IBM InfoSphere Optim

Archive Viewer Installation and Configuration Guide

Version 11 Release 7

IBM InfoSphere Optim

Archive Viewer Installation and Configuration Guide

Version 11 Release 7

Note:
Before using this information and the product it supports, read the information on the next page.

This edition applies to Version 11, Release 7 of IBM InfoSphere Optim Archive Viewer and to all subsequent releases and modifications until otherwise indicated in new editions.

IBM InfoSphere Optim Archive Viewer
Installation and Configuration Guide

IBM InfoSphere Optim Archive Viewer,
Version 11 Release 7

Notices

This information was developed for products and services offered in the USA.

IBM® may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service.

IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not grant you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 10504-1785
USA

For license inquiries regarding double-byte (DBCS) information, contact the IBM Intellectual Property Department in your country or send inquiries, in writing, to:

Intellectual Property Licensing
Legal and Intellectual Property Law
IBM Japan Ltd.
1623-14, Shimotsuruma, Yamato-shi
Kanagawa 242-8502 Japan

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law:

INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you.

This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk.

IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you.

Such information may be available, subject to appropriate terms and conditions, including in some cases, payment of a fee.

The licensed program described in this information and all licensed material available for it are provided by IBM under terms of the IBM Customer Agreement, IBM International Program License Agreement, or any equivalent agreement between us.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the

names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Trademarks

IBM, the IBM logo, ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the web at “Copyright and trademark information” at www.ibm.com/legal/copytrade.shtml

Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.

Microsoft, Windows, and Windows NT are trademarks of Microsoft Corporation in the United States, other countries, or both.

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

Contents

1. Introduction to the Optim Archive Viewer

IBM® InfoSphere® Optim™ Archive Viewer allows users to access data from Optim Archive Files via open (JDBC/ODBC) technologies. It provides automation to unload the data from the proprietary Archive File format to highly compressed open data files, and then provides high performance access with no performance tuning necessary, similar to standard industry columnar relational databases.

This guide describes how to install and configure the IBM InfoSphere Optim Archive Viewer application.

To install or use the Archive Viewer application, you must have the Optim Convert Archive File to CSV utility, which is available in the Archive File Access, the Archive and Data Growth Solution and the Solution for Application Retirement.

1.1 Components of the Archive Viewer

The Archive Viewer application is used to access data from Optim Archive Files via open (JDBC/ODBC) technologies.

The various components required for the Archive Viewer to function are listed below:

- Archive Viewer metadata database – This database is a set of tables in which the Archive Viewer stores the data that are needed for processing and tracking the status of the processes. These tables are created when the user completes the first step of the Onboarding process. So, before you use Archive Viewer, you must follow the Onboarding process to create or configure the Archive Viewer metadata database tables that are needed to access the Archive Viewer metadata database. For more details, see **Section 3. Onboarding**.
  - Supported databases are:

    Oracle, SQL Server, DB2® for Linux® and Windows, PostgreSQL, MySQL and SQLite

- Archive Viewer Conversion Server – This server unloads the data from the proprietary Archive File format to highly compressed open data files.
  - You can install the Archive Viewer Conversion Server where the Optim server is located or on a separate virtual machine (VM) where new Optim Server can be installed with Archive Viewer.
  - You can use an existing Optim directory or create a new Optim directory to store the tablemaps and convert requests that will be created by the Archive Viewer.

  - You can deploy the Archive Viewer Conversion Server in both Windows and Linux platforms.

  - If you deploy the Archive Viewer Conversion Server on Windows, you should add a shared network drive consisting of Parquet files between Archive Viewer Conversion Server and Archive Viewer Query Server deployed on Linux.

- Archive Viewer Query Server – This server leverages Apache Spark, which is an open-source unified analytics engine for large-scale data processing. It provides high performance access with no performance tuning necessary as in the standard industry columnar relational databases.

  - You must deploy the Archive Viewer Query Server in the Linux platform.

- Archive Viewer Reporting Server (optional) – This server leverages Apache Superset, which is a modern data exploration and visualization platform.

  - You must deploy the Archive Viewer Reporting Server in a Linux platform.

- All the above components can be installed on a single Linux platform.

Minimum Configuration

The minimum configuration and/or consolidated VM details for Conversion Server, Query Server and Reporting Server are listed in the below table.

Server	Configuration Details
Conversion Sever	Windows/Linux 4 core CPU 8 GB RAM 200 GB HDD
Query Server	Linux 4 core CPU 8GB RAM 4-8 disks per node Configured without RAID (just as separate mount points)
Reporting Server (optional)	Linux 2 core CPU 8 GB RAM 40 GB HDD

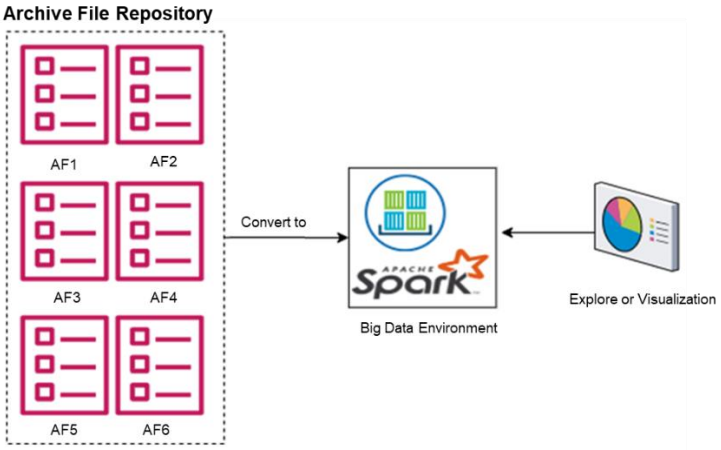
1.1.1 Architecture of the Archive Viewer

You can find the Optim Convert Archive File to CSV utility in the Archive File Access, the Archive and Data Growth Solution and the Solution for Application Retirement.

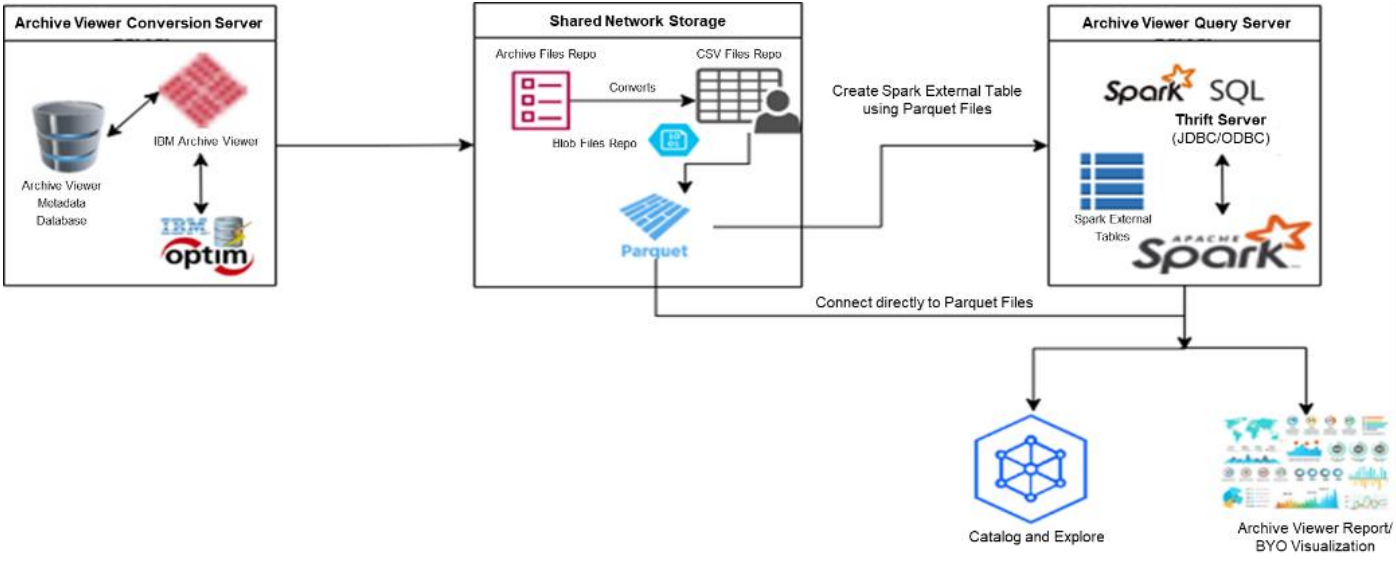
The Archive Viewer metadata database is used to store the objects that are needed for processing and tracking the status of processes.

