Engineering Excellence J1

(Hadoop Cluster Setup – Ranger Setup)

Ranger installation and Configuration

TABLE OF CONTENT

APACHE RANGER

Comprehensive security for Enterprise Hadoop

Apache Ranger delivers a comprehensive approach to security for a Hadoop cluster. It provides central security policy administration across the core enterprise security requirements of authorization, authentication, audit and data protection.

Apache Ranger offers a centralized security framework to manage fine-grained access control over Hadoop data access components like Apache Hive and Apache HBase. Using the Apache Ranger console, security administrators can easily manage policies for access to files, folders, databases, tables, or column. These policies can be set for individual users or groups and then enforced within Hadoop.

Security administrators can also use Apache Ranger to manage audit tracking and policy analytics for deeper control of the environment. The solution also provides an option to delegate administration of certain data to other group owners, with the aim of securely decentralizing data ownership.

Apache Ranger currently supports authorization, authentication, auditing, data encryption and security administration for the following HDP components:

- Apache Hadoop HDFS
- Apache Hive
- Apache HBase
- Apache Storm
- Apache Knox
- Apache Solr
- Apache Kafka
- YARN

Overview

Apache Ranger can be installed either manually using the Hortonworks Data Platform (HDP) or the Ambari 2.1 User Interface (UI). Unlike the manual installation process, which requires you to perform a number of installation steps, installing Ranger using the Ambari UI is simpler and easier. The Ranger service option will be made available through the Add Service wizard after the HDP cluster is installed using the installation wizard.

Once Ambari has been installed and configured, you can use the Add Service wizard to install the following components:

- Ranger Admin
- Ranger UserSync
- Ranger Key Management Service

After these components are installed and started, you can enable Ranger plugins by navigating to each individual Ranger service (HDFS, HBase, Hiveserver2, Storm, Knox, YARN, and Kafka) and modifying the configuration under advanced ranger-<service>-plugin-properties.

Note that when you enable a Ranger plugin, you will need to restart the component.

Note

Enabling Apache Storm or Apace Kafka requires you to enable Kerberos. To enable Kerberos on your cluster, see Enabling Kerberos Security in the Ambari Security Guide.

Installation Prerequisites

Before you install Ranger, make sure your cluster meets the following requirements:

- A MySQL, Oracle, or PostgreSQL database instance is running and available to be used by Ranger.

The Ranger installation will create two new users (default names: rangeradmin and rangerlogger) and two new databases (default names: ranger and ranger_audit).

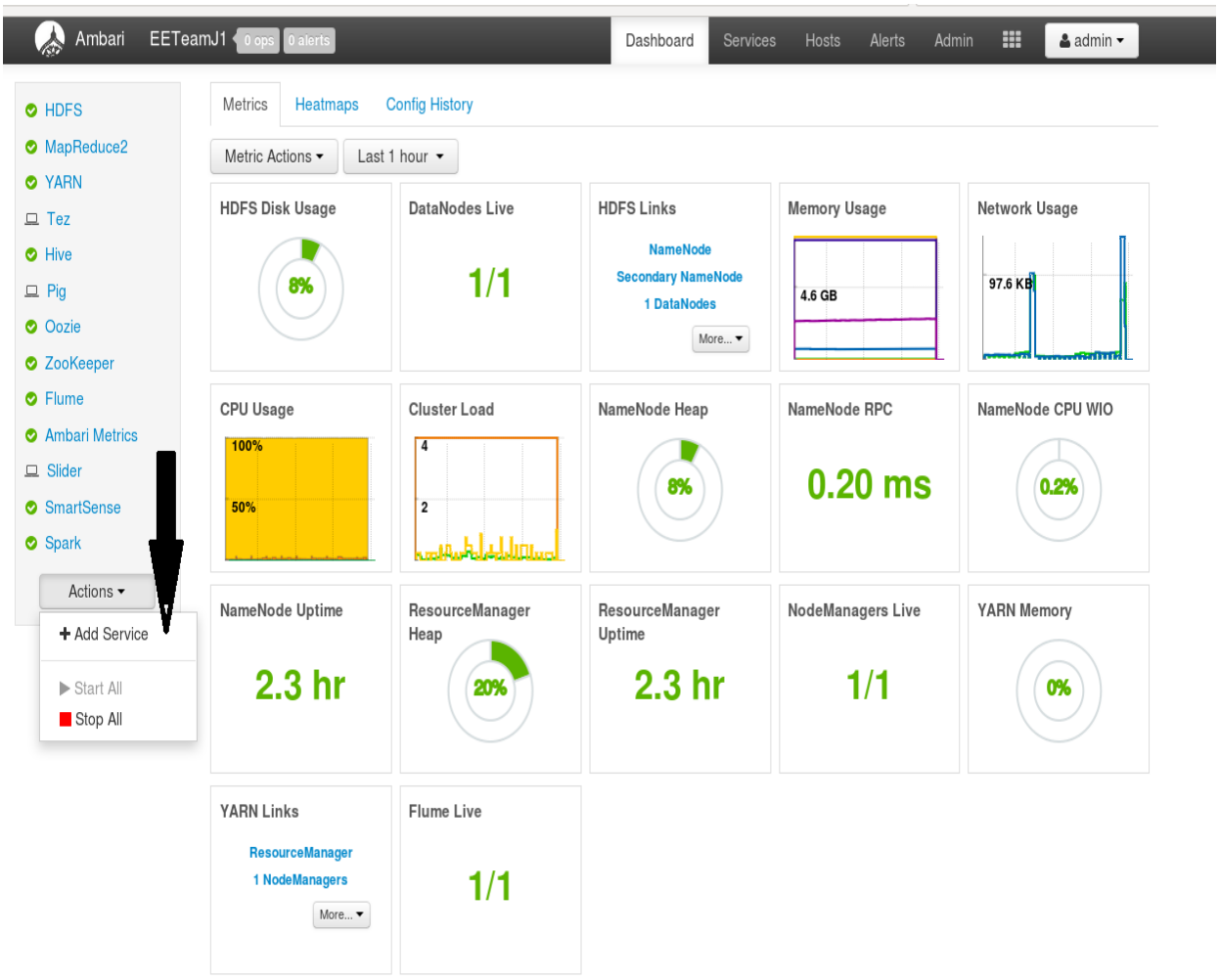
- Configure the database instance for Ranger as described in the following sections.
    - Configuring MySQL for Ranger
    - Configuring PostgreSQL for Ranger
    - Configuring Oracle for Ranger

