parcel appears.

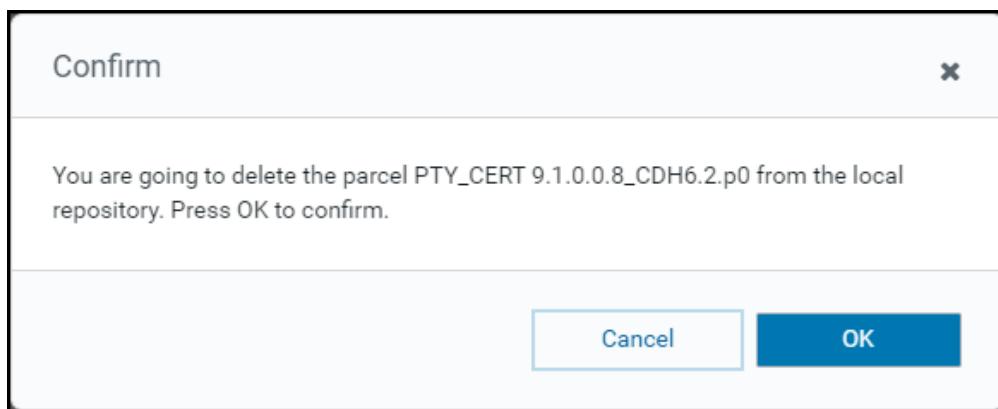


Figure 12-182: Prompt to Delete the Certificates Parcel

6. Click **OK**.
The Certificates parcel is deleted from the local repository.

7. On the Cloudera Manager **Parcels** page, besides the Fluent Bit configuration parcel, click . The drop-down menu appears.

| | | | |
|-------------------------|---------------------|--------------------|------------|
| PTY_FLUENTBIT_CONF | 9.1.0.0.8_CDH6.2.p0 | Downloaded | Distribute |
| SQOOP_NETEZZA_CONNECTOR | 1.5.1c6 | Available Remotely | Delete |

Figure 12-183: Drop-down Menu

8. Select **Delete**.
The prompt to confirm the deletion of the Fluent Bit configuration parcel appears.

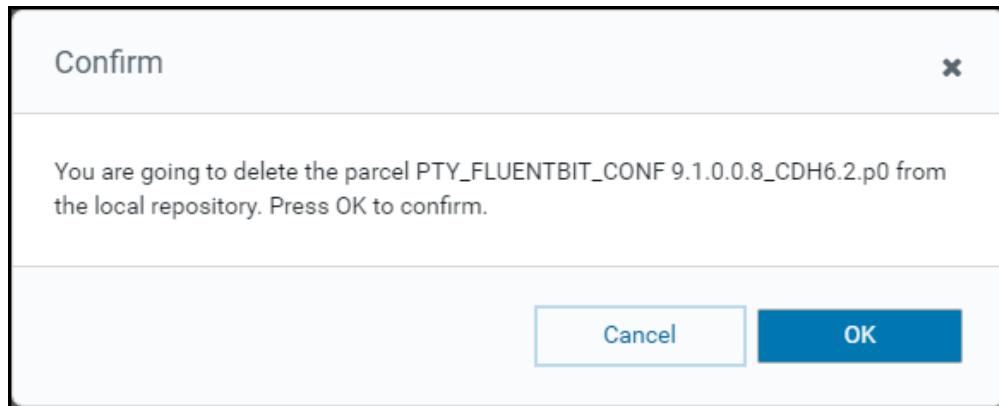


Figure 12-184: Prompt to Delete the Fluent Bit Configuration Parcel

9. Click **OK**.
The Fluent Bit configuration parcel is deleted from the local repository.
10. After all the Big Data Protector parcels are deleted from the repository, roll back the configuration updates from the cluster.

Note: For more information about rolling back the configuration updates from the cluster, refer to section *Rolling back the Configuration Updates* in the *Protegility Big Data Protector Guide 9.1.0.0*.

12.6.2.16.5 Deleting the .JAR Files

► To delete the *BDP_PEP* and the *PTY_PROXYJAR* files:

1. On the Cloudera Manager server node, navigate to the */opt/cloudera/csd/* directory.
2. Delete the *BDP_PEP-9.1.0.0.x.jar* and *PTY_PROXY-9.1.0.0.x.jar* files.
3. Restart the Cloudera Manager server.
4. Restart the Cloudera Management services on the Cloudera Manager web interface after the Cloudera Manager server starts up.

12.6.3 Installing the Big Data Protector using Ambari Native Installer

This section describes the procedures to install and uninstall Protegility Big Data Protector using the Ambari Native installer.

Note:

The Ambari Native installer for Big Data Protector is certified till the 7.1 release only.

If you need the Ambari Native installer binaries for Big Data Protector 7.2.1, then contact Protegility Support.

Note:

The HDFSFP-related components in the Ambari Native installer (such as the *BDPHDFSFP-AMARI-MPACK* management pack, the *BDPHDFSFP* service, and the Talend-related files) will not be available starting from the Big Data Protector 7.2.0 release, as the HDFS File Protector (HDFSFP) is deprecated.

For the Big Data Protector 7.2.1 release, the following versions are used for reference:

- Ambari, version 2.6
- HDP, version 2.6

For more information about Ambari, refer to <https://docs.cloudera.com/HDPDocuments/Ambari/Ambari-2.6.0.0/index.html>.

12.6.3.1 Verifying Prerequisites for Installing Big Data Protector

Ensure that the following prerequisites are met, before installing Big Data Protector in Ambari:

- ESA appliance, version 7.2.0, is installed, configured, and running.
- The following ports are configured on the ESA and the nodes in the cluster, which will run Big Data Protector:
 - ESA: Ensure that the Big Data Protector nodes can communicate with the ESA on the port *8443*.
 - Big Data Protector nodes: Ensure that on each node (localhost), the port *16700* is open and not allocated to any other process.
 - Proxy node: If a proxy node is used between the ESA and Big Data Protector nodes, then the port *8443* needs to be open on the proxy node.
- The user who is installing Big Data Protector has the required permissions to perform the following tasks:
 - Install the Big Data Protector management packs in HDP
 - Add and install the Big Data Protector services in HDP
 - Restart the Ambari Server service
- If you are installing Big Data Protector using Cloudbreak on Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure, then ensure that the following prerequisites are met:
 - Ensure familiarity with the Cloudbreak environment.

For more information about Cloudbreak, refer to [Cloudbreak documentation](#).

- Credentials to authenticate your cloud provider account and provision cloud resources are known.
- The required permissions and roles are set in the management consoles of the respective cloud providers.
- The Cloudbreak deployer is installed and running.

Note:

In the 7.1 MR2 release, Cloudbreak has been certified with the AWS platform.

- If you are configuring the Big Data Protector with a Kerberos-enabled Hadoop cluster, then ensure that the HDFS superuser (*hdfs*) has a valid Kerberos ticket.
- If you are installing the Big Data Protector on a cluster, then ensure that it is installed on all the nodes in the cluster.
- The group *ptyitusr* and the user *ptyitusr*, which are responsible to manage the Big Data Protector-related services, are not present in the Hadoop cluster nodes, which are managed by Ambari.

12.6.3.2 Installing Big Data Protector

This section describes the tasks for installing Big Data Protector on the nodes in a Hadoop cluster.

Caution: Ensure that you are logged in as the required user with the relevant privileges for installing Big Data Protector and running the Ambari services.

12.6.3.2.1 Downloading and Extracting the Big Data Protector Package

After receiving the Big Data Protector installation package from Protegility, copy it to any user defined directory on any node that has ESA connectivity. You need to extract the Big Data Protector package to access the Big Data Protector Configurator script to proceed with the installation of Big Data Protector on the nodes in the Hadoop cluster.

► To extract the files from the installation package:

1. Login to the CLI on any node that has ESA connectivity.
2. Copy the Big Data Protector package to a directory, such as the `/opt/bigdata` directory.
3. Extract the `BDPConfigurator_<OS>_Linux_<arch>_hdp-2.x-<arch>_7.2.0.x.sh` file from the Big Data Protector installation package using the following command.
`tar -xvf BigDataProtector_<OS>-<arch>_HDP-2.6-64_7.2.0.x.tgz`
4. Press ENTER.

The `BDPConfigurator_HDP-2.6_7.2.0.x.sh` file is extracted from the Big Data Protector installation package.

12.6.3.2.2 Running the Big Data Protector Configurator Script

You need to run the Big Data Protector configurator script to download certificates from the ESA, and create the Big Data Protector management packs.

► To create the Big Data Protector management packs:

1. Run the `BDPConfigurator_HDP-2.6_7.2.0.x.sh` script from the directory where it is extracted.
A prompt to install or update Big Data Protector appears.
2. Type `I` for installing Big Data Protector.
3. Press ENTER.
A prompt for the Big Data Protector installation directory appears.
4. Enter the directory to install Big Data Protector on all the nodes in the cluster.
The Big Data Protector is installed in the `/opt/protegility` directory by default.

Note: If the installation directory already exists and contains any files or directories, then a backup of the same is created as a `PROTEGILITY-<TIMESTAMP>.tar` file, and placed in the parent directory of the installation path.

5. Press ENTER.
A prompt for the ESA IP address appears.
6. Enter the ESA IP address.
7. Press ENTER.
A prompt for the ESA listening port appears.
8. Enter the ESA listening port.
9. Press ENTER.
A prompt for the ESA user name appears.
10. Enter the ESA user name.
11. Press ENTER.
A prompt for the ESA password appears.



12. Enter the ESA password.

13. Press ENTER.

The certificates are downloaded from the ESA and the following Big Data Protector management packs are created:

- *BDPPEP-AMBARI-MPACK-7.2.0.x.tar.gz*: Big Data Protector management pack
- *BDPCERTS-AMBARI-MPACK-7.2.0.x.tar.gz*: Certificates management pack

12.6.3.2.3 Installing the Big Data Protector Management Packs

You need to install the following Big Data Protector management packs on the node in the cluster, which is hosting the Ambari server, before starting the Big Data Protector services on the nodes in the cluster:

- Big Data Protector management pack: *BDPPEP-AMBARI-MPACK-7.2.0.x.tar.gz*
- Certificates management pack: *BDPCERTS-AMBARI-MPACK-7.2.0.x.tar.gz*

Note: If the Big Data Protector configurator script was run on a node that did not host the Ambari server, then ensure that you copy the Big Data Protector-related management packs to the node, which is hosting the Ambari server.

If you are using Cloudbreak to create an HDP cluster and install Big Data Protector, then refer to section [Using Cloudbreak to Install the Big Data Protector](#).

► To install Big Data Protector Management Packs on the Node hosting the Ambari Server:

1. On the Ambari server host machine, install the Big Data Protector management pack using the following command.

`ambari-server install-mpack --mpack=/opt/bigdata/BDPPEP-AMBARI-MPACK-7.2.0.x.tar.gz`

2. Press ENTER.

The Big Data Protector management pack is installed on the nodes.

Note: You need to install the Big Data Protector management pack first so that the Certificates management pack, which are dependent on the Big Data Protector management pack, can be installed.

3. Install the Certificates management pack using the following command.

`ambari-server install-mpack --mpack=/opt/bigdata/BDPCERTS-AMBARI-MPACK-7.2.0.x.tar.gz`

The Certificates management pack is installed on the nodes.

4. Restart the Ambari server to detect the new management packs, using the following command.

`ambari-server restart`

5. Press ENTER.

The Ambari server is restarted and the Big Data Protector services are loaded in the Ambari environment.

12.6.3.2.4 Using Cloudbreak to Install the Big Data Protector

If you are using Cloudbreak to create an HDP cluster, then you need to host the Big Data Protector management packs initially. You can then use Cloudbreak to install the following Big Data Protector management packs on all the nodes in the cluster, and start the respective Big Data Protector services on the cluster:

- Big Data Protector management pack: *BDPPEP-AMBARI-MPACK-7.2.1.x.tar.gz*

- Certificates management pack: *BDPCERTS-AMBARI-MPACK-7.2.1.x.tar.gz*

Note: In this section, Cloudbreak, version 2.7.1, is used for reference.

► To install Big Data Protector using Cloudbreak:

- Host the required Big Data Protector management packs on an HTTP URL-based repository, that is accessible to Cloudbreak.
- Navigate to the Cloudbreak login screen.

The Cloudbreak login screen appears.

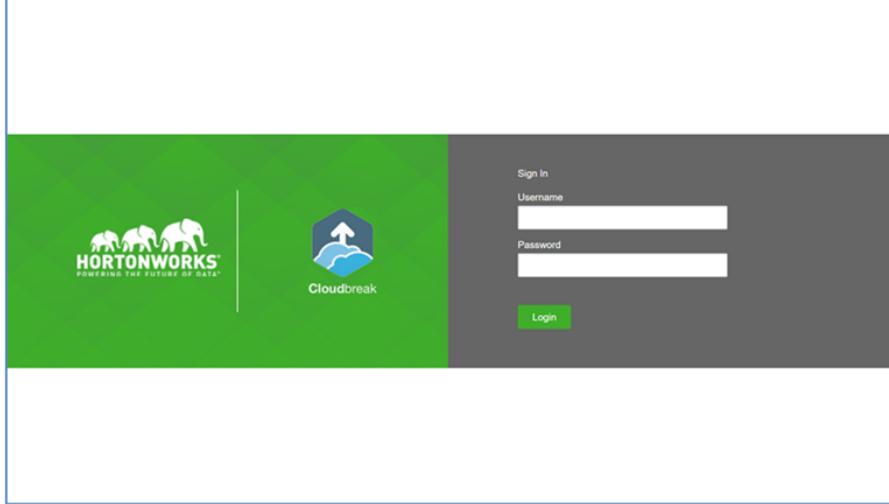


Figure 12-185: Cloudbreak Login Screen

- Enter the required credentials and click **Login** to log in to Cloudbreak.

The Cloudbreak Web UI with the Clusters screen appears.

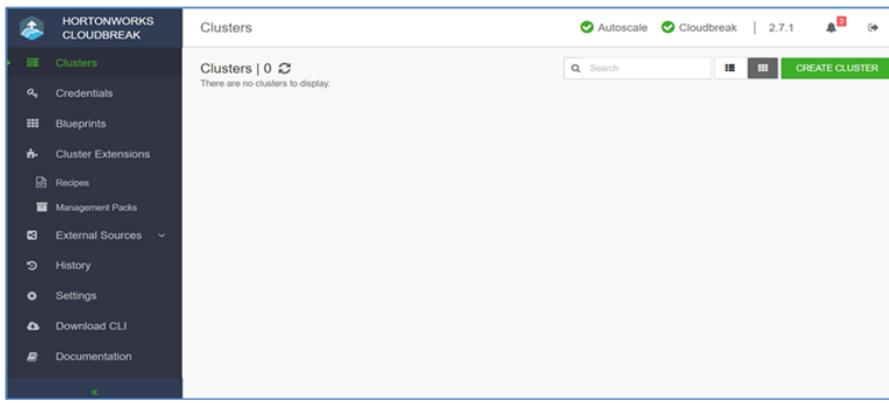


Figure 12-186: Cloudbreak Web UI Screen

- On the Cloudbreak Web UI, navigate to **Cluster Extensions > Management Packs**.

The *Management Packs* pane appears.

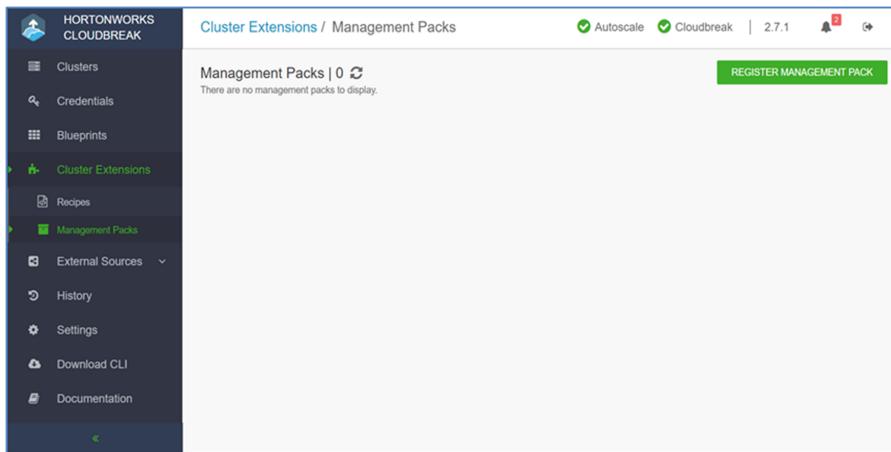


Figure 12-187: Management Packs pane

5. Click **Register Management Pack**.

The *Create Management Packs* pane appears.

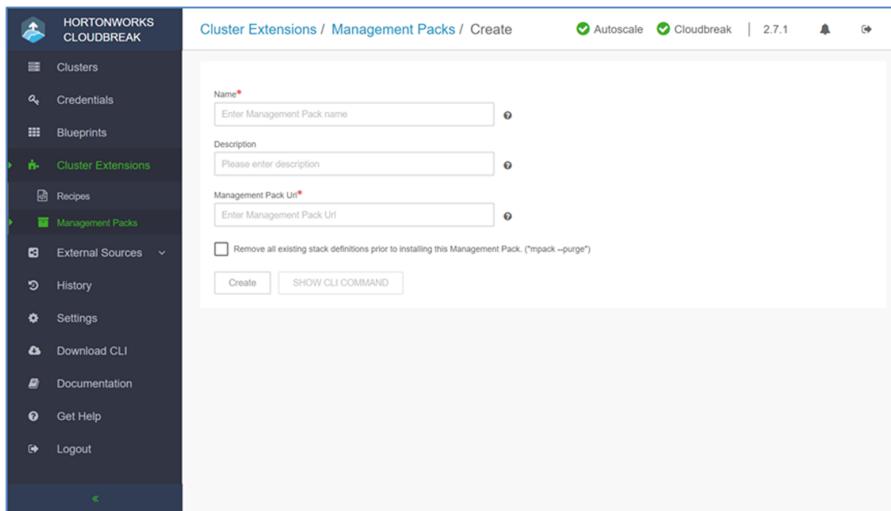


Figure 12-188: Create Management Packs pane

6. Enter the name of the management pack in the *Name* text box.
 7. If required, then enter a description for the management pack in the *Description* text box.
 8. Enter the URL of the management pack, that is hosted in the repository, in the *Management Pack URL* text box
 9. Click **Create**.
- The management pack is registered with Cloudbreak.
10. Repeat the steps 4 through 9 for registering all the required Big Data Protector management packs.
 11. On the Cloudbreak UI, click **Blueprints**.

The *Blueprints* pane appears.

The screenshot shows the 'Blueprints' section of the Hortonworks Cloudbreak interface. On the left is a sidebar with navigation links like Clusters, Credentials, Blueprints (which is selected and highlighted in green), Cluster Extensions, Recipes, Management Packs, External Sources, History, Settings, Download CLI, and Documentation. The main area is titled 'Blueprints | 9' and contains a table with 9 rows. Each row represents a blueprint with columns for Name, Description, Platform, Group Count, and Tags. Some tags are marked as 'Data Lake Ready' or 'Built-In'. At the bottom right of the table are pagination controls for 'Items per page: 25' and '1 - 9 of 9'.

Figure 12-189: Blueprints pane

12. Click CREATE BLUEPRINT.

The *Create Blueprints* pane appears.

The screenshot shows the 'Blueprints / Create' pane. It has a similar sidebar on the left as the previous screen. The main form has fields for 'Name*' (with placeholder 'Enter blueprint name') and 'Description' (with placeholder 'Please enter description'). Below these is a 'Blueprint Source' section with three radio button options: 'File' (selected), 'Url', and 'Text'. A 'Upload JSON File' button is also present. At the bottom of the form are two buttons: 'CREATE' and 'SHOW CLI COMMAND'.

Figure 12-190: Create Blueprints pane

13. Enter the name of the blueprint in the *Name* text box.
14. If required, then enter a description for the blueprint in the *Description* text box.
15. Depending on the requirements, in the *Blueprint Source* area, select the blueprint source as an existing blueprint text file, URL containing the blueprint, or text in JSON format containing blueprint information.
16. Update the blueprint of the HDP cluster to install the Big Data Protector components in the required host groups.

The Big Data Protector consists of the following components:

- *PEP*: Required to be installed on all host groups
- *BDPCERTS*: Required to be installed on all host groups
- *PEP_PROXY*: Required to be installed if you are using a proxy

The following example is a sample blueprint snippet for Cloudbreak.

```
"host_groups": [
  {
    "name": "master",
    "configurations": [
      {
        "hdfs-site": {
          "dfs.datanode.data.dir": "/hadoopfs/fs1/hdfs/datanode"
        }
      },
    ]
  }
]
```



```

    "yarn-site": {
      "yarn.nodemanager.local-dirs": "/hadoopfs/fs1/yarn/nodemanager",
      "yarn.nodemanager.log-dirs": "/hadoopfs/fs1/yarn/nodemanager/log"
    }
  },
  {
    "core-site": {
      "fs.s3a.buffer.dir": "/hadoopfs/fs1/s3-${user.name}"
    }
  }
],
"components": [
  {
    "name": "APP_TIMELINE_SERVER"
  },
  {
    "name": "HCAT"
  },
  ....
  {
    "name": "PEP"
  },
  {
    "name": "BDPCERTS"
  },
  {
    "name": "PEP_PROXY"
  }
],
"cardinality": "1"
}

```

17. Click **Create**.

The blueprint for the HDP cluster is created.

18. On the Cloudbreak UI, click **Clusters**.

19. Click **CREATE CLUSTER**.

The *Create Cluster* wizard appears with the **BASIC** pane displayed by default.

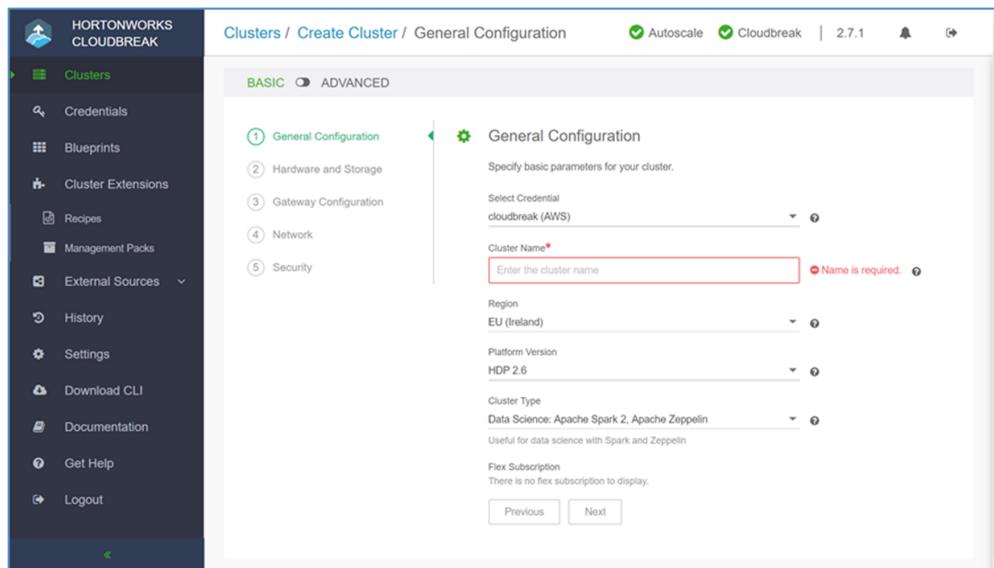


Figure 12-191: Create Cluster Wizard Screen with Basic Pane

20. Click the **ADVANCED** toggle button.

The **ADVANCED** pane appears with the *General Configuration* pane displayed by default.

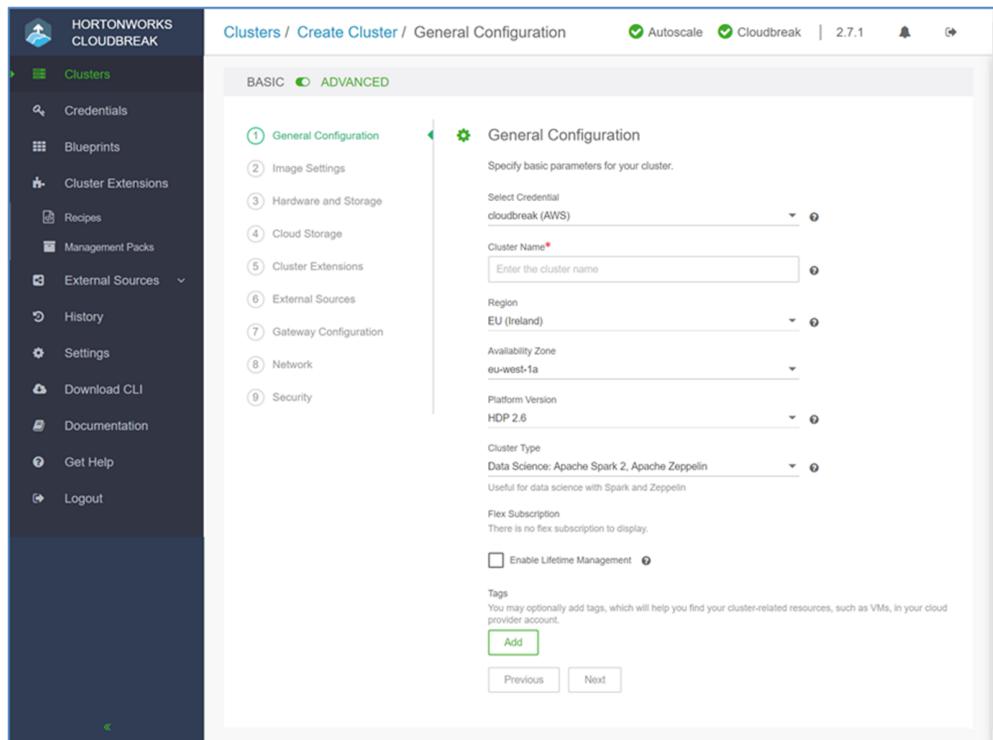


Figure 12-192: Create Cluster Wizard Screen with Advanced pane

21. Select the required cloud provider credentials in the *Select Credential* drop-down list.
22. Enter the name of the cluster in the *Cluster Name* text box.
23. Select the region nearest to the cluster requirement in the *Region* drop-down list.
24. Select the location nearest to the cluster requirement in the *Availability Zone* drop-down list.
25. Select the required HDP version in the *Platform Version* drop-down list.
26. Select the blueprint that was created earlier in the *Cluster Type* drop-down list.
27. Click **Next**.

The *Image Settings* pane appears.

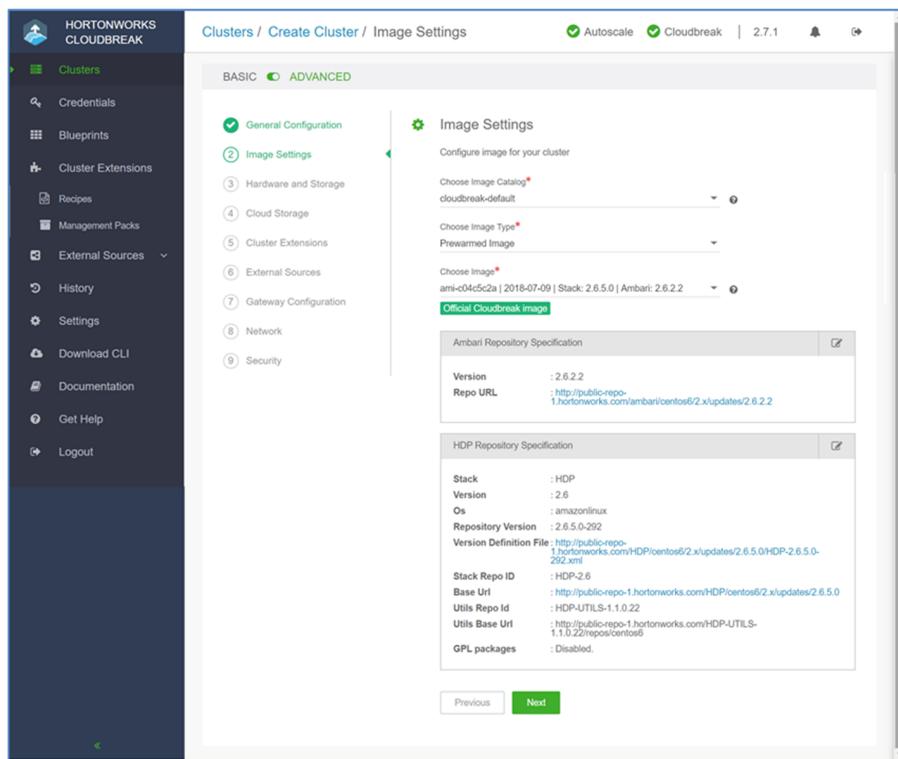


Figure 12-193: Image Settings pane

28. Verify the required settings on the *Image Settings* pane and click **Next**.

The *Hardware and Storage* pane appears.

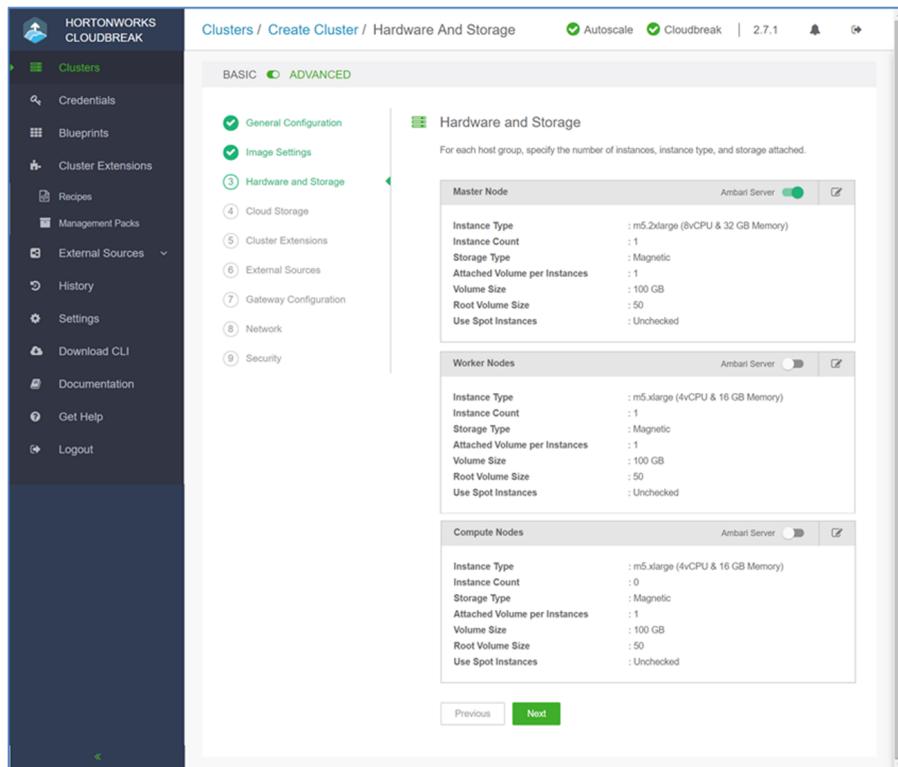


Figure 12-194: Hardware and Storage pane

29. Verify the required settings on the *Image Settings* pane and click **Next**.

The *Cloud Storage* pane appears.

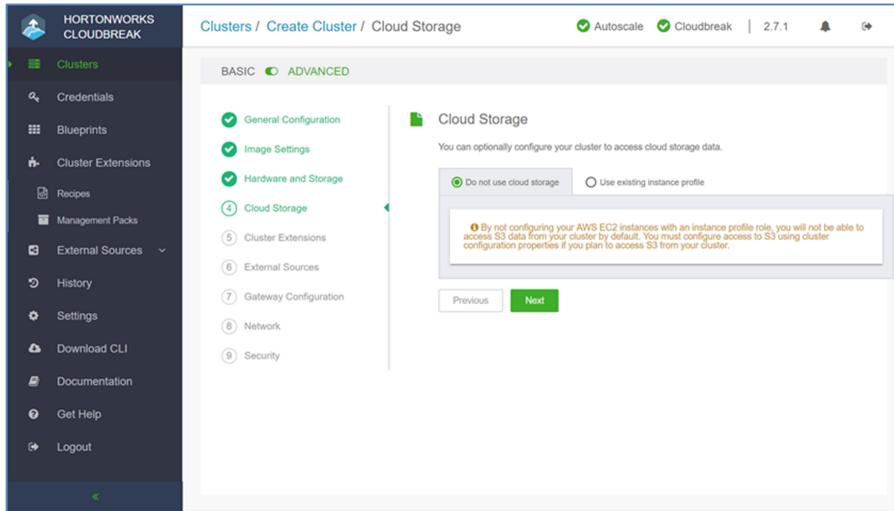


Figure 12-195: Cloud Storage pane

- Verify the required settings on the *Cloud Storage* pane and click **Next**.

The *Cluster Extensions* pane appears.

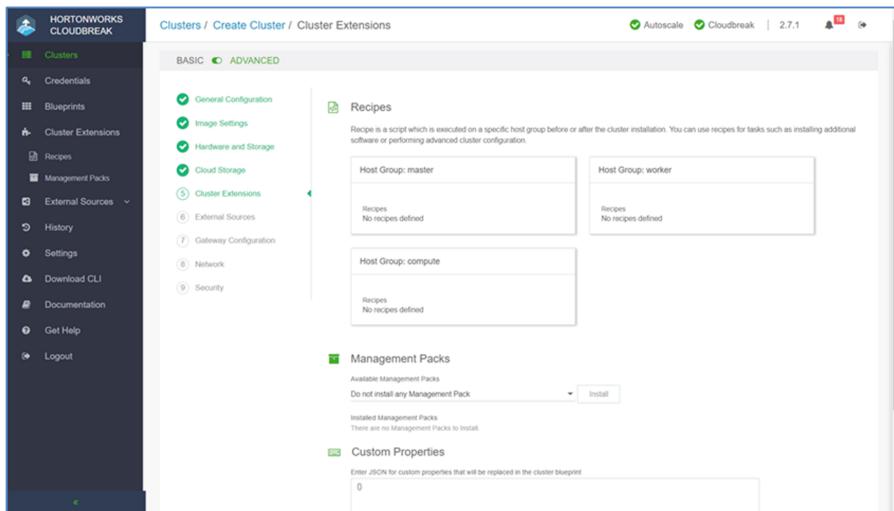


Figure 12-196: Cluster Extensions pane

- Under the **Management Packs** section, select the required Big Data Protector management packs, that are registered with Cloudbreak, for installation on the HDP cluster.
- Verify the required settings on the *Cluster Extensions* pane and click **Next**.

The *External Sources* pane appears.

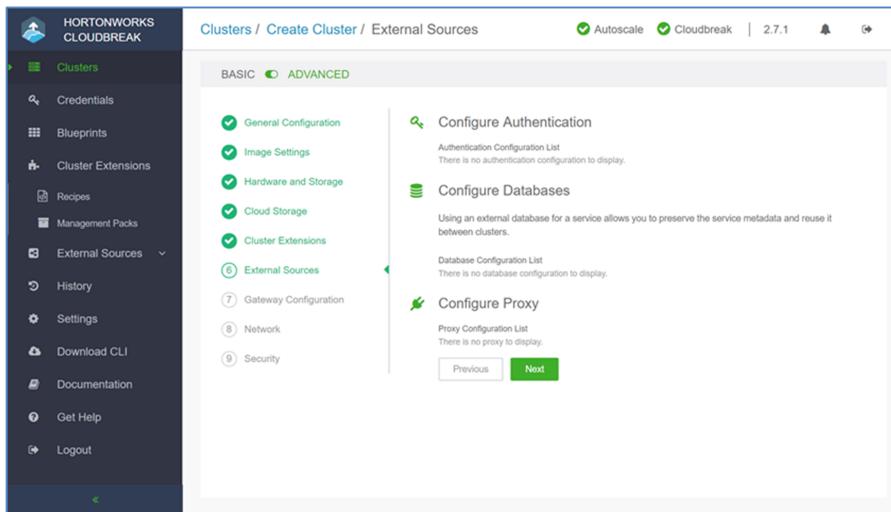


Figure 12-197: External Sources pane

- Verify the required settings on the *External Sources* pane and click **Next**.

The *Gateway Configuration* pane appears.

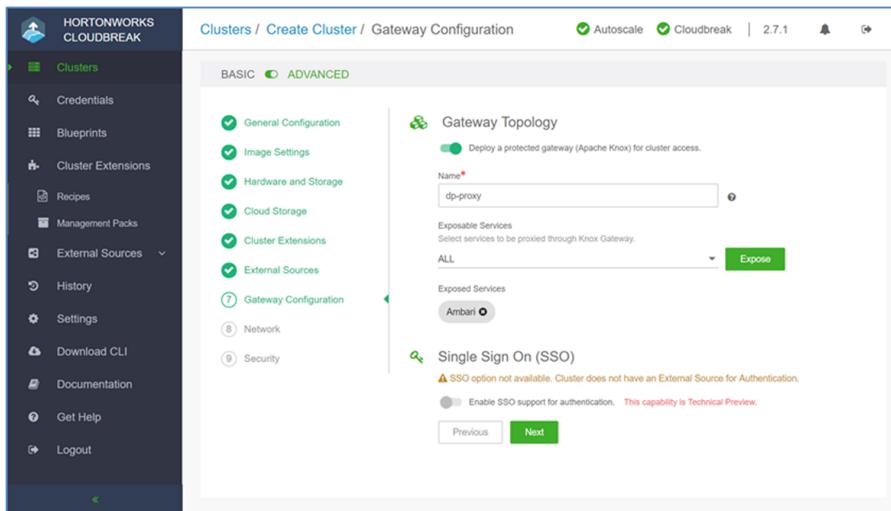


Figure 12-198: Gateway Configuration pane

- Verify the required settings on the *Gateway Configuration* pane and click **Next**.

The *Network* pane appears.

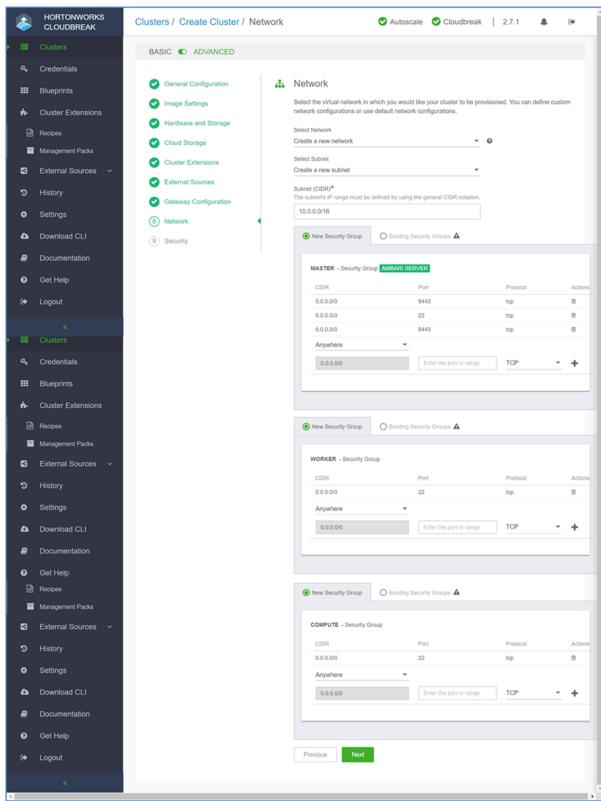


Figure 12-199: Network pane

35. Verify the required settings on the *Network pane* and click **Next**.

The *Security pane* appears.

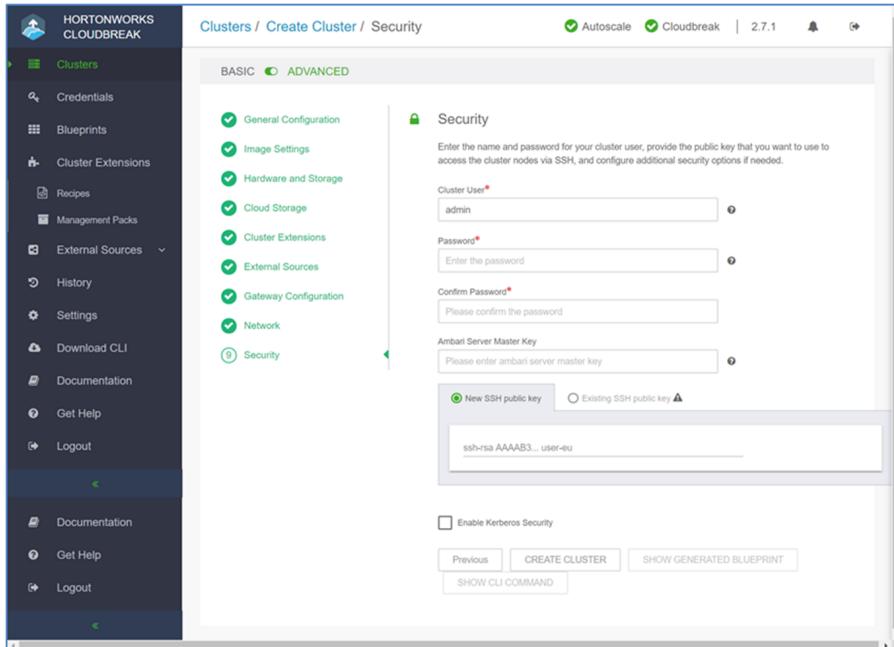


Figure 12-200: Security pane

36. Verify the required settings on the *Security pane* and click **CREATE CLUSTER**. The HDP cluster is created based on the blueprint provided, the Big Data Protector management packs are installed, and the corresponding services are started on the nodes in the cluster.

You can now utilize the Ambari UI to manage the Big Data Protector services.

For more information about managing the Big Data Protector services, refer to section [Managing the Big Data Protector on HDP](#).

Before using the Big Data Protector, the configuration parameters need to be set to the recommended values through Cloudbreak, depending on the components installed on the cluster.

For more information about setting the configuration parameters, refer to section [3.6 Setting the Configuration for Big Data Protector](#) in the [Protegility Big Data Protector Guide 9.0.0.0](#).

12.6.3.2.5 Starting the Big Data Protector Services on the Nodes

After installing the Big Data Protector management packs on the nodes, you need to add the following Big Data Protector-related services on the cluster.

- Big Data Protector service: Adds and starts the *BDPPEP* service on the nodes
- Certificates service: Adds and starts the *BDPCERTS* service on the nodes
- Proxy service: Adds and starts the *PTYPROXY* service on the nodes

12.6.3.3 Adding the Big Data Protector PEP Service

To use Big Data Protector, you need to add and start the Big Data Protector PEP service on the nodes.

Note:

The Big Data Protector PEP service (*BDPPEP*) depends on the Certificates service (*BDPCERTS*).

After you add the *BDPPEP* by following the steps in this section, ensure that you start the *BDPCERTS* service too by referring to the section [Starting the Certificates Service](#).

► To add and start the Big Data Protector PEP Service on the Nodes:

1. Navigate to the Ambari screen.



Figure 12-201: Ambari screen

2. Enter the required user name for Ambari.

3. Enter the required password for Ambari.

4. Click **Log In**.

The Ambari Home screen appears.

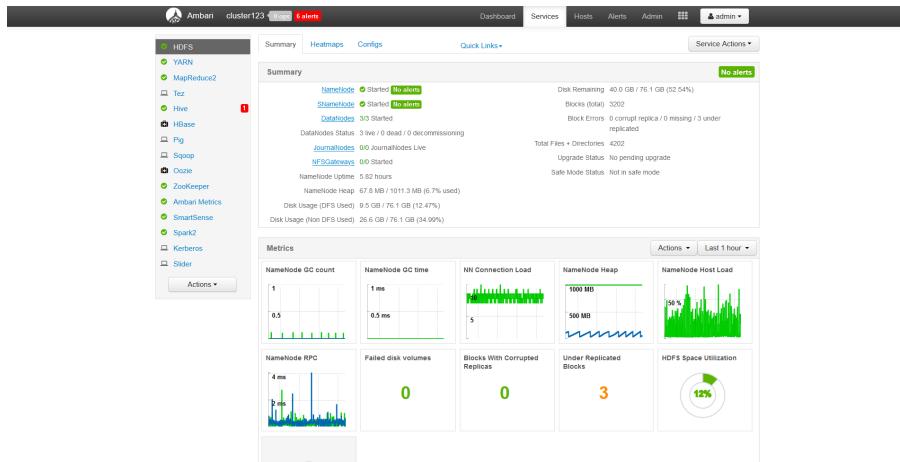


Figure 12-202: Ambari Home screen

5. Click the **Actions** button.

The Actions drop down appears.

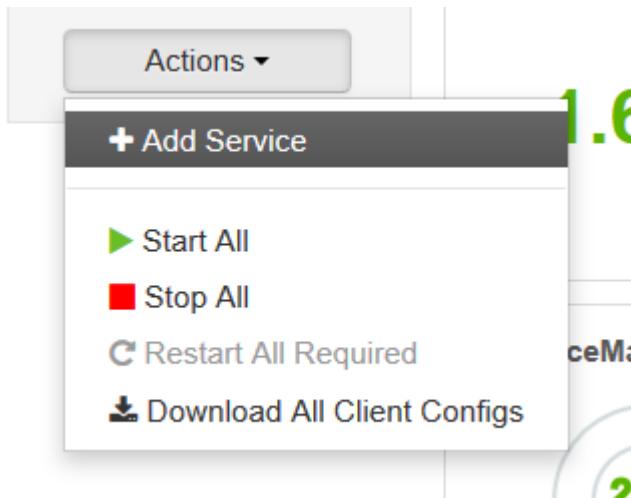


Figure 12-203: Actions Drop Down

6. Click **Add Service**.

The **Choose Service** screen appears listing the Big Data Protector-related services.

| | | |
|-----------------------------------|---------|---|
| <input type="checkbox"/> BDPCERTS | 7.2.0.6 | Protegility provided Authentication enabler for BDPPEP Service |
| <input type="checkbox"/> BDPPEP | 7.2.0.6 | Protegility provided distributed service for comprehensive data protection across various components in hadoop ecosystem. |

Figure 12-204: Choose Service screen

7. Select **BDPPEP**.

8. Click **Next**.

9. Select the hosts for the Protegility PEP Server (*Pty PepServer*) component.

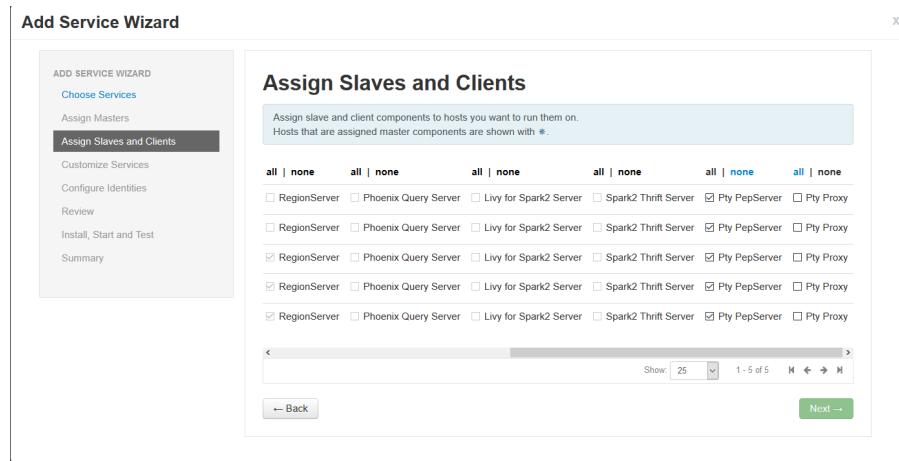


Figure 12-205: Protegility PEP Server hosts

10. If you need to use the Proxy service or ensure that the required number of nodes are accessible from the ESA, then select the Protegility Proxy (*Pty Proxy*) component.
11. Click **Next**.

The **Customize Services** screen appears.

Figure 12-206: Customize Services screen

12. Click the **BDPPEP** tab.
- The **BDPPEP** tab appears, and the *pepperserver.cfg* configuration file, is displayed.

Note: The *service_installation_dir* parameter considers the installation directory parameter from the input provided while running the Big Data Protector Configurator script. Ensure that you do not modify this value.

The screenshot shows the 'Customize Services' interface. At the top, there's a message: 'We have come up with recommended configurations for the services you selected. Customize them as you see fit.' Below this are tabs for various services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Oozie, ZooKeeper, Ambari Metrics, SmartSense, Spark2, BDPPEP, Slider, and Misc. The BDPPEP tab is selected. Underneath are buttons for 'Group', 'Default (5)', 'Manage Config Groups', and a 'Filter...' dropdown. The main area contains two expandable sections: 'Advanced pepconfig-env' and 'Advanced ptyproxyconfig'. The 'Advanced pepconfig-env' section shows the 'pepserver-cfg' file content, which includes configuration for the pepserver, application configuration, and temporary file storage. The 'service_installation_dir' field is set to '/opt/protegility/'. The 'Advanced ptyproxyconfig' section shows fields for 'ESA_IP' and 'PTYProxy_Port', both currently set to '8443'.

Figure 12-207: pepserver.cfg File

13. If required, then you can customize some parameters in the *pepserver.cfg* file. For more information about customizing the parameters in the *pepserver.cfg* file, refer to section *6.5.2.7.1 Updating the PEP Server Parameters for the BDP PEP Service*.
14. If you are setting up a Proxy, then enter the following details.
 - a. Proxy IP in the *ESA_IP* text box.
 - b. Proxy port in the *PTYProxy_Port* text box.
 - c. Replace all the instances of the ESA IP address in the *pepserver-cfg* section with the Proxy IP.
 - d. If the Proxy port is other than *8443*, then replace the same in the *pepserver-cfg* section with the required port number.

This screenshot shows the 'Customize Services' interface with the 'Advanced pepconfig-env' section expanded. The 'pepserver-cfg' file content is displayed, showing several commented-out sections and parameters. A cursor is positioned over the line '# Oracle - Configurable. Must match the value specified in 'transportbaricall''. Below the configuration file is the 'service_installation_dir' field, which is set to '/opt/protegility/'. The 'Advanced ptyproxyconfig' section is also visible at the bottom.

Figure 12-208: ESA or Proxy IP Address in the ptyproxyconfig section

15. Click **Next**.

A dialog box appears listing all the configuration recommendations, which are typical configuration settings that would be required to be set for the Big Data Protector PEP services before using Big Data Protector.

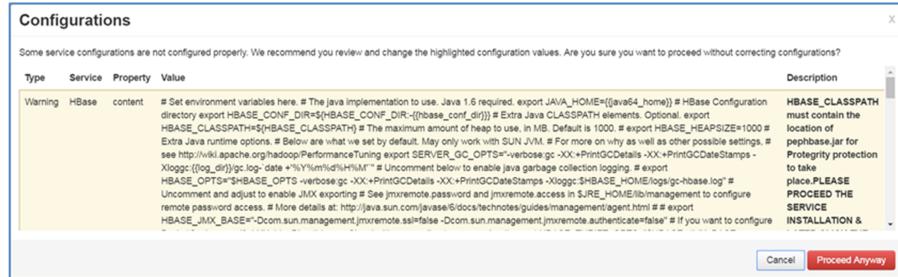


Figure 12-209: Configuration Recommendations for Big Data Protector

- Click **Proceed Anyway**.

The **Configure Identities** tab appears.

Figure 12-210: Configure Identities tab

- Click **Next**.

The **Review** tab appears listing the Big Data Protector PEP services added to HDP.

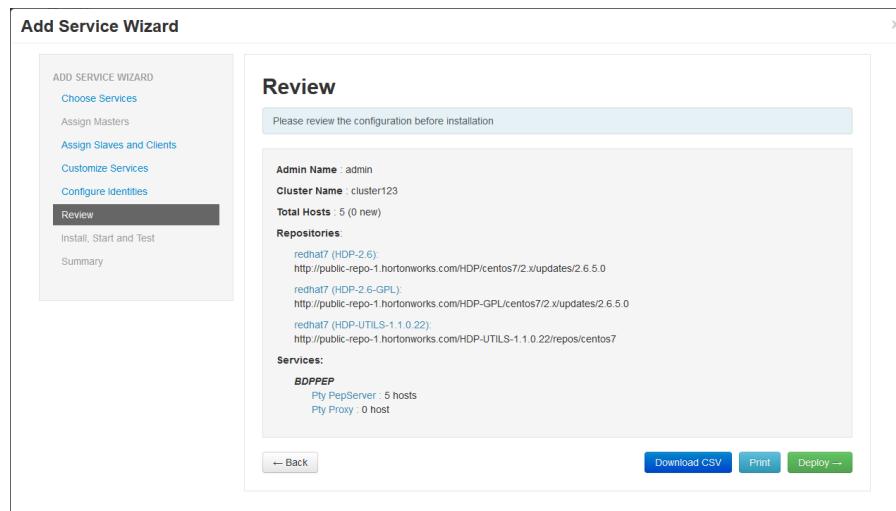


Figure 12-211: Review tab

18. Click **Deploy**.

If the cluster is a Kerberos-enabled Hadoop cluster, then the prompt for the Kerberos credentials appears.

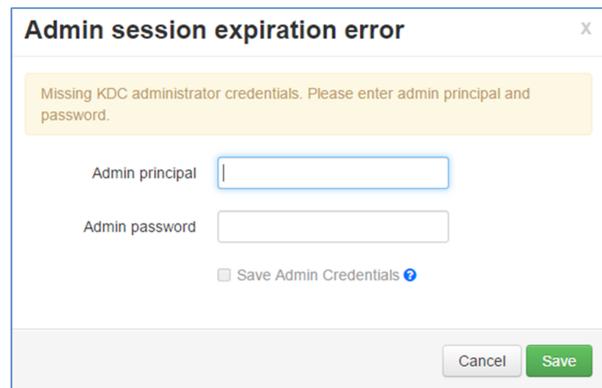


Figure 12-212: Kerberos Credentials

19. Enter the user name for the *Admin principal*.

20. Enter the Admin password.

21. Click **Save**.

The *BDPPEP* service is deployed on the nodes.

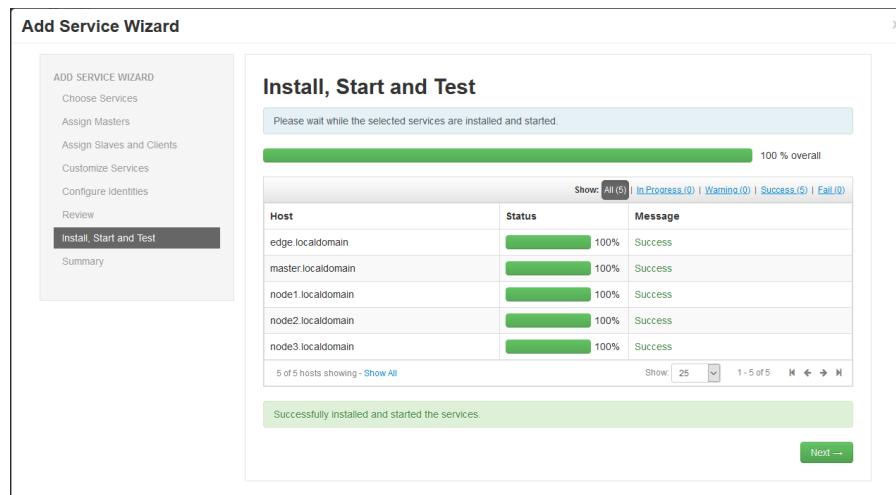


Figure 12-213: BDPPEP Service Status

22. Click **Next.**

The **Summary** tab appears.

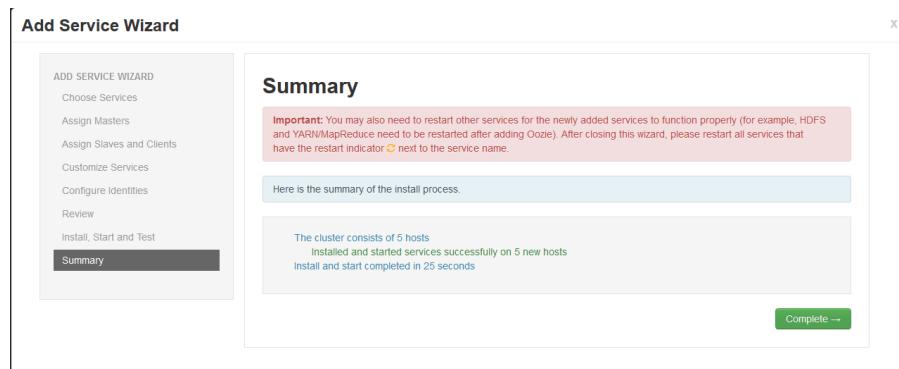


Figure 12-214: Summary tab

23. Click **Complete.**

The **BDPPEP** service appears on the Ambari Home screen. The Big Data Protector PEP service is not started as the **BDPCERTS** service is not installed.

12.6.3.4 Starting the Certificates Service

► To start the Certificates Service on the Nodes:

1. On the Ambari Home screen, click the **Actions** button.

The *Actions* drop down appears.

2. Click **Add Service**.

The **Choose Service** screen appears listing the Big Data Protector-related service.

| | | |
|--|---------|---|
| <input type="checkbox"/> BDPCERTS | 7.2.0.6 | Protegility provided Authentication enabler for BDPPEP Service |
| <input checked="" type="checkbox"/> BDPPEP | 7.2.0.6 | Protegility provided distributed service for comprehensive data protection across various components in hadoop ecosystem. |

Figure 12-215: Choose Service screen

3. Select **BDPCERTS**.

4. Click **Next**.

5. Select the hosts for the *ESA CERTS* component.

Ensure that all the hosts with the PEP server component are selected.

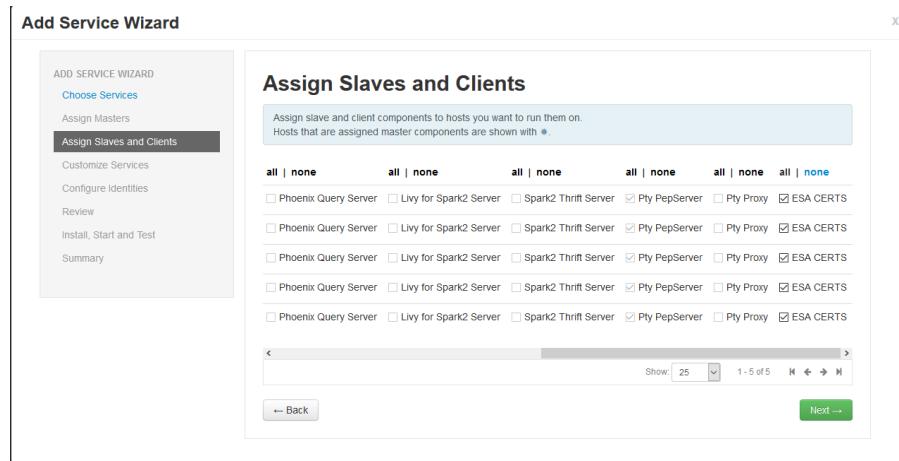


Figure 12-216: ESA CERTS hosts

- Click **Next**.

The **Customize Services** screen appears, and the **BDPCERTS** tab is displayed by default.

Caution: Ensure that you do not modify the value of the *BDPPEP_Version* parameter.

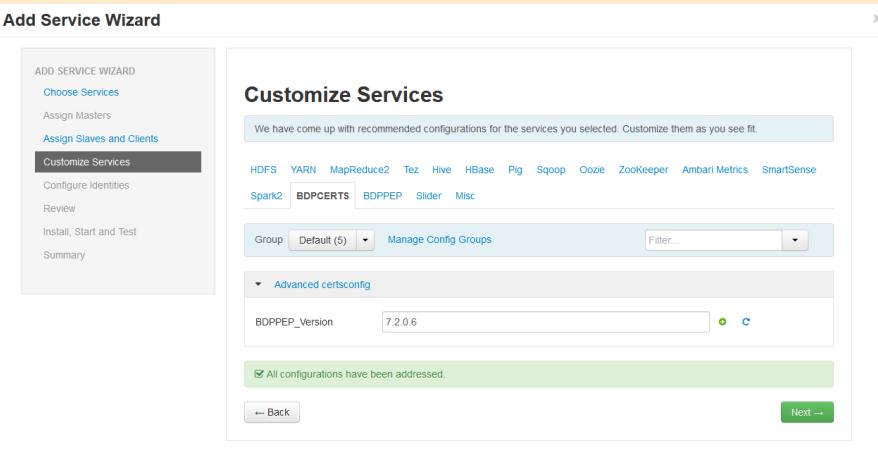


Figure 12-217: Customize Services screen

- Click **Next**.

A dialog box appears listing all the configuration recommendations, which are typical configuration settings that would be required to be set before using the Big Data Protector.

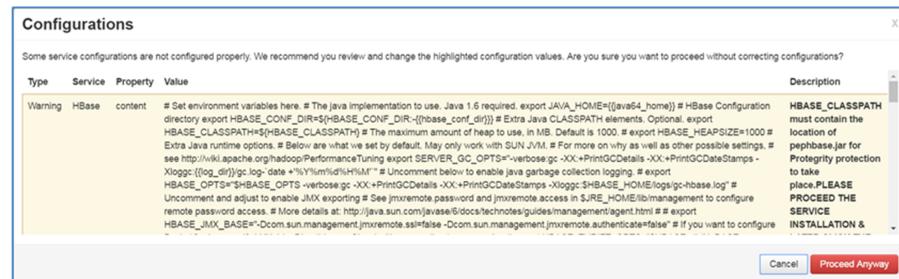


Figure 12-218: Configuration Recommendations for Big Data Protector

- Click **Proceed Anyway**.

The **Configure Identities** tab appears.

Configure Identities

Configure principal name and keytab location for service users and hadoop service components.

General Advanced

Global

| | |
|-------------------|--------------------------------------|
| Keytab Dir | /etc/security/keytabs |
| Realm | EXAMPLE.COM |
| Additional Realms | |
| Principal Suffix | -\${cluster_name} Lower() |
| Spnego Keytab | \$(keytab_dir)/spnego.service.keytab |
| Spnego Principal | HTTP/_HOST@\${realm} |

Ambari Principals

| | |
|-----------------------|---|
| Smoke user keytab | \$(keytab_dir)/smokeuser.headless.keytab |
| Smoke user principal | \$(cluster-env/smokeuser)\${principal_suffix}@\${realm} |
| Ambari Keytab | \$(keytab_dir)/ambari server keytab |
| Ambari Principal Name | ambari-server\${principal_suffix}@\${realm} |
| HBase user principal | \$(hbase-env/hbase_user)\${principal_suffix}@\${realm} |
| HDFS user keytab | \$(keytab_dir)/hbase headless keytab |
| HDFS user principal | \$(hadoop-env/hdfs_user)\${principal_suffix}@\${realm} |
| HDFS user keytab | \$(keytab_dir)/hdfs headless keytab |
| Spark user keytab | \$(keytab_dir)/spark.headless.keytab |
| Spark user principal | \$(spark2-env/spark_user)\${principal_suffix}@\${realm} |

All configurations have been addressed.

[← Back](#) [Next →](#)

*Figure 12-219: Configure Identities tab*9. Click **Next**.

The **Review** tab appears with the BDPCERTS configuration for Big Data Protector.

Add Service Wizard

Review

Please review the configuration before installation

Admin Name: admin

Cluster Name: cluster123

Total Hosts: 5 (0 new)

Repositories:

- redhat7 (HDP-2.6)
 - http://public-repo-1.hortonworks.com/HDP/centos7/2.x/updates/2.6.5.0
 - redhat7 (HDP-2.6-GPL)
 - http://public-repo-1.hortonworks.com/HDP-GPL/centos7/2.x/updates/2.6.5.0
 - redhat7 (HDP-UTILS-1.1.0.22):
 - http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.22/repos/centos7

Services:

- BDPCERTS**
 - ESA CERTS : 5 hosts

[← Back](#) [Download CSV](#) [Print](#) [Deploy →](#)

*Figure 12-220: Review tab*10. Click **Deploy**.

11. If the cluster is a Kerberos-enabled Hadoop cluster, then the prompt for the Kerberos credentials appears.

12. Enter the user name for the *Admin principal*.

13. Enter the Admin password.

14. Click **Save**.

The *BDPCERTS* service is deployed on the nodes.

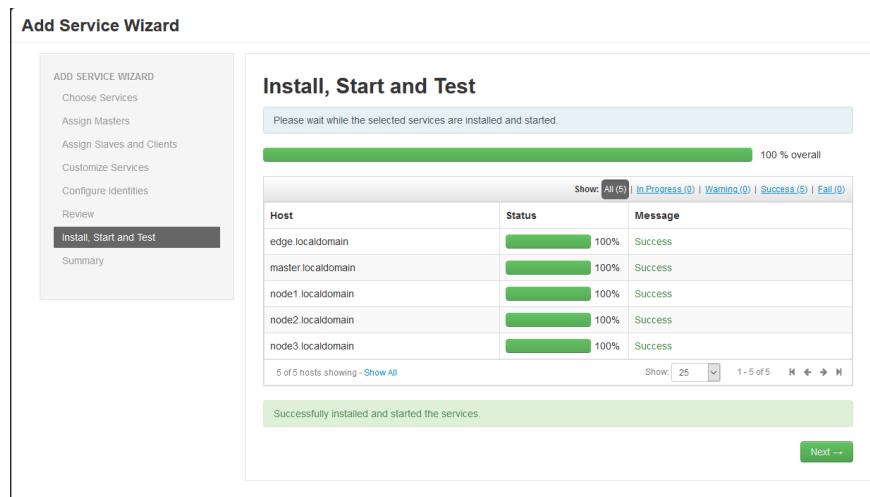


Figure 12-221: BDPCERTS Service Status

15. Click **Next**.

The **Summary** tab appears.

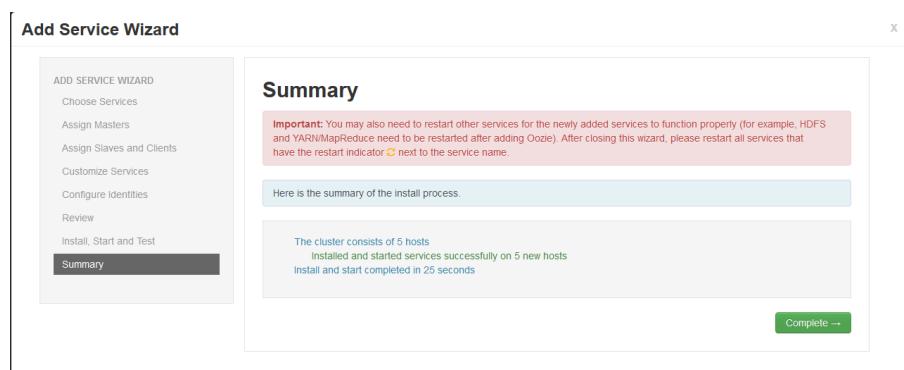


Figure 12-222: Summary tab

16. Click **Complete**.

The **BDPCERTS** service is started. The **BDPPEP** service on the Ambari Home screen can be restarted as the **BDPCERTS** service is installed.

Caution: After all the Big Data Protector-related services are added and activated on the nodes in the cluster, you need to set the configuration parameters to the recommended values.

For instance, if you need to set the configuration parameters for MapReduce, then perform the following steps.

- From the Ambari Home screen, click on the **MapReduce2** service from the left pane to display the MapReduce-related summary.
- Click the *Configs* tab to open the configuration parameters for MapReduce.
- Enter the name of the configuration parameter that needs to be updated, in the *Filters* search box.
- Click the **C** icon to set the recommended configuration value for the parameter.

Ensure that the Protegility recommended values are set in the configuration parameter.

- Click **Save** to save the modifications made to the configuration parameters.
- If you need to set the recommended values for other configuration parameters, go to step 3.

Similarly, you need to set the values for the configuration parameters to the recommended values depending on the components that you need to use with the Big Data Protector.

For more information about setting the configuration parameters, refer to section *3.6 Setting the Configuration for Big Data Protector* in the *Protegility Big Data Protector Guide 9.0.0.0*.

12.6.3.5 Installing Big Data Protector on a New Node

If a new node is added to the Hadoop cluster and you need to install Big Data Protector on it, then perform the following steps.

► To install the Big Data Protector on a New Node:

1. Ensure that the BDPCERTS management pack is installed on the cluster.
2. Add the new node to the cluster.
3. Add the Big Data Protector PEP service to the new node.

For more information about adding and starting the Big Data Protector service, refer to section [Starting the Big Data Protector Services on the Nodes](#).

4. Add the Big Data Protector Certificates service to the new node.

For more information about adding and starting the Big Data Protector Certificates service, refer to section [Starting the Certificates Service](#).

5. If the Proxy needs to be configured, then add the Proxy service to the new node.

12.6.3.6 Installing Big Data Protector on a New Node using Cloudbreak

If a new node is added to the HDP cluster that has been created using Cloudbreak and you need to install Big Data Protector on it, then perform the following steps.

► To install the Big Data Protector on a New Node using Cloudbreak:

1. From the Cloudbreak Web UI, click **Clusters**.

The *Clusters* screen appears.

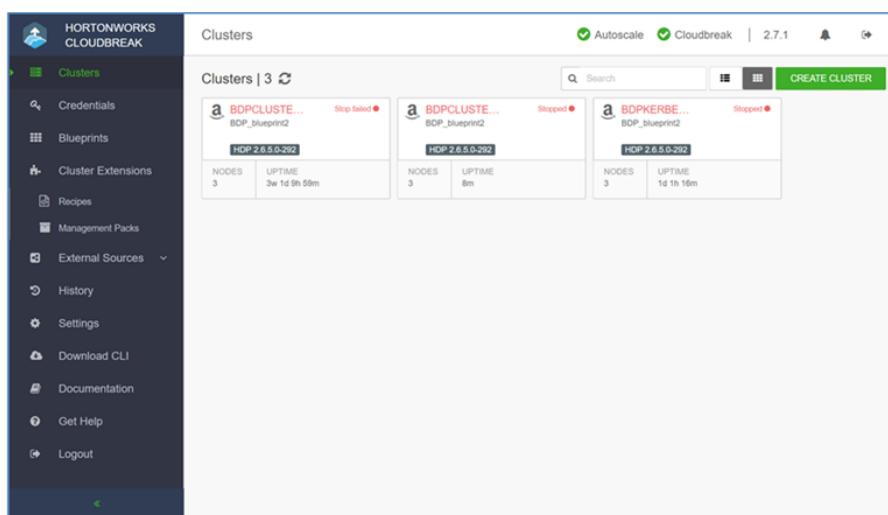


Figure 12-223: Clusters Screen

2. Select an existing cluster on which the nodes need to be added.

The details of the cluster appear listing the existing host groups.

The screenshot shows the Hortonworks CloudBreak interface. On the left, there's a sidebar with navigation links: Clusters, Credentials, Blueprints, Cluster Extensions, Recipes, Management Packs, External Sources, History, Settings, Download CLI, Documentation, Get Help, and Logout. The main content area is titled "Clusters / bdpcluster2 / Hardware". It displays basic cluster information: Cluster User (admin), Credential (cloudbreak), Status (Running), Nodes (3), Uptime (3w 1d 13h 5m), and Created (Nov 19, 2018). Below this is a "CLUSTER INFORMATION" section with details like Region (us-east-1), Availability Zone (us-east-1a), Blueprint (BDP_blueprint2), Created With (2.7.1), Ambari Version (2.6.2.2), and HDP Version (2.6.5.0-292). A "HARDWARE" section lists three hosts: compute (ID i-00db45110c98b7dd, stopped, FQDN ip-10-175-139-40.ec2.internal, Private IP 10.175.139.40, Public IP 34.224.212.7), master (ID i-0242a514f1ca15eab, stopped, FQDN ip-10-175-132-69.ec2.internal, Private IP 10.175.132.69, Public IP 100.25.126.204), and worker (ID i-063057e43709a4ee6, stopped, FQDN ip-10-175-143-87.ec2.internal, Private IP 10.175.143.87, Public IP 54.146.240.61). An "EVENT HISTORY" section shows a log of events from the cluster's start, including Ambari services starting and infrastructure successfully starting. At the bottom right, there's a "DOWNLOAD" button.

Figure 12-224: Existing Cluster Screen

3. Click **ACTIONS**.

The ACTIONS drop-down list appears.

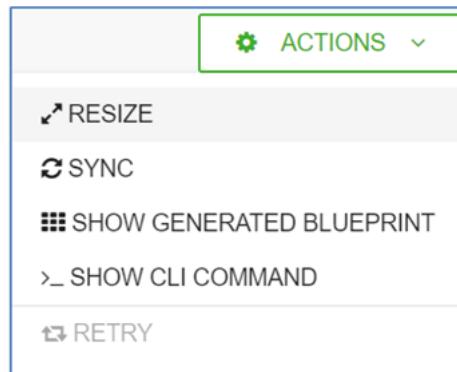


Figure 12-225: ACTIONS Drop-Down List

4. Click **RESIZE**.

The *Cluster Resize* dialog box appears.

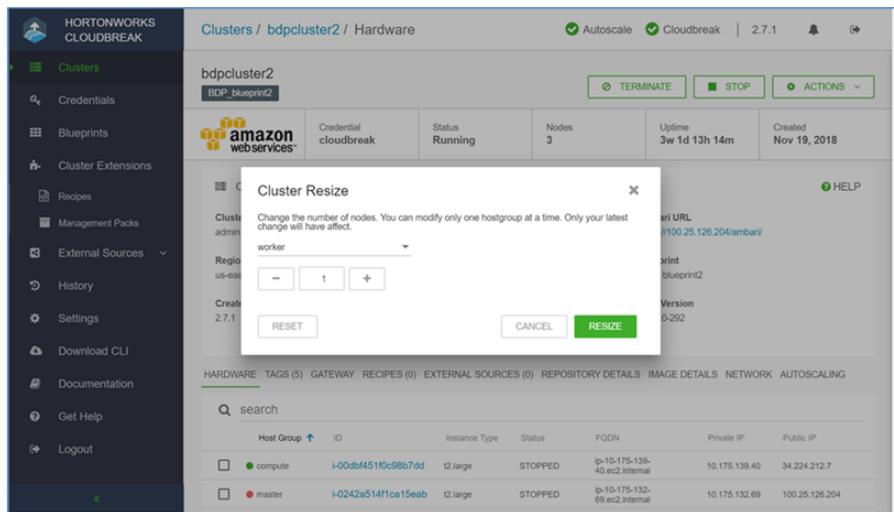


Figure 12-226: Cluster Resize Dialog Box

5. Depending on the requirements, select the host group to which nodes need to be added. In this case, the type of nodes available are *worker* or *compute*.
6. Depending on the requirements, select the number of nodes to be added.
7. Click **RESIZE**.

The required type and number of nodes are added, and Big Data Protector is installed on it. The details of the cluster listing the updated number of nodes in each host group appears.

| Host Group | ID | Instance Type | Status | FQDN | Private IP | Public IP |
|------------|---------------------|---------------|------------|-------------------------------|----------------|----------------|
| compute | i-00db451f0c98b7dd | t2.large | STOPPED | ip-10-175-139-40.ec2.internal | 10.175.139.40 | 34.224.212.7 |
| master | i-0242a514f1ca15eab | t2.large | STOPPED | ip-10-175-132-69.ec2.internal | 10.175.132.69 | 100.25.126.204 |
| worker | i-0054377ad8ccfade | t2.large | CREATED | N/A | 10.175.135.117 | 52.204.157.61 |
| worker | i-063057e43709a4ee6 | t2.large | REGISTERED | ip-10-175-143-87.ec2.internal | 10.175.143.87 | 54.146.240.61 |

Figure 12-227: Cluster Screen with Updated Number of Nodes

12.6.3.7 Managing the Big Data Protector on HDP

You can manage the Big Data Protector on HDP by modifying the PEP server parameters for the Big Data Protector PEP service.

12.6.3.7.1 Updating the PEP Server Parameters for the BDP PEP Service

You can modify the PEP server configuration parameters in the Advanced pepconfig-env section on the BDPPEP tab for the BDP PEP service.

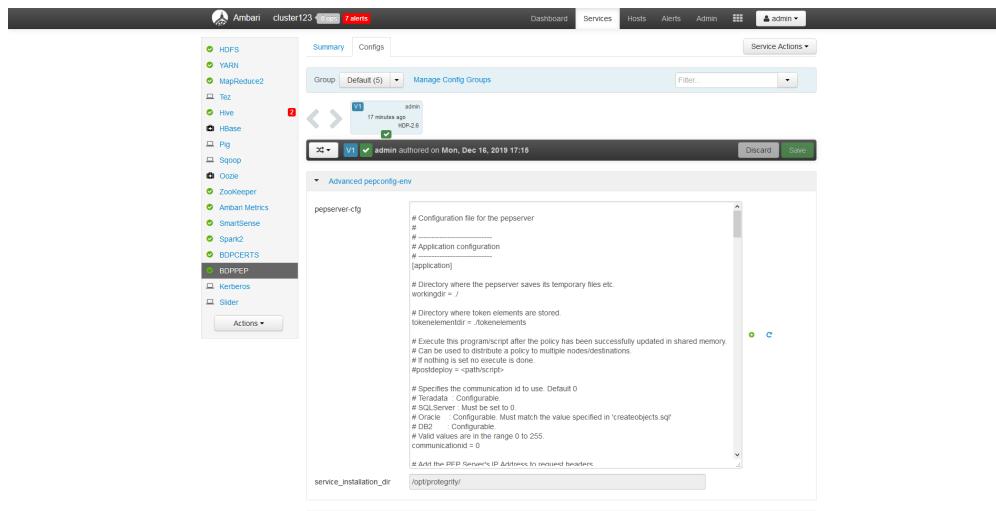


Figure 12-228: BDP PEP Service Configuration Parameters

The following table lists the configuration parameters for the BDP PEP service.

Table 12-45: Configuration Parameters for the BDP PEP Service

| Parameter | Description | Possible Values |
|-----------|---|---|
| level | Specify the level of details for the logs. | <ul style="list-style-type: none"> OFF - No logging Severe Warning Info Config ALL(Default) |
| filename | Log file name and directory to store the log entries. | The default log file, <i>pepservice.log</i> , is saved in the <i>/var/log/protegility/</i> directory. |
| append | Appends the log entries to the existing log file. | <ul style="list-style-type: none"> Yes (Default) No |
| url | URL of the deployment API | <a href="https://<ESA_IP_Address>:8443">https://<ESA_IP_Address>:8443 |

12.6.3.8 Updating the Certificates Management Pack

If customers need to utilize their own CA signed certificates, then the Certificates management pack would be updated with the new certificates. The updated Certificates management pack needs to be utilized by the nodes in the cluster.

► To utilize updated Certificates:

1. Login to the Ambari Server host machine, which contains the Big Data Protector Configurator script.
A prompt to install or update Big Data Protector appears.
2. Type 2 for updating the ESA certificates.
3. Press ENTER.
A prompt for the Big Data Protector installation directory appears.
4. Enter the directory to install Big Data Protector on all the nodes in the cluster.
The Big Data Protector is installed in the `/opt/protegility` directory by default.

Note: If the installation directory already exists and contains any files or directories, then a backup of the same is created as a `PROTEGERY-<TIMESTAMP>.tar` file, and placed in the parent directory of the installation path.

5. Press ENTER.
A prompt for the ESA IP address appears.
6. Enter the ESA IP address.
7. Press ENTER.
A prompt for the ESA listening port appears.
8. Enter the ESA listening port.
9. Press ENTER.
A prompt for the ESA user name appears.
10. Enter the ESA user name.
11. Press ENTER.
A prompt for the ESA password appears.
12. Enter the ESA password.
13. Press ENTER.
A prompt for the current version of the Certificates management pack appears.
14. Enter the current version of the Certificates management pack in the format `7.2.0.x`.
15. Press ENTER.
The certificates are downloaded from the ESA and the `BDPCERTS-AMARI-MPACK-7.2.0.x.tar.gz` management pack is created.
16. Start the Certificates service on the nodes using the updated Certificates management pack.
For more information about starting the new Certificate service, refer to section [Starting the Certificates Service](#).

12.6.3.9 Configuring Ambari to Set-up Multiple External Audit Stores

After BDPPEP installation is completed successfully, you can add an additional audit store or disable the existing audit store.

12.6.3.9.1 Adding a New Audit Store alongside Protegility Elasticsearch

1. After setting and starting up BDPPEP and BDPCERT service using the Ambari UI, login to the CLI on any node that has external audit store connectivity.
2. Create a conf file pertaining to the type and tool that you want to configure as the audit store.

Note:

For more information, refer <https://docs.fluentbit.io/manual/pipeline/outputs>.

Example: Contents of file-xxxx.conf file for file audit store that stores log in /opt/logs.

```
[OUTPUT]
Name file
Match *
Path /opt/logs
```

- To copy this file to the folder from where fluentbit picks up the configuration files, execute the following command:

```
cp file-xxxx.conf /opt/protegility/9.0.0.0.x/fluent-bit/data/config.d
```

Note:

This file must be copied to this directory to all the nodes in the cluster and the extension of the configuration file should be *.conf

- After copying the file, login to the Ambari UI using valid credentials.

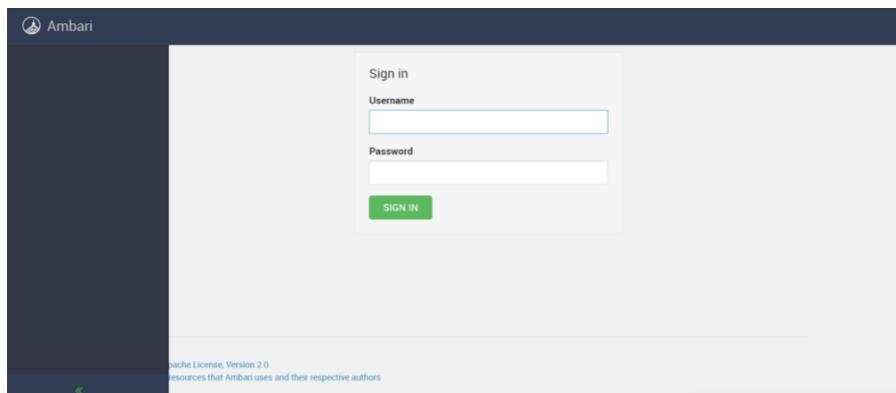


Figure 12-229: Ambari Login Screen

- Select the **BDPPEP** service.
- Click **Actions** and select **Restart All**.

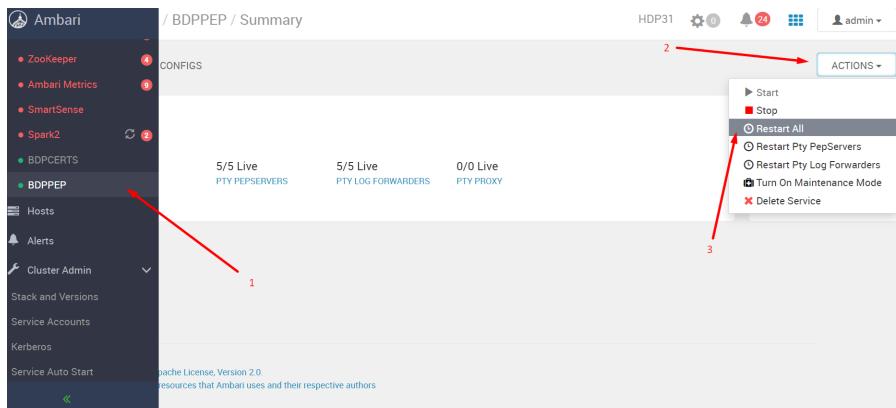


Figure 12-230: Restarting the BDPPEP Service

You should now have an external audit store configured successfully along with the Protegility Elasticsearch Audit Store.

12.6.3.9.2 Adding a New Audit Store and Disabling Protegility Elasticsearch

- After setting and starting up BDPPEP and BDPCERT service using the Ambari UI, login to the CLI on any node that has external audit store connectivity.

2. Create a conf file pertaining to the type and tool that you want to configure as the audit store.

Note:

For more information, refer <https://docs.fluentbit.io/manual/pipeline/outputs>.

Example: Contents of file-xxxx.conf file for file audit store that stores log in /opt/logs.

```
[OUTPUT]
  Name file
  Match *
  Path /opt/logs
```

3. To copy this file to the folder from where fluentbit picks up the configuration files, execute the following command:

```
cp    file-xxxx.conf    /opt/protegility/9.0.0.0.x/fluent-bit/data/config.d
```

Note:

This file must be copied to this directory to all the nodes in the cluster and the extension of the configuration file should be **.conf*

4. After copying the file, login to the Ambari UI using valid credentials.

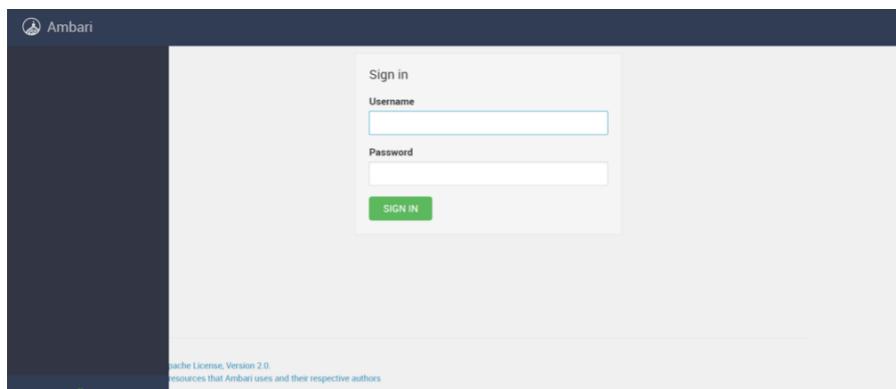


Figure 12-231: Ambari Login Screen

5. Select the **BDPPEP** service and click **CONFIGS**.
6. Expand **Advanced logforwarderconfig** to view the configuration files managed by Protegility.
7. To comment the line in the *out_elastic.conf* file, add # at the start of the uncommented line.

```

config.d/out_elastic.conf

# An Output plugin that normally flushes the records to Elasticsearch.
# In our case they are instead sent to upstream\_es.cfg that enables the possibility to
# send the records to multiple Elasticsearch instances.
# There are three different sections, one for the aggregated data from protectors,
# # one for tailing the internal fluent-bit log and one for application log, separated by the Match attribute.
# They are separated due to if there was a match on wildcard ('*'), then both the aggregated data and non-aggregated
# data would be sent to Elasticsearch.
#
# Match: The tag to match the records.
# Retry\_Limit: The number of retries that should be done if Fluent Bit was not able to process and flush the data.
# If set to False means that there is no limit of number of retries. When tailing the internal logfile
# the Retry\_Limit is set to 1, this is due to that Fluent Bit writes to the internal logfile if it would fail
# and to avoid sending the same message several times.
# Index: Index name, must be pty\_insight\_audit for the record to be recognized by Elasticsearch.
# Type: Type name, to be compliant with Elastic Search it should be set to _doc.
# Time\_Key The name of the timestamp field.
#
#[OUTPUT]
# Name es
# Match egg
# Retry\_Limit False
# Index pty\_insight\_audit
# Type _doc
# Time\_Key ingest_time_utc
# Upstream /opt/protegility/8.1.0.0.12/fluent-bit/data/config.d/upstream\_es.cfg
#
#[OUTPUT]
# Name es
# Match fulog
# Retry\_Limit 1
# Index pty\_insight\_audit
# Type _doc
# Time\_Key ingest_time_utc
# Upstream /opt/protegility/8.1.0.0.12/fluent-bit/data/config.d/upstream\_es.cfg
#
#[OUTPUT]
# Name es
# Match applog
# Retry\_Limit False
# Index pty\_insight\_audit
# Type _doc
# Time\_Key ingest_time_utc
# Upstream /opt/protegility/8.1.0.0.12/fluent-bit/data/config.d/upstream\_es.cfg

```

Figure 12-232: Commenting lines in the *out_elastic.conf* file

8. Click **Save**.
9. Restart the BDPPEP service to configure all the changes and set up the new audit store.

Note:

If you want to revert to the original configuration of the Protegility Elasticsearch, Ambari provides the **Set Recommended** option to roll back the contents in the *out_elastic.conf* file to the original value.

```

config.d/out_elastic.conf

# An Output plugin that normally flushes the records to Elasticsearch.
# In our case they are instead sent to 'upstream_es.cfg' that enables the possibility to
# send the records to multiple Elasticsearch instances.
# There are three different sections, one for the aggregated data from protectors,
# one for tailing the internal fluent-bit.log and one for application log, separated by the Match attribute.
# They are separated due to if there was a match on wildcard (*), then both the aggregated data and non-aggregated
# data would be sent to Elasticsearch.
#
# Match: The tag to match the records.
# Retry_Limit: The number of retries that should be done if Fluent Bit was not able to process and flush the data.
# If set to False means that there is no limit of number of retries. When tailing the internal logfile
# the Retry_Limit is set to 1, this is due to that Fluent Bit writes to the internal logfile if it would fail
# and to avoid sending the same message several times.
# Index: Index name, must be pty_insight_audit for the record to be recognized by Elasticsearch.
# Type: Type name, to be compliant with Elastic Search it should be set to _doc.
# Time_Key: The name of the timestamp field.

[OUTPUT]
Name es
Match applog
Retry_Limit False
Index pty_insight_audit
Type _doc
Time_Key ingest_time_utc
Upstream /opt/protegity/8.1.0.0.12/fluent-bit/data/config.d/upstream_es.cfg

[OUTPUT]
Name es
Match flulog
Retry_Limit 1
Index pty_insight_audit
Type _doc
Time_Key ingest_time_utc
Upstream /opt/protegity/8.1.0.0.12/fluent-bit/data/config.d/upstream_es.cfg

[OUTPUT]
Name es
Match applog
Retry_Limit False
Index pty_insight_audit
Type _doc
Time_Key ingest_time_utc
Upstream /opt/protegity/8.1.0.0.12/fluent-bit/data/config.d/upstream_es.cfg

```

Figure 12-233: Set Recommended option in Ambari

You should now have an external audit store configured successfully. Depending on how you have configured the *out_elastic.conf* file, the Protegity Elasticsearch will be disabled.

Note:

- It is recommended to comment out the contents of the *out_elastic.conf* file from the Ambari UI rather than deleting the entire content.
- You can check for failure logs after you have restarted the BDPPEP service to understand the reason for the error.
- Ensure that the newly created conf file has the correct access permissions for Fluentbit to detect and use.

12.6.3.10 Uninstalling the Big Data Protector Services from the Nodes

If you need to uninstall the Big Data Protector from all the nodes, then perform the following tasks.

1. Stop the Big Data Protector services on the nodes
2. Remove the Big Data Protector services from all the nodes

12.6.3.10.1 Stopping the Big Data Protector Services on the Nodes

Before you uninstall Big Data Protector, stop the following Big Data Protector services on all the nodes:

- **BDPPEP:** Big Data Protector PEP service
- **BDPCERTS:** Certificates service

► To stop the Big Data Protector Services on the Nodes in the Cluster:

1. On the Ambari screen, select the BDPPEP service.

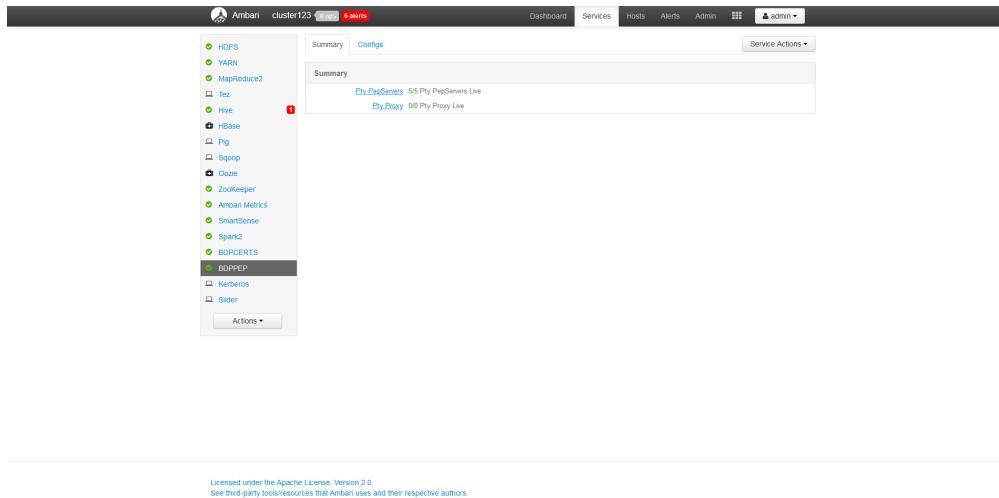


Figure 12-234: BDPPEP Service

2. Click the **Service Actions** button.

The Service Actions drop-down appears.

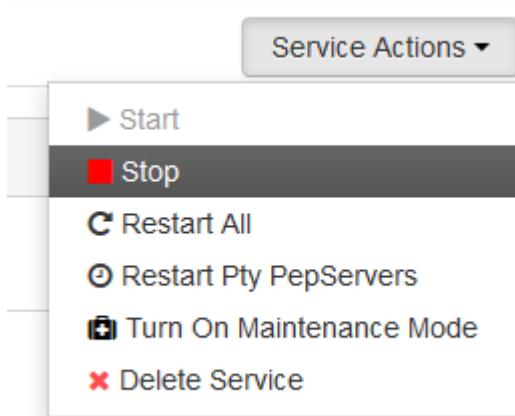


Figure 12-235: BDPPEP Service Actions drop-down menu

3. Click **Stop**.

A confirmation message for stopping the BDPPEP service appears.

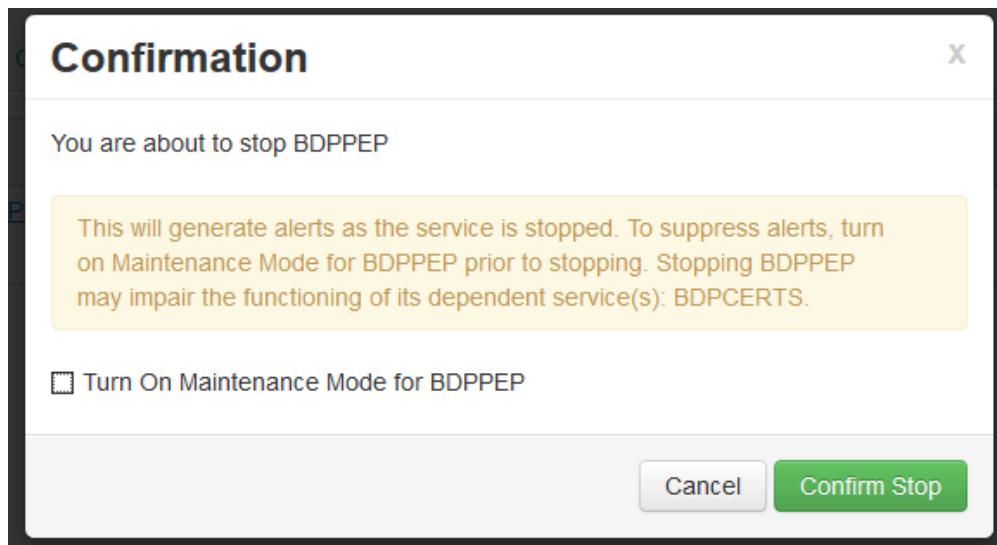


Figure 12-236: Confirmation for Stopping the BDPPEP Service

- Click **Confirm Stop**.

The BDPPEP service and its dependent services are terminated and the following screen appears.

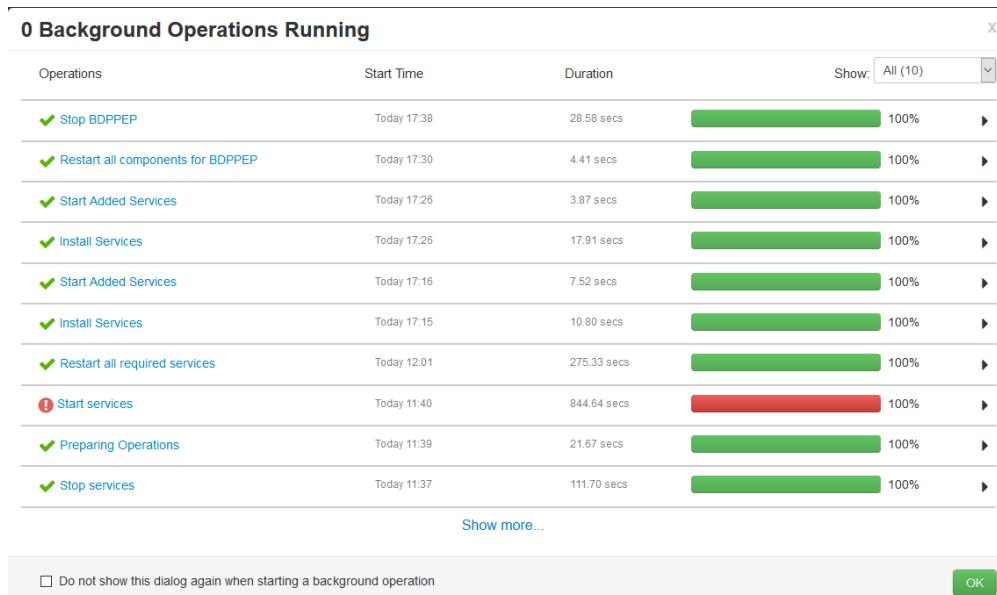
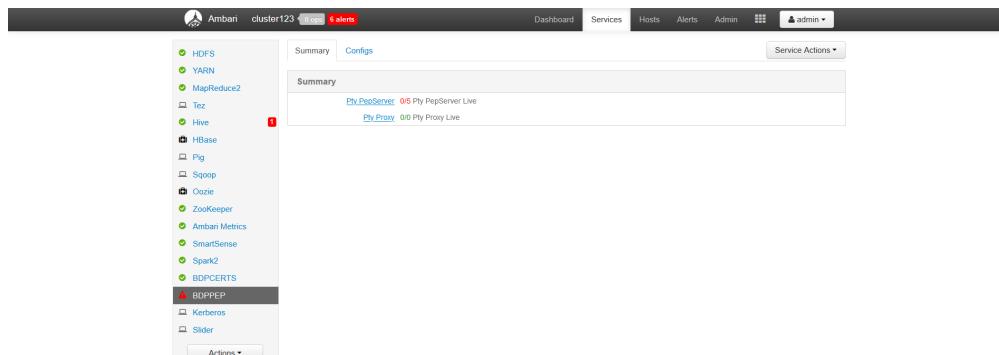


Figure 12-237: BDP PEP Service Stopped

- Click **OK**.

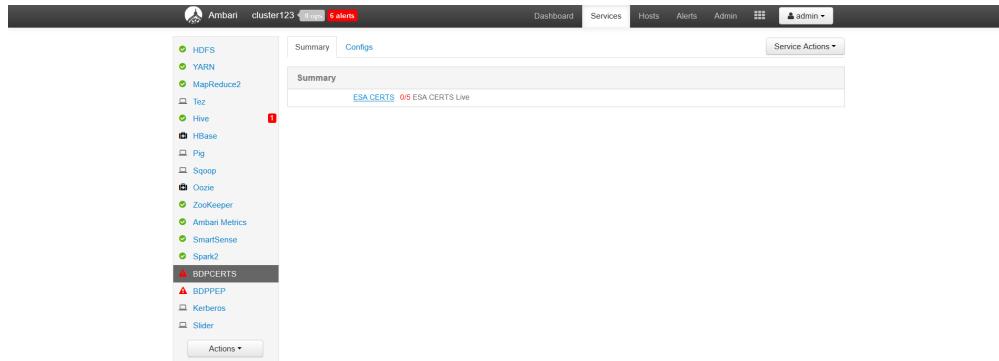
The *BDP PEP* service is stopped and the status of the service is updated.



Licensed under the Apache License, Version 2.0
See third-party tools/resources that Ambari uses and their respective authors

Figure 12-238: BDP PEP Service Stopped in the Cluster

6. Select the BDPCERTS service.
7. Click the **Service Actions** button.
The Service Actions drop-down appears.
8. Click **Stop**.
A confirmation message for stopping the BDPCERTS service appears.
9. Click **Confirm Stop**.
The BDPCERTS service and its dependent services are terminated.
10. Click **OK**.
The BDPCERTS service is stopped and the status of the service is updated.



Licensed under the Apache License, Version 2.0
See third-party tools/resources that Ambari uses and their respective authors

Figure 12-239: BDPCERTS Service Stopped in the Cluster

12.6.3.10.2 Deleting the Big Data Protector Services from the Nodes

After stopping the Big Data Protector services on the nodes, delete the following Big Data Protector services from all the nodes to uninstall Big Data Protector.