A typical and a detailed architecture of the Archive Viewer are shown below.

Typical Architecture of the Optim Archive Viewer



Detailed Architecture of the Optim Archive Viewer



Prerequisites for the supported databases

DB2 for Linux, Unix and Windows

For DB2, you must install ibm_db_sa==0.3.8 to connect local and remote DB2 instances by running the below command.

pip install ibm_db_sa==0.3.8

It is recommended to create a new user and a database and make the user as the database owner with DBADM, IMPLICIT_SCHEMA, BINDADD, CREATE_NOT_FENCED_ROUTINE privileges. Make sure that the DB2 page size is set to 8k.

In Windows, follow the steps given if you encounter the below error:

```
Connection Refused due to
ibm_db_dbi::ProgrammingError: [IBM][CLI Driver]
SQL1042C  An unexpected system error occurred.
SQLSTATE=58004  SQLCODE=-1042"
```

1. Run the below command to find the location of your site-packages directory.

   pip show ibm_db_sa

   For example, the directory is C:\python\lib\site-packages.

2. Modify the PATH environment variable to append the following directory to the PATH:

   C:\python\lib\site-packages\clidriver\bin\amd64.VC12.CRT

   Note: If you do not have permissions to modify the PATH, you can copy the contents (2 .dll files) of the amd64.VC12.CRT directory into the directory, where the process takes place when python runs.

3. After modifying the PATH, test the database connection. The error message will not occur.

   For more details about this issue, see the following site:

   https://github.com/ibmdb/python-ibmdb/issues/599

SQL Server

For SQL Server, the python and the ODBC system libraries are required.

For instructions to install and configure the ODBC library, see the Download ODBC Driver for SQL Server section at the following site:

https://learn.microsoft.com/en-us/sql/connect/odbc/download-odbc-driver-for-sql-server?view=sql-server-ver16

Note: The Archive Viewer application supports Microsoft ODBC Driver 18 for SQL Server only.

Run the below command to install the python library, pyodbc:

pip install pyodbc==4.0.34

After the installation and configuration, you should create a user, the member of db_datareader, bd_datawriter and db_ddladmin.

MySQL

For MySQL, the system library is not required, however, the application connects to the DB with pymysql.

Run the below command to install the DB:

pip install pymysql==1.0.2

It is recommended to create a new user and a database and make the user as the database owner.

Oracle

For Oracle database, you should use **cx_oracle**, which is required for oracle thin client libraries. Alternatively, you can use the **python-oracledb** package, which does not require client installation.

Run the below command to install this package:

pip install oracledb==1.1.1

It is recommended to create different users with CREATE TABLE, CREATE SEQUENCE, UNLIMITED TABLESPACE and CREATE SESSION privileges.

PostgreSQL

For PostgreSQL, the installation of any system library is not required because pip package, psycopg2-binary==2.9.3, connects to DB without an external DB thin client.

Run the below command to install this package:

pip install psycopg2-binary==2.9.3

It is recommended to create a new user and a database and make the user as the database owner.

Run the series of commands using postgres admin user, as follows:

1. Create a database.

   Create database <DbName>

2. Create a user with the following details.

   CREATE ROLE <UserName> WITH
   LOGIN
   NOSUPERUSER
   NOCREATEDB
   NOCREATEROLE
   NOINHERIT
   NOREPLICATION

CONNECTION LIMIT 10
PASSWORD 'passwd';

3. Change the database owner from an existing user to the new user.

   ALTER DATABASE <DbName> OWNER TO <UserName>;

4. Switch to the user.

   create schema <UserName>;

SQLite

The SQLlite driver is in-built in pyhton3 and hence, the installation of any pip library is not required.

If the hosting server is Linux and python3.9 must be installed manually, then ensure that the sqlite-devel package is installed before installing python3.9.

The supported versions and required privileges for these databases are listed in the below table.

Database	Version	Privileges
Oracle	12c 18c 19c 21c	CREATE TABLE CREATE SEQUENCE UNLIMITED TABLESPACE CREATE SESSION
MySQL	8.0.31	DB owner
SQL Server	2017 2019 2022	Schema owner Member of db_datareader, bd_datawriter and db_ddladmin

Database	Version	Privileges
		Not supported for Windows Authentication
Postgres	15 14	DB owner and a schema with the same name as user
DB2	DB2® for Linux and Windows, Version 9.1 Fix Pack 2 and later	DBADM IMPLICIT_SCHEMA BINDADD CREATE_NOT_FENCED_ROUTINE
SQLite	3	NA

1.1.2 Deployment Model for the Archive Viewer

The single server setup of the Optim Archive Viewer should have the minimum configuration. For more details, see **Minimum Configuration**.

Here is an example of a single-server setup.

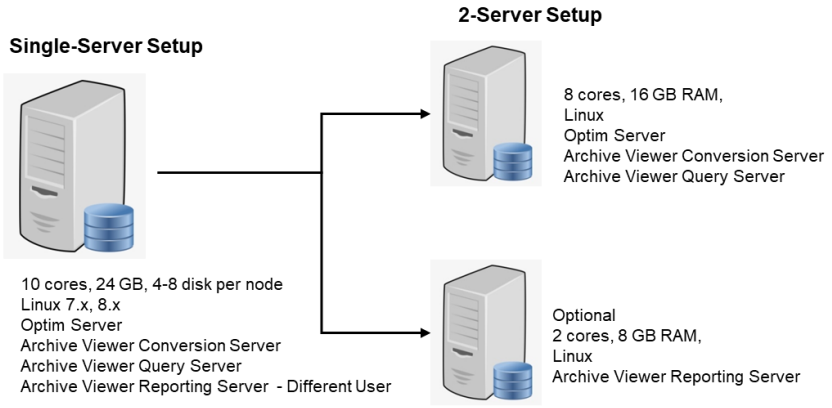
Deployment Model for Archive Viewer (single-server setup)

10 cores, 24 GB, 4-8 disks per node
Linux 7.x, 8.x
Optim Server
Archive Viewer Conversion Server
Archive Viewer Query Server

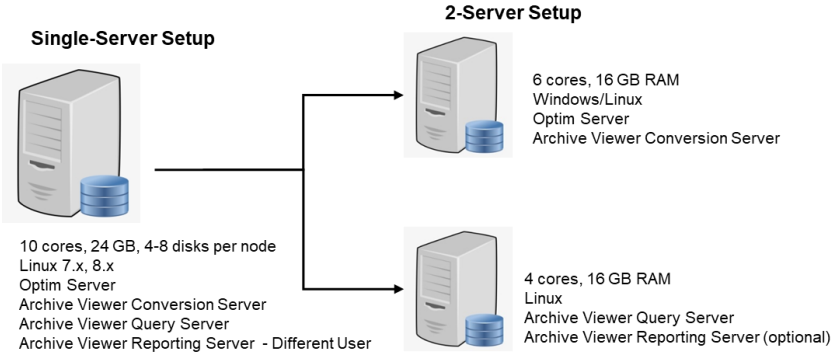
Archive Viewer Reporting Server - Different User

You can split the single server setup into 2-server setup models. The deployment models for 2-server setups are shown below.

Deployment Model for Archive Viewer (2-server setup - 1)



Deployment model for Archive Viewer (2-server setup - 2)



1.2 Hardware and Software Requirements for Archive Viewer Conversion Server

This section describes the hardware and software requirements for the Archive Viewer Conversion Server.

1.2.1 Hardware Requirements for Archive Viewer Conversion Server

If you deploy the Archive Viewer Conversion Server on the Windows system, you must deploy both Query and Reporting servers in Linux. You should store the Apache Parquet file in the shared drive.

You should mount a drive from Windows to Linux where the query server is installed and all the Parquet files are located.