Ranger Installation

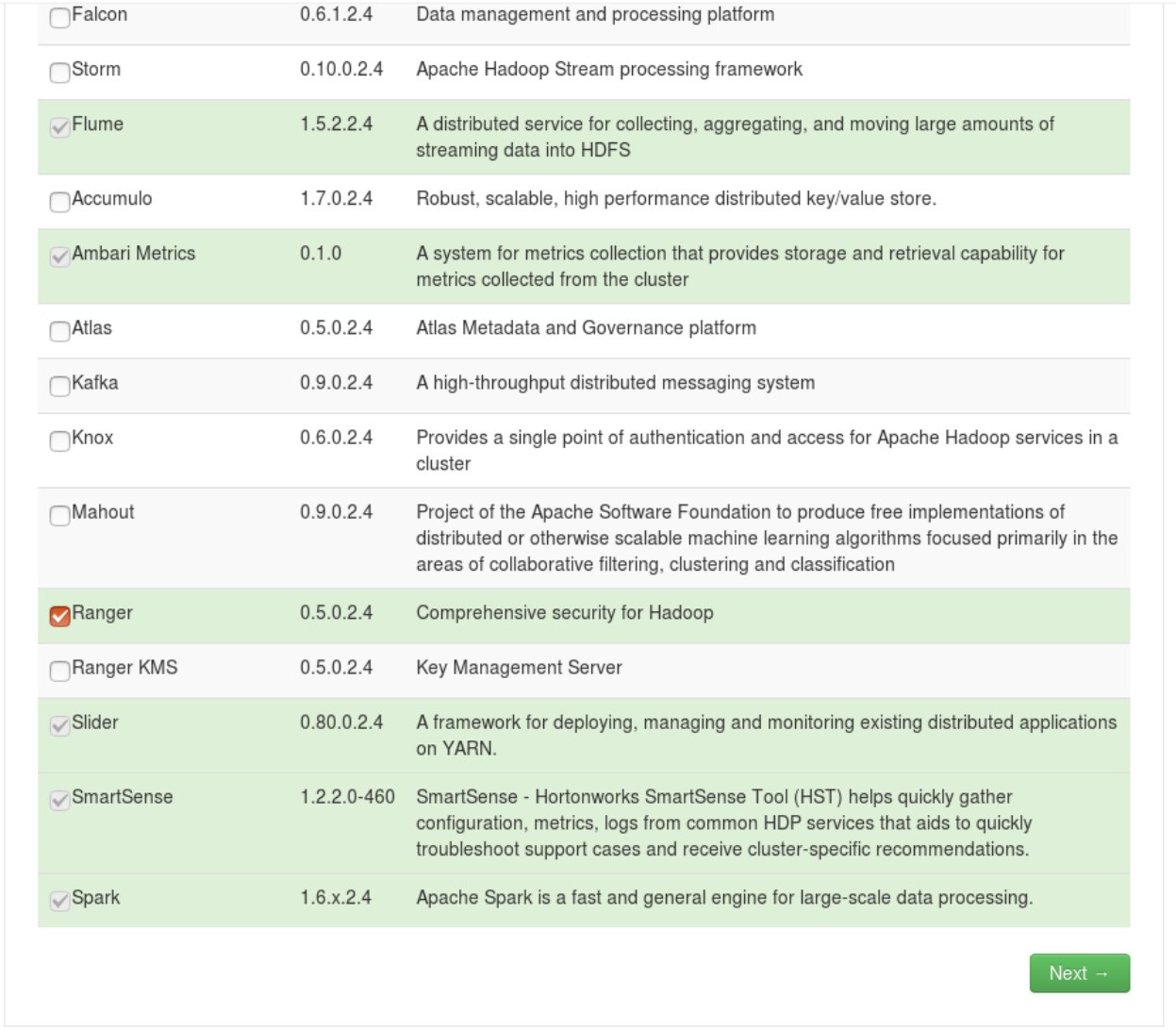
Use the following steps to install Ranger using Ambari.

Start the Installation

1. Log into your Ambari cluster with your user credentials. The main Ambari Dashboard page will be displayed.
2. In the left navigation menu, click **Actions**, then select **Add Service**.



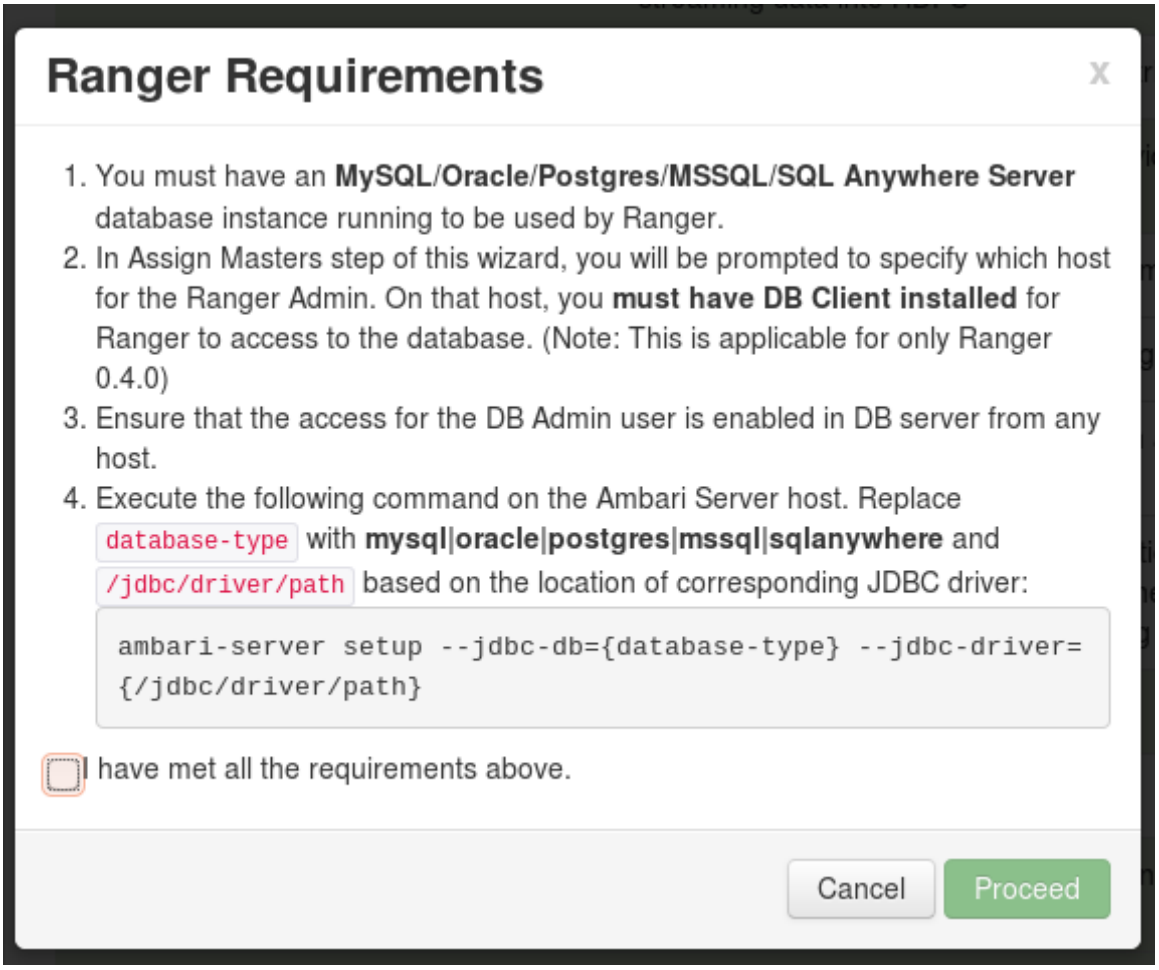
3. On the Choose Services page, select Ranger, then click Next



