Hortonworks Data Platform

Apache Solr Search Installation

docs.cloudera.com

Hortonworks Data Platform: Apache Solr Search Installation

Copyright © 2012-2019 Hortonworks, Inc. Some rights reserved.

The Hortonworks Data Platform, powered by Apache Hadoop, is a massively scalable and 100% open source platform for storing, processing and analyzing large volumes of data. It is designed to deal with data from many sources and formats in a very quick, easy and cost-effective manner. The Hortonworks Data Platform consists of the essential set of Apache Hadoop projects including MapReduce, Hadoop Distributed File System (HDFS), HCatalog, Pig, Hive, HBase, ZooKeeper and Ambari. Hortonworks is the major contributor of code and patches to many of these projects. These projects have been integrated and tested as part of the Hortonworks Data Platform release process and installation and configuration tools have also been included.

Unlike other providers of platforms built using Apache Hadoop, Hortonworks contributes 100% of our code back to the Apache Software Foundation. The Hortonworks Data Platform is Apache-licensed and completely open source. We sell only expert technical support, training and partner-enablement services. All of our technology is, and will remain free and open source.

Please visit the Hortonworks Data Platform page for more information on Hortonworks technology. For more information on Hortonworks services, please visit either the Support or Training page. Feel free to Contact Us directly to discuss your specific needs.

Table of Contents

List of Tables

1. Introduction

HDP Search is a full-text search server, designed for enterprise-level performance, flexibility, scalability, and fault-tolerance. HDP Search exposes REST-like HTTP/XML and JSON APIs for use with a wide range of programming languages.

HDP Search includes:

- Apache Solr 7.4.0 (for HDP Search 5.0)

- Banana 1.6.12

- JARs for integration with Hadoop, Spark, Hive and Pig.

> ### Note
>
> **HDP Search is a separate product, not packaged with the HDP platform.**

The high-level steps for using HDP search are as follows:

1. Install and deploy HDP Search, either manually or by using Ambari.

2. Ingest documents from sources such as HDFS.

3. Index the data. Documents, and updates to documents, will be available for search almost immediately after being indexed.

4. Perform a wide range of basic and advanced operations on the indexed documents.

Resources:

- This document describes software requirements for HDP Search, followed by installation instructions for specific operating systems.

- For information about using Solr, see the Solr Reference Guide and the Solr tutorial.

2. HDP Search 5.0

2.1. HDP Search 5.0 Release Notes

The HDP Search 5.0 Release Notes summarize and describe the following information released in HDP Search 5.0:

- HDP Search 5.0 New Features [2]

- HDP Search 5.0 Behavioral Changes [2]

- HDP Search 5.0 Apache Solr Version Information [2]

- HDP Search 5.0 Known Issues [4]

Note

HDP Search is a separate product and does NOT come with the HDP platform.

2.1.1. HDP Search 5.0 New Features

HDP 5.0 includes no significant new features.

2.1.2. HDP Search 5.0 Behavioral Changes

HDP Search 5.0 introduces no significant behavioral changes.

2.1.3. HDP Search 5.0 Apache Solr Version Information

HDP Search 5.0 is based on Apache Solr 7.4 (Apache Solr Release Notes).

The following patches are applied on top of Apache Solr 7.4.

Apache Patch	Link to the Jira
SOLR-12343	https://issues.apache.org/jira/browse/SOLR-12343 JSON Field Facet refinement can return incorrect counts/stats for sorted buckets.
SOLR-12516	https://issues.apache.org/jira/browse/SOLR-12516 JSON range facets can incorrectly refine subfacets for buckets
SOLR-12541	https://issues.apache.org/jira/browse/SOLR-12541 Metrics handler throws an error if there are transient cores.
SOLR-12594	https://issues.apache.org/jira/browse/SOLR-12594 MetricsHistoryHandler.getOverseerLeader fails when hostname contains hyphen.
SOLR-12683	https://issues.apache.org/jira/browse/SOLR-12683 HashQuery will throw an exception if more than 4 partitionKeys is specified.
SOLR-12704	https://issues.apache.org/jira/browse/SOLR-12704 Guard AddSchemaFieldsUpdateProcessorFactory against null field names and field values.
SOLR-12750	https://issues.apache.org/jira/browse/SOLR-12750 Migrate API should lock the collection instead of shared.