Hardware requirements is listed below:

- 16 GB RAM or more
- 4-8 disks per node, configured without RAID
- Four or more cores per machine

1.2.2 Software Requirements for Archive Viewer Conversion Server

Linux

You must install the system library before installing the Archive Viewer application. You can install the system library using the yum, dnf, or apt-get utility depending on the distribution of Linux.

You must also install Python 3.9.x before installing the Archive Viewer application. It is recommended to use Python 3.9.13. You should also update the pip package before running the installation by running the below command:

python3.9 -m pip install --upgrade pip

For RedHat 7:

Library	Repository Name
gcc	rhel-7-server-rpms
gcc-c++	rhel-7-server-rpms
libffi-devel	rhel-7-server-rpms
python3-devel	rhel-7-server-optional-rpms
python3-pip	rhel-7-server-rpms
python3-wheel	rhel-7-server-rpms
openssl-devel	rhel-7-server-rpms
openldap-devel	rhel-7-server-rpms
nginx	epel

For RedHat 8:

Library	Repository Name
gcc	rhel-8-for-x86_64-appstream-rpms
gcc-c++	rhel-8-for-x86_64-appstream-rpms
libffi-devel	rhel-8-for-x86_64-appstream-rpms
python39-devel	rhel-8-for-x86_64-appstream-rpms
python39-pip	rhel-8-for-x86_64-appstream-rpms
python39-wheel	rhel-8-for-x86_64-appstream-rpms
openssl-devel	rhel-8-for-x86_64-baseos-rpms
openldap-devel	rhel-8-for-x86_64-baseos-rpms
nginx	rhel-8-for-x86_64-appstream-rpms

Note: You should set the environment variable, JAVA_HOME, properly as PySpark uses Java 8 internally.

Windows

You must install Python 3.9.x before installing the Archive Viewer application. It is recommended to use Python 3.9.13. You should also update the pip package before running the installation by running the below command:

python -m pip install --upgrade pip

You should have the native libraries for PySpark when installing the Archive Viewer on Windows.

The library should include accessing the file://filesystem, whereas Hadoop uses some Windows APIs to implement POSIX-like file

access permissions. This is implemented in hadoop.dll and winutils.exe files. The path **%HADOOP_HOME%\BIN\WINUTILS.EXE** must be searchable or executable.

1. Download the winutil package.

   To download, follow step 6 in the Learn how to build your .NET for Apache Spark application on Windows section at the following site:

   https://learn.microsoft.com/en-us/dotnet/spark/how-to-guides/windows-instructions

   For more details, see https://cwiki.apache.org/confluence/display/HADOOP2/WindowsProblems and https://github.com/cdarlint/winutils.

2. Set **HADOOP_HOME** to the appropriate Hadoop version and add %HADOOP_HOME%\bin to **PATH**.

3. Verify that the Python39\Scripts directory is available in **PATH**.

   Here is an example of the Python39 scripts directory.

   C:\Users\<username>\AppData\Local\Programs\Python\Python39\Scripts\

4. Set **PYSPARK_PYTHON** to the directory location where the python.exe exist.

   For Python 3.9.x, set the following values:

   - **HADOOP_HOME** = <Copied folder location>\winutils-master\hadoop-3.2.2
   - **PATH** = %HADOOP_HOME%\bin

- **PYSPARK_PYTHON** = <Python3.9.x installed location> \pyhton.exe

Note: You should set the environment variable, JAVA_HOME, properly as PySpark uses Java 8 internally.

Pip Dependency Library

You should have the python library installed to run the Archive Viewer application.

You can install the python libraries using the requirements.txt file during application onboarding.

The python-pip commands and the library are listed below:

- bs4==0.0.1
- crc32c==2.3
- fastapi==0.87.0
- gunicorn==20.1.0
- impyla==0.18.0
- PyJWT==2.4.0
- jwt==1.3.1
- ldap3==2.9.1
- lxml==4.9.1
- passlib==1.7.4
- pyspark==3.3.1
- SQLAlchemy==1.4.44

- thrift-sasl==0.4.3
- uvicorn==0.20.0

1.3 Hardware and Software Requirements for Archive Viewer Query Server

This section describes the hardware and software requirements for the Archive Viewer Query Server.

1.3.1 Hardware Requirements for Archive Viewer Query Server

Storage Systems

It is important to place the Query Server as close as possible to the external storage systems because most of the query jobs should read input data from an external storage system.

It is recommended that:

- If the Query Server can be placed as close as possible, run the query server installer on the same nodes as HDFS. It is easier to set up a Query Server standalone mode cluster on the same nodes, configure it and use Hadoop's memory and CPU to avoid interference.

  For more details on standalone mode cluster, see the below site:

  https://spark.apache.org/docs/latest/spark-standalone.html

- If not, run the query server installer on different nodes in the local area network.

Local Disks

While the Query Server uses its memory to perform most of the computation, it also uses local disks to store data that is not acceptable in RAM and to preserve intermediate output between stages.

It is recommended to have 4-8 disks per node, configured without RAID (as separate mount points).

In Linux, you should mount the disks with the **noatime** option to reduce unnecessary writes.

For more details on configuration, see the below site:

https://spark.apache.org/docs/latest/configuration.html

The **local.dir** variable lists the local disks separated by a comma.

Memory

In general, the Query Server runs effectively if the memory per machine is 8 GB through hundreds of GBs.

In all cases, it is recommended to allocate 75% of the memory for the Query Server and leave the rest for the operating system and buffer cache.

The amount of memory required is dependent on your application. You can determine the memory size by loading part of the dataset in a Query Server RDD and clicking the **Storage** tab of the Query Server's monitoring UI (http://<driver-node>:4040).

Note: The storage level and serialization format affect the memory usage. For more details on tuning, see the below site:
https://spark.apache.org/docs/latest/tuning.html

Java VMs run effectively if the RAM is less than 200 GB. If you have machines with more than 200 GB RAM, you can launch multiple executors in a single node.

In Query Server's standalone mode, the user is responsible for launching multiple executors according to the available memory and cores, and each executor is launched in a separate Java VM.

Network

When the data is in memory, many Query Server applications are network-bound. It is recommended to use a network with 10 Gigabit or higher to run these applications effectively. This is especially true for "distributed reduce" applications such as group-bys, reduce-bys,

and SQL joins. In any given application, you can see how much data a Query Server shuffles across the network from the application's monitoring UI (http://<driver-node>:4040).

CPU Cores

Query Server can contain tens of CPU cores per machine because it performs minimal sharing between threads. You should set up at least 8-16 cores per machine. More cores are necessary depending on the CPU cost of your workload. Once the data is in memory, most applications are either CPU- or network-bound.

For the hardware requirements related to spark, see the Hardware Provisioning section at the following site:

https://spark.apache.org/docs/3.3.0/hardware-provisioning.html

1.3.2 Software Requirements for Archive Viewer Query Server

Linux

You can deploy the Archive Viewer Query Server only on a Linux platform. You must have Java 8 installed and set **JAVA_HOME** properly.

1.4 Hardware and Software Requirements for Archive Viewer Reporting Server

This section describes the hardware and software requirements for the Archive Viewer Reporting Server.

1.4.1 Hardware Requirements for Archive Viewer Reporting Server

It is recommended to have at least 8 GB RAM and 40 GB disk space for the Archive Viewer Reporting Server. You can increase the sizes based on the report and query performances.

1.4.2 Software Requirements for Archive Viewer Reporting Server

The Archive Viewer Reporting Server is used to query the data from Parquet files. This is optional and you can install only on the Linux platform.

You must install the Reporting Server in a different server or different user than that of the Conversion Server. Otherwise, there will be a conflict on the python library of the Conversion and Reporting servers.

The system library of the Reporting Server when it is installed in a different server, as follows:

Library	Repository Name
gcc	rhel-7-server-rpms
gcc-c++	rhel-7-server-rpms
libffi-devel	rhel-7-server-rpms
python3-devel	rhel-7-server-optional-rpms
python3-pip	rhel-7-server-rpms
python3-wheel	rhel-7-server-rpms
openssl-devel	rhel-7-server-rpms
openldap-devel	rhel-7-server-rpms
xz-devel	Rhel-7-server-rpms
nginx	epel

You should install the below listed pip library to query the Parquet file. These libraries are installed before installing the Reporting Server as it connects the Reporting Sever to the thrift server.