Falcon	0.6.1.2.4	Data management and processing platform
Storm	0.10.0.2.4	Apache Hadoop Stream processing framework
Flume	1.5.2.2.4	A distributed service for collecting, aggregating, and moving large amounts of streaming data into HDFS
Accumulo	1.7.0.2.4	Robust, scalable, high performance distributed key/value store.
Ambari Metrics	0.1.0	A system for metrics collection that provides storage and retrieval capability for metrics collected from the cluster
Atlas	0.5.0.2.4	Atlas Metadata and Governance platform
Kafka	0.9.0.2.4	A high-throughput distributed messaging system
Knox	0.6.0.2.4	Provides a single point of authentication and access for Apache Hadoop services in a cluster
Mahout	0.9.0.2.4	Project of the Apache Software Foundation to produce free implementations of distributed or otherwise scalable machine learning algorithms focused primarily in the areas of collaborative filtering, clustering and classification
Ranger	0.5.0.2.4	Comprehensive security for Hadoop
Ranger KMS	0.5.0.2.4	Key Management Server
Slider	0.80.0.2.4	A framework for deploying, managing and monitoring existing distributed applications on YARN.
SmartSense	1.2.2.0-460	SmartSense - Hortonworks SmartSense Tool (HST) helps quickly gather configuration, metrics, logs from common HDP services that aids to quickly troubleshoot support cases and receive cluster-specific recommendations.
Spark	1.6.x.2.4	Apache Spark is a fast and general engine for large-scale data processing.

Next →

4. The Ranger Requirements page appears. Ensure that you have met all of the installation requirements, then select the "I have met all the requirements above" check box and click **Proceed**.



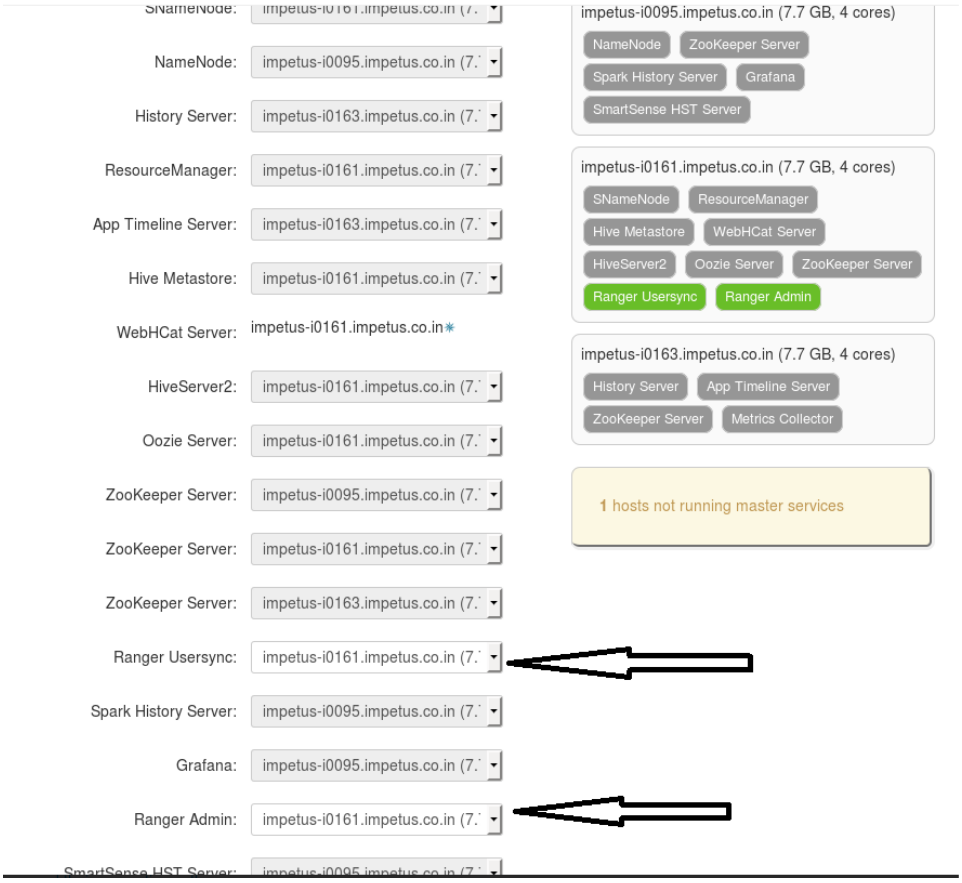
Run the below command using root or postgres user on **impetus-i0161.impetus.co.in** as that's where ambari is running. Once you get the setup successful message then we can proceed to the next step.

```
# ambari-server setup --jdbc-db=postgres --jdbc-driver=/usr/lib/ambari-
server/postgresql-9.3-1101-jdbc4.jar
Using python  /usr/bin/python
Setup ambari-server
Copying /usr/lib/ambari-server/postgresql-9.3-1101-jdbc4.jar to
/var/lib/ambari-server/resources
JDBC driver was successfully initialized.
Ambari Server 'setup' completed successfully.
```

5. You are then prompted to select the host where Ranger Admin will be installed. This host should have DB admin access to the Ranger DB host and UserSync. Notice in the figure below that both the Ranger Admin and Ranger Usersync services will be installed on the host **impetus-i0161.impetus.co.in** as postgres is already configured on that host.

   Make a note of the Ranger Admin host for use in subsequent installation steps. Click **Next** when finished to continue with the installation.

Note : The Ranger Admin and Ranger Usersync services must be installed on the same cluster node.