Apache Patch	Link to the Jira
SOLR-12765	https://issues.apache.org/jira/browse/SOLR-12765 Incorrect format of JMX cache stats.
SOLR-12836	https://issues.apache.org/jira/browse/SOLR-12836 ZkController creates a cloud solr client with no connection or read timeouts.
SOLR-12615	https://issues.apache.org/jira/browse/SOLR-12615 HashQParserPlugin won't throw an NPE for string hash key and documents with empty value.
SOLR-12674	https://issues.apache.org/jira/browse/SOLR-12674 RollupStream should not use the HashQueryParser for 1 worker.
SOLR-2834	https://issues.apache.org/jira/browse/SOLR-2834 Handle CharacterFilters in Solr.
SOLR-8207	https://issues.apache.org/jira/browse/SOLR-8207 Collections with underscores in name no longer cause a crash the Cloud->Nodes UI .
SOLR-8207	https://issues.apache.org/jira/browse/SOLR-8207 Add "Nodes" view to the Admin UI "Cloud" tab, listing nodes and key metrics .
SOLR-8207	https://issues.apache.org/jira/browse/SOLR-8207 Nodes view support for shard_1_1_1 format and replica1, replica_1 format. Show core state in label if not "active" .
SOLR-12570	https://issues.apache.org/jira/browse/SOLR-12570 OpenNLPExtractNamedEntitiesUpdateProcessor cannot support multi fields because pattern replacement does not work correctly.
SOLR-12597	https://issues.apache.org/jira/browse/SOLR-12597 Migrate API should fail requests that do not specify split.key parameter.
SOLR-12649	https://issues.apache.org/jira/browse/SOLR-12649 CloudSolrClient retries requests unnecessarily exception from server.
SOLR-12670	https://issues.apache.org/jira/browse/SOLR-12670 RecoveryStrategy logs wrong wait time when retrying recovery.
SOLR-12679	https://issues.apache.org/jira/browse/SOLR-12679 MiniSolrCloudCluster.stopJettySolrRunner should remove jetty from the internal list.
SOLR-12679	https://issues.apache.org/jira/browse/SOLR-12679 MiniSolrCloudCluster.startJettySolrRunner method should not add a duplicate jetty instance to the list.
SOLR-12770	https://issues.apache.org/jira/browse/SOLR-12770 Make it possible to configure a host whitelist for distributed search.
SOLR-12776	https://issues.apache.org/jira/browse/SOLR-12776 Setting of TMP in solr.cmd causes invisibility of Solr to JDK tools.
SOLR-12723	https://issues.apache.org/jira/browse/SOLR-12723 Reduce object creation in HashBasedRouter.
SOLR-13793	https://issues.apache.org/jira/browse/SOLR-13793 Limiting number of forwards to total replicas in collection to avoid deadly forwarding loops.
SOLR-13971	https://issues.apache.org/jira/browse/SOLR-13971 Velocity custom template RCE vulnerability

Cloudera Internal Patch	Description
HDPSRC-1	SolrSslContextFactory is not on the classpath of Solr service and fails to load when SSL is enabled.
HDPSRC-2	Solr jaas.conf file name mismatch.
HDPSRC-5	Boolean literals case mismatch in params.py.
HDPSRC-6	A few of the advanced configuration entries have missing display names.
HDPSRC-12	Banana UI added as exception to Kerberos.
HDPSRC-16	Hive Serde dependencies updated with HDP 3.1.x.

2.1.4. HDP Search 5.0 Known Issues

Issue Description:

When SSL or Kerberos is enabled Solr services fail to start with MPack version 5.0.0.

Workaround:

Download MPack version 5.0.0.7 or later and upgrade your MPack with Ambari.