- PyHive==0.6.5
- sasl==0.3.1
- thrift-sasl==0.4.3
- requests==2.28.1
- sqlalchemy_trino==0.5.0
- markupsafe==2.0.1
- Werkzeug==2.0.3

If the Conversion and Query servers are installed on different servers, you should have a shared network disk between the Conversion and Query servers and then store the Parquet files in the shared network location.

If the Archive Viewer Conversion Server is deployed in Windows, then you should add a shared network drive between the Conversion and Query servers.

2. Installing the Archive Viewer

The Archive Viewer application consists of the components based on the requirement and library dependency, as follows:

1. Archive Viewer Conversion Server
2. Archive Viewer Query Server
3. Archive Viewer Reporting Server (optional)

This chapter explains how to install the Conversion, Query and Reporting servers.

You can install the Conversion Server in a Windows or Linux platform. The Conversion Server is the core component of the Archive Viewer application, which is used to manage, monitor and track all the processes. See **Section 2.1 Installing the Archive Viewer Conversion Server** for guidelines on installing the Conversion Server.

You should install the Query and Reporting servers only in a Linux platform. See **Section 2.2 Installing the Archive Viewer Query Server** and **Section 2.3 Installing the Archive Viewer Reporting Server** for guidelines on installing the Query and Reporting servers.

2.1 Installing the Archive Viewer Conversion Server

Prerequisite: Before you install the Conversion Server, you must ensure that the hardware and software requirements are met. See **Section 1.2 Hardware and Software Requirements for Archive Viewer Conversion Server** for more details.

This section describes:

- **Installing the Conversion Server on Linux platform**
- **Installing the Conversion Server on Windows platform**

2.1.1 Installing the Conversion Server on Linux platform

Do the following to install the Conversion Server on Linux:

1. Copy the software on Linux server.
2. Navigate to the …/installation/linux folder.
3. Navigate to the Linux folder.
4. Run the command **python3 appInstall.py**.

   Note: You should use the appropriate python version if there are multiple python3 variants, that is, **python3.9 queryserverInstall.py**.

   The **appInstall.py** script validates for the following:

   - Validates if the user has sudo access or not.

     Note: If a user has the sudo access, the **appInstall.py** script installs the missing libraries and configures the nginx and gunicorn services.

   - Validates the platform used.
   - Checks the system-dependent libraries and lists the missing libraries.
   - Checks the pip libraries.
5. When prompted, type the directory where the application should be installed and press Enter.

   If you press Enter without providing the directory, system uses the default location.

/home/user/IBM/InfoSphere/Optim/archiveviewer

6. Type **Yes** if you have procured the SSL certification for the server, otherwise, type **No** and press Enter.

   - If you choose **No**, then do the following to create the SSL certification for your server:

     1. Enter the 2-letter country code, for example, IN for India, and press Enter.

     2. Enter the state or province name and press Enter.

     3. Enter the locality name and press Enter.

     4. Enter the organization name and press Enter.

        This creates the self-signed certificate and configures the gunicorn and nginx services automatically.

   - If you choose **Yes**, then do the following:

     1. Type the SSL certification file directory along with the file name and press Enter.

     2. Type the SSL certification private key file directory along the file name and press Enter.

   This configures the gunicorn and nginx services and This completes the Conversion Server installation on a Linux platform.

2.1.2 Installing the Conversion Server on Windows platform

Prerequisite: You must have procured the SSL certification for the server or you can also use the self-signed certificate. To create the self-signed certificate, the system should have the OpenSSL installed.

For more information on how to create self-signed certificate, see the Generating a self-signed certificate using OpenSSL section, steps 1 and 2 only, at the following site:

https://www.ibm.com/docs/en/api-connect/5.0.x?topic=profiles-generating-self-signed-certificate-using-openssl

Do the following to install the Conversion Server on Linux:

1. Copy the software on Windows server.
2. Download the stable version of nginx from the below site.

   http://nginx.org/en/download.html

3. Navigate to the …\installation\windows folder path.
4. Run the command **python appInstall.py**.

   The **appInstall.py** script validates the dependent python library. It allows user select either to install via script or skip the installation to install the dependent libraries manually.

5. When prompted, type the SSL certification file directory along with the file name and press Enter.
6. When prompted, type the SSL certification private key file directory along the file name and press Enter.
7. When prompted, type the directory where the nginx.exe file is located.

8. When prompted, type the directory where the application should be installed and press Enter.

   If you press Enter without providing the directory, system uses the default location.

   C:\IBM\InfoSphere\Optim\archiveviewer

This copies the application, copies the log folder location and configures nginx service. The scripts folder is also created.

This completes the Conversion Server installation on a Windows platform.

2.2 Installing the Archive Viewer Query Server

Prerequisite: Before you install the Query Server, you must ensure that the hardware and software requirements are met. See **Section 1.3 Hardware and Software Requirements for Archive Viewer Query Server** for more details.

Query Server is used to create a table for the Parquet files that are created by the Conversion Server. This helps users to connect via ODBC/JDBC and query the data by using SQL.

Do the following to install the Archive Viewer Query Server:

1. Copy the software in a Linux platform.
2. Navigate to the …/installation/linux folder.
3. Run the command **python3 queryserverInstall.py**.

   Note: You should use the appropriate python version if there are multiple python3 variants, that is, **python3.9 queryserverInstall.py**.

   The **queryserverInstall.py** script validates for the following:

   - Validates if the user has sudo access or not.
   - Validates the platform used.

   Note: If a user has the sudo access, the **queryserverInstall.py** script installs the missing libraries.

4. When prompted, type **Yes** to install Query Server and press Enter.
5. Type the directory where the query server should be installed and press Enter.

This downloads the query server installer package from Website, extracts the files and installs the Query Server.

You should have a shared path between Archive Viewer Conversion Server and Archive Viewer Query Server when these servers are installed on different platforms.

You can secure the query server by adding LDAP authentication, that is, by copying the hive-site.xml file to the $SPARK_HOME/conf folder.

Here is an example of the hive-site.xml file:

```
<configuration>
    <property>
        <name>hive.server2.authentication</name>
        <value>LDAP</value>
        <description>
        Expects one of [nosasl, none, ldap, kerberos, pam,
        custom].
        Client authentication types.
            NONE: no authentication check
            LDAP: LDAP/AD based authentication
            KERBEROS: Kerberos/GSSAPI authentication
            CUSTOM: Custom authentication provider
            (Use with property
            hive.server2.custom.authentication.class)
            PAM: Pluggable authentication module
            NOSASL: Raw transport
        </description>
    </property>
    <property>
        <name>hive.server2.authentication.ldap.url</name>
        <value>ldap://hostname:port</value>
        <description>
        LDAP connection URL(s),
        this value could contain URLs to mutiple LDAP servers
```

```
            instances for HA, each LDAP URL is separated by a
            SPACE character. URLs are used in the order specified
            until a connection is successful.
            </description>
        </property>
        <property>
            <name>hive.server2.authentication.ldap.baseDN</name>
            <value>ou=users,ou=system</value>
            <description>LDAP base DN</description>
        </property>
        <property>
            <name>hive.server2.authentication.ldap.groupDNPattern</
            name>
            <value>uid=%s,ou=users,ou=system</value>
            <description> COLON-separated list of patterns to use to
            find DNs for group entities in this directory. Use %s where
            the actual group name is to be substituted for. For
            example:
            CN=%s,CN=Groups,DC=subdomain,DC=domain,DC=com.
            </description>
        </property>
</configuration>
```

For more information, see the Data Access section at the following site:

https://docs.cloudera.com/HDPDocuments/HDP2/HDP-2.6.1/bk_data-access/content/ch02s05s02.html

After copying the hive-site.xml file, navigate to the the directory where the application is installed and edit the properties.json file by changing auth_machanism from **PLAIN** to **LDAP** under queryserver.

2.3 Installing the Archive Viewer Reporting Server

Prerequisite: Before you install the Reporting Server, you must ensure that the hardware and software requirements are met. See **Section 1.4 Hardware and Software Requirements for Archive Viewer Reporting Server** for more details.