- **BDPCERTS:** Certificates service
- **BDPPEP:** Big Data Protector PEP service

► To delete the Big Data Protector Services from the Nodes in the Cluster:

1. On the Ambari screen, select the *BDPCERTS* service.

The screenshot shows the Ambari interface for a cluster named 'cluster123'. The sidebar on the left lists various services: HDFS, YARN, MapReduce2, Tez, Hive, HBase, Pig, Sqoop, Ozone, ZooKeeper, Ambari Metrics, SmartSense, Spark2, BDPCERTS (which is highlighted in red), BDPEP, Kerberos, and Slider. The main content area shows a summary for the 'ESA CERTS' service, which has 0/5 ESA CERTS Live. A 'Service Actions' button is visible at the top right of the main content area. A note at the bottom states: 'Licensed under the Apache License, Version 2.0. See third-party tools/resources that Ambari uses and their respective authors.'

Figure 12-240: BDPCERTS Service

2. Click the **Service Actions** button.

The Service Actions drop-down appears.

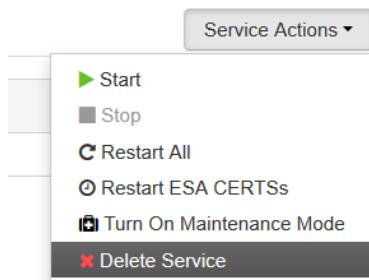


Figure 12-241: BDPCERTS Service Actions drop-down menu

3. Click **Delete Service**.

A message for deleting the *BDPCERTS* service appears.

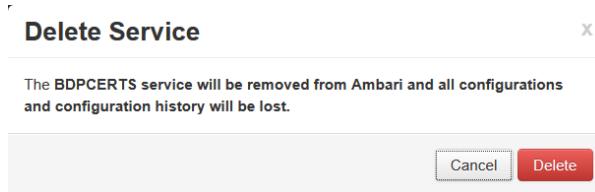


Figure 12-242: Message for Deleting the BDPCERTS Service

4. Click **Delete**.

A confirmation message for deleting the *BDPCERTS* service appears.

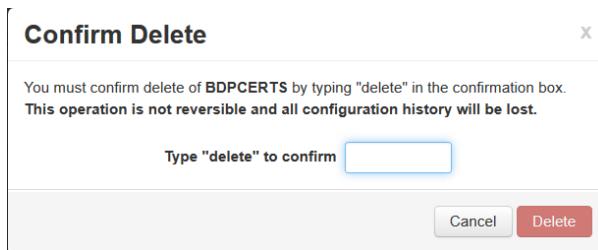


Figure 12-243: Confirmation for Deleting the BDPCERTS Service

5. Type **delete** in the text box to confirm the deletion of the **BDPCERTS** service.

6. Click **Delete**.

The BDPCERTS service is deleted from the nodes in the cluster.

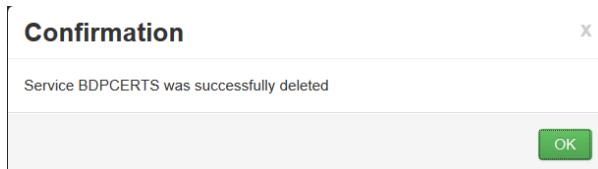


Figure 12-244: BDPCERTS Service Deleted from the Nodes in the Cluster

7. Select the **BDPPEP** service.

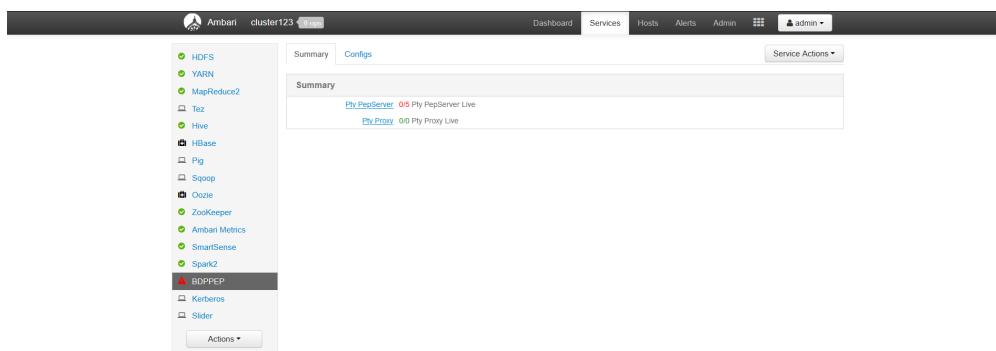


Figure 12-245: BDPPEP Service

8. Click the **Service Actions** button.

The Service Actions drop-down appears.

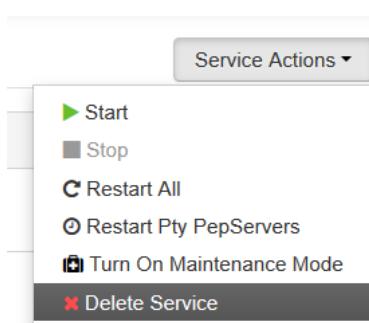


Figure 12-246: BDPPEP Service Actions drop-down menu

9. Click **Delete Service**.

A message for deleting the *BDPPEP* service appears.

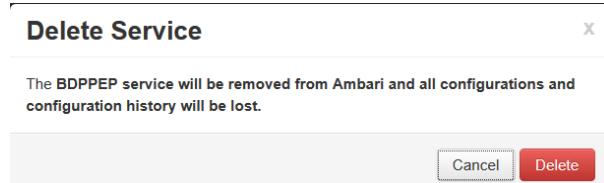


Figure 12-247: Message for Deleting the BDPPEP Service

- Click **Delete**.

A confirmation message for deleting the *BDPPEP* service appears.

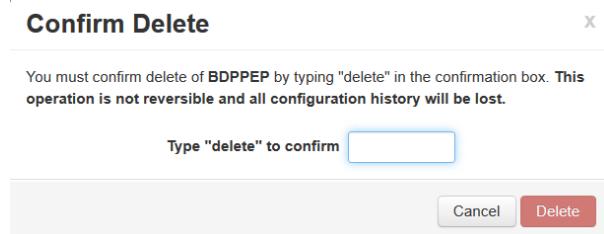


Figure 12-248: Confirmation for Deleting the BDPPEP Service

- Type *delete* in the text box to confirm the deletion of the *BDPPEP* service.

- Click **Delete**.

The *BDPPEP* service is deleted from the nodes in the cluster and the following confirmation appears.

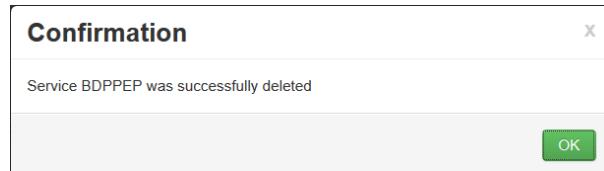


Figure 12-249: BDPPEP Service Deleted from the Nodes in the Cluster

- Perform the following steps to delete the directories containing Big Data Protector-related files from all the nodes in the cluster.

a. On each node in the HDP cluster, delete the *<PROTEGILITY_DIR>* directory, which is the installation directory for the Big Data Protector.

b. If you are using Ambari, version 2.6 and lower, then perform the following steps on the node containing the Ambari Server.

i. Verify the directories containing the BDPPEP and BDPCERTS-related files by running the following commands.

```
sudo find / -name "BDPPEP"
```

```
sudo find / -name "BDPCERTS"
```

ii. Delete the directories containing the BDPPEP-related files by running the following commands.

```
sudo rm -rf /var/lib/ambari-server/resources/common-services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/stacks/HDP/2.6/services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-
mpack-7.2.1.<x>/common-services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-
mpack-7.2.1.<x>/custom-services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/cache/BDPPEP-AMBARI-MPACK-7.2.1.<x>.tar.gz
```

- iii. Delete the directories containing the BDPCERTS-related files by running the following commands.

```
sudo rm -rf /var/lib/ambari-server/resources/common-services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/stacks/HDP/2.6/services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-mpack-7.2.1.<x>/common-services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-mpack-7.2.1.<x>/custom-services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/cache/BDPCERTS-AMBARI-MPACK-7.2.1.<x>.tar.gz
```

- iv. Delete the directories containing the name BDP*** using the following command.

```
sudo rm -rf /var/lib/ambari-server/data/tmp/BDP*
```

- c. If you are using Ambari, version 2.6 and lower, then perform the following steps on the nodes containing the Ambari Agent.

- i. Verify the directories containing the BDPPEP and BDPCERTS-related files by running the following commands.

```
sudo find / -name "BDPPEP"
```

```
sudo find / -name "BDPCERTS"
```

- ii. Delete the directories containing the BDPPEP-related files by running the following command.

```
sudo rm -rf /var/lib/ambari-agent/cache/common-services/BDPPEP
```

- iii. Delete the directories containing the BDPCERTS-related files by running the following command.

```
sudo rm -rf /var/lib/ambari-agent/cache/common-services/BDPCERTS
```

- d. If you are using Ambari, version 2.6 and higher, then perform the following steps on the node containing the Ambari Server.

- i. Run the following commands to uninstall the Management packs for BDPPEP and BDPCERTS.

```
ambari-server uninstall-mpack --mpack-name="bdp-ambari-mpack"
```

```
ambari-server uninstall-mpack --mpack-name="certs-ambari-mpack"
```

- ii. Verify the directories containing the BDPPEP and BDPCERTS-related files by running the following commands.

```
sudo find / -name "BDPPEP"
```

```
sudo find / -name "BDPCERTS"
```

- iii. Delete the directories containing the BDPPEP-related files by running the following commands.

```
sudo rm -rf /var/lib/ambari-server/resources/stacks/HDP/2.<x>/services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-mpack-7.2.1.<x>/common-services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-
mpack-7.2.1.<x>/custom-services/BDPPEP
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/cache/BDPPEP-AMBARI-
MPACK-7.2.1.<x>.tar.gz
```

- iv. Delete the directories containing the BDPCERTS-related files by running the following commands.

```
sudo rm -rf /var/lib/ambari-server/resources/stacks/HDP/2.<x>/services/
BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-
mpack-7.2.1.<x>/common-services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/bdp-ambari-
mpack-7.2.1.<x>/custom-services/BDPCERTS
```

```
sudo rm -rf /var/lib/ambari-server/resources/mpacks/cache/BDPCERTS-AMBARI-
MPACK-7.2.1.<x>.tar.gz
```

- v. Delete the directories containing the name BDP*** using the following command.

```
sudo rm -rf /var/lib/ambari-server/data/tmp/BDP*
```

- e. If you are using Ambari, version 2.6 and higher, then perform the following steps on the nodes containing the Ambari Agent.

- i. Verify the directories containing the BDPPEP and BDPCERTS-related files by running the following commands.

```
sudo find / -name "BDPPEP"
```

```
sudo find / -name "BDPCERTS"
```

- ii. Delete the directories containing the BDPPEP-related files by running the following command.

```
sudo rm -rf /var/lib/ambari-agent/cache/common-services/BDPPEP
```

- iii. Delete the directories containing the BDPCERTS-related files by running the following command.

```
sudo rm -rf /var/lib/ambari-agent/cache/common-services/BDPCERTS
```

Note: After all the Big Data Protector-related services are stopped and removed from the nodes in the cluster, you need to roll back the configuration parameters to their previous values.

For instance, if you need to roll back the configuration parameters for MapReduce, then perform the following steps.

1. From the Ambari Home screen, click on the **MapReduce2** service from the left pane to display the MapReduce-related summary.
2. Click the *Configs* tab to open the configuration parameters for MapReduce.
3. Enter the name of the configuration parameter that needs to be updated, in the *Filters* search box.
4. Click the **C** icon to roll back the recommended configuration value for the parameter to the previous value.
Ensure that the Protegility recommended values are rolled back to their previous values in the configuration parameter.
5. Click **Save** to save the modifications made to the configuration parameters.
6. If you need to roll back the recommended values to the previous values for other configuration parameters, go to step 3.

Similarly, you need to roll back the values for the configuration parameters to the previous values depending on the components that you used with the Big Data Protector.

For more information about rolling back the configuration parameters to their previous values, refer to section *3.6 Setting the Configuration for Big Data Protector* in the [Protegility Big Data Protector Guide 9.0.0.0](#).

Note: If you are using HBase, then remove the following HBase coprocessor region classes property value from the *hbase-site.xml* file in all the respective region server groups, using the Ambari UI.

```
com.protegility.hbase.PTYRegionObserver
```

12.6.4 Installing Big Data Protector using the Shell-based Installer

This section describes the procedures to install and uninstall Protegility Big Data Protector using the shell-based installer.

Note: It is recommended to use native installers for the Cloudera and Ambari environments, as they simplify the task of installing, configuring, and managing Big Data Protector using Cloudera Manager or the Ambari UI.

Note: The Shell-based installer for Big Data Protector is certified till the 7.1 release only. If you need the Shell-based installer binaries for Big Data Protector 7.2.0, then contact Protegility Support.

Note: The HDFSFP-related components in the Shell-based installer (such as the *PepHdfsFp_Setup* shell script, *XCPep2Ini_Setup* shell script, Protegility Cache Control (*cluster_cachesrvctl.sh*) utility, Recover utility, and the Talend-related files) will not be available starting from the Big Data Protector 7.2.0 release, as the HDFS File Protector (HDFSFP) is deprecated.

Note: Starting from the Big Data Protector 6.6.4 release, you do not require *root* access to install Big Data Protector on a cluster.

You need a *sudoer* user account to install Big Data Protector on a cluster.

12.6.4.1 Verifying Prerequisites for Installing Big Data Protector

Ensure that the following prerequisites are met, before installing Big Data Protector:

- The Hadoop cluster is installed, configured, and running.
- ESA appliance version 7.2.1 is installed, configured, and running.
- The following ports are configured on the ESA and the nodes in the cluster, which will run Big Data Protector:
 - ESA: Ensure that the Big Data Protector nodes can communicate with the ESA on the port *8443*.
 - Big Data Protector nodes: Ensure that on each node (localhost), the port *16700* is open and not allocated to any other process.
 - Proxy node: If a proxy node is used between the ESA and Big Data Protector nodes, then the port *8443* needs to be open on the proxy node.
- A *sudoer* user account with privileges to perform the following tasks:
 - Update the system by modifying the configuration, permissions, or ownership of directories and files.
 - Perform third party configuration.
 - Create directories and files.
 - Modify the permissions and ownership for the created directories and files.
 - Set the required permissions to the create directories and files for the Protegility Service Account.
 - Permissions for using the SSH service.
- The *sudoer* password is the same across the cluster.
- The following user accounts to perform the required tasks:



- ***ADMINISTRATOR_USER***: It is the *sudoer* user account that is responsible to install and uninstall the Big Data Protector on the cluster.

This user account needs to have *sudo* access to install the product.

- ***EXECUTOR_USER***: It is a user that has ownership of all Protegity files, directories, and services.
- ***OPERATOR_USER***: It is responsible for performing tasks such as, starting or stopping tasks, monitoring services, updating the configuration, and maintaining the cluster while the Big Data Protector is installed on it.

If you need to start, stop, or restart the Protegity services, then you need *sudoer* privileges for this user to impersonate the *EXECUTOR_USER*.

Note: Depending on the requirements, a single user on the system may perform multiple roles.

If a single user is performing multiple roles, then ensure that the following conditions are met:

- The user has the required permissions and privileges to impersonate the other user accounts, for performing their roles, and perform tasks as the impersonated user.
- The user is assigned the highest set of privileges, from the required roles that it needs to perform, to execute the required tasks.

For instance, if a single user is performing tasks as *ADMINISTRATOR_USER*, *EXECUTOR_USER*, and *OPERATOR_USER*, then ensure that the user is assigned the privileges of the *ADMINISTRATOR_USER*.

- The management scripts provided by the installer in the *cluster_utils* directory should be run only by the user (*OPERATOR_USER*) having privileges to impersonate the *EXECUTOR_USER*.
- If the value of the *AUTOCREATE_PROTEGITY_IT_USR* parameter in the *BDP.config* file is set to *No*, then ensure that a service group containing a user for running the Protegity services on all the nodes in the cluster already exists. Ensure that the service group created is the primary group for the service account user to run the Protegity services.
- If the Hadoop cluster is configured with LDAP or AD for user management, then ensure that the *AUTOCREATE_PROTEGITY_IT_USR* parameter in the *BDP.config* file is set to *No* and that the required service account user is created on all the nodes in the cluster. Ensure that the service group created is the primary group for the service account user to run the Protegity services.
- If Big Data Protector, version 6.6.3, with build version 6.6.3.15, or lower, was previously installed and the following Spark protector APIs for Encryption/Decryption are utilized:
 - *public void protect(String dataElement, List<Integer> errorIndex, short[] input, byte[][] output)*
 - *public void protect(String dataElement, List<Integer> errorIndex, int[] input, byte[][] output)*
 - *public void protect(String dataElement, List<Integer> errorIndex, long[] input, byte[][] output)*
 - *public void unprotect(String dataElement, List<Integer> errorIndex, byte[][] input, short[] output)*
 - *public void unprotect(String dataElement, List<Integer> errorIndex, byte[][] input, int[] output)*
 - *public void unprotect(String dataElement, List<Integer> errorIndex, byte[][] input, long[] output)*

For more information, refer to the [Advisory for Spark Protector APIs](#), before installing Big Data Protector, version 6.6.5.

- If the Big Data Protector was previously installed then uninstall it. In addition, delete the *<PROTEGITY_DIR>* directory from the Lead node. If the */var/log/protegity/* directory exists on any node in the cluster, then ensure that it is empty.
- Password based authentication is enabled in the *sshd_config* file before installation. After the installation is completed, this setting might be reverted back by the system administrator.
- The *lsb_release* library is present on the client machine, at least on the Lead node.

The **Lead node** can be any node, such as the Name node, Data node, or Edge node, that can access the Hadoop cluster. The Lead node would be driving the installation of the Big Data Protector across the Hadoop cluster and is responsible for managing the Big Data Protector services throughout the cluster.



If the *lsb_release* library is not present, then the installation of the Big Data Protector fails. This can be verified by running the following command.

lsb_release

- If you are configuring the Big Data Protector with a Kerberos-enabled Hadoop cluster, then ensure that the HDFS superuser (*hdfs*) has a valid Kerberos ticket.

For more information about creating data elements, security policies, and user roles, refer to *Protegity Enterprise Security Administrator Guide 9.1.0.5* and *Protegity Policy Management Guide 9.1.0.5*.

12.6.4.2 Extracting Files from the Installation Package

► To extract the files from the installation package:

1. After receiving the installation package from Protegity, copy it to the Lead node in any temporary directory, such as */opt/bigdata*.
2. Extract the files from the installation package using the following command:

```
tar -xf BigDataProtector_<OS>-<arch>_<Big data distribution>-64_7.2.0.x.tgz
```

The following files are extracted:

- *BDP.config*
- *BdpInstallx.x.x_Linux_<arch>_7.2.0.x.sh*
- *JpepLiteSetup_Linux_<arch>_7.2.0.x.sh*
- *node_uninstall.sh*
- *PepHbaseProtectorx.x.xSetup_Linux_<arch>_<distribution>-x.x_7.2.0.x.sh*
- *PepHivex.x.xSetup_Linux_<arch>_<distribution>-x.x_7.2.0.x.sh*
- *PepImpalax.xSetup_<OS>-x86-<arch>_7.2.0.x.sh*, only if it is a Cloudera or HPE Ezmeral Data Fabric distribution
- *PepMapreduce.x.xSetup_Linux_<arch>_<distribution>-x.x_7.2.0.x.sh*
- *PepPigx.x.xSetup_Linux_<arch>_<distribution>-x.x_7.2.0.x.sh*
- *PepServer_Setup_Linux_<arch>_7.2.0.x.sh*
- *PepSparkx.x.xSetup_Linux_<arch>_<distribution>-x.x_7.2.0.x.sh*
- *ptyLogAnalyzer.sh*
- *ptyLog_Consolidator.sh*
- *uninstall.sh*

12.6.4.3 Updating the *BDP.config* File

Note: Ensure that the *BDP.config* file is updated before the Big Data Protector is installed.

Do not update the *BDP.config* file when the installation of the Big Data Protector is in progress.

► To update the *BDP.config* file:

1. Create a file containing a list of all nodes in the cluster, except the Lead node, and specify it in the *BDP.config* file. This file is used by the installer for installing Big Data Protector on the nodes.
2. Open the *BDP.config* file in any text editor and modify the following parameter values:
 - HADOOP_DIR – The installation home directory for the Hadoop distribution.
 - PROTEGRITY_DIR – The directory where the Big Data Protector will be installed.

The samples and examples used in this document assume that the Big Data Protector is installed in the */opt/protegility/* directory.

 - HOSTS – This file contains the host name or IP addresses all the nodes in the cluster, except the Lead node, listing one host name and IP address per line.

Ensure that you specify the file name with the complete path.

 - SPARK_PROTECTOR – Specifies one of the following values, as required:
 - Yes – The installer installs the Spark protector.
 - No – The installer does not install the Spark protector.
 - AUTOCREATE_PROTEGRITY_IT_USR – This parameter determines the Protegility service account. The service group and service user name specified in the *PROTEGRITY_IT_USR_GROUP* and *PROTEGRITY_IT_USR* parameters respectively will be created if this parameter is set to *Yes*. One of the following values can be specified, as required:
 - Yes – The installer creates a service group *PROTEGRITY_IT_USR_GROUP* containing the user *PROTEGRITY_IT_USR* for running the Protegility services on all the nodes in the cluster.

If the service group or service user are already present, then the installer exits.

If you uninstall the Big Data Protector, then the service group and the service user are deleted.

 - No – The installer does not create a service group *PROTEGRITY_IT_USR_GROUP* with the service user *PROTEGRITY_IT_USR* for running the Protegility services on all the nodes in the cluster.

Ensure that a service group containing a service user for running Protegility services has been created, as described in section [Verifying Prerequisites for Installing Big Data Protector](#).

 - PROTEGRITY_IT_USR_GROUP – This service group is required for running the Protegility services on all the nodes in the cluster. All the Protegility installation directories are owned by this service group.
 - PROTEGRITY_IT_USR – This service account user is required for running the Protegility services on all the nodes in the cluster and is a part of the group *PROTEGRITY_IT_USR_GROUP*. All the Protegility installation directories are owned by this service user.
 - HADOOP_NATIVE_DIR – The Hadoop native directory. This parameter needs to be specified if you are using HPE Ezmeral Data Fabric.
 - HADOOP_SUPER_USER – The Hadoop super user name. This parameter needs to be specified if you are using HPE Ezmeral Data Fabric.

12.6.4.4 Setting up the Proxy

If you need to configure a proxy, which connects to the ESA on one end, and communicates with the Big Data Protector cluster nodes containing on the other end, then perform the following task.

► To setup the Proxy for the Big Data Protector Nodes in the Cluster:

1. Copy the *ProxySetup_<OS>_<arch>_7.2.0.x.tgz* to the node, which will act as the proxy, and communicate with ESA and the nodes in the Big Data Protector cluster.

2. Extract the Proxy archive using the following command.

```
tar -xvf ProxySetup_<OS>_<arch>_7.2.0.x.tgz
```

3. Press ENTER.

The *ProxySetup_<OS>_<arch>_7.2.0.x.sh* file is extracted.

4. Setup the proxy on the required node using the following command.

```
./ProxySetup_<OS>_<arch>_7.2.0.x.sh -esa host {-dir <Installation_Directory>} {-port <Proxy_Port>}
```

The *-dir <Installation_Directory>* parameter is optional. If it is not specified, then the Proxy is installed in *<PROTEGERITY_DIR>/proxy directory*.

The *-port <Proxy_Port>* parameter is optional. If it is not specified, then the Proxy port is considered as *8443* by default.

5. Navigate to the Proxy directory, such as *<Installation_Directory>/proxy/bin/*.

6. Start the Proxy using the following command.

```
./proxycctl start
```

7. Ensure that all the proxy services are up and running after setting up the proxy for the Big Data Protector cluster using the following command.

```
./proxycctl status
```

The status of the Proxy service appears.

12.6.4.5 Installing Big Data Protector

► To install the Big Data Protector:

1. As a *sudoer* user, run *BdpInstallx.x.x_Linux_<arch>_7.2.1.x.sh* from the directory where it is extracted.

A prompt to confirm or cancel the Big Data Protector installation appears.

2. Type *yes* to continue with the installation.

The Big Data Protector installation starts.

If you are using a Cloudera or HPE Ezmeral Data Fabric distribution, then the presence of the HDFS connection is also verified.

A prompt to enter the *sudoer* password for the *ADMINISTRATOR* user appears.

3. Enter the *sudoer* password.

A prompt to enter the ESA IP address or proxy IP address appears.

4. If a proxy is configured, then enter the proxy IP address.

Alternatively, enter the ESA IP address.

A prompt to enter the ESA user name appears.

5. Enter the ESA user name (Security Officer).

The PEP Server Installation wizard starts.

6. When prompted, perform the following steps to download the ESA certificates.

- a. Specify the Security Officer user with administrative privileges.

- b. Specify the Security Officer password for the ESA certificates.

The installer then installs the Big Data Protector on all the nodes in the cluster.

The status of the installation of the individual components appears, and the log files for all the required components on all the nodes in the cluster are stored on the Lead node in the `<PROTEGILITY_DIR>/cluster_utils/logs` directory.

Verify the installation report, that is generated at `<PROTEGILITY_DIR>/cluster_utils/installation_report.txt` to ensure that the installation of all the components is successful on all the nodes in the cluster.

Verify the `bdp_setup.log` file confirm if the Big Data Protector was installed successfully on all the nodes in the cluster.

7. Restart the MapReduce (MRv1) or Yarn (MRv2) services on the Hadoop cluster.

The installer installs the following components in the installation directory of the Big Data Protector:

- PEP server in the `<PROTEGILITY_DIR>/defiance_dps` directory
- Jpeplite in the `<PROTEGILITY_DIR>/jpeplite` directory
- MapReduce protector in the `<PROTEGILITY_DIR>/pepmapreduce/lib` directory
- Hive protector in the `<PROTEGILITY_DIR>/pephive/lib` directory
- Pig protector in the `<PROTEGILITY_DIR>/peppig/lib` directory
- HBase protector in the `<PROTEGILITY_DIR>/pephbase-protector/lib` directory
- Impala protector in the `<PROTEGILITY_DIR>/pepimpala` directory, if you are using a Cloudera or HPE Ezmeral Data Fabric distribution
- `pepspark-xxx.jar` in the `<PROTEGILITY_DIR>/pepspark/lib` directory, only if the value of the SPARK parameter in the `BDP.config` file is specified as Yes
- Cluster Utilities in the `<PROTEGILITY_DIR>/cluster_utils` directory

The following files and directories are present in the `<PROTEGILITY_DIR>/cluster_utils` directory:

- `BdpInstallx.x.x_Linux_<arch>_7.2.1.x.sh` utility to install the Big Data Protector on any node in the cluster.

For more information about using the `BdpInstallx.x.x_Linux_<arch>_7.2.1.x.sh` utility, refer to section [Installing Big Data Protector on New Nodes added to a Hadoop Cluster](#).

- `cluster_pepsrvctrl.sh` utility for managing PEP servers on all nodes in the cluster.
- `uninstall.sh` utility to uninstall the Big Data Protector from all the nodes in the cluster.
- `node_uninstall.sh` to uninstall the Big Data Protector from any nodes in the cluster.

For more information about using the `node_uninstall.sh` utility, refer to section [Uninstalling Big Data Protector from Selective Nodes in the Hadoop Cluster](#).

- `update_cluster_policy.sh` utility for updating PEP servers when a new policy is deployed.
- `BDP.config` file
- `HOSTS` file, which is a file containing a list of all the nodes, except the Lead node.
- `installation_report.txt` file that contains the status of installation of all the components in the cluster.
- `logs` directory that contains the consolidated setup logs from all the nodes in the cluster.

8. Starting with the Big Data Protector, version 6.6.4, the Bulk APIs in the MapReduce protector will return the detailed error and return codes instead of `0` for *failure* and `1` for *success*.

For more information about the error codes for Big Data Protector, version 8.0.0.0, refer to *Table 9-2 PEP Log Return Codes* and *Table 9-3 PEP Result Codes* in the [Protegility Big Data Protector Guide 9.1.0.0](#).

If the older behaviour from the Big Data Protector, version 6.6.3 or lower with the Bulk APIs in the MapReduce protector is desired, then perform the following steps to enable the Backward compatibility mode to retain the same error handling capabilities.

- a. If you are using HDP, version 2.2 or higher (Hortonworks), or PHD, version 3.0 or higher (Pivotal Hadoop), then append the following entry to the `mapreduce.admin.reduce.child.java.opts` property in the `mapred-site.xml` file.

-Dpty.mr.compatibility=old

- If you are using HDP, version 2.2 or higher (Hortonworks), or PHD, version 3.0 or higher (Pivotal Hadoop), then append the following entry to the *mapreduce.admin.map.child.java.opts* property in the *mapred-site.xml* file.

-Dpty.mr.compatibility=old

- If you are using CDH, then add the following values to the *Yarn Service Mapreduce Advanced Configuration Snippet (Safety Valve)* parameter in the *mapred-site.xml* file.

```
<property>
<name>mapreduce.admin.map.child.java.opts</name>
<value>-Dpty.mr.compatibility=old</value>
</property>

<property>
<name>mapreduce.admin.reduce.child.java.opts</name>
<value>-Dpty.mr.compatibility=old</value>
</property>
```

- If you are using HDP, version 2.2 or higher (Hortonworks), or PHD, version 3.0 or higher (Pivotal Hadoop), then perform the following steps.

- Ensure that the *mapreduce.application.classpath* property in the *mapred-site.xml* file contains the following entries.

```
<PROTEGRITY_DIR>/pepmapper/lib/*
<PROTEGRITY_DIR>/pephive/lib/*
<PROTEGRITY_DIR>/peppig/lib/*
```

Ensure that the above entry is before all other entries in the *mapreduce.application.classpath* property.

- Ensure that the *yarn.application.classpath* property in the *yarn-site.xml* file contains the following entries.

```
<PROTEGRITY_DIR>/pepmapper/lib/*
<PROTEGRITY_DIR>/pephive/lib/*
<PROTEGRITY_DIR>/peppig/lib/*
```

Ensure that the above entry is before all other entries in the *yarn.application.classpath* property.

- Restart the Yarn service.
- Restart the MRv2 service.
- Ensure that the *tez.cluster.additional.classpath.prefix* property in the *tez-site.xml* file contains the following entries.

```
<PROTEGRITY_DIR>/pepmapper/lib/*
<PROTEGRITY_DIR>/pephive/lib/*
<PROTEGRITY_DIR>/peppig/lib/*
```

Ensure that the above entry is before all other entries in the *tez.cluster.additional.classpath.prefix* property.

- Restart the Tez services.
- If you need to use the Hive protector, then perform the following steps.
- Specify the following value for the *hive.exec.pre.hooks* property in the *hive-site.xml* file.

```
hive.exec.pre.hooks=com.protegity.hive.PtyHiveUserPreHook
```

- Restart the Hive services to ensure that the updates are propagated to all the nodes in the cluster.

Note: If you are using Beeline or Hue, then ensure that Protegity Big Data Protector is installed on the following machines:

- For Beeline: The machines where Hive Metastore, and HiveServer2 are running.

- For Hue: The machines where HueServer, Hive Metastore, HiveServer2 are running.

Note: If you require the PEP Server service to start automatically after every reboot of the system, then define the PEP Server service in the startup with the required run levels.

For more info about starting the PEP Server service automatically, refer to [Protegility Installation Guide 9.1.0.5](#).

12.6.4.6 Installing or Uninstalling Big Data Protector on Specific Nodes

This section describes the following procedures:

- Installing Big Data Protector on New Nodes added to a Hadoop cluster
- Uninstalling Big Data Protector from a Nodes in the Hadoop cluster

12.6.4.6.1 Installing Big Data Protector on New Nodes added to a Hadoop Cluster

If you need to install Big Data Protector on new nodes added to a Hadoop cluster, then use the `BdpInstallx.x.x_Linux_<arch>_7.2.0.x.sh` utility in the `<PROTEGRITY_DIR>/cluster_utils` directory.

Note: Ensure that you install the Big Data Protector from an *ADMINISTRATOR* user having full *sudoer* privileges.

► To install Big Data Protector on New Nodes added to a Hadoop Cluster:

- Login to the Lead Node.
- Navigate to the `<PROTEGRITY_DIR>/cluster_utils` directory.
- Add additional entries for each new node, on which the Big Data Protector needs to be installed, in the `NEW_HOSTS_FILE` file.

The new nodes from the `NEW_HOSTS_FILE` file will be appended to the `HOSTS` file.

- Execute the following command utility to install Big Data Protector on the new nodes.

`./BdpInstallx.x.x_Linux_<arch>_7.2.0.x.sh -a <NEW_HOSTS_FILE>`

The Protegility Big Data Protector is installed on the new nodes.

12.6.4.6.2 Uninstalling Big Data Protector from Selective Nodes in the Hadoop Cluster

If you need to uninstall Big Data Protector from selective nodes in the Hadoop cluster, then use the `node_uninstall.sh` utility in the `<PROTEGRITY_DIR>/cluster_utils` directory.

Note: Ensure that you uninstall the Big Data Protector from an *ADMINISTRATOR* user having full *sudoer* privileges.

► To uninstall Big Data Protector from Selective Nodes in the Hadoop Cluster:

- Login to the Lead Node.
- Navigate to the `<PROTEGRITY_DIR>/cluster_utils` directory.
- Create a new hosts file (such as `NEW_HOSTS_FILE`).

The *NEW_HOSTS_FILE* file contains the required nodes on which the Big Data Protector needs to be uninstalled.

4. Add the nodes from which the Big Data Protector needs to be uninstalled in the new hosts file.
5. Execute the following command to remove the Big Data Protector from the nodes that are listed in the new hosts file.

```
./node_uninstall.sh -c NEW_HOSTS_FILE
```

The Big Data Protector is uninstalled from the nodes listed in the new hosts file.

6. Remove the nodes from which the Big Data Protector is uninstalled in Step 5 from the HOSTS file.

12.6.4.7 Uninstalling Big Data Protector from a Cluster

This section describes the procedure for uninstalling the Big Data Protector from the cluster.

12.6.4.7.1 Verifying the Prerequisites for Uninstalling Big Data Protector

If you are configuring the Big Data Protector with a Kerberos-enabled Hadoop cluster, then ensure that the HDFS superuser (*hdfs*) has a valid Kerberos ticket.

12.6.4.7.2 Removing the Cluster from the ESA

Before uninstalling Big Data Protector from the cluster, the cluster should be deleted from the ESA.

For more information about deleting the cluster from the ESA, refer to section *Removing a Cluster* in the *Protegility Big Data Protector Guide 9.1.0.0*.

12.6.4.7.3 Rolling Back the Configuration Updates from the Cluster

Depending on the requirements, perform the following tasks to uninstall the Big Data Protector from the cluster.

For more information about rolling back the configuration updates from the cluster, refer to section *Rolling back the Configuration Updates* in the *Protegility Big Data Protector Guide 9.1.0.0*.

12.6.4.7.4 Running the Uninstallation Script

► To run the scripts for uninstalling the Big Data Protector on all nodes in the cluster:

1. Login as the *sudoer* user and navigate to the *<PROTEGILITY_DIR>/cluster_utils* directory on the Lead node.
2. Run the following script to stop the PEP servers on all the nodes in the cluster.

```
./cluster_pepsrvctrl.sh
```

3. Run the *uninstall.sh* utility.
A prompt to confirm or cancel the Big Data Protector uninstallation appears.
4. Type *yes* to continue with the uninstallation.
5. When prompted, enter the *sudoer* password.

The uninstallation script continues with the uninstallation of Big Data Protector.

If you are using a Cloudera or HPE Ezmeral Data Fabric distribution, then the presence of an HDFS connection and a valid Kerberos ticket is also verified.

Note: The *<PROTEGILITY_DIR>/cluster_utils* directory continues to exist on the Lead node.

This directory is retained to perform a cleanup in the event of the uninstallation failing on some nodes, due to unavoidable reasons, such as *host being down*.

6. After Big Data Protector is successfully uninstalled from all nodes, manually delete the <PROTEGILITY_DIR> directory from the Lead node.
7. If the <PROTEGILITY_DIR>/defiance_dps_old directory is present on any of the nodes in the cluster, then it can be manually deleted from the respective nodes.
8. Restart all Hadoop services.

12.6.5 Installing the Big Data Protector on an Amazon EMR Cluster

Amazon Web Services (AWS) is a cloud-based computing service, which provides several services, such as computing power through Amazon Elastic Compute Cloud (EC2), storage through Amazon Simple Storage Service (S3), Amazon Elastic MapReduce (EMR) and so on. AWS runs on servers, which are located in several data centers, which are organized by geographical regions, across the world.

You can install the Big Data Protector on an Amazon EMR cluster, using the AWS cloud-based service, to efficiently run analytics and scale in a cost-effective manner.

Note:

- Amazon EMR supports the EMR File System (EMRFS), HDFS, and S3 for storing data.
- This section considers S3 as reference for the input and output data.
- If you want to use HDFS for input and output data, then contact Protegility Support.

This section describes the tasks that you must perform to install the Big Data Protector on an Amazon EMR cluster.

12.6.5.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before installing the Big Data Protector on an Amazon EMR cluster:

- It is recommended to be familiar with the following parts:
 - The Amazon EMR environment
 - Storage bucket, used to store the Big Data Protector installation files
 - Bootstrap Action, used to invoke the installation of Big Data Protector
 - Amazon Virtual Private Cloud (VPC)
- An ESA 9.1.0.0 is installed and running on an EC2 instance.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-46: List of Ports for the Big Data Protector

| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|----------------------|-----------|--|-----------------------------------|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |



| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|--|-----------|--|---|--|
| 15780 | | Protector on the Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses localhost port 16700. |
| Port Requirement for Proxy Node (Optional) | | | | |
| 8443 | TCP | PEP server on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The PEP server communicates with the Proxy node through port 8443, which substitutes the stream to ESA. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The Log Forwarder communicates with the Proxy node through port 9200, which substitutes the stream to Protegility Audit Store Appliance. |

- If you are installing the Big Data Protector on an existing EMR cluster, then ensure the following pre-requisites are met:
 - The EMR cluster is setup and running.
 - The Private Key file is available on the Master node in the EMR cluster to communicate with the other nodes in the cluster.

12.6.5.2 Creating an S3 Bucket on AWS

If you are installing Big Data Protector on a new EMR cluster, then you need to create an S3 bucket to copy the Big Data Protector installation files, which are created using the Configurator script.

The S3 bucket provides an isolated space for storing the files that are required to install Big Data Protector on the new EMR cluster. These files are then utilized by the Amazon EMR environment when setting up the new EMR cluster. Even if the EMR cluster is terminated, the S3 bucket continues to retain the files required to install the Big Data Protector.

Note: For more information about creating an S3 bucket, refer to the Amazon documentation for [Creating an S3 bucket](#).

12.6.5.3 Downloading the Big Data Protector Package

After receiving the Big Data Protector installation package from Protegility, copy it to any Amazon EC2 instance or any node that has connectivity to the ESA.

12.6.5.4 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Configurator script and install the Big Data Protector on all the nodes on an Amazon EMR cluster.

- To extract the Configurator script from the installation package:

1. Login to the CLI on a machine or an Amazon EC2 node that has connectivity to the ESA.
2. Copy the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>-64_9.1.0.0.x.tgz* to any directory.

For example, */opt/bigdata*.

3. To extract the files from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the following files:

- *BDP_Configurator_EMR-<EMR_version>_9.1.0.0.x.sh*
- *ProxySetup_Linux_x64_<core_version>.sh*

12.6.5.5 Setting up the Proxy

You must configure a proxy if the nodes in the cluster are prohibited from connecting to the external network. The proxy will connect to the ESA and communicate with the nodes in the EMR cluster containing the Big Data Protector.

► To setup the Proxy for the EMR Cluster:

1. To install the proxy before you install the Big Data Protector, perform the following steps:

Note: Use this approach to install the Big Data Protector using a bootstrap action.

- a. Copy the *ProxySetup_Linux_x64_<core_version>.sh* script to the node that will act as the proxy.

Note: The node where you configure the proxy will communicate with the ESA and all other nodes in the EMR cluster.

- b. To install the proxy, run the following command.

```
./ProxySetup_Linux_x64_<core_version>.sh -esa host (-elastic elastic_host) (-dir installation_directory) (-port proxy_port) (-port_es proxy_port_es)
```

Table 12-47: Parameters in the proxy setup command

| Parameter | Description |
|--|--|
| <i>-esa <ESA_Host></i> | Specifies the ESA hostname or IP address that the proxy service will use for communication. |
| <i>-dir <Installation_Directory></i> | Specifies the location to install the proxy service. This is an optional parameter. If you fail to specify this parameter, then the script will install the proxy service in the <i><PROTEGITY_DIR>/proxy</i> directory. |
| <i>-port <Proxy_Port></i> | Specifies the ESA port that the proxy service will use for communication. This is an optional parameter. If you fail to specify this parameter, then the script will use <i>8443</i> as the default port. |
| <i>-elastic <elastic_host></i> | Specifies the hostname or IP address for the Protegity Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use the ESA hostname as the default value. |

| Parameter | Description |
|---|---|
| <code>-port_es <proxy_port_es></code> | Specifies the port number for the Protegility Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use <code>9200</code> as the default value. |

2. To install the Proxy service on any node in an Amazon EMR cluster after installing the Big Data Protector, perform the following steps:

Note: Use this approach to install the Big Data Protector using the static installer.

- Copy the `ProxySetup_Linux_x64_<core_version>.sh` script to the node in the cluster that will act as the proxy.

Note: The node where you configure the proxy will communicate with the ESA and all the other nodes in the EMR cluster.

- To install the proxy, run the following command.

```
./ProxySetup_Linux_x64_<core_version>.sh -esa host (-elastic elastic_host) (-dir installation_directory) (-port proxy_port) (-port_es proxy_port_es)
```

Table 12-48: Parameters in the proxy setup command

| Parameter | Description |
|--|--|
| <code>-esa <ESA_Host></code> | Specifies the ESA hostname or IP address that the proxy service will use for communication. |
| <code>-dir <Installation_Directory></code> | Specifies the location to install the proxy service. This is an optional parameter. If you fail to specify this parameter, then the script will install the proxy service in the <code><PROTEGILITY_DIR>/proxy</code> directory. |
| <code>-port <Proxy_Port></code> | Specifies the ESA port that the proxy service will use for communication. This is an optional parameter. If you fail to specify this parameter, then the script will use <code>8443</code> as the default port. |
| <code>-elastic <elastic_host></code> | Specifies the hostname or IP address for the Protegility Audit Store appliance(ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use the ESA hostname as the default value. |
| <code>-port_es <proxy_port_es></code> | Specifies the port number for the Protegility Audit Store appliance(ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use <code>9200</code> as the default value. |

- Open the `pepservice.cfg` file in insert mode.
- In the `pepservice.cfg` file, replace the IP address of the ESA with the IP address of the proxy node.

Note:

- Replace the IP address of the ESA with the IP address of the proxy node in the `pepservice.cfg` file on all the nodes.
- If you specify the port number while installing the proxy, then you must update the same port in the `pepservice.cfg` file for all the nodes.

- Save the changes to the `pepservice.cfg` file.
 - Restart the PEP server on all the nodes.
3. Navigate to the proxy directory.
For example, `<Installation_Directory>/proxy/bin/`
- To start the proxy service, run the following command.

```
./proxyctrl start
```



5. Press ENTER.
6. To verify whether all the proxy services are up and running after setting up the proxy, run the following command.

```
./proxyctrl status
```
7. Press ENTER.
The status of the Proxy service appears.

12.6.5.6 Running the Configurator Script

You must run the configurator script to create the installation files for installing the Big Data Protector on an Amazon EMR cluster. You can install the Big Data Protector on an Amazon EMR cluster in any one of the following methods:

- New EMR cluster: The configurator script will:
 - Download the certificates and key encryption files from the ESA
 - Create the Big Data Protector installation files for a new EMR cluster
 - Create the bootstrap installer and classpath configurator script for a new EMR cluster
 - Copy the Big Data Protector installation files, bootstrap installer, and the classpath configurator script to the S3 bucket
- Existing EMR cluster: The configurator script will generate the installation package to install the Big Data Protector on an existing EMR cluster.

► To run the Big Data Protector Configurator Script:

1. Login to the staging environment.
2. Navigate to the directory that contains the *BDP_Configurator_EMR-<EMR_version>_9.1.0.0.x.sh* script.
3. To execute the *BDP_Configurator_EMR-<EMR_version>_9.1.0.0.x.sh* script, run the following command.

```
./BDP_Configurator_EMR-<EMR_version>_9.1.0.0.x.sh
```

4. Press ENTER.
The prompt to continue the installation of the Big Data Protector appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS EMR.
Do you want to continue? [yes or no]
```

5. To continue the installation of the Big Data Protector, type *yes*.
6. Press ENTER.
The prompt to create the Big Data Protector installation package, depending on the EMR cluster, appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS EMR.
Do you want to continue? [yes or no]
yes

Protegility Big Data Protector Configurator started...

Enter the EMR cluster for which the Big Data Protector installation package needs to be
created:
[ 1 ] : New EMR Cluster
[ 2 ] : Existing EMR cluster
[ 1 or 2 ]:
```

7. Depending on your requirement, select any one of the following options:
 - To create the Big Data Protector installation package for a new EMR cluster, type *1*.
 - To generate the Big Data Protector installation package, in a local directory, for an existing EMR cluster, type *2*.

Note: For more information about installing the Big Data Protector on an existing EMR cluster, refer to section [Installing the Big Data Protector on an Existing EMR Cluster](#).

8. To create the Big Data Protector installation package for a new EMR cluster, type *1*.

9. Press ENTER.

The prompt to enter the S3 URI to upload the Big Data Protector installation files appears.

```
Generating Big Data Protector for a new EMR cluster.....  
Enter the S3 URI where the BDP Installation files are to be uploaded.  
(E.g. s3://examplebucket/folder) :
```

10. Type the path of the S3 storage bucket.

Ensure that the path of the S3 storage bucket is in the following format:

```
s3://<bucket_name>/<folder_in_the_bucket>
```

where,

- <bucket_name> - specifies the name of the storage bucket.
- <folder_in_the_bucket> - specifies the directory within the bucket.

11. Press ENTER.

The prompt to either upload the installation files to the S3 bucket or generate them locally appears.

```
Choose one option among the following for BDP Installation files:  
[1] -> Upload files to 's3://<bucket_name>/<folder_in_the_bucket>' S3 URI.  
[2] -> Generate files locally to current working directory. (You would have to manually  
upload the files to the specified S3 URI)  
[ 1 or 2 ]:
```

12. To upload the installation files to the S3 storage bucket, type *1*.

13. Press ENTER.

The prompt to select the type of AWS access key appears.

```
Choose the Type of AWS Access Keys from the following options:  
[1] -> IAM User Access Keys (Permanent access key id & secret access key)  
[2] -> Temporary Security Credentials (Temporary access key id, secret access key &  
session token)  
[ 1 or 2 ]:
```

14. Depending on the type of AWS Access Keys you want to use, type *1* or *2*.

For example, to use the temporary security credentials, type *2*.

15. Press ENTER.

The prompt to enter the access key ID appears.

```
Enter the Access Key ID:
```

16. Enter the access key ID.

17. Press ENTER.

The prompt to enter the secret access key appears.

```
Enter the Secret Access Key:
```

18. Enter the secret access key.

19. Press ENTER.

The prompt to enter the security session token appears.

Note: The configurator script will prompt for the security session token only when you select *Temporary Security Credentials* as the type of AWS access keys.

Enter the Security Session Token:

20. Enter the security session token.

21. Press ENTER.

The prompt to enter the ESA hostname or IP address appears.

Enter the ESA Host or IP Address[]:

22. Enter the hostname or the IP address of the ESA.

23. Press ENTER.

The prompt to enter the listening port for the ESA appears.

Enter ESA host listening port [8443]:

24. Enter the listening port for the ESA.

Alternatively, to use the default listening port, press ENTER.

25. Press ENTER.

The prompt to enter the user name for the ESA appears.

Enter ESA Username:

26. Enter the user name.

27. Press ENTER.

The prompt to select the audit store type appears.

Select the Audit Store type where Log Forwarder(s) should send logs to.

[1] : Protegrity Audit Store
 [2] : External Audit Store
 [3] : Protegrity Audit Store + External Audit Store

Enter the no.:

28. To select the Audit Store type, select any one of the following options:

Table 12-49: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting using the Protegrity Audit Store appliance, type 1. If you enter 1, then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegrity Audit Store appliances. |
| 2 | To use an external audit store, type 2. If you enter 2, then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type 3. If you enter 3, then the default Fluent Bit configuration files used for the Protegrity Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/</i> |

| Option | Description |
|--------|--|
| | <code>data/config.d/</code> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <code>/opt/protegility/fluent-bit/data/config.d/</code> directory. |

29. Press ENTER.

The prompt to enter the comma separated list of hostname or IP addresses appears.

```
Enter comma-separated list of Hostnames/IP Addresses and/or Ports of Protegility Audit Store.
Allowed Syntax: hostname[:port][,hostname[:port],hostname[:port]...] (Default Value - <ESA_hostname/IP>:9200)
Enter the list:
```

30. Enter the comma-separated IP addresses/ports in the correct syntax.

31. Press ENTER.

The prompt to enter the local directory path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this node that stores the custom Fluent-Bit configuration files for External Audit Store:
```

Note: The configurator script will display this prompt only if you select option 2 or 3 in step 27. When you select option 2 or 3 in step 27, the custom configuration files are copied to the `/<Installation_directory>/fluent-bit/data/config.d/` directory during the execution of bootstrap script on the EMR nodes.

32. Enter the local directory path that stores the custom Fluent Bit configuration files.

33. Press ENTER.

The configurator script downloads the certificates from the ESA, generates the installation files, and uploads them to the S3 storage bucket.

```
*****
Welcome to the Pep Server Setup Wizard.
*****

Unpacking.....
Extracting files...
Unpacked pepserver compressed file...
Temporarily setting up PepServer defiance_dps directory structure on current node...
Please enter the password for downloading certificates[]:

Unpacking...
Extracting files...
Downloading certificates from x.x.x.x:8443...
% Total    % Received % Xferd  Average Speed   Time     Time      Current
          Dload  Upload Total   Spent    Left Speed
100 20480  100 20480    0      0  109k      0 --::-- --::-- --::-- 109k

Extracting certificates...
Certificates successfully downloaded and stored in /<installation_dir>/defiance_dps/data

Protegility PepServer installed in /<installation_dir>/defiance_dps.
```

Retrieving the S3 bucket's AWS Region via AWS S3 REST API...
Successfully retrieved S3 bucket's AWS region: <AWS_region_name>

Started Uploading the generated installation files via AWS S3 REST API.....

Uploading bdp_bootstrap_installer.sh to the S3 bucket.
File uploaded to s3://<bucket_name>/<folder_in_the_bucket>/bdp_bootstrap_installer.sh

Uploading bdp_classpath_configurator.py to the S3 bucket.

```

File uploaded to s3://<bucket_name>/<folder_in_the_bucket>/bdp_classpath_configurator.py

Uploading BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>_9.1.0.0.x.tgz to the S3
bucket.
File uploaded to s3://<bucket_name>/<folder_in_the_bucket>/BigDataProtector_Linux-
ALL-64_x86-64_EMR-<EMR_version>_9.1.0.0.x.tgz

Successfully Uploaded BigDataProtector_Linux-ALL-64_x86-64_EMR-
<EMR_version>_9.1.0.0.x.tgz, bdp_bootstrap_installer.sh, bdp_classpath_configurator.py to
S3 bucket 's3://<bucket_name>/<folder_in_the_bucket>'

Successfully Generated installation files at ./Installation_Files/ directory.

Successfully configured Big Data Protector for a new EMR cluster..

```

If you select the option to generate the installation files in a local directory in step 7, then the configurator script will generate the installation files in a local directory.

```

*****
***** Welcome to the Pep Server Setup Wizard.
***** *****

Unpacking.....
Extracting files...
Unpacked pepserver compressed file...
Temporarily setting up PepServer defiance_dps directory structure on current node...
Please enter the password for downloading certificates
[]:

Unpacking...
Extracting files...
Downloading certificates from X.X.X.X:8443...
  % Total    % Received % Xferd  Average Speed   Time     Time      Current
               Dload  Upload Total   Spent    Left Speed
 100 20480  100 20480    0      0  116k      0 --:--:-- --:--:-- --:--:-- 116k

Extracting certificates...
Certificates successfully downloaded and stored in /<installation_dir>/defiance_dps/data

Protegility PepServer installed in /<installation_dir>/defiance_dps.

Successfully Generated installation files at ./Installation_Files/ directory.

Please upload these files to S3 bucket 's3:<bucket_name>/<folder_in_the_bucket>' manually.

Successfully configured Big Data Protector for a new EMR cluster..

```

34. After all the Big Data Protector files are copied to the S3 storage bucket, on the Amazon S3 page, refresh the contents of the S3 storage bucket.
The Big Data Protector installation files appear in the S3 storage bucket.

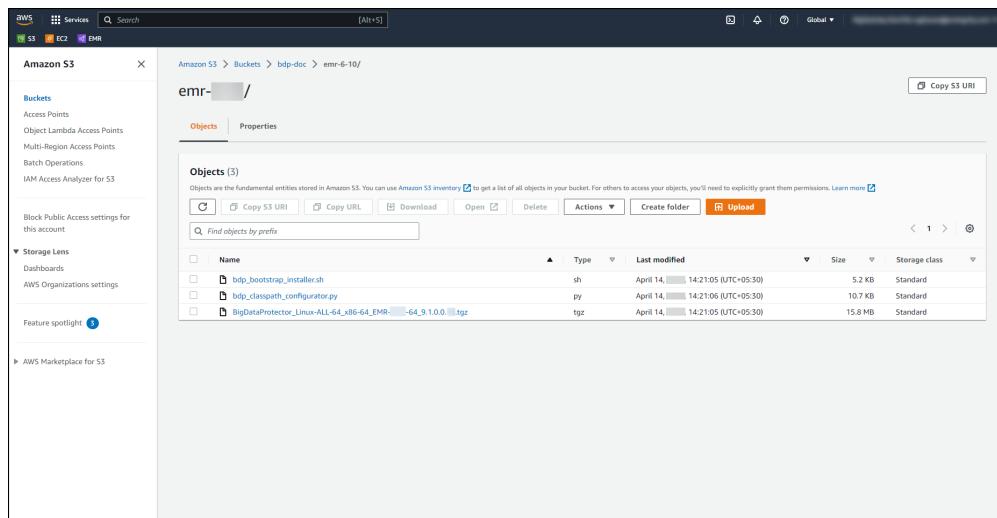


Figure 12-250: Big Data Protector Installation Files in the S3 Storage Bucket

The following files must be available in the S3 storage bucket:

- *BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>-64_9.1.0.0.x.tgz* - is the Big Data Protector installation package for a new EMR cluster.
- *bdp_bootstrap_installer.sh* - is the bootstrap installer script for installing the Big Data Protector on a new Amazon EMR cluster.
- *bdp_classpath_configurator.py* - is the script that contains the required classpath settings for running the Big Data Protector on a new Amazon EMR cluster.

12.6.5.7 Installing Big Data Protector on a New EMR Cluster

This section describes the tasks that need to be performed for installing Big Data Protector on a new EMR cluster.

12.6.5.7.1 Creating a New Cluster on an Amazon EMR Environment

You can create a new EMR cluster on AWS and install the Big Data Protector on all the nodes.

► To create a new EMR cluster and install the Big Data Protector:

1. On the AWS management console, expand **Services** and click **Analytics**. The sub-menu appears.
2. From the sub-menu, click **EMR**. The **Amazon EMR** page appears.

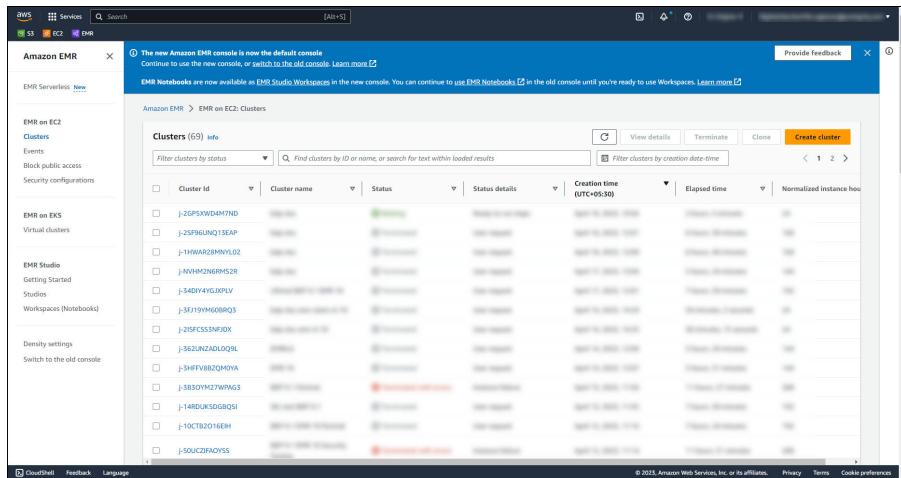


Figure 12-251: Amazon EMR Page

3. Click **Create Cluster**.

The **Create Cluster** page appears.

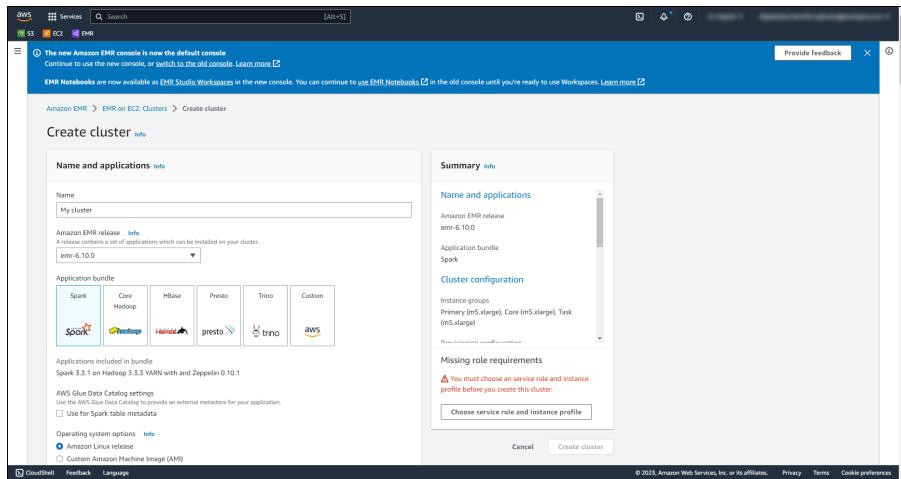


Figure 12-252: Create Cluster Page

4. Under **Name and applications**, enter the details as mentioned in the following steps.

- In the **Name** box, enter a name to identify the cluster.
- From the **Amazon EMR release** list, select the required version of EMR.
- Under **Application bundle**, select the required applications that you want to install on the EMR cluster.
- To customize the applications that you want to install on the EMR cluster, under **Application bundle**, select **Custom**. The available list of applications appears.

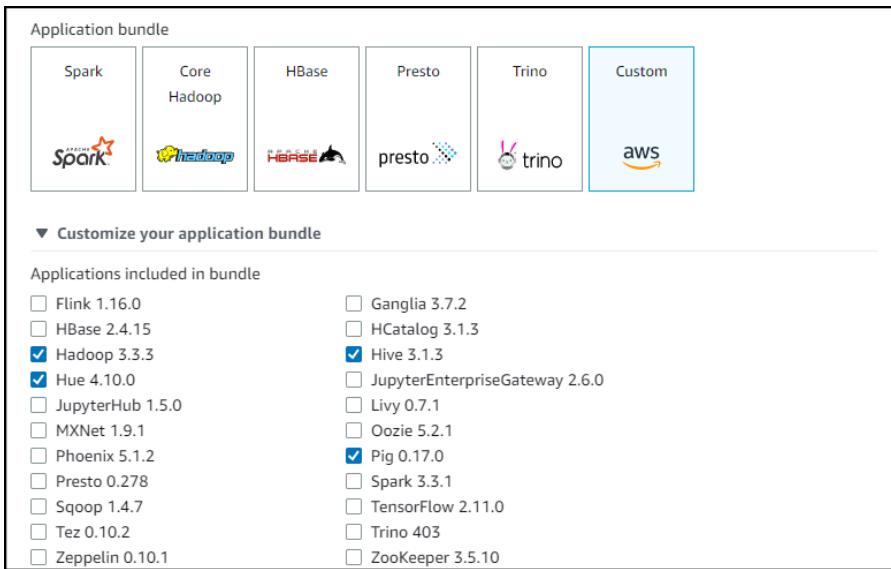


Figure 12-253: List of Applications

Note: The list of components, available for installation, will vary depending on the version of EMR that you select. In addition, the options available for configuration will also vary depending on the components that you select.

- e. Under **Applications included in bundle**, select the check box against the component that you want to install.
5. Under **Cluster configuration**, select the details as mentioned in the following steps.
 - a. Under **Instance groups**, select the required EC2 instance type for each of the nodes, such as, the **Primary**, **Core**, and the **Task** node.
 - b. If you want to provide the configuration for a node, then expand **Node configuration - optional** to enter the required configuration.
 - c. If you want to provide a custom size for the root volume, then in the **EBS root volume size** box, enter the required size in Gigabytes (GB).
6. Under **Networking**, select the options as mentioned in the following steps.
 - a. To select the virtual private cloud (VPC) for the cluster, click **Browse**.
The **Choose VPC** page appears.
 - b. Select the required VPC.
 - c. Click **Choose**.
The VPC that you selected appears under **Virtual Private Cloud (VPC)**.
 - d. To select the subnet for the virtual private cloud (VPC) for the cluster, click **Browse**.
The **Choose subnet** page appears.
 - e. Select the required subnet.
 - f. Click **Choose**.
The subnet that you selected appears under **Subnet**.
7. Under **EC2 security groups (firewall)**, select the required security group for the Primary node.
8. Under **EC2 security groups (firewall)**, select the required security group for the Core and the Task node.
9. Under **EC2 security groups (firewall)**, from the **Service access (Private subnet)** list, select the required private subnet.
10. To prevent automatic termination of the cluster, under **Cluster termination**, select the **Manually terminate cluster** option.
11. To configure bootstrap actions that will run on every node in the cluster during installation, perform the following steps.
 - a. Under **Bootstrap Actions**, click **Add**.

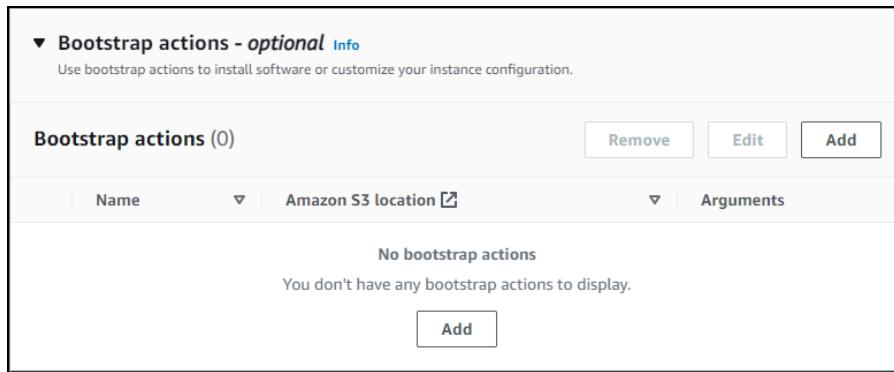


Figure 12-254: Bootstrap Actions

The **Add bootstrap action** page appears.

The screenshot shows the 'Add bootstrap action' dialog box. It has fields for 'Name' (with placeholder 'Enter name'), 'Script location' (with placeholder 's3://bucket/prefix/object'), and 'Arguments - optional' (with placeholder 'Specify the arguments your scripts take'). At the bottom are 'Cancel' and 'Add bootstrap action' buttons, with the latter being orange.

Figure 12-255: Add bootstrap action Page

- In the **Name** box, enter a name to identify the bootstrap action.
- In the **Script location** box, enter the URI of the bootstrap installer script.

Note: Browse and select the *bdp_bootstrap_installer.sh* script.

- Click **Add bootstrap action**.
- The bootstrap script appears under the Bootstrap actions section.

The screenshot shows the 'Bootstrap actions - optional' section again. It lists one entry: 'btstrp' with the 'Amazon S3 location' as 's3://bdp-doc/emr-6-10/bdp_bootstrap_installer.sh'. The same set of buttons (Remove, Edit, Add) is present.

Figure 12-256: Bootstrap actions

- Under **Security configuration and EC2 key pair - optional**, select the required **Security configuration** and the **Amazon EC2 key pair for SSH to the cluster**.
- Under **Identity and Access Management (IAM) roles**, perform the following steps.
 - To choose an IAM role, from the **Service role** list, select the required role.
 - To choose an **EC2 instance profile for Amazon EMR**, from the **Instance profile** list, select the required profile.

14. Review the security settings for the cluster.

15. Click **Create Cluster**.

AWS provisions the cluster and the cluster appears on the EMR page.

You can also use the CLI to create a new EMR cluster and install the Big Data Protector on the nodes in the cluster. For example:

```
aws emr create-cluster --termination-protected --applications Name=Hadoop Name=Hive
Name=Pig Name=Hue --ec2-attributes
'{"KeyName": "<KEY_NAME>", "InstanceProfile": "<Instance_Profile_Name>", "ServiceAccessSecurityGroup": "<Service_Access_Security_Group_Name>", "SubnetId": "<Subnet_ID>", "EmrManagedSlaveSecurityGroup": "<EMR_Managed_Slave_Security_Group_Name>", "EmrManagedMasterSecurityGroup": "<EMR_Managed_Master_Security_Group_Name>"}' --release-label <EMR_version> --log-uri
'<s3_bucket_URI>' --instance-groups '[{"InstanceCount":1,"EbsConfiguration":
{"EbsBlockDeviceConfigs": [{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp3"}, "VolumesPerInstance":2}}}, {"InstanceGroupType": "MASTER",
"InstanceType": "m4.xlarge", "Name": "Master - 1"}, {"InstanceCount":2, "EbsConfiguration":
{"EbsBlockDeviceConfigs": [{"VolumeSpecification":
{"SizeInGB":32,"VolumeType":"gp3"}, "VolumesPerInstance":2}}}, {"InstanceGroupType": "CORE", "InstanceType": "m4.xlarge", "Name": "Core - 2"}]' --custom-ami-id <AMI_instance_ID> --
bootstrap-actions '[{"Path": "<S3_URI_for_installer.sh>", "Name": "<btstrp_action_name>"}]' --
auto-scaling-role <EMR_AutoScaling_Role> --ebs-root-volume-size 15 --service-role
<Service_Role_Name> --repo-upgrade-on-boot SECURITY --name '<Name_of_the_cluster>' --
scale-down-behavior TERMINATE_AT_TASK_COMPLETION --<name_of_the_region>
```

where,

- *<S3_Path_For_BootstrapInstaller>* - specifies the S3 bucket path containing the Big Data Protector bootstrap installer script.
- *<Script_Name>* - specifies the script to install the Big Data Protector.
- *<KEY_NAME>* - specifies the private key file, on the Master node in the EMR cluster, which is used to communicate with the other nodes in the cluster.
- *<Cluster_Name>* - specifies the name of the new EMR cluster.

Attention: When Big Data Protector is installed via the EMR bootstrap actions, the Protegility PEP server and the Log Forwarder will be integrated with *systemd*. This integration will set a configuration for the PEP server and the Log Forwarder to restart them automatically when the cluster nodes restart.

12.6.5.7.2 Managing the Nodes on an Amazon EMR Cluster

Depending on the workload on the EMR cluster, you can add or remove the Big Data Protector nodes. You can either set the cluster to automatically scale or manually add or remove nodes in the EMR cluster. You can add or remove nodes in the EMR cluster either while you create the cluster or after you have created the cluster. Before you add or remove the nodes from the cluster, ensure that you save all your data to S3, as standard practice, to avoid any data loss.

Note: This section covers the procedure to add or remove nodes from an Amazon EMR cluster after you have created it.

► To add or remove nodes from an Amazon EMR cluster:

1. On the AWS management console, expand **Services** and click **Analytics**.
The sub-menu appears.
2. From the sub-menu, click **EMR**.
The **Amazon EMR** page appears.

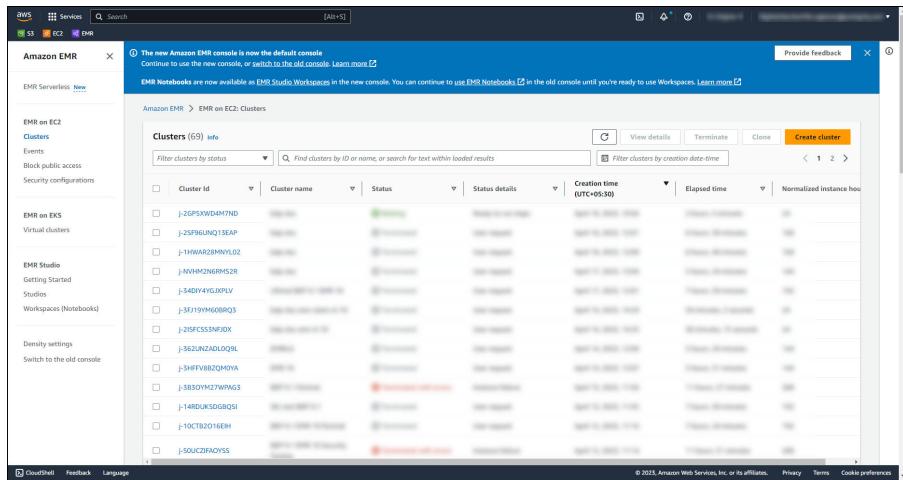


Figure 12-257: Amazon EMR Page

3. Click the required cluster.

The **Properties** tab of the cluster appears.

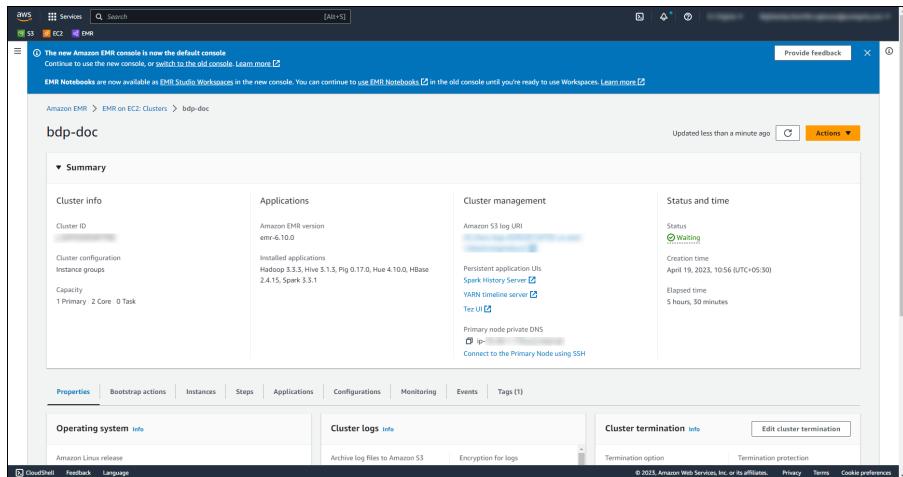


Figure 12-258: Properties Tab of the Cluster

4. Click the **Instances** tab.

The **Instances** tab appears.

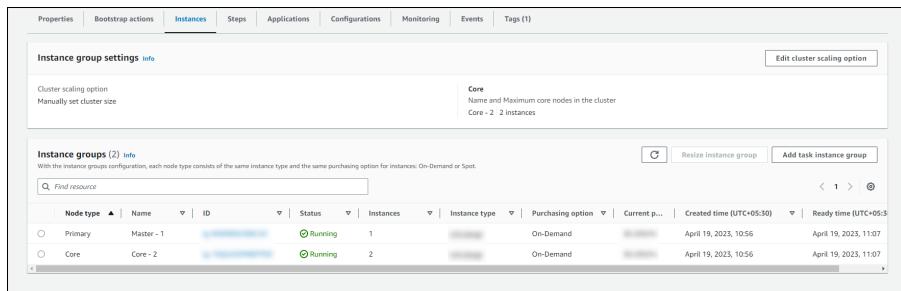


Figure 12-259: Instances Tab

5. To add an instance, perform the following steps.

- a. Under **Instance groups**, click **Add task instance group**.

The **Add task instance group** page appears.

Amazon EMR > EMR on EC2: Clusters > bdp-doc > Add task instance group for cluster j-2GP5XWD4M7ND

Add task instance group for cluster j-2GP5XWD4M7ND

Task instance group [Info](#)

Add task instance group to increase Amazon EC2 instances in response to increasing workload.

Name

Choose EC2 instance type

m5.xlarge
4 vCore 16 GiB memory EBS only storage
On-Demand price: -
Lowest Spot price: \$0.094 (us-east-1a)

[Actions ▾](#)

► Node configuration - optional

Instance group size instance(s)

Use Spot purchasing option

[Cancel](#) [Add task instance group](#)

Figure 12-260: Add task instance group Page

- b. In the **Name** box, enter the name to identify the node.
 - c. From the **Choose EC2 instance type** list, select the required storage type.
 - d. In the **Instance group size** box, enter the required number of instances.
 - e. Click **Add task instance group**.
The new instance is added to the node and appears on the **Instances** tab.
6. To resize an instance, perform the following steps.
- a. Under **Instance groups**, select the required instance that you want to resize.
 - b. Click **Resize instance group**.
The **Resize** page appears.

Resize Core - 2 for Cluster j-2GP5XWD4M7ND [X](#)

Instance group size Instance(s)

[ⓘ Current size: 2 instance\(s\)](#)

Learn more about [Amazon EC2 On-Demand pricing](#)

[Cancel](#) [Resize](#)

Figure 12-261: Resize Page

- c. In the **Instance group size** box, enter the required number of instances.
 - d. Click **Resize**.
- The instance is resized as per the inputs and appears on the **Instances** tab.

12.6.5.7.3 Verifying the Big Data Protector Parameters in EMR

Before you start using the Big Data Protector, you must configure the required Protegity-related parameters in EMR. The Big Data Protector configuration parameters are set for the EMR cluster when you install it on all the nodes in the cluster.

The following table provides the parameters that are set for the new Amazon EMR cluster before using Big Data Protector.

Table 12-50: Big Data Protector Classpath Configuration for a New EMR Cluster

| Name | Configuration File | Updated Classpath Parameter |
|-----------|--|---|
| MapReduce | /var/aws/emr/bigtop-deploy/puppet/modules/hadoop/templates/mapred-site.xml | <pre>mapreduce.application.classpath : /opt/protegity/ pepmapper/lib/*,/opt/protegity/pephive/lib/*,/opt/ protegity/bdp_version/ mapreduce.admin.user.env : LD_LIBRARY_PATH=/opt/ protegity/jpeplite/lib</pre> |
| Hive | <pre>/var/aws/emr/bigtop-deploy/puppet/modules/ hadoop_hive/templates/hive-site.xml</pre> <pre>/var/aws/emr/bigtop-deploy/puppet/modules/tez/ templates/tez-site.xml</pre> <pre>/var/aws/emr/bigtop-deploy/puppet/modules/ hadoop_hive/templates/hive-env.sh</pre> | <pre>hive.exec.pre.hooks : com.protegity.hive.PtyHiveUserPreHook</pre> <pre>tez.cluster.additional.classpath.prefix : /opt/protegity/ pephive/lib/*:/opt/protegity/bdp_version/</pre> <pre>tez.am.launch.env :</pre> <pre>LD_LIBRARY_PATH=/opt/protegity/jpeplite/lib/</pre> <pre>export HIVE_CLASSPATH=\${HIVE_CLASSPATH}:./opt/ protegity/pephive/lib/*:/opt/protegity/bdp_version/</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegity/jpeplite/lib/</pre> |
| Pig | /var/aws/emr/bigtop-deploy/puppet/modules/hadoop_pig/templates/pig-env.sh | <pre>PIG_CLASSPATH="/opt/protegity/peppig/lib/*:/opt/ protegity/bdp_version/"</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegity/jpeplite/lib/</pre> |
| HBase | <pre>/var/aws/emr/bigtop-deploy/puppet/modules/ hadoop_hbase/templates/hbase-site.xml</pre> <pre>/var/aws/emr/bigtop-deploy/puppet/modules/ hadoop_hbase/templates/hbase-env.sh</pre> | <pre>hbase.coprocessor.region.classes : com.protegity.hbase.PTYRegionObserver</pre> <pre>export HBASE_CLASSPATH=\$ {HBASE_CLASSPATH}:./opt/protegity/pephbase/lib/*:/opt/ protegity/bdp_version/</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegity/jpeplite/lib/</pre> |
| Spark | /var/aws/emr/bigtop-deploy/puppet/modules/spark/templates/spark-defaults.conf | <pre>spark.driver.extraClassPath=/opt/protegity/pephive/lib/*:/opt/ protegity/pepspark/lib/*:/opt/protegity/bdp_version/</pre> <pre>spark.executor.extraClassPath=/opt/protegity/ pephive/lib/*:/opt/protegity/pepspark/lib/*:/opt/protegity/ bdp_version/</pre> <pre>spark.executor.extraLibraryPath= /opt/protegity/jpeplite/lib</pre> |



| Name | Configuration File | Updated Classpath Parameter |
|------|--------------------|--|
| | | <p><code>spark.driver.extraLibraryPath= /opt/protegility/jpeplite/lib</code> For Spark Version >= 2.4.0 and < 3.0.0</p> <p><code>spark.executor.plugins com.protegility.spark.PtyExecSparkPlugin</code></p> <p>For Spark Version >= 3.0.0</p> <p><code>spark.plugins com.protegility.spark.PtyExecSparkPlugin</code></p> |

12.6.5.7.4 Uninstalling the Big Data Protector

The uninstallation of the Big Data Protector is not supported when a new EMR cluster is created. However, when the workload on the nodes in the EMR cluster is reduced and the nodes are decommissioned automatically by AWS or manually removed, then the decommissioned or removed nodes are not available for use.

Note: Before you delete nodes from the cluster, ensure that you save your data to S3 as standard practice to avoid any data loss.

12.6.5.8 Installing the Big Data Protector on an Existing EMR Cluster

This section describes the tasks that need to be performed for installing Big Data Protector on an existing EMR cluster.

The following are the overall steps for installing the Big Data Protector on an existing EMR cluster.

1. Verify the prerequisites
2. Extract the files from the installation package
3. Update the *BDP.config* file
4. Install the Big Data Protector on all the nodes in the cluster
5. Install the Big Data Protector on the New Nodes Added to the Cluster
6. Verify the parameters for the Big Data Protector

12.6.5.8.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before installing the Big Data Protector:

- The EMR cluster is installed, configured, and running.
- An ESA 9.1.0.0 instance is installed, configured, and running.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-51: List of Ports for the Big Data Protector

| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|----------------------|-----------|--|-----------------------------------|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a Policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |



| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|--|-----------|--|---|--|
| 15780 | | Protector on the Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port <i>15780</i> . The PEP server Application Logs are also written to localhost through port <i>15780</i> . The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses localhost port <i>16700</i> . |
| Port Requirement for Proxy Node (Optional) | | | | |
| 8443 | TCP | PEP server on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The PEP server communicates with the Proxy node through port <i>8443</i> , which substitutes the stream to the ESA. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The Log Forwarder communicates with the Proxy node through port <i>9200</i> , which substitutes the stream to the Protegility Audit Store Appliance. |

- The static installer for EMR uses utilities, such as, *pssh* (parallel ssh) and *pscp* (parallel scp). These utilities require Python to be installed on the Primary node.

To verify whether Python is installed on the Primary node, run the following command:

```
/usr/bin/env python --version
```

The command returns the version of Python installed on the system.

If you are unable to detect Python on the Primary node, then ensure that you have a compatible version of Python installed on the lead node (preferably Python 3.x). Ensure that the utilities are able to detect the version of Python using the following command:

```
/usr/bin/env python
```

- A *sudoer* user account with privileges to perform the following tasks:
 - Update the system by modifying the configuration, permissions, or ownership of directories and files.
 - Perform third party configuration.
 - Create directories and files.
 - Modify the permissions and ownership for the created directories and files.
 - Set the required permissions to the create directories and files for the Protegility Service Account.
 - Permissions for using the SSH service.
- The following user accounts are present to perform the required tasks:
 - ADMINISTRATOR_USER**: It is the *sudoer* user account that is responsible to install and uninstall the Big Data Protector on the cluster.

This user account needs to have *sudo* access to install the product.

- EXECUTOR_USER**: It is a user that has ownership of all Protegility files, directories, and services.



- ***OPERATOR_USER***: It is responsible for performing tasks, such as, starting or stopping tasks, monitoring services, updating the configuration, and maintaining the cluster while the Big Data Protector is installed on it.
- If you need to start, stop, or restart the Protegity services, then you need *sudoer* privileges for this user to impersonate the ***EXECUTOR_USER***.

Note: Depending on the requirements, a single user on the system may perform multiple roles.

If a single user is performing multiple roles, then ensure that the following conditions are met:

- The user has the required permissions and privileges to impersonate the other user accounts, for performing their roles, and perform tasks as the impersonated user.
 - The user is assigned the highest set of privileges, from the required roles that it needs to perform, to execute the required tasks.
- For instance, if a single user is performing tasks as ***ADMINISTRATOR_USER***, ***EXECUTOR_USER***, and ***OPERATOR_USER***, then ensure that the user is assigned the privileges of the ***ADMINISTRATOR_USER***.

- A Private Key file (*.pem* file) for the *sudoer* user, which is used for enabling key-based authentication, and for communicating with all the nodes in the EMR cluster, is present on the Master node.
- As key-based authentication for the *sudoer* user is provided, which is required for installing and using Big Data Protector on the EMR cluster, ensure that the ***ADMINISTRATOR_USER*** or ***OPERATOR_USER*** have the value of the ***NOPASSWD*** parameter set to ***ALL*** in the *sudoer*'s file.
- The management scripts provided by the installer in the *cluster_utils* directory should be run only by the user (***OPERATOR_USER***) having privileges to impersonate the ***EXECUTOR_USER***.
 - If the value of the ***AUTOCREATE_PROTEGITY_IT_USR*** parameter in the *BDP.config* file is set to ***No***, then ensure that a service group containing a user for running the Protegity services on all the nodes in the cluster already exists.
 - If the Hadoop cluster is configured with AD or LDAP for user management, then ensure that the ***AUTOCREATE_PROTEGITY_IT_USR*** parameter in the *BDP.config* file is set to ***No*** and that the required service account user is created on all the nodes in the cluster.

12.6.5.8.2 Extracting the Files from the Installation Package

Before you begin

Note: Ensure that you copy the installation package *BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>-64_9.1.0.0.x.tgz*, which is generated by the configurator script, to the Master node on the EMR cluster in any temporary directory, such as, */opt/bigdata/*.

► To extract the files from the installation package:

1. Login to the Master node of the existing EMR cluster.
2. Navigate to the directory that contains the installation package.
3. To extract the files from the installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_EMR-<EMR_version>-64_9.1.0.0.x.tgz
```

The command extracts the following files:

- *uninstall.sh*
- *ptyLogAnalyzer.sh*
- *ptyLog_Consolidator.sh*

- *PepPigx.x.xSetup_Linux_emr-<EMR_version>.0_9.1.0.0.x.sh*
- *bdp_classpath_deconfigurator.py*
- *PepHivex.x.xSetup_Linux_emr-<EMR_version>.0_9.1.0.0.x.sh*
- *BdpInstallx.x.x_Linux_9.1.0.0.x.sh*
- *JpepLiteSetup_Linux_x64_9.1.0.0.x.sh*
- *Logforwarder_Setup_Linux_x64_<core_version>.sh*
- *PepSparkx.x.xSetup_Linux_emr-<EMR_version>.0_9.1.0.0.x.sh*
- *PepHbaseProtectorx.x.xSetup_Linux_emr-<EMR_version>.0_9.1.0.0.x.sh*
- *node_uninstall.sh*
- *bdp_classpath_configurator.py*
- *bdp_version/*
- *bdp_version/bdp_version.properties*
- *PepServer_Setup_Linux_x64_<core_version>.sh*
- *BDP.config*
- *PepMapreducex.x.xSetup_Linux_emr-<EMR_version>.0_9.1.0.0.x.sh*

12.6.5.8.3 Updating the *BDP.config* File

Before you begin

Note: Ensure that you update the *BDP.config* file before you install the Big Data Protector.

Warning: Do not update the *BDP.config* file when the installation of the Big Data Protector is in progress.

To update the *BDP.config* file:

1. Create a *hosts* file that contains the IP addresses of all the nodes in the cluster, except the Lead node.
2. Open the *BDP.config* file in any text editor.
3. In the *BDP.config* file, specify the full path of the *hosts* file.

Note: The installer uses the host file to install the Big Data Protector on all the nodes in the cluster.

4. In the *BDP.config* file, modify the values of the following parameters:

Table 12-52: Parameters in the bdp.config file

| Parameter | Description |
|-------------------------|---|
| <i>HADOOP_DIR</i> | Specifies the installation home directory for the Hadoop distribution. |
| <i>PROTEGERITY_DIR</i> | Specifies the directory where you will install the Big Data Protector. The examples used in this document consider <i>/opt/protegility/</i> as the default installation directory. |
| <i>CLUSTERLIST_FILE</i> | Specifies the file that contains the host name or IP addresses all the nodes in the cluster. This file will not contain an entry for the Lead node and will list one host name and IP address per line. Ensure that you specify the file name with the complete path. |
| <i>SPARK_PROTECTOR</i> | Specify any one of the following values: |

| Parameter | Description |
|------------------------------------|---|
| | <ul style="list-style-type: none"> • <i>Yes</i> – instructs the installer to install the Spark protector. Set the value of this parameter to <i>Yes</i>, if you want to run Hive UDFs with Spark SQL. • <i>No</i> – instructs the installer to skip the installation of the Spark protector. |
| <i>AUTOCREATE_PROTEGITY_IT_USR</i> | <p>Specifies the Protegity service account. The service group and service user name specified in the <i>PROTEGITY_IT_USR_GROUP</i> and <i>PROTEGITY_IT_USR</i> parameters respectively will be created if you set the value of this parameter to <i>Yes</i>. Specify any one of the following values:</p> <ul style="list-style-type: none"> • <i>Yes</i> – instructs the installer to create a service group <i>PROTEGITY_IT_USR_GROUP</i> containing the user <i>PROTEGITY_IT_USR</i> for running the Protegity services on all the nodes in the cluster. <p>If the service group or service user are already present, then the installer exits.</p> <p>If you uninstall the Big Data Protector, then the service group and the service user are deleted.</p> <ul style="list-style-type: none"> • <i>No</i> – instructs the installer to skip creating the <i>PROTEGITY_IT_USR_GROUP</i> service group with the <i>PROTEGITY_IT_USR</i> service user for running the Protegity services on all the nodes in the cluster. <p>Ensure that a service group containing a service user for running Protegity services are created, as described in section Verifying Prerequisites for Installing Big Data Protector.</p> |
| <i>PROTEGITY_IT_USR_GROUP</i> | <p>Indicates the service group that is required for running the Protegity services on all the nodes in the cluster. This service group owns all the Protegity installation directories.</p> |
| <i>PROTEGITY_IT_USR</i> | <p>Specifies the name of the service account user that is required for running the Protegity services on all the nodes in the cluster. This user account is a part of the <i>PROTEGITY_IT_USR_GROUP</i> group. This service user owns all the Protegity installation directories.</p> |

12.6.5.8.4 Installing the Big Data Protector on all the Nodes in an EMR Cluster

► To install the Big Data Protector:

1. Navigate to the directory that contains the *BdpInstallx.x.x_Linux_9.1.0.0.x.sh* script.
2. To run the installer, execute the following script.

```
./BdpInstallx.x.x_Linux_9.1.0.0.x.sh
```

3. Press ENTER.
The prompt to continue the installation of the Big Data Protector appears.

```
*****
Welcome to the Hadoop Big Data Protector Setup Wizard
*****
This will install the Hadoop Big Data Protector on your system.

This installation requires a Private Key file for communicating with other nodes in the
```

cluster.

Do you want to continue? [yes or no]:

4. To continue with the installation, type *yes*.

5. Press ENTER.

The prompt to enter path of the Private Key file (*.pem* file) appears.

```
Big Data Protector installation started
Enter the path of the Private Key (.PEM) file :
```

6. Type the path of the Private Key file.

7. Press ENTER.

The prompt to enter the ESA user name or IP address appears.

```
Unpacking...
Extracting files...

Preparing for cluster deploy, Please wait...
```

```
Enter ESA Hostname or IP Address:
```

8. If you have installed a proxy, then enter the IP address of the proxy node.

Alternatively, enter the host name or IP address of the ESA.

9. Press ENTER.

The prompt to enter the listening port for the ESA appears.

```
Enter ESA host listening port [8443]:
```

10. Enter the port for the ESA.

11. Press ENTER.

The prompt to enter the user name for the ESA appears.

```
Enter ESA Username:
```

12. Type the user name.

13. Press ENTER.

The prompt to enter the password for downloading the certificate appears.

```
*****
***** Welcome to the Pep Server Setup Wizard. *****
***** *****

Unpacking.....
Extracting files....
Unpacked pepserver compressed file...
PepServer Installing in Lead Node...
Please enter the password for downloading certificates[]:
```

14. Enter the password.

15. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Unpacking...
Extracting files...
Downloading certificates from X.X.X.X:8443...
  % Total    % Received % Xferd  Average Speed   Time     Time      Time  Current
               Dload  Upload   Total   Spent    Left  Speed
 100 20480  100 20480     0      0  103k      0  --::--  --::--  --::--  104k

Extracting certificates...
Certificates successfully downloaded and stored in /opt/protegility/defiance_dps/data
```

```

Protegrity PepServer installed in /opt/protegrity/defiance_dps.

PepServer installed on Lead node at location /opt/protegrity/defiance_dps.

Performing install on other nodes...

PepServer installed on other nodes at location /opt/protegrity/defiance_dps.

Check the status in /opt/protegrity/logs/pepper_server_setup.log

Select the Audit Store type where Log Forwarder(s) should send logs to.

[ 1 ] : Protegrity Audit Store
[ 2 ] : External Audit Store
[ 3 ] : Protegrity Audit Store + External Audit Store

Enter the no.:

```

16. To select the Audit Store type, select any one of the following options:

Table 12-53: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting using the Protegrity Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegrity Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegrity Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory. |

17. Press ENTER.

The prompt to enter the comma separated list of hostnames/IP addresses appears.

```

Enter comma-separated list of Hostnames/IP Addresses and/or Ports of Protegrity Audit
Store.
Allowed Syntax: hostname[:port][,hostname[:port],hostname[:port]...] (Default Value -
<ESA_IP_Address>:9200)
Enter the list:

```

18. To use the default value, press ENTER.

The prompt to enter the location of the Fluent Bit configuration file appears.

```
Enter the local directory path on this node that stores the custom Fluent-Bit
configuration files for External Audit Store:
```

Note: The configurator script will display this prompt only if you select option 2 or 3 in step 15. When you select option 2 or 3 in step 15, the custom configuration files are copied to the /<Installation directory>/fluent-bit/data/config.d/ directory on all the EMR nodes selected for installation.

19. Enter the path that contains the Fluent Bit configuration file.

20. Press ENTER.

The installer installs all the components required for the Big Data Protector.

```
*****
          Welcome to the LogForwarder Setup Wizard.
*****  
  

Unpacking.....  

Extracting files...  

Unpacked logforwarder compressed file...  

Logforwarder Installing in Lead Node...  

Unpacking...  

Extracting files...  
  

Logforwarder installed in /opt/protegility/fluent-bit.  
  

LogForwarder installed on Lead node at location /opt/protegility/fluent-bit.  
  

Performing install on other nodes...  
  

Logforwarder installed on other nodes at location /opt/protegility/fluent-bit.  
  

Check the status in /opt/protegility/logs/logforwarder_setup.log  
*****  

          Welcome to the Jpeplite Setup Wizard.  
*****  
  

Unpacking.....  

Extracting files...  

Unpacked jpeplite compressed file...  

Installing Jpeplite ....  
  

Jpeplite installed on lead node at location /opt/protegility/jpeplite.  
  

Performing install on other nodes...  
  

Jpeplite installed on other nodes at location /opt/protegility/jpeplite.  
  

Check the status in /opt/protegility/logs/jpeplite_setup.log  
*****  

          Welcome to the Hive Protector Setup Wizard.  
*****  
  

Unpacking.....  

Extracting files...  

Unpacked pephive compressed file...  
  

Hive Big Data Protector installed on lead node at location /opt/protegility/pephive.  
  

Performing install on other nodes...  
  

Hive Big Data Protector installed on other nodes at location /opt/protegility/pephive.  
  

Check the status in /opt/protegility/logs/pephive_setup.log  
*****  

          Welcome to the Pig Protector Setup Wizard.  
*****
```

```
Unpacking.....  
Extracting files...  
Unpacked peppig compressed file...  
  
Pig Big Data Protector installed on lead node at location /opt/protegity/peppig.  
  
Performing install on other nodes...  
  
Pig Big Data Protector installed on other nodes at location /opt/protegity/peppig.  
  
Check the status in /opt/protegity/logs/peppig_setup.log  
*****  
Welcome to the MapReduce Protector Setup Wizard.  
*****  
  
Unpacking.....  
Extracting files...  
Unpacked pepmapreduce compressed file...  
  
Mapreduce Big Data Protector installed on lead node at location /opt/protegity/  
pepmapreduce.  
  
Performing install on other nodes...  
  
Mapreduce Big Data Protector installed on other nodes at location /opt/protegity/  
pepmapreduce.  
  
Check the status in /opt/protegity/logs/pepmapreduce_setup.log  
*****  
Welcome to the Hbase Protector Setup Wizard.  
*****  
  
Unpacking.....  
Extracting files...  
Unpacked pepfhbase compressed file...  
  
Hbase Big Data Protector installed on lead node at location /opt/protegity/pephbase.  
  
Performing install on other nodes...  
  
Hbase Big Data Protector installed on other nodes at location /opt/protegity/pephbase.  
  
Check the status in /opt/protegity/logs/pephbase_setup.log  
*****  
Welcome to the Spark Protector Setup Wizard.  
*****  
  
Unpacking.....  
Extracting files...  
Unpacked pepspark compressed file...  
  
Spark Big Data Protector installed on lead node at location /opt/protegity/pepspark.  
  
Performing install on other nodes...  
  
Spark Big Data Protector installed on other nodes at location /opt/protegity/pepspark.  
  
Check the status in /opt/protegity/logs/pepspark_setup.log  
  
Starting Logforwarder on lead node...  
  
Starting Logforwarder on other nodes...  
  
Starting PepServer on lead node...  
  
Starting PepServer on other nodes...  
  
Hadoop Big Data Protector installed in /opt/protegity.  
  
Generating Big Data Protector installation status report ...  
  
Clearing previous logs files ...
```

```
Installation Status report generated in /opt/protegility/cluster_utils/
installation_report.txt
```

21. Restart the Hadoop, Hive, and HBase service daemon processes to start using the updated configuration.

Tip: You can delete the extracted files from temporary directory (`/opt/bigdata/`) because it is no longer required.

12.6.5.8.5 Installing the Big Data Protector on the New Nodes Added to an Existing EMR Cluster

Protegility provides the `BdpInstallx.x.x_Linux_<arch>_9.1.0.0.x.sh` script in the `<PROTEGILITY_DIR>/cluster_utils` directory. You can use this script to install the Big Data Protector on the new nodes that you add to an existing EMR cluster.

Before you begin

Note: Ensure that you install the Big Data Protector from an *ADMINISTRATOR* user having full *sudoer* privileges.

► To install Big Data Protector on the new nodes added to an existing EMR cluster:

1. Login to the Lead Node on the EMR cluster.
2. Navigate to the `<PROTEGILITY_DIR>/cluster_utils` directory.
3. In the `NEW_HOSTS_FILE` file, add an additional entry for each new node in the EMR cluster, on which you want to install the Big Data Protector.
The new nodes from the `NEW_HOSTS_FILE` file will be appended to the `CLUSTERLIST_FILE`.
4. To install the Big Data Protector on the new nodes, run the the following command.

```
./BdpInstallx.x.x_Linux_<arch>_9.1.0.0.x.sh -a <NEW_HOSTS_FILE>
```

5. Press ENTER.
The prompt to enter the path of the Private Key file (`.pem` file) appears.
6. Type the path of the Private Key file.
7. Press ENTER.
The script installs the Big Data Protector on the new nodes in the EMR cluster.

12.6.5.8.6 Verifying the Big Data Protector Parameters in EMR

Before you start using Big Data Protector, you must configure the required Protegility-related parameters in EMR. The Big Data Protector configuration parameters are set for the EMR cluster when it is installed on all the nodes in the cluster.

The following table provides the parameters that are set for the existing Amazon EMR cluster before using the Big Data Protector.

Table 12-54: Big Data Protector Classpath Configuration for an Existing EMR Cluster

| Name | Configuration File | Updated Classpath Parameter |
|-----------|---|--|
| MapReduce | <code>/etc/hadoop/conf/mapred-site.xml</code> | <code>mapreduce.application.classpath : /opt/protegility/pepmapreduce/lib/*,/opt/protegility/pephive/lib/*,/opt/protegility/bdp_version/</code> <code>mapreduce.admin.user.env : LD_LIBRARY_PATH=/opt/protegility/jpeplite/lib</code> |

| Name | Configuration File | Updated Classpath Parameter |
|-------|--|--|
| Hive | /etc/hive/conf/hive-site.xml /etc/tez/conf/tez-site.xml /etc/hive/conf/hive-env.sh | <pre>hive.exec.pre.hooks : com.protegrity.hive.PtyHiveUserPreHook</pre> <pre>tez.cluster.additional.classpath.prefix : /opt/protegrity/ pephive/lib/*:/opt/protegrity/bdp_version/</pre> <pre>tez.am.launch.env :</pre> <pre>LD_LIBRARY_PATH=/opt/protegrity/jpeplite/lib/</pre> <pre>export HIVE_CLASSPATH=\${HIVE_CLASSPATH}:./opt/ protegrity/pephive/lib/*:/opt/protegrity/bdp_version/</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegrity/jpeplite/lib/</pre> |
| Pig | /etc/pig/conf/pig-env.sh | <pre>PIG_CLASSPATH="/opt/protegrity/peppig/lib/*:/opt/ protegrity/bdp_version/"</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegrity/jpeplite/lib/</pre> |
| HBase | /etc/hbase/conf/hbase-site.xml /etc/hbase/conf/hbase-env.sh | <pre>hbase.coprocessor.region.classes : com.protegrity.hbase.PTYRegionObserver</pre> <pre>export HBASE_CLASSPATH=\$ {HBASE_CLASSPATH}:./opt/protegrity/pephbase/lib/*:/opt/ protegrity/bdp_version/</pre> <pre>export JAVA_LIBRARY_PATH=\$ {JAVA_LIBRARY_PATH}:./opt/protegrity/jpeplite/lib/</pre> |
| Spark | /etc/spark/conf/spark-defaults.conf | <pre>spark.driver.extraClassPath=/opt/protegrity/pephive/lib/*:/opt/ protegrity/pepspark/lib/*:/opt/protegrity/bdp_version/</pre> <pre>spark.executor.extraClassPath=/opt/protegrity/ pephive/lib/*:/opt/protegrity/pepspark/lib/*:/opt/protegrity/ bdp_version/</pre> <pre>spark.executor.extraLibraryPath= /opt/protegrity/jpeplite/lib</pre> <pre>spark.driver.extraLibraryPath= /opt/protegrity/jpeplite/lib</pre> <p>For Spark Version >= 2.4.0 and < 3.0.0</p> <pre>spark.executor.plugins com.protegrity.spark.PtyExecSparkPlugin</pre> <p>For Spark Version >= 3.0.0</p> <pre>spark.plugins com.protegrity.spark.PtyExecSparkPlugin</pre> |

12.6.5.8.7 Uninstalling the Big Data Protector from the EMR Cluster

This section outlines the procedures to uninstall the Big Data Protector from the EMR cluster. You can use any one of the following methods to remove the Big Data Protector from the EMR cluster:



- *Uninstalling the Big Data Protector from all the Nodes on the EMR Cluster*
- *Uninstalling the Big Data Protector from Selective Nodes on the EMR Cluster*

12.6.5.8.7.1 Uninstalling the Big Data Protector from all the Nodes on the EMR Cluster

1. Login to the Lead node as the *sudoer* user.
2. Navigate to the *<PROTEGILITY_DIR>/cluster_utils* directory on the Lead node.
3. To remove the Big Data Protector from all the nodes in the cluster, run the following script.