• Obtain the latest MPack

```
wget https://archive.cloudera.com/p/HDP-SOLR/5.0.0.7-401/repos/centos7/
mpack/cloudera-hdp-solr-mpack-5.0.0.7.tar.gz
```

• Upgrade the installed MPack to 5.0.0.7

```
ambari-server install-mpack --mpack=cloudera-hdp-solr-mpack-5.0.0.7.tar.gz
```

• Restart Ambari server

```
ambari-server restart
```

Issue Description:

When Serde jar is added using ADD JAR command in Hive CLI/Beeline, External table created in Hive for Solr stores NULL in place of all data inserted. ie: Select query on this external table returns NULL.

Workaround:

To get around this issue, you must add Serde jar in Hive's class path. This can be done in multiple ways. The one we recommend is :

• create an "auxlib" directory in /usr/hdp/current/hive-server2

• cp serde jar (/usr/cloudera-hdp-solr/current/cloudera-hdp-solr/hive/solr-hive-serde-4.0.1.jar) to auxlib directory

• Restart Hive

Issue Description:

• On kerberized clusters, accessing the Banana UI gives a Kerberos Replay Error and fails to load.

Workaround:

- Disable the jvm kerberos replay cache for the solr process instance. Note this does not affect the global kerberos replay cache for the KDC, and therefore other services.

To accomplish this:

Go to Ambari UI -> Solr -> Configs -> Advanced solr-config-env -> solr.in.sh.template, then add:

```
SOLR_OPTS="$SOLR_OPTS -Dsun.security.krb5.rcache=none"
```

Then restart the Solr service.

2.2. HDP Search 5.0 - Getting Ready

This section describes information and materials that you should get ready before installing HDP Search 5.0.

- HDP Search 5.0 HDP Requirements [5]

- Using a Local Repository [6]

2.2.1. HDP Search 5.0 Minimum System Requirements

To use HDP Search, your system must meet the following minimum requirements.

2.2.1.1. HDP Search 5.0 Operating System Requirements

HDP Search 5.0 is supported on the following operating systems:

- 64-bit CentOS 7

- 64-bit Red Hat Enterprise Linux (RHEL) 7

2.2.1.2. JDK Requirements - HDP Search 5.0

HDP Search 5.0 requires Oracle or OpenJDK Java 1.8 or higher.

Make sure your `$JAVA_HOME` and `$PATH` variables are set to the correct version; for example:

```
export JAVA_HOME=/usr/java/default
export PATH=$JAVA_HOME/bin:$PATH
```

2.2.1.3. HDP Search 5.0 HDP Requirements

HDP Search 5.0 is tested and certified with HDP 3.0.x and the following HDP components.

Component	Version
Apache Hadoop	3.1.0
Apache Hive*	3.0.0

Component	Version
Apache Pig	0.16.0
Apache Spark	2.3.1

* We have certified Hive with Tez only and not Map Reduce.

2.2.2. Using a Local Repository

Local repositories are frequently used in enterprise clusters that have limited outbound internet access. In these scenarios, local packages provide more governance and better installation performance. Local repositories are used during installation and for post-installation cluster operations such as service start and restart operations.

The following sections describe steps for setting up and using a local repository for HDP Search.

> ### Important
>
> The following instructions assume that you have already obtained and prepared the appropriate Ambari repository and other HDP Stack repositories as described in Using a Local Repository in the *Ambari Installation Guide*.

2.2.2.1. Obtaining the HDP Search Repository

Use the appropriate link to download or reposync the HDP Search 5.0 repository.

OS	Format	URL
RedHat or CentOS 7	Repo File	https://archive.cloudera.com/p/HDP-SOLR/5.0.0.7-401/repos/centos7/cloudera-hdp-solr.repo
	Tarball	https://archive.cloudera.com/p/HDP-SOLR/5.0.0.7-401/repos/centos7/CLOUDERA-HDP-SOLR-5.0.0.7-401-centos7.tar.gz (md5, asc)

2.2.2.2. Setting Up the HDP Search 5.0 Local Repository

The following instructions assume that you have obtained the appropriate HDP Search repository for your operating system, and that you have created an HTTP server and a web server directory as described in Getting Started Setting Up a Local Repository in the *Ambari Installation Guide*.