You should install the Reporting Server in a different server or login as a different user to avoid conflicts between the application and Reporting Server libraries.

For installing and configuring the Archive Viewer Reporting Server, see the Installing Superset from Scratch section at the following site:

https://superset.apache.org/docs/installation/installing-superset-from-scratch

2.4 Post-installation Steps

Linux

After the successful installation of the application, you can stop, start or restart the services using the systemctl command.

The gunicorn and nginx services should be up and running for the application to function.

The commands to start, check status, stop or restart the services are listed in the below table:

Action	Commands
start	sudo systemctl start gunicorn sudo systemctl start nginx
check status	sudo systemctl status gunicorn sudo systemctl status nginx
stop	sudo systemctl stop gunicorn sudo systemctl stop nginx
restart	sudo systemctl restart gunicorn sudo systemctl restart nginx

The logs for nginx are stored under /var/log/nginx/error.log and access logs are stored under /var/log/nginx/access.log.

The logs for gunicorn are stored under archiveviewer_logs, which is located outside the application installation location.

For example, if the application location is, /home/linux/IBM/InfoSphere/Optim/archiveviewer/

then the logs are stored in the path:

/home/linux/IBM/InfoSphere/Optim/archiveviewer_log/gunicorn

The property file should be stored under the application installation location, as shown below:

/home/linux/IBM/InfoSphere/Optim/archiveviewer/properties/version

Windows

The logs are stored under archiveviewer_logs, which is located outside the application installation location.

For example, if the application location is, \IBM\InfoSphere\Optim\archiveviewer\

then the logs are stored in the path:

\IBM\InfoSphere\Optim\archiveviewer_logs

The property file should be stored under the application installation location, as shown below:

\IBM\InfoSphere\Optim\archiveviewer\properties\version

All the process scripts are located in the scripts folder that was created during the installation.

The available scripts are listed in the below table:

Scripts	Commands
startNginx.cmd	This script starts the nginx service in terminal mode.
startUvicorn.cmd	This script starts the uvicorn server in terminal mode. The uvicorn service will be terminated if you close the window.

Scripts	Commands
runNginx.cmd	This script allows user to run the nginx service in the background.
runUvicorn.cmd	This script runs the uvicorn server in the background.
stopServices.cmd	This script terminates both nginx and uvicorn services.

2.5 Uninstalling the Archive Viewer

Prerequisite: Before uninstalling the Archive Viewer application, you must have stopped the services.

Linux

Do the following to uninstall the Archive Viewer application:

1. Navigate to the …/installation/linux/ folder.
2. Run the below command:

   python3 uninstallApp.py

   Note: You should use the appropriate python version if there are multiple python3 variants, that is, **python3.9 queryserverInstall.py**.

3. When prompted, select either of the following options:

   - **Remove only application** to remove the application folder
   - **Remove Application and Pip library** to remove the application and pip libraries that were used for this application
   - **Remove Application, Pip library and Metadata database** to remove the database table including the library and the application folder.

Windows

Do the following to uninstall the Archive Viewer application:

1. Navigate to the …\installation\windows folder.
2. Run the below command:

   python uninstallApp.py

3. When prompted, select either of the following options:
   - **Remove only application** to remove the application folder
   - **Remove Application and Pip library** to remove the application and pip libraries that were used for this application
   - **Remove Application, Pip library and Metadata database** to remove the database table including the library and the application folder.

3. Onboarding Process

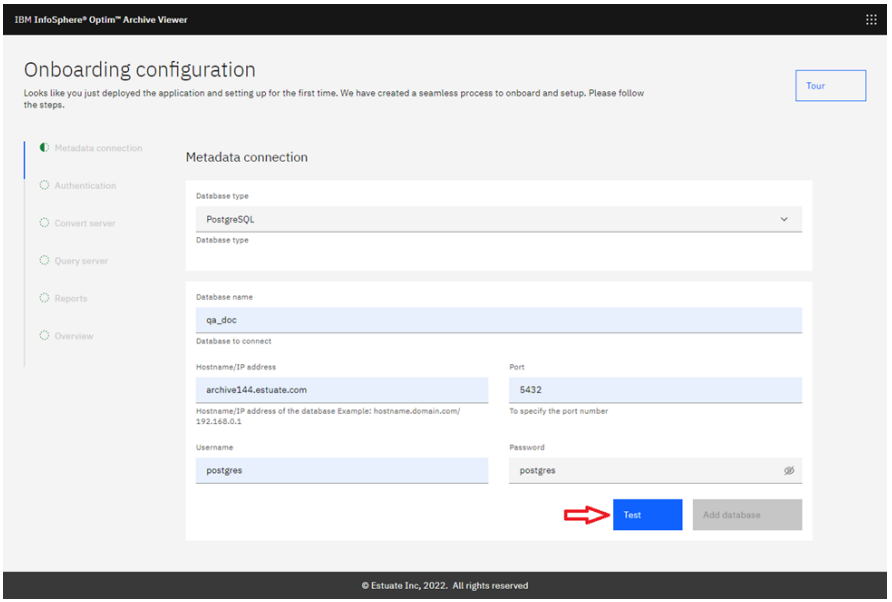
Once you deploy the application and performing set up for the first time, you should follow the onboarding process.

Follow the onboarding process in the below sequence:

1. **Configuring Metadata Connection**
2. **Authenticating Users**
3. **Configuring Conversion Server**
4. **Configuring Query Server**
5. **Configuring Reports**

3.1 Configuring Metadata Connection

You must configure the metadata connection details of the database to track all the jobs.



Do the following values to configure the metadata connection on the Onboarding configuration page.

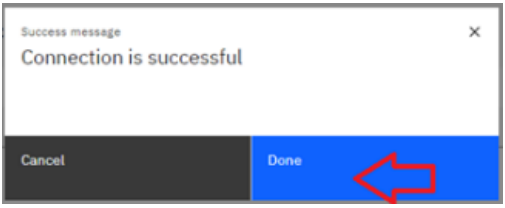
Note: The steps are given for the PostgreSQL database type.

1. Select the database type from the **Database type** to which you want to connect with the application.
2. Type the database name in the **Database name** field.
3. Type the hostname or IP address of the database in the **Hostname/IP address** field.
4. Type the port number of the database in the **Port** field.

5. Type the username and password of the database in the **Username** and **Password** fields respectively.

6. After you specify the appropriate values for the database connection, click **Test** to establish the connection.

7. Click **Done** when the connection is successful.



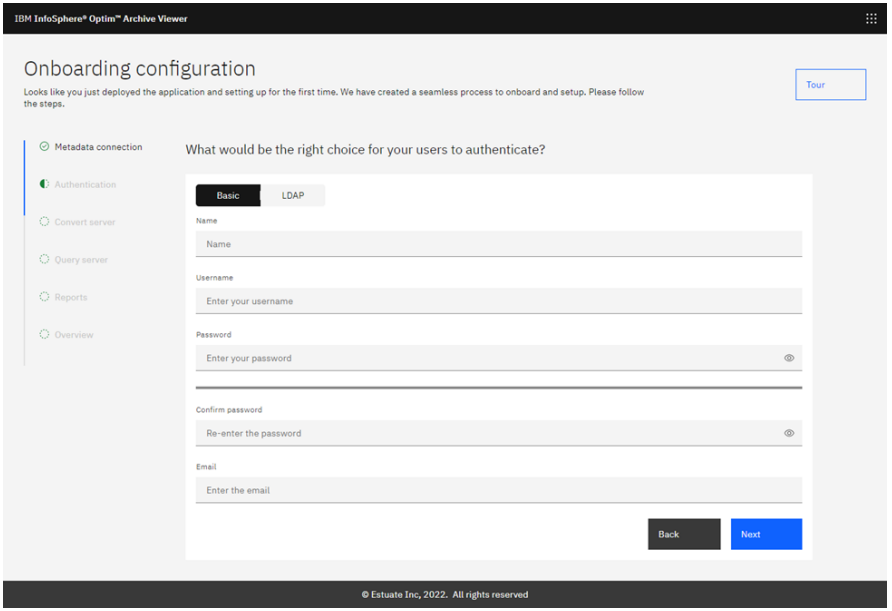
Note: For Oracle, select either SID or Service Name from the **Connection type** drop-down list based on the connection information.
In SQLite, provide the directory path where you want to place the DB file. Optionally, you can provide details, such as **Username**, **Password** and **Hostname**. If you provide any one of these field information, you must enter all the other information.