```
./uninstall.sh
```

4. Press ENTER.

The prompt to continue the uninstallation of the Big Data Protector appears.

```
*****
Welcome to the Hadoop Big Data Protector Uninstallation Wizard
*****
This will uninstall the Hadoop Big Data Protector on your system.
Do you want to continue? [yes or no]
```

5. To continue with the install, type *yes*.

6. Press ENTER.

The prompt to enter the path of the private key appears.

```
*****
Welcome to the Hadoop Big Data Protector Uninstallation Wizard
*****
This will uninstall the Hadoop Big Data Protector on your system.
Do you want to continue? [yes or no]
yes

Big Data Protector uninstallation started
Enter the path of the Private Key (.PEM) file :-
```

7. Type the path of the Private Key file.

8. Press ENTER.

The script stops the services and proceeds to uninstall the Big Data Protector from all the nodes in the cluster.

```
*****
Welcome to the Pep Server Setup Wizard.
*****

Uninstalling PepServer...
Stopping pep server. Please wait...

PepServer uninstalled on Lead node at location /opt/protegility/defiance_dps.

Performing uninstall on other nodes...

PepServer uninstalled on other nodes at location /opt/protegility/defiance_dps.

Check the status in /opt/protegility/logs/pepper_server_setup.log
*****
Welcome to the LogForwarder Setup Wizard.
*****

Uninstalling LogForwarder....
Stopping Logforwarder. Please wait...

LogForwarder uninstalled on Lead node at location /opt/protegility/fluent-bit.
```

```
Performing uninstall on other nodes...

Logforwarder uninstalled on other nodes at location /opt/protegity/fluent-bit.

Check the status in /opt/protegity/logs/logforwarder_setup.log
*****
        Welcome to the JpepLite Setup Wizard.
*****

Uninstalling JpepLite .....

JpepLite uninstalled on lead node at location /opt/protegity/jpeplite.

Performing uninstall on other nodes...

JpepLite uninstalled on other nodes at location /opt/protegity/jpeplite.

Check the status in /opt/protegity/logs/jpeplite_setup.log
*****
        Welcome to the Hive Protector Setup Wizard.
*****

Uninstalling PepHive .....

Hive Big Data Protector uninstalled on lead node at location /opt/protegity/pephive.

Performing uninstall on other nodes...

Hive Big Data Protector uninstalled on other nodes at location /opt/protegity/pephive.

Check the status in /opt/protegity/logs/pephive_setup.log
*****
        Welcome to the Pig Protector Setup Wizard.
*****

Uninstalling PepPig .....

Pig Big Data Protector uninstalled on lead node at location /opt/protegity/peppig.

Performing uninstall on other nodes...

Pig Big Data Protector uninstalled on other nodes at location /opt/protegity/peppig.

Check the status in /opt/protegity/logs/peppig_setup.log
*****
        Welcome to the MapReduce Protector Setup Wizard.
*****

Uninstalling PepMapreduce .....

Mapreduce Big Data Protector uninstalled on lead node at location /opt/protegity/
pepmapreduce.

Performing uninstall on other nodes...

Mapreduce Big Data Protector uninstalled on other nodes at location /opt/protegity/
pepmapreduce.

Check the status in /opt/protegity/logs/pepmapreduce_setup.log
*****
        Welcome to the Hbase Protector Setup Wizard.
*****

Uninstalling PepHbaseProtector .....

Hbase Big Data Protector uninstalled on lead node at location /opt/protegity/pephbase.

Performing uninstall on other nodes...

Hbase Big Data Protector uninstalled on other nodes at location /opt/protegity/pephbase.

Check the status in /opt/protegity/logs/pephbase_setup.log
*****
```

```
Welcome to the Spark Protector Setup Wizard.
*****
```

```
Spark Big Data Protector uninstalled on lead node at location /opt/protegity/pepspark.
Performing uninstall on other nodes...
Spark Big Data Protector uninstalled on other nodes at location /opt/protegity/pepspark.
Check the status in /opt/protegity/logs/pepspark_setup.log
Clearing previous log files ...
Uninstallation Status report generated in /opt/protegity/cluster_utils/
uninstallation_report.txt
Removing Protegity service user from all nodes...
Uninstallation process done.
```

9. After the script removes the Big Data Protector successfully from all the nodes, manually delete the <PROTEGITY_DIR> directory from the Lead node.
10. Restart all the Hadoop services.

12.6.5.8.7.2 Uninstalling the Big Data Protector from Selective Nodes on the EMR Cluster

To uninstall Big Data Protector from selective nodes in the EMR cluster, use the *node_uninstall.sh* utility in the <PROTEGITY_DIR>/cluster_utils directory.

Note: Ensure that you uninstall the Big Data Protector from an *ADMINISTRATOR* user having full *sudoer* privileges.

► To uninstall the Big Data Protector from selective nodes on the EMR Cluster:

1. Login to the Lead node.
2. Navigate to the <PROTEGITY_DIR>/cluster_utils directory.
3. Create a new hosts file.

For example, *NEW_HOSTS_FILE*

The *NEW_HOSTS_FILE* file contains the required nodes in the EMR cluster on which the Big Data Protector needs to be uninstalled.

4. Add the nodes on the EMR cluster, from which the Big Data Protector needs to be uninstalled in the new hosts file.
5. To remove the Big Data Protector from the nodes that are listed in the new hosts file, run the following command.

```
./node_uninstall.sh -c NEW_HOSTS_FILE
```

6. Press ENTER.
The prompt to enter the path of the Private Key file (*.pem* file) appears.
7. Type the path of the private key file.
8. Press ENTER.
The Big Data Protector is uninstalled from the nodes in the EMR cluster, which are listed in the new hosts file.
9. Check whether the nodes from which the Big Data Protector is uninstalled in Step 5 are removed from the *CLUSTERLIST_FILE* file.

12.6.5.8.8 Utilities

This section provides information about the following utilities:

- PEP Server Control (*cluster_pepsrvctrl.sh*) – Manages the PEP servers across the cluster
- LogForwarder Control (*cluster_logforwarderctrl.sh*) – Manages the Log Forwarders across the cluster
- Update Cluster Policy (*update_cluster_policy.sh*) – Updates the configurations and the certificates of the PEP servers across the cluster.

Note: Ensure that you run the utilities with a user (*OPERATOR_USER*) having sudo privileges for impersonating the service account (*EXECUTOR_USER* or *PROTEGRITY_IT_USR*, as configured).

When prompted, enter the path of the Private Key file (.pem file).

12.6.5.8.8.1 PEP Server Control

This utility (*cluster_pepsrvctrl.sh*), in the *<PROTEGRITY_DIR>/cluster_utils* directory, manages the PEP server services on all the nodes in the cluster that are listed in the BDP hosts file available in *<PROTEGRITY_DIR>/cluster_utils* directory.

The utility provides the following options:

- Start – Starts the PEP servers in the cluster.
- Stop – Stops the PEP servers in the cluster.
- Restart – Restarts the PEP servers in the cluster.
- Status – Reports the status of the PEP servers.

The utility (*pepsrvctrl.sh*), in the *<PROTEGRITY_DIR>/defiance_dps/bin*/directory, manages the PEP server services on the Lead node.

Note: When you run the PEP Server Control utility, you will be prompted to enter the path of the SSH private key file to securely login into the cluster nodes.

12.6.5.8.8.2 Log Forwarder Control

This utility (*cluster_logforwarderctrl.sh*), in the *<PROTEGRITY_DIR>/cluster_utils* directory, manages the Log Forwarder services on all the nodes in the cluster that are listed in the BDP hosts file available in *<PROTEGRITY_DIR>/cluster_utils* directory.

The utility provides the following options:

- Start – Starts the Log Forwarder on all the nodes in the cluster.
- Stop – Stops the Log Forwarder on all the nodes in the cluster.
- Restart – Restarts the Log Forwarder on all the nodes in the cluster.
- Status – Reports the status of the Log Forwarder on all the nodes.

The utility (*logforwarderctrl.sh*), in the *<PROTEGRITY_DIR>/fluent-bit/bin*/directory, manages the Log Forwarder services on the Lead node.

Note: When you run the Log Forwarder Control utility, you will be prompted to enter the path of the SSH private key file to securely login into the cluster nodes.

12.6.5.8.8.3 Update Cluster Policy

This utility (*update_cluster_policy.sh*), in the *<PROTEGRITY_DIR>/cluster_utils* directory, updates the configurations and the certificates of the PEP servers across the cluster.

For example, if you need to make any changes to the PEP server configuration, make the changes on the Lead node and then propagate the change to all the PEP servers in the cluster using the *update_cluster_policy.sh* utility.

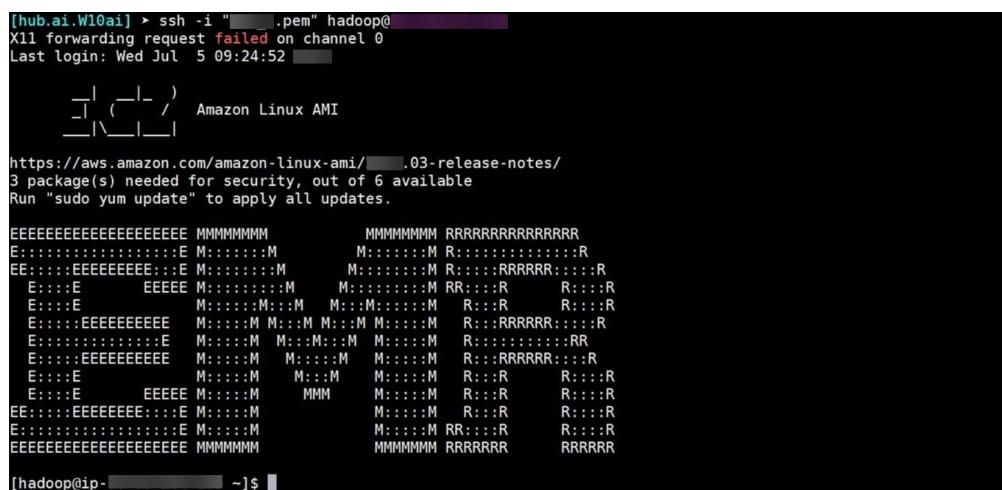
Note: Ensure that all the PEP servers in the cluster are stopped before running the *update_cluster_policy.sh* utility.

Note: When you run the Update Cluster Policy utility, you will be prompted to enter the path of the SSH private key file to securely login into the cluster nodes.

12.6.5.9 Using Hive UDFs with Amazon EMR

This section illustrates how Hive protector UDFs are used in the Amazon EMR environment with Big Data Protector installed.

The following figure displays the Amazon EMR screen that appears post authentication.



```
[hub.ai.W10ai] > ssh -i "----- .pem" hadoop@[REDACTED]
X11 forwarding request failed on channel 0
Last login: Wed Jul  5 09:24:52 [REDACTED]
[REDACTED] ( [REDACTED] ) - Amazon Linux AMI
[REDACTED]\_\_|_____|

https://aws.amazon.com/amazon-linux-ami/ [REDACTED].03-release-notes/
3 package(s) needed for security, out of 6 available
Run "sudo yum update" to apply all updates.

EEEEEEEEEEEEEEEEEE MMMMMMM MRRRRRRRRRRRRRR
E:::::::::::E M:::::M M:::::M R:::::R R:::::R
EE:::::EEEEE:M:::M M:::::M R:::::R R:::::R
E:::::E EEEEE:M:::M M:::M:::M R:::R R:::R
E:::::EEEEEEEEE M:::M M:::M M:::M R:::RRRRR:::R
E:::::::::::E M:::M M:::M:::M M:::::M R:::::RR
E:::::EEEEE:::::M M:::M M:::M R:::::R R:::::R
E:::::E EEEEE:M:::M M:::M R:::::R R:::::R
EE:::::EEEEE:::::E M:::M M:::M R:::::R R:::::R
E:::::::::::E M:::M M:::::M R:::::R R:::::R
EEEEEEEEEEEEEEEEEE MMMMMMM MRRRRRRR RRRRRR
[hadoop@ip-[REDACTED] ~]$
```

Figure 12-262: Amazon EMR Authentication screen

Create a *TEST_DEMO* table in Hive with the values listed in the following table.

Table 12-55: Values used in the *TEST_DEMO* table

| |
|---------------------|
| Hultgren Caylor |
| Bourne Jose |
| Sorce Hatti |
| Lorie Garvey |
| Belva Beeson |
| Bucky Hapte |
| Lorrie Joselyn |
| MacDonell Hutchings |
| Dachia Bisset |
| Calabrese Normi |

The following figure displays the data contained in the *TEST_DEMO* table.

```

0: jdbc:hive2://localhost:10000/default> select * from TEST_DEMO;
INFO : Compiling command(queryId=hive_20170705101135_d99f648d-5353-47eb-b0ef-db25e062ff38): select * from TEST_DEMO
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:test_demo.col, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20170705101135_d99f648d-5353-47eb-b0ef-db25e062ff38); Time taken: 0.11 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20170705101135_d99f648d-5353-47eb-b0ef-db25e062ff38): select * from TEST_DEMO
INFO : Completed executing command(queryId=hive_20170705101135_d99f648d-5353-47eb-b0ef-db25e062ff38); Time taken: 0.0 seconds
INFO : OK
+-----+-----+
| test_demo.col |
+-----+-----+
| Hultgren Taylor |
| Bourne Jose |
| Sorce Hatti |
| Lorie Garvey |
| Belva Beeson |
| Bucky Hapte |
| Lorrie Joselyn |
| MacDonell Hutchings |
| Dachia Bisset |
| Calabrese Normi |
+-----+-----+
10 rows selected (0.448 seconds)
0: jdbc:hive2://localhost:10000/default>

```

Figure 12-263: Contents of the *TEST_DEMO* table

Execute the *ptyProtectStr()* UDF to protect the data contained in the *TEST_DEMO* table and save it to a new table named *PROTECTED_TABLE*.

The following figure displays the protected data contained in the *PROTECTED_TABLE* table.

```

0: jdbc:hive2://localhost:10000/default> select * from PROTECTED_TABLE;
INFO : Compiling command(queryId=hive_20170705101222_30faa5bf-5f19-44ff-8c62-a3aace42d239): select * from PROTECTED_TABLE
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:protected_table.c0, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20170705101222_30faa5bf-5f19-44ff-8c62-a3aace42d239); Time taken: 0.07 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20170705101222_30faa5bf-5f19-44ff-8c62-a3aace42d239): select * from PROTECTED_TABLE
INFO : Completed executing command(queryId=hive_20170705101222_30faa5bf-5f19-44ff-8c62-a3aace42d239); Time taken: 0.001 seconds
INFO : OK
+-----+-----+
| protected_table.c0 |
+-----+-----+
| HDQtSwlux IPPLor |
| Bv1l0v Vfse |
| SMswW WLxti |
| LGjEq HWtPey |
| BPZUX OmRlon |
| BWaso SuDte |
| LiYCag oclaiyn |
| MLTNRgMma ihFaqycgs |
| DPKFnz hLTRet |
| CdKGXBxeFB qpGmi |
+-----+-----+
10 rows selected (0.1 seconds)
0: jdbc:hive2://localhost:10000/default>

```

Figure 12-264: Protected Data in the *PROTECTED_TABLE* table

Execute the *ptyUnprotectStr()* UDF to unprotect the data contained in the *PROTECTED_TABLE* table.

The following figure displays the unprotected data from the *PROTECTED_TABLE* table.

```

0: jdbc:hive2://localhost:10000/default> select ptyUnprotectStr(c0, 'TE_A_S13_L1R2_Y') from protected_table;
INFO : Compiling command(queryId=hive_20170705101346_630bee01-ee3a-4f4c-9d21-9885944e8038): select ptyUnprotectStr(c0, 'TE_A_S13_L1R2_Y') from protected_table
INFO : Semantic Analysis Completed
INFO : Returning Hive schema: Schema(fieldSchemas:[FieldSchema(name:c0, type:string, comment:null)], properties:null)
INFO : Completed compiling command(queryId=hive_20170705101346_630bee01-ee3a-4f4c-9d21-9885944e8038); Time taken: 0.113 seconds
INFO : Concurrency mode is disabled, not creating a lock manager
INFO : Executing command(queryId=hive_20170705101346_630bee01-ee3a-4f4c-9d21-9885944e8038): select ptyUnprotectStr(c0, 'TE_A_S13_L1R2_Y') from protected_table
INFO : Completed executing command(queryId=hive_20170705101346_630bee01-ee3a-4f4c-9d21-9885944e8038); Time taken: 0.001 seconds
INFO : OK
+-----+-----+
| c0 |
+-----+-----+
| Hultgren Taylor |
| Bourne Jose |
| Sorce Hatti |
| Lorie Garvey |
| Belva Beeson |
| Bucky Hapte |
| Lorrie Joselyn |
| MacDonell Hutchings |
| Dachia Bisset |
| Calabrese Normi |
+-----+-----+
10 rows selected (0.151 seconds)
0: jdbc:hive2://localhost:10000/default>

```

Figure 12-265: Unprotected Data from the *PROTECTED_TABLE* table

12.6.5.10 Best Practices for Using Big Data Protector on EMR

The following table lists some best practices for using Protegility Big Data Protector on Amazon EMR.



Table 12-56: Best Practices for using Protegility Big Data Protector on Amazon EMR

| Practice | Description |
|--|---|
| Select the appropriate instance size | Based on the anticipated workload on the EMR cluster, choose the right instance size for the EMR cluster. Some workloads would consume more CPU, disk space, or memory, or a combination of these factors. |
| Select the appropriate number of instances | Based on the size of your data set, select the right number of instances to improve the parallel processing of data. |
| Select the appropriate memory on the instances | Based on the data that needs to be processed, use instances with more memory as Hadoop tries to use as much memory as possible. To improve performance, it is beneficial to process data in memory versus disk space. |
| Select the right number of instances | Based on the size of your data set, select the right number of instances to improve the parallel processing of data. |
| Structure the data better | To optimize performance, structure your data better to limit the amount of data processed by Hadoop. |

12.6.6 Installing the Big Data Protector on a Dataproc Cluster

The Google Cloud Platform (GCP) is a cloud-based computing service, which provides several services, such as, computing power through Compute Engine, storage through Cloud Storage, Hadoop Cloud Dataproc, and so on. The GCP runs on servers that are located in several data centers, which are organized by geographical regions, across the world.

You can run the Big Data Protector 9.1.0.0 on Google Dataproc, using GCP, to efficiently run analytics and scale in a cost effective manner.

Note: The GCP supports Storage Buckets for storing data.

This section considers the storage buckets as reference for the input and output data.

If you want to use HDFS for the input and output data, then contact Protegility Support.

This section describes the tasks that need to be performed for installing the Big Data Protector on a new Dataproc cluster.

12.6.6.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before installing the Big Data Protector on a Dataproc cluster:

- It is recommended to be familiar with the following parts:
 - The Google Dataproc environment
 - Storage bucket, used to store the Big Data Protector installation files
 - Initialization Action, used to invoke the installation of Big Data Protector
 - Google Virtual Private Cloud (VPC)
- The ESA 9.1.0.0, installed and running on-premise.
- The following ports are configured on the ESA and the nodes in the Dataproc cluster, which will run the Big Data Protector:

Table 12-57: List of Ports for the Big Data Protector

| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|----------------------|-----------|---|--------------|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a policy. |



| Destination Port No. | Protocols | Sources | Destinations | Descriptions |
|----------------------|-----------|--|--|--|
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance (Elasticsearch) through port 9200. |
| 15780 | | Protector on the Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that port. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses localhost port 16700. |

Port Requirement for Proxy Node (Optional)

| | | | | |
|------|-----|--|---|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The PEP server communicates with the Proxy node through port 8443, which substitutes the stream to the ESA. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | PTY Proxy Node (Similar or different Big Data Protector cluster node) | The Log Forwarder communicates with the Proxy node through port 9200, which substitutes the stream to the Protegility Audit Store Appliance. |

- The Public-Private Key pair file is created and saved as a .ppk file and the Public SSH key is configured on the GCP at **Compute Engine > Metadata** in the Dataproc cluster to communicate with all nodes in the cluster.
- A storage bucket and an Access Token for authenticating and accessing the storage bucket is available.

12.6.6.2 Creating a Storage Bucket on the Google Cloud Platform

If you are installing the Big Data Protector on a new Dataproc cluster, then you must create a storage bucket to copy the Big Data Protector installation files, which are created using the Configurator script.

The storage bucket provides an isolated space for storing the files required to install the Big Data Protector on the new Dataproc cluster. These files are then utilized by the GCP environment when setting up the new Dataproc cluster. Even if the Dataproc cluster is terminated, the storage bucket continues to retain the files required to install the Big Data Protector.

Note: For more information about creating a storage bucket, refer to the GCP documentation for [Creating a Storage Bucket](#).

After a storage bucket is created, an Access Token for authenticating and accessing the bucket is required.

Note: For more information about generating an Access Token for Google Cloud Storage, refer to <https://cloud.google.com/storage/docs/authentication#apiauth>.

12.6.6.3 Downloading the Big Data Protector Package

After receiving the Big Data Protector installation package from Protegility, copy it to any instance or node that has connectivity to the ESA.

12.6.6.4 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script to continue the installation of the Big Data Protector on all the nodes in the Dataproc cluster.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a machine or a Google Dataproc node that has connectivity to the ESA.
2. Copy the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz* to the required directory.
For example, */opt/bigdata* directory.
3. To extract the *BDP_Configurator_DATAPROC-<DATAPROC_Version>_9.1.0.0.x.sh* script from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the following files:

- *BDP_Configurator_DATAPROC-<DATAPROC_Version>_9.1.0.0.x.sh*
- *ProxySetup_Linux_x64_<core_version>.sh*

12.6.6.5 Setting up the Proxy

If you want to configure a proxy, which connects to the ESA on one end, and communicates with the Dataproc cluster nodes containing Big Data Protector on the other end, then perform the following task.

► To setup the Proxy for the Dataproc Cluster:

1. To install the proxy before you install the Big Data Protector, perform the following steps.
 - a. Copy the *ProxySetup_Linux_x64_<core_version>.sh* script to the node that will act as the proxy.

Note: The node where you configure the proxy will communicate with the ESA and all other nodes in the Dataproc cluster.

- b. To install the proxy, run the following command.

```
./ProxySetup_Linux_x64_<core_version>.sh -esa host (-elastic elastic_host) (-dir installation_directory) (-port proxy_port) (-port_es proxy_port_es)
```

Table 12-58: Parameters in the proxy setup command

| Parameter | Description |
|------------------------------------|---|
| <code>-esa <ESA_Host></code> | Specifies the ESA hostname or IP address that the proxy service will use for communication. |

| Parameter | Description |
|--|--|
| <code>-dir <Installation_Directory></code> | Specifies the location to install the proxy service. This is an optional parameter. If you fail to specify this parameter, then the script will install the proxy service in the <code><PROTEGILITY_DIR>/proxy</code> directory. |
| <code>-port <Proxy_Port></code> | Specifies the ESA port that the proxy service will use for communication. This is an optional parameter. If you fail to specify this parameter, then the script will use <code>8443</code> as the default port. |
| <code>-elastic <elastic_host></code> | Specifies the hostname or IP address for the Protegility Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use the ESA hostname as the default value. |
| <code>-port_es <proxy_port_es></code> | Specifies the port number for the Protegility Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use <code>9200</code> as the default value. |

2. To install the Proxy service on any node in the Dataproc cluster after installing the Big Data Protector, perform the following steps.

- a. Copy the `ProxySetup_Linux_x64_<core_version>.sh` script to the node in the cluster that will act as the proxy.

Note: The node where you configure the proxy will communicate with the ESA and all the other nodes in the Dataproc cluster.

- b. To install the proxy, run the following command.

```
./ProxySetup_Linux_x64_<core_version>.sh -esa host (-elastic elastic_host) (-dir installation_directory) (-port proxy_port) (-port_es proxy_port_es)
```

Table 12-59: Parameters in the proxy setup command

| Parameter | Description |
|--|--|
| <code>-esa <ESA_Host></code> | Specifies the ESA hostname or IP address that the proxy service will use for communication. |
| <code>-dir <Installation_Directory></code> | Specifies the location to install the proxy service. This is an optional parameter. If you fail to specify this parameter, then the script will install the proxy service in the <code><PROTEGILITY_DIR>/proxy</code> directory. |
| <code>-port <Proxy_Port></code> | Specifies the ESA port that the proxy service will use for communication. This is an optional parameter. If you fail to specify this parameter, then the script will use <code>8443</code> as the default port. |
| <code>-elastic <elastic_host></code> | Specifies the hostname or IP address for the Protegility Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use the ESA hostname as the default value. |
| <code>-port_es <proxy_port_es></code> | Specifies the port number for the Protegility Audit Store appliance (ESA/PSU). This is an optional parameter. If you fail to specify this parameter, then the script will use <code>9200</code> as the default value. |

- c. Open the `pepserver.cfg` file in insert mode.
d. In the `pepserver.cfg` file, replace the IP address of the ESA with the IP address of the proxy node.

Note:

- Replace the IP address of the ESA with the IP address of the proxy node in the `pepserver.cfg` file on all the nodes.
- If you specify the port number while installing the proxy, then you must update the same port in the `pepserver.cfg` file on all the nodes.



- e. Save the changes to the *pepservice.cfg* file.
- f. Restart the PEP server on all the nodes.
3. Navigate to the proxy directory.
For example, <*Installation_Directory*>/proxy/bin/
4. To start the proxy service, run the following command.

```
./proxyctrl start
```

5. Press ENTER.
6. To verify whether all the proxy services are up and running after setting up the proxy, run the following command.

./proxyctrl status
7. Press ENTER.
The status of the Proxy service appears.

12.6.6.6 Running the Configurator Script

You must execute the configurator script to create the installation files required to install the Big Data Protector on a new Dataproc cluster.

The configurator script performs the following tasks:

- Downloads the certificates from the ESA.
- Updates the Big Data Protector installation package for a new Dataproc cluster.
- Creates the bootstrap installer and classpath configurator scripts for the new Dataproc cluster.
- Copies the Big Data Protector installation package, bootstrap installer, and classpath configurator scripts to the storage bucket.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_DATAPROC-<DATAPROC_Version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command.

```
./BDP_Configurator_DATAPROC-<DATAPROC_Version>_9.1.0.0.x.sh
```

2. Press ENTER.
The prompt to continue the installation of the Big Data Protector appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will create the Big Data Protector Installation files for GCP Dataproc.
Do you want to continue? [yes or no]:
```

3. To continue the installation of the Big Data Protector, type *yes*.
4. Press ENTER.
The prompt to enter the Google Cloud Storage URI to upload the Big Data Protector installation files appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will create the Big Data Protector Installation files for GCP Dataproc.
Do you want to continue? [yes or no]: yes

Protegility Big Data Protector Configurator started...
```

Generating Big Data Protector for a new Google Dataproc cluster.....
 Enter the Google Cloud Storage URI where the BDP Installation files are to be uploaded.
 (E.g. gs://examplebucket/folder):

- Type the path of the storage bucket, which was created in section [Creating a Storage Bucket on the Google Cloud Platform](#), in the following format.

```
gs://<bucket_name>
```

where:

<bucket_name>: Name of the storage bucket

- Press ENTER.

The prompt to select the method to generate the installation file appears.

```
Choose one option among the following for BDP Installation files:  

[1] : Upload files to 'gs://<google_cloud_storage_location>' Google Cloud Storage URI.  

[2] : Generate files locally to current working directory. (You would have to manually  

upload the files to the specified Google Cloud Storage URI)  

[ 1 or 2 ]:
```

- Press ENTER.

The prompt to enter the Access Token appears.

```
Enter the Access Token:
```

- Type the Access Token for the bucket, which was created in section [Creating a Storage Bucket on the Google Cloud Platform](#).

- Press ENTER.

The prompt to enter the installation directory appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.  

[default:- /opt/protegility]:
```

- Enter the directory where you want to install the Big Data Protector.

Note: To install the Big Data Protector in the default directory, `/opt/protegility`, press ENTER.

- Press ENTER.

The prompt to enter the hostname or IP address of the ESA appears.

```
Enter the ESA Hostname or IP Address:
```

- If a proxy is setup, then enter the IP address of the proxy node. Alternatively, enter the IP address of the ESA.

- Press ENTER.

The prompt to enter the listening port for the ESA appears.

```
Enter ESA host listening port [8443]:
```

- Enter the listening port for the ESA.

- Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

- Enter the ESA user name.

- Press ENTER.

The prompt to select the Audit Store type appears.

```
Select the Audit Store type where Log Forwarder(s) should send logs to.  
[ 1 ] : Protegility Audit Store  
[ 2 ] : External Audit Store  
[ 3 ] : Protegility Audit Store + External Audit Store
```

Enter the no.:

- To select the Audit Store type, select any one of the following options:

Table 12-60: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting using the Protegility Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

- Press ENTER.

The prompt to enter the comma-separated list of hostnames or IP address appears.

```
Enter comma-separated list of Hostnames/IP Addresses and/or Ports of Protegility Audit Store.  
Allowed Syntax: hostname[:port][,hostname[:port],hostname[:port]...] (Default Value - X.X.X.X:9200)  
Enter the list:
```

- Enter the IP address of the Protegility Audit Store appliance.

- Press ENTER.

The prompt to enter the local directory path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this node that stores the custom Fluent-Bit configuration files for External Audit Store:
```

Note: The configurator script will display this prompt only if you select option *2* or *3* for the Audit Store type.

- Enter the path of the location that stores the custom Fluent Bit configuration files for an external audit store.

- Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

24. To generate the logs for the PEP server in a file, type *yes*.

25. Press ENTER.

The script extracts the installation files and the prompt to enter the password to download the certificates appears.

```
PepServer's log will be generated in a file.
*****
Welcome to the Pep Server Setup Wizard.
*****  
  
Unpacking.....  
Extracting files...  
Unpacked pepserver compressed file...  
Temporarily setting up PepServer defiance_dps directory structure on current node...  
Please enter the password for downloading certificates  
[ ]:
```

26. Enter the password for the ESA.

27. Press ENTER.

The configurator script downloads the certificates from the ESA and uploads the following installation files to the Google Cloud Storage:

- *BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz*: The Big Data Protector installation package for a new Dataproc cluster.
- *bdp_bootstrap_installer.sh*: The Bootstrap installer script for installing the Big Data Protector on a new Dataproc cluster.
- *bdp_classpath_configurator.py*: The script containing the required classpath settings for running the Big Data Protector on the Dataproc cluster.

```
Unpacking...  
Extracting files...  
Downloading certificates from X.X.X.X:8443...  
% Total % Received % Xferd Average Speed Time Time Current  
          Dload Upload Total Spent Left Speed  
100 20480 100 20480 0 0 39148 0 --::-- --::-- --::-- 39158  
  
Extracting certificates...  
Certificates successfully downloaded and stored in /<installation_dir>/defiance_dps/data  
Protegility PepServer installed in /<installation_dir>/defiance_dps.
```

Started Uploading the generated installation files via Google Cloud Storage REST API.....

Uploading *bdp_bootstrap_installer.sh* to Google Storage.
bdp_bootstrap_installer.sh uploaded to Google Storage.

Uploading *bdp_classpath_configurator.py* to Google Storage.
bdp_classpath_configurator.py uploaded to Google Storage.

Uploading *BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz* to Google Storage.
BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz uploaded to Google Storage.

Successfully Uploaded *BigDataProtector_Linux-ALL-64_x86-64_DATAPROC-<DATAPROC_Version>-64_9.1.0.0.x.tgz*, *bdp_bootstrap_installer.sh*, *bdp_classpath_configurator.py* to Google Cloud Storage bucket 'gs://<Google_cloud_storage_location>'

Successfully Generated installation files at ./Installation_Files/ directory.

Successfully configured Big Data Protector for a new Google Dataproc cluster..

If you select option 2 to generate the installation files, then the configurator script will generate the installation files in a local directory.

```
*****
Welcome to the Pep Server Setup Wizard.
*****

Unpacking..... .
Extracting files.... .
Unpacked pepserver compressed file...
Temporarily setting up PepServer defiance_dps directory structure on current node...
Please enter the password for downloading certificates
[]:

Unpacking... .
Extracting files... .
Downloading certificates from X.X.X.X:8443... .
    % Total      % Received   % Xferd  Average Speed   Time     Time     Time  Current
          Dload     Upload   Total   Spent   Left  Speed
100 20480  100 20480     0      0   138k      0 --::-- --::-- --::--  138k

Extracting certificates... .
Certificates successfully downloaded and stored in /<installation_directory>/defiance_dps/
data

Protegility PepServer installed in /<installation_directory>/bdp-doc/defiance_dps.

Successfully Generated installation files at ./Installation_Files/ directory.

Upload these files to Google Cloud Storage bucket 'gs://<Google_cloud_storage_location>' manually.

Successfully configured Big Data Protector for a new Google Dataproc cluster..
```

28. After all the Big Data Protector files are copied to the storage bucket, on the **Google Cloud Storage** page, refresh the contents of the storage bucket.

The installation files generated and uploaded by the configurator script appear in the storage bucket.

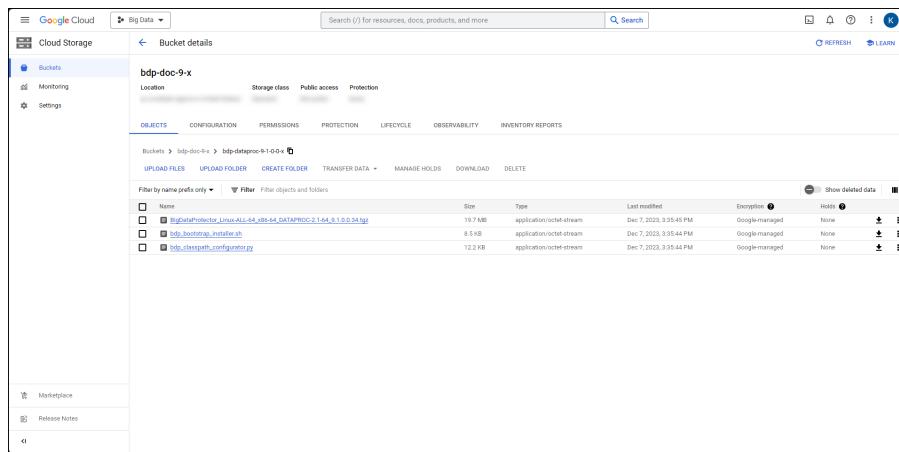


Figure 12-266: Google Cloud Storage Bucket Containing the Big Data Protector Installation Files

Note: The build number, in the installation files, will reflect the version that you have downloaded from the [My.Protegility](#) portal.

12.6.6.7 Installing the Big Data Protector

Perform the following steps to create a Dataproc cluster on GCP and install the Big Data Protector on all the nodes.

► To install the Big Data Protector on a new Dataproc cluster:

1. Login to your Google Cloud account.
2. To access the navigation menu, click . The navigation menu appears.
3. From the navigation menu, navigate to **Dataproc > Clusters**. The **Clusters** page appears.

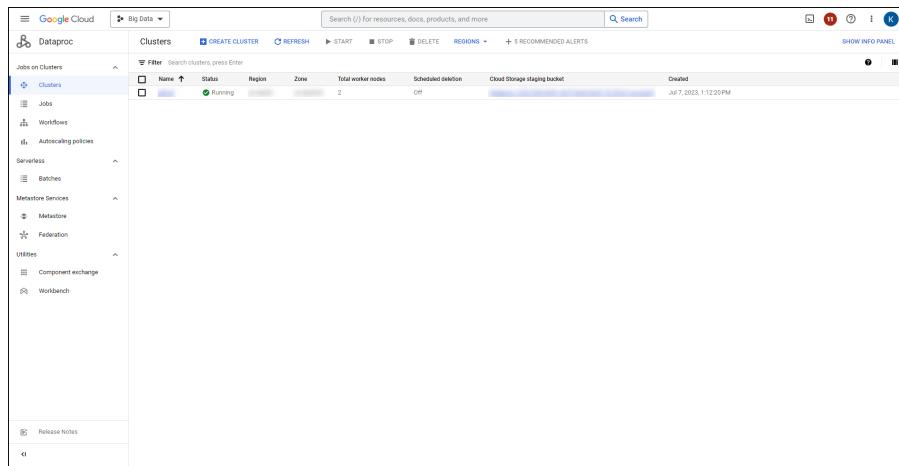


Figure 12-267: Clusters Page

4. To create a new cluster and install the Big Data Protector, click **CREATE CLUSTER**. The option to select the type of cluster appears.

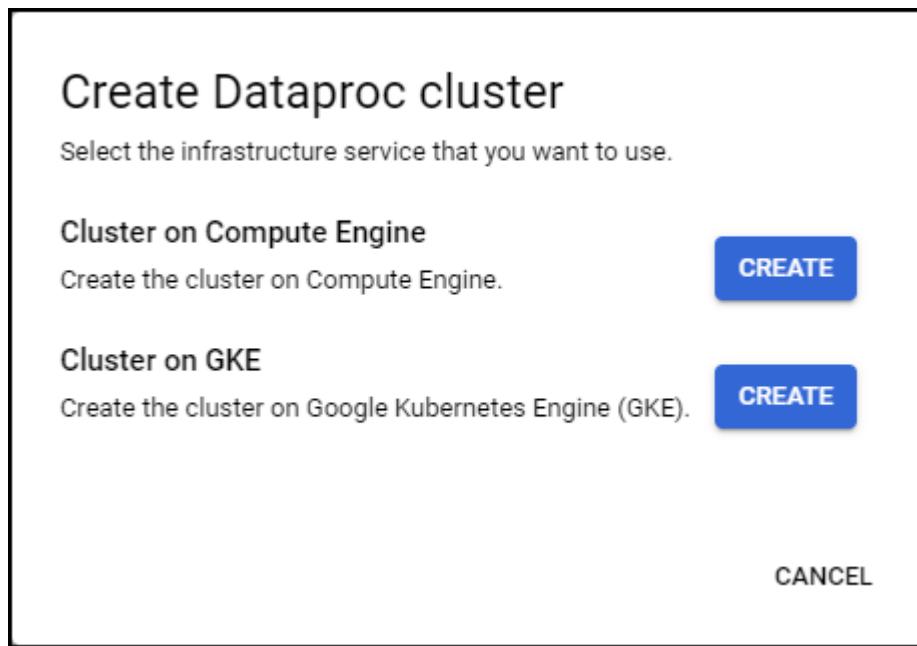


Figure 12-268: Option to select the Cluster Type

5. To create a cluster, besides the **Cluster on Compute Engine**, click **CREATE**. The **Set up cluster** tab of the **Create a Dataproc cluster on Compute Engine** wizard appears.

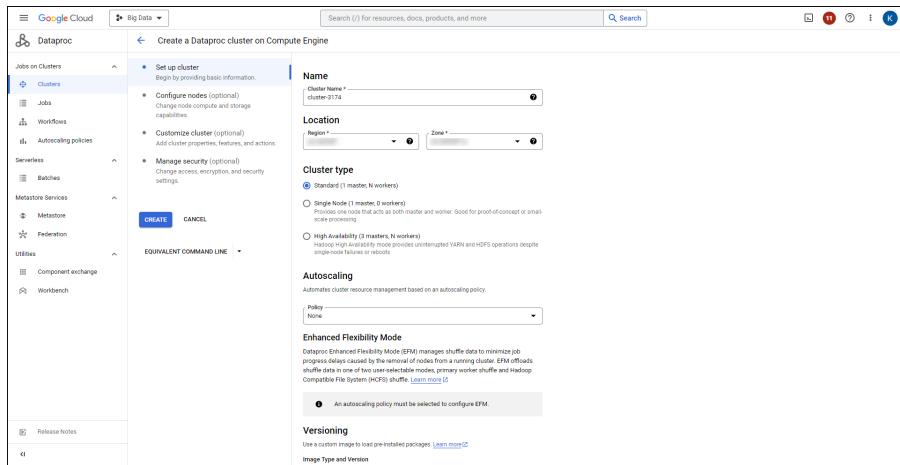


Figure 12-269: Set up cluster Tab

6. In the **Cluster Name** box, enter a name to identify the cluster.
7. From the **Region** list, select the required region.
8. From the **Zone** list, select the required zone.
9. Under **Cluster type**, select the required configuration for the cluster.
10. To use a custom image, under **Versioning**, click **CHANGE**.
The **Choose Image Version** pane appears.

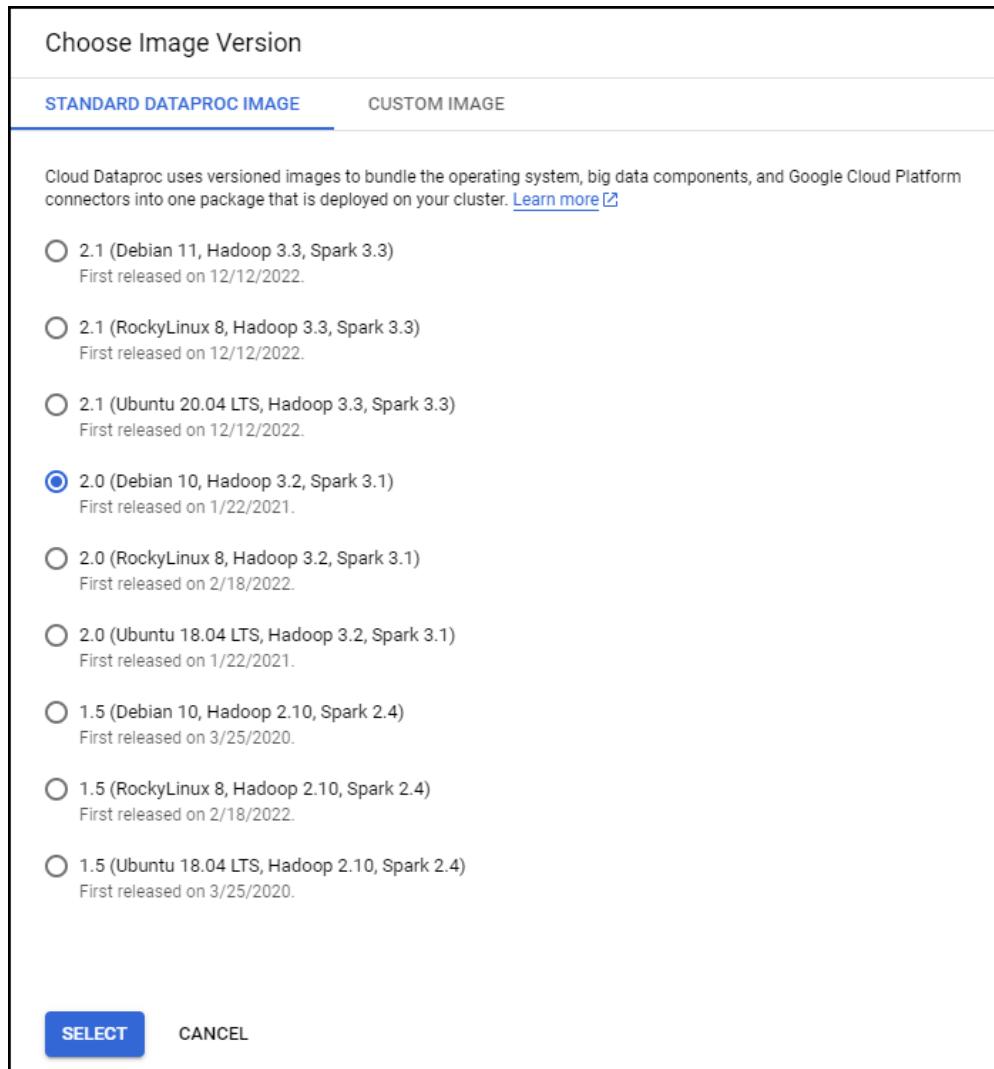


Figure 12-270: Choose Image Version Pane

11. From the **Choose Image Version** pane, select the required operating system image.
12. Click **SELECT**.
The selected image appears under the **Versioning** section.
13. Under **Network Configuration**, select the required connection type.
14. From the left, click the **Customize cluster (optional)** tab.
The **Customize cluster (optional)** tab appears.

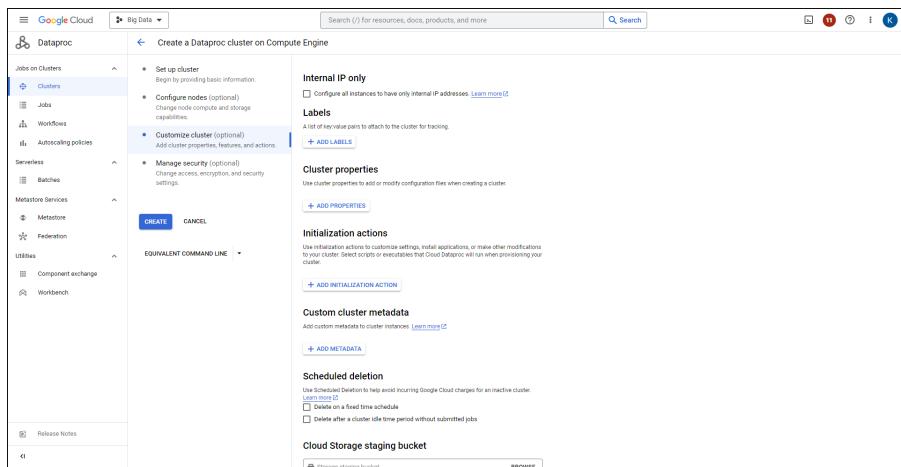


Figure 12-271: Customize cluster options

- To select a script to run while provisioning the cluster, under Initialization actions, click **ADD INITIALIZATION ACTION**.

The option to browse the storage bucket and select the script appears.

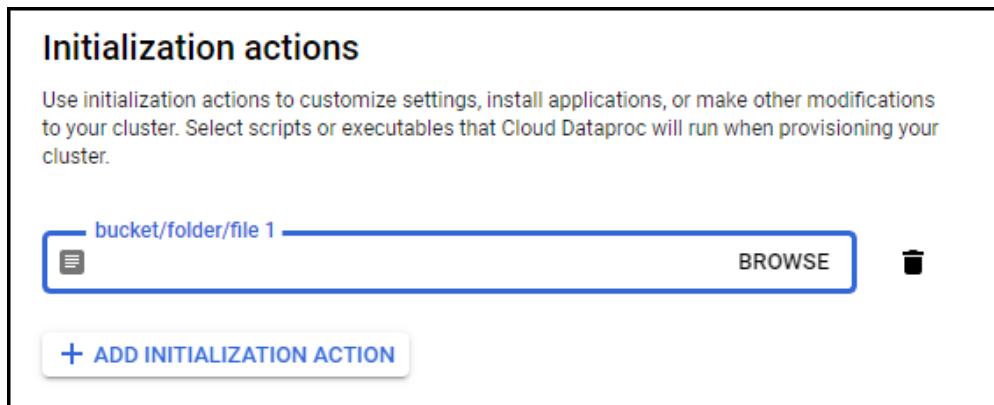


Figure 12-272: Option to select a script

- To select the script, click **BROWSE**.
The **Select object** pane appears.

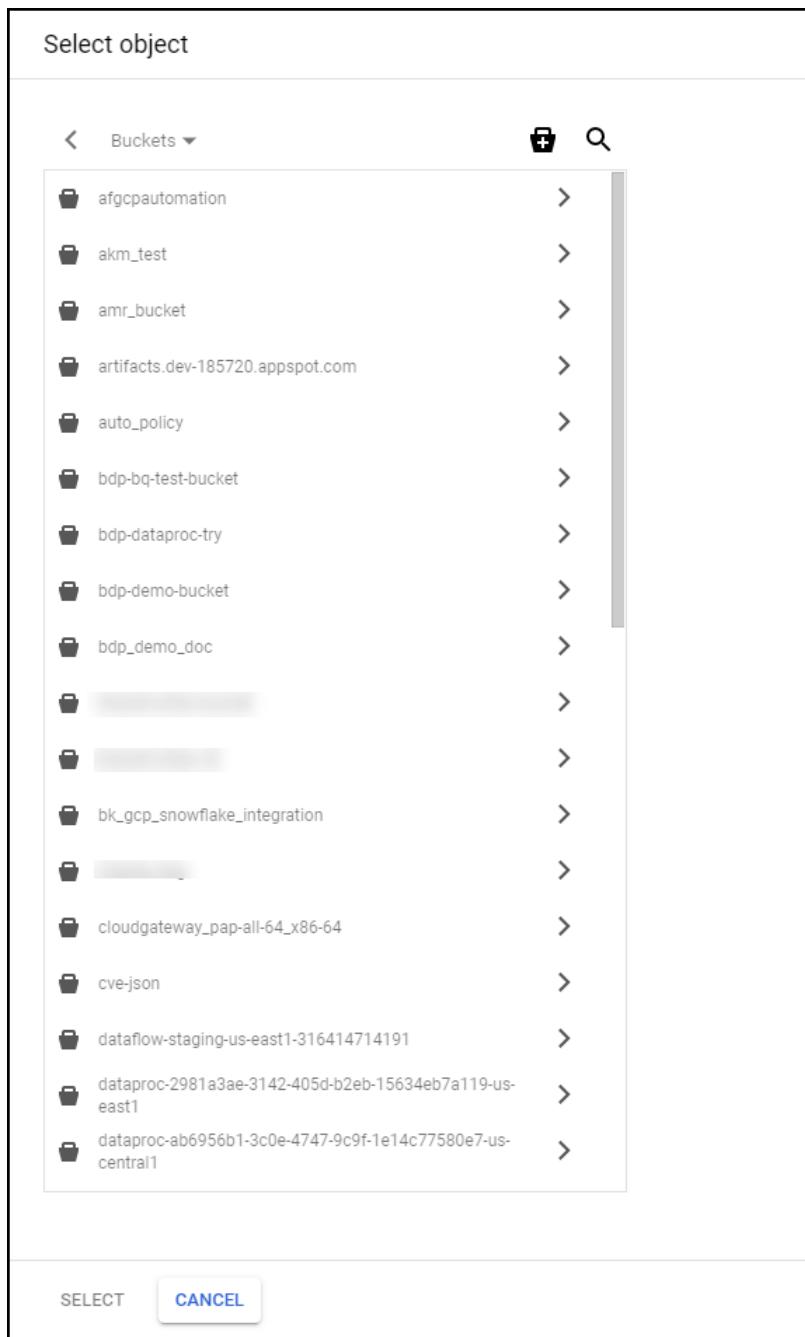


Figure 12-273: Select Object Pane

17. Browse to the required bucket location that contains the script.
18. To use the script, click **SELECT**.
The location and the name of the script appears under **Initialization actions**.

Initialization actions

Use initialization actions to customize settings, install applications, or make other modifications to your cluster. Select scripts or executables that Cloud Dataproc will run when provisioning your cluster.

bucket/folder/file 1 —

bdp-doc-9-x/bdp-dataproc-9-1-0-0-x/bdp_bootstrap_installer. [BROWSE](#)



[+ ADD INITIALIZATION ACTION](#)

Figure 12-274: Script Details

19. To enable API access to Google Cloud services, perform the following steps.
 - a. Click the **Manage security (optional)** tab.
 - b. Under **Project access**, select the **Enables the cloud-platform scope for this cluster** check box.
20. To create the cluster with the specified settings and components, click **CREATE**.

Note: The process of provisioning the cluster might take some time depending on the configuration and the options that you select.

The Google Cloud Platform provisions the cluster and it appears on the **Clusters** page.

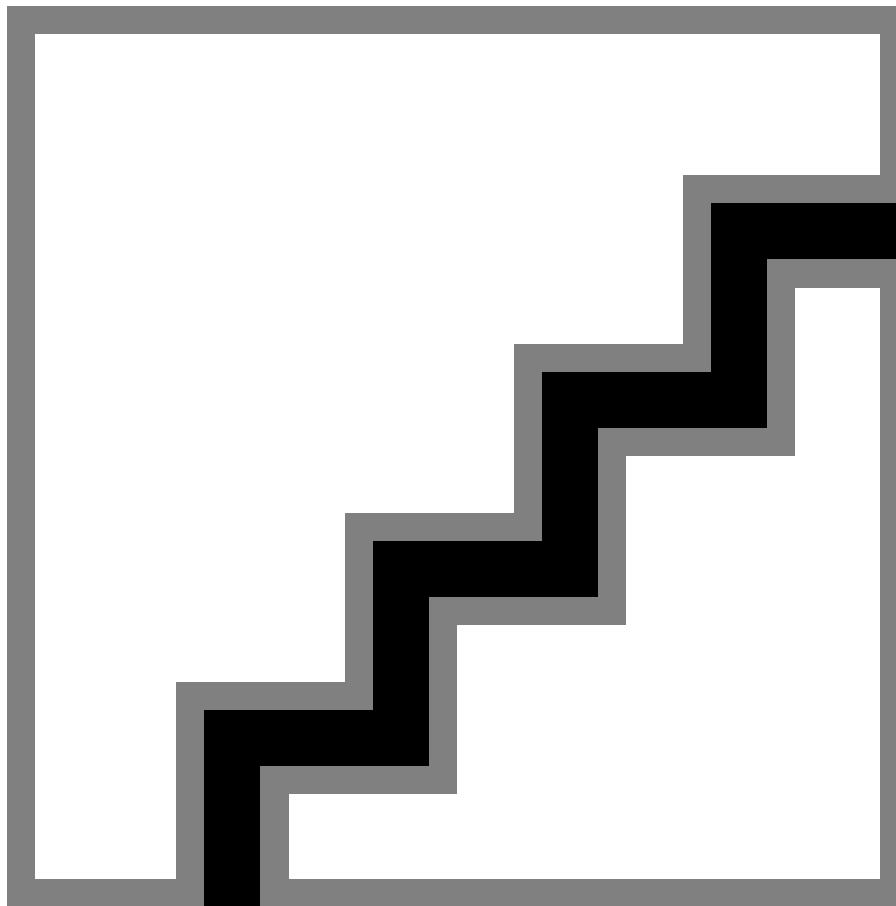


Figure 12-275: Cluster Provisioned

21. To display an overview of the cluster, click the Dataproc cluster.
The details of the cluster appear.

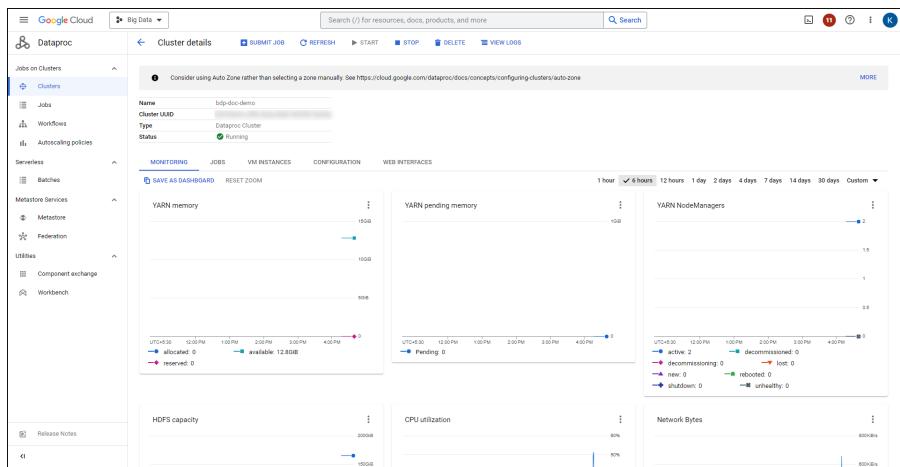


Figure 12-276: Cluster Details

22. To verify the number of nodes in the cluster, click the **VM INSTANCES** tab.

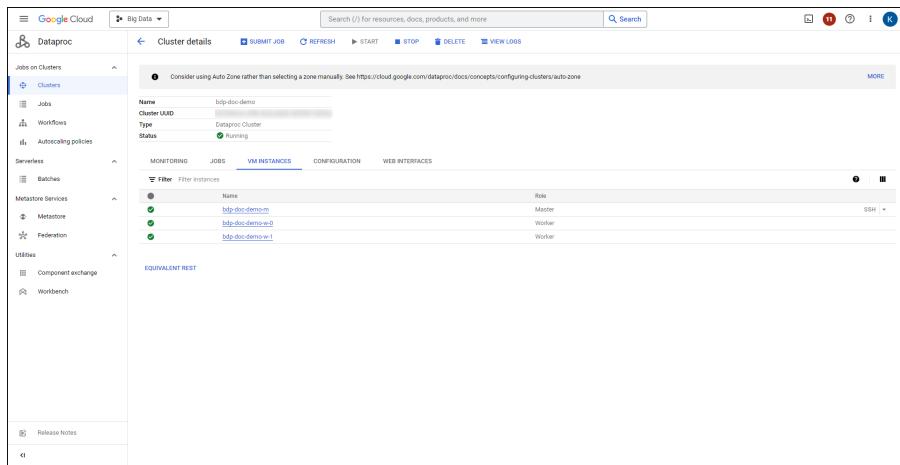


Figure 12-277: Nodes in the Cluster

23. To verify the configuration of the cluster, click the **CONFIGURATION** tab.

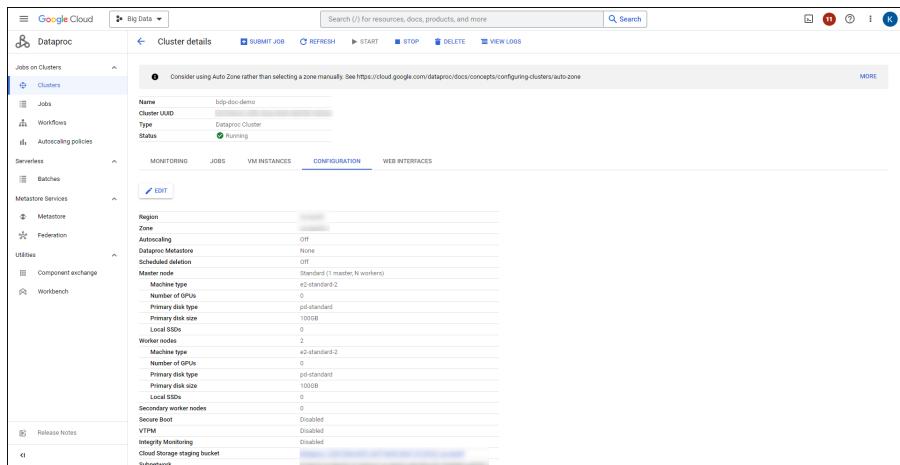


Figure 12-278: Cluster Configuration Page

You can also create a new Dataproc cluster and install the Big Data Protector on the nodes using the following command on the Command Line Interface:

```
BDP_INSTALLER_PATH='gs://<bucket_name>/<build/bdp_bootstrap_installer.sh'
gcloud dataproc --region ${REGION} clusters create ${CLUSTER_NAME} --subnet default
--zone ${ZONE} --master-machine-type n1-standard-2 --master-boot-disk-size 500 --num-
```

```
workers ${WORKER_COUNT} --worker-machine-type n1-standard-2 --worker-boot-disk-size 500 --
project warm-scout-184122 --initialization-actions ${BDP_INSTALLER_PATH}
```

where:

<bucket_name>: specifies the storage bucket path that contains the Big Data Protector bootstrap installer script

<CLUSTER_NAME>: specifies the name of the new Dataproc cluster

Note: For more information about the CLI parameters, refer to [GCP documentation](#).

12.6.6.8 Uninstalling the Big Data Protector

The uninstallation of the Big Data Protector is not supported when a new Dataproc cluster is created. However, when the workload on the nodes in the Dataproc cluster is reduced and the nodes are decommissioned automatically or manually removed, then the decommissioned or removed nodes are not available for use.

Note: Before you delete the nodes from the cluster, ensure that you save your data to the required storage bucket as a standard practice to avoid any data loss.

12.6.7 Installing Big Data Protector on an Azure HDInsight Cluster

The Microsoft Azure platform offers cloud-based computing services, which include computing services, virtual machines, data storage, analytics, networking services, identity and access management (IAM), and so on.

Azure HDInsight is a Hadoop service that allows you to run Hadoop clusters, which are located in several data centres, which are organized by geographical regions, across the world.

You can run Big Data Protector 7.2.0 on HDInsight to efficiently run analytics and scale in a cost-effective manner.

Note: The HDInsight platform supports Azure Blob Storage and Azure Data Lake Storage (ADLS) locations for storing data.

This section considers Azure Blob Storage as reference for storing the Big Data Protector installer files, which are needed for the Script action feature of HDInsight.

We do not support ADLS storage for storing the Big Data Protector installer files.

This section describes the tasks that you need to perform for installing Big Data Protector on a new HDInsight cluster.

12.6.7.1 Verifying Prerequisites for Installing Big Data Protector

Ensure that the following prerequisites are met, before installing Big Data Protector on an HDInsight cluster:

- It is recommended to be familiar with the following parts:
 - The Azure HDInsight environment
 - Storage locations, used to store the Big Data Protector installation files
 - Script Actions, used to invoke the installation of Big Data Protector
 - Azure Virtual Network
- An ESA 7.2.1 is installed and running either on-premise or on the cloud.
- Ensure that the settings for the Virtual Network are accurate so that the ESA is accessible to all the nodes in the cluster.



- The following ports are configured on the ESA and the nodes in the HDInsight cluster, which will run Big Data Protector:
 - ESA: Ensure that the Big Data Protector nodes can communicate with the ESA on the port *8443*.
 - Big Data Protector nodes: Ensure that on each node (localhost), the port *16700* is open and not allocated to any other process.
 - Proxy node: If a proxy node is used between the ESA and Big Data Protector nodes, then the port *8443* needs to be open on the proxy node.
- An Azure Blob storage location and the required access key for authenticating and accessing the storage location is available.

12.6.7.2 Creating a Storage Location on the Azure Platform

If you are installing Big Data Protector on a new HDInsight cluster, then you must create a storage location to copy the Big Data Protector installation files, which are created using the Configurator script.

The storage location provides an isolated space for storing the files that are required to install Big Data Protector on the new HDInsight cluster. These files are then used by the Script Action feature of the HDInsight environment when setting up a new cluster. Even if the HDInsight cluster is terminated, the storage location continues to retain the files required to install Big Data Protector.

For more information about creating a storage location, refer to the Azure documentation for [Creating an Azure Blob Storage Location](#).

After a storage location is created, an Access Key for authenticating and accessing the location is required.

For more information about Access Key, refer to [Azure Key Vault Storage Account Keys](#).

Note: The installation of Big Data Protector from ADLS Storage is not supported.

Note: If you are using Big Data Protector on an HDInsight cluster, then HDFSFP is not supported.

12.6.7.3 Downloading the Big Data Protector Package

After receiving the Big Data Protector installation package from Protegility, copy it to any HDInsight instance or node that has ESA connectivity.

12.6.7.4 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to update the Big Data Protector installation package, on all the nodes in the HDInsight cluster.

► To extract the Big Data Protector Configurator file from the installation package:

- Login to the CLI on a machine or an HDInsight node that has ESA connectivity.
- Copy the Big Data Protector package *BigDataProtector_Ubuntu-16-64_x86-64_HDInsight-x.x-64_9.0.0.0.x.tgz* to a directory, such as the */opt/bigdata* directory
- Extract the *BDPConfigurator_HDInsight.sh* file from the Big Data Protector installation package using the following command.
`tar -xvf BigDataProtector_Ubuntu-16-64_x86-64_HDInsight-x.x-64_9.0.0.0.x.tgz`
- Press **ENTER**.

The following files are extracted:

- *BDPConfigurator_HDInsight.sh*
- *PTYProxy_Setup_<OS>_<arch>_9.0.0.0.x.tgz*

12.6.7.5 Setting up the Proxy

If you need to configure a proxy, which connects to the ESA on one end, and communicates with the HDInsight cluster nodes containing Big Data Protector on the other end, then perform the following task.

► To setup the Proxy for the HDInsight Cluster:

1. Copy the *PTYProxy_Setup_<OS>_<arch>_7.2.x.x.tgz* file to the HDInsight node, which will act as the proxy, and communicate with ESA and the nodes in the HDInsight cluster.
2. Extract the Proxy archive using the following command.
`tar -xvf PTYProxy_Setup_<OS>_<arch>_7.2.x.x.tgz`
3. Press **ENTER**.
The *ProxySetup_<OS>_<arch>_7.2.x.x.sh* file is extracted.
4. Setup the proxy on the required node using the following command.
`./ProxySetup_<OS>_<arch>_7.2.1.x.sh -esa host (-dir installation_directory) (-port proxy_port)`
5. Press **ENTER**.
The Proxy service is installed.
6. Navigate to the *<Proxy_Installation_directory>/Proxy/bin* directory.
7. Start the Proxy service using the following command.
`./proxyctrl start`
8. Press **ENTER**.
The Proxy service is started.

12.6.7.6 Running the Configurator Script

You need to run the configurator script to create the installation package for installing Big Data Protector on a new HDInsight cluster.

The configurator script performs the following tasks:

- Downloads the certificates from an ESA.
- Updates the Big Data Protector installation package for a new HDInsight cluster.
- Creates the Script Action and Big Data Protector classpath configurator scripts for the new HDInsight cluster.
- Uploads the Big Data Protector installation package, and Script Action and Big Data Protector classpath configurator scripts to the storage location.

► To run the Big Data Protector Configurator Script:

1. Run the *BDPConfigurator_HDInsight.sh* script from the folder where it is extracted using the following command.
`./BDPConfigurator_HDInsight.sh`

2. Press **ENTER**.

A prompt to continue the installation of Big Data Protector appears.

3. Type *yes* to continue the installation of Big Data Protector.

4. Press **ENTER**.

A prompt for the Blob Storage location to upload the Big Data Protector files appears.

5. Type the path of the Blob Storage location, which was created in the section *Creating a Storage Location on the Azure Platform*, in the following format.

`https://<storageaccountname>.blob.core.windows.net/<container_name>`

where:

`<storageaccountname>`: Name of the Storage Account on Azure

`<container_name>`: Name of the storage location

6. Press **ENTER**.

A prompt for the Access Key appears.

7. Type the Access Key for the storage location.

8. Press **ENTER**.

A prompt for the installation directory appears.

9. Enter the directory for installing Big Data Protector.

10. Press **ENTER**.

A prompt for the ESA Host or IP address appears.

11. If a proxy is setup, then enter the IP address of the proxy node.

- Alternatively, enter the ESA IP address.

12. Press **ENTER**.

A prompt for the ESA user name appears.

13. Enter the ESA user name.

14. Press **ENTER**.

A prompt for the ESA password appears.

15. Enter the ESA password.

16. Press **ENTER**.

The ESA Certificates are downloaded.

The Big Data Protector configurator script updates the installation package to include the Certificates downloaded from the ESA and uploads the following files to the storage location, which was created in the section *Creating a Storage Location on the Azure Platform*:

- `BigDataProtector_Ubuntu-16-64_x86-64-nCPU_HDInsight-x.x-64_7.2.1.x.tgz`: The Big Data Protector installation package for a new HDInsight cluster.
- `bdp_script_action.sh`: The action script for installing Big Data Protector on a new HDInsight cluster.
- `bdp_classpath_configurator.py`: The script containing the required classpath settings for running Big Data Protector on the HDInsight cluster.

17. After all the Big Data Protector files are uploaded to the storage location, on the Azure screen, refresh the contents of the storage location.

The storage location lists the Big Data Protector files contained in it.

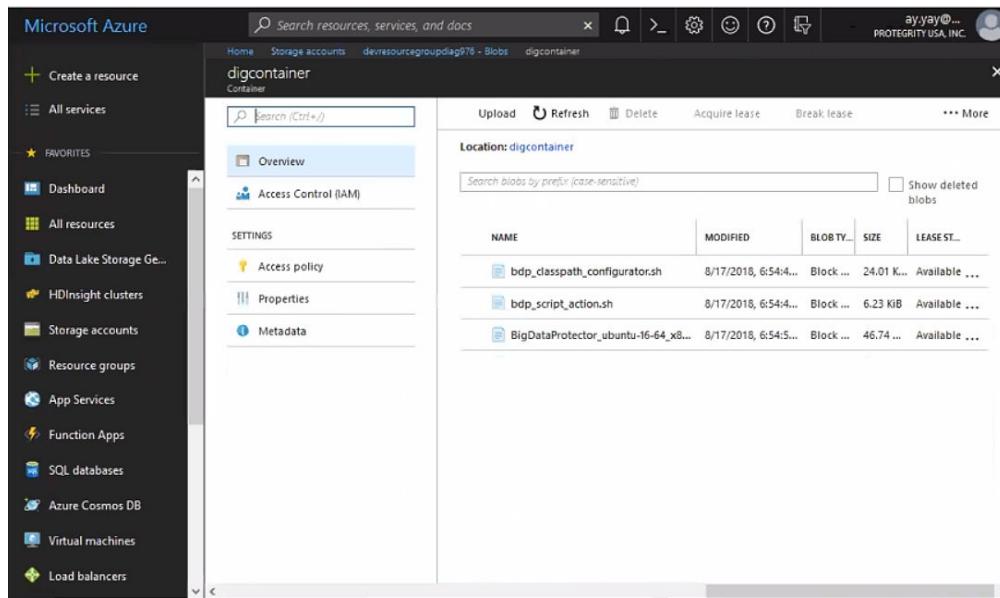


Figure 12-279: Storage Location with the Big Data Protector files on the Azure screen

12.6.7.7 Installing Big Data Protector on a New HDInsight Cluster

Perform the following steps to create an HDInsight cluster on Azure and install Big Data Protector on the required nodes in the HDInsight cluster.

► To install Big Data Protector on a New HDInsight Cluster:

1. Navigate to the Azure HDInsight home page.
2. Login to the Azure dashboard after providing the credentials.
After successful authentication, the Azure dashboard appears.

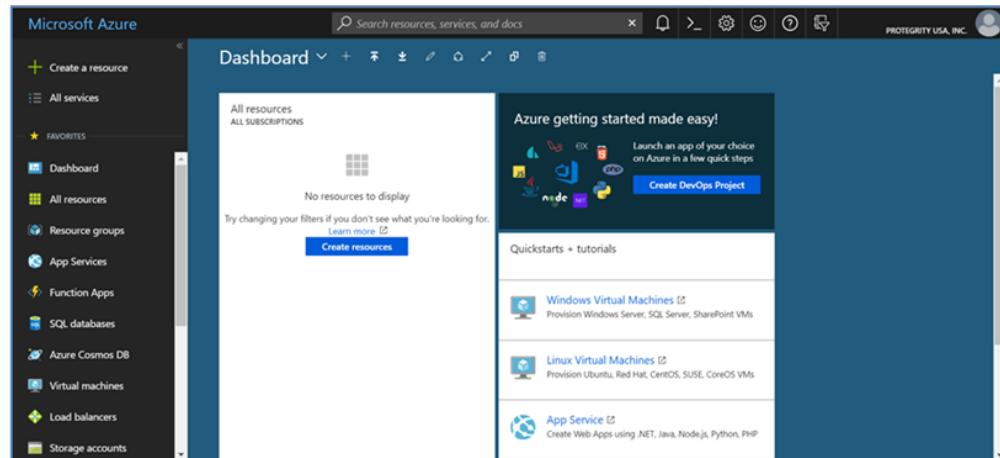


Figure 12-280: Azure Dashboard screen

3. On the Azure Dashboard screen, click **All services** on the left pane.
The *All services* pane appears.

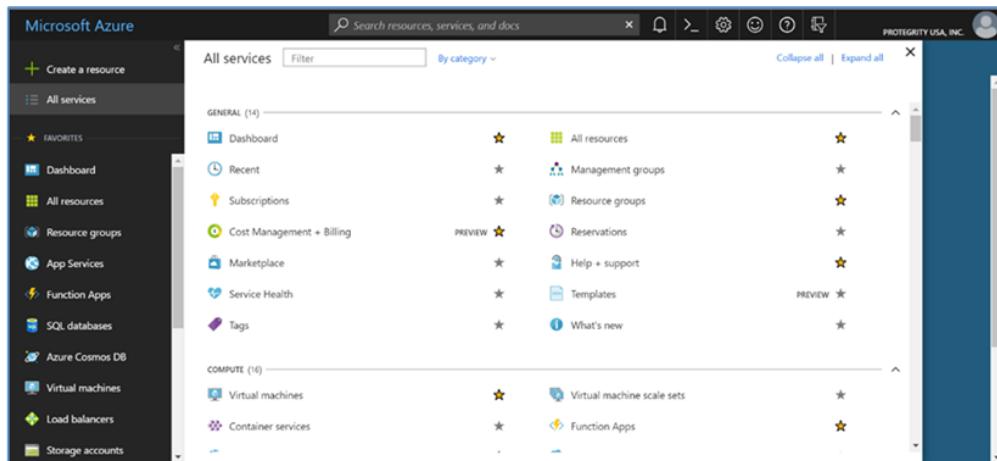


Figure 12-281: All Services Pane

- Click **HDInsight clusters** under the *ANALYTICS* section.

The *HDInsight clusters* pane appears.

The screenshot shows the 'HDInsight clusters' pane. The left sidebar has the same structure as Figure 12-281. The main pane displays a table with three items, each representing an HDInsight cluster. The columns are NAME, RESOURCE GROUP, LOCATION, and SUBSCRIPTION. The clusters listed are 'clusspark' (Resource Group: Dev-ResourceGroup, Location: East US, Subscription: Azure Cloud Platform), 'smsazure' (Resource Group: Dev-ResourceGroup, Location: East US, Subscription: Azure Cloud Platform), and 'spark-hadoop' (Resource Group: east-us-2-development-rg, Location: East US 2, Subscription: Azure Cloud Platform). There are also buttons for Add, Edit columns, Refresh, and Assign tags.

Figure 12-282: HDInsight Clusters Pane

- Click **Add**.

The *HDInsight* pane appears with the *Basics* option, selected by default, to configure basic settings of the HDInsight cluster.

The screenshot shows the 'HDInsight' pane with the 'Basics' tab selected. The left sidebar shows the 'HDInsight clusters' list is empty. The main pane has a 'Quick create' section with three steps: 1. Basics (Configure basic settings), 2. Storage (Set storage settings), and 3. Summary (Confirm configurations). Step 1 is currently active. It includes fields for Cluster name (with placeholder 'Enter new cluster name' and suffix '.azurehdinsight.net'), Subscription (set to 'Azure Cloud Platform'), Cluster type (set to 'Loading subscription capability...'), Cluster login username ('admin'), Cluster login password (''), Secure Shell (SSH) username ('sshuser'), and a checkbox for 'Use same password as cluster login'. Below these are options for Resource group ('Create new') or 'Use existing'. A note at the bottom says 'This cluster may take up to 20 minutes to create.' A 'Next' button is at the bottom right.

Figure 12-283: HDInsight pane with Basics option selected

- On the *HDInsight* pane, click **Custom (size, settings, apps)**.
- On the *Basics* pane, type the name of the cluster in the *Cluster name* text box.
- Select the required type of cluster in the *Cluster type* field.

The *Cluster configuration* pane appears.

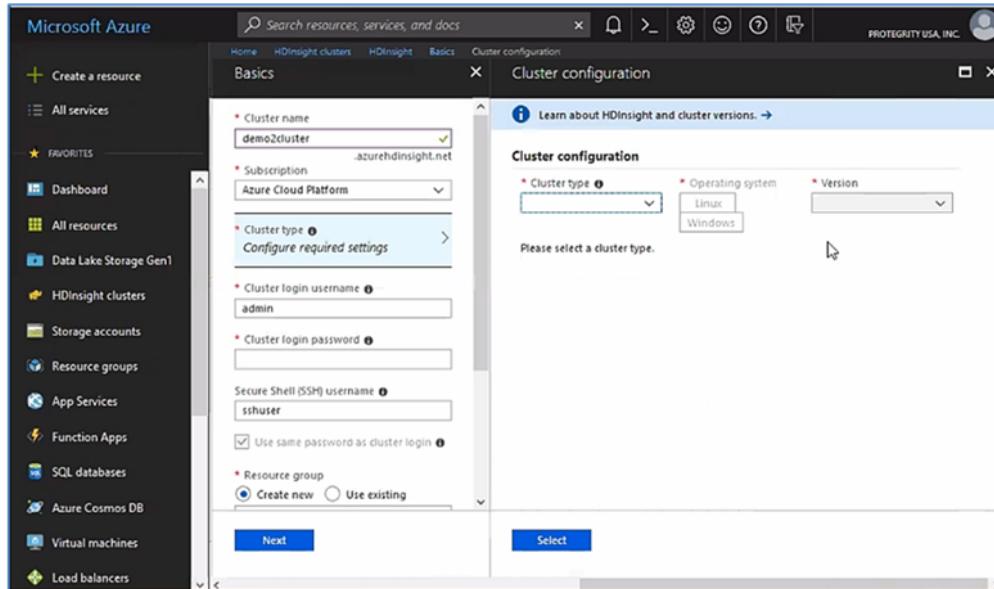


Figure 12-284: Cluster configuration pane

- On the *Cluster configuration* pane, in the **Cluster type** drop down, select *Hadoop*.

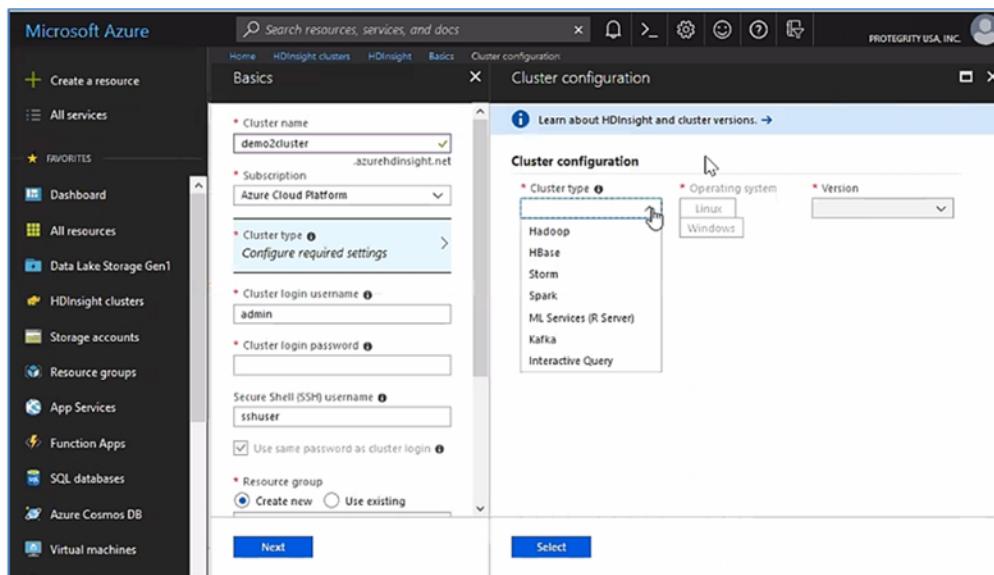


Figure 12-285: Cluster type drop down

- On the *Cluster configuration* pane, the *Operating system* is selected, and the *Version* drop down lists the version of Hadoop.

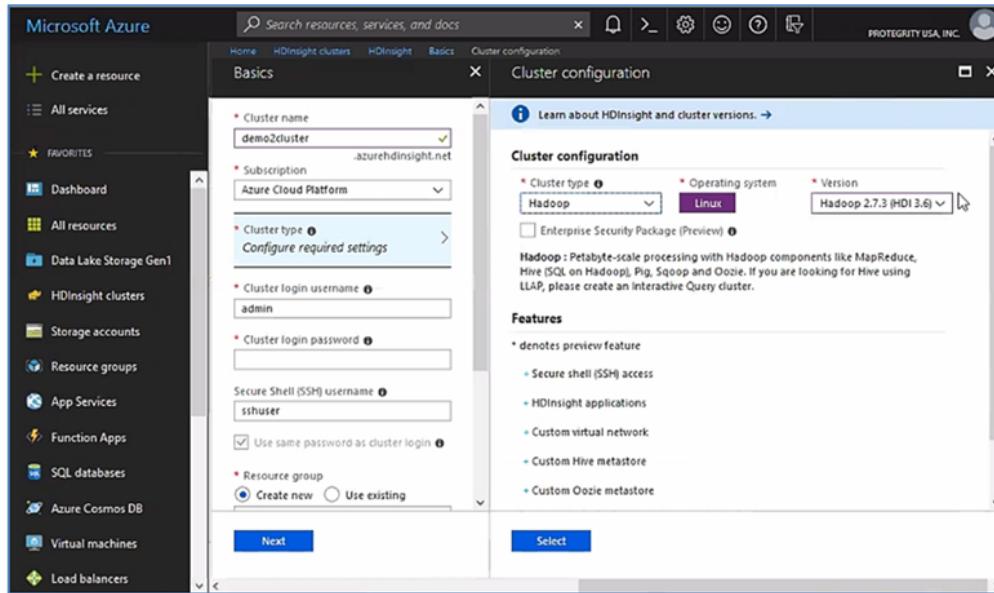


Figure 12-286: Cluster configuration details

11. If required, you can select the required version of Hadoop from the *Version* drop down.
12. On the Cluster configuration pane, click **Select**.

The *Basics* pane appears with the *Cluster type* field populated.

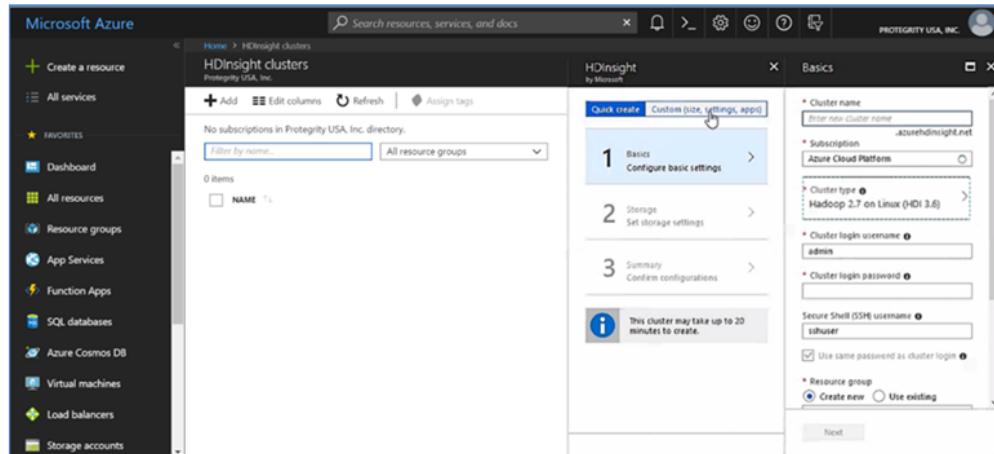


Figure 12-287: HDInsight pane with the Cluster type option selected

13. Enter the user name for logging in to the cluster in the *Cluster login username* text box.
14. Enter the password for logging in to the cluster in the *Cluster login password* text box.
15. Enter the user that will connect to the cluster using SSH in the *Secure Shell (SSH) username* text box.
16. Depending on the requirements, if you are creating a new Resource group, then select the *Create new* option. Alternatively, if you are using an existing Resource group, then select the *Use existing* option.
17. Depending on the requirements, select the preferred location for the HDInsight cluster from the *Location* drop down.
18. On the *Basics* pane, click **Next**.

The *HDInsight* pane appears with the *Storage* option selected, to configure the storage settings for the HDInsight cluster.

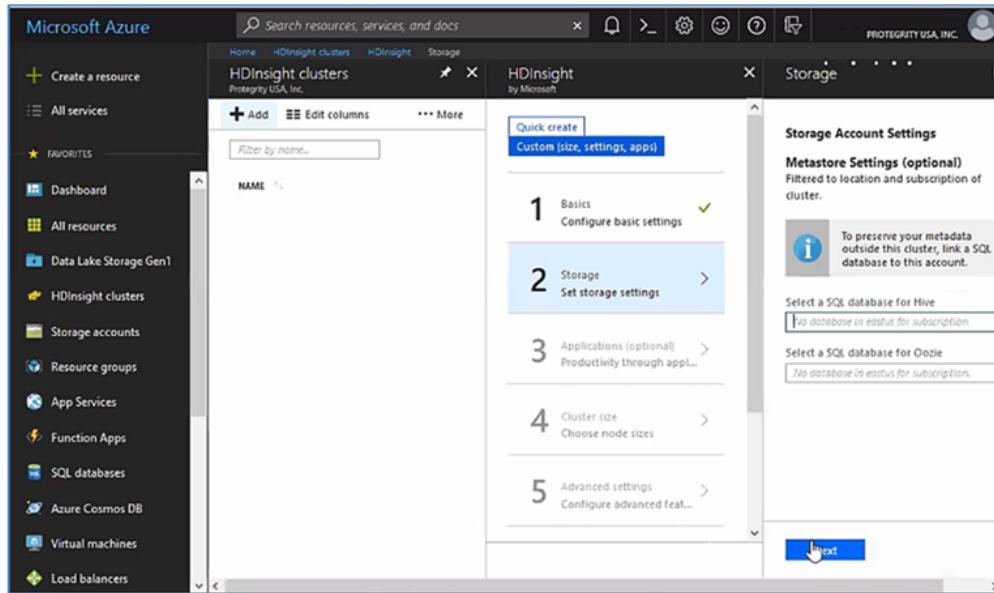


Figure 12-288: HDInsight pane with Storage option selected

- If an Azure storage location is configured for the HDInsight cluster, then the **Storage Account Settings** are populated in the **Storage** pane.

The *Storage* pane is updated to display the storage location mapped for the HDInsight cluster.

In this case, the *Primary storage type* selected is **Azure Storage**, which is an Azure Blob storage location.

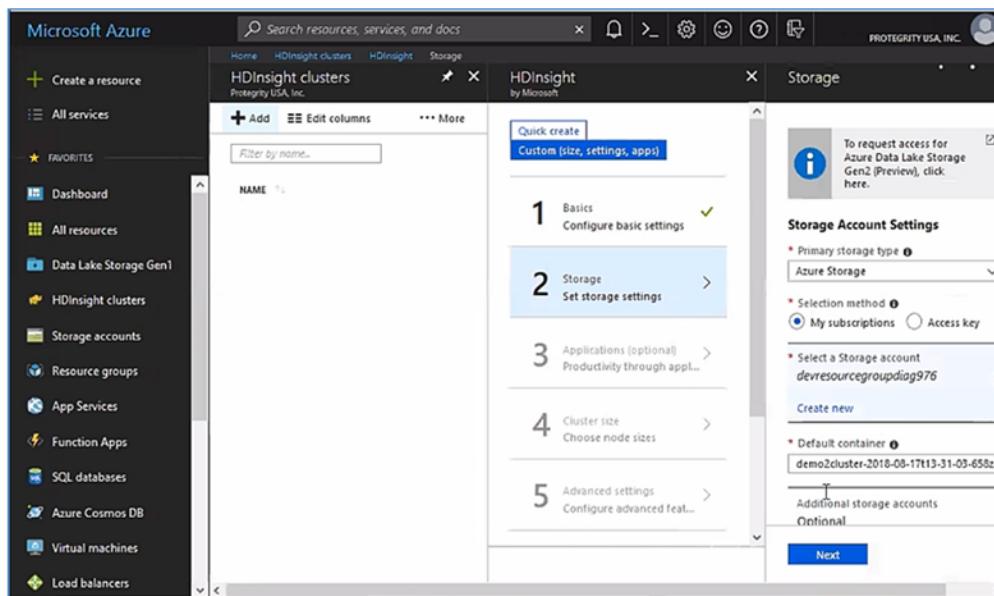


Figure 12-289: HDInsight pane with Storage pane populated

- Depending on the requirements, select the storage account to be used in the *Select a Storage account* field.
- If you have selected the Azure Blob storage location account, then in the *Default container* field, the path of the Big Data Protector configurator script is automatically populated.
- On the *Storage* pane, click **Next**.

The *HDInsight* pane appears with the *Applications (optional)* option selected, to configure some predefined applications for the HDInsight cluster.

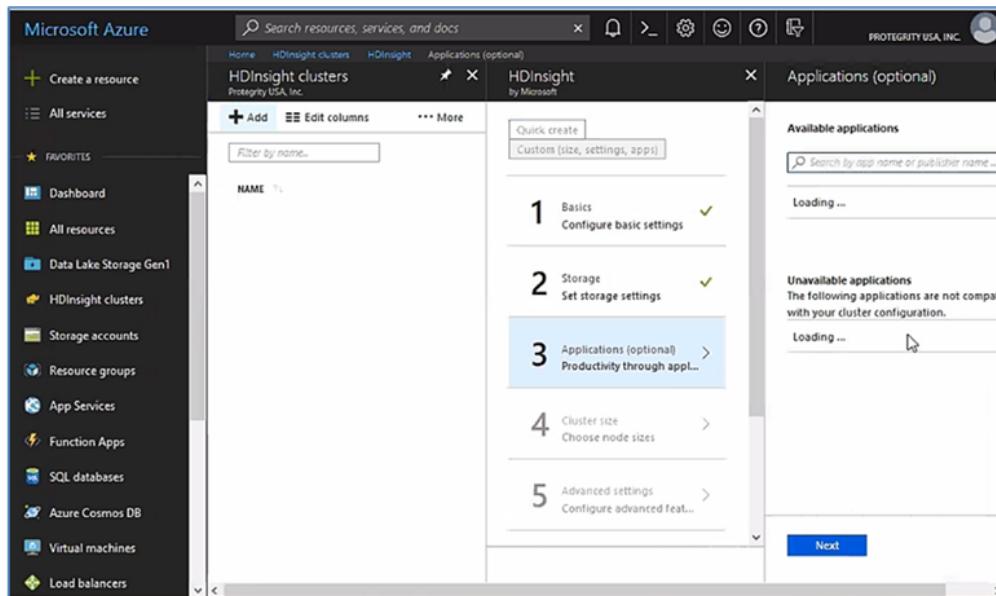


Figure 12-290: HDInsight pane with Applications (optional) option selected

23. If you need to use any predefined applications for the HDInsight cluster, then select the same in the *Applications (optional)* pane.
24. On the *Applications (optional)* pane, click **Next**.

The *HDInsight* pane appears with the *Cluster size* option selected, to configure the sizing parameters for the HDInsight cluster.

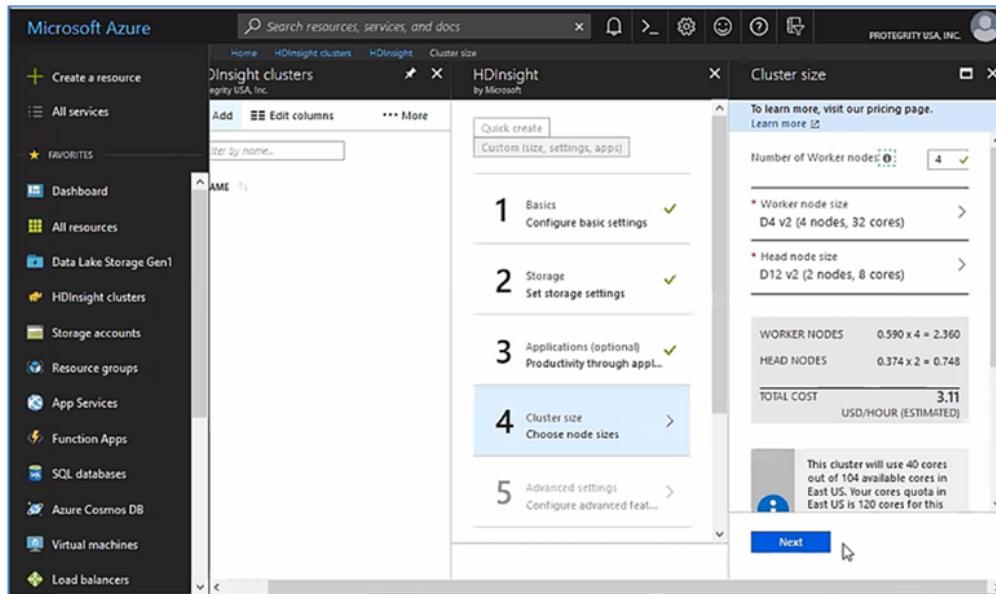


Figure 12-291: HDInsight pane with Cluster size option selected

25. On the *Cluster size* pane, depending on the requirements, select the required number of worker nodes in the *Number of Worker nodes* drop down.
26. Depending on the requirements, select the configuration of the worker and head nodes, which are similar to Master nodes, in the *Worker node size* and *Head node size* fields respectively.
27. On the *Cluster size* pane, click **Next**.

The *HDInsight* pane appears with the *Advanced settings* option selected, to configure advanced settings, such as script actions, virtual network, and subnet settings for the HDInsight cluster.

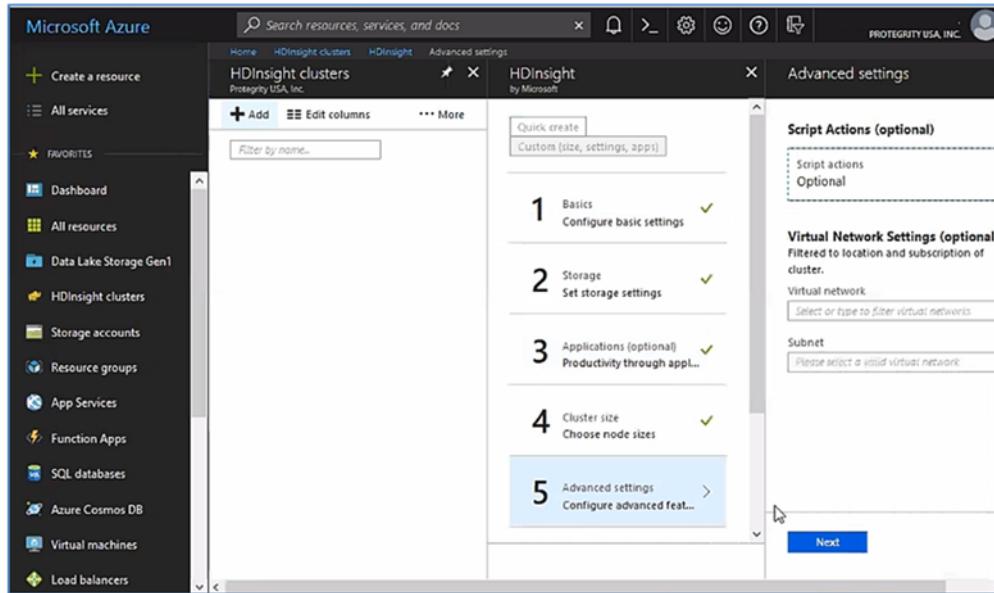


Figure 12-292: HDInsight pane with Advanced settings option selected

28. On the *Advanced settings* pane, click **Script actions**.

The *Script actions* pane appears.

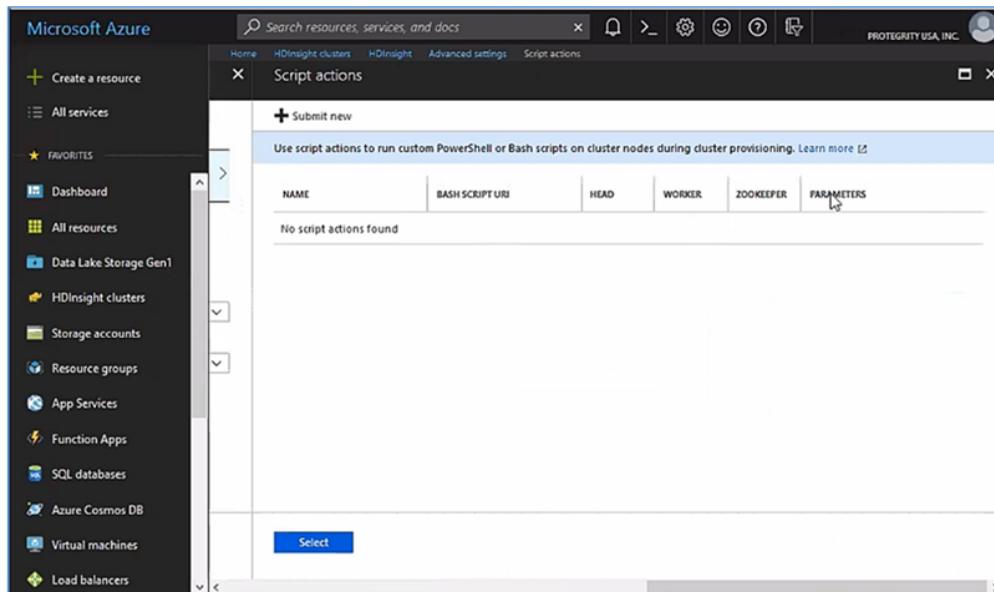


Figure 12-293: Script actions pane

29. On the *Script actions* pane, click **Submit new**.

The *Submit script action* pane appears.

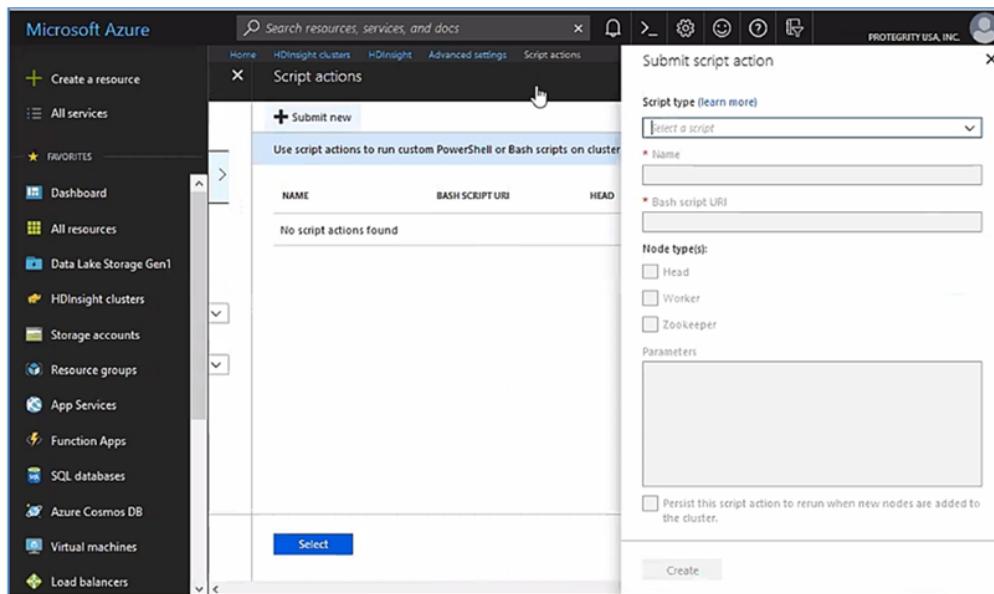


Figure 12-294: Submit script action pane

30. On the *Submit script action* pane, in the **Script type** drop down, select **Custom**.
31. Enter the name of the script in the *Name* text box.
32. Enter the Big Data Protector Script Action URI, which will install Big Data Protector on the HDInsight cluster nodes, including the storage location, in the *Bash script URI* text box, in the following format.

`https://<storageaccountname>.blob.core.windows.net/<Container_name>/
bdp_script_action.sh`

In this case, the path of the Big Data Protector configurator script, that is located in the storage location, which was created in the section *Creating a Storage Location on the Azure Platform*, is populated in the *Bash script URI* text box.

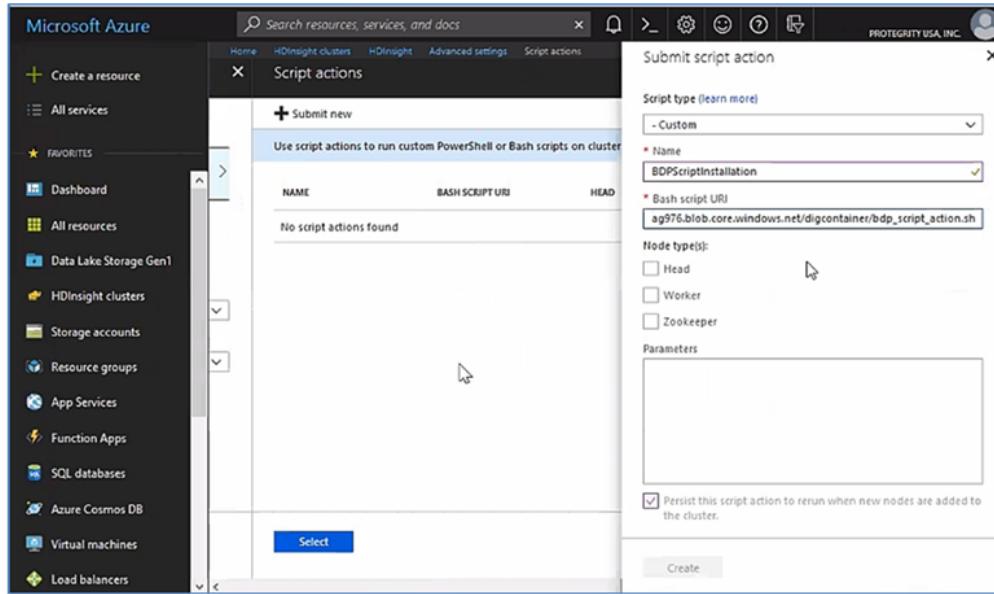


Figure 12-295: Submit script action pane populated

33. In the *Node type(s)* options, select the *Head* and *Worker* node check boxes to run the Big Data Protector script action on the head and worker nodes respectively.
34. Click **Create**.

The *Script actions* pane is updated to display the Big Data Protector script action.

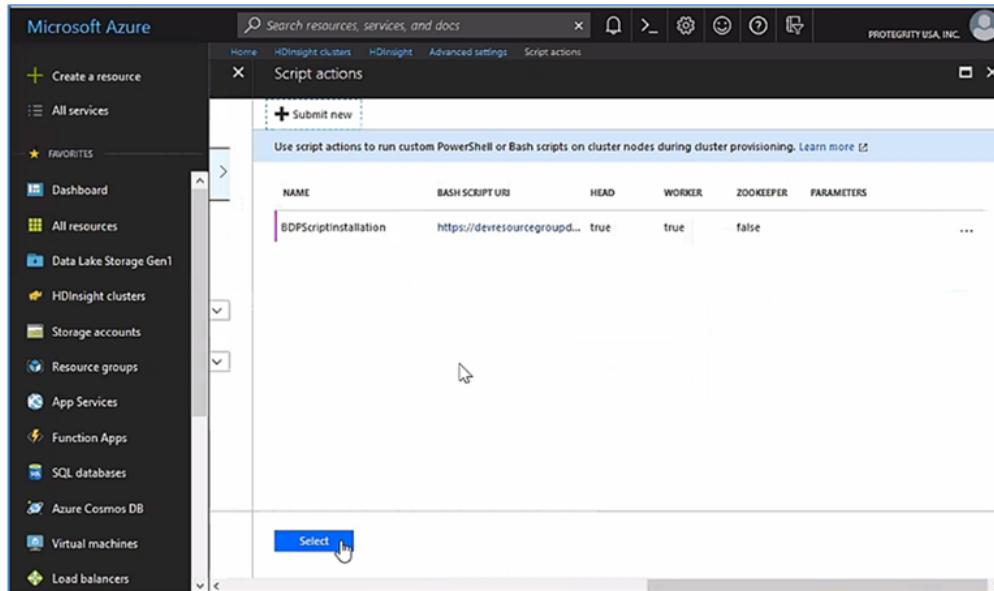


Figure 12-296: Script actions pane populated

35. On the *Script actions* pane, click **Select**.

The HDInsight pane appears with the *Advanced settings* option selected, and the **Script Actions ()** section populated for the HDInsight cluster.

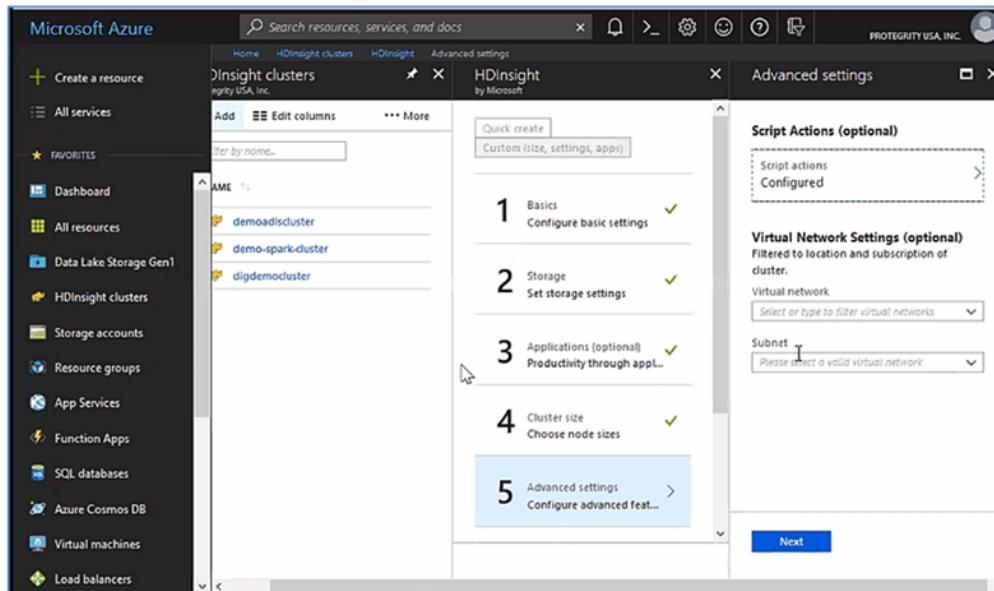


Figure 12-297: Script actions pane populated

36. Depending on the requirements, select the virtual network for the HDInsight cluster in the *Virtual network* drop down.
37. Depending on the requirements, select the subnet for the HDInsight cluster in the *Subnet* drop down.
38. On the *Advanced settings* pane, click **Next**.

The *Cluster summary* pane appears listing a summary of the HDInsight cluster.

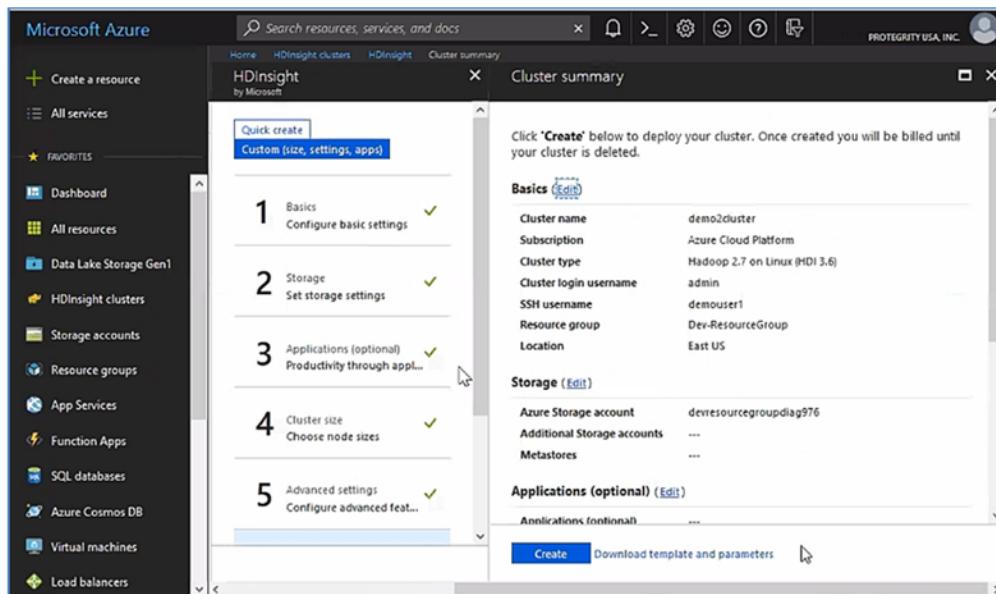


Figure 12-298: Cluster summary pane

39. After you have confirmed the parameters of the HDInsight cluster, click **Create**.

After the nodes in the cluster are provisioned, the *HDInsight clusters* pane is populated with the active cluster that is available for use.

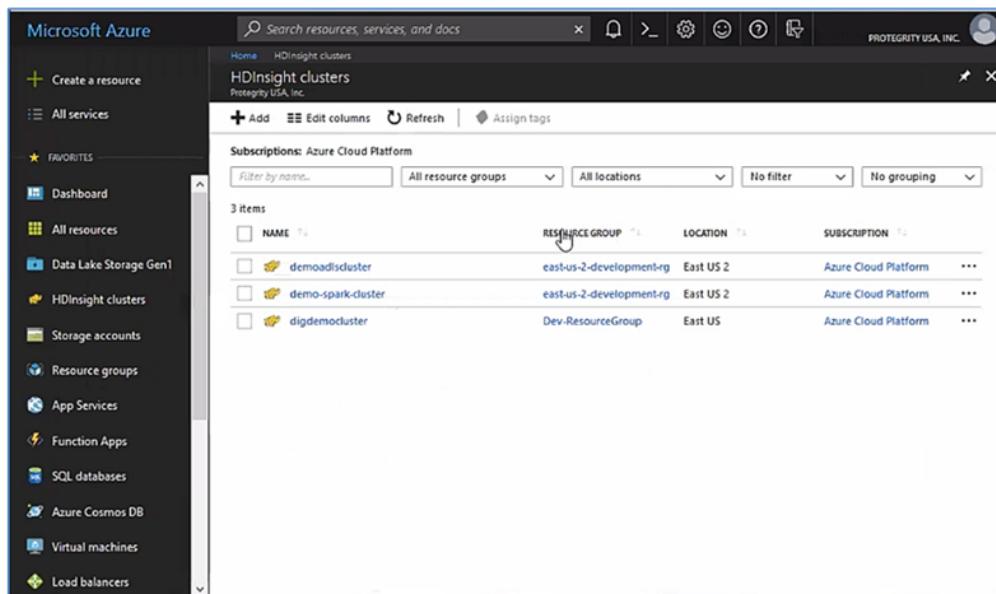


Figure 12-299: Cluster is Ready on the HDInsight clusters screen

12.6.7.8 Installing Big Data Protector on an Existing HDInsight Cluster

Perform the following steps to install Big Data Protector on the required nodes in an existing HDInsight cluster.

► To install Big Data Protector on an Existing HDInsight Cluster:

1. Navigate to the Azure HDInsight home page.
2. Login to the Azure dashboard after providing the credentials.

After successful authentication, the Azure dashboard appears.

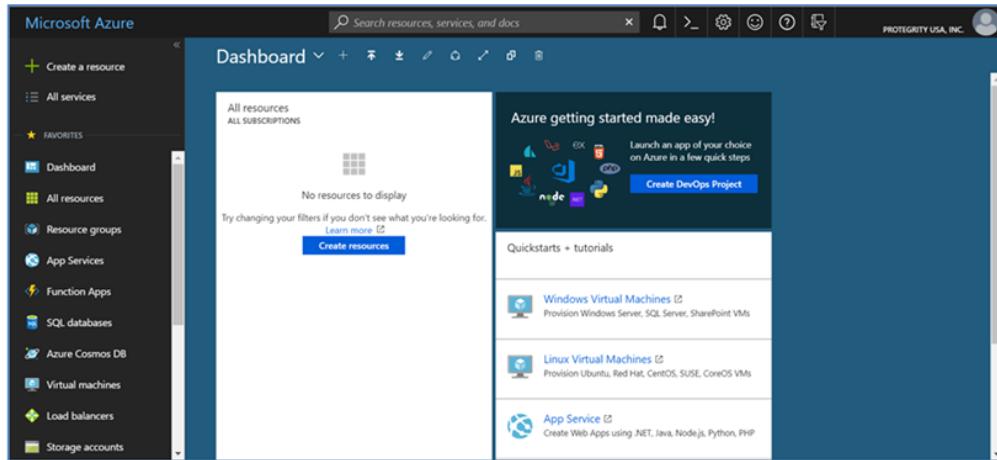


Figure 12-300: Azure Dashboard screen

3. On the Azure Dashboard screen, click **All services** on the left pane.
- The *All services* pane appears.

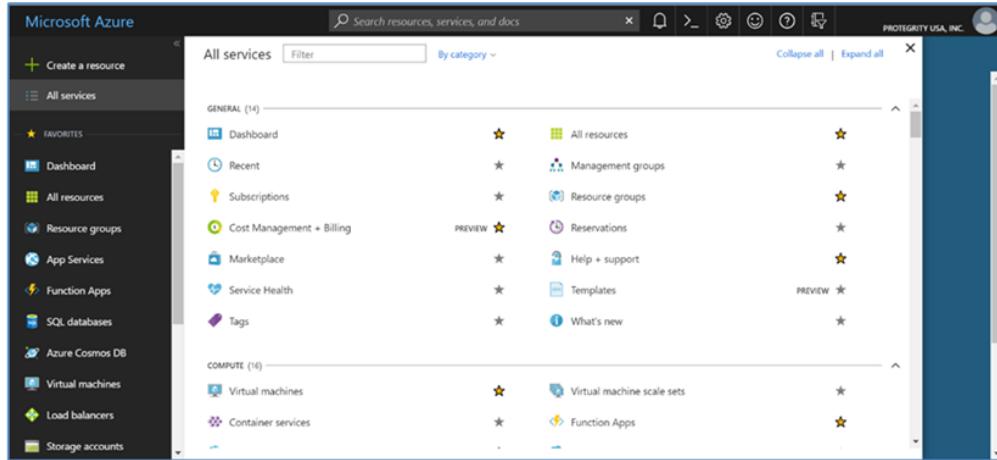


Figure 12-301: All services pane

4. Click **HDInsight clusters** under the *ANALYTICS* section.
- The *HDInsight clusters* pane appears listing the existing clusters.

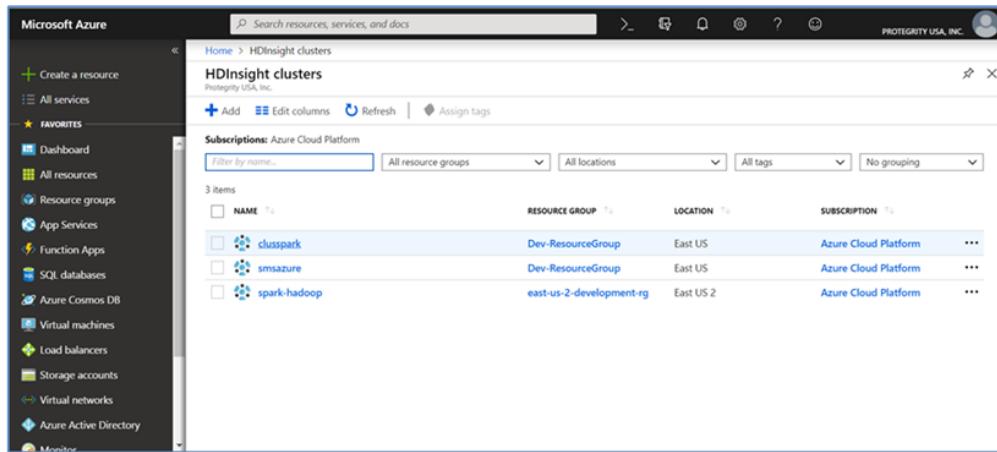


Figure 12-302: HDInsight clusters pane

5. Click an existing HDInsight cluster to install Big Data Protector.

The cluster pane for the selected cluster appears.

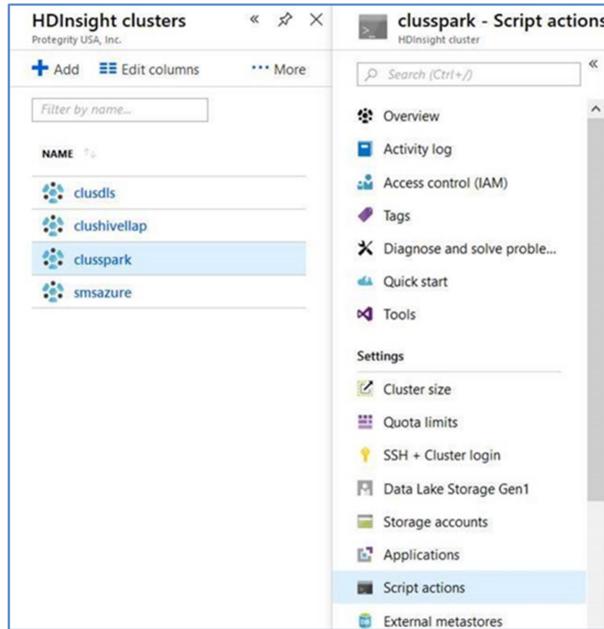


Figure 12-303: Selected cluster pane

6. Click **Script actions** in the selected cluster pane.

The *Script actions* pane appears.

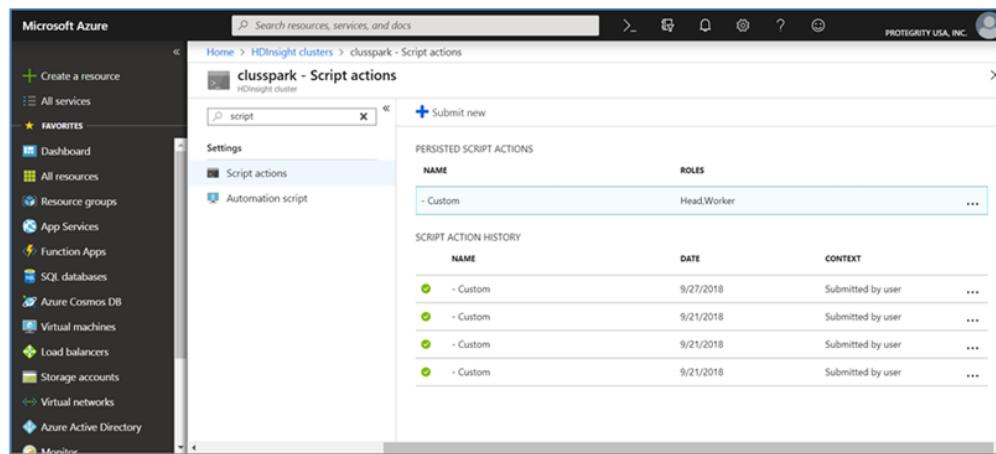


Figure 12-304: Script actions pane

7. Click **Submit new**.

The *Submit script action* pane appears.

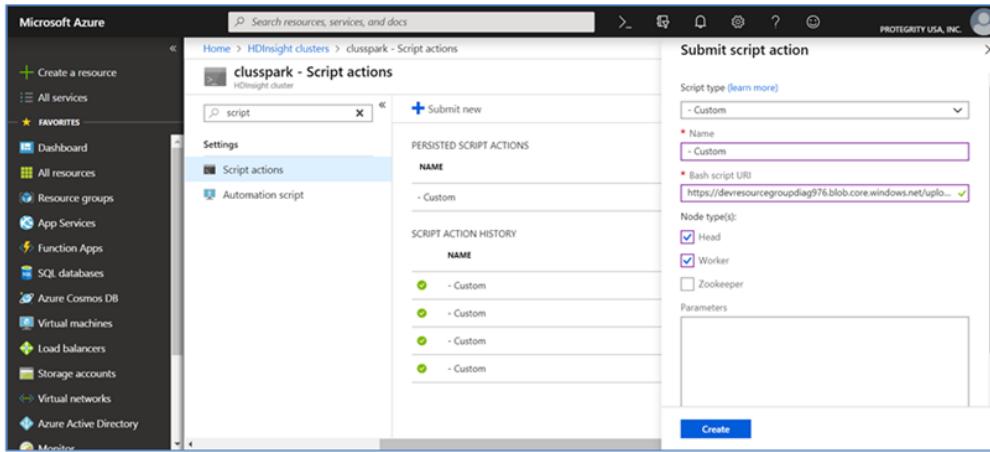


Figure 12-305: Submit script action pane

8. In the **Script type** drop down, select **Custom**.
9. Enter the name of the script in the *Name* text box.
10. Enter the Big Data Protector Script Action URI, which will install Big Data Protector on the HDInsight cluster nodes, including the storage location, in the *Bash script URI* text box, in the following format.

`https://<storageaccountname>.blob.core.windows.net/<Container_name>/bdp_script_action.sh`

In this case, the path of the Big Data Protector configurator script, that is located in the storage location, which was created in the section *Creating a Storage Location on the Azure Platform*, is populated in the *Bash script URI* text box.

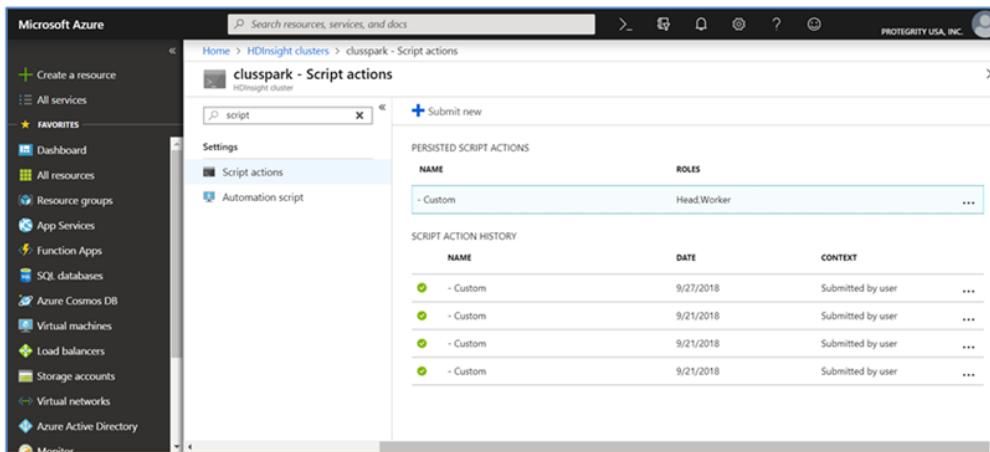


Figure 12-306: Submit script action pane populated

11. In the *Node type(s)* options, select the *Head* and *Worker* node check boxes to run the Big Data Protector script action on the head and worker nodes respectively.
12. Click **Create**.

The *Script actions* pane is updated to display the Big Data Protector script action and Big Data Protector is installed on the required nodes in the HDInsight cluster.

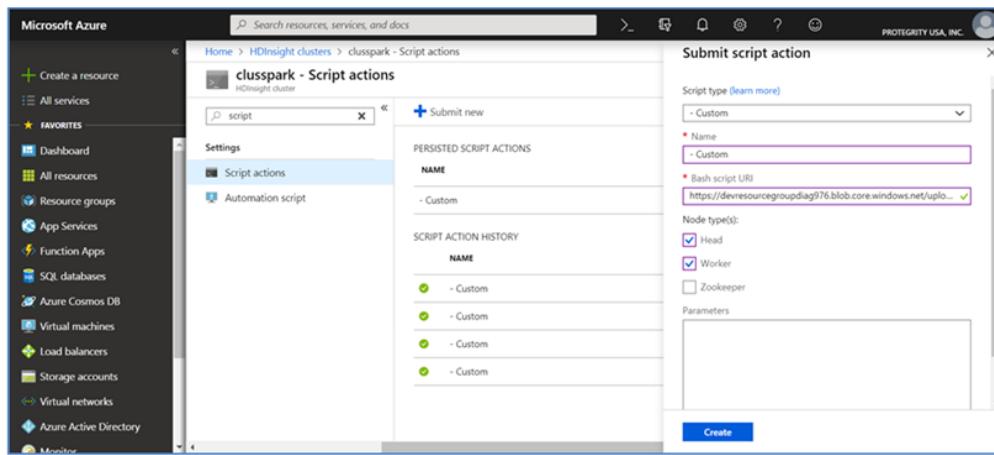


Figure 12-307: Script actions pane populated

12.6.7.9 Scaling and Shrinking of the Nodes in the Cluster

Depending on the requirements for increasing or decreasing the number of nodes in the cluster, nodes are added or removed from a cluster, either through triggers based on workload, or CPU utilization, and so on, or manually by the users.

Note: Ensure that the Storage Location, that contains the Script Action and installation files for Big Data Protector is available and accessible to the nodes in the cluster. In addition, ensure that the location of the Script Action file, that is present in the Storage Location is specified for the HDInsight cluster nodes.

If additional nodes are being provisioned for the HDInsight cluster, then Big Data Protector is installed on the new nodes from the Storage Location.

12.6.7.10 Best Practices for Using Big Data Protector on the HDInsights Cluster

The following table lists some best practices for using Big Data Protector on the HDInsight cluster.

Table 12-61: Best Practices for using Big Data Protector on the HDInsight Cluster

| Practice | Description |
|--|--|
| Ensuring ESA Connectivity | Ensure that the Virtual Network that the ESA resides on is accessible to the all the nodes in the HDInsight cluster. |
| Verifying the installation of Big Data Protector | Use an SSH-based utility to connect to the nodes in the HDInsight cluster to verify if the Big Data Protector services are operational. |
| Setting a Minimum Cluster Configuration | Ensure that you start with a cost-effective cluster configuration and scale it based on your requirements. |
| Differing the Subnets for the Machine and ESA | If you are using a machine that is connected to the ESA, then ensure that the machine and the ESA are on different subnets. |
| Naming the Cluster | Ensure that you create clusters with unique names, which includes the initial six characters of the cluster name being unique. If the initial six characters in the cluster names are not unique across clusters, then you might be unable to use the clusters. |

12.6.7.11 Uninstalling Big Data Protector

The uninstallation of Big Data Protector is not supported when a new Azure HDInsight cluster is created. However, when the workload on the nodes in the Azure HDInsight cluster is reduced and the nodes are decommissioned automatically or manually removed, then the decommissioned or removed nodes are not available for use.

Caution: Before you delete nodes from the cluster, ensure that you save your data to the required storage location, as standard practice, to avoid any data loss.

12.6.8 Installing the Big Data Protector on an AWS Databricks Cluster using the Amazon S3 Bucket

This section describes the steps for installing the Big Data Hive and Spark Protectors on the AWS Databricks platform using the S3 bucket provided by Amazon Simple Storage Service.

Warning: You will be unable to use the initialization scripts on DBFS, starting 01-December-2023. Databricks is ending support for initialization scripts on DBFS and the feature will not function after that date.

For more information, refer to <https://docs.databricks.com/en/init-scripts/legacy-global.html>.

12.6.8.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using the S3 Bucket

Ensure that the following prerequisites are met, before installing the Big Data Protector on the AWS Databricks platform:

- You have a valid AWS account.
- You have access to the AWS Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, installed, configured, and running, and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-62: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |



| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|---|--|
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |

- You must have an AWS S3 bucket.
- You must generate any one of the following types of AWS Access keys that have read/write permissions on the S3 bucket:
 - Permanent IAM User Access keys (*AWS Access Key ID*, *AWS Secret Access Key*)

Note: For more information about generating the permanent IAM user access keys, refer to https://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_access-keys.html.

- Temporary Security Credentials (*AWS Access Key ID*, *AWS Secret Access Key*, *AWS Session Token*)

Note: For more information about generating the temporary security credentials, refer to https://docs.aws.amazon.com/IAM/latest/UserGuide/id_credentials_temp.html.

- You have a Databricks-backed secret scope.

Note: For more information about creating a Databricks-backed secret scope, refer to <https://docs.databricks.com/en/security/secrets/secret-scopes.html#create-a-databricks-backed-secret-scope>.

- You have added the Access keys as Databricks secrets to the Databricks-backed secret scope.

Note: For more information about added the access keys as Databricks secret to the Databricks-backed secret scope, refer to <https://docs.databricks.com/en/security/secrets/secrets.html#create-a-secret-in-a-databricks-backed-scope>.

- You have created an AWS instance profile with read permissions on the S3 bucket and added it to your AWS Databricks Workspace.

Note:

For more information about creating an AWS instance profile with read permissions on the S3 bucket and adding it to your AWS Databricks workspace, refer to the following links:

- https://docs.aws.amazon.com/IAM/latest/UserGuide/id_roles_use_switch-role-ec2_instance-profiles.html
- <https://docs.databricks.com/en/aws/iam/instance-profile-tutorial.html>

12.6.8.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks

This section describes a brief overview of the installation steps that you must perform to install the Big Data Protector on the AWS Databricks platform using the Amazon Simple Storage Service.

The following diagram represents the installation workflow of the Big Data Protector on the AWS Databricks platform using the Amazon S3 bucket.

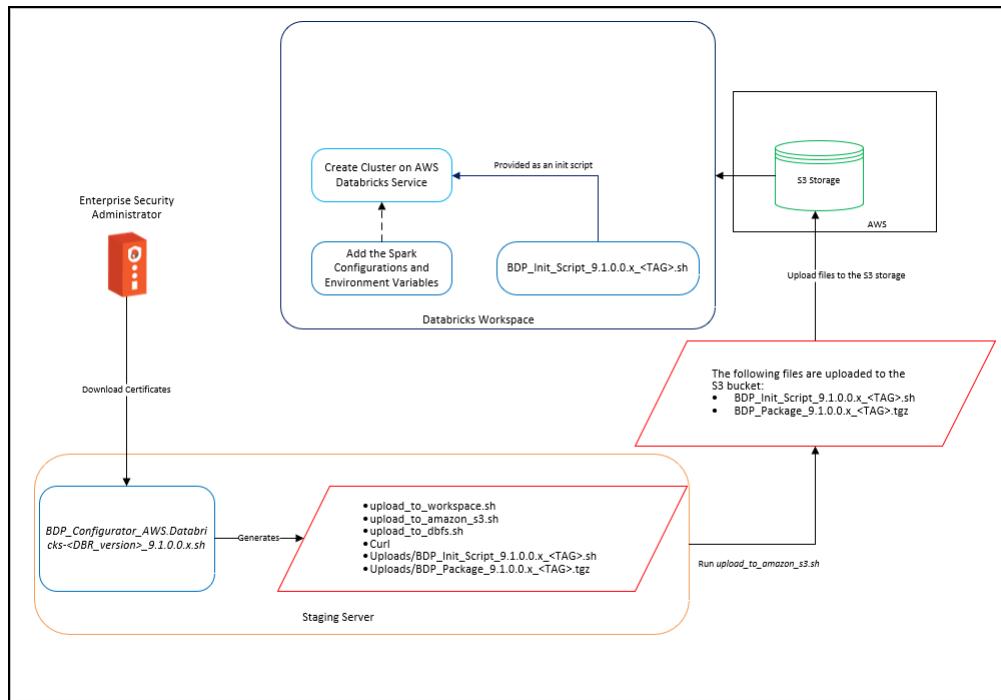


Figure 12-308: Installation Workflow

1. On a staging machine, perform the following steps.

- a. Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- b. Run the `BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh` script to download the certificates from the ESA and create the following files in the `/Installation_Files`/directory:

- `upload_to_workspace.sh`
- `upload_to_amazon_s3.sh`
- `upload_to_dbfs.sh`
- `curl`
- `Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh`
- `Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz`

Note: For more information about running the `BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh` script, refer to section [Running the Configurator Script](#).

- c. Run the `upload_to_amazon_s3.sh` script to upload the following files to the `S3` storage:

- `BDP_Init_Script_9.1.0.0.x_<TAG>.sh`
- `BDP_Package_9.1.0.0.x_<TAG>.tgz`

Note: For more information about running the `upload_to_amazon_s3.sh` script, refer to section [Uploading the Installation Files to the Amazon S3 Bucket](#).

2. On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* during cluster creation with the Spark configuration and the environment variables.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector](#).

12.6.8.3 Creating an S3 Bucket on AWS

If you want to upload the Big Data Protector installation files to an Amazon S3 bucket, then you must create an S3 bucket to store the Big Data Protector installation files, which are created using the Configurator script.

The S3 bucket provides an isolated space for storing the files that are required to install the Big Data Protector on the Databricks cluster. These files are then utilized by the Databricks workspace while provisioning the new cluster.

Note: For more information about creating an S3 bucket, refer to the Amazon documentation for [Creating an S3 bucket](#).

12.6.8.4 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because that IP address or hostname is written by the Configurator script in the *pepper.cfg* file.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the local file system.
3. To extract the *BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.
The command extracts the *BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh* file.

12.6.8.5 Running the Configurator Script

You need to run the Big Data Protector configurator script to download certificates from the ESA, and create the installation files for the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command.