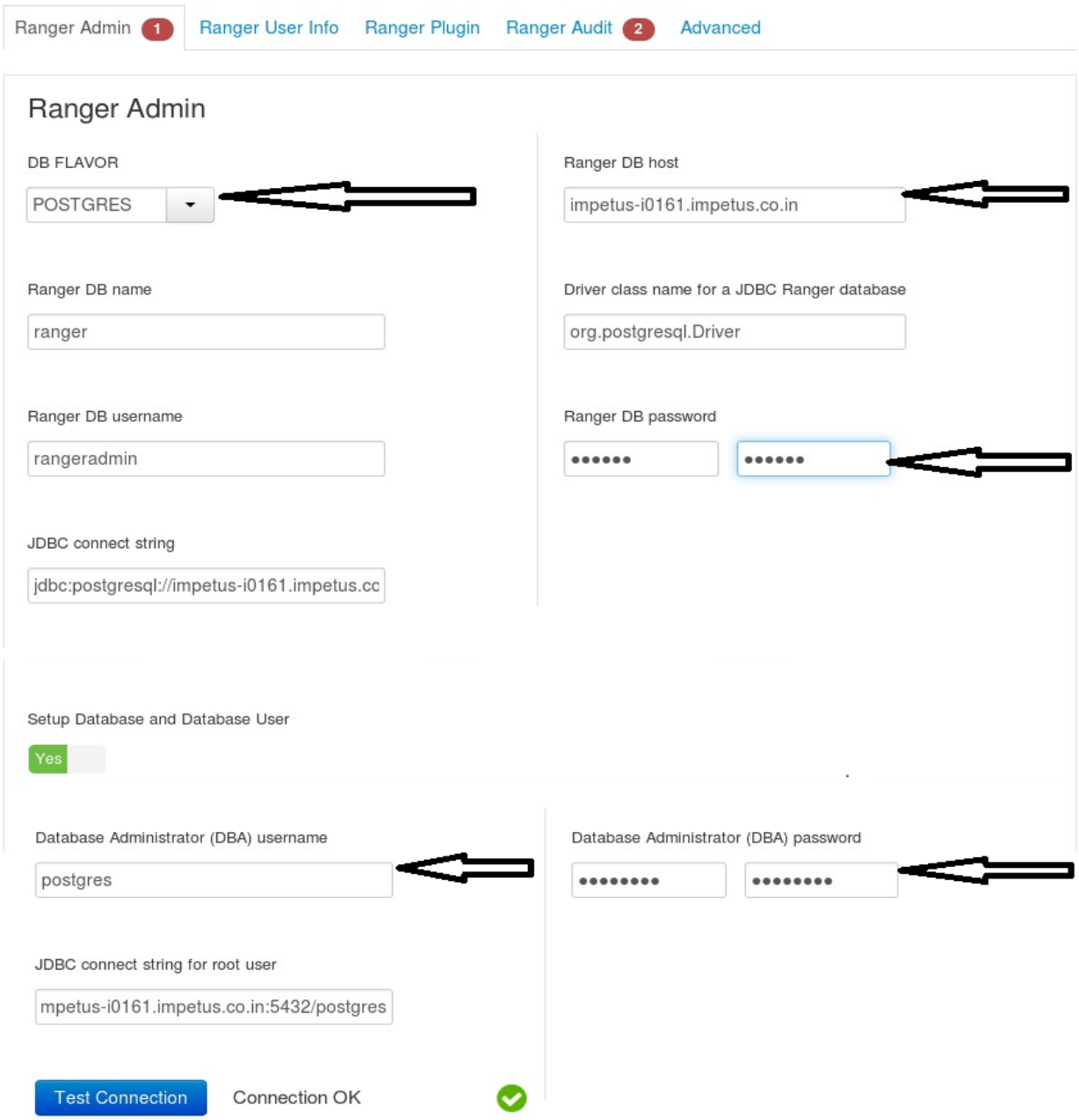
Customize Services

The next step in the installation process is to specify Ranger settings on the Customize Services page. You must specify all of the following settings on the Customize Services page before clicking Next at the bottom of the page to continue with the installation.

Admin Settings

1. Under Admin Settings on the Customize Services page, type in the password for the user account used by Ambari(This setting is only on a non-ambari host) . This password will only be used by the Ambari Agent, and will be used with the user name specified in the Ranger configuration asranger_admin_username under "Advanced ranger-env".

Use **postgres** as DB_Flavour and enter the Ranger DB host as **impetus-i0161.impetus.co.in** and set the Ranger DB password as "**ranger**"