8. Click **Add database** to complete the metadata connection.

3.2 Authenticating Users

Basic authentication

You can create a super user with basic credentials.



Do the following to create a super user with basic credentials.

1. Type the name of a basic user in the **Name** field.
2. Type the username and password of the database in the **Username** and **Password** fields respectively.

   Note: The password should satisfy below conditions:

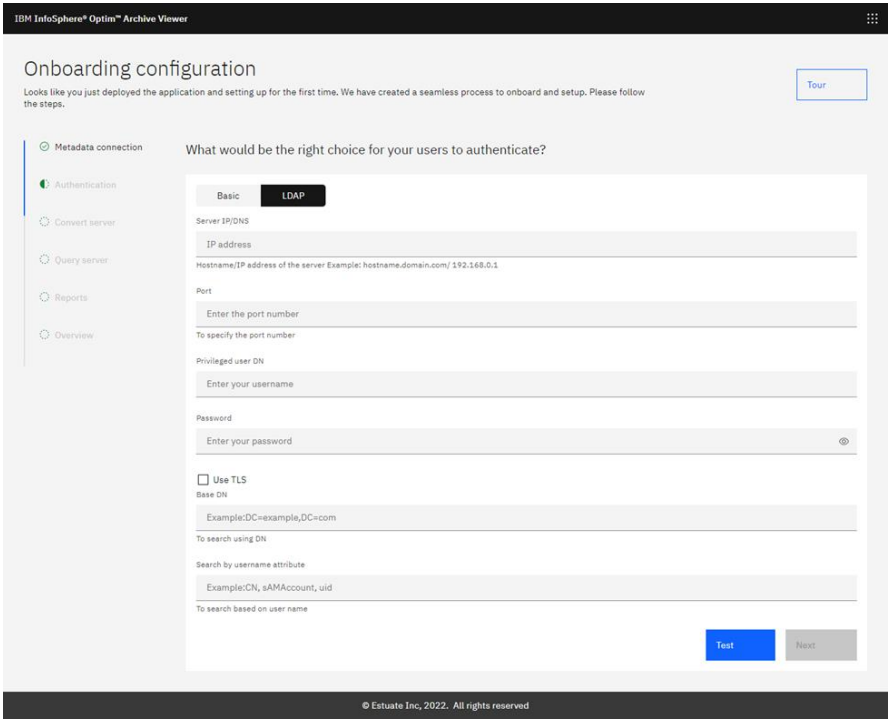
   - 8-20 characters
   - At least one uppercase character

- At least one lowercase character
- At least one number
- No spaces
- At least one special character from the following: !~@#$%^&*.-

3. Type the same password in the **Confirm password** field.
4. Type the email address.
5. Click **Next** after you specify all the above details to navigate to the **Convert server** tab.

LDAP authentication

You can create a super user with LDAP credentials.



Do the following to create a super user with LDAP credentials.

1. Type the hostname or the IP address of the server in the **Server IP/DNS** field.
2. Type the port number in the **Port** field.
3. Type the username in the **Privileged user DN** field.
4. Type the password of the user in the **Password** field.

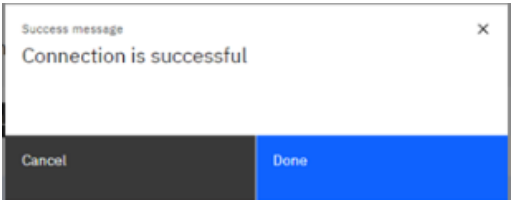
5. Select the **Use TLS** checkbox to connect via LDAPs, otherwise, leave it unselected.

6. Type the DN username in the **Base DN** field from where the application should search for the user.

7. Type the attribute value in the **Search by username attribute** field based on which the username will be validated.

8. Click **Test** to authenticate users.

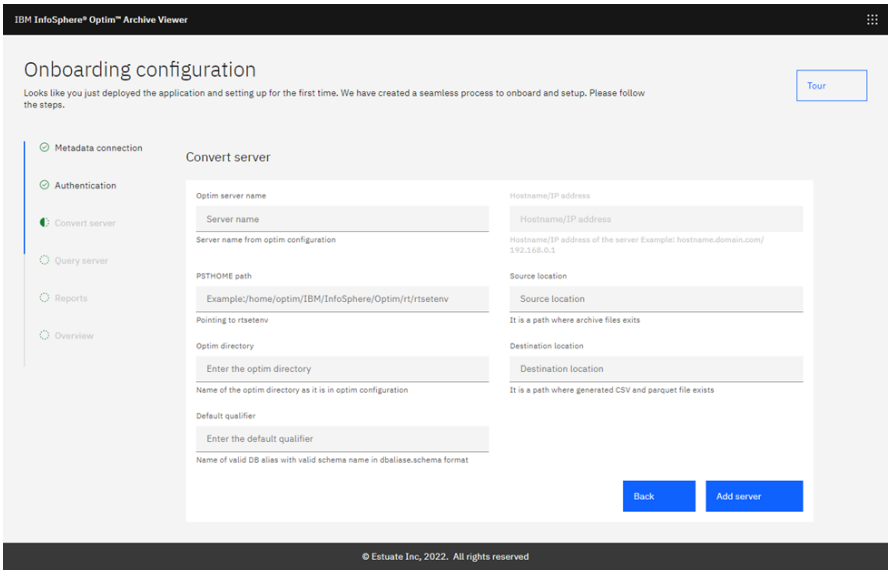
9. Click **Done** when the connection is successful or **Cancel** to discard the changes.



10. Click **Next** after you specify all the above details to navigate to the **Convert server** tab.

3.3 Configuring Conversion Server

On successful user authentication, you will be navigated to the **Convert server** tab.



Do the following steps to configure a Conversion Server:

1. Type the server name from Optim configuration in the **Optim server name** filed.

2. Type the corresponding path that points to rtsetenv for Linux and %PSTHOME%/bin for Windows in the **PSTHOME path**.

3. Type the source directory, where the archive file exists, in the **Source location** path.

4. Type the Optim directory name as it is in the Optim configuration in the **Optim directory** field.

5. Type the destination directory, where the generated CSV and Parquet files exist, in the **Destination location** path.

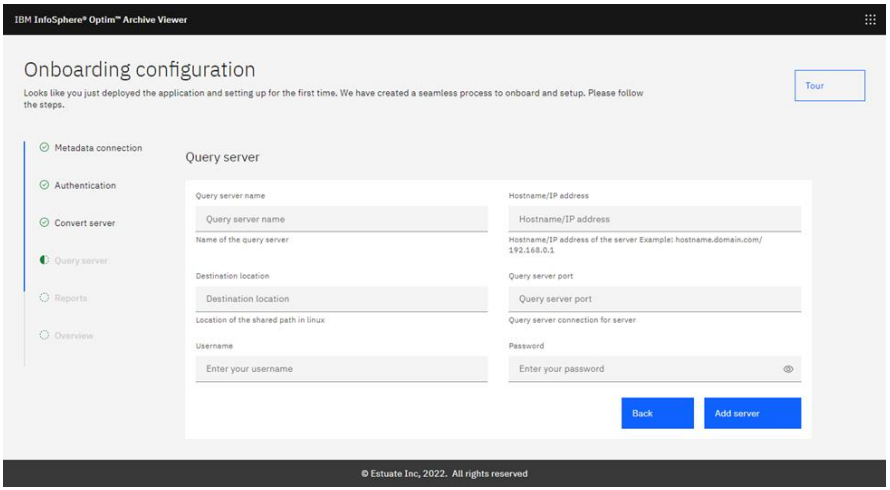
   Note: The destination directory should be shared with the server where the Query Server has been deployed. If the Query Server is in the same server, then the user initiating the Query Server should have access to the files.

6. Type the valid DB alias name with valid schema name in the dbalias.schema format in the **Default qualifier** field.

7. Click **Add server** to add the conversion server details.

3.4 Configuring Query Server

You will be navigated to the **Query server** tab only if all the details provided to configure Conversion Server are valid.