```
./BDP_Configurator_AWS.Databricks-<DBR_version>_9.1.0.0.x.sh
```

2. Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS Databricks

Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Amazon Simple Storage Service (Amazon S3), or Databricks Workspace.
This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

Enter a Unique Alpha-Numeric Tag:

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

11. Enter the listening port for the ESA.

12. Press ENTER.



The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

```
Enter ESA Username:
```

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current | | | | | |
|---------|------------|---------|---------------|------|-------|-------|---------|---------|---------|---------|---------|-------|
| Dload | Upload | Total | Spent | Left | Speed | | | | | | | |
| 100 | 30720 | 100 | 30720 | 0 | 0 | 32370 | 0 | --::--- | --::--- | --::--- | --::--- | 32336 |

```
Please select the Audit Store type where Log Forwarder(s) should send logs to.
```

```
[ 1 ] : Protegility Audit Store
```

```
[ 2 ] : External Audit Store
```

```
[ 3 ] : Protegility Audit Store + External Audit Store
```

```
Enter the no.:
```

17. To select the Audit Store type, select any one of the following options:

Table 12-63: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting using the Protegility Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |



When you select option **2** or **3**, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

18. Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

19. To generate the logs for the PEP server in a file, type *yes*.

20. Press ENTER.

The installer generates the installation files as per the options you selected.

```
PepServer's log will be generated in a file.
```

```
Successfully generated Big Data Protector files for AWS Databricks in ./Installation_Files/
```

```
All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Amazon Simple Storage Service (Amazon S3), or Databricks Workspace.
```

For DBFS:

The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./Installation_Files/Uploads/ to DBFS root directory (dbfs:/)

Usage: ./upload_to_dbfs.sh

Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload failure.

For Amazon S3:

The ./Installation_Files/upload_to_amazon_s3.sh script can be used to upload the files under ./Installation_Files/Uploads/ to Amazon Simple Storage Service (Amazon S3).

Usage: ./upload_to_amazon_s3.sh

For Workspace:

The ./Installation_Files/upload_to_workspace.sh script can be used to upload the files under ./Installation_Files/Uploads/ to Databricks' Workspace.

Usage: ./upload_to_workspace.sh

Warning: The upload_to_workspace.sh script uses Workspace REST API that has a limitation of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload failure.

```
Installation_Files
  Uploads
    BDP_Init_Script_9.1.0.0.x_<TAG>.sh
    BDP_Package_9.1.0.0.x_<TAG>.tgz
  curl
  upload_to_amazon_s3.sh
  upload_to_dbfs.sh
  upload_to_workspace.sh
```

12.6.8.6 Modifying the *pepper.cfg* File

This section explains the process of modifying the *pepper.cfg* file. You will be unable to modify the *pepper.cfg* file when the cluster is running. You must extract the Big Data Protector archive generated by the configurator script, update the *pepper.cfg* file, re-package the installation files and then upload the updated Big Data Protector archive to the *S3* storage using the helper script.

Attention: Modifying the *pepper.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepper.cfg* file.

► To modify the *pepperserver.cfg* file:

1. To navigate to the *./Installation_Files/Uploads/* directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

2. To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the *pepperserver.cfg* file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepperserver.cfg* file.

6. Modify the *pepperserver.cfg* file according to your requirements.

7. Save the changes to the *pepperserver.cfg* file.

8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. To upload the updated Big Data Protector archive to *S3*, use the *upload_to_amazon_s3.sh* script.

Note: For more information about uploading the files to S3 using the script, refer to section [Uploading the Installation Files to the Amazon S3 Bucket](#).

12.6.8.7 Uploading the Installation Files to the Amazon S3 Bucket

You must upload the installation files generated by the configurator script to the Amazon Simple Storage Service or the S3 bucket.

► To upload the files using the helper script:

1. To execute the helper script, run the following command.

```
./upload_to_amazon_s3.sh
```

The prompt to continue with the upload appears.

```
*****
This Script will upload the files generated under ./Uploads/ to Amazon S3.
*****
```

Do you wish to continue? [yes or no]:

2. To continue with the upload, type *yes*.

3. Press ENTER.

The prompt to enter the storage location appears.

```
Execution of upload_to_amazon_s3.sh script is started.

Checking if all required files are present under ./Uploads/
BDP_Package*.tgz file is present under ./Uploads/
BDP_Init_Script*.sh file is present under ./Uploads/.

All required files are present under ./Uploads/.

Enter the Amazon S3 URI for uploading the files.
(S3 URI Format: s3://<bucket_name>/<optional_folder_structure>)
NOTE: Whitespaces in the path are not allowed.
Enter S3 URI:
```

4. Type the path of the S3 storage bucket.

Ensure that the path of the S3 storage bucket is in the following format:

```
s3://<bucket_name>/<folder_in_the_bucket>
```

where,

- <*bucket_name*> - specifies the name of the storage bucket.
- <*folder_in_the_bucket*> - specifies the directory within the bucket.

5. Press ENTER.

The prompt to select the type of AWS access keys appears.

```
Choose the AWS Access Keys Type from the following options:
[1] -> IAM User Access Keys (permanent AWS_ACCESS_KEY_ID and AWS_SECRET_ACCESS_KEY)
[2] -> Temporary Security Credentials (temporary AWS_ACCESS_KEY_ID,
AWS_SECRET_ACCESS_KEY, and AWS_SESSION_TOKEN)
[ 1 or 2 ]:
```

6. Depending on your setup, select any one of the following options.

Table 12-64: AWS Access Key Types

| Option | Description |
|--------|---|
| 1 | Prompts to enter the following permanent IAM user access keys: <ul style="list-style-type: none"> • <i>AWS_ACCESS_KEY_ID</i> • <i>AWS_SECRET_ACCESS_KEY</i> |
| 2 | Prompts to enter the following temporary security credentials: <ul style="list-style-type: none"> • <i>AWS_ACCESS_KEY_ID</i> • <i>AWS_SECRET_ACCESS_KEY</i> • <i>AWS_SESSION_TOKEN</i> |

7. Enter the required credentials.

Note: Depending on the option you select for the AWS access key type, the script will prompt for the corresponding security credentials.

8. Press ENTER.

The script uploads the following files to the S3 bucket:



- *BDP_Package_9.1.0.0.x_<TAG>.tgz*
- *BDP_Init_Script_9.1.0.0.x_<TAG>.sh*

```
Choose the AWS Access Keys Type from the following options:
[1] -> IAM User Access Keys (permanent AWS_ACCESS_KEY_ID and AWS_SECRET_ACCESS_KEY)
[2] -> Temporary Security Credentials (temporary AWS_ACCESS_KEY_ID,
AWS_SECRET_ACCESS_KEY, and AWS_SESSION_TOKEN)
[ 1 or 2 ]: 1
```

AWS Access Keys Type -> 1

Enter AWS Access Key ID:

Enter AWS Secret Access Key:

Retrieving Amazon S3 Bucket's AWS Region via REST API

Successfully retrieved Amazon S3 Bucket's AWS region: <region_name>

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Amazon S3.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Amazon S3.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Amazon S3.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Amazon S3.
```

Successfully uploaded all files to Amazon S3.

To use the uploaded BDP_Init_Script_9.1.0.0.x_<TAG>.sh file as Init Script in Databricks Cluster, follow the below steps:

1) Since "[1] IAM User Access Keys" was used to upload the files, create 2 secrets in a Databricks-backed secret scope for storing the "AWS_ACCESS_KEY_ID" and "AWS_SECRET_ACCESS_KEY".

2) Add the following environment variables in the "Environment variables" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}
here, replace <scope_name> with the secret scope name & <secret_name> with the
"AWS_ACCESS_KEY_ID" secret key.
```

```
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}
here, replace <scope_name> with the secret scope name & <secret_name> with the
"AWS_SECRET_ACCESS_KEY" secret key.
```

3) Add the following S3 URI in the "Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:
s3://<bucket_name>/<folder_in_the_bucket>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh

4) Set an AWS Instance Profile in the Databricks cluster page that has permissions to download the Init script from the above S3 URI.

The script will return the following output, if you select the *Temporary Security Credentials* option while uploading the files to the S3 bucket:

```
Choose the AWS Access Keys Type from the following options:
[1] -> IAM User Access Keys (permanent AWS_ACCESS_KEY_ID and AWS_SECRET_ACCESS_KEY)
[2] -> Temporary Security Credentials (temporary AWS_ACCESS_KEY_ID,
AWS_SECRET_ACCESS_KEY, and AWS_SESSION_TOKEN)
[ 1 or 2 ]: 2
```

AWS Access Keys Type -> 2

Enter AWS Access Key ID:

Enter AWS Secret Access Key:

```
Enter AWS Session Token:
```

```
Retrieving Amazon S3 Bucket's AWS Region via REST API
```

```
Successfully retrieved Amazon S3 Bucket's AWS region: <region_name>
```

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Amazon S3.  
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Amazon S3.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Amazon S3.  
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Amazon S3.
```

```
Successfully uploaded all files to Amazon S3.
```

To use the uploaded BDP_Init_Script_9.1.0.0.x_<TAG>.sh file as Init Script in Databricks Cluster, follow the below steps:

- 1) Since "[2] Temporary Security Credentials" was used to upload the files, create 3 secrets in a Databricks-backed secret scope for storing the "AWS_ACCESS_KEY_ID", "AWS_SECRET_ACCESS_KEY", and "AWS_SESSION_TOKEN".

- 2) Add the following environment variables in the "Environment variables" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}  
here, replace <scope_name> with the secret scope name & <secret_name> with the  
"AWS_ACCESS_KEY_ID" secret key.
```

```
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}  
here, replace <scope_name> with the secret scope name & <secret_name> with the  
"AWS_SECRET_ACCESS_KEY" secret key.
```

```
PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/<scope_name>/<secret_name>}}  
here, replace <scope_name> with the secret scope name & <secret_name> with the  
"AWS_SESSION_TOKEN" secret key.
```

- 3) Add the following S3 URI in the "Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:
s3://<bucket_name>/<folder_in_the_bucket>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh

- 4) Set an AWS Instance Profile in the Databricks cluster page that has permissions to download the Init script from the above S3 URI.

12.6.8.8 Installing the Big Data Hive and Spark Protector

This section explains the following two methods to install the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.8.8.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protectors by creating a new cluster in the AWS Databricks workspace.

Note: For more information about modifying the *pepserver.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/create.html>.
2. On the cluster creation page, click the **Advanced Options** tab.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value separated by a space.



Table 12-65: Spark Configurations

| Keys | Values |
|---------------------------------|---|
| spark.driver.extraJavaOptions | -Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm |
| spark.plugins | com.protegility.spark.PtyExecSparkPlugin |
| spark.executor.extraJavaOptions | -Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm |

The following image represents a sample Spark configuration for reference.

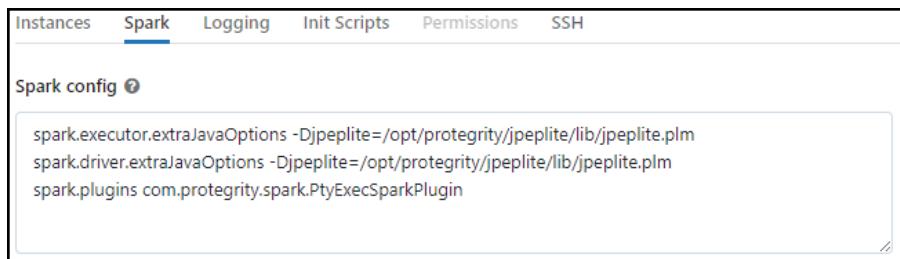


Figure 12-309: Spark Configurations

- If you select the *IAM User Access Keys* option while uploading the Big Data Protector installation files to the S3 bucket, then under **Environment variables**, add the following code snippet:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}
```

- For *PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}*, replace the *<scope_name>* with the secret scope name and the *<secret_name>* with the *AWS_ACCESS_KEY_ID* secret key.
- For *PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}*, replace the *<scope_name>* with the secret scope name and the *<secret_name>* with the *AWS_SECRET_ACCESS_KEY* secret key.

The following image represents a sample Spark Configuration and the environment variables for reference.

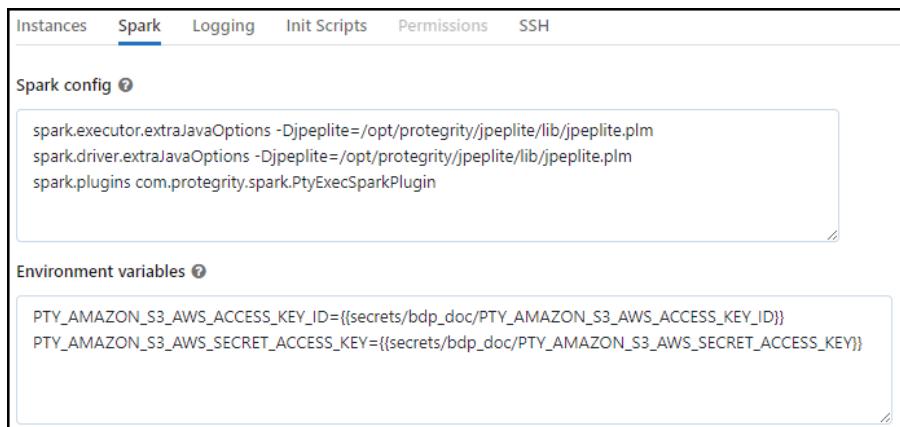


Figure 12-310: Spark Configuration and Environment Variables for IAM User Access Keys

- If you select the *Temporary Security Credentials* option while uploading the Big Data Protector installation files to the S3 bucket, then under **Environment variables**, add the following code snippet:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/<scope_name>/<secret_name>}}
```

- For *PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}*, replace the *<scope_name>* with the secret scope name and the *<secret_name>* with the *AWS_ACCESS_KEY_ID* secret key.

- For `PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}`, replace the `<scope_name>` with the secret scope name and the `<secret_name>` with the `AWS_SECRET_ACCESS_KEY` secret key.
- For `PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/<scope_name>/<secret_name>}}`, replace the `<scope_name>` with the secret scope name and the `<secret_name>` with the `AWS_SESSION_TOKEN` secret key.

The following image represents a sample Spark Configuration and the environment variables for reference.

The screenshot shows the 'Spark' tab selected in a navigation bar. Under 'Spark config', there are three lines of configuration code:

```
spark.executor.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.driver.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Under 'Environment variables', there are three environment variable definitions:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/bdp_doc/AWS_ACCESS_KEY_ID}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/bdp_doc/AWS_SECRET_ACCESS_KEY}}
PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/bdp_doc/AWS_SESSION_TOKEN}}
```

Figure 12-311: Spark Configurations and Environment Variables for Temporary Security Credentials

6. Click the **Init Scripts** tab.

The screenshot shows the 'Init Scripts' tab selected in a navigation bar. The 'Init scripts' section has a table with columns: Type, File path, and Region. A large '+' button is present above the table. Below the table, it says 'No init script set for this compute'. The 'Source' dropdown is set to 'S3'. The 'File path' input field contains 's3://'. The 'Region' dropdown is set to 'auto'. A note at the bottom states: 'The compute instance profile must have permission to read data from the S3 destination and it must include getObjectAcl permission. See the user guide for how to setup cluster instance profiles.'

Figure 12-312: The Init Scripts Tab

- From the **Source** list, select **S3**.
- In the **File Path** box, enter the location where you uploaded the initialization script to *S3* using the helper script. For example, enter the path for the initialization script as `s3://<storage_location>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh`.
- Click **Add**.

The location of the initialization script appears in the **File path** column.

The screenshot shows the 'Init Scripts' tab with one entry in the table:

| Type | File path | Region |
|------|--|--------|
| S3 | s3://bdp-doc/AWS_DBR/9-1-0-0-31/BDP_Init_Script_9.1.0.0.31_a1.sh | |

Figure 12-313: Initialization Script Path

- From the **Instance profile** list, select the instance profile (with read permissions to the *S3* bucket) that you added to the Databricks workspace.

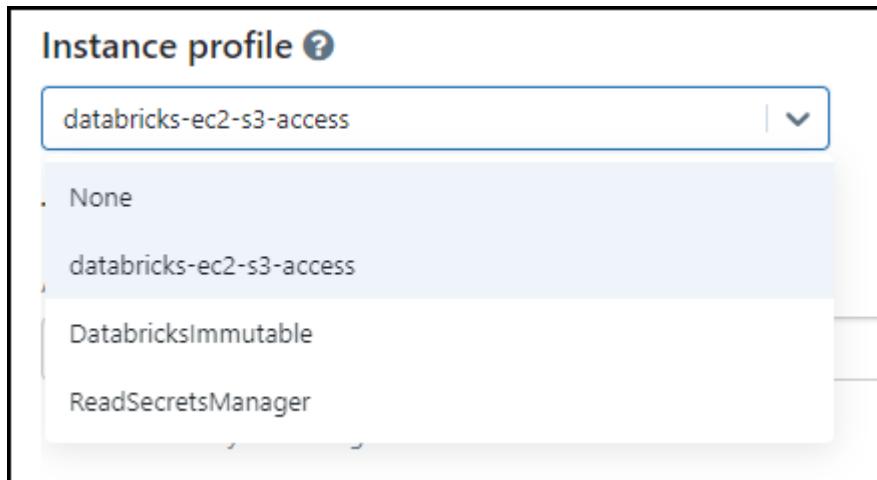


Figure 12-314: Selecting the Instance Profile

11. Click **Create compute**.

The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs, and the Spark APIs from Protegrity as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.8.8.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protectors on an existing cluster in the AWS Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/clusters-manage.html#edit-a-cluster>.
2. On the cluster configuration page, click the **Advanced Options** toggle.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format, separated by a space.

Table 12-66: Spark Configurations

| Keys | Values |
|--|--|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegrity.spark.PtyExecSparkPlugin</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</i> |

The following image represents a sample Spark configuration for reference.

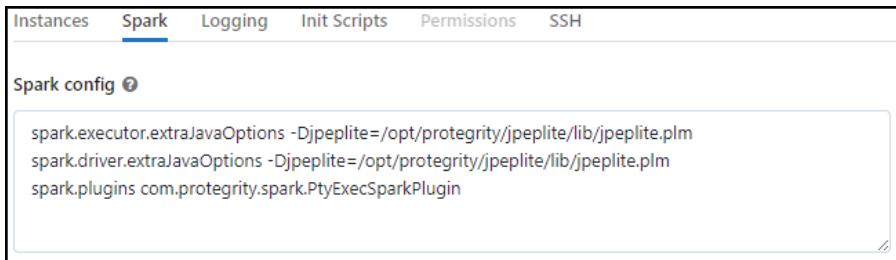


Figure 12-315: Spark Configurations

- If you select the *IAM User Access Keys* option while uploading the Big Data Protector installation files to the S3 bucket, then under **Environment variables**, add the following code snippet:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}
```

- For *PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}*, replace the <scope_name> with the secret scope name and the <secret_name> with the *AWS_ACCESS_KEY_ID* secret key.
- For *PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}*, replace the <scope_name> with the secret scope name and the <secret_name> with the *AWS_SECRET_ACCESS_KEY* secret key.

The following image represents a sample Spark configuration and the environment variables for reference.



Figure 12-316: Spark Configuration and Environment Variables for IAM User Access Keys

- If you select the *Temporary Security Credentials* option while uploading the Big Data Protector installation files to the S3 bucket, then under **Environment variables**, add the following code snippet:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}
PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/<scope_name>/<secret_name>}}
```

- For *PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/<scope_name>/<secret_name>}}*, replace the <scope_name> with the secret scope name and the <secret_name> with the *AWS_ACCESS_KEY_ID* secret key.
- For *PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/<scope_name>/<secret_name>}}*, replace the <scope_name> with the secret scope name and the <secret_name> with the *AWS_SECRET_ACCESS_KEY* secret key.
- For *PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/<scope_name>/<secret_name>}}*, replace the <scope_name> with the secret scope name and the <secret_name> with the *AWS_SESSION_TOKEN* secret key.

The following image represents a sample Spark configuration and the environment variables for reference.

The screenshot shows the 'Spark' tab of a configuration interface. Under 'Spark config', there are three lines of Java options:

```
spark.executor.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.driver.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Under 'Environment variables', there are three entries:

```
PTY_AMAZON_S3_AWS_ACCESS_KEY_ID={{secrets/bdp_doc/AWS_ACCESS_KEY_ID}}
PTY_AMAZON_S3_AWS_SECRET_ACCESS_KEY={{secrets/bdp_doc/AWS_SECRET_ACCESS_KEY}}
PTY_AMAZON_S3_AWS_SESSION_TOKEN={{secrets/bdp_doc/AWS_SESSION_TOKEN}}
```

Figure 12-317: Spark Configurations and Environment Variables for Temporary Security Credentials

- Click the **Init Scripts** tab.

The screenshot shows the 'Init Scripts' tab. It has columns for 'Type', 'File path', and 'Region'. A large '+' button is present. Below it, a message says 'No init script set for this compute'. At the bottom, there are dropdowns for 'Source' (set to 'S3'), 'File path' (containing 's3://'), and 'Region' (set to 'auto'). A note at the bottom states: 'The compute instance profile must have permission to read data from the S3 destination and it must include getObjectAcl permission. See the [user guide](#) for how to setup cluster instance profiles.'

Figure 12-318: The Init Scripts Tab

- From the **Source** list, select **S3**.
- In the **File Path** box, enter the location where you uploaded the initialization script to *S3* using the helper script.
For example, enter the path for the initialization script as *s3://<storage_location>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh*.
- Click **Add**.

The location of the initialization script appears in the **File path** column.

The screenshot shows the 'Init Scripts' tab with one entry in the table:

| Type | File path | Region |
|------|--|--------|
| S3 | s3://bdp-doc/AWS_DBR/9-1-0-0-31/BDP_Init_Script_9.1.0.0.31_a1.sh | |

Figure 12-319: Init Script Path

- From the **Instance profile** list, select the instance profile (with read permissions to the *S3* bucket) that you added to the Databricks workspace.

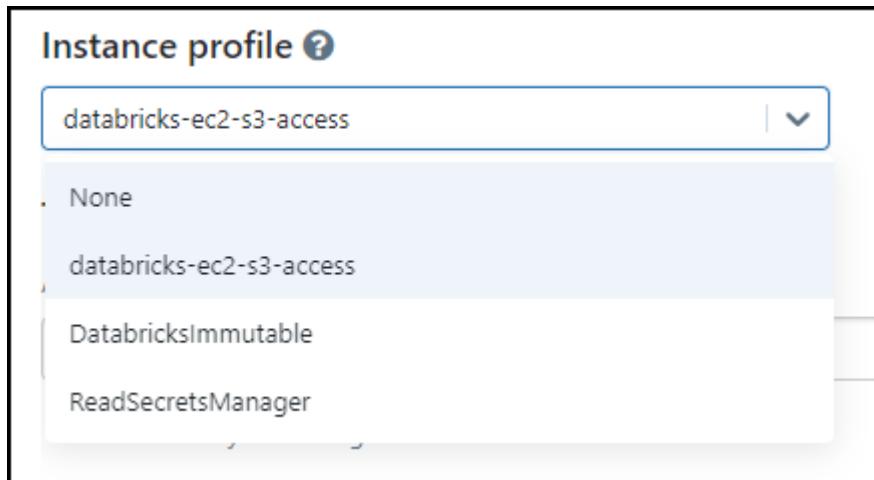


Figure 12-320: Selecting the Instance Profile

12. To restart the cluster, click **Confirm and restart**.

The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs, and the Spark APIs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.8.9 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs requires registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
```

```
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';

OK
```

When you register Hive UDFs permanently on the Unity Catalog, the registration fails when you set the default catalog and the database to custom Unity Catalog and a database inside it. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.

For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to enable the registration of permanent Hive UDFs in the Unity Catalog.

To enable the registration of permanent Hive UDFs in the Unity Catalog, perform the following steps.

1. Register UDFs permanently in the *hive_metastore* catalog.
2. Call the UDF in the SQL queries using the fully qualified name in the following format:

```
<catalog>.<database>.<UDF>
```

For example,

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
  'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
  'com.protegility.hive.udf.ptyUnprotectStr';
```

3. Press ENTER.

The registration of the UDFs completes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
  'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
  'com.protegility.hive.udf.ptyUnprotectStr';

OK
```

This approach protects the data in the tables stored under the Unity Catalog database by referring to them by their fully qualified name.



Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT
hive_metastore.default.ptyProtectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.9 Installing the Big Data Protector on an AWS Databricks Cluster using the Workspace

This section describes the steps for installing the Big Data Hive and Spark Protectors on the AWS Databricks platform using the Workspace storage.

Warning: You will be unable to use the initialization scripts on DBFS, starting 01-December-2023. Databricks is ending support for initialization scripts on DBFS and the feature will not function after that date.

For more information, refer to <https://docs.databricks.com/en/init-scripts/legacy-global.html>.

12.6.9.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using the Workspace

Ensure that the following prerequisites are met, before installing the Big Data Protector on the AWS Databricks platform:

- The user should have a valid AWS account.
- The user should have access to the AWS Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.



Table 12-67: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |

- You have a Databricks-backed secret scope.

Note: For more information about creating a Databricks-backed secret scope, refer to <https://docs.databricks.com/en/security/secrets/secret-scopes.html#create-a-databricks-backed-secret-scope>.

- You have created a Databricks personal access token.

Note: For more information about creating the Databricks personal access token, refer to <https://docs.databricks.com/en/dev-tools/auth.html#databricks-personal-access-token-authentication>.

- You have added the personal access token as Databricks secrets to the Databricks-backed secret scope.

Note: For more information about added the access keys as Databricks secret to the Databricks-backed secret scope, refer to <https://docs.databricks.com/en/security/secrets/secrets.html#create-a-secret-in-a-databricks-backed-scope>.

- You have created the Databricks workspace folder to upload the files.

Note: For more information about creating folders in the Workspace, refer to <https://docs.databricks.com/en/workspace/workspace-objects.html#folders>.

12.6.9.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks

This section describes a brief overview of the installation workflow that the user needs to perform to install the Big Data Protector on the AWS Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the AWS Databricks platform.

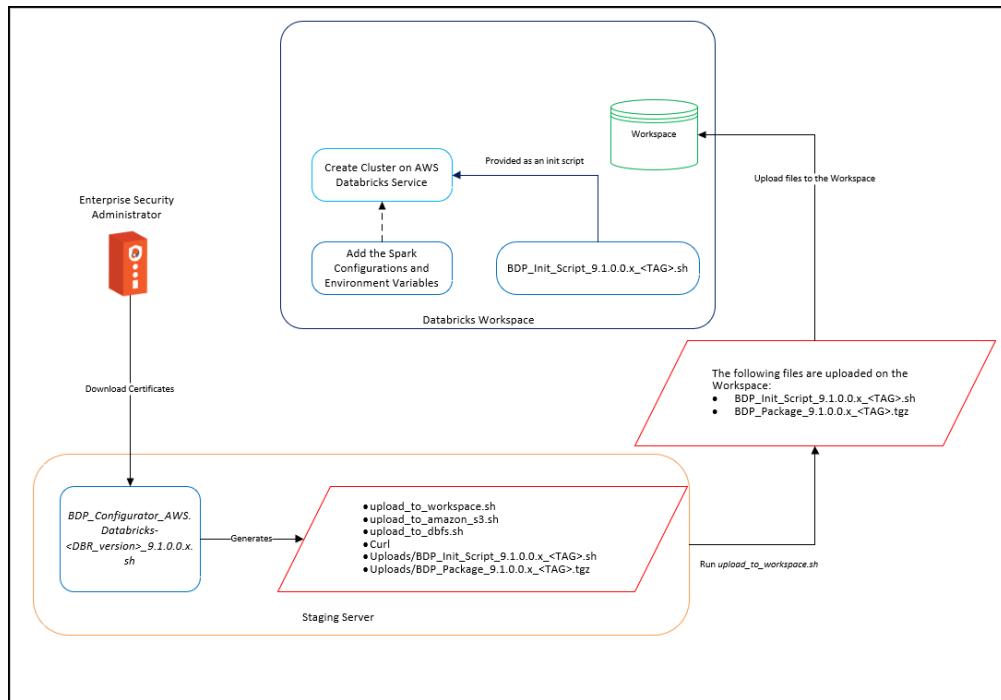


Figure 12-321: Installation Workflow

1. On a staging machine, perform the following steps.

- a. Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- b. Run the *BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* script to download the certificates from the ESA and create the following files in the */Installation_Files*/directory:

- *upload_to_workspace.sh*
- *upload_to_amazon_s3.sh*
- *upload_to_dbfs.sh*
- *curl*
- *Uploads/BDP_Init_Script_9.1.0.0.x-<TAG>.sh*
- *Uploads/BDP_Package_9.1.0.0.x-<TAG>.tgz*

Note: For more information about running the *BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* script, refer to section [Running the Configurator Script](#).

- c. Run the *upload_to_workspace.sh* script to upload the following files on the *workspace* storage:

- *BDP_Init_Script_9.1.0.0.x-<TAG>.sh*
- *BDP_Package_9.1.0.0.x-<TAG>.tgz*

Note: For more information about running the *upload_to_workspace.sh* script, refer to section [Uploading the Installation Files to the Workspace Storage](#).

- On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script during the cluster creation with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector](#).

12.6.9.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

► To extract the Big Data Protector Configurator file from the installation package:

- Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because that IP address or hostname is written by the Configurator script in the *pepserver.cfg* file.

- Download the *BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
- Extract the *BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* file from the Big Data Protector installation package using the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-
<DBR_version>-64_9.1.0.0.x.tgz
```

The command extracts the *BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* file.

12.6.9.4 Running the Configurator Script

You must run the Big Data Protector configurator script to download certificates from the ESA, and create the installation files for the Big Data Protector.

► To run the Big Data Protector Configurator Script:

- To execute the *BDP_Configurator_AWS.Databricks-<DBR_version>-9.1.0.0.x.sh* script from the directory where it is extracted, run the following command.

```
./BDP_Configurator_AWS.Databricks-<DBR_version>-9.1.0.0.x.sh
```

- Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
* This will create the Big Data Protector Installation files for AWS Databricks
*****
Do you wish to continue? [yes or no]:
```

- To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****  
This will create the Big Data Protector Installation files for AWS Databricks  
  
Do you wish to continue? [yes or no]: yes  
  
Big Data Protector service configurator started...  
  
Enter the directory path on Cluster nodes where you want to install Protegility products.  
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.  
[default:- /opt/protegility]:  
  
The BDP Installation files that will be generated needs to be uploaded to either  
DBFS root directory (dbfs:/), Amazon Simple Storage Service (Amazon S3), or Databricks  
Workspace.  
This script will append a unique tag to the generated files' names to distinguish them  
from other BDP files.  
  
Enter a Unique Alpha-Numeric Tag:
```

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

11. Enter the listening port for the ESA.

12. Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

```
Enter ESA Username:
```

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
  % Total    % Received % Xferd  Average Speed   Time     Time     Time  Current
               Dload  Upload   Total   Spent   Left  Speed
100  30720  100  30720      0       0  32370      0  --::--  --::--  --::--  32336
```

Please select the Audit Store type where Log Forwarder(s) should send logs to.

```
[ 1 ] : Protegility Audit Store
[ 2 ] : External Audit Store
[ 3 ] : Protegility Audit Store + External Audit Store
```

Enter the no.:

- To select the Audit Store type, select any one of the following options:

Table 12-68: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type 1. If you enter 1, then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type 2. If you enter 2, then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type 3. If you enter 3, then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

When you select option 2 or 3, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

- Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

- To generate the logs for the PEP server in a file, type yes.

- Press ENTER.

The installer generates the installation files as per the options you selected.

```
PepServer's log will be generated in a file.
```

```
Successfully generated Big Data Protector files for AWS Databricks in ./Installation_Files/
```

All the generated files under `./Installation_Files/Uploads/` needs to be uploaded to either DBFS root directory (`dbfs:/`), Amazon Simple Storage Service (Amazon S3), or Databricks Workspace.

For DBFS:

The `./Installation_Files/upload_to_dbfs.sh` script can be used to upload the files under `./Installation_Files/Uploads/` to DBFS root directory (`dbfs:/`)

Usage: `./upload_to_dbfs.sh`

Warning: The `upload_to_dbfs.sh` script uses DBFS REST API that has a limitation of uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload failure.

For Amazon S3:

The `./Installation_Files/upload_to_amazon_s3.sh` script can be used to upload the files under `./Installation_Files/Uploads/` to Amazon Simple Storage Service (Amazon S3).

Usage: `./upload_to_amazon_s3.sh`

For Workspace:

The `./Installation_Files/upload_to_workspace.sh` script can be used to upload the files under `./Installation_Files/Uploads/` to Databricks' Workspace.

Usage: `./upload_to_workspace.sh`

Warning: The `upload_to_workspace.sh` script uses Workspace REST API that has a limitation of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload failure.

```
Installation_Files
  Uploads
    BDP_Init_Script_9.1.0.0.x_<TAG>.sh
    BDP_Package_9.1.0.0.x_<TAG>.tgz
  curl
  upload_to_amazon_s3.sh
  upload_to_dbfs.sh
  upload_to_workspace.sh
```

12.6.9.5 Modifying the `pepserver.cfg` File

This section explains the process of modifying the `pepserver.cfg` file. You will be unable to modify the `pepserver.cfg` file when the cluster is running. Therefore, you must extract the Big Data Protector archive generated by the configurator script, update the `pepserver.cfg` file, repackage the installation files, and then upload the updated Big Data Protector archive to *DBFS* using the helper script.

Attention: Modifying the `pepserver.cfg` file is an optional step. Exercise caution when modifying the contents of the `pepserver.cfg` file.

► To modify the `pepserver.cfg` file:

- To navigate to the `./Installation_Files/Uploads/` directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

- To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

- To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

- Navigate to the directory that contains the `pepserver.cfg` file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepserver.cfg* file.
6. Modify the *pepserver.cfg* file according to your requirements.
7. Save the changes to the *pepserver.cfg* file.
8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. To upload the updated Big Data Protector archive to *Workspace*, use the *upload_to_workspace.sh* script.

Note: For more information about uploading the files to Workspace using the script, refer to section [Uploading the Installation Files to the Workspace Storage](#).

12.6.9.6 Uploading the Installation Files to the Workspace Storage

► To upload the installation files to the Workspace:

1. To execute the script to upload the Big Data Protector installation files to the Databricks workspace, run the following command:

```
./upload_to_workspace.sh
```

2. Press ENTER.

The prompt to continue with the upload process appears.

```
*****
This Script will upload the files generated under ./Uploads/ to Workspace.
*****
```

```
Do you wish to continue? [yes or no]:
```

3. To proceed with the upload, type *yes*.

4. Press ENTER.

The execution of the script starts where the script checks for the presence of the Big Data Protector installation package and the initialization script and the prompt to enter the Workspace path appears.

```
Execution of upload_to_workspace.sh script is started.

Checking if all required files are present under ./Uploads/
BDP_Package*.tgz file is present under ./Uploads/
BDP_Init_Script*.sh file is present under ./Uploads/.

All required files are present under ./Uploads/.

Enter the Workspace Path for uploading the files.
Format of the path: /Users/<directory_structure(if any)>
Example of the path: /Users/abc@xyz.com/mydirectory1/mydirectory2/mydirectory3
```



Here, "mydirectory3" is the directory where the BDP files are expected to be uploaded.

NOTE:

Whitespaces in the path are not allowed.

Enter the Workspace Path:

5. Enter the path of the directory in the Workspace where you want to upload the Big Data Protector installation package and the initialization script.

6. Press ENTER.

The prompt to enter the Databricks URL appears.

Enter Databricks URL:

7. Enter the URL that you have used to access the Databricks environment.

8. Press ENTER.

The prompt to enter the access token appears.

Enter Access Token:

9. Enter the access token.

10. Press ENTER.

The upload script splits the Big Data Protector installation package into fragments and uploads the files to the location mentioned in the *Workspace Path* prompt.

```
Starting upload of BDP_Package_9.1.0.0.x_<TAG>.tgz's fragments to Workspace. This may
take some time.
```

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_00 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_00 file to Workspace.
```

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_01 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_01 file to Workspace.
```

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_02 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_02 file to Workspace.
```

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_03 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_03 file to Workspace.
```

```
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz's fragments to Workspace.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Workspace.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Workspace.
```

Successfully uploaded all files to Workspace.

To use the BDP_Init_Script_9.1.0.0.x_<TAG>.sh file, which is just uploaded to Workspace, as an Init Script in Databricks Cluster, follow the below steps:

1. Create 1 secret in a Databricks-backed secret scope for storing the "PTY_WORKSPACE_ACCESS_TOKEN".

2. Add the below environment variable in the "Environment variables" section of the "Spark" tab of the "Advanced Options" menu of the "Configuration" tab of your Databricks Cluster:

2a. `PTY_WORKSPACE_ACCESS_TOKEN={{secrets/<scope_name>/<secret_name>}}`
here, replace <scope_name> with the name of your secret scope and <secret_name> with the name of your "PTY_WORKSPACE_ACCESS_TOKEN" secret.

3. Add the below Workspace Path in the "Init Scripts" tab of the "Advanced Options" menu of the "Configuration" tab of your Databricks Cluster:

`/<workspace_path>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh`

12.6.9.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.9.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector by creating a new cluster in the AWS Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/create.html>.
2. On the cluster creation page, click the **Advanced Options** tab.
3. Click the **Spark** tab.
4. Add the following Spark configurations, such as, the key and value separated by a space.

Table 12-69: Spark Configurations

| Keys | Values |
|--|---|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegility.spark.PtyExecSparkPlugin</i> |

The following image represents a sample Spark configuration for reference.

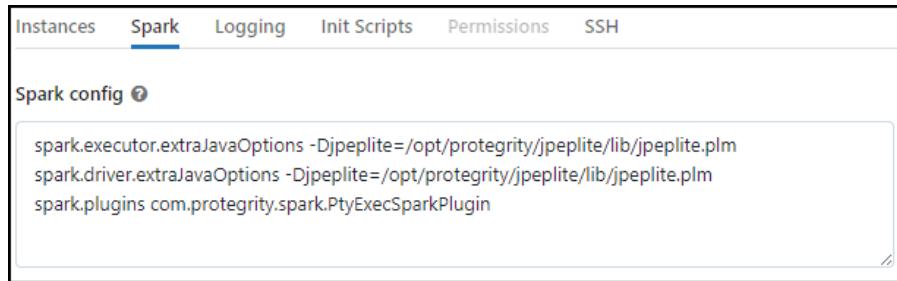


Figure 12-322: Spark Configurations

5. Under **Environment variables**, add the workspace access token in the following format:

```
PTY_WORKSPACE_ACCESS_TOKEN={ { secrets/<scope_name>/<secret_name> } }
```

Replace the *<scope_name>* with the name of your secret scope and the *<secret_name>* with the name of your *PTY_WORKSPACE_ACCESS_TOKEN* secret.

The following image represents a sample Spark configuration and the environment variables for reference.

The screenshot shows the 'Spark' tab selected in a navigation bar. Under 'Spark config', there are two code snippets:

```
spark.executor.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.driver.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Under 'Environment variables', there are two entries:

```
PYSPARK_PYTHON=/databricks/python3/bin/python3
PTY_WORKSPACE_ACCESS_TOKEN={{secrets/bdp-doc/bdp-doc-key}}
```

Figure 12-323: Spark Configuration and Environment Variables

- Click the **Init Scripts** tab.

The screenshot shows the 'Init Scripts' tab selected in a navigation bar. It displays a table with columns 'Type', 'File path', and 'Region'. A large '+' button is centered above the table, and a message below it says 'No init script set for this compute'. Below the table, there are 'Source' and 'File path' dropdowns. The 'Source' dropdown is set to 'Workspace' and the 'File path' dropdown is empty.

Figure 12-324: The Init Scripts Tab

- From the **Source** list, select **Workspace**.
- In the **File path** box, enter the location where you uploaded the initialization script to the *Workspace* using the helper script. For example, enter the path for the initialization script as <storage_location>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh
- Click **Add**.
The location of the initialization script appears in the **File path** column.

The screenshot shows the 'Init Scripts' tab with the initialization script added. The table now has one row:

| Type | File path | Region |
|-----------|---|--------|
| Workspace | /Users/.../bdp-doc/BDP_Init_Script_9.1.0.0.31_a1.sh | |

Below the table, the 'Source' and 'File path' dropdowns are shown again, with 'Source' set to 'Workspace'.

Figure 12-325: Initialization Script Path

- Click **Create compute**.
The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is `user1@example.com`, then the user `user1` must be present in the ESA policy.

12.6.9.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the AWS Databricks workspace.

Note: For more information about modifying the `pepserver.cfg` file, refer to section [Modifying the pepserver.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/clusters-manage.html#edit-a-cluster>.
2. On the cluster configuration page, click the **Advanced Options** toggle.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format separated by a space.

Table 12-70: Spark Configurations

| Keys | Values |
|--|--|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegrity.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark configuration for reference.

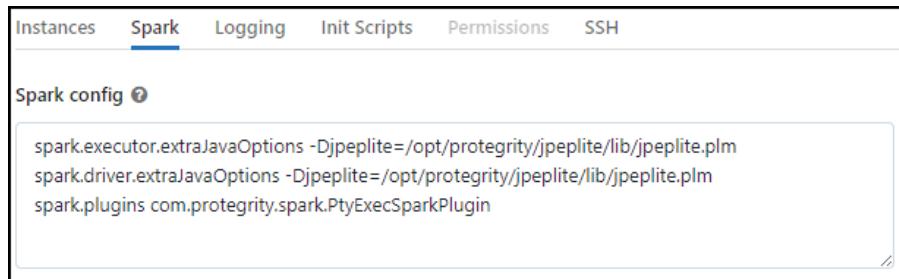


Figure 12-326: Spark Configurations

5. Under **Environment variables**, add the workspace access token in the following format:

```
PTY_WORKSPACE_ACCESS_TOKEN={ { secrets/<scope_name>/<secret_name> } }
```

Replace the `<scope_name>` with the name of your secret scope and the `<secret_name>` with the name of your `PTY_WORKSPACE_ACCESS_TOKEN` secret.

The following image represents a sample Spark configuration and the environment variables for reference.

The screenshot shows the 'Spark' tab selected in a navigation bar. Under 'Spark config', there are two code snippets:

```
spark.executor.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.driver.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Under 'Environment variables', there are two entries:

```
PYSPARK_PYTHON=/databricks/python3/bin/python3
PTY_WORKSPACE_ACCESS_TOKEN={{secrets/bdp-doc/bdp-doc-key}}
```

Figure 12-327: Spark Configuration and Environment Variables

- Click the **Init Scripts** tab.

The screenshot shows the 'Init Scripts' tab selected in a navigation bar. It displays a table with columns 'Type', 'File path', and 'Region'. A large '+' button is centered above the table, and a message below it says 'No init script set for this compute'. Below the table, there are dropdown menus for 'Source' (set to 'Workspace') and 'File path' (empty), and a 'Add' button.

Figure 12-328: The Init Scripts Tab

- From the **Source** list, select **Workspace**.
 - In the **File path** box, enter the location where you uploaded the initialization script to *Workspace* using the helper script. For example, enter the path for the initialization script as <storage_location>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh
 - Click **Add**.
- The location of the initialization script appears in the **File path** column.

The screenshot shows the 'Init Scripts' tab with the initialization script added. The table now has one row with the following data:

| Type | File path | Region |
|-----------|---|--------|
| Workspace | /Users/.../bdp-doc/BDP_Init_Script_9.1.0.0.31_a1.sh | |

Below the table, the 'Source' and 'File path' fields are shown again, with 'File path' containing the same path as the table.

Figure 12-329: Initialization Script Path

- Click **Confirm and restart**.
- The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is `user1@example.com`, then the user `user1` must be present in the ESA policy.

12.6.9.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs requires registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';
```

OK

When you register Hive UDFs permanently on the Unity Catalog, the registration fails when you set the default catalog and the database to custom Unity Catalog and a database inside it. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.

For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to enable the registration of permanent Hive UDFs in the Unity Catalog.



To enable the registration of permanent Hive UDFs in the Unity Catalog, perform the following steps.

1. Register UDFs permanently in the *hive_metastore* catalog.
2. Call the UDF in the SQL queries using the fully qualified name in the following format:

```
<catalog>. <database>. <UDF>
```

For example,

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

3. Press ENTER.

The registration of the UDFs completes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

OK

This approach protects the data in the tables stored under the Unity Catalog database by referring to them by their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT
hive_metastore.default.ptyProtectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog



- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.10 Installing the Big Data Protector on an AWS Databricks Cluster using DBFS

This section describes the steps for installing the Big Data Hive and Spark Protectors on the AWS Databricks platform using the Databricks File System (DBFS).

Warning: You will be unable to use the initialization scripts on DBFS, starting 01-December-2023. Databricks is ending support for initialization scripts on DBFS and the feature will not function after that date.

For more information, refer to <https://docs.databricks.com/en/init-scripts/legacy-global.html>.

12.6.10.1 Verifying the Prerequisites for Installing the Big Data Protector on AWS Databricks Cluster using DBFS

Ensure that the following prerequisites are met, before installing the Big Data Protector on the AWS Databricks platform:

- The user should have a valid AWS account.
- The user should have access to the AWS Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-71: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes the Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |



- You have created a Databricks personal access token.

Note: For more information about creating the Databricks personal access token, refer to <https://docs.databricks.com/en/dev-tools/auth.html#databricks-personal-access-token-authentication>.

12.6.10.2 Understanding the Installation Workflow of the Big Data Protector on AWS Databricks

This section describes a brief overview of the installation workflow that the user needs to perform to install the Big Data Protector on the AWS Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the AWS Databricks platform.

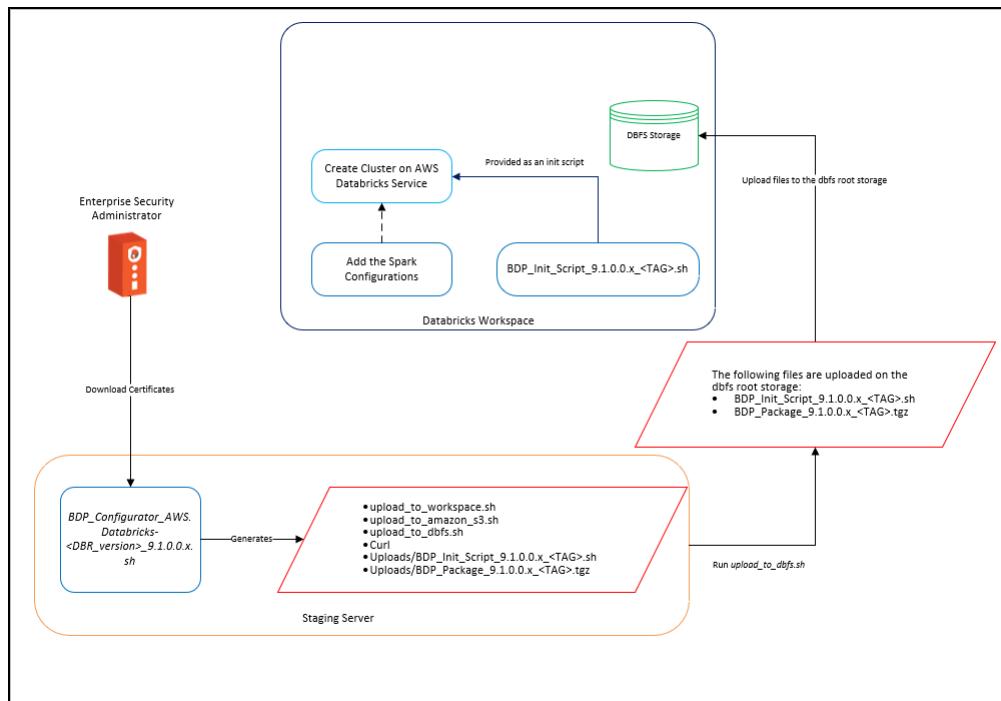


Figure 12-330: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector Installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- Run the `BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh` script to download the certificates from the ESA and create the following files in the `/Installation_Files`/directory:

- `upload_to_workspace.sh`
- `upload_to_amazon_s3.sh`
- `upload_to_dbfs.sh`
- `curl`
- `Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh`
- `Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz`

Note: For more information about running the `BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh` script, refer to section [Running the Configurator Script](#).

- c. Run the *upload_to_dbfs.sh* script to upload the following files on the *dbfs* root storage:

- *BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *upload_to_dbfs.sh* script, refer to section [Uploading the Files using the Helper Script](#).

2. On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* during cluster creation along with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector during Cluster Creation](#).

12.6.10.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because that IP address or hostname is written by the Configurator script in the *pepper.cfg* file.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
3. Extract the *BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* file from the Big Data Protector installation package using the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_AWS.Databricks-
<DBR_version>-64_9.1.0.0.x.tgz
```

The command extracts the *BDP_Configurator_AWS.Databricks-<DBR_version>-64_9.1.0.0.x.sh* file.

12.6.10.4 Running the Configurator Script

You must run the Big Data Protector configurator script to download certificates from the ESA, and create the installation files for the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_AWS.Databricks-<DBR_version>-9.1.0.0.x.sh* script from the directory where it is extracted, execute the following command.

```
./BDP_Configurator_AWS.Databricks-<DBR_version>-9.1.0.0.x.sh
```

2. Press ENTER.



The prompt to continue the installation of the Big Data Protector appears.

```
*****
        Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS Databricks

Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
        Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for AWS Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Amazon Simple Storage Service (Amazon S3), or Databricks Workspace.
This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

Enter a Unique Alpha-Numeric Tag:

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

Enter ESA Hostname or IP Address:

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

Enter ESA host listening port [8443]:

11. Enter the listening port for the ESA.

12. Press ENTER.

The prompt to enter the username for the ESA appears.

Enter ESA Username:

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

Enter ESA Username:

```
Extracting files...
Fetching Certificates.....
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
Fetching Certificates.....
Enter host password for user '<user_name>':
  % Total    % Received % Xferd  Average Speed   Time     Time     Time  Current
               Dload  Upload   Total   Spent   Left  Speed
100 30720  100 30720      0       0  32370      0 ---:--- ---:--- ---:--- 32336
```

Please select the Audit Store type where Log Forwarder(s) should send logs to.

```
[ 1 ] : Protegility Audit Store
[ 2 ] : External Audit Store
[ 3 ] : Protegility Audit Store + External Audit Store
```

Enter the no.:

17. To select the Audit Store type, select any one of the following options:

Table 12-72: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

When you select option *2* or *3*, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

18. Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

19. To generate the logs for the PEP server in a file, type *yes*.

20. Press ENTER.



The installer generates the installation files as per the options you selected.

```
PepServer's log will be generated in a file.
```

```
Successfully generated Big Data Protector files for AWS Databricks in ./  
Installation_Files/
```

All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Amazon Simple Storage Service (Amazon S3), or Databricks Workspace.

For DBFS:

```
The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./  
Installation_Files/Uploads/ to DBFS root directory (dbfs:/)
```

Usage: ./upload_to_dbfs.sh

Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload failure.

For Amazon S3:

```
The ./Installation_Files/upload_to_amazon_s3.sh script can be used to upload the files  
under ./Installation_Files/Uploads/ to Amazon Simple Storage Service (Amazon S3).
```

Usage: ./upload_to_amazon_s3.sh

For Workspace:

```
The ./Installation_Files/upload_to_workspace.sh script can be used to upload the files  
under ./Installation_Files/Uploads/ to Databricks' Workspace.
```

Usage: ./upload_to_workspace.sh

Warning: The upload_to_workspace.sh script uses Workspace REST API that has a limitation of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload failure.

```
Installation_Files  
  Uploads  
    BDP_Init_Script_9.1.0.0.x_<TAG>.sh  
    BDP_Package_9.1.0.0.x_<TAG>.tgz  
  curl  
  upload_to_amazon_s3.sh  
  upload_to_dbfs.sh  
  upload_to_workspace.sh
```

12.6.10.5 Modifying the *pepperserver.cfg* File

This section explains the process of modifying the *pepperserver.cfg* file. You will be unable to modify the *pepperserver.cfg* file when the cluster is running. Therefore, you must extract the Big Data Protector archive generated by the configurator script, update the *pepperserver.cfg* file, repackage the installation files, and then upload the updated Big Data Protector archive to *DBFS* using the helper script.

Attention: Modifying the *pepperserver.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepperserver.cfg* file.

► To modify the *pepperserver.cfg* file:

- To navigate to the ./*Installation_Files/Uploads/* directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

- To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the *pepperserver.cfg* file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepperserver.cfg* file.

6. Modify the *pepperserver.cfg* file according to your requirements.

7. Save the changes to the *pepperserver.cfg* file.

8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. To upload the updated Big Data Protector archive to *DBFS*, use the *upload_to_dbfs.sh* script.

Note: For more information about uploading the files to DBFS using the script, refer to section [Uploading the Files using the Helper Script](#).

12.6.10.6 Uploading the Files using the Helper Script

You must upload the files generated by the configurator script to the DBFS root directory (*dbfs:/*) of the AWS Databricks workspace.

Caution: The *upload_to_dbfs.sh* helper script can be optionally used to upload the files. It uses the DBFS REST API, that has a limitation of uploading a file by dividing it into chunks of 1 MB. If you encounter any failure, then upload the files manually.

Note: The *upload_to_dbfs.sh* helper script requires an API authentication token that needs to be generated by the user.

For more information about generating a personal access token on AWS Databricks Workspace UI, refer to <https://docs.databricks.com/dev-tools/api/latest/authentication.html>.

► To upload the files using the helper script:

1. To execute the helper script, run the following command:

```
./upload_to_dbfs.sh
```

2. Press ENTER.

The prompt to continue with the upload appears.

```
*****
*****  
This Script will upload the files generated under ./Uploads/ to DBFS root directory.
```

```
Note: Make sure that you have a valid API Access Token generated from the AWS Databricks Web UI.
```

Do you wish to continue? [yes or no]:

3. To continue with the upload, type *yes*.
 4. Press **ENTER**.

The script verifies the presence of the installation package and the initialization script in the `/Uploads/` directory and the prompt to enter the AWS Databricks instance URL appears.

Upload to DBFS Script started...

```
Checking if all required files are present under ./Uploads/  
All files present
```

Enter the Databricks Instance URL (E.g.: <https://dbc-a1b2c3d4-1234.cloud.databricks.com>):

5. Enter the AWS Databricks instance URL of the AWS Databricks workspace.

For more information about the AWS Databricks instance URL, refer to <https://docs.databricks.com/workspace/workspace-details.html#workspace-instance-names-urls-and-ids>.

6. Press ENTER.

The prompt to enter the API access token appears.

Enter API Access Token:

7. Enter the access token that was generated previously in the same workspace.

8. Press ENTER.

The helper script uploads the installation package and the initialization script from the `./Installation_Files/Uploads/` directory to the DBFS root directory using the DBFS REST API.

```
Started upload of ./Uploads/.tmp/BDP_Init_Script_9.1.0.0.x <TAG>.sh to DBFS.  
#####
##### 100.0%  
#####
##### 100.0%  
#####
##### 100.0%  
Finished upload of ./Uploads/.tmp/BDP Init Script 9.1.0.0.x <TAG>.sh to DBFS.
```



Successfully uploaded all files to DBFS root 'dbfs:/'

Note: Ensure that you upload the Big Data Protector installation files only to the DBFS root directory (`dbfs:/`) in the AWS Databricks workspace. There might be a possibility that similar file names might be present. To distinguish the installation files and prevent overwriting it, they are tagged with alpha-numeric tags provided by the user.

12.6.10.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.10.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector by creating a new cluster in the AWS Databricks workspace.

Note: For more information about modifying the `pepservice.cfg` file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/create.html>.
2. On the cluster creation page, click the **Advanced Options** tab.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value separated by a space.

Table 12-73: Spark Configurations

| Keys | Values |
|--|---|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark configuration for reference.

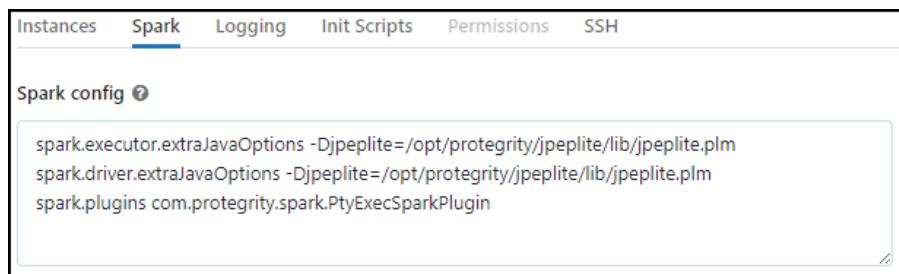


Figure 12-331: Spark Configurations

4. Click the **Init Scripts** tab.

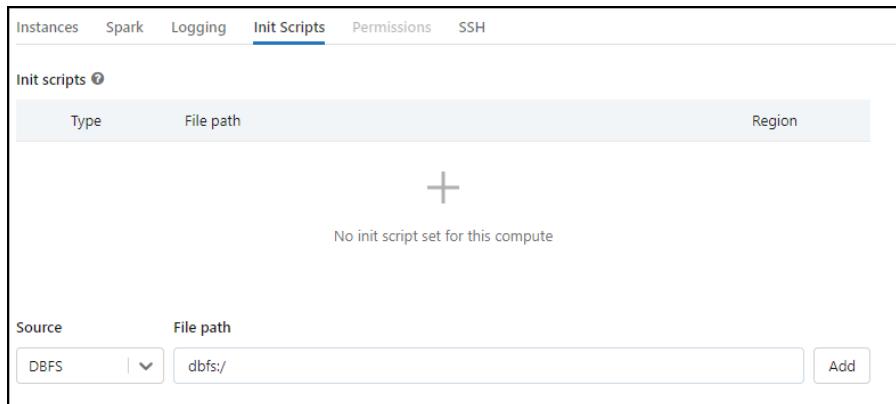


Figure 12-332: The Init Scripts Tab

5. From the **Source** list, select **DBFS**.
6. In the **File path** box, enter the location where you uploaded the initialization script to *DBFS* using the helper script. For example, enter the path for the initialization script as *dbfs:/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*
7. Click **Add**.

The location of the initialization script appears in the **File path** column.

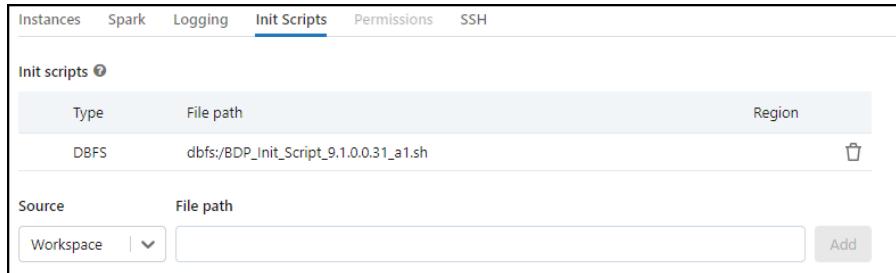


Figure 12-333: Initialization Script Path

8. Click **Create compute**.
The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.10.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the AWS Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

- To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the AWS Databricks workspace, refer to <https://docs.databricks.com/clusters/clusters-manage.html#edit-a-cluster>.

2. On the cluster configuration page, click the **Advanced Options** toggle.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format that is separated by a space.

Table 12-74: Spark Configurations

| Keys | Values |
|--|---|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark Configuration for reference.

The screenshot shows a tabbed interface with the "Spark" tab selected. Below the tabs, there's a section titled "Spark config" with a question mark icon. Inside this section, three configuration lines are listed:

```
spark.executor.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.driver.extraJavaOptions -Djpeplite=/opt/protegility/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Figure 12-334: Spark Configurations

5. Click the **Init Scripts** tab.

The screenshot shows the "Init Scripts" tab selected. The interface includes a table with columns for "Type", "File path", and "Region". A large plus sign button is present above the table. Below the table, there's a section for "Source" and "File path", with a dropdown menu set to "DBFS" and a text input field containing "dbfs:/".

| Type | File path | Region |
|------|-----------|--------|
| DBFS | dbfs:/ | |

Figure 12-335: The Init Scripts Tab

6. From the **Source** list, select **DBFS**.
7. In the **File path** box, enter the location where you uploaded the initialization script to *DBFS* using the helper script. For example, enter the path for the initialization script as *dbfs:/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*
8. Click **Add**.

The location of the initialization script appears in the **File path** column.

The screenshot shows the "Init Scripts" tab selected. The table now has one row, with the "File path" column containing "dbfs:/BDP_Init_Script_9.1.0.0.31_a1.sh". Below the table, the "Source" dropdown is set to "Workspace" and the "File path" input field is empty.

| Type | File path | Region |
|------|--|--------|
| DBFS | dbfs:/BDP_Init_Script_9.1.0.0.31_a1.sh | |

Figure 12-336: Initialization Script Path

9. Click **Confirm and restart**.

The Big Data Hive and Spark Protector are installed on the AWS Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.10.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs require registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION ptyProtectStr;
3 CREATE FUNCTION ptyProtectStr AS 'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION ptyUnprotectStr;
5 CREATE FUNCTION ptyUnprotectStr AS 'com.protegility.hive.udf.ptyUnprotectStr';
```

OK

When you register Hive UDFs permanently on the Unity Catalog, the registration fails when you set the default catalog and the database to custom Unity Catalog and a database inside it. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.



For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to enable the registration of permanent Hive UDFs in the Unity Catalog.

To enable the registration of permanent Hive UDFs in the Unity Catalog, perform the following steps.

1. Register UDFs permanently in the *hive_metastore* catalog.
2. Call the UDF in the SQL queries using the fully qualified name in the following format:

```
<catalog>. <database>. <UDF>
```

For example,

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

3. Press ENTER.

The registration of the UDFs completes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';

OK
```

This approach protects the data in the tables stored under the Unity Catalog database by referring to them by their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT
hive_metastore.default.ptyProtectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table



Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.11 Installing the Big Data Protector on an Azure Databricks Cluster Using ABFSS

This section describes the steps for installing the Big Data Hive and Spark Protector on the Azure Databricks platform using the Azure Blob File System Secure (ABFSS).

The ABFSS is a Hadoop file system driver that is compatible with the Azure Data Lake Storage Gen2. The ABFSS driver uses a URI format to manage files and directories in a Data Lake Storage Gen2 account.

The URI format for ABFSS is explained in the following syntax:

```
abfss://<container_name>@<account_name>.dfs.core.windows.net/<path>/<file_name>
```

where,

- *Scheme identifier* - indicates the protocol to be used. In this case, the protocol is ABFSS.
- *Container name* - specifies the parent location that contains the directories and files.
- *Account name* - indicates the name of the storage account.
- *Path* - is a forward slash delimited (/) representation of the directory structure.
- *File name* - indicates the name of the individual file.

Warning: This build is designed to work only for "single-user" access mode. This build should not be deployed in "shared" and "No isolation shared" access mode environments.

12.6.11.1 Verifying the Prerequisites for Installing the Big Data Protector on an Azure Databricks Cluster using ABFSS

Ensure that the following prerequisites are met, before installing the Big Data Protector on the Azure Databricks platform:

- You have a valid Microsoft Azure account.
- You have access to the Azure Databricks Workspace and must be authorized to create clusters.
- You have an Azure Data Lake Storage Gen2 account with a container and an optional directory inside the container.

Note: For more information about creating an Azure Data Lake Storage Gen2 account, refer to <https://learn.microsoft.com/en-us/azure/storage/blobs/create-data-lake-storage-account>.

- You have a Storage Account SAS token generated for the Azure Data Lake Storage Gen2 account.

Note: The *upload_to_adls.sh* helper script requires you to generate the Storage Account SAS Token for Azure Storage REST API authentication.



For more information about Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/azure/storage/common/storage-sas-overview>.

For more information about creating a Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/rest/api/storageservices/create-account-sas>.

- You have a Databricks-backed secret scope.

Note: For more information about creating a Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes#create-a-databricks-backed-secret-scope>.

- You have added the generated SAS Token as a Databricks secret to the Databricks-backed secret scope.

Note: For more information about adding a secret to the Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secrets#create-a-secret-in-a-databricks-backed-scope>.

- You have the ESA appliance, version 9.1.0.0, installed, configured, and running, and the Databricks cluster nodes are able to communicate with the ESA.
- You have a Linux machine with connectivity to the ESA.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-75: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |

12.6.11.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks with ABFSS

This section describes a brief overview of the installation workflow that you must perform to install the Big Data Protector to the ABFSS on the Azure Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the Azure Databricks platform.

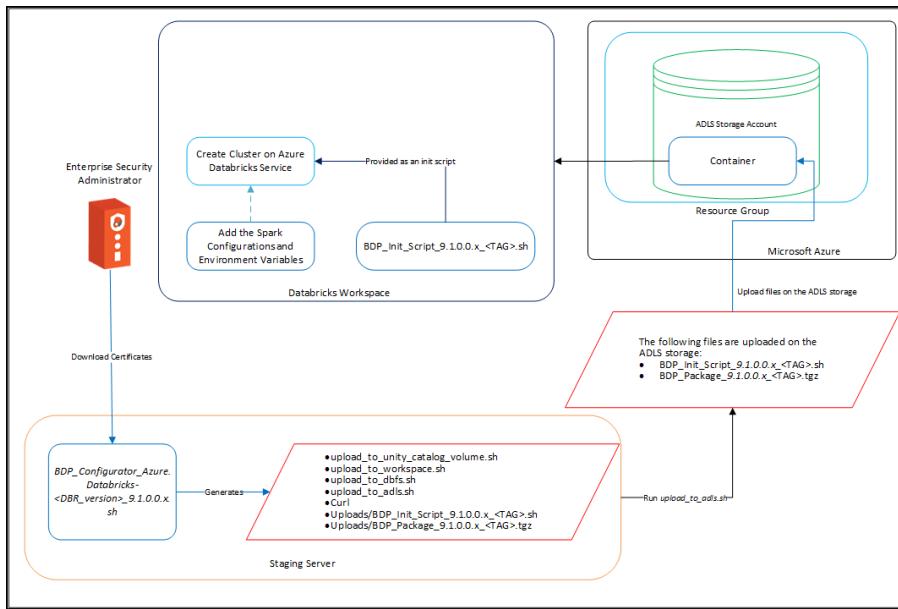


Figure 12-337: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector Installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- Run the *BDP_Configurator_Azure.Databricks-<DBR_version>.9.1.0.0.x.sh* script to download the certificates from the ESA and create the following files in the *./Installation_Files/* directory:

- *upload_to_unity_catalog_volume.sh*
- *upload_to_workspace.sh*
- *upload_to_dbfs.sh*
- *upload_to_adls.sh*
- *curl*
- *Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *BDP_Configurator_Azure.Databricks-<DBR_version>.9.1.0.0.x.sh* script, refer to section [Running the Configurator Script](#).

- Run the *upload_to_adls.sh* script to upload the following files on the *adls* storage:

- *BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *upload_to_adls.sh* script, refer to section [Uploading the Installation Files to ADLS](#).

- On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector](#).

12.6.11.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script. The Configurator script creates the Big Data Protector installation files that you must upload to ABFSS.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because the configurator script appends the same IP address or hostname in the *pepserver.cfg* file.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
3. To extract the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file.

12.6.11.4 Running the Configurator Script

You must execute the Big Data Protector configurator script to download the certificates from the ESA and create the installation files required to install the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command:

```
./BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh
```

2. Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks
Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
```

```
This will create the Big Data Protector Installation files for Azure Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), or Databricks Workspace. This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

```
Enter a Unique Alpha-Numeric Tag:
```

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

11. Enter the listening port for the ESA.

12. Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current | | | |
|---------|------------|---------|---------------|------|-------|-------|---------|---------|--------|-------|
| Dload | Upload | Total | Spent | Left | Speed | | | | | |
| 100 | 20480 | 100 | 20480 | 0 | 0 | 11257 | 0:00:01 | 0:00:01 | --::-- | 11258 |
| | | | | | | | | | | |

Select the Audit Store type where Log Forwarder(s) should send logs to.

[1] : Protegrity Audit Store
 [2] : External Audit Store
 [3] : Protegrity Audit Store + External Audit Store

Enter the no.:

- To select the Audit Store type, select any one of the following options:

Table 12-76: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegrity Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegrity Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegrity Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegrity/fluent-bit/data/config.d/</i> directory. |

- Press ENTER.

When you select option *2* or *3*, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:

- Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

Do you want PepServer's log to be generated in a file? [yes or no]:

- To generate the logs for the PEP server in a file, type *yes*.

- Press ENTER.

The installer generates the installation files based on the options you selected and the prompt to select the upload type appears.

PepServer's log will be generated in a file.

Successfully generated Big Data Protector files for Azure Databricks in ./Installation_Files/

All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), Databricks Workspace or Unity Catalog Volume.

Select an option:

[1] : Upload the Installation Files now
 [2] : Upload the Installation Files manually later

Enter the no.:

The following snippet lists the files and directories generated by the configurator script in the `./Installation_Files` directory:

```
+++ curl
+++ Uploads
|   +++ BDP_Init_Script_9.1.0.0.x_<TAG>.sh
|   +++ BDP_Package_9.1.0.0.x_<TAG>.tgz
+++ upload_to_adls.sh
+++ upload_to_dbfs.sh
+++ upload_to_workspace.sh
+++ upload_to_unity_catalog_volume.sh
```

22. To manually upload the installation files using the helper script, type 2.

If you select option 1, to upload the installation files immediately, then the configurator script will prompt to enter the location to upload the installation files.