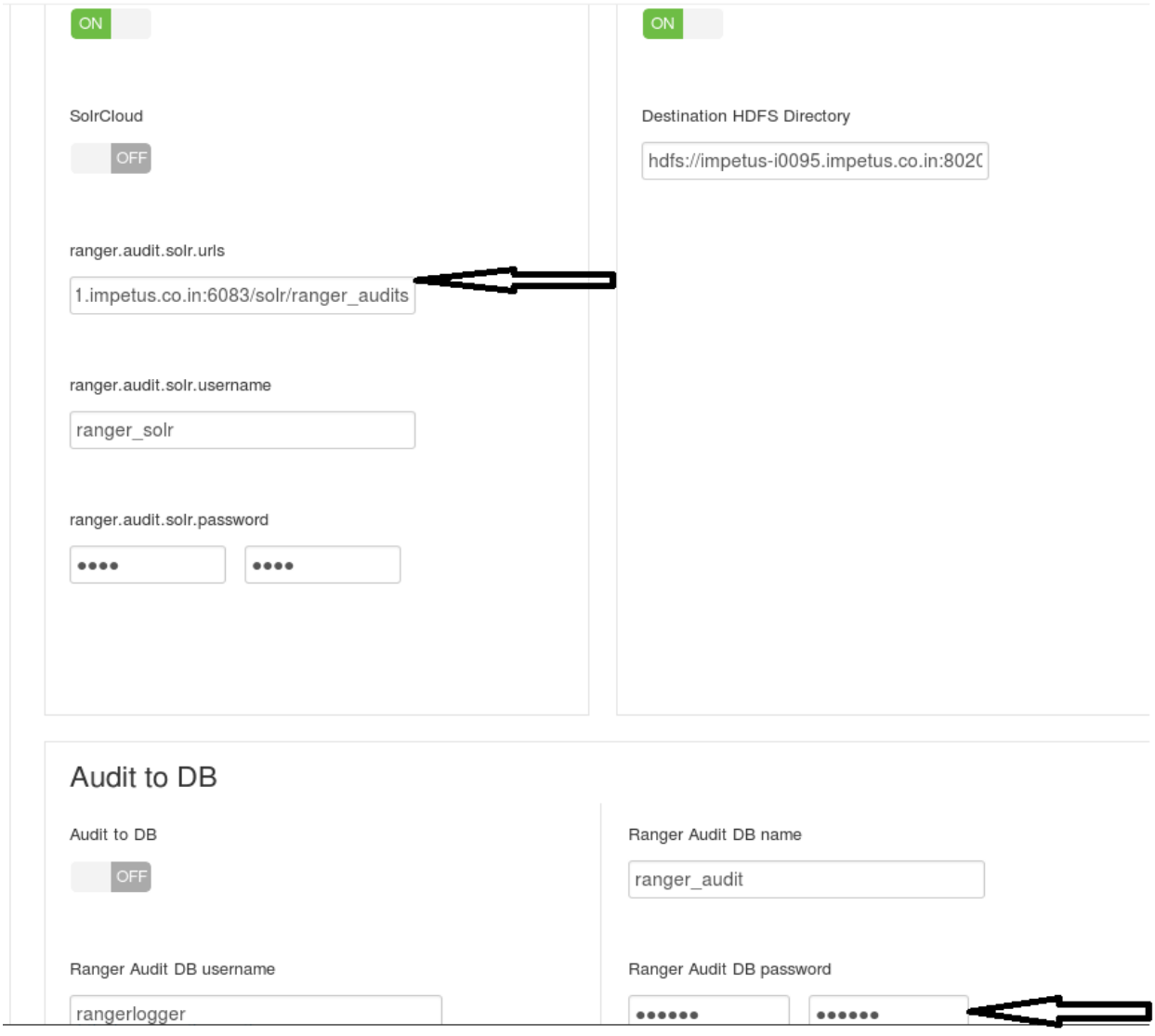


If you get error while testing connection saying postgres password is wrong then run the following on the box **impetus-i0161.impetus.co.in**

```
# sudo su - postgres
# psql
postgres=# ALTER USER postgres with encrypted password 'postgres';
```

Leave default settings for **Ranger User info** and enable all available plugins under **Ranger Plugin** tabs. In **Ranger Audits** and **advance** tabs change only those show in the figure leave others default.

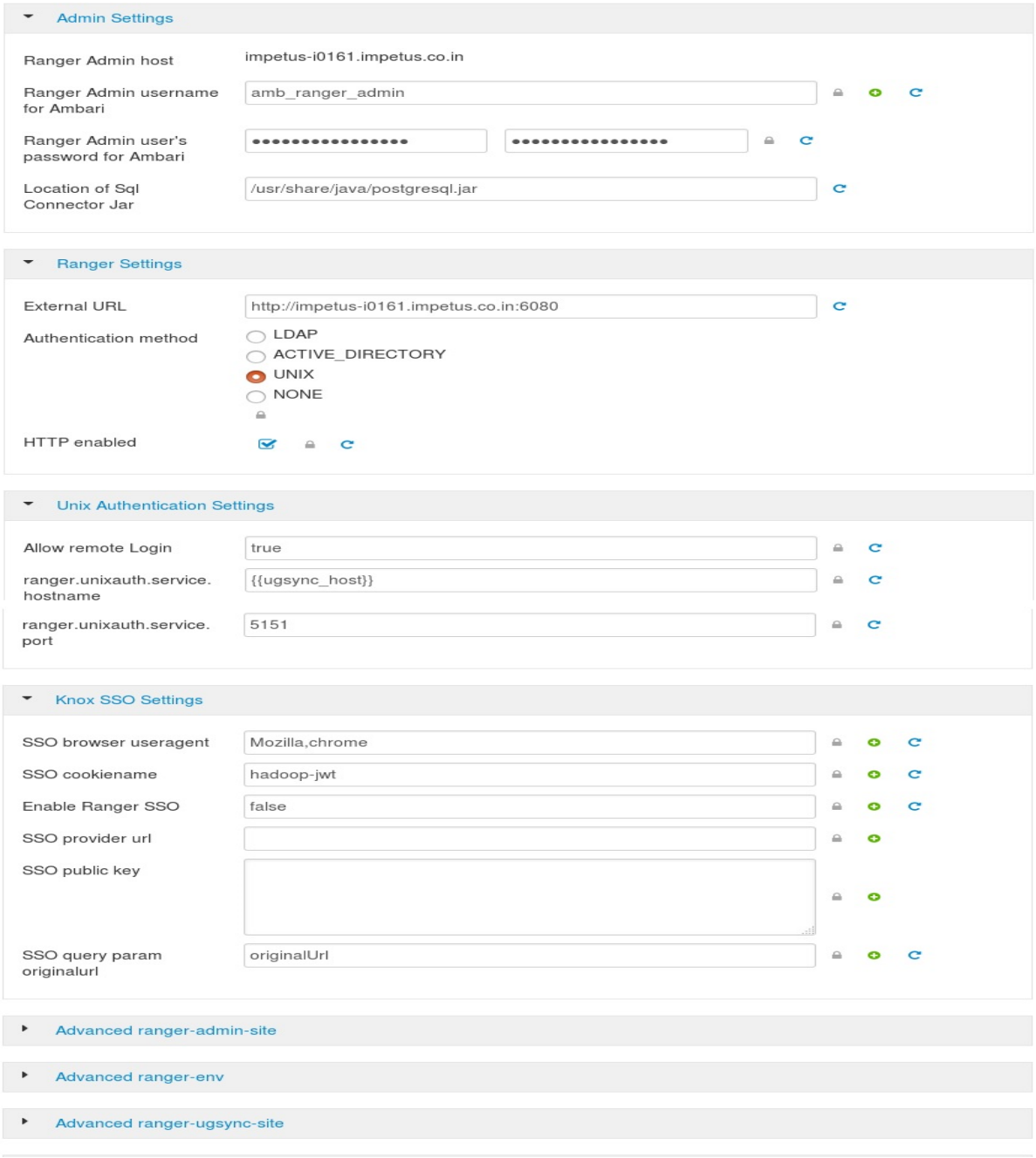
Ranger.audit.solr.username = http://impetus-i0161.impetus.co.in:6083/solr/ranger_audits
Ranger Audit DB password = ranger