Do the following steps to configure Query Server:

1. Type the Query Server name in the **Query server name** field.

2. Type the destination directory, where the generated CSV and Parquet files exist, in the **Destination location** path.

   Note: The destination directory should be shared with the server where the Query Server has been deployed. If the Query Server is in the same server, then the user initiating the Query Server should have access to the files.

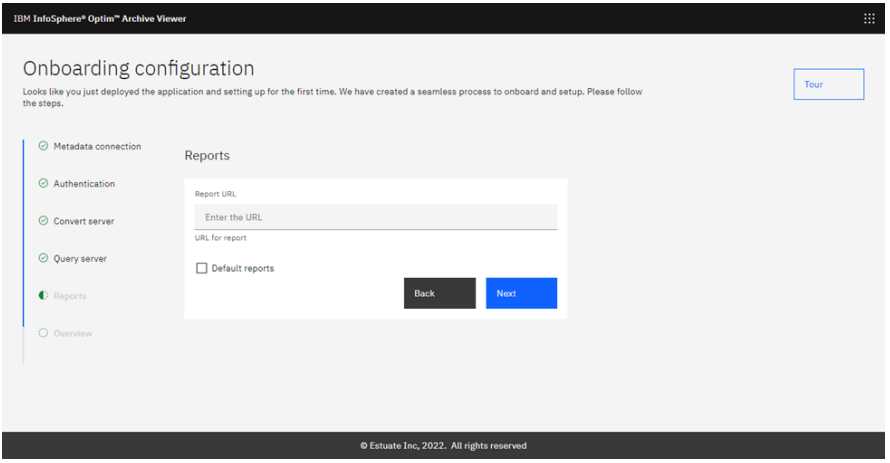
3. Type the Query Server port number in the **Query server port** field.

4. Type the username and password in the **Username** and **Password** fields, respectively.

5. Click **Add server** to add the query server details.

3.5 Configuring Reports

You will be navigated to the **Reports** tab only if all the details provided to configure Query Server are valid.



Do the following steps to set up Reports directory:

1. Type the URL where reports are to be stored in the **Report URL** field.

2. Select the **Default reports** check box, if the Reporting Server is installed in the same server as Query Server.

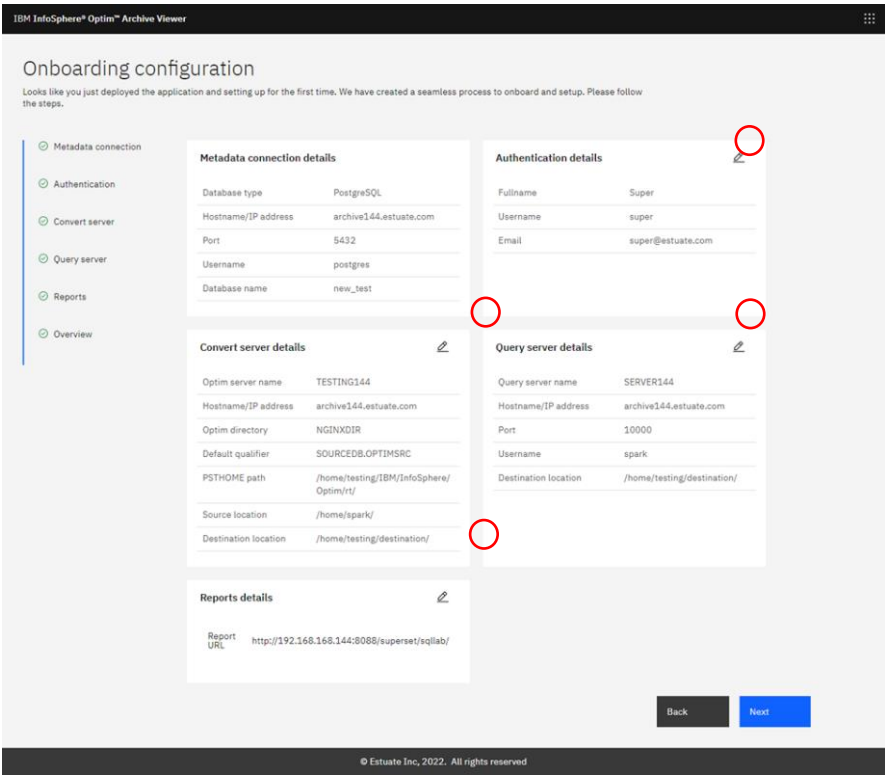
3. Click **Next** to complete the configuration. The **Overview** tab is displayed if all the configuration details are valid.

3.6 Modifying configurations

Once all the server configuration details are added, you can modify the user authentication details and the conversion, query and reports server details.

Do the following steps to modify the configuration details:

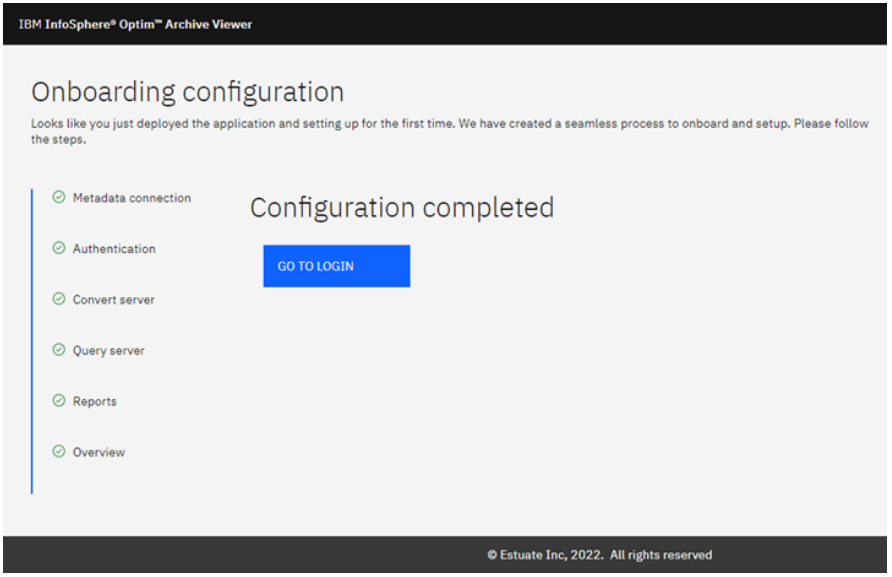
1. Navigate to the **Overview** tab.



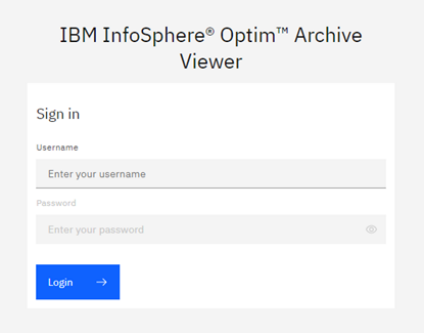
2. Click the Edit icon for which section you want to modify details.

3. Edit the required fields you want to modify.

4. Click **Next**. The **Configuration completed** message appears.

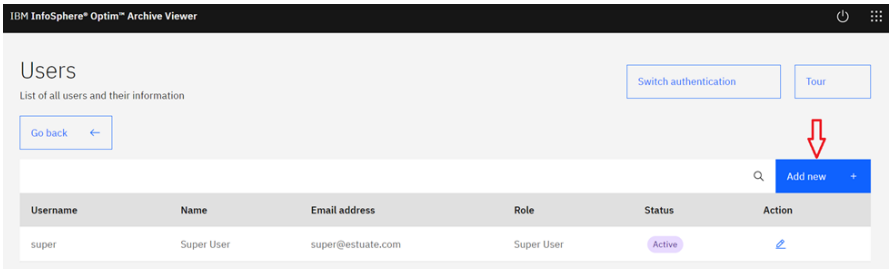


5. Click **GO TO LOGIN**. The Sign in page of the Archive Viewer application appears.

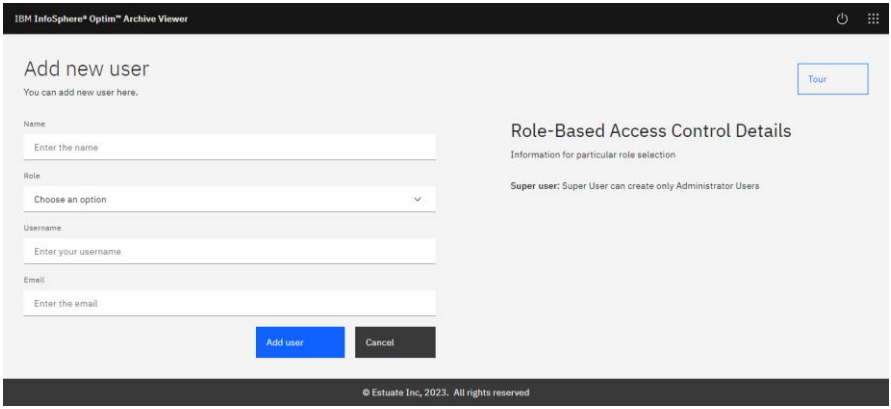


6. Type the super user credentials in the **Username** and **Password** fields and click **Login**. The application Dashboard appears.