```
Select the location where you want to upload the BDP Installation Files:
[ 1 ] : Upload to DBFS
[ 2 ] : Upload to ADLS
[ 3 ] : Upload to Workspace
[ 4 ] : Upload to Unity Catalog Volume
```

Enter the no.:

Note: Depending on the location that you select, the configurator script will trigger the corresponding helper script to upload the installation files.

- For more information about uploading the installation files to DBFS, refer to the section [Uploading the Installation Files to DBFS](#).
- For more information about uploading the installation files to ADLS, refer to the section [Uploading the Installation Files to ADLS](#).
- For more information about uploading the installation files to the Workspace, refer to the section [Uploading the Installation Files to the Workspace Storage](#).
- For more information about uploading the installation files to the Unity Catalog Volume, refer to the section [Uploading the Installation Files to the Unity Catalog Volumes](#).

23. Press ENTER.

The configurator script provides the information about uploading the installation files to the required Azure Databricks storage.

```
For DBFS:
The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./Installation_Files/Uploads/ to DBFS root directory (dbfs:/)
Usage: ./upload_to_dbfs.sh
Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of
uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload
failure.

For ADLS:
The ./Installation_Files/upload_to_adls.sh script can be used to upload the files under ./Installation_Files/Uploads/ to Azure Data Lake Storage (ADLS).
Usage: ./upload_to_adls.sh

For Workspace:
The ./Installation_Files/upload_to_workspace.sh script can be used to upload the files
under ./Installation_Files/Uploads/ to Databricks' Workspace.
Usage: ./upload_to_workspace.sh
Warning: The upload_to_workspace.sh script uses Workspace REST API that has a limitation
of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload
failure.
```

```
For Unity Catalog Volume:  
The ./Installation_Files/upload_to_unity_catalog_volume.sh script can be used to upload  
the files under ./Installation_Files/Uploads/ to Unity Catalog Volume.  
Usage: ./upload_to_unity_catalog_volume.sh
```

12.6.11.5 Modifying the *pepserver.cfg* File

This section explains the process of modifying the *pepserver.cfg* file. You will be unable to modify the *pepserver.cfg* file when the cluster is running. Therefore, you must:

1. Extract the Big Data Protector archive generated by the configurator script.
2. Update the *pepserver.cfg* file.
3. Repackage the installation files.
4. Upload the updated Big Data Protector archive to *ABFSS* using the helper script.

Attention: Modifying the *pepserver.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepserver.cfg* file.

► To modify the *pepserver.cfg* file:

1. To navigate to the *./Installation_Files/Uploads/* directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

2. To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the *pepserver.cfg* file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepserver.cfg* file.
6. Modify the *pepserver.cfg* file according to requirements.
7. Save the changes to the *pepserver.cfg* file.
8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --  
owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. To upload the updated Big Data Protector archive to *ABFSS*, use the *upload_to_adls.sh* script.

Note: For more information about uploading the files to ABFSS using the script, refer to section [Uploading the Files using the Helper Script](#).

12.6.11.6 Uploading the Installation Files to ADLS

You must upload the installation files generated by the configurator script to the Azure Data Lake Storage Gen2 (ADLS) file system (*abfss:/*) of the Azure Databricks workspace. You can use the *upload_to_adls.sh* script to upload the installation files to the ADLS file system.

Before you begin

Ensure that the following prerequisites are met before installing the Big Data Protector on the Azure Databricks platform:

- You have an Azure Data Lake Storage Gen2 account with a container and an optional directory inside the container.

Note: For more information about creating an Azure Data Lake Storage Gen2 account, refer to <https://learn.microsoft.com/en-us/azure/storage/blobs/create-data-lake-storage-account>.

- You have a Storage Account SAS token generated for the Azure Data Lake Storage Gen2 account.

Note: The *upload_to_adls.sh* helper script requires a you to generate the Storage Account SAS Token for Azure Storage REST API authentication.

For more information about Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/azure/storage/common/storage-sas-overview>.

For more information about creating a Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/rest/api/storageservices/create-account-sas>.

- You have a Databricks-backed secret scope.

Note: For more information about creating a Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes#create-a-databricks-backed-secret-scope>.

- You have added the generated SAS Token as a Databricks secret to the Databricks-backed secret scope.

Note: For more information about adding a secret to the Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secrets#create-a-secret-in-a-databricks-backed-scope>.

- You have the shared key generated for the Azure Data Lake Storage Gen2 account.

Note:

For more information about the shared key, refer to <https://learn.microsoft.com/en-us/rest/api/storageservices/authorize-with-shared-key>.

► To upload the files using the helper script:

1. To execute the helper script, run using the following command.

```
./upload_to_adls.sh
```

2. Press ENTER.

The prompt to continue with the upload appears.

```
*****
This Script will upload the files generated under ./Uploads/ to ADLS.
```



```
*****
Do you wish to continue? [yes or no]:
```

3. To continue with the upload, type *yes*.
4. Press ENTER.

The configurator script verifies whether all the required files are available and the prompt to enter the ADLS path to upload the file appears. to enter the SAS token for the Azure storage account appears.

```
Execution of upload_to_adls.sh script is started.

Checking if BDP files are present under ./Uploads/
BDP_Package*.tgz file is present under ./Uploads/
BDP_Init_Script*.sh file is present under ./Uploads/ .

BDP files are present under ./Uploads/ .

Enter the ADLS path for uploading the BDP files.
Format of the path: <storage_account_name>/<container_name>/<directory_structure(if any)>
Example of the path: mystorageaccount/mycontainer/mydirectory1/mydirectory2/mydirectory3
Here, "mydirectory3" is the directory where the BDP files are expected to be uploaded.

NOTE:
Don't add any leading or trailing forward slashes in the path.
Whitespaces in the path are not allowed.
Enter the ADLS path:
```

5. Enter the path of the location where you want to upload the installation files.

6. Press ENTER.

The prompt to select the authorization methods to upload the files to ADLS appears.

```
Select one of the following authorization methods for uploading BDP files to ADLS:
[1] -> SAS Token (Shared Access Signature)
[2] -> Shared Key (Storage Account Key)
[ 1 or 2 ]:
```

7. To upload the files to ADLS using the SAS token, type *1*.

8. Press ENTER.

The prompt to enter the SAS token for the Azure storage account appears.

```
Enter SAS Token for Azure Storage Account:
```

If you select the *Shared Key* option to upload the files to ADLS, then the prompt to enter the shared key for the Azure storage account appears.

```
Enter Shared Key for Azure Storage Account:
```

9. Depending on the authorization method you select to upload the files to ADLS, enter the SAS token or the shared key for the Azure storage account.

10. Press ENTER.

The helper script uploads the following files in the *./Installation_Files/Uploads/* directory to ADLS using the ABFSS REST API:

```
- BDP_Init_Script_9.1.0.0.x_<TAG>.sh
- BDP_Package_9.1.0.0.x_<TAG>.tgz
```

A sample output of uploading the installation files to ADLS using the SAS token authorization is listed below.

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to ADLS.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to ADLS.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to ADLS.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to ADLS.
```

Successfully uploaded BDP files to ADLS.

To use BDP_Init_Script_9.1.0.0.x_<TAG>.sh file as an init script in Databricks Cluster, follow the below steps:

1) Since "[1] SAS Token" was used to upload the files, create a secret in a Databricks-backed secret scope for storing the SAS Token, and then, replace "<scope_name>" with the name of the secret scope and "<secret_name>" with the secret key in the below mentioned steps.

2) Add the following Spark configurations in the "Spark config" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
spark.hadoop.fs.azure.account.auth.type.bdpdocstore.dfs.core.windows.net SAS
spark.hadoop.fs.azure.sas.token.provider.type.bdpdocstore.dfs.core.windows.net
org.apache.hadoop.fs.azurebfs.sas.FixedSASTokenProvider
spark.hadoop.fs.azure.sas.fixed.token.bdpdocstore.dfs.core.windows.net {{secrets/
<scope_name>/<secret_name>}}
```

3) Add the following environment variable in the "Environment variables" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
PTY_SAS_TOKEN={{secrets/<scope_name>/<secret_name>}}
```

4) Add the following ABFSS URI in the "Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
abfss://<container_name>@<storage_account_name>.dfs.core.windows.net/<directory_path>/
BDP_Init_Script_9.1.0.0.x_<TAG>.sh
```

A sample output of uploading the installation files to ADLS using the shared key authorization is listed below.

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to ADLS.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to ADLS.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to ADLS.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to ADLS.
```

Successfully uploaded BDP files to ADLS.

To use BDP_Init_Script_9.1.0.0.x_<TAG>.sh file as an init script in Databricks Cluster, follow the below steps:

1) Since "[2] Shared Key" was used to upload the files, create a secret in a Databricks-backed secret scope for storing the Shared Key, and then, replace "<scope_name>" with the name of the secret scope and "<secret_name>" with the secret key in the below mentioned steps.

2) Add the following Spark configurations in the "Spark config" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
spark.hadoop.fs.azure.account.key.bdpdocstore.dfs.core.windows.net {{secrets/<scope_name>/
<secret_name>}}
```

3) Add the following environment variable in the "Environment variables" section under "Spark" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
PTY_SHARED_KEY={{secrets/<scope_name>/<secret_name>}}
```

4) Add the following ABFSS URI in the "Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the Databricks cluster page:

```
abfss://<container_name>@<storage_account_name>.dfs.core.windows.net/<directory_path>/
BDP_Init_Script_9.1.0.0.x_<TAG>.sh
```

12.6.11.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.11.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector in the Azure Databricks workspace on a new Databricks cluster.

Note: For more information about updating the *pepservr.cfg* file, refer to section [Modifying the pepservr.cfg File](#).

1. To create a cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/configure>.
2. On the cluster creation page, expand **Advanced Options**.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value, separated by a space.

Table 12-77: Spark Configurations

| Keys | Values |
|--|--|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djepelite=<Protegrity_Dir>/jepelite/lib/jepelite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djepelite=<Protegrity_Dir>/jepelite/lib/jepelite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegrity.spark.PtyExecSparkPlugin</i> |

- a. If you have used the SAS token to upload the installation files to ADLS, then, in the **Spark** tab, append the following configurations, in the key value format, separated by a space:

Table 12-78: Spark Configurations for SAS Token Authorization

| Keys | Values |
|---|--|
| <i>spark.hadoop.fs.azure.account.auth.type.<storage_account>.dfs.core.windows.net</i> | <i>SAS</i> |
| <i>spark.hadoop.fs.azure.sas.token.provider.type.<storage_account>.dfs.core.windows.net</i> | <i>org.apache.hadoop.fs.azurebfs.sas.FixedSASTokenProvider</i> |
| <i>spark.hadoop.fs.azure.sas.fixed.token.<storage_account>.dfs.core.windows.net</i> | <i>{<secrets/><scope_name>/<secret_name>}}</i> |

- b. If you have used the shared key to upload the installation files to ADLS, then, in the **Spark** tab, append the following configurations, in the key value format, separated by a space:

Table 12-79: Spark Configurations for Shared Token Authorization

| Keys | Values |
|---|--|
| <i>spark.hadoop.fs.azure.account.key.<storage_account>.dfs.core.windows.net</i> | <i>{<secrets/><scope_name>/<secret_name>}}</i> |

4. Under **Environment variables**, add the corresponding code snippet.
 - a. If you have used the SAS token to upload the installation files to ADLS, then, under **Environment Variables** tab, add the following code snippet:

```
PTY_SAS_TOKEN={<secrets/><scope_name>/<secret_name>}
```

The following image represents a sample Spark configuration and the environment variables for reference.



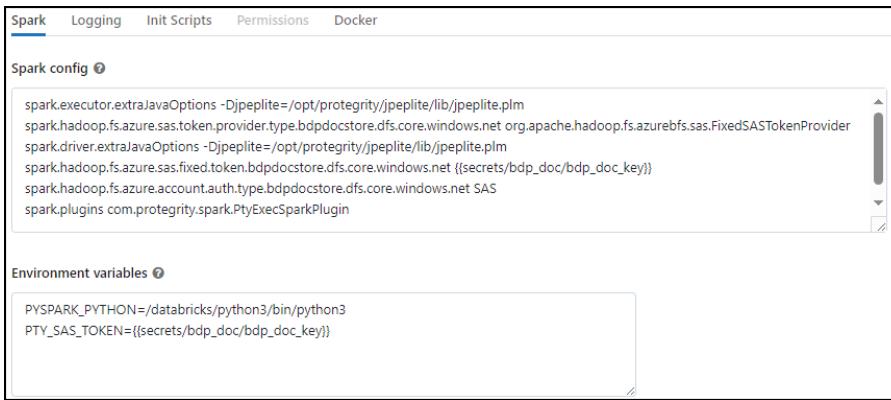


Figure 12-338: Spark Configurations and Environment Variables

- If you have used the shared token to upload the installation files to ADLS, then, under **Environment Variables** tab, add the following code snippet:

```
PTY_SHARED_KEY={{secrets/<scope_name>/<secret_name>}}
```

The following image represents a sample Spark configuration and the environment variables for reference.



Figure 12-339: Spark Configurations and Environment Variables

- Click the **Init Scripts** tab.

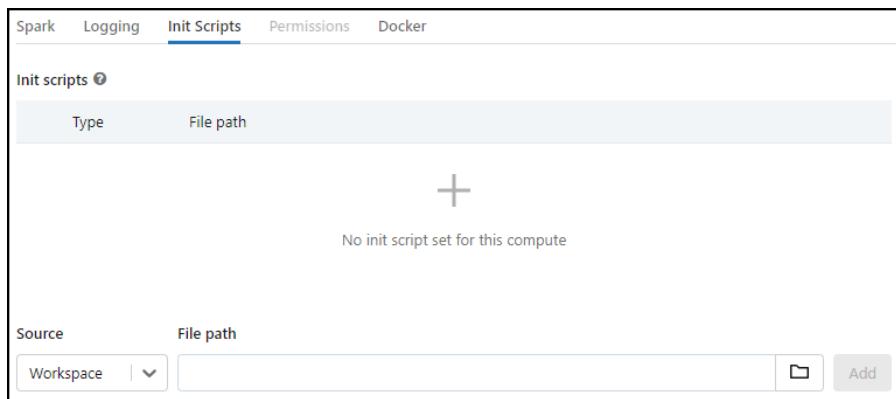


Figure 12-340: The Init Scripts Tab

- From the **Source** list, select **ABFSS**.
- In the **File path** box, enter the location where you uploaded the initialization script to the *ABFSS* using the helper script.

For example, enter the path for the initialization script as `abfss://<container_name>@<storage_account_name>/<directory_path>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh`.

8. Click **Add**.

The location of the initialization script appears in the **File path** column.

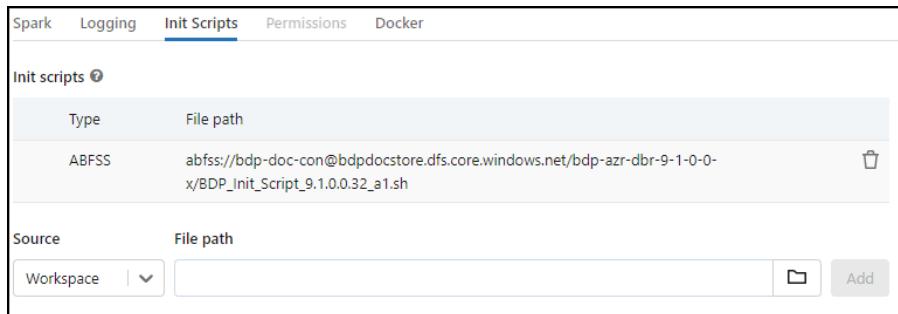


Figure 12-341: Init Script Path

9. Click **Create compute**.

The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is `user1@example.com`, then the user `user1` must be present in the ESA policy.

12.6.11.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the Azure Databricks workspace.

Note: For more information about modifying the `pepservice.cfg` file, refer to section [Modifying the pepservice.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/clusters-manage#--edit-a-cluster>.
2. On the cluster configuration page, expand **Advanced Options**.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. On the **Spark** tab, add the following Spark configurations, such as, the key and value, separated by a space.

Table 12-80: Spark Configurations

| Keys | Values |
|--|--|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

- a. If you have used the SAS token to upload the installation files to ADLS, then, in the **Spark** tab, append the following configurations, in the key value format, separated by a space:

Table 12-81: Spark Configurations for SAS Token Authorization

| Keys | Values |
|---|--|
| <code>spark.hadoop.fs.azure.account.auth.type.<storage_account>.dfs.core.windows.net</code> | <code>SAS</code> |
| <code>spark.hadoop.fs.azure.sas.token.provider.type.<storage_account>.dfs.core.windows.net</code> | <code>org.apache.hadoop.fs.azurebfs.sas.FixedSASTokenProvider</code> |
| <code>spark.hadoop.fs.azure.sas.fixed.token.<storage_account>.dfs.core.windows.net</code> | <code>{secrets/<scope_name>/<secret_name>}</code> |

- b. If you have used the shared key to upload the installation files to ADLS, then, in the **Spark** tab, append the following configurations, in the key value format, separated by a space:

Table 12-82: Spark Configurations for Shared Token Authorization

| Keys | Values |
|---|---|
| <code>spark.hadoop.fs.azure.account.key.<storage_account>.dfs.core.windows.net</code> | <code>{secrets/<scope_name>/<secret_name>}</code> |

5. Under **Environment variables**, add the corresponding code snippet.

- a. If you have used the SAS token to upload the installation files to ADLS, then, under **Environment Variables** tab, add the following code snippet:

```
PTY_SAS_TOKEN={secrets/<scope_name>/<secret_name>}
```

The following image represents a sample Spark configuration and the environment variables for reference.

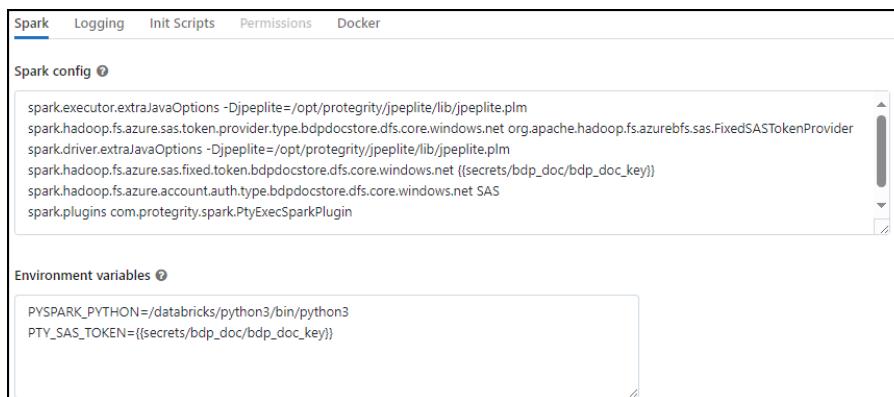


Figure 12-342: Spark Configurations and Environment Variables

- b. If you have used the shared token to upload the installation files to ADLS, then, under **Environment Variables** tab, add the following code snippet:

```
PTY_SHARED_KEY={secrets/<scope_name>/<secret_name>}
```

The following image represents a sample Spark configuration and the environment variables for reference.



Figure 12-343: Spark Configurations and Environment Variables

6. Click the **Init Scripts** tab.

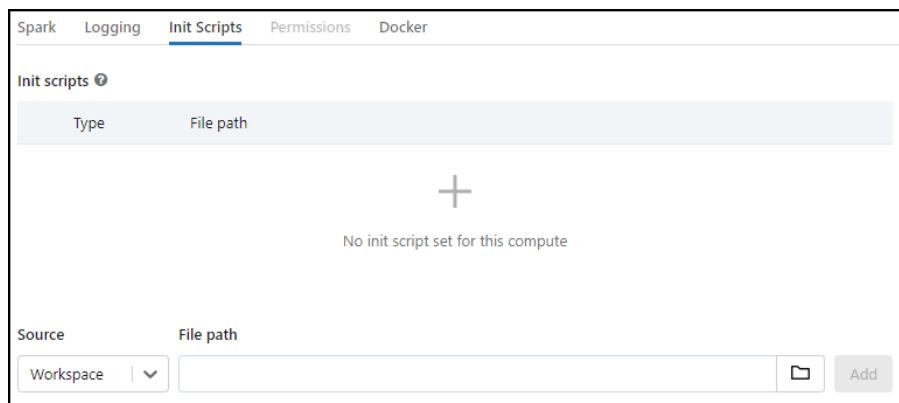


Figure 12-344: The Init Scripts Tab

7. From the **Source** list, select **ABFSS**.
8. In the **File path** box, enter the location where you uploaded the initialization script to the *ABFSS* using the helper script. For example, enter the path for the initialization script as *abfss://<container_name>@<storage_account_name>/<directory_path>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*.
9. Click **Add**.
The location of the initialization script appears in the **File path** column.

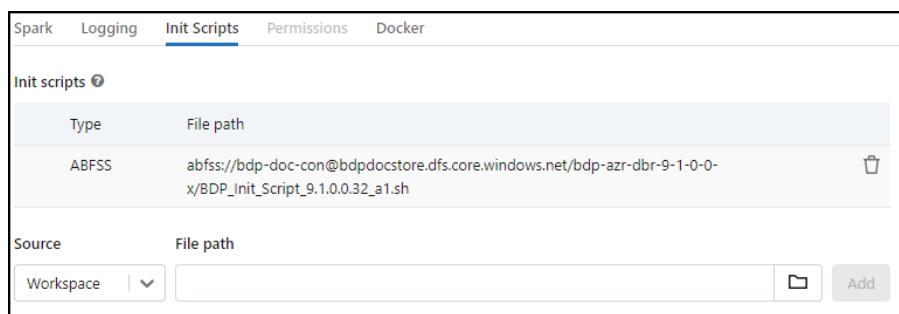


Figure 12-345: Init Script Path

10. To save the changes and restart the cluster, click **Confirm**.

The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.11.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs require registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

OK

In the Unity Catalog environment, the permanent registration for Hive UDFs fails for a custom catalog. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.



For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to use the permanently registered Hive UDFs in the Unity Catalog environment in a custom catalog.

The following examples illustrate protect and unprotect operations in the tables stored under a custom catalog from the Unity Catalog environment by referring them using their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT hive_metastore.default.ptyProtectStr(coll,
'Token_Alphanumeric') FROM dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.12 Installing the Big Data Protector on an Azure Databricks Cluster using the Workspace

This section describes the steps for installing the Big Data Hive and Spark Protector on the Azure Databricks platform using the Databricks workspace.

Warning: This build is designed to work only for "single-user" access mode. This build should not be deployed in "shared" and "No isolation shared" access mode environments.

12.6.12.1 Verifying the Prerequisites for Installing the Big Data Protector on Azure Databricks Cluster using the Workspace

Ensure that the following prerequisites are met, before installing the Big Data Protector on the Azure Databricks platform:

- The user should have a valid Azure account.
- The user should have access to the Azure Databricks Workspace and should be authorized to create clusters.



- The ESA appliance, version 9.1.0.0, is installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-83: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegity Audit Store appliance | The Log Forwarder sends all the logs to the Protegity Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |

- You have a Databricks-backed secret scope.

Note: For more information about creating a Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secret-scopes#--create-a-databricks-backed-secret-scope>.

- You have created a Databricks personal access token.

Note: For more information about creating the Databricks personal access token, refer to <https://learn.microsoft.com/en-us/azure/databricks/dev-tools/auth#--azure-databricks-personal-access-token-authentication>.

- You have added the personal access token as Databricks secrets to the Databricks-backed secret scope.

Note: For more information about added the access keys as Databricks secret to the Databricks-backed secret scope, refer to <https://learn.microsoft.com/en-us/azure/databricks/security/secrets/secrets#create-a-secret-in-a-databricks-backed-scope>.

- You have created the Databricks workspace folder to upload the files.

Note: For more information about creating folders in the Workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/workspace/workspace-objects#folders>.

12.6.12.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using the Workspace

This section describes a brief overview of the installation workflow that the user needs to perform to install the Big Data Protector on the Azure Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the Azure Databricks platform.

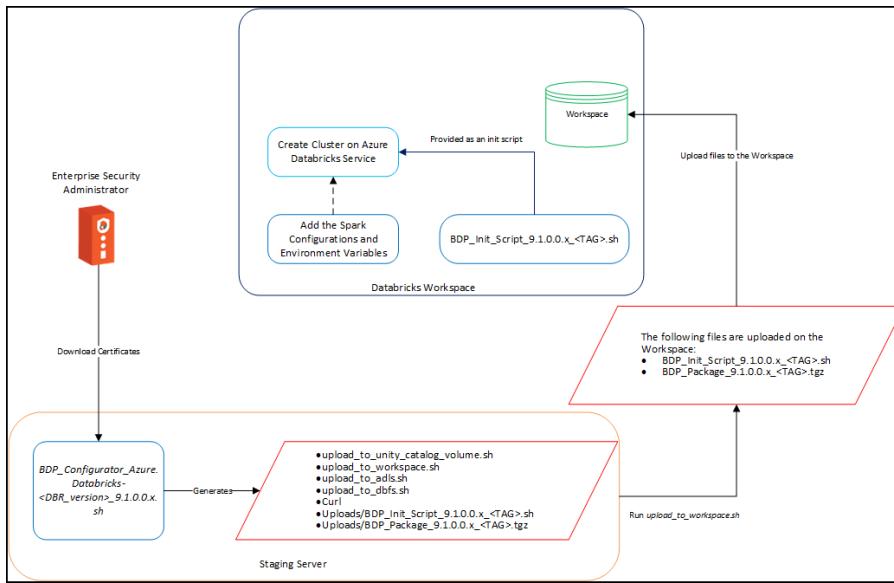


Figure 12-346: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- Run the *BDP_Configurator_Azure.Databricks-<DBR_version>.9.1.0.0.x.sh* script to download the certificates from the ESA and create the following files in the */Installation_Files*/directory:
 - upload_to_unity_catalog_volume.sh*
 - upload_to_workspace.sh*
 - upload_to_adls.sh*
 - upload_to_dbfs.sh*
 - curl*
 - Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
 - Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *BDP_Configurator_Azure.Databricks-<DBR_version>.9.1.0.0.x.sh* script, refer to section [Running the Configurator Script](#).

- Run the *upload_to_workspace.sh* script to upload the following files to the *workspace* storage:

- BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *upload_to_workspace.sh* script, refer to section [Uploading the Installation Files to the Workspace Storage](#).

- On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script during the cluster creation with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector](#).

12.6.12.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files that you must upload to ABFSS.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because the configurator script appends the same IP address or hostname in the *pepserver.cfg* file.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
3. To extract the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file.

12.6.12.4 Running the Configurator Script

You must execute the Big Data Protector configurator script to download the certificates from the ESA and create the installation files required to install the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command:

```
./BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh
```

2. Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks
Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
```

```
This will create the Big Data Protector Installation files for Azure Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

- Enter the location where you want to install the Big Data Protector.

- Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), or Databricks Workspace. This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

```
Enter a Unique Alpha-Numeric Tag:
```

- Enter a unique alphanumeric tag.

- Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

- Enter the hostname or the IP address of the ESA.

- Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

- Enter the listening port for the ESA.

- Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

- Enter the user name to connect to the ESA.

- Press ENTER.

The prompt to enter the password appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

- Enter the password.

- Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current | | | |
|---------|------------|---------|---------------|------|-------|-------|---------|---------|--------|-------|
| Dload | Upload | Total | Spent | Left | Speed | | | | | |
| 100 | 20480 | 100 | 20480 | 0 | 0 | 11257 | 0:00:01 | 0:00:01 | --::-- | 11258 |
| | | | | | | | | | | |

Select the Audit Store type where Log Forwarder(s) should send logs to.

[1] : Protegility Audit Store
 [2] : External Audit Store
 [3] : Protegility Audit Store + External Audit Store

Enter the no.:

- To select the Audit Store type, select any one of the following options:

Table 12-84: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

- Press ENTER.

When you select option *2* or *3*, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:

- Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

Do you want PepServer's log to be generated in a file? [yes or no]:

- To generate the logs for the PEP server in a file, type *yes*.

- Press ENTER.

The installer generates the installation files based on the options you selected and the prompt to select the upload type appears.

PepServer's log will be generated in a file.

Successfully generated Big Data Protector files for Azure Databricks in ./Installation_Files/

All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), Databricks Workspace or Unity Catalog Volume.

Select an option:

[1] : Upload the Installation Files now
 [2] : Upload the Installation Files manually later

Enter the no.:

The following snippet lists the files and directories generated by the configurator script in the `./Installation_Files` directory:

```
+++ curl
+++ Uploads
|   +++ BDP_Init_Script_9.1.0.0.x_<TAG>.sh
|   +++ BDP_Package_9.1.0.0.x_<TAG>.tgz
+++ upload_to_adls.sh
+++ upload_to_dbfs.sh
+++ upload_to_workspace.sh
+++ upload_to_unity_catalog_volume.sh
```

22. To manually upload the installation files using the helper script, type *2*.

If you select option *1*, to upload the installation files immediately, then the configurator script will prompt to enter the location to upload the installation files.

```
Select the location where you want to upload the BDP Installation Files:
[ 1 ] : Upload to DBFS
[ 2 ] : Upload to ADLS
[ 3 ] : Upload to Workspace
[ 4 ] : Upload to Unity Catalog Volume
```

Enter the no.:

Note: Depending on the location that you select, the configurator script will trigger the corresponding helper script to upload the installation files.

- For more information about uploading the installation files to DBFS, refer to the section [Uploading the Installation Files to DBFS](#).
- For more information about uploading the installation files to ADLS, refer to the section [Uploading the Installation Files to ADLS](#).
- For more information about uploading the installation files to the Workspace, refer to the section [Uploading the Installation Files to the Workspace Storage](#).
- For more information about uploading the installation files to the Unity Catalog Volume, refer to the section [Uploading the Installation Files to the Unity Catalog Volumes](#).

23. Press ENTER.

The configurator script provides the information about uploading the installation files to the required Azure Databricks storage.

```
For DBFS:
The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./Installation_Files/Uploads/ to DBFS root directory (dbfs:/)
Usage: ./upload_to_dbfs.sh
Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of
uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload
failure.

For ADLS:
The ./Installation_Files/upload_to_adls.sh script can be used to upload the files under ./Installation_Files/Uploads/ to Azure Data Lake Storage (ADLS).
Usage: ./upload_to_adls.sh

For Workspace:
The ./Installation_Files/upload_to_workspace.sh script can be used to upload the files
under ./Installation_Files/Uploads/ to Databricks' Workspace.
Usage: ./upload_to_workspace.sh
Warning: The upload_to_workspace.sh script uses Workspace REST API that has a limitation
of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload
failure.
```



```
For Unity Catalog Volume:  
The ./Installation_Files/upload_to_unity_catalog_volume.sh script can be used to upload  
the files under ./Installation_Files/Uploads/ to Unity Catalog Volume.  
Usage: ./upload_to_unity_catalog_volume.sh
```

12.6.12.5 Modifying the *pepserver.cfg* File

This section explains the process of modifying the *pepserver.cfg* file. You will be unable to modify the *pepserver.cfg* file when the cluster is running. Therefore, you must:

1. Extract the Big Data Protector archive generated by the configurator script.
2. Update the *pepserver.cfg* file.
3. Repackage the installation files.
4. Upload the updated Big Data Protector archive to *ABFSS* using the helper script.

Attention: Modifying the *pepserver.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepserver.cfg* file.

► To modify the *pepserver.cfg* file:

1. To navigate to the *./Installation_Files/Uploads/* directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

2. To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the *pepserver.cfg* file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepserver.cfg* file.
6. Modify the *pepserver.cfg* file according to requirements.
7. Save the changes to the *pepserver.cfg* file.
8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --  
owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. To upload the updated Big Data Protector archive to *Workspace*, use the *upload_to_workspace.sh* script.

Note: For more information about uploading the files to Workspace using the script, refer to section [Uploading the Installation Files to the Workspace Storage](#).

12.6.12.6 Uploading the Installation Files to the Workspace Storage

► To upload the installation files to the Workspace:

1. To execute the script to upload the Big Data Protector installation files to the Databricks workspace, run the following command:

```
./upload_to_workspace.sh
```

2. Press ENTER.

The prompt to continue with the upload process appears.

```
*****
This Script will upload the files generated under ./Uploads/ to Workspace.
*****
```

```
Do you wish to continue? [yes or no]:
```

3. To proceed with the upload, type *yes*.

4. Press ENTER.

The execution of the script starts where the script checks for the presence of the Big Data Protector installation package and the initialization script and the prompt to enter the Workspace path appears.

```
Execution of upload_to_workspace.sh script is started.

Checking if all required files are present under ./Uploads/
BDP_Package*.tgz file is present under ./Uploads/
BDP_Init_Script*.sh file is present under ./Uploads/ .

All required files are present under ./Uploads/ .

Enter the Workspace Path for uploading the files.
Format of the path: /Users/<directory_structure(if any)>
Example of the path: /Users/abc@xyz.com/mydirectory1/mydirectory2/mydirectory3
Here, "mydirectory3" is the directory where the BDP files are expected to be uploaded.
NOTE:
Whitespaces in the path are not allowed.
Enter the Workspace Path:
```

5. Enter the path of the directory in the Workspace where you want to upload the Big Data Protector installation package and the initialization script.

6. Press ENTER.

The prompt to enter the Databricks URL appears.

```
Enter the Databricks Workspace Instance URL (E.g.: https://
adb-1234567890123456.7.azuredatabricks.net):
```

7. Enter the URL that you have used to access the Databricks environment.

8. Press ENTER.

The prompt to enter the access token appears.

```
Enter Access Token:
```

9. Enter the access token.

10. Press ENTER.



The upload script splits the Big Data Protector installation package into fragments and uploads the files to the location mentioned in the *Workspace Path* prompt.

```

Starting upload of BDP_Package_9.1.0.0.x_<TAG>.tgz's fragments to Workspace. This may
take some time.

Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_00 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_00 file to Workspace.

Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_01 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_01 file to Workspace.

Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_02 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_02 file to Workspace.

Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_03 file to Workspace.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz_03 file to Workspace.

Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz's fragments to Workspace.

Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Workspace.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Workspace.

Successfully uploaded all files to Workspace.

To use the BDP_Init_Script_9.1.0.0.x_<TAG>.sh file, which is just uploaded to Workspace,
as an Init Script in Databricks Cluster, follow the below steps:
1. Create 1 secret in a Databricks-backed secret scope for storing the
"PTY_WORKSPACE_ACCESS_TOKEN".

2. Add the below environment variable in the "Environment variables" section of the
"Spark" tab of the "Advanced Options" menu of the "Configuration" tab of your Databricks
Cluster:
2a. PTY_WORKSPACE_ACCESS_TOKEN={{secrets/<scope_name>/<secret_name>}}
here, replace <scope_name> with the name of your secret scope and <secret_name> with the
name of your "PTY_WORKSPACE_ACCESS_TOKEN" secret.

3. Add the below Workspace Path in the "Init Scripts" tab of the "Advanced Options" menu
of the "Configuration" tab of your Databricks Cluster:
/<workspace_path>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh

```

12.6.12.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.12.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on a new cluster in the Azure Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/configure>.
2. On the cluster creation page, expand **Advanced Options**.



- On the **Spark** tab, add the following Spark configurations, such as, the key and value, separated by a space.

Table 12-85: Spark Configurations

| Keys | Values |
|--|---|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark configuration for reference.

Figure 12-347: Spark Configurations

- Under **Environment variables**, add the workspace access token in the following format:

```
PTY_WORKSPACE_ACCESS_TOKEN={{secrets/<scope_name>/<secret_name>}}
```

Replace the `<scope_name>` with the name of your secret scope and the `<secret_name>` with the name of your `PTY_WORKSPACE_ACCESS_TOKEN` secret.

The following image represents a sample Spark configuration and the environment variables for reference.

Figure 12-348: Spark Configuration and Environment Variables

- Click the **Init Scripts** tab.

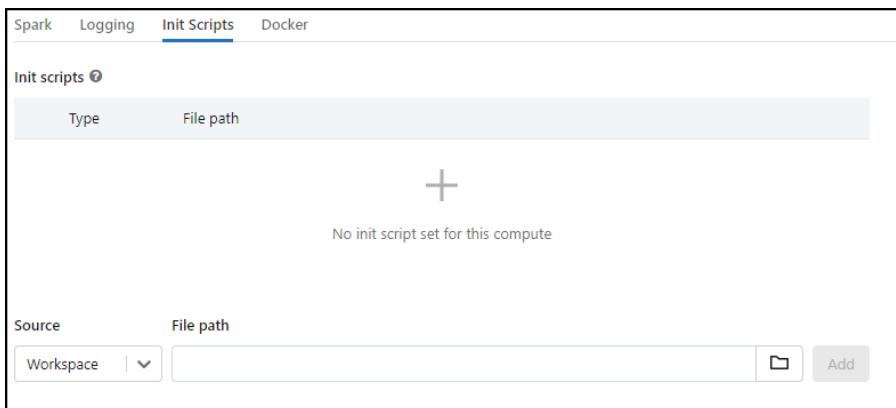


Figure 12-349: The Init Scripts Tab

6. From the **Source** list, select **Workspace**.
7. In the **File path** box, enter the location where you uploaded the initialization script to the *Workspace* using the helper script. For example, enter the path for the initialization script as <storage_location>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh
8. Click **Add**.
The location of the initialization script appears in the **File path** column.



Figure 12-350: Initialization Script Path

9. Click **Create compute**.
The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.12.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the Azure Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

- To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/clusters-manage##edit-a-cluster>.

2. On the cluster configuration page, click the **Advanced Options** toggle.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format that is separated by a space.

Table 12-86: Spark Configurations

| Keys | Values |
|--|---|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark configuration for reference.

```

spark.driver.extraJavaOptions -Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm
spark.executor.extraJavaOptions -Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin

```

Figure 12-351: Spark Configurations

5. Under **Environment variables**, add the workspace access token in the following format:

```
PTY_WORKSPACE_ACCESS_TOKEN={{secrets/<scope_name>/<secret_name>}}
```

Replace the `<scope_name>` with the name of your secret scope and the `<secret_name>` with the name of your `PTY_WORKSPACE_ACCESS_TOKEN` secret.

The following image represents a sample Spark configuration and the environment variables for reference.

```

PYSPARK_PYTHON=/databricks/python3/bin/python3
PTY_WORKSPACE_ACCESS_TOKEN={{secrets/bdp_doc/bdp-doc-key}}

```

Figure 12-352: Spark Configuration and Environment Variables

6. Click the **Init Scripts** tab.

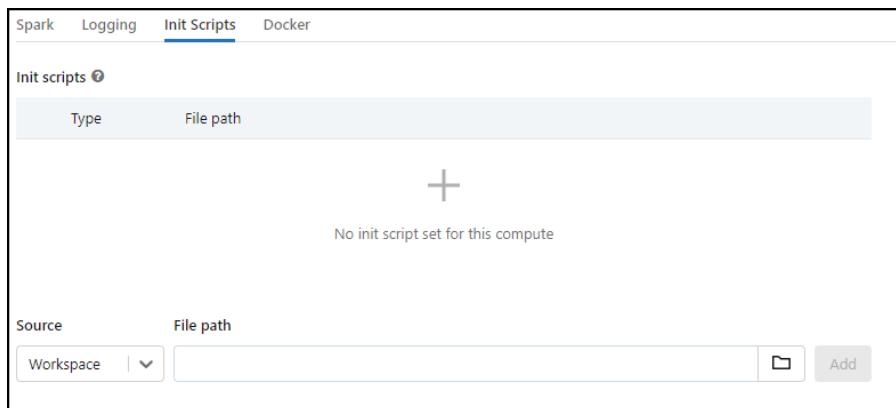


Figure 12-353: The Init Scripts Tab

7. From the **Source** list, select **Workspace**.
8. In the **File path** box, enter the location where you uploaded the initialization script to the *Workspace* using the helper script. For example, enter the path for the initialization script as <storage_location>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh
9. Click **Add**.
The location of the initialization script appears in the **File path** column.

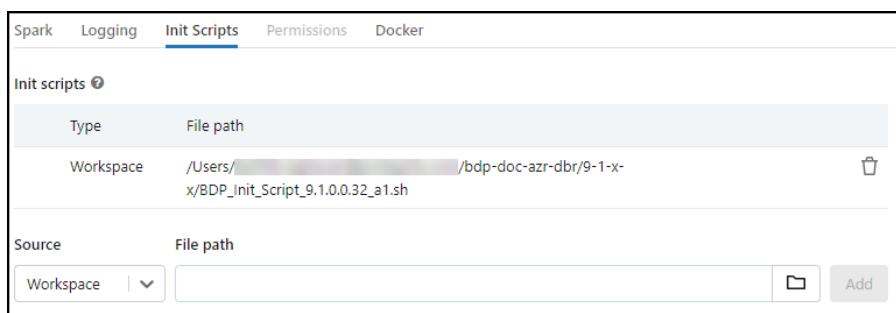


Figure 12-354: Initialization Script Path

10. To save the changes and restart the cluster, click **Confirm**.
The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.12.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs require registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

OK

In the Unity Catalog environment, the permanent registration for Hive UDFs fails for a custom catalog. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.

For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to use the permanently registered Hive UDFs in the Unity Catalog environment in a custom catalog.

The following examples illustrate protect and unprotect operations in the tables stored under a custom catalog from the Unity Catalog environment by referring them using their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT hive_metastore.default.ptyProtectStr(col1,
'Token_Alphanumeric') FROM dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table



- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.13 Installing the Big Data Protector on an Azure Databricks Cluster using the Unity Catalog Volumes

This section describes the steps for installing the Big Data Hive and Spark Protector on the Azure Databricks platform using the Databricks Unity Catalog volumes.

Warning: This build is designed to work only for "single-user" access mode. This build should not be deployed in "shared" and "No isolation shared" access mode environments.

12.6.13.1 Verifying the Prerequisites for Installing the Big Data Protector on Azure Databricks Cluster using the Unity Catalog Volumes

Ensure that the following prerequisites are met, before installing the Big Data Protector on the Azure Databricks platform:

- The user should have a valid Azure account.
- The user should have access to the Azure Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-87: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|---|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are |



| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|---|--|
| | | | | also written to localhost through port <i>15780</i> . The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port <i>16700</i> . |

- You have created a Databricks personal access token.

Note: For more information about creating the Databricks personal access token, refer to <https://learn.microsoft.com/en-us/azure/databricks/dev-tools/auth#-azure-databricks-personal-access-token-authentication>.

- You have created the Databricks volume to upload the files.

Note: For more information about creating volumes, refer to <https://learn.microsoft.com/en-us/azure/databricks/connect/unity-catalog/volumes>.

12.6.13.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using the Unity Catalog Volume

This section describes a brief overview of the installation workflow that the user needs to perform to install the Big Data Protector on the Azure Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the Azure Databricks platform.

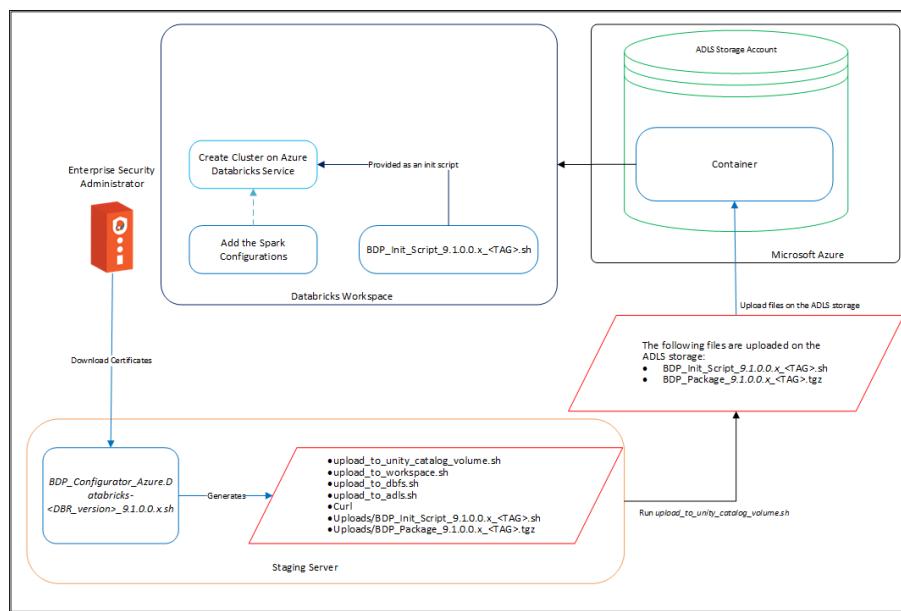


Figure 12-355: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

Note: For more information about extracting the Big Data Protector package, refer to the section [Extracting the Big Data Protector Package](#).

- Run the *BDP_Configurator_Azure.Databricks-<DBR_version>.9.1.0.0.x.sh* script to download the certificates from the ESA and create the following files in the */Installation_Files*/directory:

- *upload_to_unity_catalog_volume.sh*
- *upload_to_workspace.sh*
- *upload_to_adls.sh*
- *upload_to_dbfs.sh*
- *curl*
- *Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* script, refer to the section [Running the Configurator Script](#).

- c. Run the *upload_to_unity_catalog_volume.sh* script to upload the following files to the *Unity Catalog Volume* storage:
 - *BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
 - *BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *upload_to_unity_catalog_volume.sh* script, refer to the section [Uploading the Installation Files to the Unity Catalog Volume](#).

2. On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script during the cluster creation with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to the section [Installing the Big Data Hive and Spark Protector](#).

12.6.13.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files that you must upload to ABFSS.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because the configurator script appends the same IP address or hostname in the *pepper.cfg* file.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
3. To extract the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz
```

4. Press ENTER.

The command extracts the *BBDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file.

12.6.13.4 Running the Configurator Script

You must execute the Big Data Protector configurator script to download the certificates from the ESA and create the installation files required to install the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. To execute the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command:

```
./BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh
```

2. Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks

Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.

4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), or Databricks Workspace. This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

```
Enter a Unique Alpha-Numeric Tag:
```

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

11. Enter the listening port for the ESA.

12. Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
```

```
Enter host password for user '<user_name>':
```

| % Total | % Received | % Xferd | Average Speed | Time Dload | Time Upload | Total | Time Spent | Time Left | Current Speed | |
|---------|------------|---------|---------------|------------|-------------|-------|------------|-----------|---------------|----------------|
| 100 | 20480 | 100 | 20480 | 0 | 0 | 11257 | 0 | 0:00:01 | 0:00:01 | --:--:-- 11258 |

Select the Audit Store type where Log Forwarder(s) should send logs to.

[1] : Protegility Audit Store

[2] : External Audit Store

[3] : Protegility Audit Store + External Audit Store

```
Enter the no.:
```

17. To select the Audit Store type, select any one of the following options:

Table 12-88: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type 1. If you enter 1, then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type 2. If you enter 2, then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type 3. If you enter 3, then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and |

| Option | Description |
|--------|---|
| | <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

18. Press ENTER.

When you select option *2* or *3*, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

19. Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

20. To generate the logs for the PEP server in a file, type *yes*.

21. Press ENTER.

The installer generates the installation files based on the options you selected and the prompt to select the upload type appears.

```
PepServer's log will be generated in a file.
```

```
Successfully generated Big Data Protector files for Azure Databricks in ./Installation_Files/
```

```
All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), Databricks Workspace or Unity Catalog Volume.
```

```
Select an option:
```

```
[ 1 ] : Upload the Installation Files now
[ 2 ] : Upload the Installation Files manually later
```

```
Enter the no.:
```

The following snippet lists the files and directories generated by the configurator script in the *./Installation_Files* directory:

```
+++ curl
+++ Uploads
|   --- BDP_Init_Script_9.1.0.0.x_<TAG>.sh
|   --- BDP_Package_9.1.0.0.x_<TAG>.tgz
+++ upload_to_adls.sh
+++ upload_to_dbfs.sh
+++ upload_to_workspace.sh
+++ upload_to_unity_catalog_volume.sh
```

22. To manually upload the installation files using the helper script, type *2*.

If you select option *1*, to upload the installation files immediately, then the configurator script will prompt to enter the location to upload the installation files.

```
Select the location where you want to upload the BDP Installation Files:
[ 1 ] : Upload to DBFS
[ 2 ] : Upload to ADLS
[ 3 ] : Upload to Workspace
[ 4 ] : Upload to Unity Catalog Volume
```

Enter the no.:

Note: Depending on the location that you select, the configurator script will trigger the corresponding helper script to upload the installation files.

- For more information about uploading the installation files to DBFS, refer to the section [Uploading the Installation Files to DBFS](#).
- For more information about uploading the installation files to ADLS, refer to the section [Uploading the Installation Files to ADLS](#).
- For more information about uploading the installation files to the Workspace, refer to the section [Uploading the Installation Files to the Workspace Storage](#).
- For more information about uploading the installation files to the Unity Catalog Volume, refer to the section [Uploading the Installation Files to the Unity Catalog Volumes](#).

23. Press ENTER.

The configurator script provides the information about uploading the installation files to the required Azure Databricks storage.

```
For DBFS:  
The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./  
Installation_Files/Uploads/ to DBFS root directory (dbfs:/)  
Usage: ./upload_to_dbfs.sh  
Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of  
uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload  
failure.  
  
For ADLS:  
The ./Installation_Files/upload_to_adls.sh script can be used to upload the files under ./  
Installation_Files/Uploads/ to Azure Data Lake Storage (ADLS).  
Usage: ./upload_to_adls.sh  
  
For Workspace:  
The ./Installation_Files/upload_to_workspace.sh script can be used to upload the files  
under ./Installation_Files/Uploads/ to Databricks' Workspace.  
Usage: ./upload_to_workspace.sh  
Warning: The upload_to_workspace.sh script uses Workspace REST API that has a limitation  
of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload  
failure.  
  
For Unity Catalog Volume:  
The ./Installation_Files/upload_to_unity_catalog_volume.sh script can be used to upload  
the files under ./Installation_Files/Uploads/ to Unity Catalog Volume.  
Usage: ./upload_to_unity_catalog_volume.sh
```

12.6.13.5 Modifying the *pepperserver.cfg* File

This section explains the process of modifying the *pepperserver.cfg* file. You will be unable to modify the *pepperserver.cfg* file when the cluster is running. Therefore, you must:

1. Extract the Big Data Protector archive generated by the configurator script.
2. Update the *pepperserver.cfg* file.
3. Repackage the installation files.
4. Upload the updated Big Data Protector archive to *ABFSS* using the helper script.

Attention: Modifying the *pepperserver.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepperserver.cfg* file.

► To modify the *pepperserver.cfg* file:

- To navigate to the `./Installation_Files/Uploads/` directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

- To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

- To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

- Navigate to the directory that contains the `pepperserver.cfg` file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

- Using an editor, open the `pepperserver.cfg` file.

- Modify the `pepperserver.cfg` file according to requirements.

- Save the changes to the `pepperserver.cfg` file.

- To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

- To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

- To upload the updated Big Data Protector archive to `Workspace`, use the `upload_to_workspace.sh` script.

Note: For more information about uploading the files to Workspace using the script, refer to section [Uploading the Installation Files to the Workspace Storage](#).

12.6.13.6 Uploading the Installation Files to the Unity Catalog Volume

You must upload the installation files generated by the configurator script to the Azure Data Lake Storage Gen2 (ADLS) file system (Unity Catalog Volume) of the Azure Databricks workspace. You can use the `upload_to_unity_catalog_volume.sh` script to upload the installation files to the ADLS file system.

Before you begin

Ensure that the following prerequisites are met before installing the Big Data Protector on the Azure Databricks platform:

- You have created the Databricks volume to upload the files.

Note: For more information about creating volumes, refer to <https://learn.microsoft.com/en-us/azure/databricks/connect/unity-catalog-volumes>

- You have an Azure Data Lake Storage Gen2 account with a container and an optional directory inside the container.

Note: For more information about creating an Azure Data Lake Storage Gen2 account, refer to <https://learn.microsoft.com/en-us/azure/storage/blobs/create-data-lake-storage-account>

- You have a Storage Account SAS token generated for the Azure Data Lake Storage Gen2 account.

Note: The `upload_to_adls.sh` helper script requires you to generate the Storage Account SAS Token for Azure Storage REST API authentication.



For more information about Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/azure/storage/common/storage-sas-overview>.

For more information about creating a Storage Account SAS Token, refer to <https://learn.microsoft.com/en-us/rest/api/storageservices/create-account-sas>.

- You have the shared key generated for the Azure Data Lake Storage Gen2 account.

Note:

For more information about the shared key, refer to <https://learn.microsoft.com/en-us/rest/api/storageservices/authorize-with-shared-key>.

► To upload the files using the helper script:

1. To execute the helper script, run using the following command.

```
./upload_to_unity_catalog_volume.sh
```

2. Press ENTER.

The prompt to continue with the upload appears.

```
*****
This Script will upload the files generated under ./Uploads/ to Unity Catalog Volume.
*****
Do you wish to continue? [yes or no]:
```

3. To continue with the upload, type *yes*.

4. Press ENTER.

The configurator script verifies whether all the required files are available and the prompt to enter the Databricks URL appears.

```
Execution of upload_to_unity_catalog_volume.sh script is started.

Checking if all required files are present under ./Uploads/
BDP_Package*.tgz file is present under ./Uploads/
BDP_Init_Script*.sh file is present under ./Uploads/ .

All required files are present under ./Uploads/ .

Enter the Databricks Workspace Instance URL (E.g.: https://
adb-1234567890123456.7.azuredatabricks.net):
```

5. Enter the Databricks URL.

6. Press ENTER.

The prompt to enter the access token for Databricks appears.

```
Enter Access Token for Databricks:
```

7. Enter the access token.

8. Press ENTER.



The prompt to enter the fully qualified name for the Unity Catalog volume appears.

```
Enter a valid Unity Catalog Volume fully qualified name using the syntax:  
<catalog>.<schema>.<volume>  
E.g.: mycatalog.myschema.myvolume  
Enter Unity Catalog Volume fully qualified name:
```

9. Enter the name of the Unity Catalog volume in the specified syntax.

10. Press ENTER.

The script retrieves the Unity Catalog volume details using the Databricks REST API and the prompt to select the authorization methods to upload the files appears.

```
Unity Catalog Volume fully qualified name-> <catalog_name>.<schema_name>.<volume_name>  
Fetching Unity Catalog Volume details using Databricks REST API...  
  
% Total % Received % Xferd Average Speed Time Time Current  
Dload Upload Total Spent Left Speed  
100 622 0 622 0 0 4975 0 --:-- --:-- --:-- 5016  
Successfully fetched Unity Catalog Volume details from Databricks.  
  
Storage Location: abfss://<container_name>@<storage_account_name>.dfs.core.windows.net/  
<directory_structure>  
Storage Account Name: <storage_account_name>  
Container Name: <container_name>  
Directory Structure: <directory_structure>  
  
Select one of the following authorization methods for uploading BDP files to ADLS:  
[1] -> SAS Token (Shared Access Signature)  
[2] -> Shared Key (Storage Account Key)  
[ 1 or 2 ]:
```

11. To upload the files to ADLS using the SAS Token, type *1*.

12. Press ENTER.

The prompt to enter the SAS token for the Azure storage account appears.

```
Enter SAS Token for Azure Storage Account:
```

If you select the *Shared Key* option to upload the files to ADLS, then the prompt to enter the shared key for the Azure storage account appears.

```
Enter Shared Key for Azure Storage Account:
```

13. Depending on the authorization method you select to upload the files to ADLS, enter the SAS token or the shared key for the Azure storage account.

14. Press ENTER.

The helper script uploads the following files in the *./Installation_Files/Uploads/* directory to ADLS using the ABFSS REST API:

```
- BDP_Init_Script_9.1.0.0.x_<TAG>.sh  
- BDP_Package_9.1.0.0.x_<TAG>.tgz
```

A sample output of uploading the installation files to ADLS using the SAS token authorization is listed below.

```
Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Unity Catalog Volume's ADLS  
location.  
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Unity Catalog Volume's ADLS  
location.
```

```
Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Unity Catalog Volume's ADLS
```

```

location.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Unity Catalog Volume's ADLS
location.

Successfully uploaded all files to Unity Catalog Volume.

To use BDP_Init_Script_9.1.0.0.x_<TAG>.sh file uploaded to Unity Catalog Volume as an
init script in Databricks Cluster, Add the following Unity Catalog Volume path in the
"Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the
Databricks cluster page:
/Volumes/<catalog_name>/<schema_name>/<volume_name>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh

```

A sample output of uploading the installation files to ADLS using the shared key authorization is listed below.

```

Started upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Unity Catalog Volume's ADLS
location.
#####
Finished upload of BDP_Package_9.1.0.0.x_<TAG>.tgz file to Unity Catalog Volume's ADLS
location.

Started upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Unity Catalog Volume's ADLS
location.
#####
Finished upload of BDP_Init_Script_9.1.0.0.x_<TAG>.sh file to Unity Catalog Volume's ADLS
location.

Successfully uploaded all files to Unity Catalog Volume.
To use BDP_Init_Script_9.1.0.0.x_<TAG>.sh file uploaded to Unity Catalog Volume as an
init script in Databricks Cluster, Add the following Unity Catalog Volume path in the
"Init Scripts" tab of the "Advanced Options" menu under the "Configuration" tab of the
Databricks cluster page:
/Volumes/<catalog_name>/<schema_name>/<volume_name>/BDP_Init_Script_9.1.0.0.x_<TAG>.sh

```

12.6.13.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
- Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.13.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector in the Azure Databricks workspace on a new Databricks cluster.

Note: For more information about updating the *pepserver.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/configure>.
2. On the cluster creation page, expand **Advanced Options**.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value, separated by a space.

Table 12-89: Spark Configurations

| Keys | Values |
|--|---|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegility.spark.PtyExecSparkPlugin</i> |



The following image represents a sample Spark configuration for reference:



Figure 12-356: Sample Spark Configuration

- Click the **Init Scripts** tab.

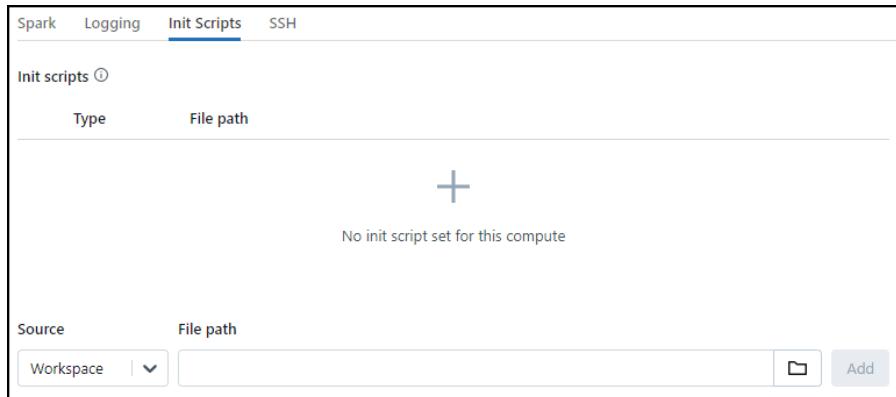


Figure 12-357: The Init Scripts Tab

- From the **Source** list, select **Volumes**.
- In the **File path** box, enter the location where you uploaded the initialization script to the Unity Catalog Volume using the helper script.
For example, enter the path for the initialization script as */Volumes/<catalog_name>/<schema_name>/<volume_name>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*.

- Click **Add**.

The location of the initialization script appears in the **File path** column.

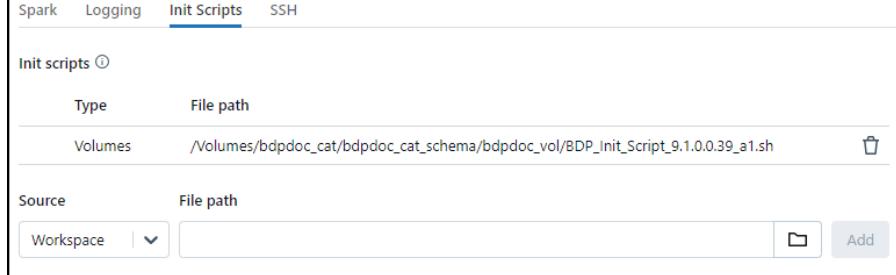


Figure 12-358: Init Script Path

- Click **Create compute**.

The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.13.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the Azure Databricks workspace.

Note: For more information about modifying the *pepserver.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/clusters-manage#--edit-a-cluster>.
2. On the cluster configuration page, expand **Advanced Options**.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. On the **Spark** tab, add the following Spark configurations, such as, the key and value, separated by a space.

Table 12-90: Spark Configurations

| Keys | Values |
|--|--|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegility.spark.PtyExecSparkPlugin</i> |

The following image represents a sample Spark configuration for reference:



Figure 12-359: Sample Spark Configuration

5. Click the **Init Scripts** tab.

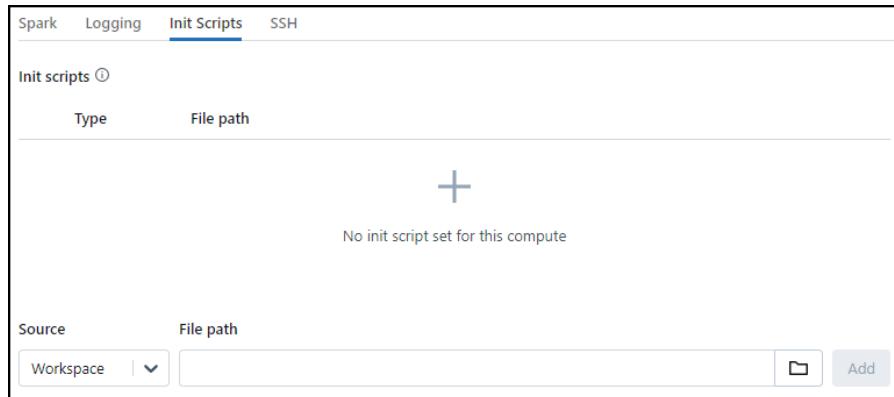


Figure 12-360: The Init Scripts Tab

6. From the **Source** list, select **Volumes**.
7. In the **File path** box, enter the location where you uploaded the initialization script to the Unity Catalog Volume using the helper script.

For example, enter the path for the initialization script as `/Volumes/<catalog_name>/<schema_name>/<volume_name>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh`.

8. Click **Add**.

The location of the initialization script appears in the **File path** column.



Figure 12-361: Init Script Path

9. To save the changes and restart the cluster, click **Confirm**.

The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is `user1@example.com`, then the user `user1` must be present in the ESA policy.

12.6.13.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs require registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
  'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
```

```
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';

OK
```

In the Unity Catalog environment, the permanent registration for Hive UDFs fails for a custom catalog. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.

For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to use the permanently registered Hive UDFs in the Unity Catalog environment in a custom catalog.

The following examples illustrate protect and unprotect operations in the tables stored under a custom catalog from the Unity Catalog environment by referring them using their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT hive_metastore.default.ptyProtectStr(col1,
'Token_Alphanumeric') FROM dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(col1, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table



12.6.14 Installing the Big Data Protector on an Azure Databricks Cluster Using DBFS

This section describes the steps for installing the Big Data Hive and Spark Protector on the Azure Databricks platform using the Databricks File System (DBFS).

Warning: You will be unable to use the initialization scripts on DBFS, starting 01-December-2023. Databricks is ending support for initialization scripts on DBFS and the feature will not function after that date.

For more information, refer to <https://learn.microsoft.com/en-us/azure/databricks/init-scripts/legacy-global>.

Warning: This build is designed to work only for "single-user" access mode. This build should not be deployed in "shared" and "No isolation shared" access mode environments.

12.6.14.1 Verifying the Prerequisites for Installing the Big Data Protector on an Azure Databricks Cluster using DBFS

Ensure that the following prerequisites are met, before installing the Big Data Protector on the Azure Databricks platform:

- The user should have a valid Microsoft Azure account.
- The user should have access to the Azure Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.
- The following table depicts the list of ports that are configured on the ESA and the nodes in the cluster, which will run the Big Data Protector.

Table 12-91: List of Ports for the Big Data Protector

| Destination Port No. | Protocol | Source | Destination | Description |
|----------------------|----------|--|--|--|
| 8443 | TCP | PEP server on the Big Data Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download the policy. |
| 9200 | | Log Forwarder on the Big Data Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all the logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Big Data Protector cluster node | Log Forwarder on the Big Data Protector cluster node | The Big Data Protector writes Audit Logs to localhost through port 15780. The PEP server Application Logs are also written to localhost through port 15780. The Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Big Data Protector cluster node | PEP server on the Big Data Protector cluster node | The DPS Admin client tool uses the localhost port 16700. |

- You have created a Databricks personal access token.

Note: For more information about creating the Databricks personal access token, refer to <https://learn.microsoft.com/en-us/azure/databricks/dev-tools/auth#-azure-databricks-personal-access-token-authentication>.

12.6.14.2 Understanding the Installation Workflow of the Big Data Protector on Azure Databricks using DBFS

This section describes a brief overview of the installation workflow that you must follow to install the Big Data Protector on the Azure Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the Azure Databricks platform.

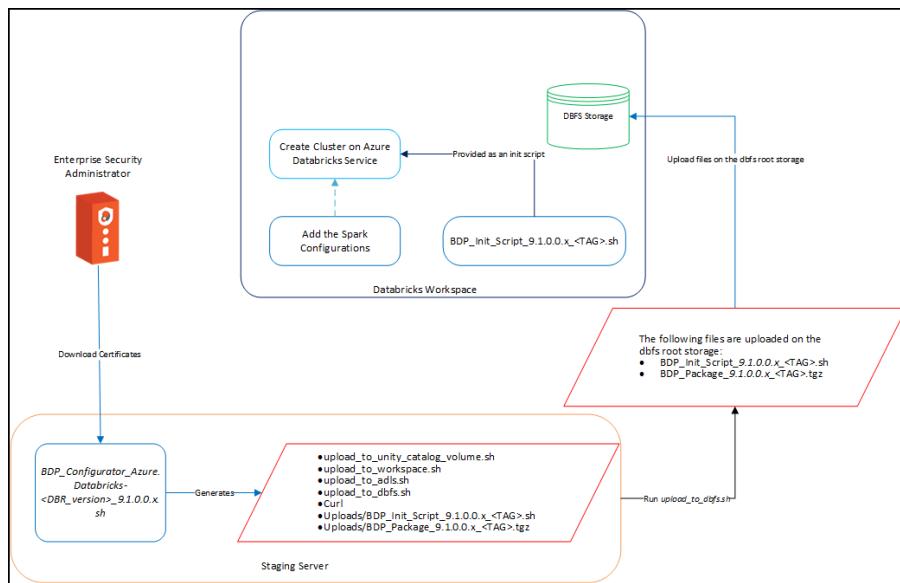


Figure 12-362: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector Installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- Run the `BDP_Configurator_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.sh` script to download the certificates from the ESA and create the following files in the `./Installation_Files/` directory:

- `upload_to_unity_catalog_volume.sh`
- `upload_to_workspace.sh`
- `upload_to_adls.sh`
- `upload_to_dbfs.sh`
- `curl`
- `Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh`
- `Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz`

Note: For more information about running the `BDP_Configurator_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.sh` script, refer to section [Running the Configurator Script](#).

- Run the `upload_to_dbfs.sh` script to upload the following files on the `dbfs` root storage:

- `BDP_Init_Script_9.1.0.0.x_<TAG>.sh`
- `BDP_Package_9.1.0.0.x_<TAG>.tgz`

Note: For more information about running the `upload_to_dbfs.sh` script, refer to section [Uploading the Installation Files to DBFS](#).

- On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* during cluster creation with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector on a New Cluster](#).

12.6.14.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector installation files.

► To extract the Big Data Protector Configurator file from the installation package:

- Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because the configurator scripts appends the same IP address or hostname in the *pepserver.cfg* file.

- Download the *BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-<DBR_version>-64_9.1.0.0.x.tgz* build to the system.
- To extract the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file from the Big Data Protector installation package, run the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_Azure.Databricks-
<DBR_version>-64_9.1.0.0.x.tgz
```

- Press ENTER.

The command extracts the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* file.

12.6.14.4 Running the Configurator Script

You must execute the Big Data Protector configurator script to download certificates from the ESA, and create the installation files for the Big Data Protector.

► To run the Big Data Protector Configurator Script:

- To execute the *BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh* script from the directory where it is extracted, run the following command:

```
./BDP_Configurator_Azure.Databricks-<DBR_version>_9.1.0.0.x.sh
```

- Press ENTER.

The prompt to continue the installation of the Big Data Protector appears.

```
*****
* Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks
Do you wish to continue? [yes or no]:
```

3. To continue the installation, type *yes*.
4. Press ENTER.

The prompt to enter the installation directory appears.

```
*****
Welcome to the Big Data Protector Service Configurator Wizard
*****
This will create the Big Data Protector Installation files for Azure Databricks

Do you wish to continue? [yes or no]: yes

Big Data Protector service configurator started...

Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

5. Enter the location where you want to install the Big Data Protector.

6. Press ENTER.

The prompt to enter a unique alphanumeric tag appears.

```
Enter the directory path on Cluster nodes where you want to install Protegility products.
[default:- /opt/protegility]:
```

The BDP Installation files that will be generated needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), or Databricks Workspace. This script will append a unique tag to the generated files' names to distinguish them from other BDP files.

```
Enter a Unique Alpha-Numeric Tag:
```

7. Enter a unique alphanumeric tag.

8. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Enter ESA Hostname or IP Address:
```

9. Enter the hostname or the IP address of the ESA.

10. Press ENTER.

The prompt to enter the listening port for the ESA host appears.

```
Enter ESA host listening port [8443]:
```

11. Enter the listening port for the ESA.

12. Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

13. Enter the user name to connect to the ESA.

14. Press ENTER.

The prompt to enter the password appears.

```
Extracting files...

Fetching Certificates.....
Enter host password for user '<user_name>':
```

15. Enter the password.

16. Press ENTER.

The installer downloads the certificates from the ESA and the prompt to select the Audit Store type appears.

```
Extracting files...
```

```
Fetching Certificates.....
Enter host password for user '<user_name>':
  % Total    % Received % Xferd  Average Speed   Time   Time     Time  Current
               Dload  Upload Total   Spent   Left Speed
100 20480  100 20480     0      0  11257       0  0:00:01  0:00:01  --::-- 11258
```

Select the Audit Store type where Log Forwarder(s) should send logs to.

```
[ 1 ] : Protegility Audit Store
[ 2 ] : External Audit Store
[ 3 ] : Protegility Audit Store + External Audit Store
```

Enter the no.:

- To select the Audit Store type, select any one of the following options:

Table 12-92: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type 1. If you enter 1, then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliances. |
| 2 | To use an external audit store, type 2. If you enter 2, then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type 3. If you enter 3, then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the External Audit Store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

- Press ENTER.

When you select option 2 or 3, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

- Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

- To generate the logs for the PEP server in a file, type yes.

- Press ENTER.

The installer generates the installation files based on the options you selected and the prompt to select the upload type appears.

```
PepServer's log will be generated in a file.
```

```
Successfully generated Big Data Protector files for Azure Databricks in ./Installation_Files/
```

```
All the generated files under ./Installation_Files/Uploads/ needs to be uploaded to either DBFS root directory (dbfs:/), Azure Data Lake Storage (ADLS), Databricks Workspace or Unity Catalog Volume.
```

Select an option:

```
[ 1 ] : Upload the Installation Files now  
[ 2 ] : Upload the Installation Files manually later
```

Enter the no.:

The following snippet lists the files and directories generated by the configurator script in the *./Installation_Files* directory:

```
+-- curl  
+-- Uploads  
|   +-- BDP_Init_Script_9.1.0.0.x_<TAG>.sh  
|   +-- BDP_Package_9.1.0.0.x_<TAG>.tgz  
+-- upload_to_adls.sh  
+-- upload_to_dbfs.sh  
+-- upload_to_workspace.sh  
+-- upload_to_unity_catalog_volume.sh
```

22. To manually upload the installation files using the helper script, type *2*.

If you select option *1*, to upload the installation files immediately, then the configurator script will prompt to enter the location to upload the installation files.

```
Select the location where you want to upload the BDP Installation Files:  
[ 1 ] : Upload to DBFS  
[ 2 ] : Upload to ADLS  
[ 3 ] : Upload to Workspace  
[ 4 ] : Upload to Unity Catalog Volume
```

Enter the no.:

Note: Depending on the location that you select, the configurator script will trigger the corresponding helper script to upload the installation files.

- For more information about uploading the installation files to DBFS, refer to the section [Uploading the Installation Files to DBFS](#).
- For more information about uploading the installation files to ADLS, refer to the section [Uploading the Installation Files to ADLS](#).
- For more information about uploading the installation files to the Workspace, refer to the section [Uploading the Installation Files to the Workspace Storage](#).
- For more information about uploading the installation files to the Unity Catalog Volume, refer to the section [Uploading the Installation Files to the Unity Catalog Volumes](#).

23. Press ENTER.

The configurator script provides the information about uploading the installation files to the required Azure Databricks storage.

```
For DBFS:  
The ./Installation_Files/upload_to_dbfs.sh script can be used to upload the files under ./Installation_Files/Uploads/ to DBFS root directory (dbfs:/)  
Usage: ./upload_to_dbfs.sh  
Warning: The upload_to_dbfs.sh script uses DBFS REST API that has a limitation of uploading a file by dividing it into chunks of 1MB. So, there is a chance of upload failure.
```

```
For ADLS:  
The ./Installation_Files/upload_to_adls.sh script can be used to upload the files under ./Installation_Files/Uploads/ to Azure Data Lake Storage (ADLS).  
Usage: ./upload_to_adls.sh
```

For Workspace:

The `./Installation_Files/upload_to_workspace.sh` script can be used to upload the files under `./Installation_Files/Uploads/` to Databricks' Workspace.

Usage: `./upload_to_workspace.sh`

Warning: The `upload_to_workspace.sh` script uses Workspace REST API that has a limitation of uploading a file by dividing it into chunks of 10MB. So, there is a chance of upload failure.

For Unity Catalog Volume:

The `./Installation_Files/upload_to_unity_catalog_volume.sh` script can be used to upload the files under `./Installation_Files/Uploads/` to Unity Catalog Volume.

Usage: `./upload_to_unity_catalog_volume.sh`

12.6.14.5 Modifying the `pepperserver.cfg` File

This section explains the process of modifying the `pepperserver.cfg` file. You will be unable to modify the `pepperserver.cfg` file when the cluster is running. Therefore, you must:

1. Extract the Big Data Protector archive generated by the configurator script.
2. Update the `pepperserver.cfg` file.
3. Repackage the installation files.
4. Upload the updated Big Data Protector archive to *DBFS* using the helper script.

Attention: Modifying the `pepperserver.cfg` file is an optional step. Exercise caution when modifying the contents of the `pepperserver.cfg` file.

► To modify the `pepperserver.cfg` file:

1. To navigate to the `./Installation_Files/Uploads/` directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

2. To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the Big Data Protector archive, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the `pepperserver.cfg` file in the path specified while extracting the contents of the Big Data Protector archive.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the `pepperserver.cfg` file.

6. Modify the `pepperserver.cfg` file according to requirements.

7. Save the changes to the `pepperserver.cfg` file.

8. To recreate the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```



- To upload the updated Big Data Protector archive to *DBFS*, use the *upload_to_dbfs.sh* script.

Note: For more information about uploading the files to DBFS using the script, refer to section *Uploading the Files using the Helper Script*.

12.6.14.6 Uploading the Installation Files to DBFS

You must upload the files generated by the configurator script to the DBFS root directory (*dbfs:/*) of the Azure Databricks workspace.

Caution: The *upload_to_dbfs.sh* helper script can be optionally used to upload the files. It uses the DBFS REST API, that has a limitation of uploading a file by dividing it into chunks of 1 MB. If you encounter any failure, then upload the files manually.

Note: The *upload_to_dbfs.sh* helper script requires an API authentication token that needs to be generated by the user.

For more information about generating a personal access token on Azure Databricks Workspace UI, refer to <https://learn.microsoft.com/en-us/azure/databricks/dev-tools/auth>.

► To upload the files using the helper script:

- To execute the helper script, run the following command:

```
./upload_to_dbfs.sh
```

- Press ENTER.

The prompt to continue with the upload appears.

```
*****
*****  
This Script will upload the files generated under ./Uploads/ to DBFS root directory.
```

Note: Make sure that you have a valid API Access Token generated from the Azure Databricks Web UI.

Do you wish to continue? [yes or no]:

- To continue upload, type *yes*.

- Press ENTER.

The script verifies whether the installation package and the initialization script are available in the *./Uploads/* directory and the prompt to enter the Databricks instance URL appears.

Upload to DBFS Script started...

Checking if all required files are present under ./Uploads/

All files present

Enter the Databricks Instance URL (E.g.: https://adb-1234567890123456.7.azure.databricks.net):

- Enter the Azure Databricks instance URL of the Azure Databricks workspace.



For more information about the Azure Databricks instance URL, refer to <https://learn.microsoft.com/en-us/azure/databricks/workspace/workspace-details#workspace-instance-names-urls-and-ids>.

6. Press ENTER.

The prompt to enter the API access token appears.

Enter API Access Token:

7. Enter the access token that was generated previously in the same workspace.

8. Press ENTER.

The helper script uploads the following files in the `./Installation_Files/Uploads/` directory to the DBFS root directory using the DBFS REST API:

- BDP_Init_Script_9.1.0.0.x_<TAG>.sh
- BDP_Package_9.1.0.0.x_<TAG>.tgz

```
Started upload of ./Uploads/.tmp/BDP_Init_Script_9.1.0.0.x_<TAG>.sh to DBFS.  
#####
##### 100.0%  
#####
##### 100.0%  
#####
##### 100.0%  
Finished upload of ./Uploads/.tmp/BDP_Init_Script_9.1.0.0.x_<TAG>.sh to DBFS.
```

Successfully uploaded all files to DBFS root 'dbfs://'

Note: The Big Data Protector installation files must be uploaded only on the DBFS root directory (`dbfs:/`) of the Azure Databricks workspace. There might be a possibility that similar file names might be present. To distinguish the installation files and prevent overwriting it, they are tagged with alpha-numeric tags provided by the user.

12.6.14.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector:

- Installing the Big Data Hive and Spark Protector on a New Cluster
 - Installing the Big Data Hive and Spark Protector on an Existing Cluster

12.6.14.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on a new cluster in the Azure Databricks workspace.

Note: For more information about updating the *pepserver.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

1. To create a cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/configure>.
2. On the cluster creation page, expand **Advanced Options**.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value separated by a space.

Table 12-93: Spark Configurations

| Keys | Values |
|--|---|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegility.spark.PtyExecSparkPlugin</i> |

The following image represents a sample Spark configuration for reference.

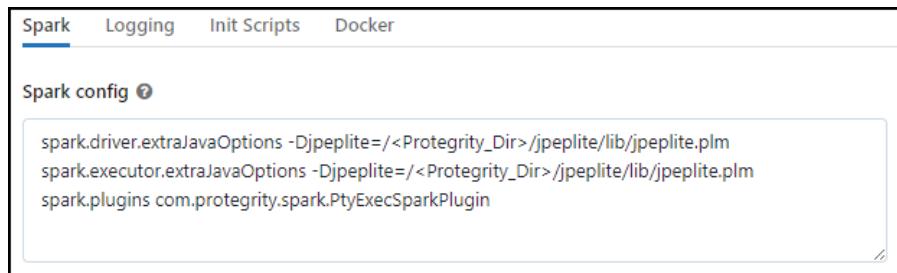


Figure 12-363: Spark Configurations

4. Click the **Init Scripts** tab.

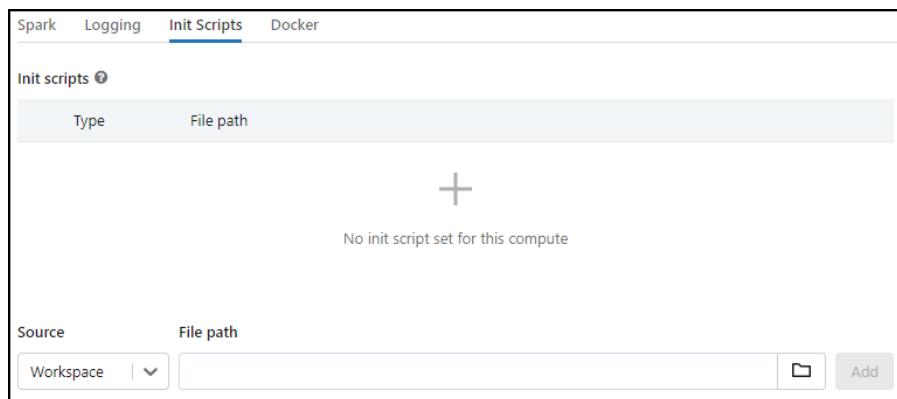


Figure 12-364: The Init Scripts Tab

5. From the **Source** list, select **DBFS**.
6. In the **File path** box, enter the location where you uploaded the initialization script to the *DBFS* using the helper script. For example, enter the path for the initialization script as *dbfs:/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*
7. Click **Add**.
The location of the initialization script appears in the **File path** column.

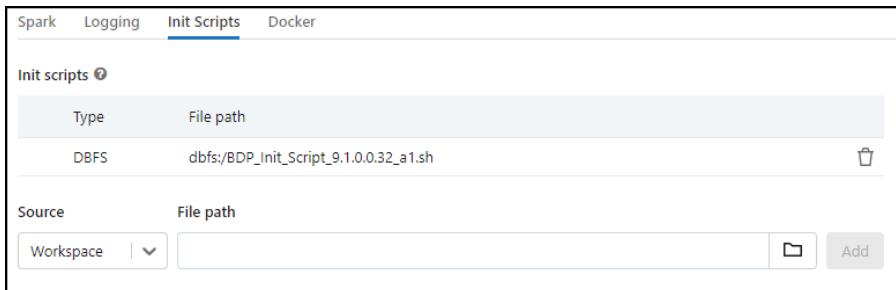


Figure 12-365: Initialization Script Path

8. Click **Create compute**.

The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is `user1@example.com`, then the user `user1` must be present in the ESA policy.

12.6.14.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the Azure Databricks workspace.

Note: For more information about updating the `pepserver.cfg` file, refer to section [Modifying the pepserver.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the Azure Databricks workspace, refer to <https://learn.microsoft.com/en-us/azure/databricks/clusters/clusters-manage#--edit-a-cluster>.
2. On the cluster configuration page, expand **Advanced Options**.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format that is separated by a space.

Table 12-94: Spark Configurations

| Keys | Values |
|--|---|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegility_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegility.spark.PtyExecSparkPlugin</code> |

The following image represents a sample Spark configuration for reference.

The screenshot shows the 'Spark' tab selected in a navigation bar. Below it, a section titled 'Spark config' contains three lines of configuration code:

```
spark.driver.extraJavaOptions -Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm
spark.executor.extraJavaOptions -Djpeplite=/<Protegility_Dir>/jpeplite/lib/jpeplite.plm
spark.plugins com.protegility.spark.PtyExecSparkPlugin
```

Figure 12-366: Spark Configurations

- Click the **Init Scripts** tab.

The screenshot shows the 'Init Scripts' tab selected. It displays a table with two columns: 'Type' and 'File path'. A large plus sign (+) is centered in the middle of the table area. Below the table, a message reads 'No init script set for this compute'. At the bottom, there are dropdown menus for 'Source' and 'File path', along with a 'Workspace' dropdown, a file browser icon, and an 'Add' button.

Figure 12-367: The Init Scripts Tab

- From the **Source** list, select **DBFS**.
 - In the **File path** box, enter the location where you uploaded the initialization script to the *DBFS* using the helper script. For example, enter the path for the initialization script as *dbfs:/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*
 - Click **Add**.
- The location of the initialization script appears in the **File path** column.

The screenshot shows the 'Init Scripts' tab with one entry in the table. The 'Type' column shows 'DBFS' and the 'File path' column shows 'dbfs:/BDP_Init_Script_9.1.0.0.32_a1.sh'. There is a delete icon next to the file path. Below the table, there are dropdown menus for 'Source' and 'File path', along with a 'Workspace' dropdown, a file browser icon, and an 'Add' button.

Figure 12-368: Initialization Script Path

- To save the configuration and restart the cluster, click **Confirm**. The Big Data Hive and Spark Protector are installed on the Azure Databricks platform and you can now execute the Hive and Spark UDFs from Protegility as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks job are performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then the user *user1* must be present in the ESA policy.

12.6.14.8 Registering Hive User Defined Functions (UDFs) with the Unity Catalog

This section provides information about permanently registering Hive UDFs with the Unity Catalog in Databricks. In the Databricks environment, you can permanently register the Hive UDFs in the Hive metastore. The permanent UDFs requires registration only once and can be used for all the sessions. In addition, the permanent UDFs are available even after the session is terminated.

To permanently register a UDF in the Hive metastore, perform the following steps.

1. To register a UDF, specify the name of the catalog and the database.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

2. Press ENTER.

The command executes successfully.

```
1 %sql
2 USE CATALOG hive_metastore;
3 USE DATABASE default;
```

OK

3. To use custom UDFs, run the following command:

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';
```

4. Press ENTER.

The command executes successfully.

```
1 %sql
2 DROP FUNCTION hive_metastore.default.ptyProtectStr;
3 CREATE FUNCTION hive_metastore.default.ptyProtectStr AS
'com.protegility.hive.udf.ptyProtectStr';
4 DROP FUNCTION hive_metastore.default.ptyUnprotectStr;
5 CREATE FUNCTION hive_metastore.default.ptyUnprotectStr AS
'com.protegility.hive.udf.ptyUnprotectStr';

OK
```

In the Unity Catalog environment, the permanent registration for Hive UDFs fails for a custom catalog. The error occurs because the Unity Catalog is yet to provide support for registering permanent UDFs in Hive.

For more information about the support, refer to <https://community.databricks.com/t5/data-governance/quot-create-external-hive-function-is-not-supported-in-unity/m-p/10227>.

Workaround

A workaround is available to use the permanently registered Hive UDFs in the Unity Catalog environment in a custom catalog.

The following examples illustrate protect and unprotect operations in the tables stored under a custom catalog from the Unity Catalog environment by referring them using their fully qualified name.

Example of a Protect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.protected_hive SELECT hive_metastore.default.ptyProtectStr(coll,
'Token_Alphanumeric') FROM dev_cat.dev_db.clear_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *clear_hive* – is the name of the table
- *protected_hive* – is the name of the table

Example of an Unprotect Operation:

```
1 %sql
2 INSERT INTO dev_cat.dev_db.unprotected_hive SELECT
hive_metastore.default.ptyUnprotectStr(coll, 'Token_Alphanumeric') FROM
dev_cat.dev_db.protected_hive;
```

where,

- *dev_cat* – is the name of the catalog
- *dev_db* – is the name of the database
- *unprotected_hive* – is the name of the table

12.6.15 Installing the Big Data Protector on the GCP Databricks Cluster

This section describes the steps for installing the Big Data Hive and Spark Protector on the GCP Databricks platform.

12.6.15.1 Verifying the Prerequisites for Installing the Big Data Protector on a GCP Databricks Cluster

Ensure that the following prerequisites are met, before installing the Big Data Protector on the GCP Databricks platform:

- The user should have a valid Google Cloud Platform account.
- The user should have access to the Google Cloud Platform Databricks Workspace and should be authorized to create clusters.
- The ESA appliance, version 9.1.0.0, is installed, configured, and running and the Databricks cluster nodes should be able to communicate with the ESA.
- A Linux machine with connectivity to the ESA is available.

Note:

The Big Data Protector build for Databricks Runtime 9.1 is used as the example in the following sections.

12.6.15.2 Understanding the Installation Workflow of the Big Data Protector on GCP Databricks

This section describes a brief overview of the installation workflow that the user needs to perform to install the Big Data Protector on the GCP Databricks platform.

The following diagram represents the installation workflow of the Big Data Protector on the GCP Databricks platform.

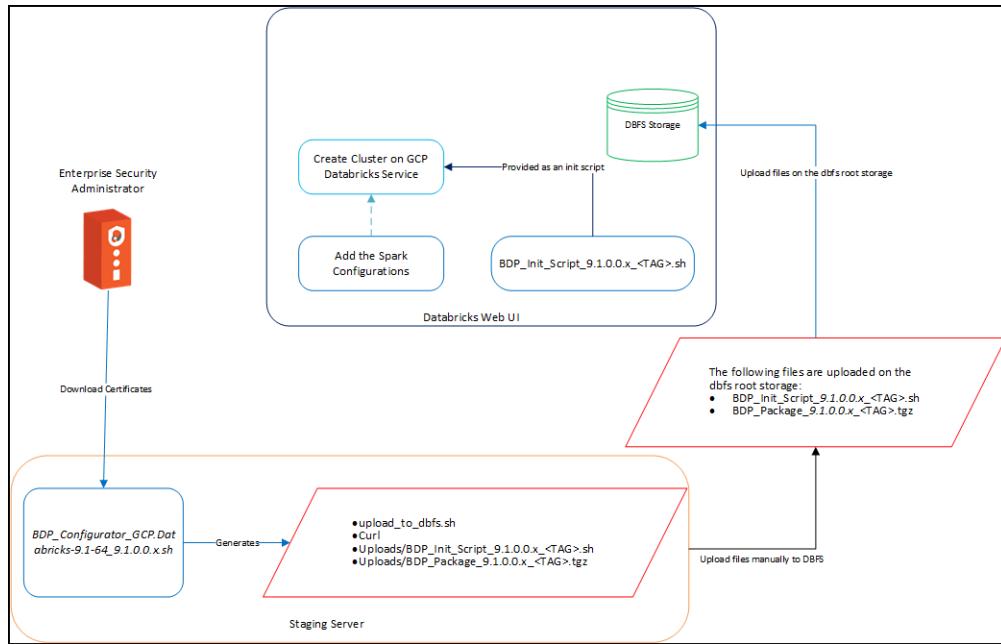


Figure 12-369: Installation Workflow

- On a staging machine, perform the following steps.

- Extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector Installation files.

Note: For more information about extracting the Big Data Protector package, refer to section [Extracting the Big Data Protector Package](#).

- Run the *BDP_Configurator_GCP.Databricks-9.1-64_9.1.0.0.x.sh* script to download the certificates from the ESA and create the following files under the *./Installation_Files/* directory:

- *upload_to_dbfs.sh*
- *curl*
- *Uploads/BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *Uploads/BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about running the *BDP_Configurator_GCP.Databricks-9.1-64_9.1.0.0.x.sh* script, refer to section [Running the Configurator Script](#).

- Upload the following files manually to the *dbfs* root storage:

- *BDP_Init_Script_9.1.0.0.x_<TAG>.sh*
- *BDP_Package_9.1.0.0.x_<TAG>.tgz*

Note: For more information about manually uploading the files to DBFS, refer to section [Uploading the Files Manually to DBFS](#).

- On the Databricks Workspace UI, add the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* during the cluster creation along with the Spark configuration.

Note: For more information about running the *BDP_Init_Script_9.1.0.0.x_<TAG>.sh* script, refer to section [Installing the Big Data Hive and Spark Protector during Cluster Creation](#).

12.6.15.3 Extracting the Big Data Protector Package

You must extract the Big Data Protector package to access the Big Data Protector Configurator script, which is used to create the Big Data Protector Installation files.

► To extract the Big Data Protector Configurator file from the installation package:

1. Login to the CLI on a Linux machine that has connectivity to the ESA.

Note: The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because that IP address or hostname is written by the Configurator script in the *pepperserver.cfg* file too.

2. Download the *BigDataProtector_Linux-ALL-64_x86-64_GCP.Databricks-9.1-64_9.1.0.0.x.tgz* build to the local file system.
3. To extract the contents of the *BigDataProtector_Linux-ALL-64_x86-64_GCP.Databricks-9.1-64_9.1.0.0.x.sh* file from the Big Data Protector installation package using the following command.

```
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_GCP.Databricks-9.1-64_9.1.0.0.x.tgz
```

4. Press **ENTER**.

The *BDP_Configurator_GCP.Databricks-9.1-64_9.1.0.0.x.sh* file is extracted.

12.6.15.4 Running the Configurator Script

You need to run the Big Data Protector configurator script to download certificates from the ESA, and create the installation files for the Big Data Protector.

► To run the Big Data Protector Configurator Script:

1. Run the *BDP_Configurator_GCP.Databricks-9.1-64_9.1.0.0.x.sh* script, from the directory where it is extracted, using the following command:

```
./BDP_Configurator_GCP.Databricks-9.1-64_9.1.0.0.x.sh
```

2. Enter the following inputs when prompted:

- Installation directory
- Unique alphanumeric tag

Note: The Big Data Protector installation files are appended with this tag.

- ESA host name or IP address

Note:

The Linux machine must be able to communicate with the ESA using the same ESA IP address or hostname as that from the Databricks node because that IP address or the hostname is written by the Configurator script in the *pepperserver.cfg* file also.

- ESA listening port
- ESA credentials
- Audit store type - select any one of the following options to manage the logs:

- To use the default setting using the Protegility Elasticsearch appliance, enter *1*. If you enter *1*, then the default Fluent Bit configuration files are retained as is and Fluent Bit will forward the logs to the Protegility ElasticSearch Appliances.
- To use an external audit store, enter *2*. If you enter *2*, then the default Fluent Bit configuration files used for Protegility ElasticSearch (*out_elastic.conf* and *upstream_es.cfg* in the */opt/protegility/fluent-bit/data/config.d/* directory) are renamed (*out_elastic.conf.bkp* and *upstream_es.cfg.bkp*) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for an external audit store are copied to the */opt/protegility/fluent-bit/data/config.d/* directory.
- To use a combination of the default setting with an external audit store, enter *3*. If you enter *3*, then the default Fluent Bit configuration files used for Protegility ElasticSearch (*out_elastic.conf* and *upstream_es.cfg* in the */opt/protegility/fluent-bit/data/config.d/* directory are not renamed. However, the custom Fluent Bit configuration files for an external audit store are copied to the */opt/protegility/fluent-bit/data/config.d/* directory.
- Host name of the Protegility Elasticsearch appliance.

Note:

To add multiple entries for appliances, use the following syntax:

```
comma-separated hostname/IP and/or Ports of Protegility ElasticSearch appliances
```

- Local directory path on the staging server that stores the Fluent Bit configuration for the external audit store.

Note:

- The system prompts for the local directory path only when you select an external audit store or you want to use an external audit store along with the default settings.
- You must add the required custom Fluent Bit configuration files to be used with the External Audit Store inside the directory. Ensure that the configuration file name has the *.conf* extension so that it will be used when Fluent Bit starts.

- The DBFS directory path to upload the installation files.

```
Please enter the DBFS directory path where you will upload the Installation files.  
The path must begin with /dbfs/ (DBFS FUSE Mount path) [default:- /dbfs/FileStore]:
```

Note:

Ensure that the path provided is a DBFS root path.

For more information, refer <https://docs.gcp.databricks.com/data/databricks-file-system.html#dbfs-root>.

3. The configurator script downloads the ESA certificates and generates the following directory and files:

```
./Installation_Files/  
curl  
Uploads/  
    BDP_Init_Script_9.1.0.0.x_<TAG>.sh  
    BDP_Package_9.1.0.0.x_<TAG>.tgz  
upload_to_dbfs.sh
```

4. The files that are generated in the *./Installation_Files/Uploads/* directory must be manually uploaded to the DBFS root directory of the GCP Databricks workspace.

12.6.15.5 Modifying the *pepserver.cfg* File

This section explains the process of modifying the *pepserver.cfg* file. You will be unable to modify the *pepserver.cfg* file when the cluster is running. Therefore, you must extract the BDP_Package tarball generated by the configurator script, update the *pepserver.cfg* file, re-package the installation files and then manually upload the updated BDP_Package tarball to *DBFS*.

Attention: Modifying the *pepserver.cfg* file is an optional step. Exercise caution when modifying the contents of the *pepserver.cfg* file.

► To modify the *pepserver.cfg* file:

1. To navigate to the *./Installation_Files/Uploads/* directory, run the following command.

```
cd ./Installation_Files/Uploads/
```

2. To create a directory to store the extracted files, run the following command.

```
mkdir extraction_dir/
```

3. To extract the contents of the BDP_Package tarball, run the following command.

```
tar -xf BDP_Package_<version>_<tag>.tgz -C extraction_dir/
```

4. Navigate to the directory that contains the *pepserver.cfg* file in the path specified while extracting the contents of the BDP_Package tarball.

For example,

```
cd extraction_dir/defiance_dps/data/
```

5. Using an editor, open the *pepserver.cfg* file.

6. Modify the *pepserver.cfg* file according to requirements.

7. Save the changes to the *pepserver.cfg* file.

8. To re-create the Big Data Protector package, run the following command.

```
tar -zcf BDP_Package_<version>_<tag>.tgz -C extraction_dir/ $(ls extraction_dir) --owner=0 --group=0
```

9. To remove the directory where you extracted the package files, run the following command.

```
rm -rf extraction_dir/
```

10. Manually upload the updated installation files to *DBFS*.

Note: For more information about manually uploading the files to DBFS, refer to section [Uploading the Files Manually to DBFS](#).

12.6.15.6 Uploading the Files Manually to DBFS

You must manually upload the installation files, generated by the configurator script under the *./Installation_Files/Uploads/* directory, to the DBFS root directory of the GCP Databricks workspace.

Warning: The *upload_to_dbfs.sh* script cannot be used to upload the installation files to DBFS because the DBFS REST API feature is disabled in GCP Databricks. Therefore, you must manually upload the files from the *./Installation_Files/Uploads/* directory to DBFS using the DBFS File Browser on the Databricks Workspace Web UI. Ensure to upload the files to the DBFS path that you entered at the configurator script prompt.

For more information about uploading files to DBFS, refer: <https://docs.gcp.databricks.com/data/databricks-file-system.html#upload-data-to-dbfs-from-the-file-browser>

12.6.15.7 Installing the Big Data Hive and Spark Protector

This section explains the following two methods of installing the Hive and Spark protector.

- Installing the Big Data Hive and Spark Protector in a New Cluster
- Installing the Big Data Hive and Spark Protector in an Existing Cluster

12.6.15.7.1 Installing the Big Data Hive and Spark Protector on a New Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on a new cluster in the GCP Databricks workspace.

1. To create a cluster on the GCP Databricks workspace, refer to <https://docs.gcp.databricks.com/clusters/create.html>.
2. On the cluster creation page, click the **Advanced Options** tab.
3. On the **Spark** tab, add the following Spark configurations, such as, the key and value separated by a space.

Table 12-95: Spark Configurations

| Keys | Values |
|--|--|
| <code>spark.driver.extraJavaOptions</code> | <code>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.executor.extraJavaOptions</code> | <code>-Djpeplite=<Protegrity_Dir>/jpeplite/lib/jpeplite.plm</code> |
| <code>spark.plugins</code> | <code>com.protegrity.spark.PtyExecSparkPlugin</code> |
| <code>spark.driver.extraClassPath</code> | <code>/<Protegrity_Dir>/bdp_version</code> |
| <code>spark.executor.extraClassPath</code> | <code>/<Protegrity_Dir>/bdp_version</code> |

The following image represents a sample Spark Configuration for reference.



Figure 12-370: Spark Configurations

4. On the **Init Scripts** tab, set the Init Script Path as `dbfs:/<dbfs_root_path>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh` and click **Add**.

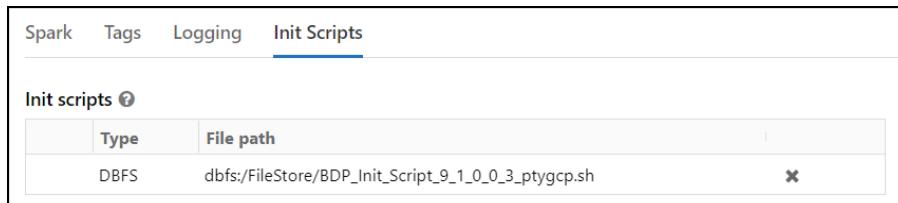


Figure 12-371: Init Script Path

5. Click **Create Cluster**.

The Big Data Hive and Spark Protector is installed on the GCP Databricks platform and you can now execute the Hive and Spark UDFs from Protegrity as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks Job is performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then user *user1* must be present in the ESA policy.

12.6.15.7.2 Installing the Big Data Hive and Spark Protector on an Existing Cluster

This section describes the steps to install the Big Data Hive and Spark Protector on an existing cluster in the GCP Databricks workspace.

Note: For more information about modifying the *pepservice.cfg* file, refer to section [Modifying the pepservice.cfg File](#).

► To install the Big Data Hive and Spark Protector on an existing cluster:

1. To edit an existing cluster on the GCP Databricks workspace, refer to <https://docs.gcp.databricks.com/clusters/clusters-manage.html#edit-a-cluster>.
2. On the cluster configuration page, click the **Advanced Options** toggle.
3. If you have not added the Spark configurations, then click the **Spark** tab.
4. Add the following configurations in the key and value format that is separated by a space.

Table 12-96: Spark Configurations

| Keys | Values |
|--|---|
| <i>spark.driver.extraJavaOptions</i> | <i>-Djepelite=<Protegrity_Dir>jepelite/lib/jepelite.plm</i> |
| <i>spark.executor.extraJavaOptions</i> | <i>-Djepelite=<Protegrity_Dir>jepelite/lib/jepelite.plm</i> |
| <i>spark.plugins</i> | <i>com.protegrity.spark.PtyExecSparkPlugin</i> |
| <i>spark.driver.extraClassPath</i> | <i><Protegrity_Dir>/bdp_version</i> |
| <i>spark.executor.extraClassPath</i> | <i><Protegrity_Dir>/bdp_version</i> |

The following image represents a sample Spark Configuration for reference.