Based on your Internet access, choose one of the following options:

• **No Internet Access**: This option involves downloading the repository tarball, moving the tarball to the selected mirror server in your cluster, and extracting files to create the repository.

• **Temporary Internet Access**: This option involves using your temporary Internet access to synchronize (using reposync) the software package to your selected mirror server and creating the repository.

The following subsections describe how to set up each option.

2.2.2.2.1. Setting Up a Local Repository with No Internet Access

If you are setting up a local repository with no internet access, complete the following steps:

1. Copy the repository tarball to the web server directory, and untar the file.

   a. Browse to the web server directory you created.

      ```
      cd /var/www/html/
      ```

   b. Untar the repository tarballs to the following location:

      ```
      <web.server.directory>/CLOUDERA-HDP-SOLR-<latest.version>/
      repos/<OS>
      ```

      where `<web.server.directory>`, `<OS>`, and `<latest.version>` represent the web server document root directory, the latest version of CLOUDERA-HDP-SOLR, the operating system type, and the release version, respectively.

2. Confirm that you can browse to the newly created local repository:

   ```
   http://<web.server>/CLOUDERA-HDP-SOLR-<latest.version>/<OS>/
   ```

   where `<web.server>` is the fully-qualified domain name of your web server host, and `<OS>` is centos6, centos7, sles11, ubuntu12, ubuntu14, debian6, or debian7.

   Important: Be sure to record this URL. You will need it when installing HDP Search.

3. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all nodes in your cluster.

   a. Install the plug-in.

      ```
      yum install yum-plugin-priorities
      ```

   b. Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

      ```
      [main]
      enabled=1
      gpgcheck=0
      ```

2.2.2.2.2. Setting up a Local Repository With Temporary Internet Access

If you are setting up a local repository with temporary internet access, complete the following steps:

1. Put the repository configuration files for HDP Search in place on the host.

2. Confirm availability of the repositories.

   ```
   yum repolist
   ```

3. Synchronize the repository contents to your mirror server.

     a. Browse to the web server directory:

```
cd /var/www/html
```

     b. Create a directory for HDP Search, and then change to the new directory:

```
mkdir -p cloudera-hdp-solr/<OS>
cd cloudera-hdp-solr/<OS>
```

     `<OS>` is centos7 or RHEL 7

     c. `reposync` the directory:

```
reposync -r CLOUDERA-HDP-SOLR-<latest.version>
```

4. Generate the repository metadata for HDP Search:

```
createrepo <web.server.directory>/CLOUDERA-HDP-SOLR-
<latest.version>/<OS>/
```

5. Using the following URL, confirm that you can browse to the newly created repository:

```
http://<web.server>/CLOUDERA-HDP-SOLR-<latest.version>/<OS>/
```

where `<web.server>` is the fully-qualified domain name of the web server host, and `<OS>` is centos7 or RHEL7.

Important: Be sure to record the Base URL. You will need it when installing HDP Search.

6. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all the nodes in your cluster.

    • Install the plug-in.

```
yum install yum-plugin-priorities
```

    • Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

```
[main]
enabled=1
gpgcheck=0
```

2.2.2.3. Installing, Configuring, and Deploying HDP Search 5.0

When finished with the preceding steps, run the Ambari Install Wizard to install, configure, and deploy HDP Search.

2.3. Installing HDP Search 5.0 Management Pack

Prerequisites

Before installing HDP Search 5.0, you must disable the yum repo priorities plug-in on all hosts that will have the HDP Search Solr component installed, as follows:

```
# vi /etc/yum/pluginconf.d/priorities.conf
```

```
[main]
enabled = 0
```

Complete the following steps to download and install the HDP Search 5.0 Management Pack.

1. Download the Ambari management pack to the Ambari Server host.

   In this example, `/tmp` is a temporary directory that stores the management pack before it is installed.