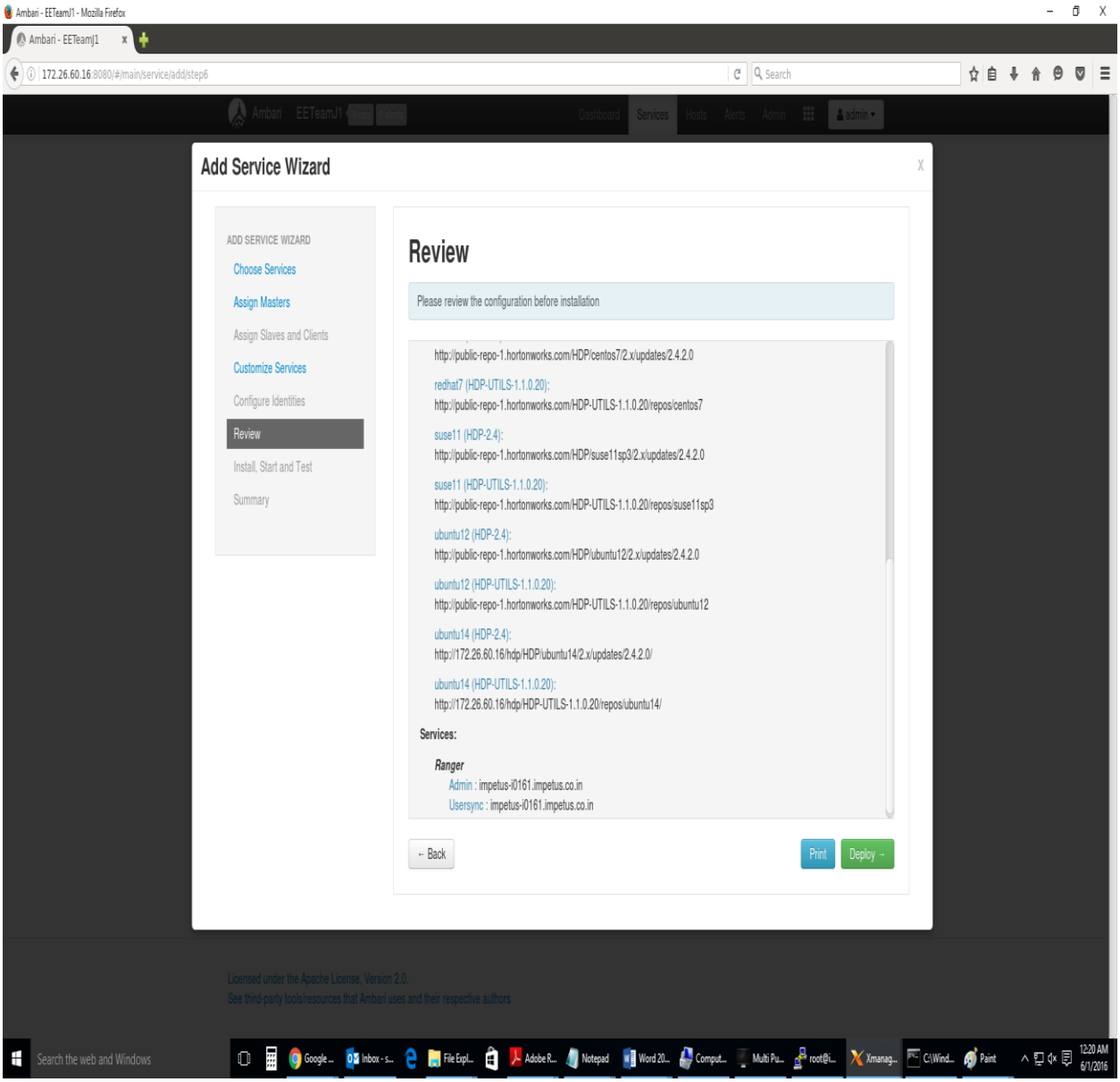
Using Apache Solr for Ranger Audits : Apache Solr is an open-source enterprise search platform. Apache Ranger can use Apache Solr to store audit logs, and Solr can also to provide a search capability of the audit logs through the ranger Admin UI