```

spark.plugins com.protegrity.spark.PtyExecSparkPlugin
spark.executor.extraClassPath /opt/protegrity/bdp_version
spark.driver.extraClassPath /opt/protegrity/bdp_version
spark.driver.extraJavaOptions -Djpeplite=/opt/protegrity/jpeplite/lib/jpeplite.plm
spark.executor.extraJavaOptions -Djpeplite=/opt/protegrity/jpeplite/lib/jpeplite.plm

```

Figure 12-372: Spark Configurations

5. Click the **Init Scripts** tab.
6. Set the **File Path** to the location where you manually uploaded the initialization script to *dbfs*.
For example, set the path for the initialization script as *dbfs:/<dbfs_root_path>/BDP_Init_Script_9.1.0.0.0.x_<TAG>.sh*.

| Type | File path | X |
|------|---|---|
| DBFS | dbfs:/FileStore/BDP_Init_Script_9_1_0_0_3_ptygcp.sh | X |

Figure 12-373: Init Script Path

7. Click **Add**.
8. Restart the cluster.

The Big Data Hive and Spark Protector is installed on the GCP Databricks platform and you can now execute the Hive and Spark UDFs from Protegrity as required.

Note: All the data protection operations are performed by the Databricks workspace user running the commands in the Databricks Notebook. The data protection operations that are performed in any Databricks Job is performed by the creator of that job.

For example, if the logged-in workspace user is *user1@example.com*, then user *user1* must be present in the ESA Policy.

12.6.16 Installing the Big Data Protector on a CDP AWS Data Hub Platform

This section describes the tasks for installing Big Data Protector on a CDP AWS Data Hub platform.

For the Big Data Protector 9.0.0.0 release, the CDP AWS Data Hub platform, which includes Cloudera Runtime and Cloudera Manager, version 7.2.1, is used for reference.

For more information about the CDP AWS Data Hub platform 7.2.1, refer to <https://docs.cloudera.com/runtime/7.2.1/index.html>.

The following list describes the installation workflow to install Big Data Protector on a CDP AWS Data Hub Platform.

- a. [Extract the Big Data Protector Package](#)
- b. [Run the Big Data Protector Configurator Script](#)
- c. [Register the Recipe Scripts](#)
- d. [Create and Register the Custom Cluster Template](#)
- e. [Create a Data Hub Cluster](#)
- f. [Updating the Certificates Parcel](#)

g. *Updating the Fluent Bit Configuration Parcel*

12.6.16.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before creating the Data Hub cluster with the Big Data Protector:

- The user must have access to the CDP Management Console.
- The user must have the *PowerUser* role in CDP to create the Data Hub clusters.
- The user must have the *EnvironmentAdmin* resource role in the CDP AWS environment.

For more information about the CDP role and resource role, refer to <https://docs.cloudera.com/management-console/cloud-user-management/topics/mc-understanding-roles-resource-roles.html>.

- An S3 bucket is available to upload the Big Data Protector installation files. The Data Hub cluster instances should be able to read the objects from this bucket by default. If you plan to upload the Big Data Protector installation files using the configurator script, then ensure that you have the Access Keys for authentication.
- If you plan to use the Spark Protector using the Spark client tools like *spark-shell* and *spark-submit* in the CDP AWS Data Hub, then ensure that the CDP users and groups in the CDP environment are accurately mapped to the IAM roles to access cloud storage through the IDBroker mappings.

To understand and set the IDBroker Mappings, refer to the following links:

- <https://docs.cloudera.com/cdp/latest/requirements-aws/topics/mc-edit-idb-mappings.html#mc-edit-idb-mappings>
- https://docs.cloudera.com/runtime/7.2.0/cdp-security-overview/topics/security_how_identity_federation_works_in_cdp.html

12.6.16.2 Extracting the Big Data Protector Package

You need to extract the Big Data Protector package to access the Big Data Protector configurator script to generate the Big Data Protector installation files.

► To extract the files from the installation package:

1. Login to the Linux machine that has connectivity to the ESA.
2. Download the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_CDP-AWS-7-64_9.0.0.0.x.tgz* to a directory.
3. Extract the *BDPConfigurator_CDP-AWS-DataHub-7_9.0.0.0.x.sh* file from the Big Data Protector installation package using the following command.
`tar -xvf BigDataProtector_Linux-ALL-64_x86-64_CDP-AWS-7-64_9.0.0.0.x.tgz`
4. Press ENTER.

The Big Data Protector Configurator script *BDPConfigurator_CDP-AWS-DataHub-7_9.0.0.0.x.sh* is extracted.

Note: The Configurator Script uses the *xxd* command when it uploads files to the S3 bucket. Ensure that the *xxd* tool is installed on the Linux machine. The *xxd* tool is installed with the *vim* (Vi IMproved) utility from the *vim-common* package.

12.6.16.3 Running the Big Data Protector Configurator Script

You need to run the Big Data Protector configurator script to download the certificates from the ESA, and create the installation files for the Big Data Protector.

► To generate the Big Data Protector installation files:



- Run the configurator script.

```
./BDPConfigurator_CDP-AWS-DataHub-7_9.0.0.0.x.sh
```

The following prompt appears.

```
*****
***** Welcome to the Big Data Protector Configurator Wizard *****
***** This will setup the Big Data Protector Installation Files for CDP AWS Data Hub.
Do you want to continue? [yes or no]
```

- Type *yes*.
- Press ENTER.

The following prompt appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ] : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ] : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.
[ 1, 2 or 3 ]:
```

- Type *1* to create the installation files of Big Data Protector.
- If you want to update the certificate parcel after updating the ESA certificates, then type *2*.

Note: For more information about updating the certificate parcel, refer to the section [Updating the Certificates Parcel on a Running Data Hub Cluster](#).

- If you want to update the Fluent Bit configuration parcel, then type *3*.

Note:

For more information on updating the Fluent Bit configuration parcel, refer to the section [Updating the Fluent Bit Configuration Parcel on a Running Data Hub Cluster](#).

- Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el6 ] : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ] : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ] : SuSE Linux Enterprise Server 12.x

Please enter the no.:
```

Note: Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

8. Type *1* or *2* or *3* depending on the OS version used in the CDP Data Hub cluster.

9. Press *ENTER*.

A prompt to enter the S3 bucket path to upload the Big Data Protector installation files appears.

```
Enter the S3 URI where the BDP Installation files are to be uploaded.
```

```
(E.g. s3://examplebucket/folder) :
```

10. When prompted for the S3 URI (Bucket path), enter the accurate S3 URI, which you want to use to upload the installation files.

For example, s3://examplebucket/folder

Note: Ensure that the CDP Data Hub cluster EC2 instances are able to read from this S3 bucket by checking the S3 permissions in the default instance profile or role attached to the instances.

11. Press *ENTER*.

The following prompt appears.

```
Choose one option among the following for BDP Installation files:
```

```
[ 1 ] : Upload files to 's3://examplebucket/folder' S3 URI.
```

```
[ 2 ] : Generate files locally to current working directory. (You would have to manually upload the files to the specified S3 URI)
```

```
[ 1 or 2 ]:
```

12. Depending on the location where you want to store the Big Data Protector files, perform one of the following steps.

a. To create the Big Data Protector installation files locally and upload them to the S3 bucket, and continue with next step, type *1*.

b. To create the Big Data Protector installation files on the local directory path and manually upload to the S3 bucket later, and skip to Step *16*, type *2*.

13. Press *ENTER*.

A prompt to select an option for the type of AWS Access Keys appears.

```
Choose the Type of AWS Access Keys from the following options:
```

```
[ 1 ] : IAM User Access Keys (Permanent access key id & secret access key)
```

```
[ 2 ] : Temporary Security Credentials (Temporary access key id, secret access key & session token)
```

```
[ 1 or 2 ]:
```

14. Type *1* to use IAM User Access Keys.

a. Press *ENTER*.

A prompt for the Access Key ID appears.

b. Type the Access Key ID which has write access to the S3 bucket.

c. Press *ENTER*.

A prompt for the Secret Access Key appears.

d. Type the Secret Access Key which has write access to the S3 bucket.

15. Alternatively, if you select option *2*, then enter the Session Token when prompted.

16. Press *ENTER*.

A prompt to enter the ESA hostname or IP address appears.

17. Enter the ESA host or IP address.

18. Press *ENTER*.

A prompt to enter the ESA Listening Port appears. The default value is *8443*.

19. Enter the ESA Listening Port.

If you want to use the default value of the ESA Listening Port, such as, *8443*, then leave the value empty.

20. Press ENTER.

A prompt for the ESA admin username appears.

21. Enter the ESA admin username.

22. Press ENTER.

A prompt for the ESA admin password appears.

23. Enter the ESA admin password.

24. Press ENTER.

The certificates are fetched from the ESA and a prompt to create the Fluent Bit configuration parcel appears.

```
Fetching Certificates from ESA....  
  
Enter host password for user 'admin':  
% Total % Received % Xferd Average Speed Time Time Time Current  
Dload Upload Total Spent Left Speed  
100 30720 100 30720 0 0 120k 0 --:--:-- --:--:-- --:--:-- 120k  
  
-----  
  
Do you want to package any custom Fluent-Bit configuration files for External Audit Store?  
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used  
with External Audit Store.  
[ no ] : Skip this step.  
[ yes or no ] :
```

25. To create the Fluent Bit configuration parcel, type *yes*.

Note:

If you type *no* at the prompt to create the *PTY_FLUENTBIT_CONF* parcel, then the installer will skip the creation of the Fluent Bit configuration parcel and proceed to generate the remaining installation files.

26. Press ENTER.

The prompt to enter the local directory path to store the Fluent Bit configuration file appears.

```
Creation of PTY_FLUENTBIT_CONF parcel is enabled.
```

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration  
files for External Audit Store:
```

27. Type the local directory path that stores the Fluent Bit configuration files.

Note:

The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across the CDP nodes through the Cloudera Manager. Ensure that you name the custom Fluent Bit configuration file(s) for the external audit store with the extension **.conf*.

28. Press ENTER.

The installation files are uploaded to the S3 URI path under the *./Installation_Files/* directory.

```
Installation_Files/  
CSDandParcels  
BDP_PEP-9.0.0.0.x.jar  
PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel  
PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha  
PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel  
PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha  
PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel
```

```

PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
pepimpala
  pepimpala32_RHEL.so
  sqlscripts
    createobjects.sql
    dropobjects.sql
RecipesAndTemplates
  BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh
  BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh
  custom_properties_template.json
  guide_to_create_cluster_template_with_bdp.txt

```

Note:

- If you select the upload option in step [12](#), then the installation files are uploaded to the S3 bucket path and maintain the same directory structure as mentioned in step [28](#).
- If you select the local directory path in step [12](#), then manually upload the Installation files that are present in the `./Installation_Files/` directory to the S3 bucket path and maintain the same directory structure as mentioned in step [28](#). For example, the `./Installation_Files/pepimpala/pepimpalaXX_RHEL.so` file should be uploaded to `s3://<bucket_path>/pepimpala/pepimpalaXX_RHEL.so` S3 bucket directory.

12.6.16.4 Registering the Recipe Scripts

Note: The `BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh` script downloads the Big Data Protector CSD and Parcels from the S3 bucket to the Cloudera Manager local CSD and Parcel repository before the Cloudera Manager server starts.

The `BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh` script runs after the Cloudera Manager Server starts. It creates and executes secondary scripts as background processes for each available Protegility Parcel. The background processes will check when the Cloudera Manager Server API endpoint would be open and then sends the requests to distribute and activate the `PTY_BDP`, `PTY_CERT`, and `PTY_FLUENTBIT_CONF`(if present) parcels.

The Recipe scripts execution logs can be found in the `/var/log/recipes/` directory by default and the execution logs of the secondary scripts (executing in the background) can be found in the `/tmp/protegility/` directory.

► Perform the following steps to register each recipe script:

1. On the Cloudera Management Console screen, navigate to **Shared Resources > Recipes**.
2. Click the **Register Recipe** button.
3. Enter the Recipe name.
4. Click the **Type** drop-down.
5. Select the recipe script type.

Note: You need to register both the `pre-cloudera-manager-start` and `post-cloudera-manager-start` recipe scripts.

6. Enter the optional recipe description.
7. Select the **File** option to upload the file, to the Cloudera manager UI, that contains the recipe script.
To upload the `pre-cloudera-manager-start` recipe script, select the `BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh` script. To upload the `post-cloudera-manager-start` recipe script, select the `BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh` script.
8. Click the **Register** button.

The registration of the `pre-cloudera-manager-start` and `post-cloudera-manager-start` recipe scripts are completed.



Refer the following images for registering the recipes.

bdp-pre-cm-start-recipe

Type
pre-cloudera-manager-start

Enter the description

Recipe Source
 File Text

Upload File

```
#!/bin/bash
# Copyright (c) 2021 Protegility, Inc. All rights reserved.
# pre-cloudera-manager-start
BUCKET_PATH="s3://pty-cdp-bucket/bdp-9.0.0.0.8-"
lsc=$echo -n "$BUCKET_PATH" | tail -c 1
if [ "$lsc" == '/' ]; then
    BUCKET_PATH=${BUCKET_PATH%?}
fi
echo "Started Execution of Pre CM Server Start Recipe Script..."
echo "Printing CM Server status..." 
service cloudera-scm-server status
echo "Creating temporary directory..." 
mkdir -p /tmp/protegility/aws_virtual_env
... (script continues)
```

Register

Figure 12-374: bdp-pre-cm-start-recipe

bdp-post-cm-start-recipe

Type
post-cloudera-manager-start

Enter the description

Recipe Source
 File Text

Upload File

```
#!/bin/bash
# Copyright (c) 2021 Protegility, Inc. All rights reserved.
# post-cloudera-manager-start
gateway_path="{{{gateway_path}}}"
knox_gateway="{{{general.primaryGatewayInstanceDiscoveryFQDN}}}"
cloudera_manager_ip="{{{general.clusterManagerIp}}}"
cm_username="{{{general.cmUserName}}}"
cm_password="{{{general.cmPassword}}}"
cluster_name="{{{general.clusterName}}}"
pty_bdp_product_version='9.0.0.0.8_CDP7.p0'
pty_cert_product_version='9.0.0.0.8_CDP7.p0'
pty_fluentbit_conf_product_version='9.0.0.0.8_CDP7.p0'
# curl https://${knox_gateway}/${gateway_path}/cdp-proxy-api/cm-api/
... (script continues)
```

Register

Figure 12-375: bdp-post-cm-start-recipe

12.6.16.5 Creating and Registering the Custom Cluster Template

You need to create and register the custom cluster template to add *BDP_PEP* service and required service configurations to the Data Hub cluster.

► To Create the Custom Cluster Template with the *BDP PEP* service:

1. On the Cloudera Management Console screen, navigate to **Shared Resources > Cluster Templates**.

| Name | Platform | Group Count | Description | Tags | Time Created |
|--|-------------------------|-------------|--|----------|----------------------------------|
| datamart_0008 | Cloudera Runtime 7.2.6 | 3 | | | 12/9/2021, 10:10:41 PM GMT+5:30 |
| 7.2.6-BDP 9.0.0.x with Data Engineering | Cloudera Runtime 7.2.6 | 4 | | | 12/9/2021, 11:31:02 PM GMT+5:30 |
| 7.2.10-BDP 9.0.0.x In Data Engineering | Cloudera Runtime 7.2.10 | 4 | | | 12/9/2021, 7:48:56 PM GMT+5:30 |
| bdp-9.0.0.x-cluster-template | Cloudera Runtime 7.2.12 | 4 | | | 12/9/2021, 7:49:56 PM GMT+5:30 |
| cdp-on-aws-9-peest | Cloudera Runtime 7.2.6 | 4 | Testing cdp on aws | | 11/12/2021, 1:45:00 PM GMT+5:30 |
| 7.2.12-SDX Micro Duty: Apache Hive Metastore, Apache Ranger, Apache Atlas | Cloudera Runtime 7.2.12 | 2 | 7.2.12-Micro SDX Template with Atlas, HAMS, Ranger and other services they are dependent on | built-in | 11/17/2021, 8:18:49 PM GMT+5:30 |
| 7.2.10-DataEngineering with BDP 9.0.0 and custom conf | Cloudera Runtime 7.2.10 | 4 | | | 11/10/2021, 1:21:57 PM GMT+5:30 |
| test3 | Cloudera Runtime 7.2.6 | 4 | | | 11/12/2021, 4:51:26 PM GMT+5:30 |
| 7.2.6-DataEngineering with BDP 9.0.0 and custom conf | Cloudera Runtime 7.2.6 | 4 | 7.2.6-DataEngineering with BDP 9.0.0 and all configs for Hive and Spark | | 11/12/2021, 12:49:44 PM GMT+5:30 |
| LogForwarderRoleTemplate_5 | Cloudera Runtime 7.2.6 | 4 | LogForwarderRoleTemplate_5 | | 11/2/2021, 3:48:12 PM GMT+5:30 |
| LogForwarderRoleTemplate_4 | Cloudera Runtime 7.2.6 | 4 | | | 11/2/2021, 2:39:54 PM GMT+5:30 |
| LogForwarderRoleTemplate_3 | Cloudera Runtime 7.2.6 | 4 | LogForwarderRoleTemplate_3 | | 11/2/2021, 1:42:29 PM GMT+5:30 |
| LogForwarderRoleTemplate_2 | Cloudera Runtime 7.2.6 | 4 | LogForwarderRoleTemplate_2 | | 11/2/2021, 1:12:54 PM GMT+5:30 |
| LogForwarderRoleTemplate_1 | Cloudera Runtime 7.2.6 | 4 | | | 11/2/2021, 1:37:38 AM GMT+5:30 |
| LogForwarderRole_Template | Cloudera Runtime 7.2.6 | 4 | | | 10/29/2021, 9:29:24 AM GMT+5:30 |
| 7.2.12-Streams Messaging Heavy Duty: Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | Cloudera Runtime 7.2.12 | 6 | 7.2.12-Streams Messaging Heavy Duty with Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | built-in | 9/5/2021, 4:07:35 PM GMT+5:30 |
| 7.2.12-Streams Messaging Light Duty: Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | Cloudera Runtime 7.2.12 | 3 | 7.2.12-Streams Messaging Light Duty with Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | built-in | 9/5/2021, 4:07:35 PM GMT+5:30 |

Figure 12-376: Cluster Templates Screen

2. Copy the JSON content of the required cluster template that you want to use in a text editor.
For example, Data Engineering, Data Mart, and so on.
3. On the cluster template JSON, search for the `services` key, whose value is a JSON array of JSON objects.
4. Add the following JSON objects to that array of services:

```
{
  "refName": "bdp_pep",
  "serviceType": "BDP_PEP",
  "roleConfigGroups": [
    {
      "refName": "bdp_pep-PEP_SERVER-BASE",
      "roleType": "PEP_SERVER",
      "base": true,
      "configs": [
        {
          "name": "esa_address",
          "value": "{{{esa_address}}}"
        }
      ]
    },
    {
      "refName": "bdp_pep-PEP_LOGFORWARDER-BASE",
      "roleType": "PEP_LOGFORWARDER",
      "base": true,
      "configs": [
        {
          "name": "elasticsearch_ip_port_list",
          "value": "{{{elasticsearch_ip_port_list}}}"
        },
        {
          "name": "auditstore_type",
          "value": "{{{auditstore_type}}}"
        }
      ]
    }
  ]
}
```

```
        ]
    }
```

Note: The service object is position-independent within the array and can be placed at the beginning or end of the array.

Note: Adding the *BDP_PEP* service to the array of services will ensure that the service is added to the Data Hub cluster during the cluster creation, when the Cloudera Manager imports the cluster template.

Note: Ensure that the values, such as the `{{{esa_address}}}` should be written as it is (called Mustache template "`{{{...}}}`"). The actual value is set by adding the custom properties during the creation of the CDP data hub cluster.

For more information about the format of the custom properties, check the *custom_properties_template.json* file in the S3 bucket containing installation files of Big Data Protector.

For more information about the installation files of Big Data Protector, refer to step [28](#) in the section [Running the Configurator Script](#).

5. After adding the *bdp_pep* service object, search for the *hostTemplates* key in the cluster template, whose value is an array of *hostTemplate* objects for master, worker, and compute nodes etc.
6. For each *hostTemplate* in that array, search for the key *roleConfigGroupsRefNames*, which has a value of array of strings.
7. Add the *bdp_pep-PEP_SERVER-BASE* and the *bdp_pep-PEP_LOGFORWARDER-BASE* strings in the *roleConfigGroupsRefNames* array.

```
{
  ...
  "hostTemplates": [
    {
      "refName": "...",
      "cardinality": "...",
      "roleConfigGroupsRefNames": [
        ...,
        "bdp_pep-PEP_SERVER-BASE",
        "bdp_pep-PEP_LOGFORWARDER-BASE"
      ],
      ...
    },
    {
      ...
    },
    ...
  ],
  ...
}
```

Note:

- The *PTY Pep Server* and *PTY Logforwarder* roles are installed on each of the *hostTemplate* instances.
- If you are planning to use the Hive on Tez service in the Data Hub cluster, then refer to section [Setting the Hive on Tez Service Configuration](#).
- If you are planning to use the HBase service in the Data Hub cluster, then refer to section [Setting the HBase Configuration](#).
- If you are planning to use the Spark on Yarn service in the Data Hub cluster, then refer to section [Setting the Spark on Yarn Service Configuration](#).

8. Click the **Create Cluster Template** button.

The **Cluster Template Register** screen appears.

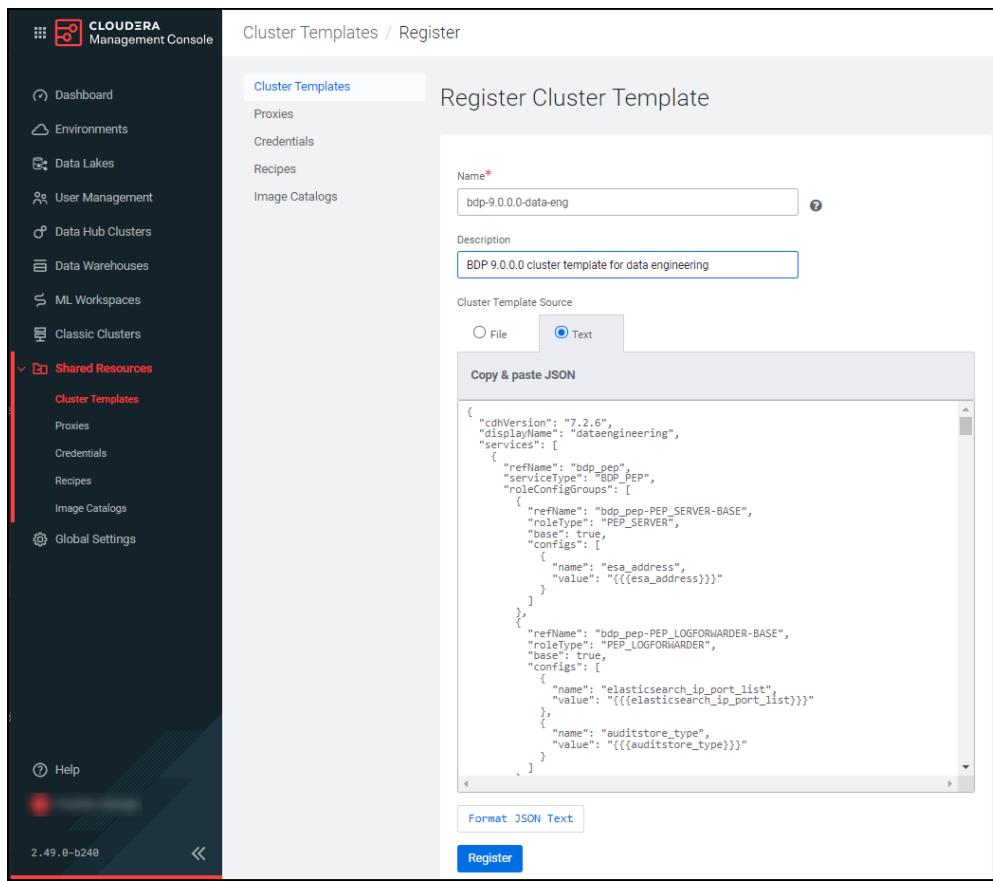


Figure 12-377: Cluster Template Register

9. Enter a cluster template name.
10. Enter an optional description for the cluster template.
11. Select any of the following options to select the Cluster Template Source.
 - a. Select the **Text** option to paste cluster template in JSON format.
 - b. Select the **File** option to upload a file that contains the cluster template.
 - c. Select the **URL** option to select the URL for the cluster template.
12. Click the **Register** button to register the custom Cluster Template on the CDP Management Console.

12.6.16.6 Creating a Data Hub Cluster

To create a new Data Hub cluster with Big Data Protector, use the registered cluster template with the two Recipes generated and custom properties.

Note:

- For more information about using the registered cluster template, refer to the section [Creating and Registering the Custom Cluster Template](#).
- For more information about registering the recipe scripts, refer to the section [Registering the Recipe Scripts](#).

► To create a Data Hub Cluster:

1. On the CDP Management Console, click the **Data Hub Clusters** tab.

The Data Hub screen appears.

The screenshot shows the Cloudera Management Console interface. On the left, there's a sidebar with various navigation options like Dashboard, Environments, Data Lakes, User Management, Data Hub Clusters (which is currently selected), Data Warehouses, ML Workspaces, Classic Clusters, Shared Resources, and Global Settings. Below the sidebar is a help section and a status bar indicating '2.49.0-b240' and the URL 'https://console.us-west-1.cdp.cloudera.com/cloud/workloads/details/parimal-damamart/hardware'. The main content area is titled 'Data Hubs' and contains a table listing six Data Hubs:

| CDH 7.2.6 | CDH 7.2.6 | CDH 7.2.6 |
|---|---|---|
| aws DATAENG-BDP-9-DEMO Running ✓ 7.2.6-DataEngineering with BDP 9.0.0.a... | aws -DATAMART Running ✓ datamart...-9008 | aws cdp-on-aws-9-psptest Stopped ⚡ |
| Environment Show Nodes 6 Created At 12/06/21, 01:06 PM GMT+5:30 | Environment Show Nodes 4 Created At 12/09/21, 12:27 PM GMT+5:30 | Environment Show Nodes 6 Created At 12/08/21, 11:45 AM GMT+5:30 |
| aws SP cdp-on-aws-9-psptest Stopped ⚡ | aws TEST3 Stopped ⚡ 7.2.10-DataEngineering with BDP 9.0.0.a... | aws TESTDATAMART Running ✓ 7.2.6 - Data Mart, Apache Impala, Hue |
| Environment Show Nodes 4 Created At 12/07/21, 06:26 PM GMT+5:30 | Environment Show Nodes 4 Created At 11/16/21, 06:59 PM GMT+5:30 | Environment Show Nodes 4 Created At 12/09/21, 02:14 PM GMT+5:30 |

Figure 12-378: Data Hub Screen

- Click the **Create Data Hub** button.

The **Provision Data Hub** screen appears.

The screenshot shows the 'Provision Data Hub' screen. The left sidebar is identical to the one in Figure 12-378. The main area has a title 'Provision Data Hub' and a sub-section 'Provision on-demand workload clusters with the combination of applications for various business needs such as enterprise data warehouse management and data science operations.' It includes a dropdown for 'Selected Environment with running Data Lake' set to 'aws pty-cdp726', a radio button for 'Cluster Definition' (selected) or 'Custom', and a 'Services' section showing installed services like Data Analytics Studio, Hdfs, Hive, Hue, Livy, Oozie, Queue Manager, Spark, Yarn, and Zeppelin, along with ZooKeeper. Below this is a 'General Settings' section with a 'Cluster Name*' field containing 'bdp-data-eng-test', a 'Tags' note about adding cluster-related resources, and buttons for 'Provision Cluster', 'Save As New Definition', 'Show CLI Command', and 'Show Generated Cluster Template'.

Figure 12-379: Provision Data Hub Screen

- Select the CDP AWS environment.
- Select the **Custom** tab.
- In the **Cluster Template** drop-down, select the previously created customized cluster template with the Big Data Protector.

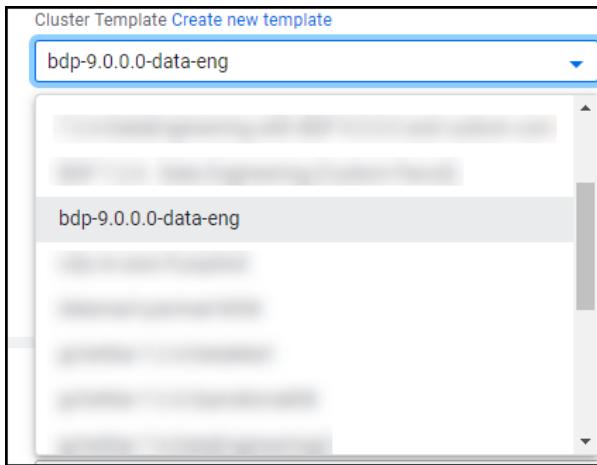


Figure 12-380: Cluster Template Drop-down

6. Set the cluster name.
7. Click **Advanced Options**.
8. Select the **Hardware and Storage** tab.

Note the host group which will install the Cloudera Manager (CM) Server. Usually, the host is the Master node.

9. Select the **Cluster Extensions** tab.

A screenshot of the 'Cluster Extensions' tab. On the left, there is a sidebar with tabs: 'Image Catalog', 'Network And Availability', 'Hardware And Storage', 'Cloud Storage', and 'Cluster Extensions', with 'Cluster Extensions' being the active tab. The main area is titled 'Recipes' and contains four sections: 'Master Node', 'Worker Nodes', 'Compute Node', and 'Gateway Node'. Each section has a 'Please select a recipe' input field and an 'Attach' button. Below each input field, it says 'Attached Recipes' and 'There are no attached Recipes.'

Figure 12-381: Cluster Extensions Tab

10. Attach the two previously registered Recipes to the host group that would host the Cloudera Manager Server.

Note: For more information about registering the recipe, refer to section [Registering the Recipe Scripts](#). Both the Recipes should be executed on the Cloudera Manager Server node of the cluster.

11. Add the contents of the `custom_properties_template.json` file to the Custom Properties section generated in step [28](#) in the section [Running the Configurator Script](#). Replace the placeholder strings of the ESA hostname / IP address, Protegility ElasticSearch appliance list of IP/ports, and Audit Store type with the actual values.

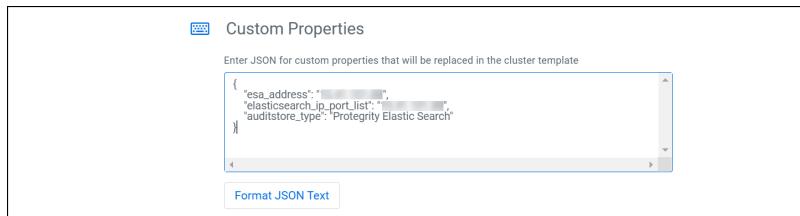


Figure 12-382: Custom Properties

Warning: After the cluster startup, the Cloudera manager username and password are seen in clear in the log files generated in the `/var/log/recipes/` directory.

12. Click Provision Cluster.

The cluster creation process starts.

12.6.16.7 Setting the Big Data Protector Configuration for the CDP AWS Data Hub Platform

While creating a custom cluster template, ensure that you add the required service configurations as mentioned in the following sections.

The configurations of the Big Data Protector can be set using any one of the following options:

- Setting the configuration using a cluster template
- Setting the configuration after creating the cluster

12.6.16.7.1 Setting the Configurations Using a Cluster Template

Set the Big Data Protector configuration in the custom cluster template by modifying the cluster template.

12.6.16.7.1.1 Setting the Hive on Tez Service Configuration

If you want to use the Hive on Tez service in the Data Hub cluster (cluster template), then set the custom Big Data Protector Hive properties.

► To configure Hive on Tez service in the cluster template:

1. On the cluster template JSON, in the services JSON array, search for the JSON object with the key service type as `HIVE_ON_TEZ`.
2. Add the following JSON object to the serviceConfigs array.

```
{
  "name": "HIVE_ON_TEZ_service_env_safety_valve",
  "value": "HIVE_CLASSPATH=/opt/cloudera/parcels/PTY_BDP/
pephive/lib/pephive-*.jar:/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:$${HIVE_CLASSPATH}
\nJAVA_LIBRARY_PATH=/opt/cloudera/parcels/PTY_BDP/jpeplite/lib:$${JAVA_LIBRARY_PATH}"
}
```

3. If the Data Analytics Studio service (DAS) is installed with Hive, then append the following string in the value of serviceConfig JSON object having name as `hive_service_config_safety_valve`.

```
<property><name>hive.exec.pre.hooks</
name><value>com.protegility.hive.PtyHiveUserPreHook,org.apache.hadoop.hive.ql.hooks.HivePro
toLoggingHook</value></property>
```

For example, after appending, the final serviceConfig *hive_service_config_safety_valve* object should be as follows.

```
{
    "name": "hive_service_config_safety_valve",
    "value": "<property><name>fs.s3a.ssl.channel.mode</name><value>openssl</value></
property><property><name>hive.txn.acid.dir.cache.duration</name><value>0</value></
property><property><name>hive.exec.pre.hooks</
name><value>com.protegrity.hive.PtyHiveUserPreHook,org.apache.hadoop.hive.ql.hooks.HivePro
toLoggingHook</value></property>"
}
```

4. If the Data Analytics Studio service (DAS) is not installed with Hive, then append the following string in the value of serviceConfig JSON object having name as *hive_service_config_safety_valve*.

For example, after appending the string, the final serviceConfig *hive_service_config_safety_valve* object should be as follows.

```
{
    "name": "hive_service_config_safety_valve",
    "value": "<property><name>fs.s3a.ssl.channel.mode</
name><value>openssl</value></property><property><name>hive.txn.acid.dir.cache.duration</
name><value>0</value></property><property><name>hive.exec.pre.hooks</
name><value>com.protegrity.hive.PtyHiveUserPreHook</value></property>"
}
```

5. In the roleConfigGroups array, search for the JSON object with refName *hive_on_tez-GATEWAY-BASE*.
6. Add the following JSON object to the configs array.

```
{
    "name": "hive_client_env_safety_valve",
    "value": "HIVE_CLASSPATH=/opt/cloudera/parcels/PTY_BDP/
pephive/lib/pephive-*.jar:/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:${HIVE_CLASSPATH}
\nJAVA_LIBRARY_PATH=/opt/cloudera/parcels/PTY_BDP/jpeplite/lib:${JAVA_LIBRARY_PATH}"
}
```

After setting the Hive on Tez service configuration, continue the steps mentioned in section [Setting the Tez Service Configuration](#).

12.6.16.7.1.2 Setting the Tez Service Configuration

To use the Tez service in the Data Hub cluster (cluster template), set the custom Tez properties.

► To set the Tez Service Configuration:

1. On the cluster template JSON, in the services JSON array, search for the JSON object with the key service type as *TEZ*.
2. Add the following JSON objects to the serviceConfigs array.

```
{
    "name": "tez.am.launch.env",
    "value": "LD_LIBRARY_PATH=/opt/cloudera/parcels/CDH/lib/hadoop/lib/native:/opt/
cloudera/parcels/PTY_BDP/jpeplite/lib"
},
{
    "name": "tez.cluster.additional.classpath.prefix",
    "value": "/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:/opt/cloudera/parcels/
PTY_BDP/pephive/lib/*"
}
```

12.6.16.7.1.3 Setting the Impala Service Configuration

If you want to use the Impala service in the Data Hub cluster (Data Mart cluster template), then note that the *pepimpala*.so* Protegity library is not a part of the *PTY_BDP* parcel. The *pepimpala* library must be stored in the S3 bucket and accessed by using the *s3a://* URI when registering the UDFs.

The *createobjects.sql* script displays the SQL syntax to create the Protegity UDFs.

The *dropobjects.sql* script displays the SQL syntax to drop the Protegity UDFs.

12.6.16.7.1.4 Setting the HBase Configuration

If you want to use the HBase service in the Data Hub cluster (cluster template), then set the custom Big Data Protector configurations for HBase.

► To configure the HBase service in the cluster template:

1. On the cluster template JSON, in the *roleConfigGroups* array, search for the JSON object with *refName* *hbase-REGIONSERVER-BASE*.
2. Add the following JSON object to the *Configs* array.

```
{
    "name": "hbase_coprocessor_region_classes",
    "value": "com.protegity.hbase.PTYRegionObserver"
}
```

12.6.16.7.1.5 Setting the Spark on Yarn Service Configuration

If you want to use the Spark on Yarn service in the Data Hub cluster (cluster template), then set the custom Big Data Protector configurations for Spark on Yarn. The following settings are only valid for Spark versions greater than or equal to 2.4.0 and lower than 3.0.0.

► To configure the Spark on Yarn service in the cluster template:

1. On the cluster template JSON, in the *services* JSON array, search for the JSON object with the key *serviceType* as *SPARK_ON_YARN*.
2. In the *roleConfigGroups* array, search for the JSON object with *refName* as *spark_on_yarn-GATEWAY-BASE*.
3. In the *configs* array, search for *spark-conf/spark-defaults.conf_client_config_safety_valve* and append *|nspark.executor.plugins=com.protegity.spark.PtyExecSparkPlugin* to the value.

```
{
    "name": "spark-conf/spark-defaults.conf_client_config_safety_valve",
    "value": "spark.hadoop.fs.s3a.ssl.channel.mode=openssl\nspark.hadoop.mapreduce.fileoutputcommitter.algorithm.version=
1\nspark.executor.plugins=com.protegity.spark.PtyExecSparkPlugin"
}
```

12.6.16.7.2 Setting the Configuration after Creating the Cluster

After the cluster starts all the services, then set the Big Data Protector configuration using the Cloudera Manager UI, and restart the services with stale configurations.

Table 12-97: Recommended Configuration for Big Data Protector

| Service | Protector Jar Configuration | Protegility Native Library Configuration |
|---|--|--|
| Hive on Tez | <p>In <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: HIVE_CLASSPATH Value: /opt/cloudera/ parcels/PTY_BDP/pephive/lib/ pephive-*.*jar:/opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:\$ {HIVE_CLASSPATH}</pre> <p>In <i>Hive Service Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-site.xml</i>:</p> <pre>Name: hive.exec.pre.hooks Value: com.protegility.hive.PtyHiveUserPreHook</pre> | <p>In <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: JAVA_LIBRARY_PATH Value: /opt/cloudera/ parcels/PTY_BDP/jpeplite/ lib:\${JAVA_LIBRARY_PATH}</pre> |
| Tez | <pre>Name: tez.cluster.additional.class path.prefix Value: /opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:/opt/ cloudera/parcels/PTY_BDP/ pephive/lib/*</pre> | <pre>Name: tez.am.launch.env = Value: LD_LIBRARY_PATH=/opt/ cloudera/parcels/CDH/lib/ hadoop/lib/native:/opt/ cloudera/parcels/PTY_BDP/ jpeplite/lib</pre> |
| HBase | <pre>Name: hbase.coprocessor.region.clas ses Value: com.protegility.hbase.PTYRegi onObserver</pre> | |
| Spark on Yarn (Spark versions greater than or equal to 2.4.0 and lower than 3.0.0.) | <p>In 'Spark Client Advanced Configuration Snippet (Safety Valve)' for <i>spark-conf/spark-defaults.conf</i>:</p> <pre>spark.executor.plugins=com.p rotegility.spark.PtyExecSpark Plugin</pre> | |

12.6.16.8 Updating the Certificates Parcel on a AWS Data Hub Cluster

If the customers have updated the certificates on the ESA, with which the Big Data Protector is configured, then the Certificates parcel must be updated with the new certificates. The updated Certificates parcel will be utilized by the nodes in the cluster.



► To update the *PTY_CERT* parcel on a running Data Hub cluster:

- Run the Configurator script.

```
./BDPConfigurator_CDP-AWS-DataHub-7_9.0.0.0.x.sh
```

The following prompt appears.

```
*****
          Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP AWS Data Hub.

Do you want to continue? [yes or no]
```

- Type *yes*.

- Press ENTER.

The following prompt appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ]      : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                         Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]           : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                         Use this if you want to set Custom Fluent-Bit configuration
files to
                         forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

- Type *2* to update the Certificate parcel after updating the ESA certificates.

- Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.
```

```
[ 1: el6 ]      : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ]    : SuSE Linux Enterprise Server 12.x
```

```
Please enter the no.:
```

Note: Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

- Type *1* or *2* or *3* depending on the OS version used in the CDP Data Hub cluster.

- Press ENTER.

A prompt to enter the S3 bucket path to upload the Big Data Protector installation files appears.

- Enter the S3 bucket path where you want to upload the Big Data Protector installation files.

A prompt to select an option to upload the Big Data Protector installation packages to the S3 bucket or a local directory path appears.

Note: Ensure that the CDP Data Hub cluster EC2 instances are able to read from the S3 bucket path by checking the S3 permissions in the instance profile or role attached to the instances.

9. Depending on the location where you want to upload the Big Data Protector files, perform one of the following steps.
 - a. Type *1* to generate the Big Data Protector installation files locally and upload to the S3 bucket and continue with next step.
 - b. Type *2* to create the Big Data Protector installation files on the local directory path, manually upload to the S3 bucket later, and skip to Step [13](#).
10. Press *ENTER*.

A prompt to select an option for the type of AWS Access Keys appears.

```
Choose the Type of AWS Access Keys from the following options:  
[ 1 ] : IAM User Access Keys (Permanent access key id & secret access key)  
[ 2 ] : Temporary Security Credentials (Temporary access key id, secret access key & session token)
```

```
[ 1 or 2 ]:
```

11. Type *1* to use IAM User Access Keys.
 - a. Press *ENTER*.
A prompt for the Access Key ID appears.
 - b. Type the Access Key ID which has write access to the S3 bucket.
 - c. Press *ENTER*.
A prompt for the Secret Access Key appears.
 - d. Type the Secret Access Key which has write access to the S3 bucket.
12. Alternatively, if you select option *2*, then enter the Session Token when prompted.
13. Press *ENTER*.
A prompt to enter the ESA hostname or IP address appears.
14. Enter the ESA host or IP address.
15. Press *ENTER*.
A prompt to enter the ESA Listening Port appears. The default value is *8443*.
16. Enter the ESA Listening Port.
If you want to use the default value of the ESA Listening Port, such as, *8443*, then leave the value empty.
17. Press *ENTER*.
A prompt for the ESA admin username appears.
18. Enter the ESA admin username.
19. Press *ENTER*.
A prompt for the ESA admin password appears.
20. Enter the ESA admin password.

The ESA certificates are successfully downloaded and the prompt to enter the patch version of the *PTY_CERT* parcel appears.

```
Enter host password for user 'admin':  
% Total % Received % Xferd Average Speed Time Time Time Current  
          Dload Upload Total Spent Left Speed  
100 30720 100 30720    0     0  138k      0 --:--:-- --:--:-- --:--:-- 138k
```

```
-----  
Generating Installation files...
```

NOTE:

You can verify the version of the activated PTY_CERT parcel from the parcel name, such as PTY_CERT-x.x.x.x_CDPx.x.p<version>-<os>.parcel, where the <version> parameter denotes the patch version of the PTY_CERT parcel.

For Example: If the current activated PTY_CERT parcel is PTY_CERT-x.x.x.x_CDPx.x.p0-<os>.parcel, the patch version of the PTY_CERT parcel will be 0. Please do NOT include 'p' while specifying the version.

Enter the <version> of the current PTY_CERT Parcel as specified in the parcel name [0]:

21. Enter the patch version of the currently activated *PTY_CERT* parcel on the Data Hub cluster.

Note:

The patch version is the number after the *.p* in the parcel name. For example, the *PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel* has a patch version *0*. The updated *PTY_CERT* parcel and SHA checksum with an incremented patch version is created locally in the same working directory, such as, *./Installation_Files/UpdatedCERTParcel* / directory.

22. If you select option *1* in step [9](#), then the updated *PTY_CERT* parcel and the SHA checksum files are uploaded to the S3 bucket for ease of access.
23. If you select option *2* in step [9](#), then the updated *PTY_CERT* parcel and the SHA checksum files are locally generated in the *./Installation_Files/UpdatedCERTParcel* / directory.

Enter the <version> of the current PTY_CERT Parcel as specified in the parcel name [0]:

```
*****
*****
```

* The updated PTY_CERT parcel 'PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel' and sha1
checksum are locally generated in ./Installation_Files/UpdatedCERTParcel/ directory.
-> Please manually copy the PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel and .sha files
to Cloudera Manager Server's local parcel repository on the existing running Data Hub
cluster.

```
*****
*****
```

Successfully configured the Updated PTY_CERT parcel for CDP AWS DataHub.

24. Perform the following steps on the running Data Hub Cluster Cloudera Manager Server Node.
 - a. Copy the updated *PTY_CERT* parcel and *.sha* file to the running Data Hub cluster's Local parcel repository (*/opt/cloudera/parcel-repo*).
 - b. Change the ownership of the copied files to *cloudera-scm:cloudera-scm*.
 - c. On Cloudera Manager UI, navigate to **Parcels** and click the **Check for New Parcels** button.
 - d. Distribute and activate the new parcel manually using the Cloudera Manager UI and replace the older *PTY_CERT* parcel.
 - e. Restart any dependent services, such as, *BDP_PEP*.

12.6.16.9 Updating the Fluent Bit Configuration Parcel on a AWS Data Hub Cluster

If the user wants to use a newer set of custom Fluent Bit configuration files for sending the logs to an external audit store, then the Fluent Bit parcel must be updated, distributed, and activated across the cluster nodes through the Cloudera Manager.

1. Login to the host machine, which contains the Big Data Protector configurator script.
2. Run the *./BDPConfigurator_CDP-AWS-DataHub-7.9.0.0.0.x.sh* script.

A prompt to continue the configuration of Big Data Protector appears.



3. Type *yes* to start the configuration of the Big Data Protector.

4. Press ENTER.

The following prompt appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ]      : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                         Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                         : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                         Use this if you want to set Custom Fluent-Bit configuration
files to
                         forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. Type *3* to update the Fluent Bit configuration parcel.

6. Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el6 ]      : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ]   : SuSE Linux Enterprise Server 12.x

Please enter the no.:
```

Note:

Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

7. Type *1* or *2* or *3* depending on the OS version used in the CDP Data Hub cluster.

8. Press ENTER.

A prompt to enter the S3 bucket path to upload the Big Data Protector installation files appears.

9. Enter the S3 bucket path where you want to upload the Big Data Protector installation files.

A prompt to select an option to upload the Big Data Protector installation packages to the S3 bucket or a local directory path appears.

Note: Ensure that the CDP Data Hub cluster EC2 instances are able to read from the S3 bucket path by checking the S3 permissions in the instance profile or role attached to the instances.

10. Depending on the location where you want to upload the Big Data Protector files, perform one of the following steps.

- a. Type *1* to generate the Big Data Protector installation files locally and upload to S3 bucket and continue with next step.
- b. Type *2* to create the Big Data Protector installation files on the local directory path, manually upload to the S3 bucket later, and skip to Step [14](#).

11. Press ENTER.

A prompt to select an option for the type of AWS Access Keys appears.

```
Choose the Type of AWS Access Keys from the following options:
[ 1 ] : IAM User Access Keys (Permanent access key id & secret access key)
[ 2 ] : Temporary Security Credentials (Temporary access key id, secret access key &
session token)
```

[1 or 2]:

12. Type *1* to use IAM User Access Keys.
 - a. Press *ENTER*.
A prompt for the Access Key ID appears.
 - b. Type the Access Key ID which has write access to the S3 bucket.
 - c. Press *ENTER*.
A prompt for the Secret Access Key appears.
 - d. Type the Secret Access Key which has write access to the S3 bucket.
13. Alternatively, if you select option 2, then enter the Session Token when prompted.
14. Press *ENTER*.

A prompt to enter the local directory path on this machine that stores the Fluent Bit configuration file appears.

Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:

15. Type the local directory path.

Note:

The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across the CDP nodes through the Cloudera Manager. Ensure that you name the custom Fluent Bit configuration file(s) for the external audit store with the extension **.conf*.

16. Press *ENTER*.

The process to generate the installation files starts and a prompt to enter the activated version of the *PTY_FLUENTBIT_CONF* parcel appears.

Generating Installation files...

NOTE:

You can verify the version of the activated *PTY_FLUENTBIT_CONF* parcel from the parcel name, such as *PTY_FLUENTBIT_CONF-x.x.x.x_CDPx.x.p<version>-<os>.parcel*, where the <version> parameter denotes the patch version of the *PTY_FLUENTBIT_CONF* parcel.

For Example: If the current activated *PTY_FLUENTBIT_CONF* parcel is *PTY_FLUENTBIT_CONF-x.x.x.x_CDPx.x.p0-<os>.parcel*, the patch version of the *PTY_FLUENTBIT_CONF* parcel will be 0. Please do NOT include 'p' while specifying the version.

Enter the <version> of the current *PTY_FLUENTBIT_CONF* Parcel as specified in the parcel name [0]:

17. Type the version of the activated *PTY_FLUENTBIT_CONF* parcel.

18. Press *ENTER*.

The installation files are generated in the following hierarchy:

```
Installation_Files/
CSDandParcels
  BDP_PEP-9.0.0.0.x.jar
  PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel
  PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
  PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel
  PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
  PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel
  PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
peimpala
  peimpala32_RHEL.so
  sqlscripts
    createobjects.sql
    dropobjects.sql
```



```

RecipesAndTemplates
  BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh
  BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh
  custom_properties_template.json
  guide_to_create_cluster_template_with_bdp.txt
UpdatedCERTParcel
  PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel
  PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel.sha
Updated_FLUENTBIT_CONF_Parcel
  PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p1-<os_version>.parcel
  PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p1-<os_version>.parcel.sha

```

19. If you select option *1* in step [10](#), then the updated *PTY_FLUENTBIT_CONF* parcel and the SHA checksum files are uploaded to the S3 bucket for ease of access.
20. If you select option *2* in step [10](#), then the updated *PTY_FLUENTBIT_CONF* parcel and the SHA checksum files are locally generated in the *./Installation_Files/Updated_FLUENTBIT_CONF_Parcel/* directory as mentioned in step [18](#).
21. Perform the following steps on the running Data Hub Cluster Cloudera Manager Server Node.
 - a. Copy the updated *PTY_FLUENTBIT_CONF* parcel and the *.sha* file to the running Data Hub cluster's Local parcel repository (*/opt/cloudera/parcel-repo*).
 - b. Change the ownership of the copied files to *cloudera-scm:cloudera-scm*.
 - c. On the Cloudera Manager UI, navigate to Parcels and click the **Check for New Parcels** button.
 - d. Distribute and activate the new parcel manually using the Cloudera Manager UI and replace the older *PTY_FLUENTBIT_CONF* parcel.
 - e. Restart any dependent services, such as, *BDP_PEP*.

Note:

For more information about common errors of the Big Data Protector, refer to the section *Big Data Protector Error Handling* in the [Protegility Troubleshooting Guide 9.1.0.0](#).

12.6.17 Installing the Big Data Protector on a CDP Azure Data Hub Platform

This section describes the tasks for installing Big Data Protector on a CDP Azure Data Hub platform.

For the Big Data Protector 9.0.0.0 release, the CDP Azure Data Hub platform, which includes Cloudera Runtime and Cloudera Manager, version 7.2.1, is used for reference.

For more information about CDP Azure Data Hub platform 7.2.1, refer to <https://docs.cloudera.com/runtime/7.2.1/index.html>.

The following list describes the installation workflow to install Big Data Protector on a CDP Azure Data Hub Platform.

- a. [Extract the Big Data Protector Package](#)
- b. [Run the Big Data Protector Configurator Script](#)
- c. [Register the Recipe Scripts](#)
- d. [Create and Register the Custom Cluster Template](#)
- e. [Create a Data Hub Cluster](#)
- f. [Update the Certificates Parcel](#)
- g. [Update the Fluent Bit Configuration Parcel](#)

12.6.17.1 Verifying the Prerequisites for Installing the Big Data Protector

Ensure that the following prerequisites are met, before creating the Data Hub cluster with the Big Data Protector:



- The user must have access to the CDP Management Console.
- The user must have the *PowerUser* role in CDP to create the Data Hub clusters.
- The user must have the *EnvironmentAdmin* resource role in the CDP Azure environment.

Note:

For more information about the CDP role and resource role, refer to <https://docs.cloudera.com/management-console/cloud/user-management/topics/mc-understanding-roles-resource-roles.html>.

- An Azure Storage Account with container (ADLS Gen 2) is available to upload the BDP Installation files. The Data Hub cluster instances should be able to read the Blobs from this container by default. If you plan to upload using the provided configurator script, then ensure that you have the Storage Account Access Key for authentication.
- If you plan to use the Spark Protector using the Spark client tools like *spark-shell* and *spark-submit* in the CDP Azure Data Hub, then ensure that the CDP users and groups in the CDP environment are accurately mapped to the Managed Identities to access Azure Storage using the IDBroker mappings.

Note:

To understand and set the IDBroker Mappings, refer to the following links:

- <https://docs.cloudera.com/cdp/latest/requirements-azure/topics/mc-az-onboarding-uag-for-cloud-storage.html>
- https://docs.cloudera.com/runtime/7.2.0/cdp-security-overview/topics/security_how_identity_federation_works_in_cdp.html

12.6.17.2 Extracting the Big Data Protector Package

You need to extract the Big Data Protector package to access the Big Data Protector configurator script to generate the Big Data Protector installation files.

► To extract the files from the installation package:

1. Login to the Linux machine that has connectivity to the ESA.
2. Download the Big Data Protector package *BigDataProtector_Linux-ALL-64_x86-64_CDP-Azure-7-64_9.0.0.0.x.tgz* to a directory.
3. Extract the *BDPConfigurator_CDP-Azure-DataHub-7_9.0.0.0.x.sh* file from the Big Data Protector installation package using the following command.
tar -xvf BigDataProtector_Linux-ALL-64_x86-64_CDP-Azure-7-64_9.0.0.0.x.tgz
4. Press ENTER.

The Big Data Protector Configurator script *BDPConfigurator_CDP-Azure-DataHub-7_9.0.0.0.x.sh* is extracted.

Note: The Configurator Script uses the *xxd* command when it uploads the files to Azure Storage. Therefore, ensure that the *xxd* tool is installed on the Linux machine. The *xxd* tool is installed with the *vim* (Vi IMproved) utility from the *vim-common* package.

12.6.17.3 Running the Big Data Protector Configurator Script

1. Run the configurator script.

./BDPConfigurator_CDP-Azure-DataHub-7_9.0.0.0.x.sh

The following prompt appears.

```
*****
        Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP Azure Data Hub.

Do you want to continue? [yes or no]
```

2. Type *yes*.
3. Press ENTER.

The following prompt appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ]      : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                        Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                        : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                        Use this if you want to set Custom Fluent-Bit configuration
files to
                        forward logs to an External Audit Store.
[ 1, 2 or 3 ]:
```

4. Type *1* to create the installation files of Big Data Protector.
5. If you want to update the certificate parcel after updating the ESA certificates, then type *2*.

Note: For more information about updating the certificate parcel, refer to the section [Updating the Certificates Parcel on a Running Azure Data Hub Cluster](#).

6. If you want to update the Fluent Bit configuration parcel, then type *3*.

Note:

For more information on updating the Fluent Bit configuration parcel, refer to the section [Updating the Fluent Bit Configuration Parcel on a Running Azure Data Hub Cluster](#).

7. Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el6 ]      : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ]    : SuSE Linux Enterprise Server 12.x

Please enter the no.:
```

Note: Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

8. Type *1* or *2* or *3* depending on the OS version used in CDP Data Hub cluster.
9. Press ENTER.



A prompt to enter the Azure Storage URL to upload the Big Data Protector installation files appears.

```
Enter the Azure Storage URL to upload the Big Data Protector installation files :  
(E.g.: https://storageaccount.blob.core.windows.net/container/folder) :
```

10. Enter the Azure Storage **https** URL (Blob/ADLS Gen2) which you want to use to upload the installation files.

Note:

Ensure that the CDP Data Hub cluster Azure VM instances are able to read from this Storage location by checking the permissions attached to the instances.

11. Press ENTER.

A prompt to select an option to upload the Big Data Protector installation packages to the Azure Storage or a local directory path appears.

```
Choose one option among the following for BDP Installation files:  
[ 1 ] : Upload files to Azure Storage URL <Azure Storage URL> via REST API.  
[ 2 ] : Generate files locally to current working directory. (You would have to manually  
upload the files to the specified Azure Storage)
```

```
[ 1 or 2 ]:
```

12. Depending on the location where you want to store the Big Data Protector files, perform one of the following steps.

- a. To generate the Big Data Protector installation files locally and then upload them to the Azure Storage, and continue with next step, type *1*.
- b. To generate the Big Data Protector installation files on the local directory path and manually upload to the specified Azure Storage URL later, and skip to Step *15*, type *2*.

13. Press ENTER.

The prompt to enter the access key for the Azure storage account appears.

```
Enter the Storage Account Access Key (Shared Key Authorization):
```

14. Type the access key.

15. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Please enter the ESA Host or IP Address[]:
```

16. Type the hostname or IP address for the ESA.

17. Press ENTER.

The prompt to enter the listening port for the ESA appears. The default value is *8443*.

```
Enter ESA host listening port [8443]:
```

18. Enter the ESA Listening Port.

If you want to use the default value of the ESA Listening Port, such as, *8443*, then leave the value empty.

19. Press ENTER.

A prompt for the ESA admin username appears.

20. Enter the ESA admin username.

21. Press ENTER.

A prompt for the ESA admin password appears.

22. Enter the ESA admin password.

23. Press ENTER.

The certificates are fetched from the ESA and a prompt to create the Fluent Bit configuration parcel appears.

```
Fetching Certificates from ESA....
```

| Total | % Received | % Xferd | Average Speed | Time Dload | Upload | Total | Time Spent | Time Left | Current Speed | |
|-------|------------|---------|---------------|------------|--------|-------|------------|-----------|---------------|--------------|
| 100 | 30720 | 100 | 30720 | 0 | 0 | 44522 | 0 | --::-- | --::-- | --::-- 44521 |

```
Do you want to package any custom Fluent-Bit configuration files for External Audit Store?  
[ yes ] : Create a PTY_FLUENTBIT_CONF parcel containing configuration files to be used  
with External Audit Store.  
[ no ] : Skip this step.
```

```
[ yes or no ] :
```

- To create the Fluent Bit configuration parcel, type *yes*.

Note:

If you type *no* at the prompt to create the *PTY_FLUENTBIT_CONF* parcel, then the installer will skip the creation of the Fluent Bit configuration parcel and proceed to generate the remaining installation files.

- Press ENTER.

The prompt to enter the local directory path that stores the Fluent Bit configuration file appears.

```
Creation of PTY_FLUENTBIT_CONF parcel is enabled.
```

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration  
files for External Audit Store:
```

- Type the local directory path that stores the Fluent Bit configuration files.

Note:

The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across the CDP nodes through the Cloudera Manager. Ensure that you name the custom Fluent Bit configuration file(s) for the external audit store with the extension **.conf*.

- Press ENTER.

The installation files are uploaded to the Azure storage in the *./Installation_Files/* directory.

```
Installation_Files/  
CSDandParcels  
    BDP_PEP-9.0.0.0.x.jar  
    PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel  
    PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha  
    PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel  
    PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha  
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel  
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha  
RecipesAndTemplates  
    BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh  
    BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh  
    custom_properties_template.json  
    guide_to_create_cluster_template_with_bdp.txt  
pepimpala  
    pepimpala32_RHEL.so  
    sqlscripts
```

```
createobjects.sql
dropobjects.sql
```

Note:

- If you select the upload option in step [12](#), then the installation files are generated locally and then uploaded to the Azure storage path in the same directory structure as mentioned in step [27](#).
- If you select the local directory path in step [12](#), then manually upload the Installation files that are present in the `./Installation_Files/` directory to the Azure storage path and maintain the same directory structure as mentioned in step [27](#). For example, the `./Installation_Files/pepimpala/pepimpalaXX_RHEL.so` file should be uploaded to `<Azure_storage_path>/pepimpala/pepimpalaXX_RHEL.so` directory.

12.6.17.4 Registering the Recipe Scripts

Note: The `BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh` script downloads the Big Data Protector CSD and Parcels from the Azure storage to the Cloudera Manager local CSD and Parcel repository before the Cloudera Manager server starts.

The `BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh` script runs after the Cloudera Manager Server starts. It creates and executes secondary scripts as background processes for each available Protegility Parcel. The background processes will check when the Cloudera Manager Server API endpoint would be open and then sends the requests to distribute and activate the `PTY_BDP`, `PTY_CERT`, and `PTY_FLUENTBIT_CONF`(if present) parcels.

The Recipe scripts execution logs can be found in the `/var/log/recipes/` directory by default and the execution logs of the secondary scripts (executing in the background) can be found in the `/tmp/protegility/` directory.

► Perform the following steps to register each recipe script:

1. On the Cloudera Management Console screen, navigate to **Shared Resources > Recipes**.
2. Click the **Register Recipe** button.
3. Enter the Recipe name.
4. Click the **Type** drop-down.
5. Select the recipe script type.

Note: You need to register both the `pre-cloudera-manager-start` and `post-cloudera-manager-start` recipe scripts.

6. Enter the optional recipe description.
7. Select the **File** option to upload a file that contains the recipe script.

To upload the `pre-cloudera-manager-start` recipe script, select the `BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh` script. To upload the `post-cloudera-manager-start` recipe script, select the `BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh` script.

8. Click the **Register** button.

The registration of the `pre-cloudera-manager-start` and `post-cloudera-manager-start` recipes are completed.

Refer the following images for registering the recipes.

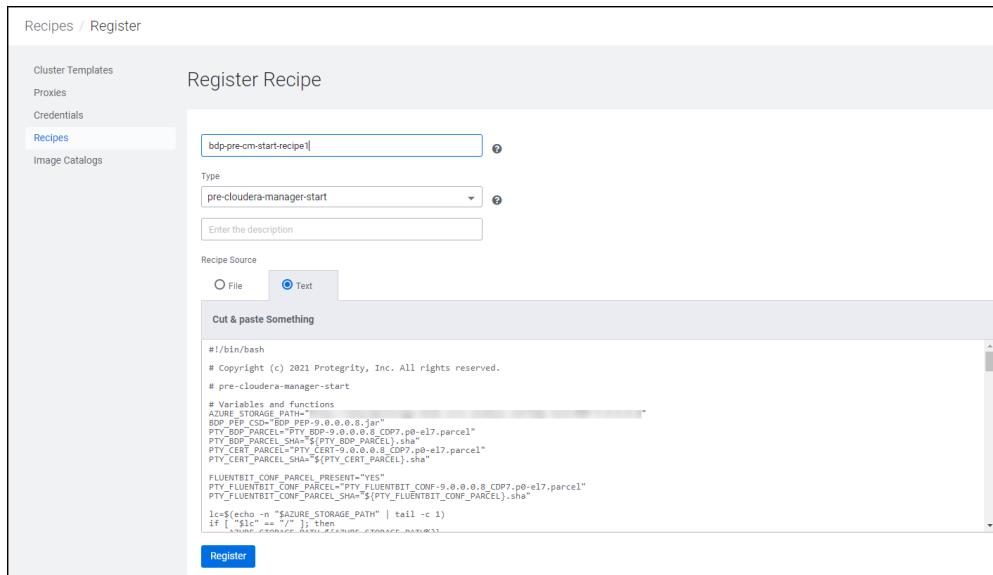


Figure 12-383: bdp-pre-cm-start-recipe

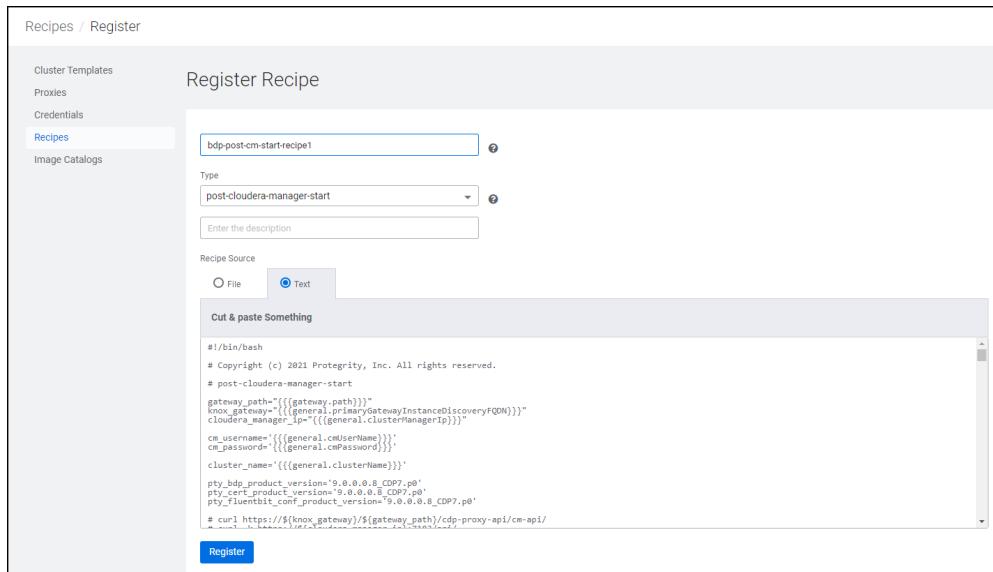


Figure 12-384: bdp-post-cm-start-recipe

12.6.17.5 Creating and Registering the Custom Cluster Template

You need to create and register the custom cluster template to add *BDP_PEP* service and required service configurations to the Data Hub cluster.

► To Create the Custom Cluster Template with the *BDP PEP* service:

1. On the Cloudera Management Console screen, navigate to **Shared Resources > Cluster Templates**.

| Name | Platform | Group Count | Description | Tags | Time Created |
|--|------------------|-------------|--------------------|--|----------------------------------|
| datamart[REDACTED]9008 | Cloudera Runtime | 7.2.6 | | | 12/9/2021, 12:10:41 PM GMT+5:30 |
| 7.2.6-BDP 9.0.0.x with Data Engineering | Cloudera Runtime | 7.2.6 | | | 12/9/2021, 12:11:02 PM GMT+5:30 |
| 7.2.10 - BDP 9.0.0.x In Data Engineering | Cloudera Runtime | 7.2.10 | | | 12/9/2021, 12:48:56 PM GMT+5:30 |
| bdp-9.0.0.x-cluster-template | Cloudera Runtime | 7.2.12 | | | 12/9/2021, 12:49:56 PM GMT+5:30 |
| cdp-on-aws-9-peest | Cloudera Runtime | 7.2.6 | | | 11/12/2021, 1:45:00 PM GMT+5:30 |
| 7.2.12 - SDX Micro Duty: Apache Hive Metastore, Apache Ranger, Apache Atlas | Cloudera Runtime | 7.2.12 | Testing cdp on aws | 7.2.12 - Micro SDX Template with Atlas, HAMS, Ranger and other services they are dependent on | 11/17/2021, 12:18:49 PM GMT+5:30 |
| 7.2.10-DataEngineering with BDP 9.0.0.0 and custom conf | Cloudera Runtime | 7.2.10 | | | 11/19/2021, 12:21:57 PM GMT+5:30 |
| test3 | Cloudera Runtime | 7.2.6 | | | 11/12/2021, 1:51:26 PM GMT+5:30 |
| 7.2.6-DataEngineering with BDP 9.0.0.0 and custom conf | Cloudera Runtime | 7.2.6 | | 7.2.6-DataEngineering with BDP 9.0.0.0 and all configs for Hive and Spark | 11/12/2021, 12:49:44 PM GMT+5:30 |
| LogForwarderRoleTemplate_5 | Cloudera Runtime | 7.2.6 | | LogForwarderRoleTemplate_5 | 11/20/2021, 1:48:12 PM GMT+5:30 |
| LogForwarderRoleTemplate_4 | Cloudera Runtime | 7.2.6 | | LogForwarderRoleTemplate_4 | 11/20/2021, 1:29:54 PM GMT+5:30 |
| LogForwarderRoleTemplate_3 | Cloudera Runtime | 7.2.6 | | LogForwarderRoleTemplate_3 | 11/20/2021, 1:42:29 PM GMT+5:30 |
| LogForwarderRoleTemplate_2 | Cloudera Runtime | 7.2.6 | | LogForwarderRoleTemplate_2 | 11/20/2021, 1:12:54 PM GMT+5:30 |
| LogForwarderRoleTemplate_1 | Cloudera Runtime | 7.2.6 | | LogForwarderRoleTemplate_1 | 11/20/2021, 1:37:38 AM GMT+5:30 |
| LogForwarderRole_Template | Cloudera Runtime | 7.2.6 | | | 10/29/2021, 1:29:24 AM GMT+5:30 |
| 7.2.12 - Streams Messaging Heavy Duty: Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | Cloudera Runtime | 7.2.12 | | 7.2.12 - Streams Messaging Heavy Duty with Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | 9/5/2021, 4:07:35 PM GMT+5:30 |
| 7.2.12 - Streams Messaging Light Duty: Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | Cloudera Runtime | 7.2.12 | | 7.2.12 - Streams Messaging Light Duty with Apache Kafka, Schema Registry, Streams Messaging Manager, Streams Replication Manager, Cruise Control | 9/5/2021, 4:07:35 PM GMT+5:30 |

Figure 12-385: Cluster Templates Screen

2. Copy the JSON content of the required cluster template that you want to use in a text editor.
For example, Data Engineering, Data Mart
3. On the cluster template JSON, search for the `services` key, whose value is a JSON array of JSON objects.
4. Add the following JSON objects to that array of services:

```
{
  "refName": "bdp_pep",
  "serviceType": "BDP_PEP",
  "roleConfigGroups": [
    {
      "refName": "bdp_pep-PEP_SERVER-BASE",
      "roleType": "PEP_SERVER",
      "base": true,
      "configs": [
        {
          "name": "esa_address",
          "value": "{{{esa_address}}}"
        }
      ]
    },
    {
      "refName": "bdp_pep-PEP_LOGFORWARDER-BASE",
      "roleType": "PEP_LOGFORWARDER",
      "base": true,
      "configs": [
        {
          "name": "elasticsearch_ip_port_list",
          "value": "{{{elasticsearch_ip_port_list}}}"
        },
        {
          "name": "auditstore_type",
          "value": "{{{auditstore_type}}}"
        }
      ]
    }
  ]
}
```

```
        ]
    }
```

Note: The service object is position-independent within the array and can be placed at the beginning or end of the array.

Note: Adding the *BDP_PEP* service to the array of services will ensure that the service is added to the Data Hub cluster during the cluster creation, when the Cloudera Manager imports the cluster template.

Note: Ensure that the values, such as the `{{ esa_address }}` should be written as it is (called Mustache template "`\"{{ ... }}\"`"). The actual value is set by adding the custom properties during the creation of the CDP data hub cluster. For more information about the format of the custom properties, check the *custom_properties_template.json* file in the Azure storage containing installation files of Big Data Protector.

For more information about the installation files of Big Data Protector, refer to step 27 in the section [Running the Configurator Script](#).

5. After adding the *bdp_pep* service object, search for the *hostTemplates* key in the cluster template, whose value is an array of *hostTemplate* objects for master, worker, and compute nodes etc.
6. For each *hostTemplate* in that array, search for the key *roleConfigGroupsRefNames*, which has a value of array of strings.
7. Add the *bdp_pep-PEP_SERVER-BASE* and the *bdp_pep-PEP_LOGFORWARDER-BASE* strings in the *roleConfigGroupsRefNames* array.

```
{
  ...
  "hostTemplates": [
    {
      "refName": "...",
      "cardinality": "...",
      "roleConfigGroupsRefNames": [
        ...
        "bdp_pep-PEP_SERVER-BASE",
        "bdp_pep-PEP_LOGFORWARDER-BASE"
      ],
      ...
    },
    ...
  ],
  ...
}
```

Note:

- The *PTY Pep Server* and *PTY Logforwarder* roles are installed on each of the *hostTemplate* instances.
- If you are planning to use the Hive on Tez service in the Data Hub cluster, then refer to section [Setting the Hive on Tez Service Configuration](#).
- If you are planning to use the HBase service in the Data Hub cluster, then refer to section [Setting the HBase Configuration](#).
- If you are planning to use the Spark on Yarn service in the Data Hub cluster, then refer to section [Setting the Spark on Yarn Service Configuration](#).

8. Click the **Create Cluster Template** button.

The **Cluster Template Register** screen appears.

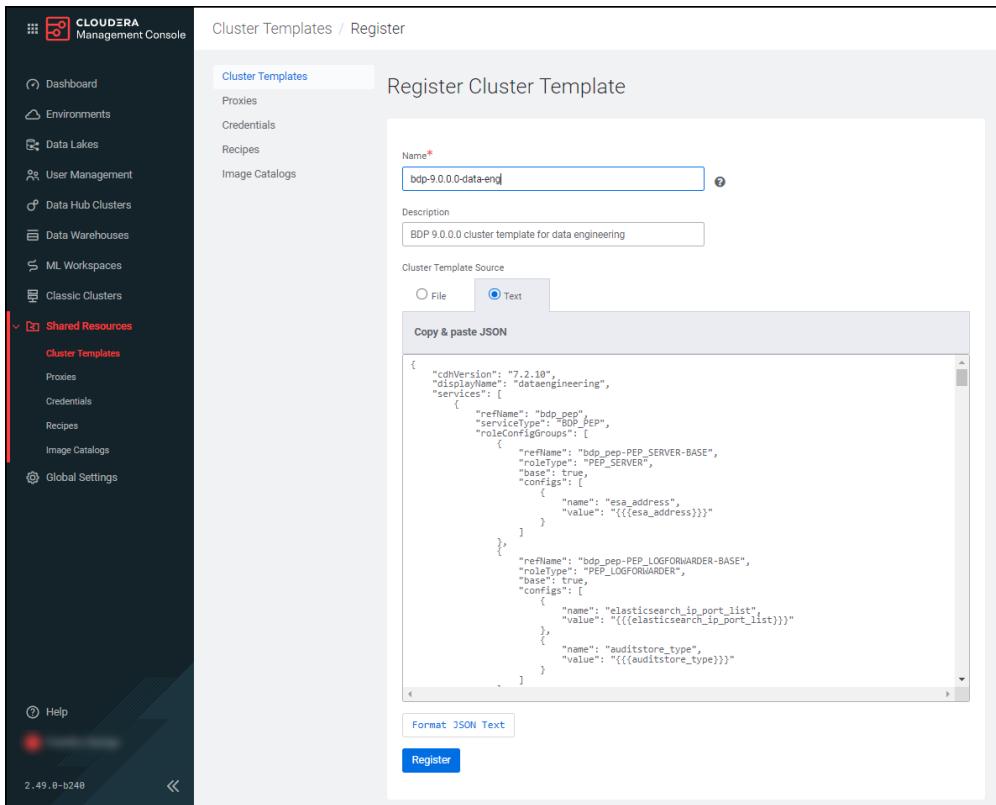


Figure 12-386: Cluster Template Register

9. Enter a cluster template name.
10. Enter an optional description for the cluster template.
11. Select any of the following options to select the Cluster Template Source.
 - a. Select the **Text** option to paste cluster template in JSON format.
 - b. Select the **File** option to upload a file that contains the cluster template.
 - c. Select the **URL** option to select the URL for the cluster template.
12. Click the **Register** button to register the custom Cluster Template on the CDP Management Console.

12.6.17.6 Creating a Data Hub Cluster

To create a new Data Hub cluster with Big Data Protector, use the registered cluster template with the two Recipes generated and custom properties.

Note:

- For more information about using the registered cluster template, refer to the section [Creating and Registering the Custom Cluster Template](#).
- For more information about registering the recipe scripts, refer to the section [Registering the Recipe Scripts](#).

► To create a Data Hub Cluster:

1. On the CDP Management Console, click the **Data Hub Clusters** tab.
The Data Hub screen appears.

The screenshot shows the Cloudera Management Console interface. On the left, a sidebar menu includes options like Dashboard, Environments, Data Lakes, User Management, Data Hub Clusters (which is currently selected), Data Warehouses, ML Workspaces, Classic Clusters, Shared Resources, and Global Settings. The main content area is titled "Data Hubs" and displays a list of six Data Hub Clusters:

- CDH 7.2.6**: Environment aws, Nodes 6, Created At 12/06/21, 01:06 PM GMT+5:30. Status: Running.
- CDH 7.2.6**: Environment aws, Nodes 4, Created At 12/09/21, 12:27 PM GMT+5:30. Status: Running.
- CDH 7.2.6**: Environment aws, Nodes 6, Created At 12/08/21, 11:45 AM GMT+5:30. Status: Stopped.
- CDH 7.2.6**: Environment aws, Nodes 4, Created At 12/07/21, 06:26 PM GMT+5:30. Status: Stopped.
- CDH 7.2.10**: Environment aws, Nodes 4, Created At 11/16/21, 06:59 PM GMT+5:30. Status: Stopped.
- CDH 7.2.6**: Environment aws, Nodes 4, Created At 12/09/21, 02:14 PM GMT+5:30. Status: Running.

A blue "Create Data Hub" button is located in the top right corner of the main content area. The URL in the browser bar is <https://console.us-west-1.cdp.cloudera.com/cloud/workloads/details/panimal-damart/hardware>.

Figure 12-387: Data Hub Screen

- Click the **Create Data Hub** button.
- The **Provision Data Hub** screen appears.

The screenshot shows the "Provision Data Hub" screen. The left sidebar is identical to Figure 12-387. The main area is titled "Provision Data Hub" and contains the following sections:

- Selected Environment**: pty-cdp-az (selected from a dropdown).
- Cluster Definition**: Custom tab is selected.
- Services**: A list of services to be installed, including Cloudera Runtime 7.2.10, Cluster Template bdp-9.0.0.0-data-eng, and various components like Data Analytics Studio, Hdfs, Hive, Hue, Livy, Oozie, Queue Manager, Spark, Yarn, and Zeppelin.
- General Settings**: Fields for Cluster Name (bdp-data-eng-test) and Tags (Add button).
- Action Buttons**: Provision Cluster, Save As New Definition, Show CLI Command, and Show Generated Cluster Template.

Figure 12-388: Provision Data Hub Screen

- Select the CDP Azure environment.
- Select the **Custom** tab.
- In the **Cluster Template** drop-down, select the previously created customized cluster template with the Big Data Protector.

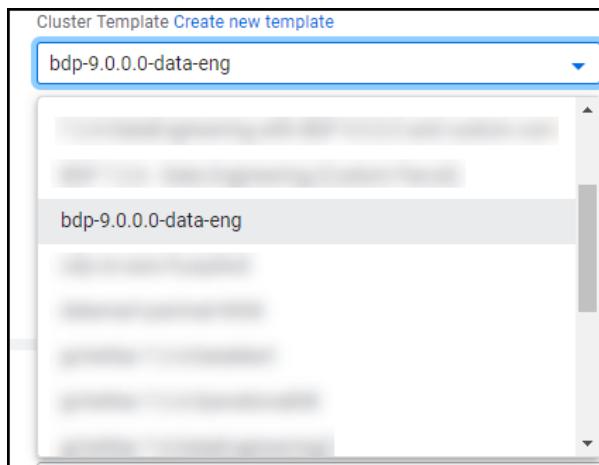


Figure 12-389: Cluster Template Drop-down

6. Set the cluster name.
7. Click **Advanced Options**.
8. Select the **Hardware and Storage** tab.

Note the host group which will install the Cloudera Manager (CM) Server. Usually, the host is the master node.

9. Select the **Cluster Extensions** tab.

A screenshot of the 'Cluster Extensions' tab. On the left, there is a sidebar with tabs: 'Image Catalog', 'Network And Availability', 'Hardware And Storage', 'Cloud Storage', and 'Cluster Extensions', with 'Cluster Extensions' being the active tab. The main area is titled 'Recipes' and contains four sections: 'Master Node', 'Worker Nodes', 'Compute Node', and 'Gateway Node'. Each section has a 'Please select a recipe' input field and an 'Attach' button. Below each input field, it says 'Attached Recipes' and 'There are no attached Recipes.'

Figure 12-390: Cluster Extensions Tab

10. Attach the two previously registered Recipes to the host group that would host the Cloudera Manager Server.

Note: For more information about registering the recipe, refer to section [Registering the Recipe Scripts](#). Both the Recipes should be executed on the Cloudera Manager Server node of the cluster.

11. Add the contents of `custom_properties_template.json` file to the Custom Properties section generated in step 27 of section [Running the Configurator Script](#). Replace the placeholder strings of the storage account access key, ESA hostname / IP address, Protegility ElasticSearch appliance list of IP/ports, and the Audit Store type with the actual values.



Figure 12-391: Custom Properties

Warning: After the cluster startup, the Cloudera manager username and password are seen in clear in the log files generated in the `/var/log/recipes/` directory.

12. Click Provision Cluster.

The cluster creation process starts.

12.6.17.7 Setting the Big Data Protector Configuration for the CDP Azure Data Hub Platform

While creating a custom cluster template, ensure that you add the required service configuration as mentioned in the sections below.

The configuration for the Big Data Protector can be set using any one of the following options:

- Setting the configuration using a cluster template
- Setting the configuration after creating the cluster

12.6.17.7.1 Setting the Configuration Using a Cluster Template

Set the Big Data Protector configuration in the custom cluster template by modifying the cluster template.

12.6.17.7.1.1 Setting the Hive on Tez Service Configuration

If you want to use the Hive on Tez service in the Data Hub cluster (cluster template), then set the custom Big Data Protector Hive properties.

► To configure Hive on Tez service in the cluster template:

1. On the cluster template JSON, in the services JSON array, search for the JSON object with the key service type as `HIVE_ON_TEZ`.
2. Add the following JSON object to the serviceConfigs array.

```
{
  "name": "HIVE_ON_TEZ_service_env_safety_valve",
  "value": "HIVE_CLASSPATH=/opt/cloudera/parcels/PTY_BDP/pephive/lib/pephive-*.jar:/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:$HIVE_CLASSPATH\nJAVA_LIBRARY_PATH=/opt/cloudera/parcels/PTY_BDP/jpeplite/lib:$JAVA_LIBRARY_PATH"
}
```

3. If the Data Analytics Studio service (DAS) is installed with Hive, then append the following string in the value of serviceConfig JSON object having name as `hive_service_config_safety_valve`.

```
<property><name>hive.exec.pre.hooks</name><value>com.protegility.hive.PtyHiveUserPreHook,org.apache.hadoop.hive ql.hooks.HiveProtoLoggingHook</value></property>
```

For example, after appending the string, the final serviceConfig *hive_service_config_safety_valve* object should be as follows.

```
{
    "name": "hive_service_config_safety_valve",
    "value": "<property><name>fs.s3a.ssl.channel.mode</name><value>openssl</value></
property><property><name>hive.txn.acid.dir.cache.duration</name><value>0</value></
property><property><name>hive.exec.pre.hooks</
name><value>com.protegility.hive.PtyHiveUserPreHook,org.apache.hadoop.hive.ql.hooks.HivePro
toLoggingHook</value></property>"
}
```

4. If the Data Analytics Studio service (DAS) is not installed with Hive, then append the following string in the value of the *serviceConfig* JSON object having name as *hive_service_config_safety_valve*.

For example, after appending the string, the final serviceConfig *hive_service_config_safety_valve* object should be as follows.

```
{
    "name": "hive_service_config_safety_valve",
    "value": "<property><name>fs.s3a.ssl.channel.mode</
name><value>openssl</value></property><property><name>hive.txn.acid.dir.cache.duration</
name><value>0</value></property><property><name>hive.exec.pre.hooks</
name><value>com.protegility.hive.PtyHiveUserPreHook</value></property>"
}
```

5. In the *roleConfigGroups* array, search for the JSON object with refName *hive_on_tez-GATEWAY-BASE*.
6. Add the following JSON object to the configs array.

```
{
    "name": "hive_client_env_safety_valve",
    "value": "HIVE_CLASSPATH=/opt/cloudera/parcels/PTY_BDP/
pephive/lib/pephive-*.jar:/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:$HIVE_CLASSPATH
\nJAVA_LIBRARY_PATH=/opt/cloudera/parcels/PTY_BDP/jpeplite/lib:$JAVA_LIBRARY_PATH"
}
```

After setting the Hive on Tez service configuration, continue the steps mentioned in section [Setting the Tez Service Configuration](#).

12.6.17.7.1.2 Setting the Tez Service Configuration

To use the Tez service in the Data Hub cluster (cluster template), set the custom Tez properties.

► To set the Tez Service Configuration:

1. On the cluster template JSON, in the services JSON array, search for the JSON object with the key service type as *TEZ*.
2. Add the following JSON objects to the serviceConfigs array.

```
{
    "name": "tez.am.launch.env",
    "value": "LD_LIBRARY_PATH=/opt/cloudera/parcels/CDH/lib/hadoop/lib/native:/opt/
cloudera/parcels/PTY_BDP/jpeplite/lib"
},
{
    "name": "tez.cluster.additional.classpath.prefix",
    "value": "/opt/cloudera/parcels/PTY_BDP/jpeplite/lib/*:/opt/cloudera/parcels/
PTY_BDP/pephive/lib/*"
}
```

12.6.17.7.1.3 Setting the Impala Service Configuration

If you want to use the Impala service in the Data Hub cluster (Data Mart cluster template), then note that the *pepimpala*.so* Protegity library is not a part of the *PTY_BDP* parcel. The *pepimpala* library must be stored in the Azure storage and accessed using the *abfs://*URI when registering the UDFs.

The *createobjects.sql* script displays the SQL syntax to create the Protegity UDFs.

The *dropobjects.sql* script displays the SQL syntax to drop the Protegity UDFs.

12.6.17.7.1.4 Setting the HBase Configuration

If you want to use the HBase service in the Data Hub cluster (cluster template), then set the custom Big Data Protector configurations for HBase.

► To configure the HBase service in the cluster template:

1. On the cluster template JSON, in the *roleConfigGroups* array, search for the JSON object with *refName* *hbase-REGIONSERVER-BASE*.
2. Add the following JSON object to the *Configs* array.

```
{
    "name": "hbase_coprocessor_region_classes",
    "value": "com.protegity.hbase.PTYRegionObserver"
}
```

12.6.17.7.1.5 Setting the Spark on Yarn Service Configuration

If you want to use the Spark on Yarn service in the Data Hub cluster (cluster template), then set the custom Big Data Protector configurations for Spark on Yarn. The following settings are only valid for Spark versions greater than or equal to 2.4.0 and lower than 3.0.0.

► To configure the Spark on Yarn service in the cluster template:

1. On the cluster template JSON, in the *services* JSON array, search for the JSON object with the key *serviceType* as *SPARK_ON_YARN*.
2. In the *roleConfigGroups* array, search for the JSON object with *refName* as *spark_on_yarn-GATEWAY-BASE*.
3. In the *configs* array, search for *spark-conf/spark-defaults.conf_client_config_safety_valve* and append *|nspark.executor.plugins=com.protegity.spark.PtyExecSparkPlugin* to the value.

```
{
    "name": "spark-conf/spark-defaults.conf_client_config_safety_valve",
    "value": "spark.hadoop.fs.s3a.ssl.channel.mode=openssl\nspark.hadoop.mapreduce.fileoutputcommitter.algorithm.version=
1\nspark.executor.plugins=com.protegity.spark.PtyExecSparkPlugin"
}
```

12.6.17.7.2 Setting the Configuration after Creating the Cluster

After the cluster starts all the services, then set the Big Data Protector configuration using the Cloudera Manager UI, and restart the services with stale configurations.

Table 12-98: Recommended Configuration for Big Data Protector

| Service | Protector Jar Configuration | Protegility Native Library Configuration |
|---|--|--|
| Hive on Tez | <p>In <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: HIVE_CLASSPATH Value: /opt/cloudera/ parcels/PTY_BDP/pephive/lib/ pephive-*.*jar:/opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:\$ {HIVE_CLASSPATH}</pre> <p>In <i>Hive Service Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-site.xml</i>:</p> <pre>Name: hive.exec.pre.hooks Value: com.protegility.hive.PtyHiveUserPreHook</pre> | <p>In <i>Hive on Tez Service Environment Advanced Configuration Snippet (Safety Valve)</i> and <i>Gateway Client Environment Advanced Configuration Snippet (Safety Valve)</i> for <i>hive-env.sh</i>:</p> <pre>Key: JAVA_LIBRARY_PATH Value: /opt/cloudera/ parcels/PTY_BDP/jpeplite/ lib:\${JAVA_LIBRARY_PATH}</pre> |
| Tez | <pre>Name: tez.cluster.additional.class path.prefix Value: /opt/cloudera/ parcels/PTY_BDP/ jpeplite/lib/*:/opt/ cloudera/parcels/PTY_BDP/ pephive/lib/*</pre> | <pre>Name: tez.am.launch.env = Value: LD_LIBRARY_PATH=/opt/ cloudera/parcels/CDH/lib/ hadoop/lib/native:/opt/ cloudera/parcels/PTY_BDP/ jpeplite/lib</pre> |
| HBase | <pre>Name: hbase.coprocessor.region.clas ses Value: com.protegility.hbase.PTYRegi onObserver</pre> | |
| Spark on Yarn (Spark versions greater than or equal to 2.4.0 and lower than 3.0.0.) | <p>In 'Spark Client Advanced Configuration Snippet (Safety Valve)' for <i>spark-conf/spark-defaults.conf</i>:</p> <pre>spark.executor.plugins=com.p rotegility.spark.PtyExecSpark Plugin</pre> | |

12.6.17.8 Updating the Certificates Parcel on a Azure Data Hub Cluster

If the customers have updated the certificates on the ESA, with which the Big Data Protector is configured, then the Certificates parcel must be updated with the new certificates. The updated Certificates parcel will be utilized by the nodes in the cluster.



► To update the *PTY_CERT* parcel on a running Data Hub cluster:

- Run the Configurator script.

```
./BDPConfigurator_CDP-Azure-DataHub-7_9.0.0.0.x.sh
```

The following prompt appears.

```
*****
Welcome to the Big Data Protector Configurator Wizard
*****
This will setup the Big Data Protector Installation Files for CDP Azure Data Hub.

Do you want to continue? [yes or no]
```

- Type *yes*.
- Press ENTER.

The following prompt appears.

```
Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ] : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                           Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ] : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                           Use this if you want to set Custom Fluent-Bit configuration
files to
                           forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

- Type *2* to update the Certificate parcel after updating the ESA certificates.
- Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```
Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.
```

```
[ 1: el6 ] : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ] : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ] : SuSE Linux Enterprise Server 12.x
```

Please enter the no.:

Note: Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

- Type *1* or *2* or *3* depending on the OS version used in CDP Data Hub cluster.
- Press ENTER.

A prompt to enter the Azure Storage URL to upload the Big Data Protector installation files appears.

```
Enter the Azure Storage URL to upload the Big Data Protector installation files :
(E.g.: https://storageaccount.blob.core.windows.net/container/folder) :
```

- Enter the Azure Storage URL where you want to upload the Big Data Protector installation files.

A prompt to select an option to upload the Big Data Protector installation packages to the Azure Storage URL or a local directory path appears.

```
Choose one option among the following for BDP Installation files:  
[ 1 ] : Upload files to Azure Storage URL <Azure_storage_URL> via REST API.  
[ 2 ] : Generate files locally to current working directory. (You would have to manually  
upload the files to the specified Azure Storage)
```

[1 or 2]:

Note: Ensure that the CDP Data Hub cluster Azure VM instances are able to read from this storage location by checking the permissions attached to the instances.

9. Depending on the location where you want to upload the Big Data Protector files, perform one of the following steps.
 - a. Type 1 to generate the Big Data Protector installation files locally and upload to Azure Storage and continue with next step.
 - b. Type 2 to create the Big Data Protector installation files on the local directory path, manually upload to the Azure Storage later, and skip to Step [12](#).

10. Press ENTER.

The prompt to enter the access key for the Azure storage account appears.

```
Enter the Storage Account Access Key (Shared Key Authorization):
```

11. Type the access key.

12. Press ENTER.

The prompt to enter the hostname or IP address for the ESA appears.

```
Please enter the ESA Host or IP Address[]:
```

13. Type the hostname or IP address for the ESA.

14. Press ENTER.

The prompt to enter the listening port for the ESA appears. The default value is *8443*.

```
Enter ESA host listening port [8443]:
```

15. Enter the ESA Listening Port.

If you want to use the default value of the ESA Listening Port, such as, *8443*, then leave the value empty.

16. Press ENTER.

A prompt for the ESA admin username appears.

17. Enter the ESA admin username.

18. Press ENTER.

A prompt for the ESA admin password appears.

19. Enter the ESA admin password.

20. Press ENTER.

The ESA certificates are successfully downloaded and the prompt to enter the patch version of the *PTY_CERT* parcel appears.

```
% Total % Received % Xferd Average Speed Time Time Time Current  
Dload Upload Total Spent Left Speed  
100 30720 100 30720 0 0 179k 0 --:--:-- --:--:-- --:--:-- 179k
```

Generating Installation files...

NOTE:

You can verify the version of the activated *PTY_CERT* parcel from the parcel name, such as *PTY_CERT-x.x.x.x_CDPx.x.p<version>-<os>.parcel*, where the

<version> parameter denotes the patch version of the PTY_CERT parcel.

For Example: If the current activated PTY_CERT parcel is PTY_CERT-x.x.x.x_CDPx.x.p0-<os>.parcel, the patch version of the PTY_CERT parcel will be 0. Please do NOT include 'p' while specifying the version.

Enter the <version> of the current PTY_CERT Parcel as specified in the parcel name [0]:

21. Enter the patch version of the currently activated *PTY_CERT* parcel on the Data Hub cluster.

Note:

The patch version is the number after the *p* in the parcel name. For example, the *PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel* has a patch version 0. The updated *PTY_CERT* parcel and SHA checksum with an incremented patch version is created locally in the same working directory, such as, *./Installation_Files/UpdatedCERTParcel*/ directory.

22. If you select option *1* in step 9, then the updated *PTY_CERT* parcel and SHA checksum files are uploaded to the Azure storage for ease of access.
23. If you select option *2* in step 9, then the updated *PTY_CERT* parcel and SHA checksum files are locally generated in the *./Installation_Files/UpdatedCERTParcel*/ directory.

```
*****
***** The updated PTY_CERT parcel 'PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel' and sha1
***** checksum are locally generated in ./Installation_Files/UpdatedCERTParcel/ directory.
***** -> Please manually copy the PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel and .sha files
***** to Cloudera Manager Server's local parcel repository on the existing running Data Hub
***** cluster.
```

```
*****
***** Successfully configured the Updated PTY_CERT parcel for CDP Azure DataHub.
```

24. Perform the following steps on the running Data Hub Cluster Cloudera Manager Server Node.
 - a. Copy the updated *PTY_CERT* parcel and *.sha* file to the running Data Hub cluster's Local parcel repository (*/opt/cloudera/parcel-repo*).
 - b. Change the ownership of the copied files to *cloudera-scm:cloudera-scm*.
 - c. On the Cloudera Manager UI, navigate to *Parcels* and click the **Check for New Parcels** button.
 - d. Distribute and activate the new parcel manually using the Cloudera Manager UI and replace the older *PTY_CERT* parcel.
 - e. Restart any dependent services, such as, *BDP_PEP*.

12.6.17.9 Updating the Fluent Bit Configuration Parcel on a Azure Data Hub Cluster

If the user wants to use a newer set of custom Fluent Bit configuration files for sending the logs to an external audit store, then the Fluent Bit parcel must be updated, distributed, and activated across the cluster nodes through the Cloudera Manager.

1. Login to the host machine, which contains the Big Data Protector configurator script.
2. Run the *./BDPConfigurator_CDP-Azure-DataHub-7_9.0.0.0.x.sh* script.
A prompt to continue the configuration of Big Data Protector appears.
3. Type *yes* to start the configuration of the Big Data Protector.
4. Press ENTER.



The following prompt appears.

```

Big Data Protector Configurator started...
Unpacking...
Extracting files...

Please select the type of Installation files you want to generate.
[ 1: Create All ]      : Creates entire Big Data Protector CSDs, Parcels, Recipes and
other files.
[ 2: Update PTY_CERT ] : Creates new PTY_CERT parcel with an incremented patch version.
                         Use this if you have updated the ESA certificates.
[ 3: Update PTY_FLUENTBIT_CONF ]
                         : Creates new PTY_FLUENTBIT_CONF parcel with an incremented patch
version.
                         Use this if you want to set Custom Fluent-Bit configuration
files to
                         forward logs to an External Audit Store.

[ 1, 2 or 3 ]:
```

5. Type *3* to update the Fluent Bit configuration parcel.
6. Press ENTER.

A prompt to select an OS version for the Cloudera Manager parcel appears.

```

Please select the OS version for Cloudera Manager Parcel.
This will be used as the OS Distro suffix in the Parcel name.

[ 1: el6 ]      : RHEL 6 and clones (CentOS, Scientific Linux, etc)
[ 2: el7 ]      : RHEL 7 and clones (CentOS, Scientific Linux, etc)
[ 3: sles12 ]    : SuSE Linux Enterprise Server 12.x

Please enter the no.:
```

Note:

Currently, the Enterprise Linux, versions 6 and 7 (RHEL and CentOS), and SLES 12 operating systems are supported.

7. Type *1* or *2* or *3* depending on the OS version used in the CDP Data Hub cluster.
8. Press ENTER.

A prompt to enter the Azure Storage URL to upload the Big Data Protector installation files appears.

9. Enter the Azure Storage URL where you want to upload the Big Data Protector installation files.

A prompt to select an option to upload the Big Data Protector installation packages to the Azure Storage URL or a local directory path appears.

```

Choose one option among the following for BDP Installation files:
[ 1 ] : Upload files to Azure Storage URL <Azure_Storage_URL> via REST API.
[ 2 ] : Generate files locally to current working directory. (You would have to manually
upload the files to the specified Azure Storage)

[ 1 or 2 ]:
```

Note: Ensure that the CDP Data Hub cluster Azure VM instances are able to read from this storage location by checking the permissions attached to the instances.

10. Depending on the location where you want to upload the Big Data Protector files, perform one of the following steps.
 - a. Type *1* to generate the Big Data Protector installation files locally and upload to the Azure Storage and continue with next step.
 - b. Type *2* to generate the Big Data Protector installation files on the local directory path, manually upload to the Azure Storage later, and skip to Step *12*.
11. Press ENTER.



A prompt to enter the access key for the storage account appears.

```
Enter the Storage Account Access Key (Shared Key Authorization):
```

12. Press ENTER.

A prompt to enter the local directory path on this machine that stores the Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

13. Type the local directory path.

Note:

The *PTY_FLUENTBIT_CONF* parcel is used to package any custom Fluent Bit configuration files that the user provides and can be distributed across the CDP nodes through the Cloudera Manager. Ensure that you name the custom Fluent Bit configuration file(s) for the external audit store with the extension **.conf*.

14. Press ENTER.

The process to generate the installation files starts and a prompt to enter the activated version of the *PTY_FLUENTBIT_CONF* parcel appears.

```
Generating Installation files...
```

NOTE:

You can verify the version of the activated *PTY_FLUENTBIT_CONF* parcel from the parcel name, such as *PTY_FLUENTBIT_CONF-x.x.x.x_CDPx.x.p<version>-<os>.parcel*, where the *<version>* parameter denotes the patch version of the *PTY_FLUENTBIT_CONF* parcel.