7. Click **Manage users** on the dashboard. The Users page appears.

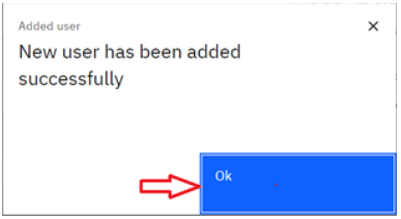


8. Click **Add new**. The Add new user page appears.



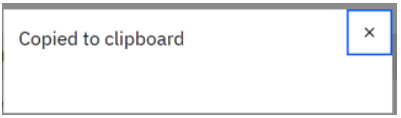
9. Type the name of the user in the **Name** field.
10. Select **Administrator** from the **Role** drop-down list.
11. Type the username and email ID in the **Username** and **Email** fields, respectively.
12. Click **Add user** to create an admin user or **Cancel** to navigate to the Users page without creating a user.

If you click **Add user**, the Added user dialog box appears as shown below.

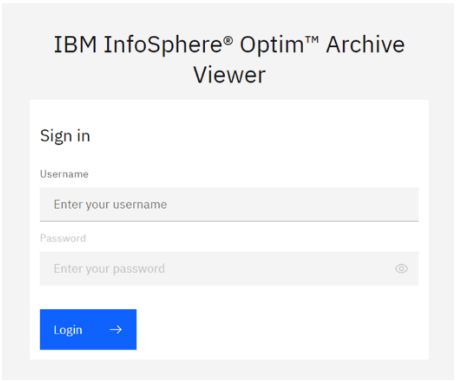


13. Click **OK**. The admin user is added to the users list.

14. Click the Copy icon to copy the auto-generated temporary password. The Copied to clipboard message dialog box appears.

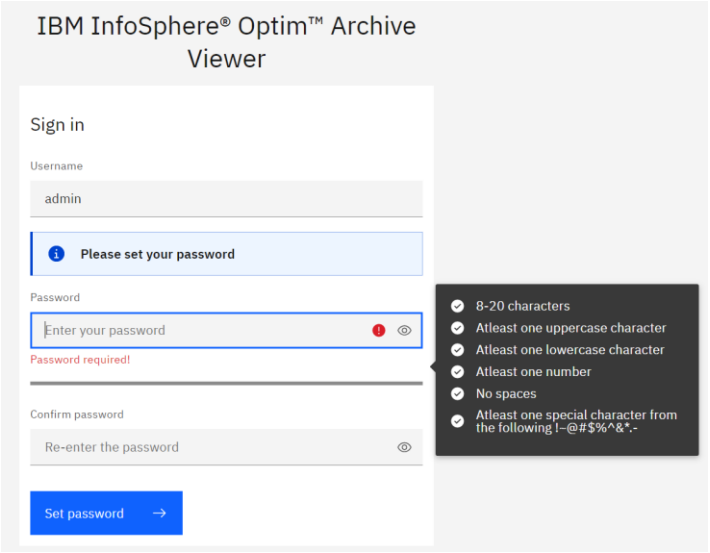


15. Click the Close ✕ icon to close the dialog box and then click the logout icon to log off from the application. The Sign in page appears.



16. Type the newly created admin user name in the **Username** field and paste the auto-generated temporary password that was copied in **Step 14** in the **Password** field.

17. Click **Login**. A note Please set your password appears along with fields to enter passwords.



18. Type a new password you want in the **Password** and **Confirm password** fields.

19. Click **Set password**. An inline message appears as Password has been set. You can login now with new password.

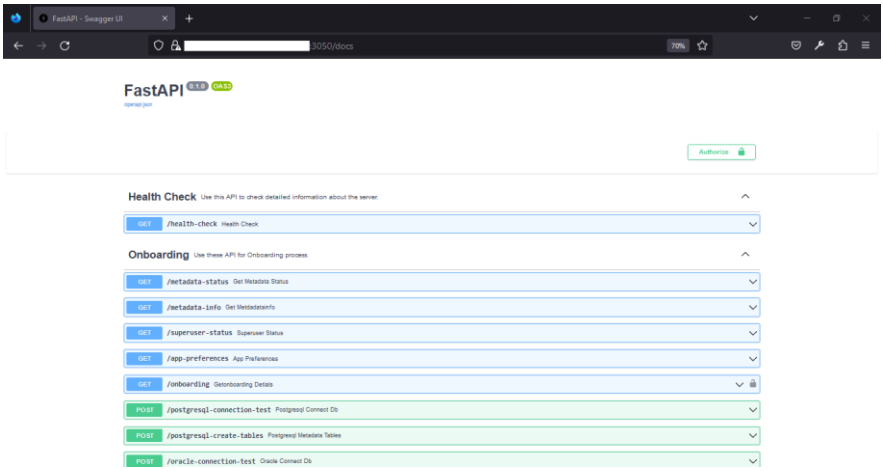
20. Click **Login**. The dashboard for an administrator appears.

4. Archive Viewer API

You can use the Archive Viewer Application Program Interface (API) if you do not want to use the Archive Viewer application user interface (UI).

You can access the swagger documents at the following site:

https://*hostname*:3050/docs

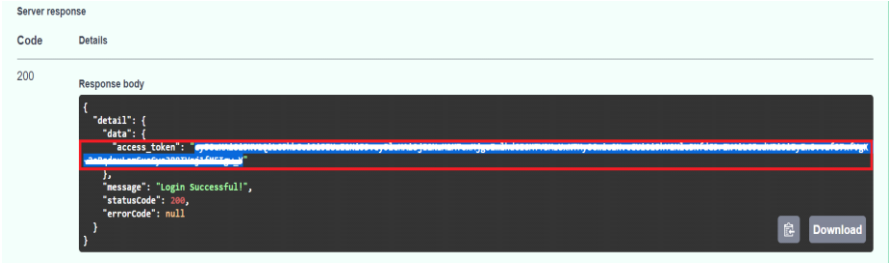


4.1 Accessing the Archive Viewer API

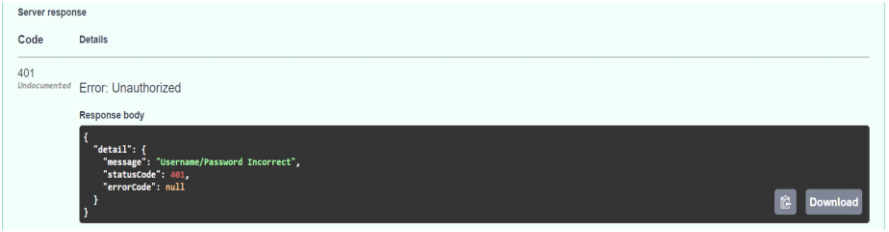
Do the following to access the Archive Viewer API:

1. Navigate to the swagger page.
2. In the Login API section, expand POST/login and click **Try it out**.
3. Type username and password and click **Execute**.

   - An access token is generated if the entered username and password are correct.



- An error code is displayed if the entered username and password are incorrect.



4. Copy the access token.
5. Click **Authorize** on the swagger page top-right corner.
6. Paste the copied access token in the **Value** field and click **Authorize**.

Now, you can access the Archive Viewer API.

Index

P

S

U

Product Number:

Printed in USA