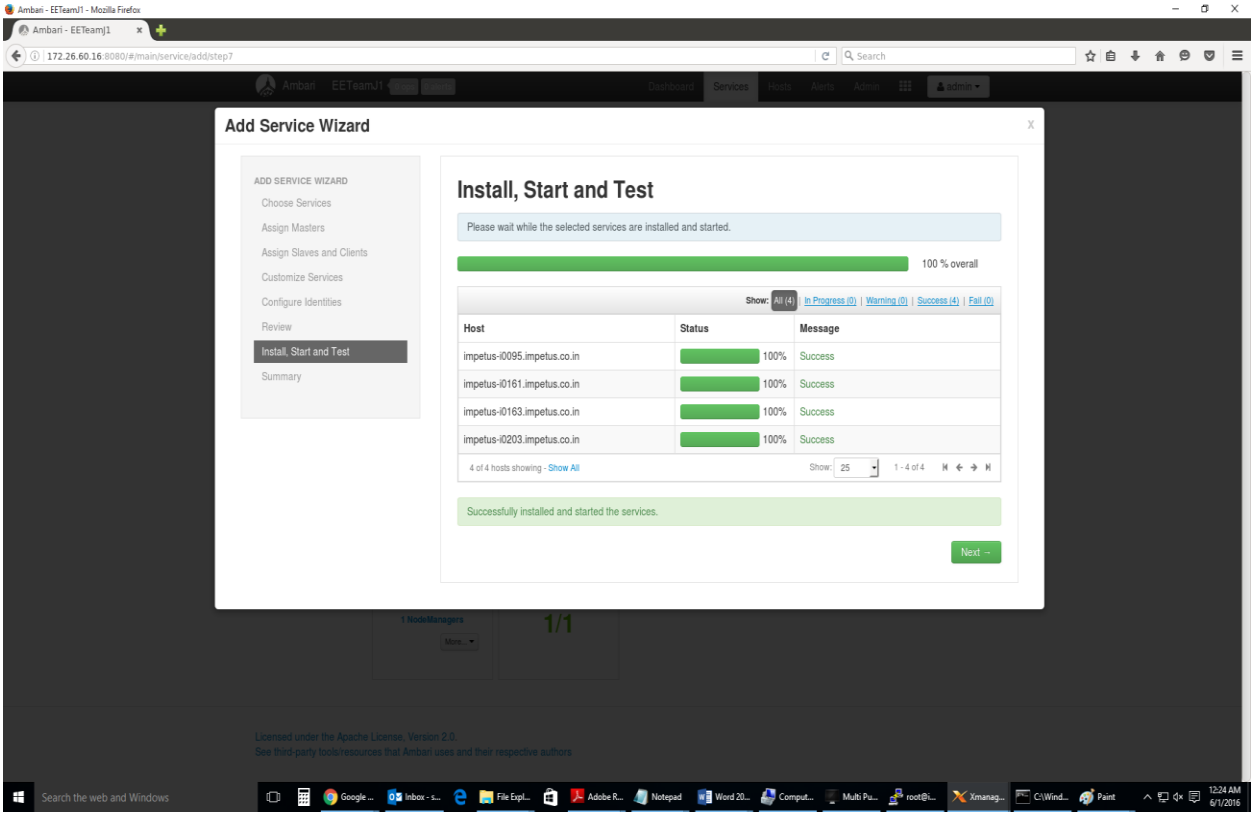
Solr Prerequisites

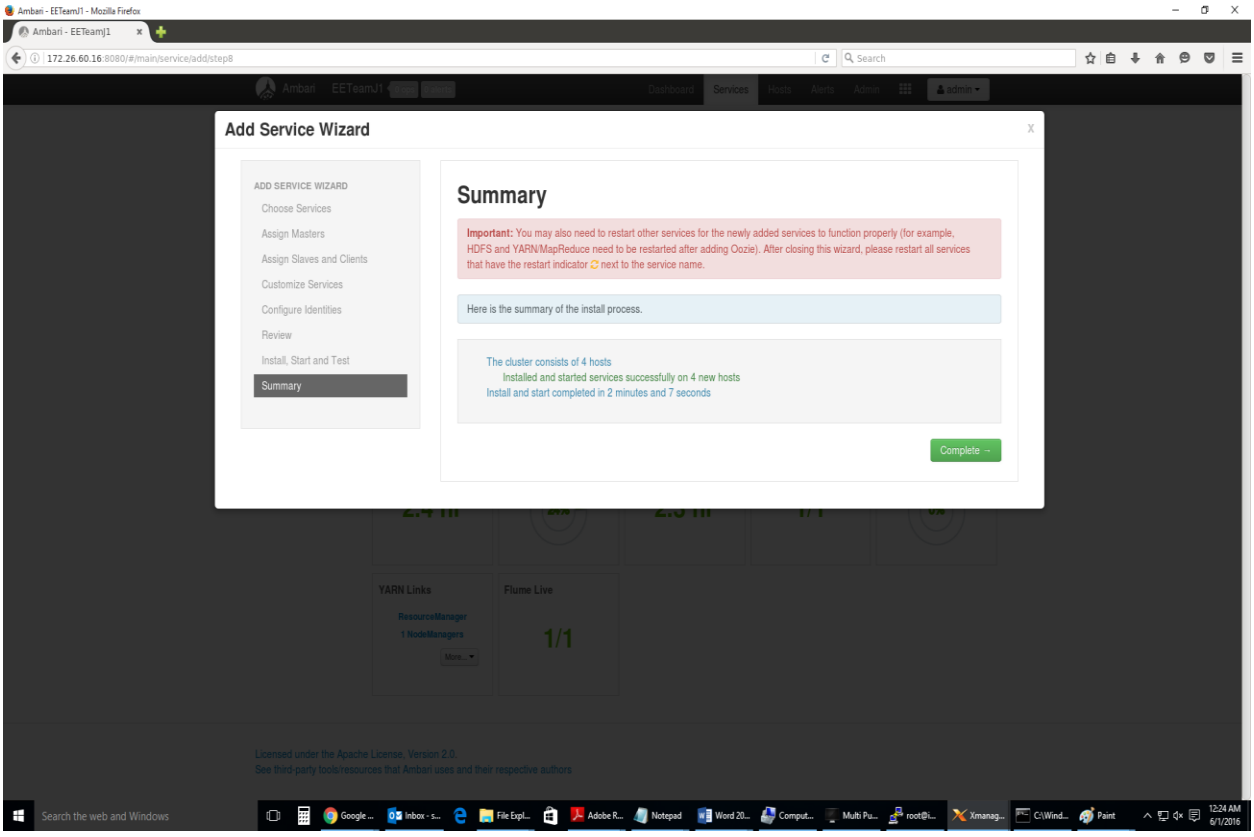
- Ranger supports Apache Solr 5.2 or higher.
- Apache Solr requires the Java Runtime Environment (JRE) version 1.7 or higher.
- 1 TB free space in the volume where Solr will store the index data.
- 32 GB RAM.