For Example: If the current activated *PTY_FLUENTBIT_CONF* parcel is *PTY_FLUENTBIT_CONF-x.x.x.x_CDPx.x.p0-<os>.parcel*, the patch version of the *PTY_FLUENTBIT_CONF* parcel will be 0. Please do NOT include 'p' while specifying the version.

```
Enter the <version> of the current PTY_FLUENTBIT_CONF Parcel as specified in the parcel name [0]:
```

15. Type the version of the activated *PTY_FLUENTBIT_CONF* parcel.

16. Press ENTER.

The installation files are generated in the following hierarchy.

```
Installation_Files/
CSDandParcels
    BDP_PEP-9.0.0.0.x.jar
    PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel
    PTY_BDP-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
    PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel
    PTY_CERT-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p0-<os_version>.parcel.sha
RecipesAndTemplates
    BDP_Post-CM-Start_Recipe_9.0.0.0.x.sh
    BDP_Pre-CM-Start_Recipe_9.0.0.0.x.sh
    custom_properties_template.json
    guide_to_create_cluster_template_with_bdp.txt
UpdatedCERTParcel
    PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel
    PTY_CERT-9.0.0.0.x_CDP7.p1-<os_version>.parcel.sha
Updated_FLUENTBIT_CONF_Parcel
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p1-<os_version>.parcel
    PTY_FLUENTBIT_CONF-9.0.0.0.x_CDP7.p1-<os_version>.parcel.sha
pepimpala
    pepimpala32_RHEL.so
    sqlscripts
        createobjects.sql
        dropobjects.sql
```

17. If you select option *1* in step [10](#), then the updated *PTY_FLUENTBIT_CONF* parcel and SHA checksum files are uploaded to the Azure storage for ease of access.
18. If you select option *2* in step [10](#), then the updated *PTY_FLUENTBIT_CONF* parcel and SHA checksum files are are locally generated in the *./Installation_Files/Updated_FLUENTBIT_CONF_Parcel/* directory as mentioned in step [16](#).
19. Perform the following steps on the running Data Hub Cluster Cloudera Manager Server Node.
 - a. Copy the updated *PTY_FLUENTBIT_CONF* parcel and *.sha* file to the running Data Hub cluster's Local parcel repository (*/opt/cloudera/parcel-repo*).
 - b. Change the ownership of the copied files to *cloudera-scm:cloudera-scm*.
 - c. On the Cloudera Manager UI, navigate to *Parcels* and click the **Check for New Parcels** button.
 - d. Distribute and activate the new parcel manually using the Cloudera Manager UI and replace the older *PTY_FLUENTBIT_CONF* parcel.
 - e. Restart any dependent services, such as, *BDP_PEP*.

Note:

For more information about common errors of the Big Data Protector, refer to the section *Big Data Protector Error Handling* in the [Protegility Troubleshooting Guide 9.1.0.0](#).

12.7 Installing and Uninstalling Database Protectors

This section provides information about installing and uninstalling the following Database Protectors:

- MS SQL Database Protector
- Oracle Database Protector
- Teradata Database Protector
- Greenplum Database Protector
- Netezza Database Protector
- DB2 Database Protector
- Presto Protector

12.7.1 Installing and Uninstalling the MS SQL Database Protector

This section describes the procedure to install and uninstall the MS SQL Database Protector.

12.7.1.1 Verifying the Prerequisites

This section describes the prerequisites including the hardware, software, and network requirements for installing the MS SQL database protector.

The following are the prerequisites for installing the MS SQL Database Protector.

- The ESA 9.1.0.0 appliance is installed, configured, and running.
- The IP address or host name of the ESA is noted.
- The administrator rights for the operating system are granted.
- The DBA rights to the MS SQL database are granted.

- Before installing the protector, ensure that the Policy Information Management (PIM) is initialized, if you are installing the ESA for the first time. This prerequisite holds true for versions 7.2.0 and later releases.

Note: For more information about initializing the PIM, refer to the section *Initializing the Policy Management* in the *Protegility Policy Management Guide 9.1.0.0*.

12.7.1.2 Installing the Log Forwarder

The Log Forwarder sends audit logs of protect, unprotect operations to the ESA. This section describes the steps to install the Log Forwarder.

► To Install the Log Forwarder:

- Download the *DatabaseProtector_WIN-ALL-64_x86-64_MSSQL-ALL-64_9.1.0.0.x.zip* installation package made available by Protegility.
- To install the MS SQL protector, create a directory on the machine where you want to install the protector.

Note: If you do not create a directory to install the MS SQL protector, then it is installed in the default *C : \Program Files\Protegility* directory.

- To extract the files from the installation package to the created directory, right-click the file.
A context menu appears.
- From the context menu, select **Extract All**.
The following files are extracted from the *DatabaseProtector_WIN-ALL-64_x86-64_MSSQL-ALL-64_9.1.0.0.x.zip* file:
 - LogforwarderSetup_Windows_x64_9.1.0.0.x.exe*
 - PepServerSetup_Windows_x64_9.1.0.0.x.exe*
 - PepSQLServerSetup_Windows_x64_9.1.0.0.x.exe*
 - U.S.Patent.No.6,321,201.Legend.txt*
- To install the Log Forwarder, run the *LogforwarderSetup_Windows_x64_9.1.0.0.x.exe* file.
The **Welcome to the Logforwarder Setup Wizard** appears.

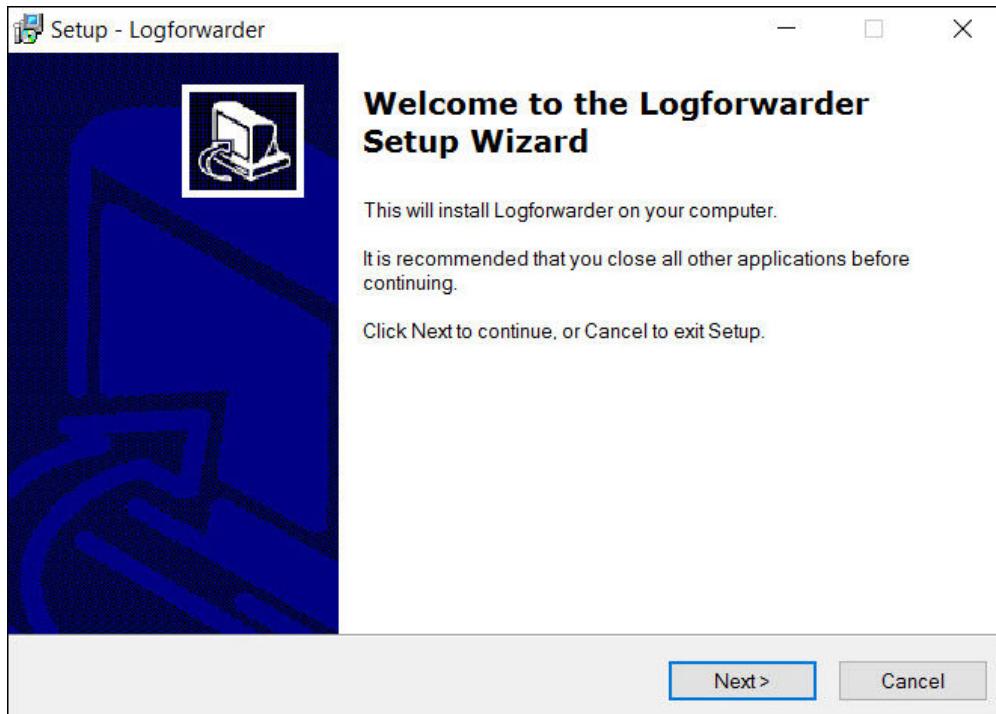


Figure 12-392: Welcome to the Logforwarder Setup Wizard

Caution: It is mandatory to install the Log Forwarder before installing the PEP server to ensure that the MS SQL Database Protector is configured correctly.

6. Click **Next**.
The **Audit Store Connectivity Information** screen appears.

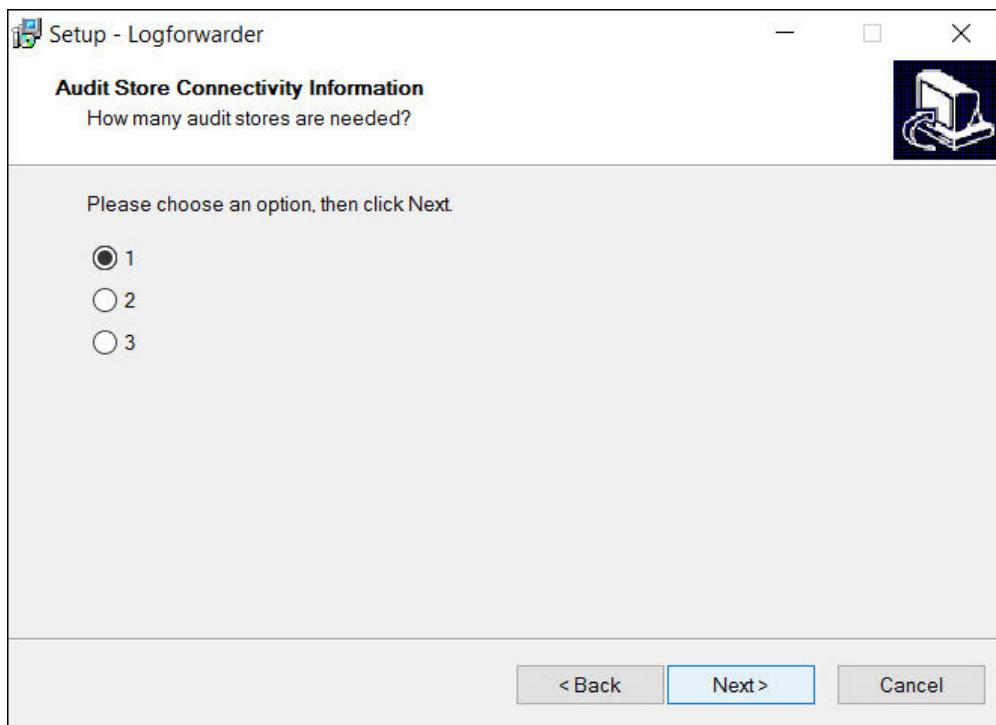


Figure 12-393: Number of Audit Stores

7. On the **Audit Store Connectivity Information** screen, select the number of Audit Stores as required.

Note: You can enter a minimum of 1 to a maximum 3 Audit Store endpoints.

8. Click **Next**.

The **Audit Store Connectivity Information** screen appears.

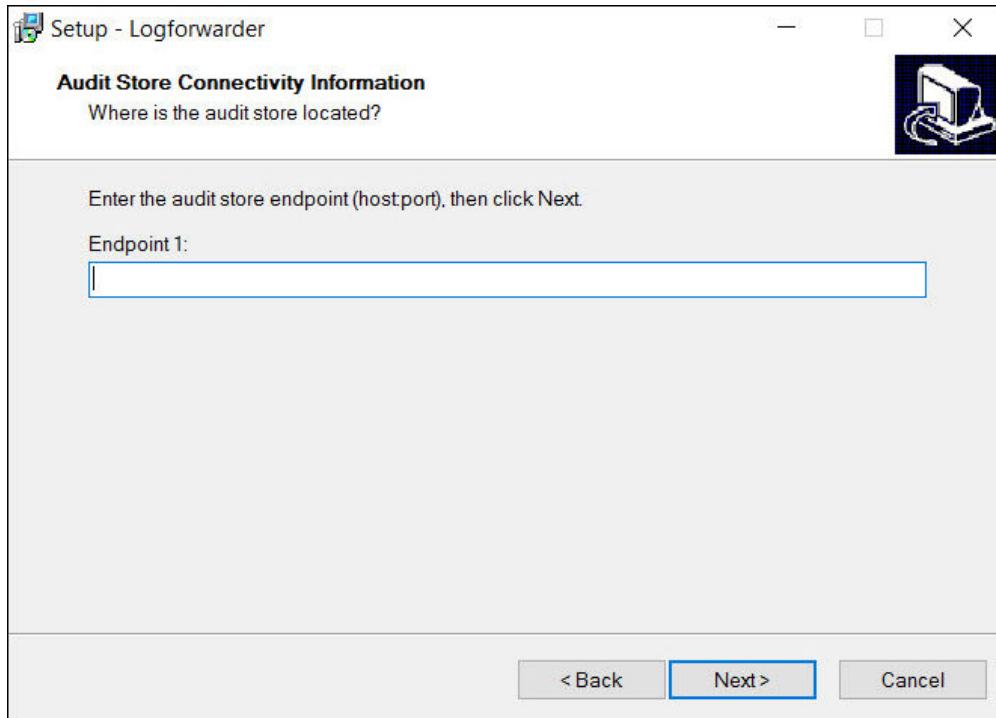


Figure 12-394: Location of Audit Store

9. On the **Audit Store Connectivity Information** screen, in the **Endpoint 1** box, enter the host and port of the ESA for the required Audit Store endpoint.

10. Click **Next**.

The **Select pepserver location** screen appears.

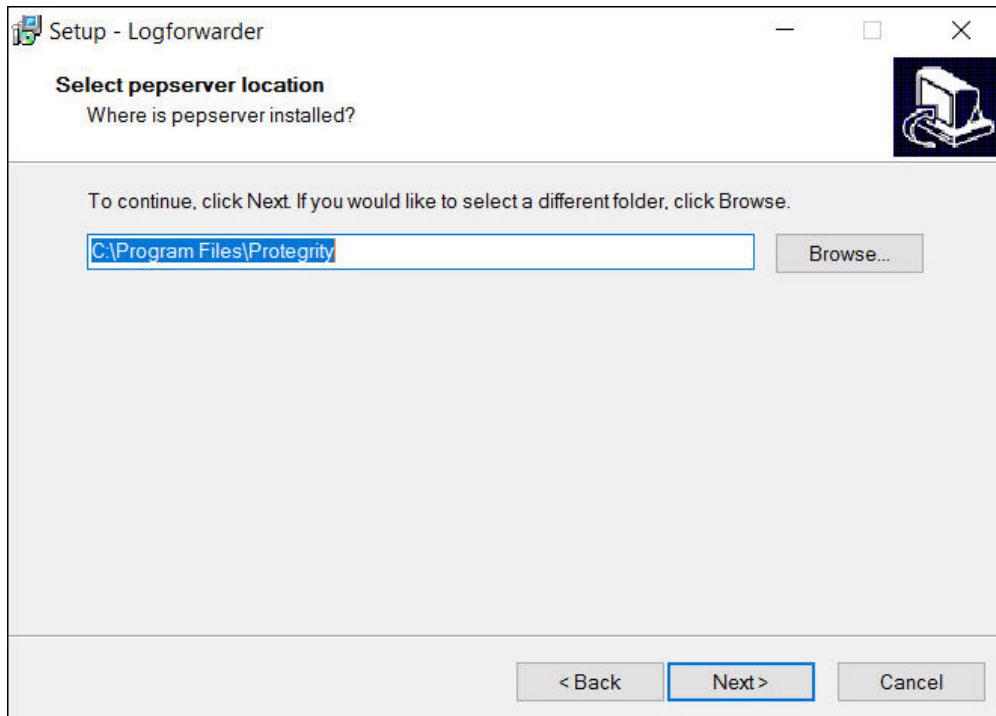


Figure 12-395: Select PEP server location

11. To change the default installation directory for the PEP server, perform the following steps.
 - a. Click **Browse**.
 - b. Select the required installation directory where you want to install the PEP server.
 - c. Click **OK**.

Note: The default installation directory is *C : \Program Files\Protegility* where the PEP server is installed.

12. Click **Next**.
The **Select Destination Location** screen appears.

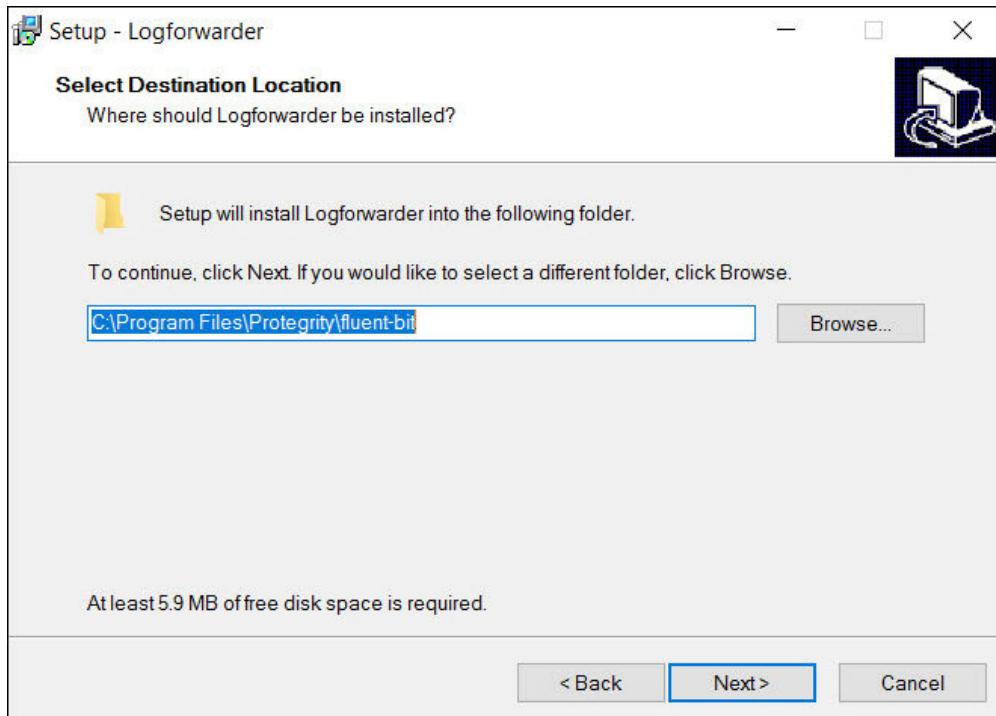


Figure 12-396: Select Destination Location for Logforwarder Setup

13. To change the default installation directory for Log Forwarder, perform the following steps.
 - a. Click **Browse**.
 - b. Select the required installation directory where you want to install the Log Forwarder.
 - c. Click **OK**.

Note:

- The default installation directory is *C:\Program Files\Protegility\fluent-bit* where the Log Forwarder is installed.
- The Log Forwarder component is herein referred to as Fluent Bit.

14. Click **Next**.
The **Ready to Install** screen appears.

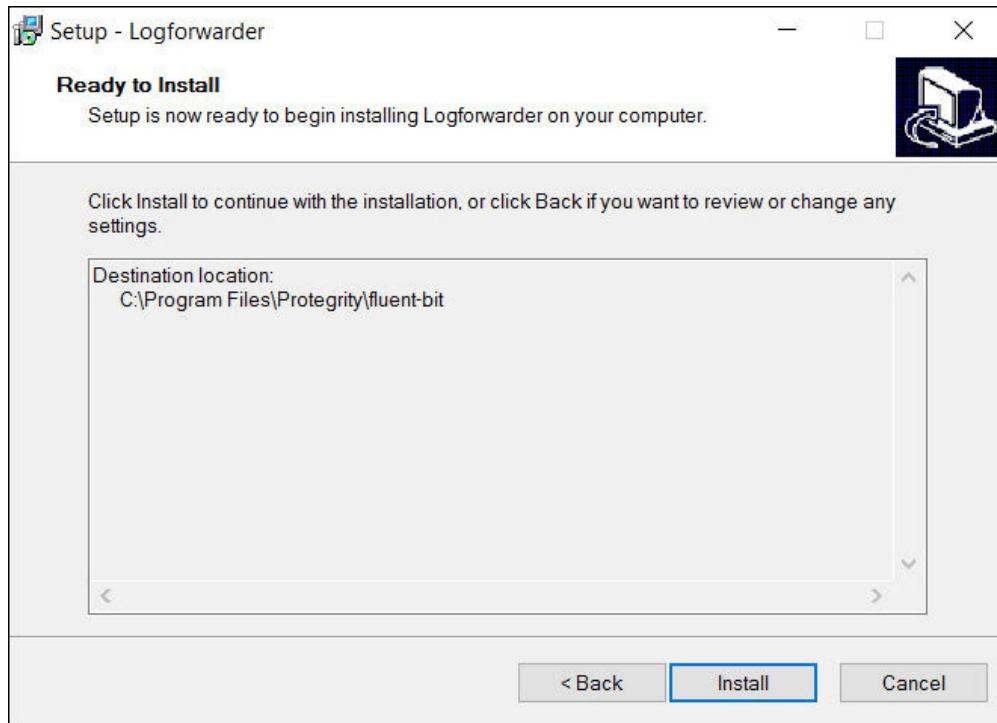


Figure 12-397: Ready to Install screen

15. Click **Install**.

The installer completes the installation of the Log Forwarder in the specified location and the **Completing the Logforwarder Setup Wizard** screen appears.

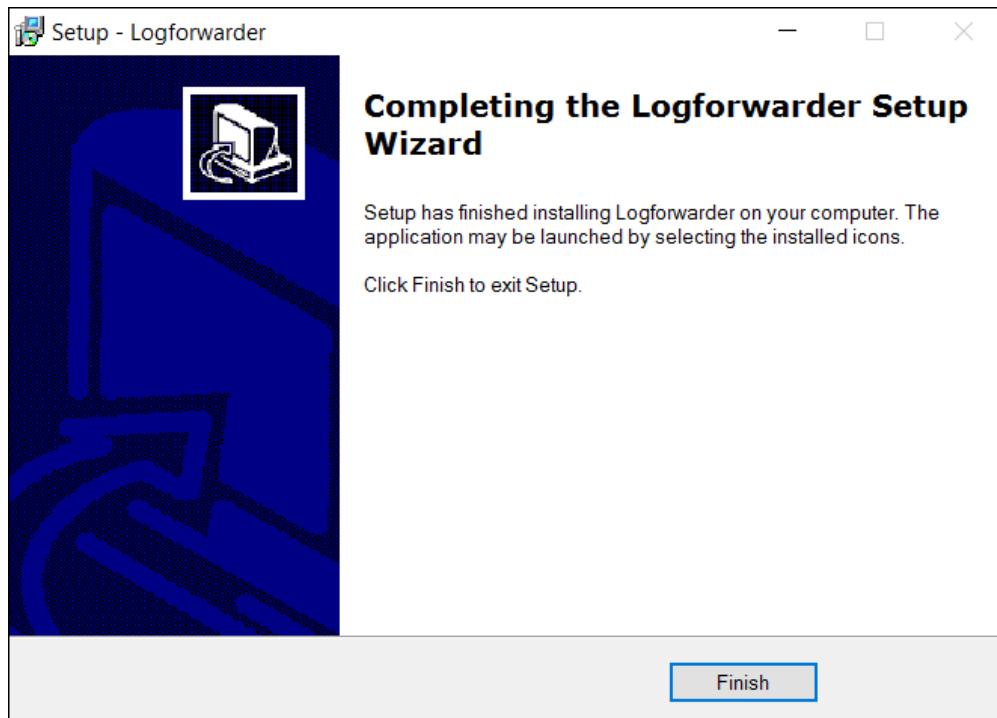


Figure 12-398: Completing the Logforwarder Setup Wizard

16. To exit the installation wizard, click **Finish**.

The **Completing the Logforwarder Setup Wizard** exits.

Note: After the Log Forwarder is installed successfully, to ensure that the Logforwarder sends logs to the ESA, verify that the Log Forwarder service is in *Running* state.

17. To verify the Logforwarder service state, perform the following step.
 - a. Navigate to **Start > Control Panel > System and Security > Administrative Tools > Services.**

Note:

- For more information about configuring the Log Forwarder, refer to the section [Appendix A: PEP Server Configuration File](#).
- After installing the MS SQL Database Protector, in the *pepservice.cfg* file, if you set the *Logging configuration* parameter as *mode=drop*, then you must drop the SQL objects and recreate them.
- For more information about starting the Log Forwarder while running multiple instances, refer to the section [Resolving the Port Collisions in the Log Forwarder](#).

18. Verify that the **Logforwarder** service is in *Running* state.

12.7.1.3 Installing the PEP Server

The PEP server is the connection between the ESA and the MS SQL Database Protector. The PEP server is responsible for accepting the policy that is deployed from the ESA. It also sends back the audit logs to the ESA. This section describes the steps to install the PEP server.

The following set of certificate files are downloaded in the *<installation_directory>\defiance_dps\data* directory:

- *authesa.plm*
- *CA.pem*
- *cert.key*
- *cert.pem*
- *certkeyup.bin*
- *keyinternal.plm*
- *pepservice.cfg*
- *pepservice.pid*

► To Install the PEP server:

1. Navigate to the directory where you have extracted the installation files.
2. To install the PEP server, run the *PepServerSetup_Windows_x64_9.1.0.0.x.exe* file.
The **Welcome to the Protegility PEP Server Setup Wizard** appears.

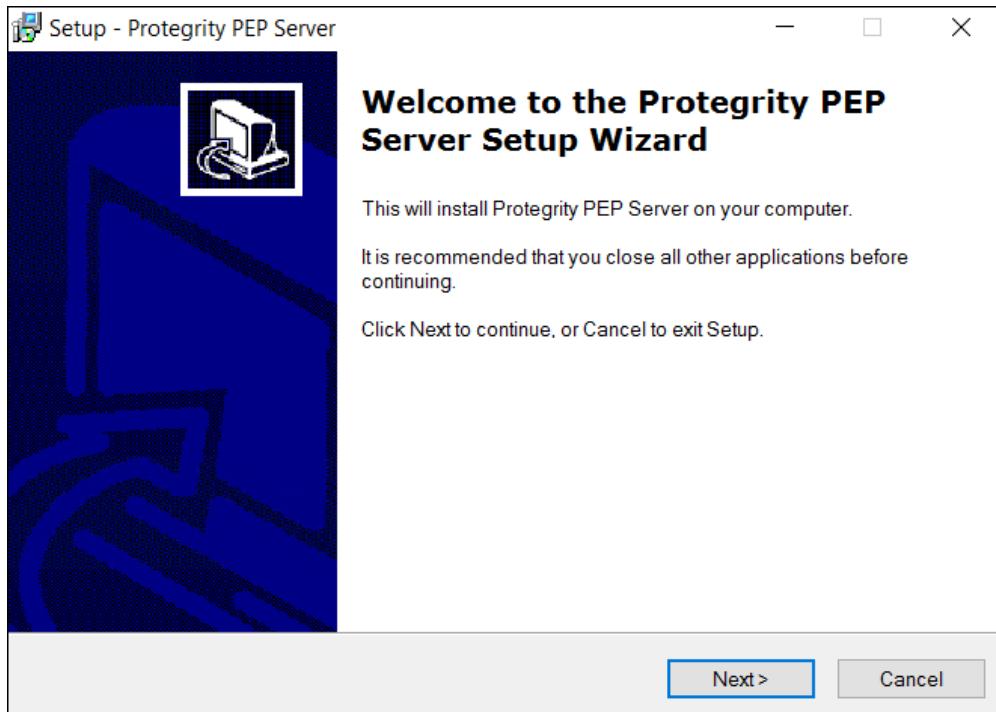


Figure 12-399: Welcome to the Protegility PEP Server Setup Wizard

3. Click **Next**.
The **ESA Connectivity Information** screen appears.

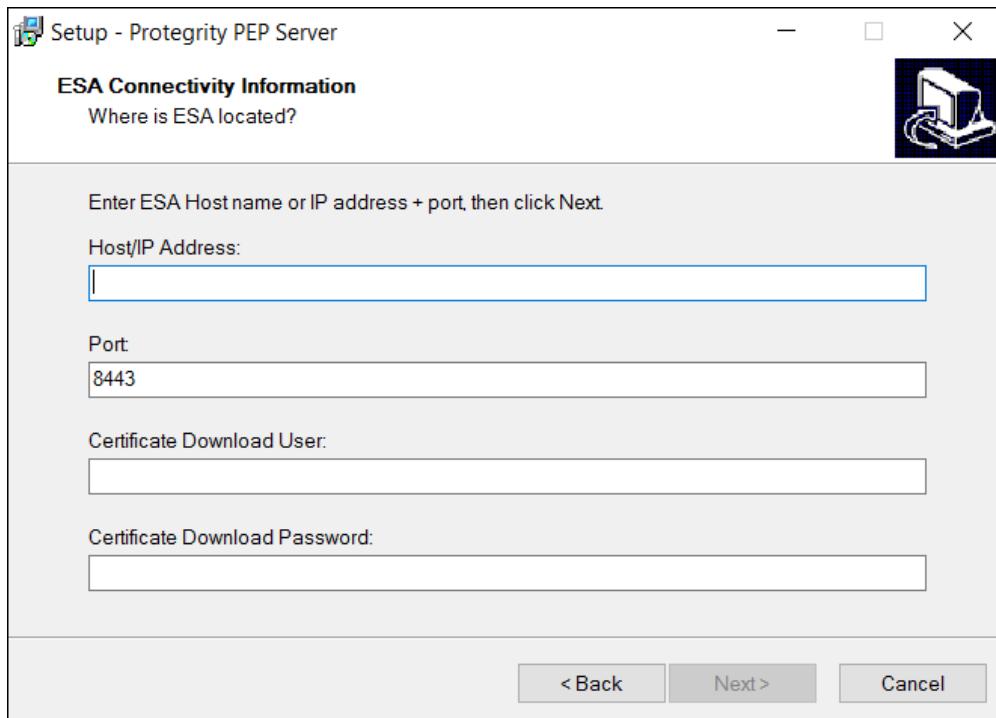


Figure 12-400: ESA Connectivity Information screen

4. In the **Host/IP Address** box, enter the host name or the IP address of the ESA.

Caution: Ensure that the ESA is up and running with the *HubController* service. The *HubController* service must be in *Running* state to enable the downloading of certificates automatically.

5. In the **Port** box, enter the port number of the ESA.

Note: You can retain the default port number of the ESA, 8443.

6. In the **Certificate Download User** box, enter the username of the ESA.

Note: It is recommended to use *Admin* user.

7. In the **Certificate Download Password** box, enter the password for the respective username.

8. Click **Next**.

The **Select Destination Location** screen appears.

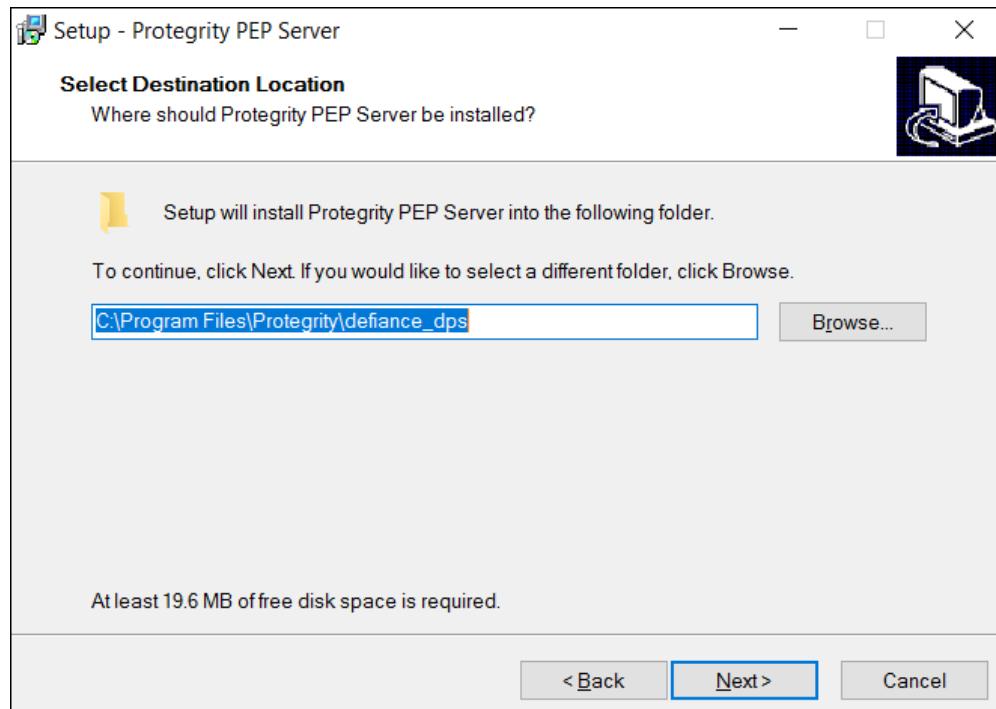


Figure 12-401: Select Destination Location screen

9. To change the default installation directory for the PEP server, perform the following steps.

- a. Click **Browse**.
- b. Select the required installation directory where you want to install the PEP server.
- c. Click **OK**.

Note: The default directory is *C:\Program Files\Protegility\defiance_dps* where the PEP server is installed.

10. Click **Next**.

The **Ready to Install** screen appears.

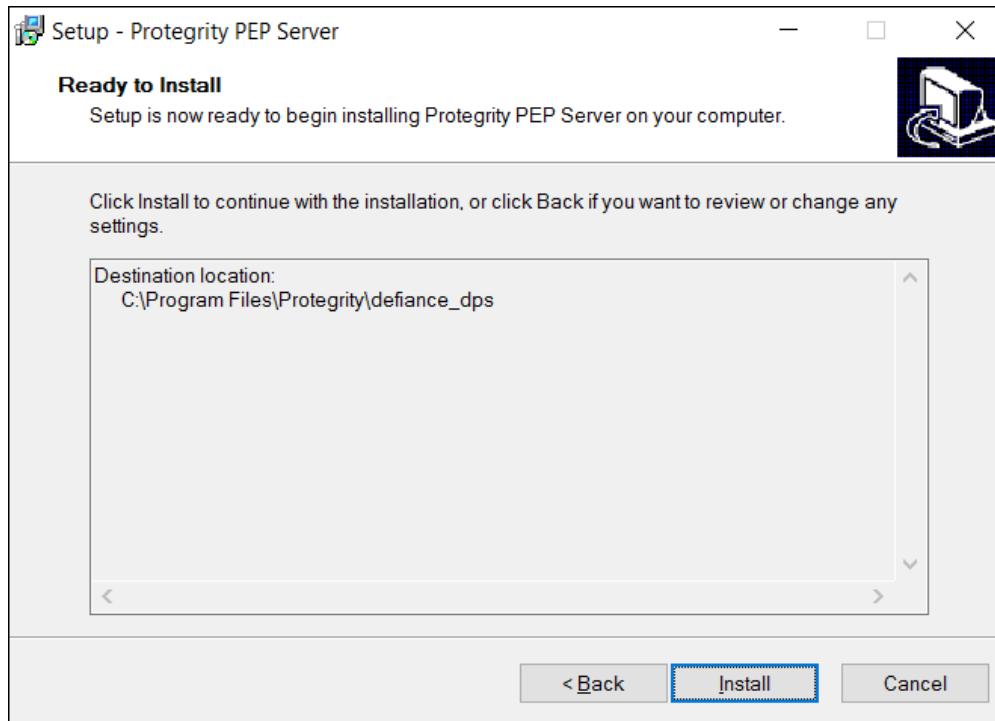


Figure 12-402: Ready to Install screen

11. Click **Install**.

The installer completes the installation of the PEP server successfully in the specified location and the **Completing the Protegility PEP Server Setup Wizard** screen appears.

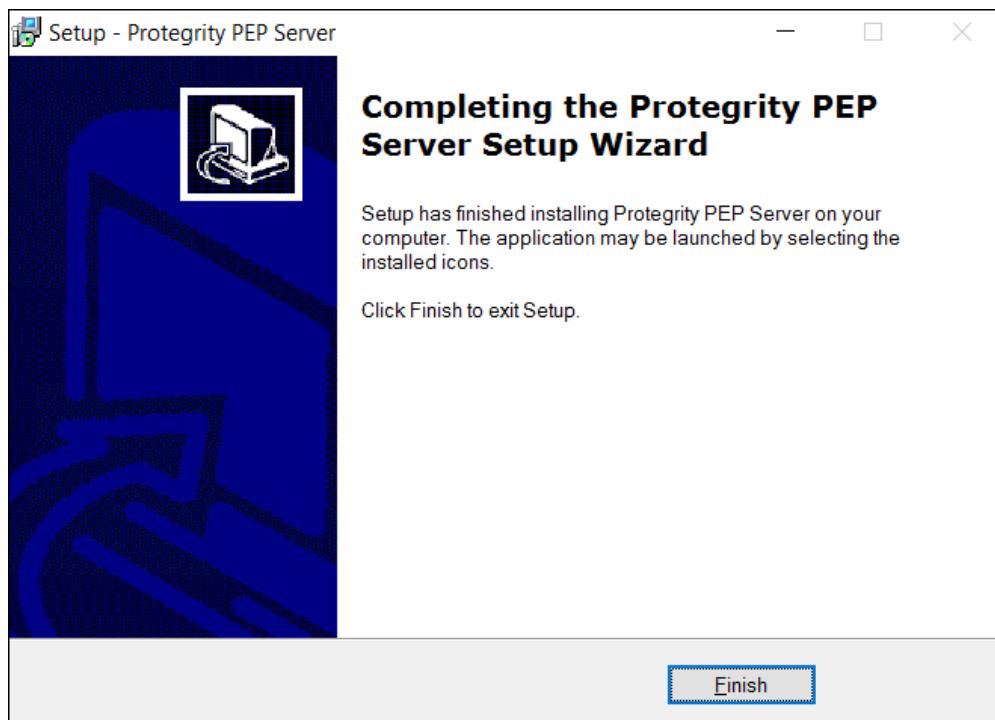


Figure 12-403: Completing the Protegility PEP Server Setup Wizard

12. To exit the installation wizard, click **Finish**.

The **Completing the Protegility PEP Server Setup Wizard** exits.

Note:

After the PEP server is installed successfully, ensure that the PEP server service is in *Running* state. The PEP server fetches the policy in the protector shared memory. If the PEP server service is not running, then the policy will not be deployed in the ESA.

13. To verify the PEP server service state, perform the following step.
 - a. Navigate to **Start > Control Panel > System and Security > Administrative Tools > Services**.
14. Verify that the **Protegility PEP Server** service is in *Running* state.

12.7.1.4 Installing the PEP for MS SQL Server

The PEP for MS SQL Server runs security operations such as protect, unprotect data. This section describes the steps to install the PEP for MS SQL Server.

After the PEP for MS SQL Server are installed successfully, the following scripts are downloaded in the `<install_dir>\Protegility\Database_Protector\pep\sqlscripts\sqlserver` directory:

- *0.0DropObjects.sql*
- *1.0.CreateAssembly.sql*
- *2.0.CreateFunctions.sql*

► To Install the PEP for MS SQL Server:

1. Navigate to the directory where you have extracted the installation files.
2. To install the PEP for MS SQL server, run the *PepSQLServerSetup_Windows_x64_9.1.0.0.x.exe* file. The **Welcome to the Protegility Data Security Platform – SQL Server UDF Setup Wizard** appears.



Figure 12-404: Welcome to the Protegility Data Security Platform - SQL Server UDF Setup Wizard

3. Click **Next**.
The **Select Destination Location** screen appears.

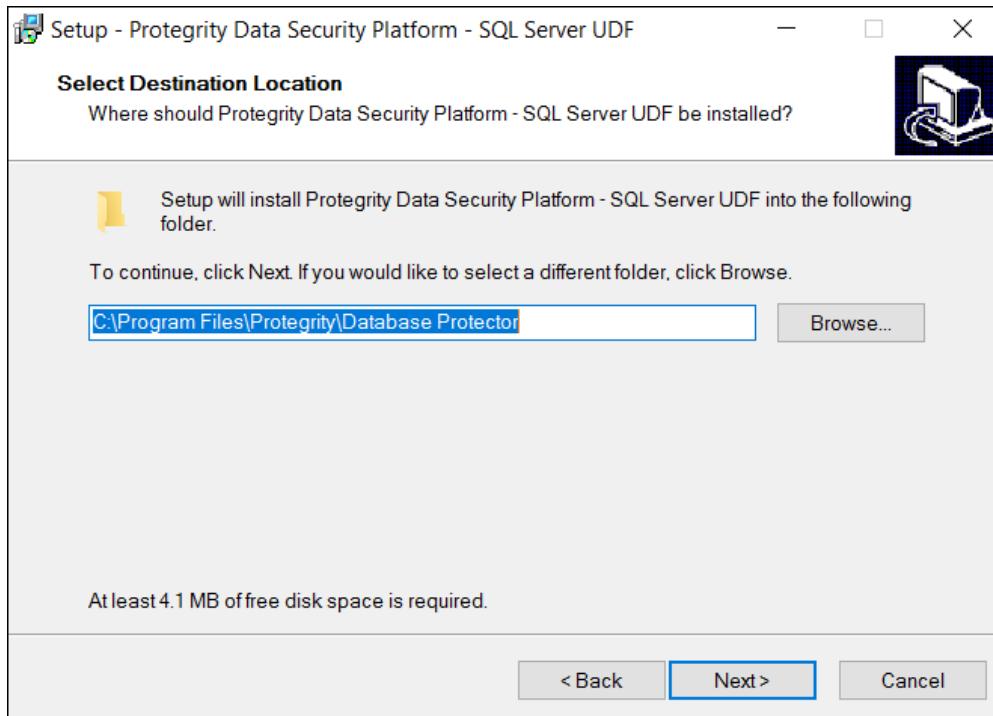


Figure 12-405: Select Destination Location screen

4. To change the default installation directory for the PEP for MS SQL Server, perform the following steps.
 - a. Click **Browse**.
 - b. Select the required installation directory where you want to install the PEP for MS SQL Server.
 - c. Click **OK**.

Note: The default directory is *C:\Program Files\Protegility\Database Protector* where the PEP for MS SQL Server is installed.

5. Click **Next**.
The **Ready to Install** screen appears.

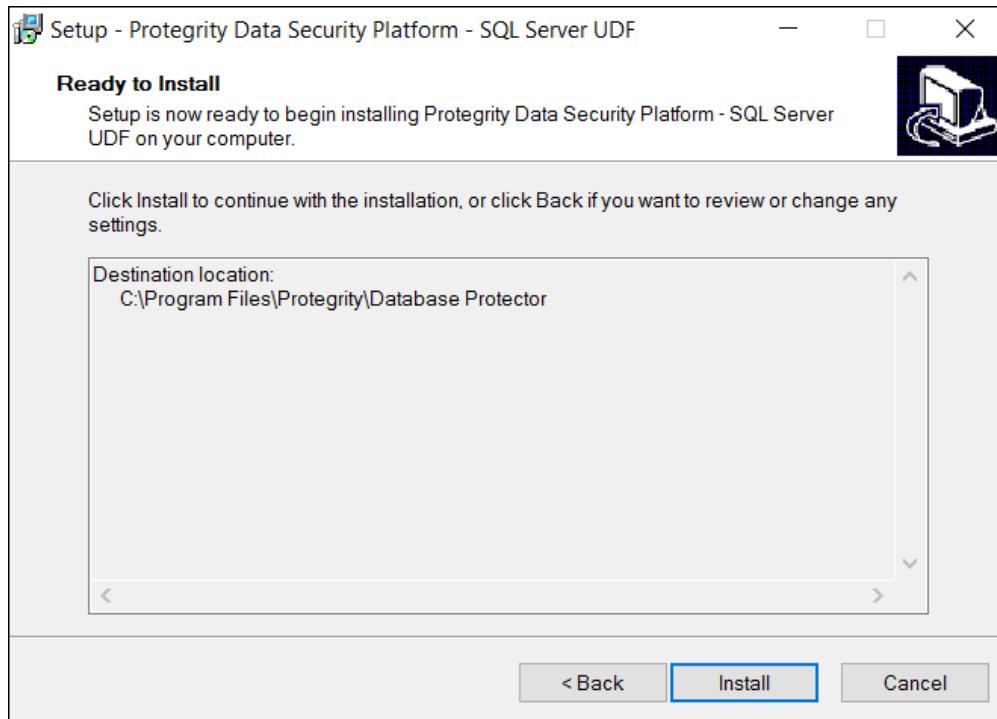


Figure 12-406: Ready to Install screen

6. Click **Install**.

The installer completes the installation of the PEP for MS SQL Server successfully in the specified location and the **Completing the Protegility Data Security Platform - SQL Server UDF Setup Wizard** screen appears.

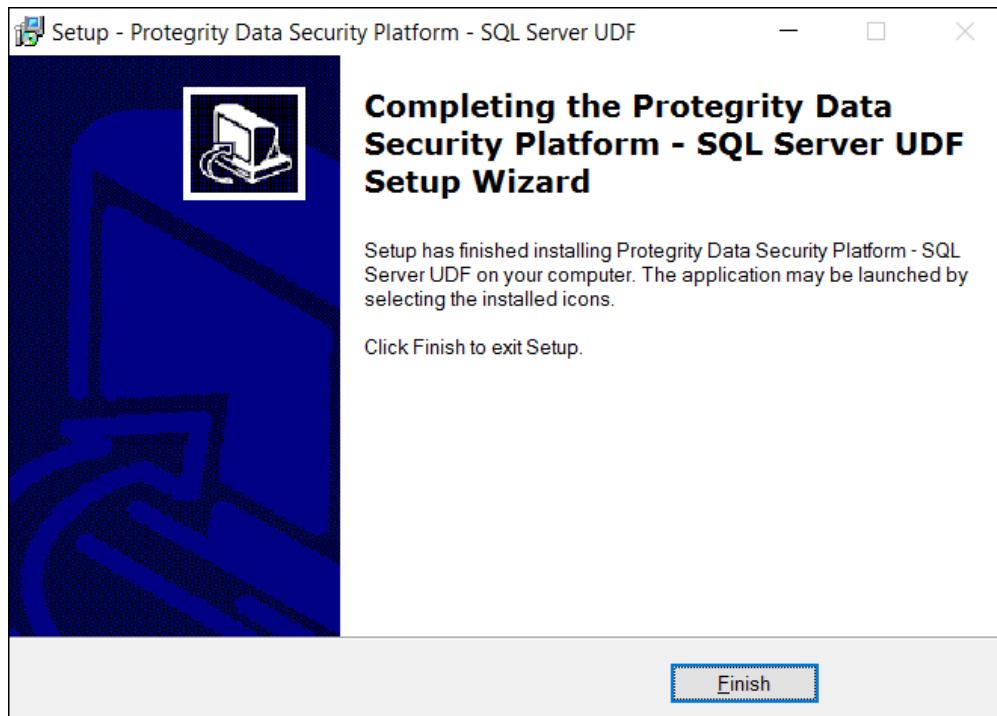


Figure 12-407: Completing the Protegility Data Security Platform - SQL Server UDF Setup Wizard

7. Click **Finish**.

The **Completing the Protegility Data Security Platform - SQL Server UDF Setup Wizard** exits.

12.7.1.5 Configuration of MS SQL Database Protector

The installation process for the MS SQL database protector automatically sets up some configuration settings. The following table describes these settings.

Table 12-99: Settings for MS SQL Server

| Setting | Description |
|------------------|---|
| Communication ID | <p>The installer sets the <i>Communication ID</i> to <i>0</i> as the default configuration settings for MS SQL Server in the following locations:</p> <ul style="list-style-type: none"> • <i>pepserver.cfg</i> file • Windows Registry - HKEY_LOCAL_MACHINE > SOFTWARE > Protegility > Defiance DPS > SQL CLR <p>Note: This parameter is no longer used and is retained for compatibility purposes only.</p> |
| Domain Name | <p>Using the LDAP member-source component, the system updates the Windows registry value in HKEY_LOCAL_MACHINE > SOFTWARE > Protegility > Defiance DPS > SQL CLR</p> <p>This helps the Administrator to create unique usernames by adding the domain name as a prefix to every username.</p> |

Note:

- A Windows authenticated user must provide the username by adding the domain or the host name as a prefix.
- A username without the domain name could lead to security vulnerabilities due to duplicate usernames.
- If you configure the SQL Server instance to perform Windows authentication, then the mixed mode authentication should be disabled.

12.7.1.6 Installation of User Defined Functions (UDFs)

This section provides information about installing the **User Defined functions (UDFs)** for the MS SQL database protector. You can install the UDFs with or without a certificate-based login. To know more about certificate-based login, refer to the section, [Understanding the Certificate-Based Login](#).

12.7.1.6.1 Understanding the Certificate-Based Login

You have to run the *CreateAssembly.sql* script before installing the UDFs for the MS SQL database protector. This section explains the configuration of *TRUSTWORTHY* property in the *CreateAssembly.sql* script.

The following figure shows the configuration of the *TRUSTWORTHY* database property for a secured connection.



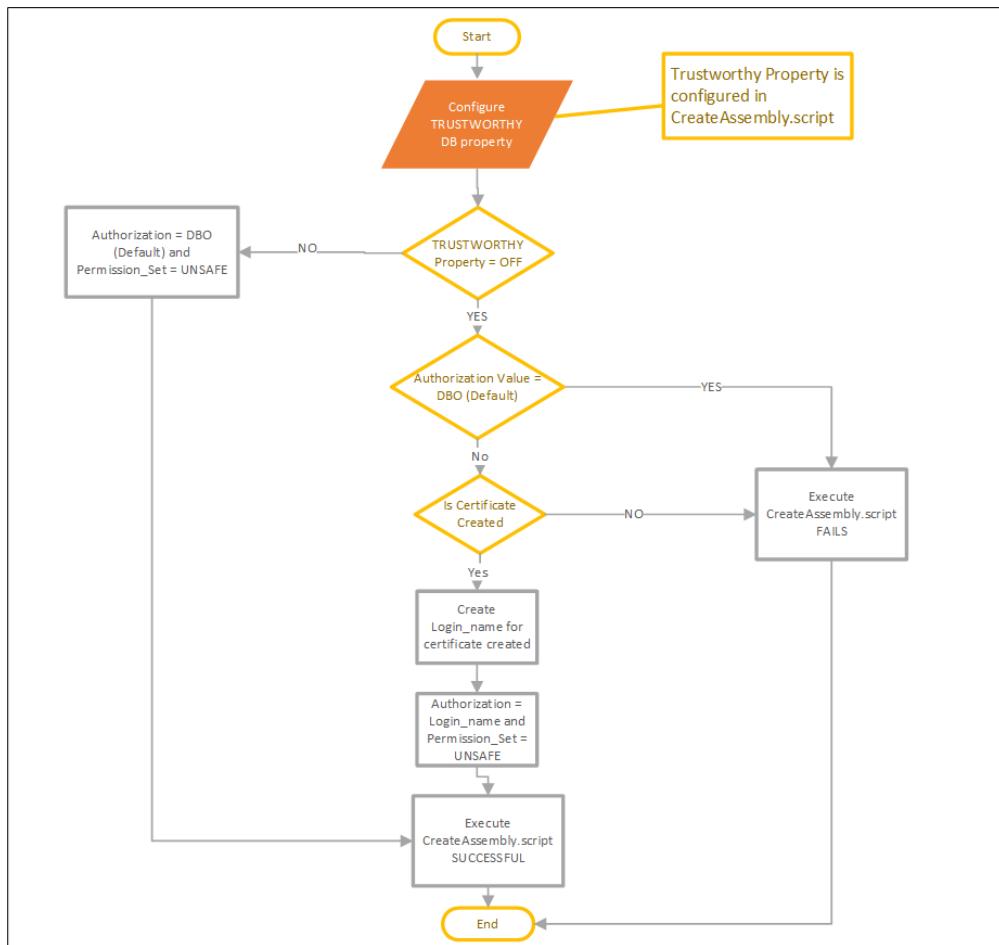


Figure 12-408: Configuration of Trustworthy Property for Certificate-Based Login

When you run the *CreateAssembly.sql* script with default value of the *TRUSTWORTHY* property set to *ON*:

- the authorization default value is set to *DBO* (Database Owner)
- the value of the *PERMISSION_SET* parameter is set to *UNSAFE*

Note: It is recommended to avoid changing the value of the *TRUSTWORTHY* property. If you set the value of the *TRUSTWORTHY* property to *ON*, then the security risk increases. To mitigate malicious threats, set the value of the *TRUSTWORTHY* property to *OFF* when the database is connected to the server.

If you run the *CreateAssembly.sql* script with the value of the *TRUSTWORTHY* property set to *OFF*, then you must create an authorized certificate using the signed *XCPepConnector.dll* file from Protegility. The digitally signed certificate validates and creates a secured connection between the database and the SQL Server.

When you run the *CreateAssembly.sql* script with the *TRUSTWORTHY* property set to *OFF*, set the following parameters for a certificate-based login:

- the authorization parameter is set to the *Login_name* who has *UNSAFE ASSEMBLY* permissions
- the value of the *PERMISSION_SET* parameter is set to *UNSAFE*

If you run the *1.0.CreateAssembly.sql* script with the value of the *TRUSTWORTHY* property set to *OFF* without any authorized certificate, then the installation fails with the following error message:

CREATE ASSEMBLY for assembly 'XCPepConnector' failed because assembly 'XCPepConnector' is not authorized for PERMISSION_SET = UNSAFE. The assembly is authorized when either of the following is true: the database owner (DBO) has UNSAFE ASSEMBLY permission and

the database has the *TRUSTWORTHY* database property on; or the assembly is signed with a certificate or an asymmetric key that has a corresponding login with *UNSAFE ASSEMBLY* permission.

Note: For more information about configuring the *TRUSTWORTHY* property and creating a certificate, refer to [TRUSTWORTHY database property](#) on Microsoft's website and create a certificate for package signing.

- If you set the *TRUSTWORTHY* database property to *OFF*, then you must create a certificate-based login before installing the UDFs. To perform the steps for creating a certificate-based login, refer to the section, [Creating Certificate-Based Login](#). Thereafter, follow the steps for installing the UDFs. To perform the steps for installing the UDFs, refer to the section, [Creating the User Defined Functions \(UDFs\)](#).
- If you retain the default value of the *TRUSTWORTHY* database property to *ON*, then you can install the UDFs without a certificate-based login. To perform the steps for installing the UDFs, refer to the section, [Creating the User Defined Functions \(UDFs\)](#).

12.7.1.6.2 Creating a Certificate-Based Login

This section describes the steps to create a certificate-based login for the MS SQL Server Database using signed *dll* from Protegility.

► To create certificate-based login:

1. Create a certificate for the database using the signed *dll* from Protegility to authenticate the identity of the server.

For example:

```
CREATE CERTIFICATE MSSQL_90009_cert
FROM EXECUTABLE FILE = '<install_dir>\Protegility\DefianceDPS\pep\XCPepConnector.dll';
GO
```

2. Create a login bound to the certificate and grant the *UNSAFE ASSEMBLY* permissions.

For example:

```
CREATE LOGIN John
FROM CERTIFICATE MSSQL_90009_cert
GO

GRANT UNSAFE ASSEMBLY TO <John>;
GO
```

3. Verify whether the certificate is created or not as shown in the highlighted location in the SQL Server Management Studio.

For example:

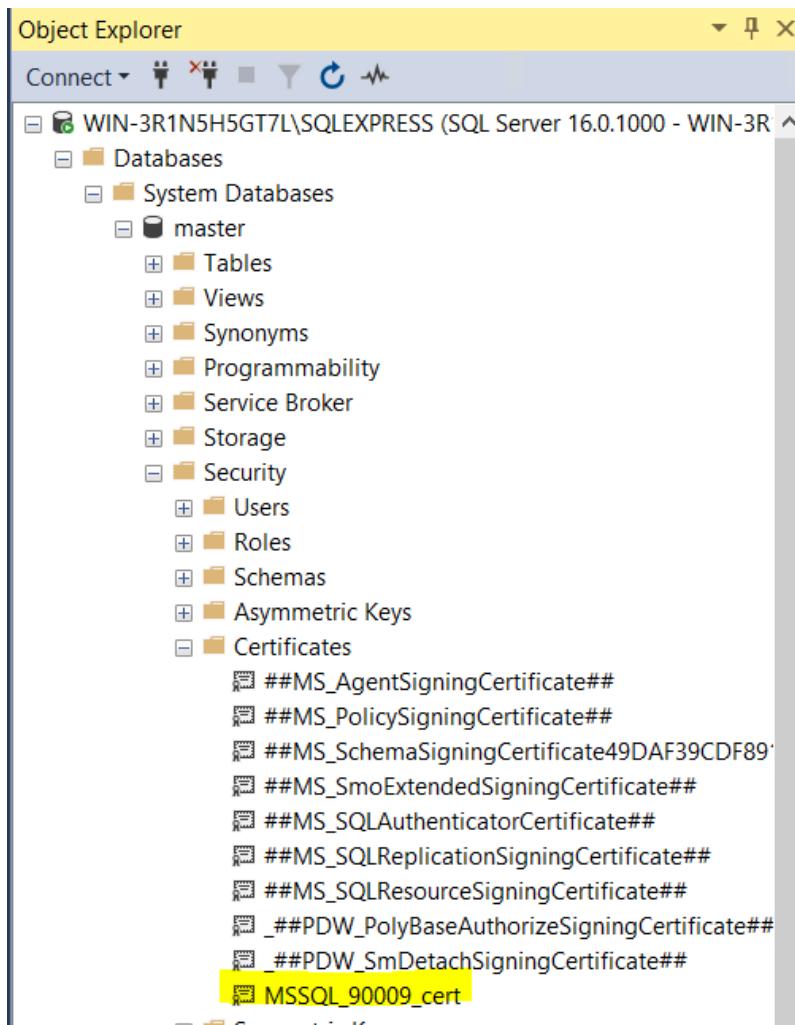


Figure 12-409: SQL Server Management Studio

4. Give access to the created user and select the type of database you need to install for the UDFs.

For example:

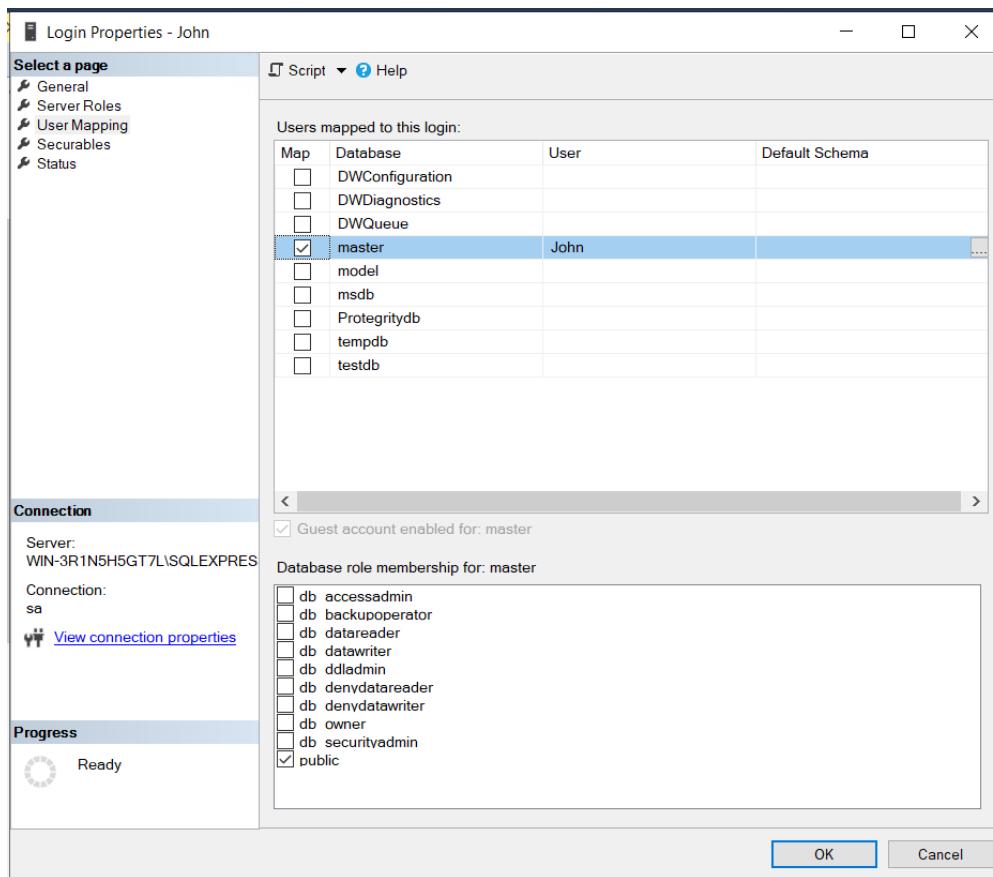


Figure 12-410: Login Properties menu

In this example, the MSSQL is installed on the *master* database for the created user *John*.

5. Edit the *create assembly* script by giving authorization name, set the trustworthy flag as *off* and with the permission as *UNSAFE*.

For example:

```
alter database [master] set trustworthy off
GO

if exists(SELECT name FROM sys.assemblies WHERE name = 'XCPepConnector')
    DROP ASSEMBLY [XCPepConnector]
GO

CREATE ASSEMBLY [XCPepConnector]
AUTHORIZATION [John]
FROM 'C:\Program Files\Protegility\Database Protector\pep\xcpепConnector.dll'
WITH PERMISSION_SET = UNSAFE
GO
```

A pop-up menu to set the assembly properties of the created user appears:

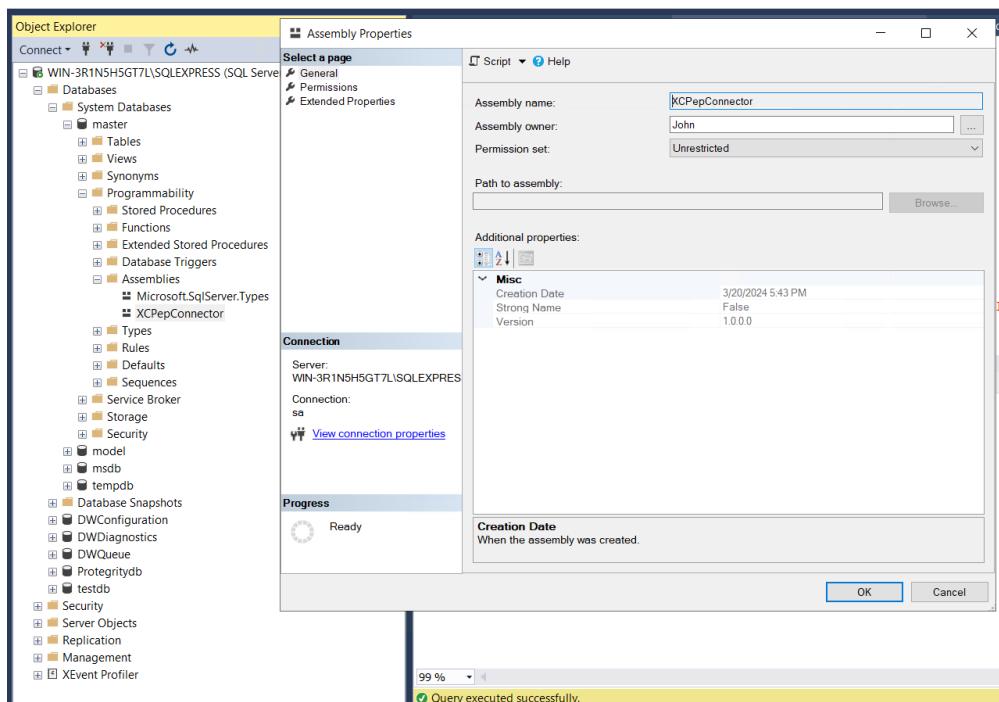


Figure 12-411: Assembly Properties menu

- Run the *create assembly* and *create objects* scripts and then run the protect/unprotect operations.

12.7.1.6.3 Creating the User Defined Functions (UDFs)

This section describes the steps to create the UDFs for the MS SQL Server database protector.

Before you begin

- If the *TRUSTWORTHY* database property is OFF, then ensure that a certificate-based login is created from the signed *dll* file provided by Protegility.
- For certificate-based login, perform the steps in the section, [Creating Certificate-Based Login](#).
- You must create a certificate-based login before the creating the UDFs.
- If the *TRUSTWORTHY* database property is ON, then perform the steps to create the UDFs without certificate-based login.

► To create the UDFs for MS SQL Server database protector:

- To connect to the database, login as the privileged user with the *CREATE ASSEMBLY* permissions.

Note: In MS SQL Server, the *sa* login is disabled by default. You must enable it to login with *sa* user for connecting to the database.

- To run the scripts, navigate to the *<install_dir>\Protegility\DatabaseProtector\pep\sqlscripts\sqlserver* directory.
- To execute the registered assembly, run the *CreateAssembly.sql* script.

```
USE [master]
GO

sp_configure 'clr enable',1
RECONFIGURE
GO
```

```
alter database [master] set trustworthy on
GO

if exists(SELECT name FROM sys.assemblies WHERE name = 'XCPepConnector')
DROP ASSEMBLY [XCPepConnector]
GO

CREATE ASSEMBLY [XCPepConnector]
AUTHORIZATION [dbo]
FROM 'C:\Program Files\Protegility\Database Protector\pep\xcpепConnector.dll'
WITH PERMISSION_SET = UNSAFE
GO
```

4. To create the standard UDFs, run the *CreateFunctions.sql* script.

12.7.1.7 Uninstalling the MS SQL Database Protector

This section describes the procedures to uninstall the MS SQL Database Protector and the PEP server.

► **To Uninstall the MS SQL DB Protector:**

1. Login to the database with a username that owns the UDF functions.
2. Run the script, *dropobjects.sql*, which is located in the *<INSTALL_DIRECTORY>\Database Protector\pep\sqlscripts\sqlserver* directory.
3. Run the uninstall utility located in the ..|*Database Protector* directory.
4. Delete the *Database Protector* directory.

12.7.1.7.1 Dropping the MS SQL Database Protector UDFs

This section explains the procedure to drop the UDFs for MS SQL database protector.

► **To Drop the UDFs for MS SQL database protector:**

1. Login to the MS SQL database with a username that owns the UDF functions.

Note: You must run the *DropObjects.sql* script. You can run this script either before or after uninstalling all the MS SQL Server database protector components.

2. To run the *DropObjects.sql* script, navigate to the *<INSTALL_DIRECTORY>\Database Protector\pep\sqlscripts\sqlserver* directory where the script is located.
3. To drop the objects, select and double-click the *DropObjects.sql* script.
The *DropObjects.sql* script drops all the objects mentioned in the script.

12.7.1.7.2 Uninstalling the PEP for MS SQL Server

This section describes the procedure to uninstall the PEP for MS SQL Server.

► **To Uninstall the PEP for MS SQL Server:**

1. To uninstall the PEP for MS SQL Server component, from the Windows menu, navigate to **Start > Control Panel > Programs > Programs and Features**.
2. From the list of programs, under the *Name* column, select **Protegility Data Security Platform - SQL Server UDF**.
3. To uninstall the PEP for MS SQL Server, click **Uninstall**.

Note: Alternatively, to uninstall the PEP for MS SQL Server, navigate to the . . \Database_Protector directory. Select *unins000* file and then double-click it.

The PEP for MS SQL Server is uninstalled and the *Database Protector* directory is deleted.

12.7.1.7.3 Uninstalling the PEP server

This section describes the procedure to uninstall the PEP server.

► **To Uninstall the PEP server:**

1. To open the Services Manager, navigate to **Start > Control Panel > System and Security > Administrative Tools > Services**.

Note: Alternatively, you can open the Services Manager, from the **Run** dialog box. From the *Windows* menu, navigate to **Start > Run**. In the **Run** window, type *services.msc* and then, click **OK**.

The **Services** window appears.

2. To select the PEP server from the list of services, select **Protegility PEP Server**.
3. To stop the PEP server service, navigate to **Action > Stop**.
4. Select **Stop**.
The PEP server service is stopped.
5. To uninstall the PEP server, perform the following steps.
 - a. From the Windows menu, navigate to **Start > Control Panel > Programs > Programs and Features**.
 - b. From the list of programs, under the *Name* column, select **Protegility PEP Server**.
 - c. Click **Uninstall**.

Note: Alternatively, navigate to the . . \defiance_dps directory. To uninstall the PEP server, select *unins000* file and then double-click it.

The PEP server is uninstalled and the *defiance_dps* directory is deleted.

12.7.1.7.4 Uninstalling the Log Forwarder

This section describes the procedure to uninstall the Log Forwarder.

► **To Uninstall the Log Forwarder:**

- To open the Services Manager, navigate to **Start > Control Panel > System and Security > Administrative Tools > Services**.

Note: Alternatively, you can open the Services Manager, from the **Run** dialog box. From the *Windows* menu, navigate to **Start > Run**. In the **Run** window, type *services.msc* and then, click **OK**.

The **Services** window appears.

- To select the Log Forwarder from the list of services, select **Logforwarder**.
- To stop the Log Forwarder service, navigate to **Action > Stop**.
- Select **Stop**.
The Logforwarder service is stopped.
- To uninstall the Log Forwarder, perform the following steps.
 - From the Windows menu, navigate to **Start > Control Panel > Programs > Programs and Features**.
 - From the list of programs, under the *Name* column, select **Logforwarder**.
 - Click **Uninstall**.

Note: Alternatively, navigate to the *.. \fluent-bit* directory. To uninstall the Log Forwarder, select *unins000* file and then double-click it.

The Log Forwarder is uninstalled.

12.7.2 Installing and Uninstalling the Oracle Database Protector

This section provides information about how to install and uninstall the Oracle Database Protector.

12.7.2.1 Installing the Oracle Database Protector

This section includes information about installing the Oracle Database Protector.

Ensure that the following order of installation is followed.

| Order of Installation | Description |
|-----------------------|--|
| 1 | Setting up the Environment Specification |
| 2 | Verifying the User Privileges |
| 3 | Configuring the Environment Variables |
| 4 | Prerequisites |
| 5 | Installing the Log Forwarder |
| 6 | Installing the PEP server |

The Oracle DB Protector can be installed by the *root* user and the Oracle *admin* user. This section discusses the installation using the *root* user. Wherever possible, the *oracle* commands for Oracle *admin* user would be provided as well.

To use the Oracle DB Protector, the environment variables need to be updated in Oracle.

For more information about the environment variables, refer to section *Configuring Environment Variables for Oracle Database* in the [Protegility Database Protector Guide 9.1.0.0](#).

12.7.2.1.1 User Privileges

The Oracle Database Protector installation can be broadly divided into installing the PEP server and installing the UDFs. The PEP server installation establishes the connection between the ESA and the Database Protector, while the UDFs use the policies to enforce protection on the data.

User for retrieving users from Oracle Database

For policies to be defined in the ESA, users can be imported from any of the multiple sources such as Active Directory (AD), file, or an Oracle database. If you want to pull users from an Oracle database, a membersource must be created. The following information applies if the users must be pulled from an Oracle database.

To retrieve users from the Member Source Server, you must either create a functional database user with create session granted or use an existing user with create session granted, and then grant the following two specific grants:

- Grant select on sys.dba_roles to *protegility*
- Grant select on sys.dba_role_privs to *protegility*

Where, *protegility* is the functional user created.

User for installing UDFs

After the PEP server is installed, the UDFs can be installed on the Oracle Database server. A functional database user with the following are the privilege rights must be created.

- Grant unlimited tablespace to USER1
- Grant create session to USER1
- Grant select any table to USER1
- Grant create library to USER1
- Grant create procedure to USER1
- Grant drop public synonym to USER1
- Grant create public synonym to USER1
- Grant create table to USER1
- Grant create view to USER1

Where, USER1 is the functional user created.

12.7.2.1.2 Prerequisites

Ensure that the following prerequisites are met:

- The Enterprise Security Administrator (ESA) appliances are installed, configured, and running.

For more information about the connection between the ESA and the Protector, refer the section [General Architecture](#).

- The IP address or host name of the ESA is noted.

- Ensure that Policy Management (PIM) has been initialized on the ESA. The initialization of PIM ensures that cryptographic keys for protecting data and the policy repository have been created.

For more information about initializing the PIM, refer to *section 4 Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.5](#).

12.7.2.1.3 Preparing the Server

This initial installation is always done on the Oracle server machine.

► To prepare the server:

1. Login to the server as the *root* user.
2. To install the Oracle DB Protector, select either your own directory or the default directory.

For example, if you select the default directory */opt*, then create */opt/protegility*, if it is not present.

Note: Grant recursive 755 permissions for the */opt/protegility* directory. If the directory does not have these permissions, then the Protegility UDFs cannot be installed and the following error is returned for the *.plm* file in the *~/defiance_dps/pep* directory.

File not found

3. Map the policies with the required Database users and groups to enable them to use the UDFs.

Before the Oracle Database UDFs are installed, the following tasks must be completed:

- Install the PEP server
- Find the status of the PEP server

12.7.2.1.4 Installing the Log Forwarder

This section describes the steps to install the Log Forwarder.

► To install the Log Forwarder:

1. Save the Oracle DB Protector package, *DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tgz*, in the */opt/protegility* directory.
2. Login to the server as the *root* user.
3. Navigate to the */opt/Protegility* directory.

- For Oracle on Linux platform, to extract the contents of the package, run the following command:

```
# tar zxf DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tgz
```

For example, if the name of the package is *DatabaseProtector_LINUX-ALL-64_x86-64_Oracle-ALL-64_9.1.0.0.xx.tgz*, then run the following command:

```
# tar zxf DatabaseProtector_LINUX-ALL-64_x86-64_Oracle-ALL-64_9.1.0.0.xx.tgz
```

Note: If you are using the Oracle RAC or the Oracle Exadata appliance, then run the command on all the Oracle Database servers.

The command extracts the following self-extracting shell executable files:

- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*
- *PepOracleSetup_Linux_x64_<version>.sh*
- *U.S.Patent.No.6,321,201.Legend.txt*

- For Oracle on AIX platform, to extract the contents of the package, run the following command:

Note: In case of AIX platform, you cannot extract the package directly using the # tar command. To convert the .tgz file to .tar file, run the following command:

```
# gunzip DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tgz
```

The command converts the *DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tgz* file to a *DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tar* file.

To extract the contents of the package, run the following command:

```
# tar -xvf DatabaseProtector_<OS>-<arch>_<Oracle distribution>-64_<version>.tar
```

For example, if the name of the package is *DatabaseProtector_AIX-ALL-64_PPC-64_Oracle-19c-64_<version>.tgz*, then run the following commands:

```
# gunzip DatabaseProtector_AIX-ALL-64_PPC-64_Oracle-19c-64_<version>.tgz
```

```
# tar xvf DatabaseProtector_AIX-ALL-64_PPC-64_Oracle-19c-64_<version>.tgz
```

The command extracts the following self-extracting shell executable files:

- PepOracle19cSetup_AIX_ppc64_<version>.sh*
- PepServerSetup_AIX_ppc64_<version>.sh*
- LogforwarderSetup_Linux_x64_<version>.sh*
- LogforwarderSetup_Windows_x64_<version>.sh*
- U.S.Patent.No.6,321,201.Legend.txt*

- To install the Log Forwarder, run the following command:

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

Caution: It is mandatory to install the Log Forwarder component before installing the PEP server component to ensure that the Oracle Database Protector is configured correctly.

Note: If you are using the Oracle RAC or the Oracle Exadata appliance, then run the command on all the Oracle Database servers.

Follow the screen prompts that appear during the installation process.

- Enter the Elasticsearch host name or the IP address.

The *Logforwarder* is installed successfully in the */opt/protegility/fluent-bit* directory.

- To start the Log Forwarder, run the following command:

```
./logforwarderctrl start
```

Note:

- For more information on how to install the Log Forwarder on the Linux machine for AIX platform, refer to the section *Installing and Configuring the Log Forwarder on the Linux machine for AIX platform* in the *Protegility Installation Guide 9.1.0.5*.
- For more information on how to install the Log Forwarder on the Windows machine for AIX platform, refer to the section *Installing and Configuring the Log Forwarder on the Windows machine for AIX platform* in the *Protegility Installation Guide 9.1.0.5*.
- If you want to change the authentication, then refer to the section *Appendix C: Configuring Security for the Log Forwarder* in the *Audit Store Guide 9.1.0.5*.

12.7.2.1.5 Installing the PEP Server

This section describes the steps to install the PEP server.

► To install the PEP server:

1. Run the PEP server installer using the following command. Follow the screen prompts that appear during the installation process.
`./PepServerSetup_Linux_x64_<version>.sh`

Caution: Ensure that the ESA is installed and running with the *HubController* service status as *Running*, to enable downloading of the certificates automatically.

Note: If you are using Oracle RAC or Oracle Exadata appliance, then run the command on all the Oracle Database servers.

2. Enter the ESA hostname or IP Address.
3. Enter the credentials for downloading the certificates.
The certificates are successfully downloaded and stored in the the `/opt/protegility/defiance_dps/data` directory and the PEP server is installed successfully in the `/opt/protegility/defiance_dps` directory.
4. Set the `EXTPROC_DLLS` parameter in the `extproc.ora` file located in the `$ORACLEHOME/hs/admin/` directory on the Oracle Database server to the following value.

```
SET EXTPROC_DLLS=ANY
```

Note: You can modify the PEP server configuration parameters as per your requirement. For more information about the usage of the parameters of the PEP Server configuration file `pepserver.cfg`, refer to section [Appendix A: PEP Server Configuration File](#).

12.7.2.1.6 Granting Permissions to the Required Files and Directories

► To grant permissions to the required files and directories:

Run the following command to ensure that the UDFs are working and access is available to all the files and directories.

```
chmod -R 755 /opt/protegility/
```

12.7.2.1.7 Finding Status of the PEP Server

► To Find the Status of the PEP Server:

1. Run the following command to find the status of the PEP server.
`/opt/protegility/defiance_dps/bin/pepsrvctrl status all`
2. If the PEP server is not running, then start the PEP server using the following command.

```
/opt/protegility/defiance_dps/bin/pepsrvctrl start
```

Note: If you are using Oracle RAC or Oracle Exadata appliance, then run the command on all the Oracle Database servers.

Note: To check the PEP server logs on the ESA, login to the ESA Web UI and navigate to **Analytics > Forensics**.

For more information about the Forensics, refer to the section *Working with Forensics* in the *Protegility Analytics Guide 9.1.0.0*.

12.7.2.1.8 UDF Installation

This section provides detailed information on the configuration of User Defined Functions (UDFs) and Extended Store (XP) procedures defined for the type of the installed database.

12.7.2.1.8.1 Prerequisites

Ensure that the following prerequisites are met before installing the UDFs:

- Verify the required User Privileges.

For more information about User Privileges, refer to section *User Privileges* in the *Protegility Database Protector Guide 9.1.0.0*.

- Configure the environment variables based on the Oracle version.

For more information about configuring the environment variables, refer to section *Configuring Environment Variables for Oracle Database* in the *Protegility Database Protector Guide 9.1.0.0*.

- The PEP server is installed.

For more information about installing the PEP server, refer to section *Installing the PEP Server*.

12.7.2.1.8.2 Oracle DB Protector UDFs

This topic describes how to create UDFs for Oracle databases from the command line interface.

12.7.2.1.8.2.1 Installing UDFs for Oracle Database Protector

This topic describes how to create UDFs for an Oracle database.

Before you begin

Before running the *createobjects.sql* script, configure the *listener.ora*, *tnsnames.ora*, and the *extproc.ora* configuration files, depending on the version of the Oracle database.

For more information about the configuration files, refer to section *Configuring Environment Variables for Oracle Database* in the *Protegility Database Protector Guide 9.1.0.0*.

Note: If the Oracle Exadata appliance is used, then install the UDFs on any one Oracle Database server only.

► To install UDFs for Oracle Database Protector:

1. Connect to the database as the *oracle* user with the database owner credentials.

Note: You can create UDFs using the *oracle* user only. The Oracle DBAs should switch to this user mode by running the following command.

```
su - oracle
```

2. Navigate to the */opt/protegility/defiance_dps/pep/sqlscripts/oracle* directory.
3. Run the following command to install the UDFs.

```
sqlplus User1/Password1 @createobjects.sql
```

Note: The *User1* and *Password1* are the credentials of the database owner.

The symbol “\” is used for Windows and “/” for UNIX environments.

12.7.2.1.8.2.2 Checking the Installation of UDFs

- To check that all UDFs have been installed:

1. Login to the database <Oracle UDFs Database>.
su - oracle
2. Run the following command to get a list of all installed Oracle DB Protector UDFs.
select PROCEDURE_NAME from user_procedures order by 1
3. To verify that the installation is successful, run either of these Oracle DB UDFs:
 - *select pty.whoami() from dual;*
Name of the user is displayed.
 - *select pty.getversion() from dual;*
Version number of the protector is displayed.
4. Logout from the user *oracle*.

12.7.2.1.8.3 Installing UDFs for Oracle XA

This topic describes additional setup steps required for the Oracle XA database.

- To install UDFs for Oracle XA:

1. Navigate to the *defiance_dps/xcoracle* directory.
2. Run the following command from the \$ prompt.
loadjava -resolve -verbose -user USER/PASSWORD xcoracle.jar com/protegility/xcoracle/XCOracle.properties
3. Navigate to the directory where the PEP server is installed. The following files are available in the directory:

| Directory | Files |
|---|---------------------------------------|
| /defiance_dps/xcoracle | <i>xcoracle.jar</i> |
| /defiance_dps/xcoracle/scripts/ | <i>createobjects.sql</i> |
| | <i>dropobjects.sql</i> |
| | <i>allowsocket.sql grant_java.sql</i> |
| /defiance_dps/xcoracle/com/protegility/xcoracle | <i>XCoracle.properties</i> |

4. Run the *allowsocket.sql* script by the *SYS* user.
5. Run the *grant_java.sql* script to allow all users to execute the Java functions by the Oracle user running XC java (for example, it can be JAVAUSR, used for the Oracle XA setup).
6. Run the *createobjects.sql* script by the PEP user (JAVAUSR in our example).
7. Ensure that one of the XC listener ports is activated (uncommented) in the *pepper.cfg* file as shown in the following code snippet.

```
# By default none of these are active.
# To activate you need to remove the comment of one or both of these rows
listener = tcp, 15910, xc
```

8. Restart the PEP server if you have made any changes to the configuration file.

12.7.2.2 Uninstalling the Oracle Database Protector

This section demonstrates the procedures to uninstall the Oracle Database Protector.

12.7.2.2.1 Uninstalling the UDFs

This section describes the procedures to uninstall the Database Protector UDFs.

► To uninstall the UDFs:

1. Login to the Oracle Database server.
2. Navigate to the *~/defiance_dps/pep/sqlscripts/oracle* directory.
3. Login to the Oracle database, with the same user that installed the UDFs, by running the following command.
sqlplus USER1/Password1 @dropobjects.sql

12.7.2.2.2 Uninstalling the PEP Server

This section describes the steps to uninstall the PEP server.

► To uninstall the PEP server:

1. Stop the PEP server by running the following command.
/opt/protegility/defiance_dps/bin/pepsrvctrl stop
2. Remove the *~/defiance_dps/bin* directory.
3. Remove the *~/defiance_dps/data* directory.
4. Remove the *~/defiance_dps/pep* directory.
5. To verify that the PEP server has been removed, check for the existence of the *~/defiance_dps/bin*, *~/defiance_dps/data*, and *~/defiance_dps/pep* directories.

The directories should not exist.

12.7.2.2.3 Uninstalling the Log Forwarder

This section describes the steps to uninstall the Log Forwarder.

► To uninstall the Log Forwarder:

1. Stop the Log Forwarder by running the following command.

```
/opt/protegility/fluent-bit/bin/logforwarderctrl stop
```

2. Navigate to the `/opt/protegility` directory.
3. Remove the `fluent-bit` directory.

12.7.3 Installing and Uninstalling the Teradata Database Protector

The following figure shows the task flow specifics for setting up the Teradata Database Protector.



Figure 12-412: Data Protector Setup Task Flow

The installation of the Teradata database protector requires some installation preparation explained in the following sections.

12.7.3.1 Prerequisites

Before setting up the Teradata Database Protector, ensure that your configuration meets the minimum requirements. The following table shows the environment requirements.

Table 12-100: Environment Requirements

| Configuration | Free Disk Space |
|--|-----------------|
| PEP Server | 50 MB per node |
| User defined functions (UDFs) and procedures | 50 MB |
| RAM | Min. 4 GB |

Starting from the version 7.2.0 release, if you are installing the ESA for the first time, then ensure that the Policy Management is initialized prior to installing the protector.

For more information about initializing the Policy Management, refer to section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.0](#).

The following are some additional requirements for each of the Teradata node:

1. Approximately 40 MB of free hard drive space should be available.

- Every node must have network connectivity, which means that it should be possible to access the node through the network using TCP/IP. If the network connection is unavailable on a node, then the Proxy module must be used to deploy policies and collect the audit records from that node.

Note: For more information about configuring the Proxy module, refer to section *Teradata Proxy Node* in the *Protegility Database Protector Guide 9.1.0.0*.

- Ensure that you have the *DBA* rights in the Teradata database.
- Ensure that you have the *root* access to the operating system.
- The Database Server must be up and running.
- The C-compiler must be installed. It is used to install the UDFs in the Teradata database.

Note: If the Teradata Parallel Upgrade Tool (PUT) is unavailable, then the Database Protector packages must be manually transferred and installed on each node.

- Configure the *pepserver.cfg* file based on your requirements.

Note:

It is recommended to set the value of the *semaphoreresources* parameter to a reasonably high value, to enable the Teradata UDFs to run in parallel. To determine the amount of resources for the Teradata platform, calculate the amount by multiplying the number of Access Module Processors (AMPs) with the number of threads per AMP.

Note:

- To determine the amount of AMPs, use the following command:

```
ps --no-header -C actmain | wc -l
```

- To determine the amount of threads per AMP, use the following command:

```
ps --no-header -C actmain -onlwp | head -1
```

- If the Teradata protector is installed on a machine, and you are planning to install a newer version of the protector, then it is advised to restart the database after you uninstall the old version of the protector using the *tpareset -f restart* command. This prevents the installer from reusing any temporary files of the older installation from the shared memory.

12.7.3.2 Installing the Teradata Database Protector

This section describes how to install the Teradata Database Protector.

► **To create a default directory:**

- Login to the Teradata server machine as a *root* user.
- Create a directory in the */opt* directory using the following command.

```
#mkdir /opt/protegility
```
- Set the access permissions for owners, groups, and users using the following command.

```
#chmod 755 /opt/protegility
```

12.7.3.2.1 Installing the Log Forwarder

This section describes the steps to install the Log Forwarder.

► To install the Log Forwarder:

1. Save the Teradata database protector package, *DatabaseProtector_<OS>-<arch>_<Teradata_distribution>-64_<version>.tgz*, in the */opt/protegility* directory.
2. Login to the server as the *root* user.
3. Navigate to the */opt/Protegility* directory.
4. To extract the contents of the installation package, run the following command:

```
# tar zxf DatabaseProtector_LINUX-ALL-64_x86-64_Teradata-ALL-64_9.1.0.0.x.tgz
```

5. Press ENTER.

The following executable files are extracted:

- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*
- *PepTeradataSetup_Linux_x64_<version>.sh*
- *PepTeradata_UDTSetup_Linux_x64_<version>.sh*
- *ProxySetup_Linux_x64_<version>.sh*

6. To install the Log Forwarder, run the following command:

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

Caution: It is mandatory to install the Log Forwarder component before installing the PEP server component to ensure that the Teradata Database Protector is configured correctly.

The prompt to enter the audit store endpoint appears.

```
Enter the audit store endpoint (host:port):
```

7. Enter the IP address of the Audit Store.

8. Press ENTER.

The prompt to enter an additional audit store appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

9. If you do not want to enter an additional audit store endpoint, then type *n*.

Note: To enter additional Audit store endpoints, type *y* for the command prompt given below:

```
Do you want to add another audit store endpoint? [y/n]:
```

Repeat the step 7 and step 8 for every additional audit store point that you want to add.

10. Press ENTER.

The installer displays the audit store endpoints to be added and the prompt to accept or abort the installation appears.

```
Type 'y' to accept or 'n' to abort installation:
```

11. To continue with the installation, type *y*.

12. Press ENTER.



The installer extracts the files and the Log Forwarder is installed in the `/opt/protegility/fluent-bit` directory.

```
Type 'y' to accept or 'n' to abort installation: y
Unpacking...
Extracting files...

Logforwarder installed in /opt/protegility/fluent-bit.
```

13. Navigate to the `/opt/protegility/fluent-bit/bin` directory.

14. To start the Log Forwarder, run the following command:

```
./logforwarderctrl start
```

Note: For more information about changing the authentication, refer to the section *Configuring Security for the Log Forwarder* in the [Audit Store Guide 9.1.0.0](#).

Note: For more information about starting the Log Forwarder while running multiple instances, refer to the section [Resolving the Port Collisions in the Log Forwarder](#).

12.7.3.2.2 Installing the PEP Server

This section describes the steps to install the PEP server.

► To install the PEP server:

1. Navigate to the `/opt/protegility/` directory.
2. To install the PEP server, run the following command.

```
./PepServerSetup_Linux_x64_<version>.sh
```

The prompt to enter the hostname or the IP address of the ESA appears.

```
Please enter ESA host name or IP address[]:
```

3. Enter the hostname or the IP address of the ESA.
4. Press ENTER.

The prompt to enter the username for downloading the certificate appears.

```
Please enter the user name for downloading certificates[]:
```

5. Enter the username to download the certificates.
6. Press ENTER.

The prompt to enter the password for downloading the certificate appears.

```
Please enter the password for downloading certificates []:
```

7. Type the password.
8. Press ENTER.

The installer extracts the files and downloads the certificates.

```
Please enter the password for downloading certificates[]:
```

```
Unpacking...
Extracting files...
```

```
Downloading certificates from X.X.X.X:8443...
```

| % Total | % Received | % Xferd | Average Speed | Time | Time | Time | Current |
|---------|------------|---------|---------------|-------|------|-------|---------|
| Dload | Upload | Total | | Spent | Left | Speed | |

```
100 30720 100 30720    0    0 58056      0 --:--:-- --:--:-- 58181

Extracting certificates...
Certificates successfully downloaded and stored in /opt/protegility/defiance_dps/data

Protegility PepServer installed in /opt/protegility/defiance_dps.
```

Note: You can modify the PEP server configuration parameters as per your requirement. For more information about the usage of the parameters of the PEP Server configuration file *pepper.cfg*, refer to the section [Appendix A: PEP Server Configuration File](#) in the *Protegility Installation Guide 9.1.0.0*.

12.7.3.2.3 Installing the PEP Package for Teradata

► To install the PEP package for Teradata:

- Run the following script.

```
./PepTeradataSetup_Linux_x64_<version>.sh
```

The execution starts and the prompt to continue appears.

```
*****
* Welcome to the Defiance DPS PEP Setup Wizard
*****
This will install the PEP on your computer.
Do you want to continue? [yes or no]
```

- To continue, type yes.
- Press ENTER.

The prompt to specify the installation directory appears.

```
Enter installation directory.
A new directory will be created in the installation directory.
[/opt/protegility]:
```

- Enter the installation directory as */opt/protegility*.
- Press ENTER.

The prompt to enter the name of the database to install the UDFs appears.

```
Enter name of database where the UDFs will be installed.
```

- Type the name of the database where you want to install the UDFs.
- Press ENTER.

The prompt to enter the varchar size appears.

```
Enter maximum size of varchar to be allocated by the UDFs.
NOTE: This is the maximum varchar size allocated by the UDFs
for latin as well as unicode character set.
Larger size will affect the performance !!!
Some applications can also have issues with larger size,
such as BTEQ, SQL Assistant.
```

- Type the maximum buffer size of varchar to be allocated by the UDFs.

Note:



The default value is 500 characters. You must modify the default value in this step, as per your requirement, for maximum character length.

9. Press ENTER.

The installer installs the PEP and a confirmation message appears.

```
[500]:
```

```
*****BUFFER LENGTH INITIALIZATION*****
UDF VARCHAR MAX INPUT BUFFER LENGTH (TOKENIZATION) : 500 Latin characters
UDF VARCHAR MAX OUTPUT BUFFER LENGTH (TOKENIZATION) : 676 Latin characters
UDF VARCHAR MAX INPUT BUFFER LENGTH (ENCRYPTION) : 500 Latin characters
UDF VARCHAR MAX OUTPUT BUFFER LENGTH (ENCRYPTION) : 538 Bytes
UDF VARCHAR_UNICODE MAX INPUT BUFFER LENGTH (TOKENIZATION) : 500 UNICODE characters
UDF VARCHAR_UNICODE MAX OUTPUT BUFFER LENGTH (TOKENIZATION) : 1356 UNICODE characters
UDF VARCHAR_UNICODE MAX INPUT BUFFER LENGTH (ENCRYPTION) : 500 UNICODE characters
UDF VARCHAR_UNICODE MAX OUTPUT BUFFER LENGTH (ENCRYPTION) : 1038 Bytes
```

```
PEP installed in /opt/protegility/defiance_dps.
```

```
Permission for /opt/protegility/defiance_dps/pep is successfully set.
```

Warning: For the data elements in the Teradata Database Protector, ensure that you do not exceed the default output buffer length for the result length parameter (*resultlen*) while executing the UDF. If you exceed the default output buffer length, then it may cause a memory overwrite issue.

12.7.3.2.4 Installing the Teradata User Defined Types (UDTs)

► To install the UDTs for Teradata:

1. Run the following script.

```
./PepTeradata_UDTSetup_Linux_x64_<version>.sh
```

The prompt to continue appears.

```
***** Welcome to the Defiance DPS PEP
Setup Wizard ***** This will install the
PEP on your computer. Do you want to continue? [yes or no]
```

2. To continue, type *yes*.

3. Press ENTER.

The prompt to enter the installation directory appears.

```
Enter installation directory.
A new directory will be created in the installation directory.
[/opt/protegility]:
```

4. Enter the installation directory */opt/protegility*.

5. Press ENTER.

The installer extracts the files and a confirmation appears.

```
Enter installation directory.
A new directory will be created in the installation directory.
[/opt/protegility]:
```

```
Unpacking...
Extracting files...
To get started with UDTs, please run /opt/protegility/defiance_dps/pep/
```



```
generate_udt_scripts.sh.  
PEP installed in /opt/protegility/defiance_dps.  
Permission for /opt/protegility/defiance_dps/pep is successfully set.
```

12.7.3.2.5 Installing the Teradata Proxy

► To install the Teradata Proxy:

1. Run the following script.

```
./ProxySetup_Linux_x64_<version>.sh
```

The command displays the syntax and prompts for parameters.

```
./ProxySetup_Linux_x64_9.1.0.0.x.sh  
Usage: ./ProxySetup_Linux_x64_9.1.0.0.x.sh -esa host (-elastic elastic_host) (-dir  
installation_directory) (-port proxy_port) (-port_es proxy_port_es)
```

2. Enter the IP address of the ESA in the command.

```
./ProxySetup_Linux_x64_9.1.0.0.x.sh -esa <ESA_IP_address>
```

3. Press ENTER.

The command extracts the files and installs the proxy.

```
Elastic host was not specified, using same as ESA host  
Unpacking...  
Extracting files...  
  
Proxy installed in /opt/protegility/proxy.
```

12.7.3.2.6 Starting the PEP Server

► To start the PEP server:

1. Navigate to the */opt/protegility/defiance_dps/bin* directory.
2. To start the PEP server, run the following command.

```
./pepsrvctrl start
```

12.7.3.2.7 Starting the Proxy

► To start the proxy:

1. Navigate to the */opt/protegility/proxy/bin* directory.

- To start the proxy, run the following command.

```
./proxyctrl start
```

Note: Ensure that the port that you are using for the proxy is not used by other services. If the port is being used by other services, then edit the port in the `/opt/protegility/proxy/conf/proxy.cfg` file.

12.7.3.3 Installing the User Defined Functions (UDFs) for Teradata

The database protector must be installed on all the nodes that are a part of the Teradata database. The database protector consists of a PEP server and the UDFs. When installing the PEP server, you must specify the maximum data size to be allocated by the UDFs. This value should not exceed 500 MB. When you calculate the data size, ensure that you factor in the space for the overheads. For example,

- For data that would be tokenized using non-length preserving tokens, add an overhead of approximately 6% to the original data size.
- For AES-encrypted data with blocks of 16 bytes, add an overhead of an additional 16 bytes to include CRC or IV.

Note: For information about installing the PEP server, refer to section [Installing the PEP Server](#).

The user who will install the UDFs must have the following privileges.

- GRANT CREATE FUNCTION ON PROTEGRITY to USER1;*
- GRANT ALTER FUNCTION ON PROTEGRITY to USER1;*

where,

- USER1* - is the user who install the UDFs
- PROTEGRITY* - is the database where the UDFs will be installed
- ROLE1* - is the group to which the *USER1* belongs

Note: Ensure that the user who installs the UDFs is part of the *ROLE1* group.

To grant privileges that a user needs to perform database administration functions, run the following query.

```
GRANT EXECUTE, SELECT, INSERT, UPDATE, DELETE, STATISTICS, DUMP, RESTORE, CHECKPOINT,
SHOW, EXECUTE PROCEDURE, ALTER PROCEDURE, EXECUTE FUNCTION, ALTER FUNCTION, ALTER EXTERNAL
PROCEDURE, CREATE OWNER PROCEDURE, CREATE TABLE, CREATE VIEW, CREATE MACRO, CREATE TRIGGER,
CREATE PROCEDURE, CREATE FUNCTION, DROP TABLE, DROP VIEW, DROP MACRO, DROP TRIGGER, DROP
PROCEDURE, DROP FUNCTION ON TESTDB TO ROLE1;
```

This section describes the UDF installation process on one Teradata node. The following UNIX commands or PUT utility can then be used to distribute the installation to all the nodes:

- psh mkdir /opt/protegility/*
- pcl -send /opt/protegility/* /opt/protegility/*

12.7.3.3.1 Creating the User Defined Functions (UDFs) for Teradata

Before you begin

- Ensure that you install the PEP package on all the nodes.
- Ensure that you create the UDFs under the *LATIN* server character set. If the default user session character set is not *LATIN*, then set it to *LATIN*.



► To create the UDFs for Teradata:

1. Start *bteq*.
2. Login to the database.

Note: Ensure that the user account has the required permissions to create the UDFs in the user-specific database.

3. To create the UDFs, execute the following script located in the */opt/protegility/defiance_dps/pep/sqlscripts/teradata* directory.

```
createobjects.sql
```

Note:

If you use the Format Preserving Encryption (FPE) with Teradata UDFs, then you can extend the maximum data length size provided by these UDFs, which is up to 47407 bytes by default. In this case, you can modify the maximum data length size allocated for the UDFs in the *createobjects.sql* file for the following functions:

- *PTY_VARCHARLATININS*
- *PTY_VARCHARLATINSEL*
- *PTY_VARCHARLATINSELEX*

The *REPLACE_UDFVARCHARTOKENMAX* parameter value for these functions can be set up to 64000. Teradata supports the maximum row size length of approximately 64000 bytes.

12.7.3.4 Uninstalling the Teradata Database Protector

This section demonstrates the procedures to uninstall the Teradata database protector.

12.7.3.4.1 Uninstalling the UDFs

This section describes the procedure to remove the UDFs for the Teradata database protector.

► To uninstall the UDFs:

1. Start *bteq*.
2. Login to the Teradata database server.

Note: Ensure that the user account has the required permissions to drop the UDFs in the user-specific database.

3. Navigate to the */opt/protegility/defiance_dps/pep/sqlscripts/teradata* directory.
4. Run the following script.

```
dropobjects.sql
```

12.7.3.4.2 Uninstalling the Log Forwarder

This section describes the steps to uninstall the Log Forwarder.

► To uninstall the Log Forwarder:

1. To stop the log forwarder, run the following command.

```
/opt/protegity/fluent-bit/bin/logforwarderctrl stop
```

2. Navigate to the `/opt/protegity` directory.
3. Remove the `fluent-bit` directory.

12.7.3.4.3 Uninstalling the PEP Server

- To uninstall the PEP server:

1. Navigate to the `/opt/protegity/defiance_dps/bin` directory.
2. To stop the PEP server, run the following command.

```
./pepsrvctrl stop
```

3. Delete the installation directory.

12.7.4 Installing and Uninstalling the Greenplum Database Protector

This section describes the installation and uninstallation process of the Greenplum Database protector.

12.7.4.1 Setup Overview for UNIX

Before setting up the Database Protector, ensure that your configuration meets the minimum requirements. The following table describes the environment requirements.

Table 12-101: Environment Requirements

| Configuration | Free Disk Space on UNIX |
|--|-------------------------|
| PEP server | 50 MB per node |
| User Defined Functions (UDFs) and Procedures | 10 MB |
| RAM | 4 GB |

Note: For systems under heavy load with auditing turned on, the space required for the PEP server (20 MB per node) can be higher.

These are the main tasks for installing the Protegity Greenplum Database Protector.

- Check the prerequisites for installing the Greenplum Database Protector are met
- Download and extract the Greenplum Database Protector package
- Install the PEP server in the database
- Install the Greenplum Database Protector
- Start the PEP server
- To verify the installation, perform the following tasks:
 - Create and deploy policy
 - Protect a table
 - Review Audit logs for successful protect operation.

The Greenplum Database Protector secures access to sensitive data on the GPDB or Greenplum databases.

12.7.4.2 Prerequisites

Ensure the following prerequisites are met:

- The Greenplum Database is installed and configured.

For more information, refer to [Greenplum® Database 4.0 Installation Guide](#) for assistance.

- Starting from the version 7.2.0 release, if you are installing the ESA for the first time, ensure that the Policy Management is initialized prior to installing the protector.

For more information about initializing the Policy Management, refer to section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.0](#).

- Download and save the Greenplum Database Protector, *DatabaseProtector_<OS>-<arch>_<Greenplum distribution>-64_x.x.x.x.tgz*, made available by Protegility.
- Even if it is not mandatory, take a backup of the databases where the Greenplum Database Protector and UDFs would be installed.
- Access as the *root* user, should be available to you.
- Access as the *gpadmin* superuser for Greenplum database, should be available to you.

12.7.4.3 Installing the Greenplum Database Protector on a Single Node

The Greenplum Database Protector can be installed by the *root* user and Greenplum *gpadmin* user. This section discusses the installation using the *root* user. Wherever possible, the *gpadmin* commands would be provided as well.

12.7.4.3.1 Creating the Directory to Install the Greenplum Database Protector

The */opt/protegility* directory is the default directory to install the Greenplum Database Protector. If the default directory is not present, then create the directory and grant *755* permissions.

12.7.4.3.2 Installing the PEP Server

 To Install the PEP Server:

- Upload the Greenplum Database Protector, *DatabaseProtector_<OS>-<arch>_<Greenplum distribution>-64_x.x.x.x.tgz*, to a temporary directory on the Greenplum server machine.
- Login to Greenplum server machine as the *root* user.
- Navigate to the temporary directory, created in Step 1, and unpack the installation package.

`tar xvzf DatabaseProtector_RHEL-6-64_x86-64_Greenplum-4.3-64_x.x.x.x.tgz`

The following files are extracted:

- PepServerSetup_<OS>-x64_x.x.x.x.sh*
- PepGreenplum<x.x>Setup_Linux_x64_x.x.x.x.sh*
- U.S.Patent.No.6,321,201.Legend.txt*

- Navigate to the */opt/protegility* directory.
- Run the following PEP server installer first in this location. Follow the screen prompts that appear during the installation process.

`./PepServerSetup_Linux_x64_x.x.x.x.sh`

- Enter the ESA IP address, as requested.

7. Without changing the location, that is `/opt/protegility` directory, run the Greenplum Database Protector installer. Follow the screen prompts that display during the installation process.

`./PepGreenplum<x.x>Setup_Linux_x64_x.x.x.x.sh`

12.7.4.3.3 Granting Permissions to the Required Files and Directories

- To Grant Permissions to the Required Files and Directories:

Run the following command to grant access to required files and directories.

`chmod -R 755 /opt/protegility/`

12.7.4.3.4 Finding Status of the PEP Server

Before the Greenplum Database UDFs are installed, it is recommended to check the status of the PEP server.

- To Find Status of PEP Server:

1. Run the following command to find the status of the PEP server.

`bash /opt/protegility/defiance_dps/bin/pepsrvctrl status`

2. If the PEP server is not running, then start the PEP server using the following command.

`bash /opt/protegility/defiance_dps/bin/pepsrvctrl start`

12.7.4.4 Installing and Uninstalling Greenplum Database Protector on Multiple Nodes

The Greenplum Database Protector can be installed both by `root` user and Greenplum `gpadmin` user. This section discusses the installation using the `root` user. Wherever possible, the `gpadmin` commands would be provided as well.

12.7.4.4.1 Creating the Directory to Install the Greenplum Database Protector on all the Nodes

The `/opt/protegility` directory is the default directory to install the Greenplum Database Protector. If the default directory is not present, then create the directory and grant 755 permissions.

Note: If you select the default `/opt/protegility` directory, then create the `/opt/protegility` directory on all the nodes.

12.7.4.4.2 Installing the PEP Server

- To Install the PEP Server:

1. Copy the Greenplum Database Protector, `DatabaseProtector_<OS>-<arch>_<Greenplum distribution>-64_x.x.x.x.tgz`, to a temporary directory on the master.
2. Login to the master node as the `root` user.
3. Navigate to the temporary directory, created in Step 1, and extract the installation package.

`tar xvzf DatabaseProtector_RHEL-6-64_x86-64_Greenplum-4.3-64_x.x.x.x.tgz`

The following files are extracted:

- `PepServerSetup_<OS>-x64_x.x.x.x.sh`



- *PepGreenplum<x.x>Setup_Linux_x64_x.x.x.x.sh*
 - *U.S.Patent.No.6,321,201.Legend.txt*
4. Navigate to the */opt/protegility* directory.
 5. Run the following PEP server installer first in this location. Follow the screen prompts that appear during the installation process.
`./PepServerSetup_Linux_x64_x.x.x.x.sh`
 6. Enter the ESA IP address, as requested.
 7. Without changing the location, that is */opt/protegility* directory, run the Greenplum Database Protector installer. Follow the screen prompts that display during the installation process.
`./PepGreenplum<x.x>Setup_Linux_x64_x.x.x.x.sh`

12.7.4.4.3 Granting Permissions to the Required Files and Directories on all the Nodes

► To Grant Permissions to the Required Files and Directories on all the Nodes:

Run the following command to ensure that the UDFs are working on all the nodes and access is available to all the files and directories.

`chmod -R 755 /opt/protegility/`

12.7.4.4.4 Copying All Files and Directories from the Source Directory to the Destination Directory

► To Copy all the Files and Directories from the Source Directory to the Destination Directory:

Run the following command to copy all the files and directories from the source directory to the destination directory.

`bash -r -h sdw1 -h sdw2 -h sdw3 -h sdw4 -h sdw5 -h sdw6 /opt/protegility/* =:/opt/protegility/`

Here source is the */opt/protegility* directory and destination comprises of all the segment nodes, for example, *sdw1* through *sdw6*.

12.7.4.4.5 Finding Status of the PEP Server on all Nodes

Before you begin

Before the Greenplum Database Protector UDFs are installed, it is recommended to check the status of the PEP server.

► To Find Status of the PEP Server on all the Nodes:

1. Run the following command to login to the *gpadmin* user.
`su - gpadmin`
2. Run the following command to find the status of the PEP server on all the nodes.
`bash /opt/protegility/defiance_dps/bin/pepsrvctrl status`
3. If the PEP server in any node is not running, then start the PEP server using the following command.
`bash /opt/protegility/defiance_dps/bin/pepsrvctrl start`

12.7.4.5 Installing UDFs of Greenplum Database Protector

This section describes how to create UDFs for Greenplum Database Protector.

► To install UDFs for Greenplum Database Protector:

1. Connect to the database as *gpadmin* user with database owner credentials.

Note: You can create UDFs using the *gpadmin* user only. The Greenplum DBAs should switch to this user mode by running the **su - gpadmin** command.

2. Navigate to the */opt/protegility/defiance_dps/pep/sqlscripts/postgres* directory.
3. Run the following command to install the UDFs.

```
psql -d <Greenplum UDFs Database> -f createobjects.sql
```

Note: The command line interface can also be used to create the Greenplum Database Protector UDFs.

```
psql -h mdw -d biadpdev -a -f /opt/protegility/defiance_dps/pep/sqlscripts/postgres/createobjects.sql -U gpadmin
```

| Name | Type |
|--|--|
| <i>mdw</i> | Master in Greenplum Database |
| <i>biadpdev</i> | Database |
| <i>gpadmin</i> | Username |
| <i>/opt/protegility/defiance_dps/pep/sqlscripts/postgres/createobjects.sql</i> | Script used to create the UDFs in the database |

12.7.4.5.1 Verifying the Installation of UDFs

► To check that all UDFs have been installed:

1. Login to the database <Greenplum UDFs Database>.

```
psql -d <Greenplum UDFs Database>
```

2. Run the following command to get a list of all the installed Greenplum DB Protector UDFs.

```
<Greenplum UDFs Database> = # \df pty*
```

3. Run either of the following Greenplum DB UDFs to verify that the installation is successful.

- **select pty_whoami();**

Name of the user is displayed.

- **select pty_getversion();**

Version number from the PEP server file is displayed.

4. Log out from *gpadmin*.

12.7.4.6 Uninstalling the Greenplum Database Protector

To uninstall the Greenplum Database Protector, the UDFs and the PEP server need to be uninstalled or removed from the main node. You should be logged in as the *root* user.

12.7.4.6.1 Uninstalling the UDFs

► To Uninstall the UDFs:

1. Login to the Greenplum file system.

```
# su - gpadmin
```

2. Navigate to the */opt/protegility/defiance_dps/pep/sqlscripts/postgres* directory.
3. Run the following query to uninstall the UDFs.

```
# psql -d <database name> -f dropobjects.sql
```

4. Delete the *~/defiance_dps/pep* directory.

12.7.4.6.2 Uninstalling the PEP Server

► To uninstall the PEP Server:

1. Navigate to the *~/defiance_dps/bin* directory.
2. Run the following command to stop the PEP server on all the segments.

```
./pepsrvctrl stop all
```

3. Delete the *<install_directory>/defiance_dps* directory.

12.7.5 Installing and Uninstalling the DB2 Database Protector

This section provides information about setting up, installing, and uninstalling the DB2 Database Protector.

The following table provides the order of installation for the DB2 Database Protector.

Table 12-102: DB2 Database Protector - Order of Installation

| Order of installation | Description | Reference |
|-----------------------|---|---|
| 1. | Install the ESA | Installing ESA Appliance |
| 2. | Verify the Prerequisites | Verifying the Prerequisites for Installing the DB2 Database Protector |
| 3. | Installation Preparation | Creating the Directory to Install the DB2 Database Protector |
| 4. | Install the PEP server and the DB2 Database Protector | Installing the PEP server and the DB2 Database Protector |

12.7.5.1 Verifying the Prerequisites for Installing the DB2 Database Protector

Ensure that the following prerequisites are met:

- The DB2 Database is installed and configured.
- Download and save the DB2 Database Protector, *DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz*, made available by Protegility.



- Even if it is not mandatory, take a backup of the databases where the DB2 Database Protector and UDFs would be installed.
- Access to the server as the *root* user, should be available to you.
- Access to the DB2 database as the *database* user, should be available to you.

12.7.5.2 Creating the Directory to Install the DB2 Database Protector

The DB2 database protector is installed either in the default directory or you can create your own directory. The default directory is */opt/protegility* directory. If the default directory is not present, then create the directory.

► To create the installation directory:

1. To create the directory, on the command prompt, run the following command:

```
mkdir -p /opt/protegility/
```

2. To grant the recursive 755 permissions, run the following command:

```
# chmod 755 /opt/protegility/
```

Note:

- To ensure that all the UDFs are working, the recursive 755 permissions should be granted to the */opt/protegility* directory.
- If the directory does not have the recursive permissions, then the Protegility UDFs cannot be installed.

12.7.5.3 Installing the Log Forwarder

This section describes the steps to install the Log Forwarder on the Linux platform using the DB2 Linux build and DB2 AIX build.

12.7.5.3.1 Installing the Log Forwarder using the DB2 Linux Build

This section describes the steps to install the Log Forwarder using the DB2 Linux build.

1. Save the DB2 Database Protector package, *DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz*, in the */opt/protegility* directory.
2. Login to the server as the *root* user.
3. Navigate to the */opt/Protegility* directory.
4. To extract the contents of the package, run the following command:

```
# tar zxf DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz
```

For example, if the name of the package is *DatabaseProtector_RHEL-7-64_x86-64_DB2-11-64_9.1.0.0.x.tgz*, then run the following command:

```
# tar zxf DatabaseProtector_RHEL-7-64_x86-64_DB2-11-64_9.1.0.0.x.tgz
```

The command extracts the following executable files:

- *LogforwarderSetup_Linux_x64_<version>.sh*
- *PepServerSetup_Linux_x64_<version>.sh*
- *PepDB2Setup_Linux_x64_<version>.sh*
- *U.S.Patent.No.6,321,201.Legend.txt*



- To run the Log Forwarder installer, in the `/opt/Protegility` directory, run the following script:

```
./LogforwarderSetup_Linux_x64_<version>.sh
```

Caution: It is mandatory to install the Log Forwarder component before installing the PEP server component to ensure that the DB2 Database Protector is configured correctly.

The prompt to enter the audit store endpoint (`host:port`) appears:

```
Enter the audit store endpoint (host:port)
```

- Enter the IP address of the Audit Store.

Warning: If you fail to specify an IP address, then the installation script will terminate the installation process.

- Press ENTER.

The installer script appends the port number to the IP address.

Note: The default value for the port is `9200`.

- Press ENTER.

The prompt to enter an additional audit store endpoint appears.

```
Do you want to add another audit store endpoint? [y/n]:
```

- If you do not want to enter an additional audit store endpoint, then type `n`.

Note: To enter additional Audit store endpoints, type `y` for the command prompt given below:

```
Do you want to add another audit store endpoint?
```

Repeat step 7 and step 8 for every additional audit store point that you want to add.

- Press ENTER.

The list of the audit store endpoints that will be added is displayed and the prompt to continue the installation appears.

```
Type 'y' to accept or 'n' to abort installation:
```

- To continue with the installation, type `y`.

- Press ENTER.

The installer extracts the files and a confirmation message that the `Logforwarder` is installed in the `/opt/protegility/fluent-bit` directory appears.

```
Type 'y' to accept or 'n' to abort installation: y  
Unpacking...  
Extracting files...
```

```
Logforwarder installed in /opt/protegility/fluent-bit.
```

- Navigate to the `/opt/protegility/fluent-bit` directory.

- To start the Log Forwarder, run the following command:

```
./logforwarderctrl start
```

Note: If you want to change the authentication, then refer to the section [Appendix C: Configuring Security for the Log Forwarder](#) in the [Audit Store Guide 9.1.0.0](#).

The Log Forwarder starts successfully and the following message appears on the prompt:

```
Fluent Bit v1.6.10-1.0.0+4.g2f46.master
* Copyright (C) 2019-2020 The Fluent Bit Authors
* Copyright (C) 2015-2018 Treasure Data
* Fluent Bit is a CNCF sub-project under the umbrella of Fluentd
* https://fluentbit.io
```

12.7.5.3.2 Installing the Log Forwarder using the DB2 AIX Build

This section describes the steps to install the Log Forwarder using the DB2 AIX build. Fluent-Bit is incompatible with AIX. For more information on the supported platforms, refer <https://docs.fluentbit.io/manual/installation/supported-platforms>

1. Save the DB2 DB Protector package, *DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz*, in the */opt/protegrity* directory.
2. Login to the server as the *root* user.
3. Navigate to the */opt/Protegrity* directory.
4. To extract the *.tar* file from the *.tgz* build, run the following command:

```
# gunzip DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz
```

The command extracts the *DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tar* file from the *DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tgz* file

5. To extract the contents of the package, run the following command:

```
# tar -xvf DatabaseProtector_<OS>-<arch>_<DB2 distribution>-64_<version>.tar
```

For example, if the name of the package is *DatabaseProtector_AIX-ALL-64_PPC-64_DB2-11-64_<version>.tgz*, then run the following commands:

```
# tar -xvf DatabaseProtector_AIX-ALL-64_PPC-64_DB2-11-64_<version>.tar
```

The command extracts the following self-extracting shell executable files:

- *LogforwarderSetup_Linux_x64_<version>.sh*
- *LogforwarderSetup_Windows_x64_<version>.exe*
- *PepOracle19cSetup_AIX_ppc64_<version>.sh*
- *PepServerSetup_AIX_ppc64_<version>.sh*
- *U.S.Patent.No.6,321,201.Legend.txt*

Note:

To install the Log Forwarder either on a Linux machine or Windows machine, use the respective script provided in the build package for installation.

- For more information on how to install the Log Forwarder on the Linux machine for AIX platform, refer to the section *Installing the Log Forwarder on Linux* in the *Protegrity Installation Guide 9.1.0.5*.
- For more information on how to configure the Log Forwarder on the Linux machine for AIX platform, refer to the section *Configuring the Log Forwarder on the Linux machine for the AIX platform* in the *Protegrity Installation Guide 9.1.0.5*.
- For more information on how to install the Log Forwarder on the Windows machine for AIX platform, refer to the section *Installing the Log Forwarder on Windows* in the *Protegrity Installation Guide 9.1.0.5*.



- For more information on how to configure the Log Forwarder on the Windows machine for AIX platform, refer to the section *Configuring the Log Forwarder on the Windows machine for the AIX platform* in the [Protegity Installation Guide 9.1.0.5](#).

12.7.5.4 Installing the PEP Server

This section describes the steps to install the PEP server.

► To install the PEP server:

1. Login to the server as the *root* user.
2. Copy the DB2 Database Protector package, *DatabaseProtector_<OS>-<arch>-<DB2 distribution>-64_<version>.tgz*, to the */opt/protegity* directory.
3. Navigate to the */opt/Protegity* directory.
4. To run the PEP server installer, in the */opt/Protegity* directory, run the following command:
./PepServerSetup_Linux_x64_<version>.sh

Caution: To enable automatic downloading of the certificates, ensure that the ESA is installed and is running. The *HubController* service state should be in *Running* state.

The prompt to enter the ESA *host name* or *IP address* appears.

```
Please enter ESA host name or IP address []:
```

5. Enter the ESA IP address.
6. Press ENTER.

The prompt to enter the username appears.

```
Please enter the user name for downloading certificates[]:
```

7. Enter the username.
8. Press ENTER.

The prompt to enter the password appears.

```
Please enter the password for downloading certificates[]:
```

9. Enter the password.
10. Press ENTER.

The certificates are downloaded successfully in the */opt/protegity/ defiance_dps/data* directory. A confirmation message that the Protegity PEP server is installed in the */opt/protegity/defiance_dps* directory appears.

```
Unpacking...
Extracting files...
Downloading certificates from xx.xx.x.xxx:8443...
  % Total    % Received % Xferd  Average Speed   Time     Time      Current
               Dload  Upload Total   Spent    Left  Speed
 100 20480  100 20480    0      0  39203      0 --:--:-- --:--:-- --:--:-- 39158
Extracting certificates...
```

```

tar: CA.pem: time stamp 2023-01-31 03:01:01 is 0.992341247 s in the future
tar: cert.pem: time stamp 2023-01-31 03:01:01 is 0.992075904 s in the future
tar: cert.key: time stamp 2023-01-31 03:01:01 is 0.99197781 s in the future
tar: certkeyup.bin: time stamp 2023-01-31 03:01:01 is 0.991858092 s in the future
Certificates successfully downloaded and stored in /opt/protegility/defiance_dps/data

Protegility PepServer installed in /opt/protegility/defiance_dps.

```

11. Navigate to the `/opt/protegility/defiance_dps` directory.

12. To start the PEP server, run the following command:

`./pepsrvctrl start`

The PEP server starts successfully and the following message appears on the prompt:

```

Protegility PEP Server
Version 1.2.0+16.g0a3b66.1.2
Copyright (c) 2004-2022 Protegility Corporation. All Rights Reserved.

```

12.7.5.5 Verifying the Status of the PEP Server

This section describes the steps to check the status of the PEP server.

► To verify the status of the PEP server:

1. To verify the status of the PEP server, run the following command:

`# sh /opt/protegility/defiance_dps/bin/pepsrvctrl status`

2. If the PEP server is not running, then start the PEP server using the following command:

`# sh /opt/protegility/defiance_dps/bin/pepsrvctrl start`

12.7.5.6 Installing the PEP for DB2

This section describes the steps to install the PEP for DB2 database.

► To install PEP for DB2:

1. To install the PEP for DB2, run the following command:

`./PepDB2Setup_Linux_x64_<version>.sh`

The prompt for PEP Setup Wizard appears.

```

*****
* Welcome to the Defiance DPS PEP Setup Wizard
*****
This will install the PEP on your computer.

```

2. To continue with the installation, type *yes*.

The prompt to enter the installation directory appears.

```

Do you want to continue? [yes or no]
yes
Enter installation directory.
A new directory will be created in the installation directory.
[/opt/protegility]:

```

- Enter the location where you want to install the PEP for DB2.

Note:

The default installation path is `/opt/protegility/`. To install the DB2 protector in the default installation path, press ENTER without entering the installation directory.

- Press ENTER.

The installer extracts the files and a confirmation message that the *PEP* is installed in the `/opt/protegility/defiance_dps` directory appears.

```
Unpacking...
Extracting files...

PEP installed in /opt/protegility/defiance_dps.
```

12.7.5.7 Installing the UDFs

This section describes how to create UDFs for a DB2 database.

► **To install UDFs:**

- Navigate to the `/opt/protegility/defiance_dps/pep/sqlscripts/db2` directory.
- To edit the *createobjects.sql* script, run the following command:

`vim createobjects.sql`

Note: In the *createobjects.sql* script, set the value of *REPLACE_DATABASE* parameter with the database name where you want to install the UDFs.

The *createobjects.sql* script opens in edit mode.

For example, to connect to *testdb* database, the value of *REPLACE_DATABASE* parameter is set to *testdb* in the *createobjects.sql* script as shown in the following snippet:

```
-- Create Protegility user-defined functions (UDFs) for DB2
-- Usage: db2 -tfcreateobjects.sql

-- Connect to database where to define UDFs

CONNECT TO TESTDB;
```

Figure 12-413: Connecting to DB2 Database

- Login to *DB2* using the database user with the database owner credentials.

Note: You can create the UDFs using the *DB2* user only. For example, to switch to the user mode, run the following command:

`su - db2inst1`

4. Navigate to the `/opt/protegility/defiance_dps/pep/sqlscripts/db2` directory.
5. To run the `createobjects.sql` script, run the following command:
`# db2 -tf createobjects.sql`

Note: For more information about the DB2 Database Protector UDFs, refer to section [4.1 DB2 Open Systems User Defined Functions in the APIs, UDFs, and Commands Reference Guide 9.1.0.0](#).

The confirmation message for the successful connection to the database appears and the UDFs are installs successfully in the DB2 database.

Database Connection Information

```
Database server      = DB2/LINUXX8664 11.5.0.0
SQL authorization ID = DB2INST1
Local database alias = TESTDB
```

12.7.5.8 Verifying the Installation of UDFs

This section describes the steps to verify whether all the UDFs have been installed.

► To verify the installation of UDFs:

To verify whether the installation of UDFs is successful, run anyone of the following DB2 database protector UDFs:

- `SELECT pty.whoami() FROM SYSIBM.SYSDUMMY1;`

The name of the user appears.

- `SELECT pty.getversion() FROM SYSIBM.SYSDUMMY1;`

The version number of the installed DB2 protector appears.

12.7.5.9 Uninstalling the DB2 Database Protector

This section describes the procedures to uninstall the components of DB2 Database Protector. To uninstall the DB2 database protector, uninstall the following components:

- The DB2 database Protector UDFs
- The Log Forwarder
- The PEP server
- The PEP for DB2

Note: You should be logged in as the `root` user.

12.7.5.9.1 Uninstalling the UDFs

This section the steps to uninstall the UDFs for the DB2 database protector.

► To Uninstall the UDFs:

1. Navigate to the `/opt/protegility/defiance_dps/pep/sqlscripts/db2` directory.
2. To edit the `dropobjects.sql` script, run the following command:

`vim dropobjects.sql`

Note: In the `dropobjects.sql` script, set the value of `REPLACE_DATABASE` parameter with the database name where the UDFs are installed.

The `dropobjects.sql` script opens in edit mode.

For example, to connect to `testdb` database, the value of `REPLACE_DATABASE` parameter is set to `testdb` in the `dropobjects.sql` script as shown in the following snippet:

```
--*****  
--  
-- usage: db2 -tfdrpfun.sql  
--  
-- drops user-defined functions from database  
--  
--*****  
  
CONNECT TO testdb;
```

Figure 12-414: Connecting to DB2 Database

3. Login to `DB2` using the database user with the database owner credentials.

Note: You can drop the UDFs using the `DB2` user only. For example, to switch to the user mode, run the following command:

`su - db2inst1`

4. Navigate to the `/opt/protegility/defiance_dps/pep/sqlscripts/db2` directory.
5. To run the `dropobjects.sql` script run the following command:

`# db2 -tfdropobjects.sql`

6. To delete the `pep` directory, including the sub-directories, run the following command:

`rmdir -r /opt/protegility/defiance_dps/pep`

The confirmation message for the successful connection to the database appears and the UDFs are uninstalled successfully.

```
Database Connection Information  
  
Database server      = DB2/LINUXX8664 11.5.0.0  
SQL authorization ID = DB2INST1  
Local database alias = TESTDB
```

12.7.5.9.2 Uninstalling the Log Forwarder

This section describes the steps to uninstall the Log Forwarder.

- To uninstall the Log Forwarder:

1. To stop the Log Forwarder, run the following command:

-
- `# /opt/protegility/fluent-bit/bin/logforwarderctrl stop`
2. To delete the `fluent-bit` directory, including the sub-directories, run the following command:
`rmdir -r /opt/protegility/fluent-bit`

12.7.5.9.3 Uninstalling the PEP Server

This section describes the steps to uninstall the PEP server.

► **To uninstall the PEP Server:**

1. Navigate to the `/opt/protegility/defiance_dps/bin` directory.
2. To stop the PEP server, run the following command:
`# ./pepsrvctrl stop all`
3. To delete the `defiance_dps` directory, including the sub-directories, run the following command:
`rmdir -r /opt/protegility/defiance_dps`

12.7.6 Installing and Uninstalling Netezza Database Protector

This section discusses the details of all the steps required for installing and uninstalling the Netezza Database Protector. The following steps broadly describe how Netezza Database Protector is installed.

- Ensure that the prerequisites for installing the Netezza Database Protector are met.
- Download and extract the Netezza Database Protector package to the right directories.
- Install the PEP Server.
- Install the Protector Proxy.
- Install the UDFs.
- Configure SPUs for PEP Server and Protector Proxy.
- Start the PEP Server.

12.7.6.1 Prerequisites

Ensure the following prerequisites are met before installing Netezza DB Protector on a Netezza setup.

1. The ESA appliance is installed, configured, and running.
2. You have the IP address or host name of the ESA noted down.
3. You have root access to the operating system.
4. The Netezza host and SPUs are installed, configured, and running.
5. You have DBA rights to the Netezza database.
6. Starting from the version 7.2.0 release, if you are installing the ESA for the first time, ensure that the Policy Management is initialized prior to installing the protector.

For more information about initializing the Policy Management, refer to section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.2](#).

12.7.6.1.1 Prerequisite tasks before installing Netezza

These are some of the other tasks which, if completed, help you during installation and configuration.

- Log in to the Netezza file system as *nz* user with the following command.

```
# su -nz
```
- Check and note the Netezza version available in your machine to verify that it is compatible with the Netezza Database Protector that you are about to install.
- Check and note the path to the directories that are shared through NFS between the Netezza Host and all SPUs. Following are the directories:
 - Directory from where files are exported by Netezza Hosts: */nz/export/tools/*
 - Directory where these files are mounted on Netezza SPU: */var/opt/nz/host/tools*
- Create a directory on the Netezza Host in the shared directory.

```
$ mkdir /nz/export/tools/protegity
```
- The SPUs are not accessible directly by the PEP server. Therefore, use the the allowed servers option, which lets you add multiple servers, that is part of the Data Store creation process on the ESA to deploy policies and Proxy servers for role_member retrieval and to collect audit records from each SPU node.

For more information about Data Store creation, refer to section *Working with Data Stores* in the *Protegity Policy Management Guide 9.1.0.2*.

12.7.6.2 Downloading and Extracting Protector Package

Before you begin

Copy the package, *DatabaseProtector_<OS>-<arch>_<Netezza Distribution>-<bit size>_<build number>.tgz*, that you have received from Protegity to a temporary directory in the Netezza host machine. One example of the Netezza Database Protector package file could be *DatabaseProtector_RHEL-5-64_x86-64_Netezza-32_7.2.0.xx.tgz*.

► To Extract the Netezza Database Protector Tar File on the Netezza Host:

1. Copy the Netezza Database Protector package, received earlier to the shared directory, */nz/export/tools/protegity* that you have already created on the Netezza host.
2. Extract the package, *DatabaseProtector_<OS>-<arch>_<Netezza Distribution>-<bit size>_<build number>.tgz* to the same directory.

After extraction, make sure that the following tar files are present. These files are required for installation:

- *PepServerSetup_Linux_<bit size>_<build-number>.sh*
- *ProxySetup_Linux_<bit size>_<build-number>.sh*
- *PepNetezza<n.n>Setup_Linux_<bit size>_<build-number>.sh*
- *NetezzaPostConfig.sh*

Note: For information about the installing Protector proxy, refer to section *Protector Proxy*.

12.7.6.3 Installing Netezza Database Protector on Netezza Host



12.7.6.3.1 Install DPS

► To install DPS:

1. To install the PEP server in `/nz/export/tools/protegility` directory, navigate to the directory and run the following command.

```
# ./PepServerSetup_Linux_<bit size>_<build number>.sh
```

Enter the ESA host name when prompted.

2. Install PEP UDF functions on the Netezza Host.

```
# ./PepNetezza<n.n>Setup_Linux_<bit size>_<build number>.sh
```

Enter the maximum data size that must be allocated by the database UDFs.

Note:

The default value is 500 characters. You must modify the default value in this step as per your requirements for maximum character length.

3. Change the access rights to the PEP server directory to `755`.
4. Install the Protector Proxy.

```
# ./ProxySetup_Linux_<bit size>_<build number>.sh -esa <ESA IP> -dir /nz/export/tools/protegility
```

For information about the installing Protector proxy, refer to section [Protector Proxy Installation](#).

5. Run this command to setup post-configuration.

```
# ./Netezza<n.n>_PostConfigSetup_Linux_<bit size>_<build number>.sh
```

6. When prompted, enter the database name where the UDT functions reside.

12.7.6.3.2 Create PEP UDFs

Before you begin

Return to Netezza user, `nz`, and execute these steps.

► To create PEP UDFs:

1. Launch `nzssql` and create the database `<database name>`.

```
# nzssql
SYSTEM(ADMIN+ => create database <database name>;
SYSTEM(ADMIN) => \q
```

2. Change the directory to `/nz/export/tools/protegility/defiance_dps/pep/sqlscripts/netezza` and run the following SQL script to create PEP functions in the database.

```
# nzssql -d <database name> -f createobjects.sql
```

To verify that all functions are installed in the database, run the following command.

```
# nzssql -c 'show function' -d <database name> | grep -i PTY
```

3. Change directory to the `/nz/export/tools/protegility/defiance_dps/data` directory.

12.7.6.4 Configuring Netezza Database Protector

12.7.6.4.1 Configure PEP Server Configuration Files

► To configure PEP server configuration files:

1. To configure the PEP server configuration file, open the `pepserver.cfg` file located in the `/nz/export/tools/protegility/defiance_dps/data` directory.
2. Edit the `pepserver.cfg` file based on your requirements.

12.7.6.4.2 Configure PEP Server and Post Configuration Files

This section describes the series of steps to configure the PEP server and the post configuration files.

Before you begin

All the post configuration files for Netezza are available in the `/nz/export/tools/protegility/defiance_dps/postconfigure` directory and its subdirectories.

► To configure PEP server and post configuration files:

1. Change the directory to `/nz/export/tools/protegility/defiance_dps/postconfigure/scripts` directory.
2. Execute the SQL script to create the Protegility user-defined table functions (UDTFs) in the database PROTEGILITY. This script verifies the existence of SPUs and creates data directories for all SPUs.

```
# ./post_config.sh
```

The output message displays the existing SPUs and the created data directories.

3. If you must increase the maximum size in bytes for the shared memory segment, then set the `SHMMAX` parameter on the Netezza host.

The following snippet displays a sample of this setting. It increases the `SHMMAX` value to 500 MB.

```
SHMMAX= cat /proc/sys/kernel/shmmmax
echo "kernel.shmmmax=$SHMMAX" >> /etc/sysctl.conf
echo 500000000 > /proc/sys/kernel/shmmmax
sysctl -w kernel.shmmmax=500000000
```

4. From the `/nz/export/tools/protegility/defiance_dps/postconfigure/scripts` directory, execute the following script to check the added event handlers.

```
./event_show.sh
```

The following output message appears:

| Field Name | Value |
|------------|-------------------|
| Name | PtyStartup |
| On | yes |
| Event Type | Sys State Changed |

```

Event Args Expr $previousState != online && $currentState == online
Notify Type Run Command
Destination /nz/export/tools/protegility/defiance_dps/postconfigure/scripts/p
CC
Message NPS system $HOST went online at $eventTimestamp $eventSource.
Body Text
Call Home no
Aggregate Count 0

```

Note: The Protector proxy must be installed on the same node where the PEP server is running in the `/nz/export/tools/protegility` directory.

12.7.6.5 Verifying Netezza Database Protector Installation and Configuration

This section is not mandatory but helps you to verify the installation and configuration steps before you start Netezza DB Protector.

► To verify Netezza DB Protector installation and configuration:

1. The directory structure and file contents of the Netezza Host and SPUs should be similar to the following.

```
cd /nz/export/tools/protegility/defiance_dps/data_spu<n>
```

2. The SPU data subdirectories mode should be *rwx* for ‘others’. Modify this mode, if required.

```
# chmod -R o+rwx data_spu*
```

3. All SPU hostnames should be listed on the text file using the following command.

```
# /nz/export/tools/protegility/defiance_dps/postconfigure/scripts/list_spus.sh
```

4. Create the user-defined table functions in the database by performing the following steps.

- a. Change the directory to `/nz/export/tools/protegility/defiance_dps/postconfigure/sqlscripts` and execute the following SQL script:

```
# nzsql -d <database name> -f create_pty_control.sql
```

- b. To verify that the table functions have been created, run the following command:

```
# nzsql -d <database name> -c 'show function' | grep -e PTY_CONTROL -e PTY_HOST
```

5. Edit the `ptyudtfctrl.sh` script file and check if the specified database connection is accurate.

```
cd /nz/export/tools/protegility/defiance_dps/postconfigure/scripts
```

```
vi ptyudtfctrl.sh
```

12.7.6.6 Using Netezza Database Protector

After the Netezza DB Protector is installed, a specific package of User Defined Table Function (UDTF) is used to start, stop, and show the status of the PEP server since the ESA cannot directly access the SPU. The function, `pty_control` is called, which executes OS commands on all connected SPUs. A limitation of this function is that it cannot generate error handling messages, and to do so an analysis of the PEP server log files is required.

In a Netezza system, the ESA cannot directly access the OS of the SPUs. To start and stop the PEP server for each SPU, the UDTF function `PTY_CONTROL` must be called. This UDTF executes OS commands on all connected SPUs. There is a limitation on error handling that can be provided so that error messages caused by start or stop operations can be handled by normal analysis of the PEP server log files.

When the Netezza Database Protector is installed, the default location is `/nz/export/tools/protegility` but the location can be customized for the Netezza installation.

A normal installation of the IBM Netezza Database Protector creates the following directory structure:

```
 ${DEFAULT_PROTEGITY_PATH}/defiance_dps
    -- bin
    -- data
    -- data_spu0101
    -- data_spu0102
    -- ...
    -- data_spuxxxyy
    -- pep
    -- postconfigure
```

The *pty_control* table function must be forced to process across all SPUs.

This can be done by running the script *ptyudtfctrl.sh*, where specific SQL statements are executed to start, stop, or check the status of the PEP servers across all SPUs. The following example shows the SQL statement to start the PEP servers across all SPUs.

```
SELECT f.* FROM _v_dual_dslice AS vds,
(SELECT MIN(ds_id) AS mds FROM _v_dslice GROUP BY ds_prihwid) AS vs,
TABLE ( pty_control(vds.dsid, 'START',null) )
AS f WHERE vs.mds = vds.dsid
```

During the post-configuration installation, another overall main start and stop script, *ptyctrl.sh*, is created and saved in the *defiance_dps/bin* directory.

This script includes the *ptyudtfctrl.sh* script to start up the PEP servers across all SPUs and also start or stop scripts for the PEP server and Proxy servers on the Netezza Host.

The status of a node can also be checked by running the following script with a parameter:

ptyctrl.sh start/stop/status

This script, *ptyctrl.sh*, should normally be used when starting, stopping, and obtaining the status of PEP servers, Protector Proxy on the master and all SPUs of a Netezza multi-node system.

During post-configuration after installation, a script named *ptyctrl.sh* is created in the */nz/export/tools/protegity/defiance_dps/bin* directory. This script can start up, show the status, and stop all the DPS servers of the Netezza DB Protector.

1. Change the directory to */nz/export/tools/protegity/defiance_dps/bin* directory.
2. Start all servers on the Netezza Host and SPUs using the following command.

./ptyctrl.sh start

The following output message appears.

```
Server status...
Service is running as PID=17595
pepper -dir /nz/export/tools/protegity/defiance_dps/data is not running!
HOSTNAME | CODE | MESSAGE
-----+-----+-----
spu0101 | 0 | PEP Server running.
(1 row)
```

Your Netezza DB Protector is now up and running.

12.7.6.6.1 Create Policy in Policy Management

The Protector requires policies to ensure that unauthorized users do not access the data and that the correct data is protected and unprotected.

For more information about creating policies, refer to the section *Creating and Deploying Policies* in the [Protegility Policy Management Guide 9.1.0.2](#).

12.7.6.7 Uninstallation

12.7.6.7.1 Uninstalling the UDFs

► To Uninstall the UDFs:

1. Login to the Netezza File System.
2. Navigate to the `/nz/export/tools/protegility/defiance_dps/pep/sqlscripts/netezza` directory.
3. Login to the database.
4. Execute the following query to drop UDF.
`# nzsql -d <database name> -f dropobjects.sql`
5. Navigate to the `~/defiance_dps/postconfigure/sqlscripts` directory and execute the following query.
`# nzsql -d <database name> -f drop_pty_control.sql`
6. Navigate to the `~/defiance_dps/postconfigure/scripts` directory and execute the `./event_delete.sh` command.
7. To verify that the UDFs have been removed, run the following command.
`# nzsql -d <Database Name, default: PROTEGRITY> -c 'show function' | grep -i PTY`
The output must display zero rows.

12.7.6.7.2 Uninstalling the PEP Server

► To uninstall the PEP server:

1. Stop the PEP server by executing the `ptyctrl.sh` command.
2. Remove the `/nz/export/tools/protegility/defiance_dps/` directory.
The `/nz/export/tools/protegility/defiance_dps/` directory must be empty.

Note: For information about uninstalling the Protector proxy, refer to the section [Installing and Uninstalling Protector Proxy](#).

12.7.7 Installing and Uninstalling the Trino Protector

This section describes the procedure to install and uninstall the Trino Protector.

Note: By default, the Installation script assumes that a user account with `sudoer` privileges is running the script and will use `sudo` to execute the commands. If `sudoer` privileges are disabled on your Trino cluster, then use the `--sudo-disabled` argument.

For more information about the argument, refer to the [Optional Arguments for the Installation Script](#) table.

The following table provides the order of installation for the Trino Protector.

Table 12-103: Trino Protector - Order of Installation

| Order of installation | Description | Reference |
|-----------------------|-----------------------------|---|
| 1. | Install the ESA | Installing the ESA On-Premise |
| 2. | Verify the Prerequisites | Prerequisites |
| 3. | Run the Configurator Script | Running the Configurator Script |
| 4. | Run the Installation Script | Running the Installation Script |

12.7.7.1 Verifying the Prerequisites for Installing the Trino Protector

Ensure that the following prerequisites are met, before installing the Trino Protector:

- The Trino cluster is installed, configured, and running.
- ESA appliance version 9.1.0.0 or higher is installed, configured, and running.
- Starting from the version 7.2.1 release, if you are installing the ESA for the first time, ensure that the Policy Management is initialized prior to installing the protector.

Note: For more information about initializing the Policy Management, refer to section *Initializing the Policy Management* in the [Protegility Policy Management Guide 9.1.0.2](#).

- A standalone Linux machine or a Trino Cluster node that has access to the ESA to run the configurator script.
- The configurator script generates a single-node installation script that expects the user executing it to be a *sudoer* user with administrative permissions in Linux.

Note: If *sudoers* is disabled on the Linux node, then please use the *--sudo-disabled* argument.

For more information about the arguments to be passed for the single-node installation script, refer to the section [Running the Installation Script](#).

- The single-node installation script will add the *ptyitusr* and the *ptyitusrgroup* group on the Linux nodes during installation.
- The ports that are configured on the ESA and the nodes in the cluster, which will run the Trino Protector, are listed in the following table:

Table 12-104: List of Ports for the Big Data Protector

| Destination Port | Protocols | Sources | Destinations | Descriptions |
|------------------|-----------|---|---|--|
| 8443 | TCP | PEP server on the Trino Protector cluster node | ESA | The PEP server communicates with the ESA through port 8443 to download a policy. |
| 9200 | | Log Forwarder on the Trino Protector cluster node | Protegility Audit Store appliance | The Log Forwarder sends all logs to the Protegility Audit Store appliance through port 9200. |
| 15780 | | Protector on the Trino Protector cluster node | Log Forwarder on the Trino Protector cluster node | The Trino Protector writes the Audit Logs to <i>localhost</i> through port 15780. The PEP server Application Logs are also written to <i>localhost</i> through port 15780. The |



| Destination Port | Protocols | Sources | Destinations | Descriptions |
|--|-----------|---|--|--|
| | | | | Log Forwarder reads the logs from that socket. |
| 16700 | | DPS Admin on the Trino Protector cluster node | PEP server on the Trino Protector cluster node | The DPS Admin client tool uses <i>localhost</i> port <i>16700</i> . |
| Port Requirement for Proxy Node (Optional) | | | | |
| 8443 | TCP | PEP server on the Trino Protector cluster node | PTY Proxy Node (Similar or different Trino Protector cluster node) | The PEP server communicates with the Proxy node through port <i>8443</i> , which substitutes the stream to the ESA. |
| 9200 | | Log Forwarder on the Trino Protector cluster node | PTY Proxy Node (Similar or different Trino Protector cluster node) | The Log Forwarder communicates with the Proxy node through port <i>9200</i> , which substitutes the stream to the Protegility Audit Store Appliance. |

12.7.7.2 Extracting the Files from the Installation Package

This section provides information about extracting the installation package for the installation of the Trino Protector.

► To extract the files from the installation package:

1. After receiving the installation package from Protegility, copy it to a standalone Linux machine or a node in the Trino cluster, which has access to the ESA, in any temporary directory, such as, */opt/trino*.
2. To extract the single-node installation script from the installation package, run the following command.

```
tar xvf DatabaseProtector_Linux-ALL-64_x86-64_Trino-<version>-64_9.1.0.0.x.tgz
```

3. Press ENTER.

The command extracts the configurator script.

```
TrinoProtectorConfigurator_9.1.0.0.x.sh
```

12.7.7.3 Running the Configurator Script

1. To execute the configurator script, run the following command:

```
./TrinoProtectorConfigurator_9.1.0.0.x.sh
```

2. Press ENTER.

The prompt to continue with the installation appears.

```
*****
***** Welcome to the Configurator for Protegility Trino Protector *****
***** This will configure and generate the Protegility Trino Protector Generic Installation *****
***** Script for a single Trino node. *****
```

```
Do you want to continue? [yes or no]:
```

3. To continue, type yes.



4. Press ENTER.

The prompt to enter the directory, on the cluster node, to install the Trino protector appears.

```
*****
      Welcome to the Configurator for Protegility Trino Protector
*****
This will configure and generate the Protegility Trino Protector Generic Installation
Script for a single Trino node.

Do you want to continue? [yes or no]: yes
Protegility Trino Protector Configurator started...
Enter the Installation Directory on cluster node
[default: /opt/protegility]:
```

5. Enter the location where you want to extract the installation files.

Note: If you want to use the default location, then press ENTER.

6. Press ENTER.

The prompt to enter the ESA Hostname or the IP address appears.

```
Enter the ESA Hostname/IP Address:
```

7. Enter the ESA hostname or the IP address.

8. Press ENTER.

The prompt to enter the listening port appears.

```
Enter ESA host listening port [8443]:
```

9. Enter the listening port for the ESA host.

10. Press ENTER.

The prompt to enter the username for the ESA appears.

```
Enter ESA Username:
```

11. Enter the username.

12. Press ENTER.

The prompt to enter the password to download the certificates appears.

```
Temporarily setting up PepServer defiance_dps directory structure on current node...
Please enter the password for downloading certificates
[]:
```

13. Enter the password.

14. Press ENTER.

The script downloads the certificates from the ESA and prompts to specify the Audit Store type.

```
Unpacking...
Extracting files...
Downloading certificates from X.X.X.X:8443...
  % Total    % Received % Xferd  Average Speed   Time     Time     Current
                                         Dload  Upload   Total   Spent    Left  Speed
 100 20480  100 20480     0      0  27130      0 --::-- --::-- --::-- 27125

Extracting certificates...
Certificates successfully downloaded and stored in /<installation_directory>/defiance_dps/
data

Protegility PepServer installed in /<installation_directory>/defiance_dps.
```

```

Repackaging defiance_dps with ESA certificates...
Fetched and Repackaged ESA Certificates successfully..

Select the Audit Store type where Log Forwarder(s) should send logs to.

[ 1 ] : Protegility Audit Store
[ 2 ] : External Audit Store
[ 3 ] : Protegility Audit Store + External Audit Store

Enter the no.:

```

15. To select the Audit Store type, select any one of the following options:

Table 12-105: Options to select the Audit Store Type

| Option | Description |
|--------|--|
| 1 | To use the default setting with the Protegility Audit Store appliance, type <i>1</i> . If you enter <i>1</i> , then the default Fluent Bit configuration files are used and Fluent Bit will forward the logs to the Protegility Audit Store appliance. |
| 2 | To use an external audit store, type <i>2</i> . If you enter <i>2</i> , then the default Fluent Bit configuration files used for the External Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are renamed (<i>out.conf.bkp</i> and <i>upstream.cfg.bkp</i>) so that they will not be used by Fluent Bit. Additionally, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |
| 3 | To use a combination of the default setting with an external audit store, type <i>3</i> . If you enter <i>3</i> , then the default Fluent Bit configuration files used for the Protegility Audit Store (<i>out.conf</i> and <i>upstream.cfg</i> in the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory) are not renamed. However, the custom Fluent Bit configuration files for the external audit store are copied to the <i>/opt/protegility/fluent-bit/data/config.d/</i> directory. |

When you select option *2* or *3*, the prompt to enter the path that stores the custom Fluent Bit configuration file appears.

```
Enter the local directory path on this machine that stores the Fluent-Bit configuration files for External Audit Store:
```

16. Press ENTER.

The prompt to generate the logs for the PEP server, in a file, appears.

```
Do you want PepServer's log to be generated in a file? [yes or no]:
```

17. To generate the logs for the PEP server in a file, type *yes*.

18. Press ENTER.

The configurator script generates the single-node installation script.

```
PepServer's log will be generated in a file.
```

```
Successfully finished configuring the Trino Protector Installation Script.
```

```
The single-node Installation Script is generated at /<current_working_directory>/Installation_Script/TrinoProtector_InstallationScript_9.1.0.0.x.sh
```

Next Steps:

- 1) Copy the Installation Script to a storage location that is reachable by the Trino cluster nodes.



- 2) You can create a shell script that will download the Installation Script and execute it by passing the correct arguments.
- 3) Ensure to pass the correct Command Line arguments to the Installation Script. Run `./TrinoProtector_InstallationScript_9.1.0.0.x.sh --help` to print Usage and Help Info.
- 4) For a new Trino cluster, you can configure the shell script to be executed at Node Startup via Bootstrap/Init Script mechanism if your cluster provides it.
- 5) For a running Trino cluster, you can execute the shell script on the existing nodes.

12.7.7.4 Running the Installation Script on Every Node in the Trino Cluster

Execute the following steps on every node (coordinator and worker) of the Trino Cluster

Important: If you want to add a new node to the Trino cluster after installing the Trino protector, then ensure that you run the installation script on the new node.

1. Login to the node, where you want to execute the installation script.
2. Copy/Download the single-node installation script previously generated by the configurator script to any directory.
3. Navigate to the directory where the single-node installation script is located.
4. To view the syntax and the usage of the installation script, run the following command.

```
./TrinoProtector_InstallationScript_9.1.0.0.x.sh --help
```

5. Press ENTER.

The command displays the syntax with the mandatory and optional arguments. The mandatory and optional arguments are explained in the following tables.

Table 12-106: Mandatory Arguments for the Installation Script

| Argument | Description |
|--|--|
| <code>--install-pepserver-and-logforwarder=<yes/no></code> | Instructs the script whether to install the PEP server and the Log Forwarder on the current node in the Trino cluster. The acceptable values are: <ul style="list-style-type: none"> • <i>yes</i> - install the PEP server and the Log Forwarder on the current node in the Trino cluster. • <i>no</i> - skip installing the PEP server and the Log Forwarder if it is already installed and running on the current node in the Trino cluster. |
| <code>--trino-plugin-dir=</path/to/plugin/></code> | Specifies the absolute path to the Trino plugin directory. <p>Note: You can set a custom plugin path using the <i>plugin.dir</i> property in the <i>node.properties</i> Trino configuration file.</p> |
| <code>--trino-service-user=<user></code> | Specifies the name of the user running the Trino server. You can use this argument to set the ownership of the <i>peptrino plugin</i> directory and also to restart the Trino server if you specify the <code>--restart-trino-server-via=launcher</code> argument. |

Table 12-107: Optional Arguments for the Installation Script

| Argument | Description |
|------------------------------|--|
| <code>--sudo-disabled</code> | <ul style="list-style-type: none"> • Use this flag if <i>sudoers</i> is disabled on the cluster. When you provide this argument, a user with elevated privileges is required to execute the script. |

| Argument | Description |
|---|--|
| | <ul style="list-style-type: none"> When you exclude this argument, a user with sudoers privilege is required to execute the script. <p>It is recommended to use a <i>NOPASSWD</i> sudoers user. Else, the script will prompt for a password.</p> |
| <code>--restart-trino-server-via=<systemd/init/launcher></code> | <p>If you specify this argument, then the script will restart the running Trino server after installing the <i>peptrino</i> plugin. The script will not restart the Trino server if the server is in the stopped state. If you want to use this argument, ensure that you enable Sudo to restart the Trino server. The acceptable values are:</p> <ul style="list-style-type: none"> <i>systemd</i> - instructs the script to use the <i>systemctl</i> command to check the status and restart the Trino server. This argument requires the <code>--trino-systemd-service-name</code> argument to be specified. <i>init</i> - instructs the script to use the <i>service</i> command to check the status and restart the Trino Server. This argument also requires you to specify the <code>--trino-init-service-name</code> argument. <i>launcher</i> - instructs the script to use the Trino launcher script to check the status and restart Trino Server. This argument requires the <code>--trino-launcher-path</code> and the optional <code>--trino-launcher-args</code> arguments. <p>If you exclude this argument, then the script will not attempt to restart the Trino server. You must manually restart the Trino server after the script execution.</p> |
| <code>--trino-systemd-service-name=<service name></code> | <p>Specifies the name of the systemd service associated with the Trino server. You must specify this argument when you use the <code>--restart-trino-server-via=systemd</code> argument.</p> |
| <code>--trino-init-service-name=<init service name></code> | <p>Specifies the name of the Sys V init service associated with the Trino server. You must pass this argument when you specify the <code>--restart-trino-server-via=init</code> argument.</p> |
| <code>--trino-launcher-path=</path/to/bin/launcher></code> | <p>Specifies the absolute path to the Trino server launcher script. For example, <i>/usr/lib/trino/bin/launcher</i>. You must specify this argument when you use the <code>--restart-trino-server-via=launcher</code> argument.</p> |
| <code>--trino-launcher-args="arg1 [arg2...]"</code> | <p>Specifies the valid command line arguments to the Trino launcher script. You can use this argument with the <code>--trino-launcher-path</code> argument. If you specify this argument, then the arguments listed between the double-quotes will be passed to the Trino launcher script for the status and restart commands. If you fail to specify this argument, then no argument will be passed to the Trino launcher script for the status and restart commands.</p> |
| <code>--reuse-jpeplite-from-path=</path/to/jpeplite/lib></code> | <p>Specifies the absolute path to the existing jpeplite/lib/ directory on the cluster node. You can use this argument with the <code>--install-pepper-and-logforwarder=no</code> argument when the JpepLite libraries packaged in this build is incompatible with the existing installed PEP server version. The existing <i>/path/to/jpeplite/lib/</i> directory must contain the <i>jpeplite.jar</i>, <i>jpeplite.properties</i>, and <i>jpeplite.plm</i> files and must be readable.</p> |
| <code>--protector-logs-output=<tcp/stdout/file></code> | <p>Instructs the Trino Protector to write the protector logs to this output. If you specify an attribute for this argument, then the update will change the output property in the <i>pepper.cfg</i> file. The acceptable values are:</p> <ul style="list-style-type: none"> <i>tcp</i> (Default) - specifies that the logs are written to the TCP socket specified in the <i>pepper.cfg</i> file. <i>stdout</i> - specifies that the logs are written to the Trino Server's <i>stdout</i> parameter. |

| Argument | Description |
|---|---|
| | <ul style="list-style-type: none"> • <i>file</i> - specifies that the logs are written to the file whose path is set in the <i>--protector-logs-output-filename</i> argument. |
| <i>--protector-logs-output-filename=</path/to/logs.txt></i> | Specifies the absolute path to the file on the cluster node to which the protector logs are written. You must use this argument with the <i>--protector-logs-output-file</i> argument. If you fail to specify this argument, then the default file name of <i>/opt/logs.txt</i> will be used. This argument will add the <i>outputfilename</i> property in the <i>pepservice.cfg</i> file. This script will create the file on the cluster node if the file is not available. Ensure that the Trino service user has write permissions to this file path. |
| <i>--wait-for-trino-installation</i> | <ul style="list-style-type: none"> • If you specify this argument, then the installation script will create and run a secondary bash script as a background process that will wait for Trino server to be installed and started on the node and only then install the plugin jars and restart the Trino Server. (To be used in scenarios where the Trino Server will be always installed and started after this installation script is executed. E.g. EMR clusters). This flag requires the <i>--restart-trino-server-via</i> argument for restarting the Trino server. You must enable Sudo for this argument. • If you fail to specify this argument, then the installation script will not wait for the Trino server to be installed and started and will go ahead with the plugin installation. Ensure that the Trino Server is installed and the plugin directory exists before the installation script is executed. |

6. Depending on the requirements, run the installation script with the required arguments.

For example, on a Starburst Trino cluster installed via RPM, one combination of the arguments to the single-node installation script is listed below.

```
./TrinoProtector_InstallationScript_9.1.0.0.x.sh \
    --install-pepservice-and-logforwarder=yes \
    --trino-plugin-dir=/usr/lib/starburst/plugin \
    --trino-service-user=starburst \
    --restart-trino-server-via=systemd \
    --trino-systemd-service-name=starburst
```

Note: If you want the Trino Server to automatically restart after installing the components, then you must specify the value for the *--restart-trino-server-via* argument for the installation script. Else, you will have to manually restart the Trino Server after the installation is complete.

7. Press ENTER.

The script installs the components as specified in the arguments.

```
./TrinoProtector_InstallationScript_9.1.0.0.x.sh \
>     --install-pepservice-and-logforwarder=yes \
>     --trino-plugin-dir=/usr/lib/starburst/plugin \
>     --trino-service-user=starburst \
>     --restart-trino-server-via=systemd \
>     --trino-systemd-service-name=starburst

Protegility Trino Protector Installation Script started...

Validating sudo permissions for root
*****
        Welcome to the Trino Protector Install Wizard.
*****
This will install the Trino Protector on your system.

Group 'ptyitusrgroup' created
User 'ptyitusr' created
```

```
PepServer installation started
*****
***** Welcome to the PepServer Setup Wizard.
*****
PepServer installed on current node at location /opt/protegility/defiance_dps/
Logforwarder installation started
*****
***** Welcome to the LogForwarder Setup Wizard.
*****
Unpacking.....
Extracting files...
Unpacked logforwarder compressed file...

LogForwarder installed on current node at location /opt/protegility/fluent-bit/

PepTrino Plugin Jars installation started
*****
***** Welcome to the PepTrino Setup Wizard.
*****
Unpacking.....
Extracting files...
Unpacked peptrino compressed file...

PepTrino installed on current node at location /opt/protegility/peptrino/

Jpeplite installation started
*****
***** Welcome to the Jpeplite Setup Wizard.
*****
Unpacking.....
Extracting files...
Unpacked jpeplite compressed file...

Jpeplite for Trino Protector installed on current node at location /opt/protegility/
peptrino/jpeplite_lib/

Moving Uninstallation Script to /opt/protegility/peptrino/scripts/
Creating cluster_utils directory in /opt/protegility/peptrino/scripts/
Creating protector_version directory in /opt/protegility/peptrino/
Starting Logforwarder on current node...
Starting PepServer on current node...

Trino Protector plugin jars and Jpeplite libraries are installed within /opt/protegility/
peptrino/ directory

Finished executing TrinoProtectorInstall<version>_Linux-ALL_9.1.0.0.x.sh script. Check
the logs at /opt/protegility/logs/

NOT waiting for Trino Server Installation and Start...

Started installation of PepTrino in plugin dir in foreground...

Checking if Trino Plugin Dir is present on node.
Trino Plugin Directory /usr/lib/starburst/plugin found.

Creating peptrino directory within Trino plugin directory.

Getting the names of plugin jars

Creating Symbolic Links within Plugin Directory...

Trino Protector jars' symbolic links created in /usr/lib/starburst/plugin/peptrino/
```

```

Checking if Trino service user exists.

Service User starburst exists on node

Changing ownership of PepTrino Plugin dir to starburst

Checking if systemctl is on PATH.
systemctl found on PATH.

Checking if 'starburst' is a valid systemd unit.
'starburst' is a valid systemd unit.

Checking if Trino Server is started and running via systemctl
Trino server is running
Restarting Trino Server...

Trino Server successfully restarted via systemctl.

Trino Protector UDFs registered on Trino Server restart. Verify via trino CLI (show
functions;) or by checking Trino's server.log

Successfully completed all steps of Installation Script

```

The installation script generates the logs in the `/<installation_directory>/logs/` directory.

12.7.7.5 Working with the Cluster Utilities

The Cluster Utilities provide scripts to perform the following actions:

Table 12-108: Cluster Utility Scripts

| Script | Description |
|--|---|
| <code>cluster_pepsrvctrl.sh</code> | <p>Manages the PEP server on all the nodes in the Trino cluster. You can use this script to perform the following actions on the PEP server:</p> <ul style="list-style-type: none"> Start the PEP server on all the nodes Stop the PEP server on all the nodes Restart the PEP server on all the nodes View the status of the PEP server on all the nodes |
| <code>cluster_logforwarderctrl.sh</code> | <p>Manages the Log Forwarder on all the nodes in the Trino cluster. You can use this script to perform the following actions on the Log Forwarder:</p> <ul style="list-style-type: none"> Start the Log Forwarder on all the nodes Stop the Log Forwarder on all the nodes Restart the Log Forwarder on all the nodes View the status of the Log Forwarder on all the nodes |
| <code>sync_defiance_dps.sh</code> | <p>Copies the files under the current node's <code>/opt/protegility/defiance_dps/data/</code> directory to all other nodes in the Trino cluster mentioned in the <code>hosts</code> file.</p> |
| <code>sync_fluent_bit.sh</code> | <p>Copies the files under the current node's <code>/opt/protegility/fluent-bit/data/</code> directory to all other nodes in the Trino cluster mentioned in the <code>hosts</code> file.</p> |

The Cluster Utilities are installed in the following location when you run the installation script:

```
<installation_directory>/peptrino/scripts/cluster_utils/
```



The scripts also require a path to manually created hosts file that contains the IP address/hostname of all the nodes in the cluster other than the current node on each line. You can use the following syntax to list the IP address/hostname of all the nodes in the Trino cluster:

```
[user@]<ip address or hostname>[:port]
```

The cluster utility scripts use *pssh* (parallel ssh) and *pscp* (parallel scp). These utilities require Python to be installed on the current node. To verify whether Python is installed on the current node, run the following command:

```
/usr/bin/env python --version
```

The command returns the version of Python installed on the system.

If you are unable to detect Python on the current node, then ensure that you have a compatible version of Python installed on the current node (preferably Python 3.x).

Ensure that the utilities are able to detect the version of Python using the following command:

```
/usr/bin/env python
```

Note: The Cluster Utilities script will only work on clusters where the *sudoer* privileges are enabled.

► To view the cluster utility scripts:

1. Login to any node in the Trino cluster.
2. To navigate to the *cluster_utils* directory, run the following command:

```
cd /<installation_directory>/peptrino/scripts/cluster_utils
```

3. To view the utility scripts, run the following command:

```
ls -la
```

4. Press ENTER.

The command lists the scripts available in the directory.

12.7.7.5.1 Working with the PEP Server Cluster Utility

The *cluster_pepsrvctrl.sh* script enables you to perform the following actions with the PEP server:

- Start the PEP server on all the nodes
- Stop the PEP server on all the nodes
- Restart the PEP server on all the nodes
- View the status of the PEP server on all the nodes

Note: The Cluster Utilities script will only work on clusters where the *sudoer* privileges are enabled.

► To use the PEP server utility script:

- To navigate to the *cluster_utils* directory, run the following command:

```
cd /<installation_directory>/peptrino/scripts/cluster_utils
```

- To view the syntax and the arguments for the *cluster_pepsrvctrl.sh* script, run the following command:

```
./cluster_pepsrvctrl.sh --help
```

- Press ENTER.

The command lists the arguments and the syntax for the *cluster_pepsrvctrl.sh* script. The arguments are explained in the following table.

Table 12-109: Arguments for the PEP Server Utility Script

| Argument | Description |
|--|--|
| --hostsfile=</path/to/hosts> | Specifies the path to the hosts file. Each line, in the hosts file, must contain the IP address or the hostname of the remaining nodes in the Trino cluster. You can also include the user that the <i>pssh</i> utility will use to connect to the node, in the following format: <i>[user@]host[:port]</i> If you fail to specify the user, then the current user running the script is used. |
| --ssh-auth-type=<password/publickey> | Specifies the SSH authentication type. The accepted values are: <ul style="list-style-type: none"> <i>password</i> - specifies to use the password based SSH authentication. <i>publickey</i> - specifies to use the Public Key Authentication. |
| --password=<actual_password> | Specifies the actual password of the current user to be passed to the <i>pssh</i> utility. You must use this argument only with the password-based SSH authentication. |
| --private-key-path=</path/to/privatekeyfile> | Specifies the path to the SSH private key file to be used by the <i>pssh</i> utility. You must use this argument only with the Public Key SSH authentication. |
| Action | Specifies the actions that the utility script must perform based on the arguments provided. The accepted values are: <ul style="list-style-type: none"> <i>start</i> - instructs the script to start the PEP server on all the nodes in the Trino cluster. <i>stop</i> - instructs the script to stop the PEP server on all the nodes in the Trino cluster. <i>restart</i> - instructs the script to stop and start the PEP server on all the nodes in the Trino cluster. <i>status</i> - instructs the script to report the status of the PEP server on all the nodes in the Trino cluster. |

- To start the PEP server on all the nodes, run the following command.

```
./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --  
private-key-path=<key_file_path>/<name_of_the_private_key_file> start
```

- Press ENTER.

The command starts the PEP server on all the nodes in the Trino cluster.

```
./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --  
private-key-path=/root/<name_of_the_private_key_file> start  
=====  
Hosts file set to '<path_of_the_hosts_file>'  
SSH Authentication Type is set to 'Public Key Authentication'
```



```

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Starting PepServer on current node, Please wait...

PepServer started on current node

Starting PepServer on all nodes, Please wait...

PepServer started on all nodes

The script's logs and operation results are logged in /opt/protegility/logs/
cluster_pepsrvctrl.log

```

- To stop the PEP server on all the nodes, run the following command.

```
./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --
private-key-path=<key_file_path>/<name_of_the_private_key_file> stop
```

- Press ENTER.

The command stops the PEP server on all the nodes in the Trino cluster.

```

./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --
private-key-path=<key_file_path>/<name_of_the_private_key_file> stop
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Stopping PepServer on current node, Please wait...

PepServer stopped on current node

Stopping PepServer on all nodes, Please wait...

PepServer stopped on all nodes

The script's logs and operation results are logged in /opt/protegility/logs/
cluster_pepsrvctrl.log

```

- To check the status of the PEP server on all the nodes, run the following command.

```
./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --
private-key-path=<key_file_path>/<name_of_the_private_key_file> status
```

- Press ENTER.

The command updates the status of the PEP server on all the nodes in the Trino cluster in the log file.

```

./cluster_pepsrvctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --
private-key-path=/root/<name_of_the_private_key_file> status
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Checking status of PepServer on current node, Please wait...

```

```
Checking status of PepServer on all nodes, Please wait...
```

```
The script's logs and operation results are logged in /opt/protegility/logs/
cluster_pepsrvctrl.log
```

12.7.7.5.2 Working with the Log Forwarder Cluster Utility

The *cluster_logforwarderctrl.sh* script enables you to perform the following actions with the Log Forwarder:

- Start the Log Forwarder on all the nodes
- Stop the Log Forwarder on all the nodes
- Restart the Log Forwarder on all the nodes
- Check the status of the Log Forwarder on all the nodes

Note: The Cluster Utilities script will only work on clusters where the *sudoer* privileges are enabled.

► To use the Log Forwarder utilities script:

1. To navigate to the *cluster_utils* directory, run the following command:

```
cd /opt/protegility/peptrino/scripts/cluster_utils
```

2. To view the syntax and arguments for the script, run the following command.

```
./cluster_logforwarderctrl.sh --help
```

3. Press ENTER.

The command displays the actions, arguments, and a detailed description. The arguments are explained in the following table

Table 12-110: Arguments for the Log Forwarder Utility Script

| Argument | Description |
|--|---|
| --hostsfile=</path/to/hosts> | Specifies the path to the hosts file. Each line, in the hosts file, must contain the IP address or the hostname of the remaining nodes in the Trino cluster. You can also include the user that the <i>pssh</i> utility will use to connect to the node, in the following format: <i>[user@]host[:port]</i> If you fail to specify the user, then the current user running the script is used. |
| --ssh-auth-type=<password/publickey> | Specifies the SSH authentication type. The accepted values are: <ul style="list-style-type: none"> • <i>password</i> - specifies to use the password based SSH authentication. • <i>publickey</i> - specifies to use the Public Key Authentication. |
| --password=<actual_password> | Specifies the actual password of the current user to be passed to the <i>pssh</i> utility. You must use this argument only with the password-based SSH authentication. |
| --private-key-path=</path/to/privatekeyfile> | Specifies the path to the SSH private key file to be used by the <i>pssh</i> utility. You must use this argument only with the Public Key SSH authentication. |
| Action | Specifies the actions that the utility script must perform based on the arguments provided. The accepted values are: |

| Argument | Description |
|----------|---|
| | <ul style="list-style-type: none"> • <i>start</i> - instructs the script to start the Log Forwarder on all the nodes in the Trino cluster. • <i>stop</i> - instructs the script to stop the Log Forwarder on all the nodes in the Trino cluster. • <i>restart</i> - instructs the script to stop and start the Log Forwarder on all the nodes in the Trino cluster. • <i>status</i> - instructs the script to report the status of the Log Forwarder on all the nodes in the Trino cluster. |

4. To stop the Log Forwarder on all the nodes, run the following command.

```
./cluster_logforwarderctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --private-key-path=<key_file_path>/<name_of_the_private_key_file> stop
```

5. Press ENTER.

The script stops the Log Forwarder on all the nodes in the cluster.

```
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Stopping Logforwarder on current node...

Logforwarder stopped on current node

Stopping Logforwarder on all nodes...

Logforwarder stopped on all nodes

The script's logs and operation results are logged in /opt/protegility/logs/
cluster_logforwarderctrl.log
```

6. To start the Log Forwarder on all the nodes, run the following command.

```
./cluster_logforwarderctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --private-key-path=<key_file_path>/<name_of_the_private_key_file> start
```

7. Press ENTER.

The script starts the Log Forwarder on all the nodes in the cluster.

```
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Starting Logforwarder on current node...

Logforwarder started on current node

Starting Logforwarder on all nodes...

Logforwarder started on all nodes
```

```
The script's logs and operation results are logged in /opt/protegity/logs/
cluster_logforwarderctrl.log
```

8. To restart the Log Forwarder on all the nodes, run the following command.

```
./cluster_logforwarderctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-
type=publickey --private-key-path=<key_file_path>/<name_of_the_private_key_file> restart
```

9. Press ENTER.

The script restarts the Log Forwarder on all the nodes in the cluster.

```
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Stopping Logforwarder on current node...

Logforwarder stopped on current node

Starting Logforwarder on current node...

Logforwarder started on current node

Stopping Logforwarder on all nodes...

Logforwarder stopped on all nodes

Starting Logforwarder on all nodes...

Logforwarder started on all nodes
```

```
The script's logs and operation results are logged in /opt/protegity/logs/
cluster_logforwarderctrl.log
```

10. To verify the status of the Log Forwarder on all the nodes, run the following command.

```
./cluster_logforwarderctrl.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-
type=publickey --private-key-path=<key_file_path>/<name_of_the_private_key_file> status
```

11. Press ENTER.

The script verifies the status of the Log Forwarder on all the nodes in the cluster and appends the status to the log file.

```
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'

Checking connectivity of cluster nodes...

Checking status of Logforwarder on current node...

Checking status of Logforwarder on all nodes...
```

```
The script's logs and operation results are logged in /opt/protegity/logs/
cluster_logforwarderctrl.log
```

12.7.7.5.3 Updating the PEP Server Configuration and Certificates on all the Nodes

The `sync_defiance_dps.sh` script enables you to replicate the PEP server configuration files and certificates (under the `/defiance_dps/data/` directory) from the current node to all the remaining nodes of the cluster. The script will also restart the PEP server on all the nodes after replicating the configuration.

Note: The Cluster Utilities script will only work on clusters where the `sudoer` privileges are enabled.

► To update the PEP server configuration and the certificates on all the nodes:

1. Navigate to the `/opt/protegility/peptrino/scripts/cluster_utils` directory.
2. To view the syntax for the script, run the following command.

```
./sync_defiance_dps.sh --help
```

3. Press ENTER.

The command displays the examples, description, and the arguments. The arguments are explained in the following table.

Table 12-111: Arguments for the PEP Server Configuration Script

| Argument | Description |
|---|---|
| <code>--hostsfile=</path/to/hosts></code> | Specifies the path to the hosts file. Each line, in the hosts file, must contain the IP address or the hostname of the remaining nodes in the Trino cluster. You can also include the user that the <code>pssh</code> utility will use to connect to the node, in the following format: <code>[user@]host[:port]</code> If you fail to specify the user, then the current user running the script is used. |
| <code>--ssh-auth-type=<password/publickey></code> | Specifies the SSH authentication type. The accepted values are: <ul style="list-style-type: none"> • <code>password</code> - specifies to use the password based SSH authentication. • <code>publickey</code> - specifies to use the Public Key Authentication. |
| <code>--password=<actual_password></code> | Specifies the actual password of the current user to be passed to the <code>pssh</code> utility. You must use this argument only with the password-based SSH authentication. |
| <code>--private-key-path=</path/to/privatekeyfile></code> | Specifies the path to the SSH private key file to be used by the <code>pssh</code> utility. You must use this argument only with the Public Key SSH authentication. |

4. To update the PEP server configuration and certificates on all the nodes in the cluster, run the following command.

```
./sync_defiance_dps.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --private-key-path=<key_file_path>/<name_of_the_private_key_file>
```

5. Press ENTER.

The command updates the PEP server configuration and the certificates on all the nodes in the cluster.

```
=====
Hosts file set to '<path_of_the_hosts_file>'

SSH Authentication Type is set to 'Public Key Authentication'

SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'
```

```

Checking connectivity of cluster nodes...

Trino Protector PepServer Configuration & Certificates cloning started

Stopping PepServer on current node...

Stopping PepServer on all nodes...

Creating defiance_dps_old/data_11-09-2023_13:29:27/new_data directory on all nodes...

Changing permission of defiance_dps_old/ on all nodes...

Removing PepServer Configuration & Certificates from all nodes...
Removed /opt/protegility/defiance_dps/data/ from all nodes

Copying current node's defiance_dps/data/ to all other nodes...

Changing ownership of defiance_dps_old/data_11-09-2023_13:29:27/new_data/data.tgz...

Changing permission of defiance_dps_old/data_11-09-2023_13:29:27/new_data/data.tgz...

Extracting defiance_dps_old/data_11-09-2023_13:29:27/new_data/data.tgz to defiance_dps/
data/...

Changing permission of defiance_dps/data/...

Removing backup directory defiance_dps_old/...

Starting PepServer on current node...

Starting PepServer on all nodes...

Successfully updated PepServer Configuration and Certificates across all cluster nodes

The script's logs and operation results are logged in /opt/protegility/logs/
sync_defiance_dps.log

```

12.7.7.5.4 Updating the Fluent Bit Configuration on all the Nodes

The *sync_fluent_bit.sh* script enables you to replicate the Log Forwarder configuration files (under the */fluent-bit/data/* directory) from the current node to all the remaining nodes of the cluster. The script will also restart the Log Forwarder on all the nodes after replicating the configuration.

Note: The Cluster Utilities script will only work on clusters where the *sudoer* privileges are enabled.

► To update the Fluent Bit configuration on all the nodes:

1. Navigate to the */opt/protegility/peptrino/scripts/cluster_utils* directory.
2. To view the syntax for the script, run the following command.

```
./sync_fluent_bit.sh --help
```

3. Press ENTER.

The command displays the examples, description, and the arguments. The arguments are explained in the following table.

Table 12-112: Arguments for the Fluent Bit Configuration Script

| Argument | Description |
|------------------------------|---|
| --hostsfile=</path/to/hosts> | Specifies the path to the hosts file. Each line, in the hosts file, must contain the IP address or the hostname of the remaining nodes in the Trino cluster. You can also include the user that the <i>pssh</i> |

| Argument | Description |
|--|---|
| | utility will use to connect to the node, in the following format: <i>[user@]host[:port]</i> If you fail to specify the user, then the current user running the script is used. |
| --ssh-auth-type=<password/publickey> | Specifies the SSH authentication type. The accepted values are: <ul style="list-style-type: none"> • <i>password</i> - specifies to use the password based SSH authentication. • <i>publickey</i> - specifies to use the Public Key Authentication. |
| --password=<actual_password> | Specifies the actual password of the current user to be passed to the <i>pssh</i> utility. You must use this argument only with the password-based SSH authentication. |
| --private-key-path=</path/to/privatekeyfile> | Specifies the path to the SSH private key file to be used by the <i>pssh</i> utility. You must use this argument only with the Public Key SSH authentication. |

4. To update the Fluent Bit configuration on all the nodes in the cluster, run the following command.

```
./sync_fluent_bit.sh --hostsfile=<path_of_the_hosts_file> --ssh-auth-type=publickey --  
private-key-path=<key_file_path>/<name_of_the_private_key_file>
```

5. Press ENTER.

The command updates the Fluent Bit configuration on all the nodes in the cluster.

```
=====  
Hosts file set to '<path_of_the_hosts_file>'  
SSH Authentication Type is set to 'Public Key Authentication'  
SSH Private Key file path is set to '<key_file_path>/<name_of_the_private_key_file>'  
Checking connectivity of cluster nodes...  
Trino Protector Logforwarder Configuration cloning started  
Stopping Logforwarder on current node...  
Stopping Logforwarder on all nodes...  
Creating fluent-bit_old/data_11-09-2023_13:31:33/new_data directory on all nodes...  
Changing permission of fluent-bit_old/ on all nodes...  
Removing Logforwarder Configuration from all nodes...  
Removed /opt/protegility/fluent-bit/data/ from all nodes  
Copying current node's fluent-bit/data/ to all other nodes...  
Changing ownership of fluent-bit_old/data_11-09-2023_13:31:33/new_data/data.tgz...  
Changing permission of fluent-bit_old/data_11-09-2023_13:31:33/new_data/data.tgz...  
Extracting fluent-bit_old/data_11-09-2023_13:31:33/new_data/data.tgz to fluent-bit/  
data/...  
Changing permission of fluent-bit/data/...  
Removing backup directory fluent-bit_old/...  
Starting Logforwarder on current node...  
Starting Logforwarder on all nodes...  
Successfully updated Logforwarder Configuration across all cluster nodes
```

The script's logs and operation results are logged in /opt/protegity/logs/
sync_fluent_bit.log

12.7.7.6 Executing the Uninstallation Script

You can use the uninstallation script to remove the Trino Protector.

Important: You must execute the uninstallation script on all the nodes in the cluster where the Trino Server is running.

Warning: Ensure that you do not run any Protegity Trino job while uninstalling the Trino Protector.

► To execute the uninstallation script:

1. Login to the Trino node from where you want to remove the Trino Protector node, as the *sudoer* user.

Note: By default, the uninstallation script assumes that a user account with *sudoer* privileges is running the script and will use *sudo* to execute the commands. If *sudoer* privileges are disabled on your Trino cluster, then use the *--sudo-disabled* argument.

For more information about the argument, refer to the *Optional Arguments for the Uninstallation Script* table.

2. Navigate to the */<installation_directory>/peptrino/scripts* directory.

3. To view the command usage, syntax, and the arguments for the uninstallation script, run the following command.

```
./TrinoProtector_UninstallationScript_9.1.0.0.x.sh --help
```

4. Press ENTER.

The command displays the usage, syntax, and the arguments for the uninstallation script. The mandatory and optional arguments are explained in the following tables.

Table 12-113: Mandatory Arguments for the Uninstallation Script

| Argument | Description |
|---|--|
| <i>--uninstall-pepper-and-logforwarder=<yes/no></i> | Instructs the script whether to remove the PEP server and the Log Forwarder from the current node in the Trino cluster. The acceptable values are: <ul style="list-style-type: none"> • <i>yes</i> - remove the PEP server and the Log Forwarder from the current node in the Trino cluster. • <i>no</i> - skip removing the PEP server and the Log Forwarder if they were already not installed by the corresponding Trino Protector installation script. |
| <i>--trino-plugin-dir=</path/to/plugin/></i> | Specifies the absolute path to the Trino plugin directory. <p>Note: You can set a custom plugin path using the <i>plugin.dir</i> property in the <i>node.properties</i> Trino configuration file.</p> |
| <i>--trino-service-user=<user></i> | Specifies the name of the user running the Trino server. You can use this argument to set the ownership of the <i>peptrino plugin</i> directory and also to restart the Trino server if you specify the <i>--restart-trino-server-via=launcher</i> argument. |
| <i>--delete-protegity-user=<yes/no></i> | Specifies whether to remove or retain the user and usergroup, that were created during the installation process, from all the nodes in the Trino cluster. Set the value for this argument to <i>no</i> when you set |



| Argument | Description |
|----------|---|
| | <p>the value for the <code>--uninstall-pepserver-and-logforwarder</code> argument to <code>no</code>. The accepted values are:</p> <ul style="list-style-type: none"> <code>yes</code> - instructs the script to remove the <code>ptyitusr</code> user and <code>ptyitusrgroup</code> from the current node in the cluster. <code>no</code> - instructs the script to skip the removal of Protegity service user and group. |

Table 12-114: Optional Arguments for the Uninstallation Script

| Argument | Description |
|---|---|
| <code>--sudo-disabled</code> | <ul style="list-style-type: none"> Use this flag if <code>sudoers</code> is disabled on the cluster. When you provide this argument, a user with elevated privileges is required to execute the script. When you exclude this argument, a user with <code>sudoers</code> privilege is required to execute the script. It is recommended to use a <code>NOPASSWD</code> sudoers user. Else, the script will prompt for a password. |
| <code>--restart-trino-server-via=<systemd/init/launcher></code> | <p>If you specify this argument, then the script will restart the running Trino server after removing the <code>peptrino</code> plugin. The script will not restart the Trino server if the server is in the stopped state. If you want to use this argument, ensure that you enable <code>sudo</code> to restart the Trino server. The acceptable values are:</p> <ul style="list-style-type: none"> <code>systemd</code> - instructs the script to use the <code>systemctl</code> command to check the status and restart the Trino server. This argument requires the <code>--trino-systemd-service-name</code> argument to be specified. <code>init</code> - instructs the script to use the <code>service</code> command to check the status and restart the Trino Server. This argument also requires you to specify the <code>--trino-init-service-name</code> argument. <code>launcher</code> - instructs the script to use the Trino launcher script to check the status and restart Trino Server. This argument requires the <code>--trino-launcher-path</code> and the optional <code>--trino-launcher-args</code> arguments. <p>If you exclude this argument, then the script will not attempt to restart the Trino server. You must manually restart the Trino server after the script execution completes.</p> |
| <code>--trino-systemd-service-name=<service name></code> | Specifies the name of the systemd service associated with the Trino server. You must specify this argument when you use the <code>--restart-trino-server-via=systemd</code> argument. |
| <code>--trino-init-service-name=<init service name></code> | Specifies the name of the Sys V init service associated with the Trino server. You must pass this argument when you specify the <code>--restart-trino-server-via=init</code> argument. |
| <code>--trino-launcher-path=</path/to/bin/launcher></code> | Specifies the absolute path to the Trino server launcher script. For example, <code>/usr/lib/trino/bin/launcher</code> . You must specify this argument when you use the <code>--restart-trino-server-via=launcher</code> argument. |
| <code>--trino-launcher-args="arg1 [arg2...]"</code> | Specifies the valid command line arguments to the Trino launcher script. You can use this argument with the <code>--trino-launcher-path</code> argument. If you specify this argument, then the arguments listed between the double-quotes will be passed to the Trino launcher script for the status and restart commands. If you fail to specify this argument, then no argument will be passed to the Trino launcher script for the status and restart commands. |

- To execute uninstallation script, provide the required arguments.



For example, on a Starburst Trino cluster installed via RPM, one combination of the arguments to the uninstallation script is listed below.

```
./TrinoProtector_UninstallationScript_9.1.0.0.x.sh \
--uninstall-pepservice-and-logforwarder=yes \
--trino-plugin-dir=/usr/lib/starburst/plugin \
--trino-service-user=starburst \
--delete-protegility-user=yes \
--restart-trino-server-via=systemd \
--trino-systemd-service-name=starburst
```

Note: If you want the Trino Server to automatically restart after uninstalling the components, then you must specify the value for the `--restart-trino-server-via` argument for the uninstallation script. Else, you will have to manually restart the Trino Server after the uninstallation is complete.

6. Press ENTER.

The script starts the uninstallation process based on the arguments specified.

```
./TrinoProtector_UninstallationScript_9.1.0.0.x.sh \
--uninstall-pepservice-and-logforwarder=yes \
--trino-plugin-dir=/usr/lib/starburst/plugin \
--trino-service-user=starburst \
--delete-protegility-user=yes \
--restart-trino-server-via=systemd \
--trino-systemd-service-name=starburst

Protegility Trino Protector Uninstallation Script started...

Validating sudo permissions for root
*****
        Welcome to the Trino Protector Uninstall Wizard.
*****
This will uninstall the Trino Protector from your system.

Stopping PepServer on current node...

Stopping Logforwarder on current node...

PepServer uninstallation started
*****
        Welcome to the PepServer Setup Wizard.
*****
Uninstalled PepServer on current node...

Logforwarder uninstallation started
*****
        Welcome to the LogForwarder Setup Wizard.
*****
Uninstalling LogForwarder .....

LogForwarder uninstalled on current node at location /opt/protegility/fluent-bit/

PepTrino Plugin Jars uninstallation started
*****
        Welcome to the PepTrino Setup Wizard.
*****
Uninstalling PepTrino .....

PepTrino uninstalled on current node at location /opt/protegility/peptrino/

Jpeplite uninstallation started
*****
        Welcome to the Jpeplite Setup Wizard.
*****
Uninstalling Jpeplite .....
```



```
JpepLite for Trino Protector uninstalled on current node at location /opt/protegity/
peptrino/jpeplite_lib/

Resetting ownership of /opt/protegity recursively.

Removing Protegity service user 'ptyitusr' and group 'ptyitusrgroup' from current node.

User 'ptyitusr' deleted

Group 'ptyitusrgroup' deleted

Trino Protector plugin jars and JpepLite libraries uninstalled from /opt/protegity/
peptrino/ directory

Finished executing TrinoProtectorUninstall<version>_Linux-ALL_9.1.0.0.x.sh script. Check
the logs at /opt/protegity/logs/

Started uninstallation of PepTrino from plugin dir in foreground...

Checking if Trino Plugin Dir is present on node.
Trino Plugin Directory /usr/lib/starburst/plugin found.

Removing peptrino plugin directory from /usr/lib/starburst/plugin

Trino Protector Plugin successfully removed from /usr/lib/starburst/plugin

Checking if Trino service user exists.

Checking if systemctl is on PATH.
systemctl found on PATH.

Checking if 'starburst' is a valid systemd unit.
'starburst' is a valid systemd unit.

Checking if Trino Server is started and running via systemctl
Trino server is running
Restarting Trino Server...

Trino Server successfully restarted via systemctl.

Trino Protector UDFs were unregistered on Trino Server restart. Verify via trino CLI
(show functions;) or by checking Trino's server.log

Successfully completed all steps of Uninstallation Script
```

The uninstallation script generates the logs in the `/<installation_directory>/logs/` directory.

Chapter 13

Appendix A: PEP Server Configuration File

The following text represents the PEP Server configuration file *pepservr.cfg*.

```
# Configuration file for the pepserver
#
# -----
# Application configuration
# -----
[application]

# Directory where the pepserver saves its temporary files etc.
workingdir = ./

# Directory where token elements are stored.
tokenelementdir = ./tokenelements

# Execute this program/script after the policy has been successfully updated in shared memory.
# Can be used to distribute a policy to multiple nodes/destinations.
# If nothing is set no execute is done.
#postdeploy = <path/script>

# Specifies the communication id to use. Default 0
# Teradata : Configurable.
# SQLServer : Must be set to 0.
# Oracle    : Configurable. Must match the value specified in 'createobjects.sql'
# DB2       : Configurable.
# Valid values are in the range 0 to 255.
communicationid = 0

# Add the PEP Server's IP Address to request headers.
# This is needed when the PEP Server is communicating with ESA via a proxy.
addipaddressheader = yes

# -----
# Logging configuration
# -----
# Logging level for pepserver application logs: OFF - No logging, SEVERE, WARNING, INFO, CONFIG,
ALL
level = ALL

# Set the output type for protections logs. Set to either tcp stdout or file.
output = tcp
# If output is set to file, set the filename here.
#outputfilename = /opt/logs.txt

# Fluentbit host and port values (mostly localhost) where logs will be forwarded from the
protector.
host = 127.0.0.1
port = 15780

# Fluentbit port for app/policy logs sent from pepserver
app_port = 15781

# In case that connection to the fluentbit is lost, set how logs must be handled.
# This setting is only for the protector logs and not application logs, sent from pepserver
# drop = Protector throws logs away if connection to the fluentbit is lost
```



```
# error = (default) Protector returns error without protecting/unprotecting data if connection
# to the fluentbit is lost
#mode = error

# Log data protection operation. Set to yes to log every data protect operation.
logallcallouts = no

# -----
# Policy management
# -----
[policymanagement]

# The base URL to the HubController service.
url = https://10.10.100.5:8443

# Path to the CA certificate.
cafile = ./CA.pem

# Path to the certificate.
certfile = ./cert.pem

# Path to the private key for the certificate.
certkeyfile = ./cert.key

# Path to the credential file used to decrypt the private key.
keycredentialfile = ./certkeyup.bin

# Number of seconds between checks to refresh policy from ESA.
# Specify a value in the range 30 to 86400 seconds (default is 60 seconds). If this
# value is set to be larger than 300 then the node status might not be proper on ESA.
# Some random bias will be added to this value to spread the load from multiple pep servers.
policyrefreshinterval = 60

# Define what value to return if data to protect is an empty string
# null = Return a null value (Default)
# encrypt = Return an encrypted value
# empty = Return an empty string
#emptystring = null

# -----
# Application Protector configuration
# -----
[applicationprotector]

# Listener port for Application Protector Client/Server.
#listener = tcp, 15910/127.0.0.1

# -----
# Administration configuration
# -----
[administration]

# Listener port for the administration interface.
# Only accessible on localhost.
#listener = tcp, 16700

# The URI to the authentication API.
# Base URL and certificates is taken from policymanagement section
uri = /api/v1/auth/login/checkcredentials

# -----
# Member management
# -----
[member]

# Specifies how policy users are checked against policy
# yes = (The default) policy users are treated in case sensitive manner
# no = policy users are treated in case insensitive manner.
#case-sensitive = yes
```



```
# -----
# Shared Memory management
#
# This section appears only for the DSG. For other protectors, you must add the section manually.
[sharedmemory]
groupname = dsggroup
worldreadable = no
```

The following table helps you to understand the usage of the parameters listed in the *pepperserver.cfg* configuration file.

Important: It is recommended that only the parameters listed in the following table are edited as per your requirement.

Table 13-1: Parameter Description and Usage

| Appliance/ Protectors | Section | Parameter Name | Description |
|--------------------------|---------------------------|-----------------------|---|
| All Protectors | Application configuration | <i>postdeploy</i> | Set the path of any script that must be executed after the policy is deployed. |
| | Logging configuration | <i>level</i> | Specifies the logging level set. The log level set in this parameter how the data protection logs appear in the ESA forensics. |
| | | <i>host</i> | Set the host IP of the Log Forwarder, generally localhost, where the protector will send the data protection logs. |
| | | <i>port</i> | Set the port number of the Log Forwarder, generally localhost, where the protector will send the data protection logs. |
| | | <i>output</i> | Set the to either <i>tcp</i> , <i>stdout</i> , and <i>file</i> . The default output is set as <i>tcp</i> . |
| | | <i>outputfilename</i> | If output is set to file, set the filename. |
| | | <i>mode</i> | Set how the logs must be handled in a situation where the connection to the Log Forwarder in the protector is lost. <ul style="list-style-type: none"> drop: Set to drop the logs when the connection to Log Forwarder is lost error: Set to stop the data security operations and generate an error when the connection to Log Forwarder is lost |
| | | <i>app_port</i> | Set the port number of the Log Forwarder where the protector will send the application and policy-related logs. |
| | | <i>logallcallouts</i> | Set to manage logging of data security operations. <ul style="list-style-type: none"> no: Default. Set to log first successful data security operation only. yes: Set to log every successful data security operation. <p>Note: Logging every data security operation might impact performance.</p> |
| | Policy management | <i>emptystring</i> | Defines the behavior when the data to protect is an empty string. The default value is <i>null</i> . The following are the possible values: <ul style="list-style-type: none"> empty encrypt |



| Appliance/ Protectors | Section | Parameter Name | Description |
|--------------------------|--------------------------|---|---|
| | | | <ul style="list-style-type: none"> • null (Default) <p>For more information about empty string handling by protectors, refer to the section <i>Appendix C: Empty String Handling by Protectors</i> in the <i>Protection Methods Reference Guide 9.0.0.0</i>.</p> |
| | Member management | <i>case-sensitive</i> | <p>If this parameter is set to <i>no</i>, then the PEP Server considers the policy user names that are case insensitive.</p> <p>If this parameter is set to <i>yes</i> or if it is commented in the file, then the PEP Server considers the policy user names that are case-sensitive. The default value is <i>yes</i>.</p> |
| | Shared Memory management | <p><i>groupname</i></p> <p>Note: This section is seen in the <i>pepperver.cfg</i> for DSG. For other protectors, you must add the section to the <i>pepperver.cfg</i> file.</p> <p>For more information about the Shared Memory management in Big Data Protector, refer to section <i>Updating the Configuration Parameters for the BDP PEP Service in an Open Hadoop Network</i>.</p> | <p>Set the group name. For DSG, this is set to <i>dsggroup</i>.</p> <p><i>worldreadable</i></p> <p>Set to <i>no</i> as default.</p> |
| DSG | Policy management | <p><i>shufflecodebooks</i></p> <p>Note: Enabling this parameter requires careful consideration. For more information about codebook reshuffling, refer to <i>Codebook Re-shuffling in the PEP Server</i> in the <i>Protegility Data Security Gateway User Guide 3.1.0.5</i>.</p> | <p>Set to <i>yes</i> when codebook reshuffling must be enabled. The default value is <i>no</i>.</p> <p><i>randomfile</i></p> <p>Path to the file that contains the random bytes for shuffling codebooks.</p> |
| | PKCS#11 configuration | <p><i>provider_library</i></p> <p>Important: You must edit values under this section only if <i>shufflecodebooks</i> is enabled.</p> | <p>Path to the PKCS#11 provider library.</p> <p>Note: For more information about codebook reshuffling, refer to <i>Codebook Re-shuffling in the PEP Server</i> in the <i>Protegility Data Security Gateway User Guide 3.1.0.5</i>.</p> |
| | | <i>slot</i> | <p>The slot number to use on the HSM.</p> <p>Note: For more information about codebook reshuffling, refer to <i>Codebook Re-shuffling in the PEP</i></p> |



| Appliance/ Protectors | Section | Parameter Name | Description |
|--------------------------|--|----------------------|--|
| | | | <i>Server in the Protegility Data Security Gateway User Guide 3.1.0.5.</i> |
| | | <i>userpin</i> | The scrambled user pin file. Note: For more information about codebook reshuffling, refer to <i>Codebook Re-shuffling in the PEP Server</i> in the Protegility Data Security Gateway User Guide 3.1.0.5 . |
| Application Protector | Shared Memory management Note: This section is seen in the <i>pepserver.cfg</i> for Application Protector. | <i>groupname</i> | Set the group name. For Application Protector, the group name must be the same as that of the user who is authorized to perform the data security operations. |
| | | <i>worldreadable</i> | By default, this parameter is set to <i>yes</i> , i.e., shared memory segment permissions are set to 666, which is world-readable. Set this parameter to <i>no</i> , to make it non world-readable. As a result, the permissions are changed to 660. |
| Database Protector | Shared Memory management Note: This section is seen in the <i>pepserver.cfg</i> file for Database Protector. Note: To modify the Shared Memory management settings, perform the following steps in sequence: 1. Stop the PEP server. 2. Change the Shared Memory management parameters in the <i>pepserver.cfg</i> file. For example, <code>groupname = oinstall worldreadable = Yes</code> Here, <i>oinstall</i> is the authorized group name to perform the data security operations in Database Protector. | <i>groupname</i> | Set the group name. For Database Protector, the group name must be the same as that of the user who is authorized to perform the data security operations. |
| | | <i>worldreadable</i> | By default, this parameter is set to <i>yes</i> , i.e., shared memory permissions are set to 666, which is world-readable. Set this parameter to <i>no</i> , to make it non world-readable. As a result, the permissions are changed to 660. |



| Appliance/ Protectors | Section | Parameter Name | Description |
|--------------------------|--|----------------|-------------|
| | <pre>ipcrm -S 0x000faffa ipcrm -S 0x000abba4 ipcrm -S 0x000beda0</pre> <p>4. Start the PEP server.</p> | | |

Chapter 14

Appendix: Configuring a Trusted Appliance Cluster (TAC) without Consul Integration

If you are using a cluster and do not want to continue with the *Consul Integration* services, then you can configure the cluster by uninstalling the *Consul Integration* services followed by creating the TAC.

For more information about creating a TAC, refer to the section *Trusted Appliances Cluster (TAC)* in the *Protegility Appliances Overview Guide 9.1.0.5*.

Note: If the node contains scheduled tasks associated with it, then you cannot uninstall the cluster services on it. Ensure that you delete all the scheduled tasks before uninstalling the cluster services.

Note: If you are uninstalling the *Consul Integration* services, then the *Consul* related ports and certificates are not required.

To uninstall cluster services, perform the following steps.

1. Remove the appliance from the TAC.
2. In the CLI Manager, navigate to **Administration > Add/Remove Services**.
3. Press **ENTER**.
4. Select **Remove already installed applications**.
5. Select **Cluster-Consul-Integration v0.2** and select **OK**.
The integration service is uninstalled.
6. Select **Consul v1.0** and select **OK**.
The Consul product is uninstalled from your appliance.

After the *Consul Integration* is successfully uninstalled, then the Cluster labels, such as, *Consul-Client* and *Consul-Server* are not available.

To manage the communication between various nodes in a TAC, you can use the communication blocking mechanism.

For more information about the communication blocking mechanism, refer to the section *Connection Settings* in the *Protegility Appliances Overview Guide 9.1.0.5*.

Chapter 15

Appendix: Verifying the Immutable Service on the ESA to Export Policy for Immutable Protectors

For any immutable protector to be able to export policies from an ESA, the Immutable service must be running on the ESA.

Verify that the IMP service is installed on the ESA by performing the following steps.

1. Login to the ESA Web UI.
2. Navigate to **System > Services**.
3. Verify that the IMP service is available and running under the **Misc** area.

| Misc | | | |
|---------------------|---------|-----------|--|
| Web-Services Engine | Running | Automatic | |
| Service Dispatcher | Running | Automatic | |
| td-agent | Running | Automatic | |
| Analytics | Running | Automatic | |
| IMPS | Running | Automatic | |

Figure 15-1: IMP Service

Chapter 16

Appendix B: Using Go Module with Private GitLab Repository

This section describes the steps to use the Go module with a private GitLab Repository.

To set up the Go module in a private GitLab repository:

1. Create a GitLab Personal Access Token with at least *read_api*, *read_repository*, and *write_repository* scopes.

For more information about creating a personal access token, refer to the section [Creating a personal access token](#) in the GitLab documentation.

2. Create a *.netrc* file and place it in your home directory.

```
machine privaterepo.example.com
  login user.name@example.com
  password <PERSONAL_ACCESS_TOKEN>
```

3. Run the following Go environment command:

```
go env -w GOPRIVATE=privaterepo.example.com/app/*
```

Chapter 17

Appendix: Configuring the IP address for the Docker Interface

Network settings allows you to configure the network details for the appliance, such as, host name, default gateway, name servers, and so on.

From ESA v9.0.0.0, the default IP addresses assigned to the docker interfaces are between *172.17.0.0/16* and *172.18.0.0/16*. If your have a VPN or your organization's network configured with the IP addresses are between *172.17.0.0/16* and *172.18.0.0/16*, then this might cause conflict with your organization's private or internal network resulting in loss of network connectivity.

Note:

Ensure that the IP addresses assigned to the docker interface must not conflict with the organization's private or internal network.

In such a case you can reconfigure the IP addresses for the docker interface by performing the following steps.

To configure the IP address of the docker interfaces:

1. Leave the docker swarm using the following command.

```
docker swarm leave --force
```

2. Remove *docker_gwbridge* network using the following command.

```
docker network rm docker_gwbridge
```

3. In the */etc/docker/daemon.json* file, enter the non-conflicting IP address range.

Note:

You must separate the entries in the *daemon.json* file using a *comma (,)*. Before adding new entries, ensure that the existing entries are separated by a *comma (,)*. Also, ensure that the entries are enlisted in the correct as shown in the following example .

```
"bip": "10.200.0.1/24",
"default-address-pools":
[
  { "base": "10.201.0.0/16", "size": 24}
]
```

Warning:

If the the entries in the file are not mentioned in the format specified in step 3, then the restart operation for docker service fails.

4. Restart the docker service using the following command.

```
/etc/init.d/docker restart
```

The docker service is restarted successfully.

5. Check the status of the docker service using the following command.

```
/etc/init.d/docker status
```

6. Initialize docker swarm using the following command.

```
docker swarm init --advertise-addr=ethMNG --listen-addr=ethMNG --data-path-addr=ethMNG
```

The IP address of the docker interfaces are changed successfully.

Chapter 18

Audit Store Performance Analysis

The Audit Store cluster saves the logs on a single node or multiple nodes according to the cluster setup. Accordingly, you can have a single ESA or multiple ESAs in the Audit Store cluster. Use the information provided in this section to analyze and determine the number of ESAs that you require for your Audit Store cluster.

Note: The values displayed here are obtained during performance testing. The actual number of logs that the cluster can hold might vary. Use these values as an approximate amount for analyzing your requirements.

The average size of a log generated by a single ESA machine with the recommended system requirements and no Protectors is approximately *300 bytes*. The following table lists the additional number of logs that the Audit Store cluster can hold when nodes are added.

| Nodes | Additional Number of Logs Per Node |
|-------------------|------------------------------------|
| 1 | 300 million |
| 2 - 3 | 1.5 billion |
| 4 | 2.5 billion |
| 5 nodes and above | 2 billion |

Consider the following examples:

- The Audit Store cluster consisting of a single ESA can hold approximately *300 million* logs.
- The Audit Store cluster consisting of three ESAs can hold approximately *3.5 billion* logs.
- The Audit Store cluster consisting of seven ESAs can hold approximately *12 billion* logs.

Similarly, you can estimate the approximate logs that your Audit Store cluster can hold. You can add or remove nodes from the cluster according to the number of logs that you require in the Audit Store cluster.

The maximum number of logs determines the number of logs the Audit Store cluster can hold without any issues. When the number of logs cross the maximum limit, then the Audit Store performance begins to deteriorate and there might be a loss of logs.

It is recommended to archive the old logs or add more ESA nodes before the total log count crosses the maximum limit on the Audit Store cluster.

For more information about archiving logs, refer to the section *Exporting Logs* in the [Protegility Analytics Guide 9.1.0.5](#).

For more information about adding an ESA to the Audit Store cluster, refer to the section *Adding an ESA to the Audit Store Cluster* in the [Protegility Installation Guide 9.1.0.5](#).

For more information about removing an ESA from the Audit Store cluster, refer to the section *Removing an ESA from the Audit Store Cluster* in the [Protegility Installation Guide 9.1.0.5](#).