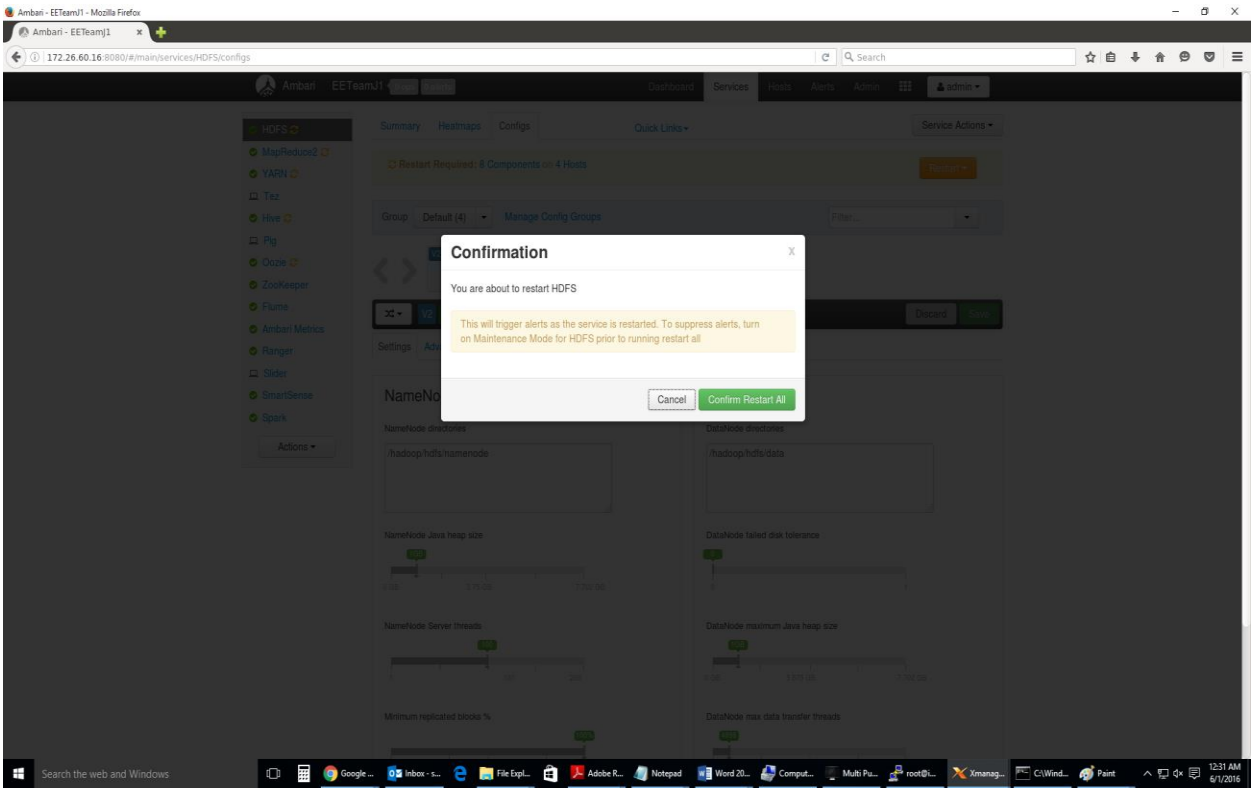


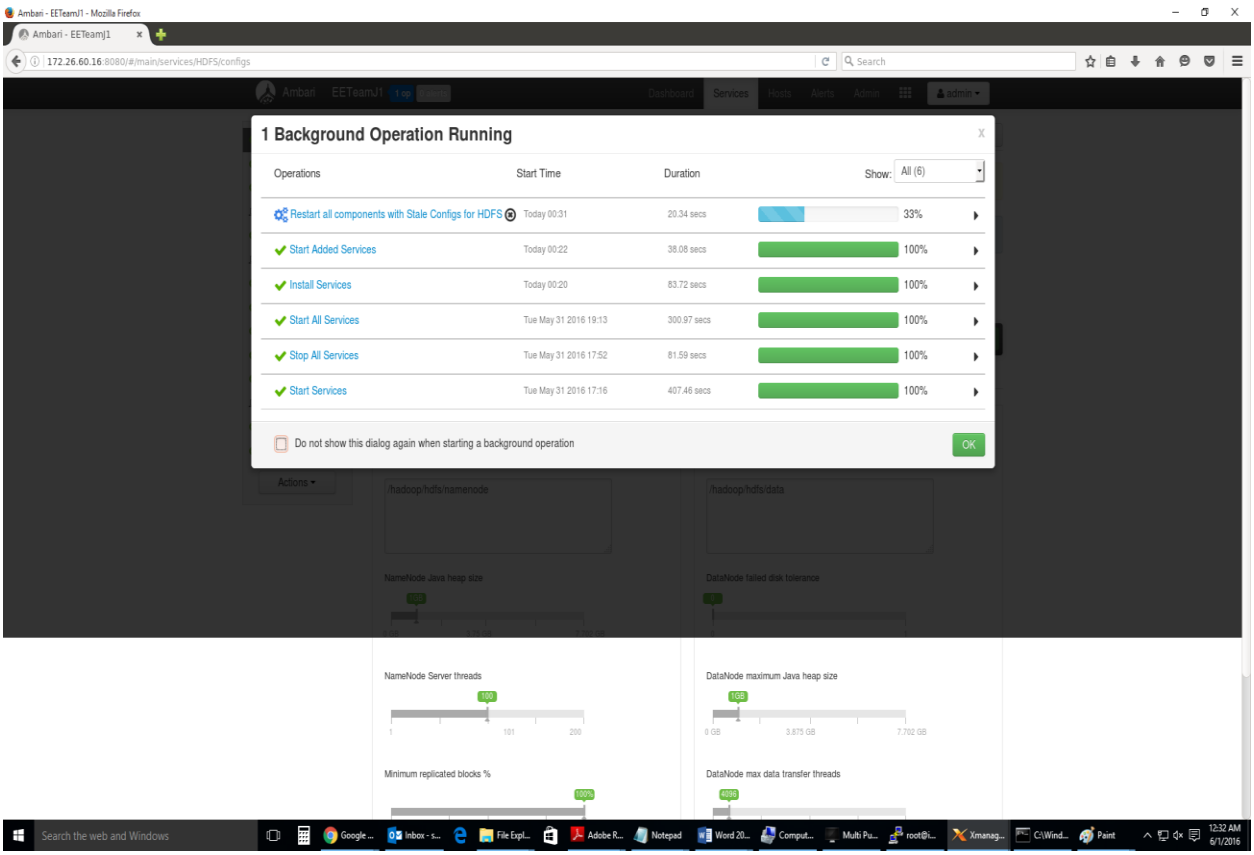
Review and Finalize installation




Ambari - EETeamJ1 - Mozilla Firefox
Add Service Wizard
ADD SERVICE WIZARD
Choose Services
Assign Masters
Assign Slaves and Clients
Customize Services
Configure Identities
Review
Install, Start and Test
Summary
Install, Start and Test
Please wait while the selected services are installed and started.
100 % overall
Show: All (4) | In Progress (0) | Warning (0) | Success (4) | Fail (0)
Host
Status
Message
impetus-i0095.impetus.co.in 100% Success
impetus-i0161.impetus.co.in 100% Success
impetus-i0163.impetus.co.in 100% Success
impetus-i0203.impetus.co.in 100% Success
4 of 4 hosts showing - Show All
Show: 25
1 - 4 of 4
Successfully installed and started the services.
Next →
Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors


Ambari - EETeamJ1 - Mozilla Firefox
Add Service Wizard
ADD SERVICE WIZARD
Choose Services
Assign Masters
Assign Slaves and Clients
Customize Services
Configure Identities
Review
Install, Start and Test
Summary
Summary
Important: You may also need to restart other services for the newly added services to function properly (for example, HDFS and YARN/MapReduce need to be restarted after adding Oozie). After closing this wizard, please restart all services that have the restart indicator next to the service name.
Here is the summary of the install process.
The cluster consists of 4 hosts
Installed and started services successfully on 4 new hosts
Install and start completed in 2 minutes and 7 seconds
Complete →
YARN Links
Flume Live
1/1
Licensed under the Apache License, Version 2.0.
See third-party tools/resources that Ambari uses and their respective authors


Ambari - EETeamJ1 - Mozilla Firefox
Ambari - EETeamJ1
172.26.60.16:8080/#/main/services/HDFS/configs
Confirmation
You are about to restart HDFS
This will trigger alerts as the service is restarted. To suppress alerts, turn on Maintenance Mode for HDFS prior to running restart all
Cancel
Confirm Restart All


Ambari - EETeamJ1 - Mozilla Firefox
Ambari - EETeamJ1
172.26.60.16:8080/#/main/services/HDFS/configs
1 Background Operation Running
Operations
Start Time
Duration
Show: All (6)
Restart all components with Stale Configs for HDFS
Today 00:31
20.34 secs
33%
Start Added Services
Today 00:22
38.08 secs
100%
Install Services
Today 00:20
83.72 secs
100%
Start All Services
Tue May 31 2016 19:13
300.97 secs
100%
Stop All Services
Tue May 31 2016 17:52
81.59 secs
100%
Start Services
Tue May 31 2016 17:16
407.46 secs
100%
Do not show this dialog again when starting a background operation
OK