   ```
   cd /tmp
   wget https://<username>:<password>@archive.cloudera.com/p/HDP-SOLR/5.0.0.
   7-401/repos/centos7/mpack/cloudera-hdp-solr-mpack-5.0.0.7.tar.gz
   ```

2. Install the management pack on the Ambari Server host, using the following command:

   ```
   # ambari-server install-mpack --mpack=/tmp/cloudera-hdp-solr-mpack-5.0.0.7.
   tar.gz
   ```

   You should see the following output:

   ```
   Using python /usr/bin/python
   Installing management pack
   Ambari Server 'install-mpack' completed successfully.
   ```

   The management pack has now been added to Ambari.

3. Restart Ambari server.

   ```
   # ambari-server restart
   ```

4. Add the Solr service, either during initial cluster installation using the Ambari installation wizard or after cluster deployment.

2.4. Installing HDP Search 5.0 Manually

HDP Search 5.0 packages are located in the CLOUDERA-HDP-SOLR 5.0 repository.

To install HDP Search 5.0, run the appropriate commands for your operating system on all cluster nodes that will run Solr.

Note

RPM packages for CentOS/RHEL Linux are signed, so you will need to add the gpg key to your server. That step is included in the following instructions.

```
rpm --import https://<username>:<password>@archive.cloudera.com/p/HDP-SOLR/5.
0.0.7-401/repos/centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget https://<username>:<password>@archive.cloudera.com/p/HDP-SOLR/5.0.0.
7-401/repos/centos7/cloudera-hdp-solr.repo
yum install cloudera-hdp-solr
```

3. HDP Search 4.0

3.1. HDP Search 4.0 Release Notes

The HDP Search 4.0 Release Notes summarize and describe the following information released in HDP Search 4.0:

> **Note**
>
> **HDP Search is a separate product and does NOT come with the HDP platform.**

3.1.1. HDP Search 4.0 New Features

HDP 4.0 includes the following new feature:

Table 3.1. HDP Search

Feature	Description
Apache Solr 7.4	Upgraded Solr to Solr 7.4
Updated Hadoop Solr connectors to support HDP 3.0	• HDFS (updated to support Hadoop 3.1.0) • Hive (updated to support Hive 3.0) • Spark (updated to support Spark 2.3.1)

3.1.2. HDP Search 4.0 Behavioral Changes

HDP Search 4.0 introduces the following change in behavior as compared to previous HDP Search versions:

Table 3.2. HDP Search

Description	Reference
Add HBase Solr Connector	The HBase Solr connector from Lucidworks is no longer supported, and has not been included in this release.

More Information

Ambari 2.7.0 Behavioral Changes

3.1.3. HDP Search 4.0 Apache Solr Version Information

HDP Search 4.0 is based on Apache Solr 7.4 (Apache Solr Release Notes).

3.1.4. HDP Search 4.0 Known Issues

Issue Description:

When Serde jar is added using ADD JAR command in Hive CLI/Beeline, External table created in Hive for Solr stores NULL in place of all data inserted. ie; Select query on this external table returns NULL.

Workaround:

To get around this issue, you must add Serde jar in Hive's class path. This can be done in multiple ways. The one we recommend is :

- create an "auxlib" directory in /usr/hdp/current/hive-server2

- cp serde jar (/opt/lucidworks-hdpsearch/hive/solr-hive-serde-4.0.0.jar) to auxlib directory

- Restart Hive

Issue Description:

- On kerberized clusters, accessing the Banana UI gives a Kerberos Replay Error and fails to load.

Workaround:

- Disable the jvm kerberos replay cache for the solr process instance. Note this does not affect the global kerberos replay cache for the KDC, and therefore other services.

To accomplish this:

Go to Ambari UI -> Solr -> Configs -> Advanced solr-config-env -> solr.in.sh.template, then add:

```
SOLR_OPTS="$SOLR_OPTS -Dsun.security.krb5.rcache=none"
```

Then restart the Solr service.

3.2. HDP Search 4.0 - Getting Ready

This section describes information and materials that you should get ready before installing HDP Search 4.0.

- HDP Search 4.0 Minimum System Requirements [12]

- Using a Local Repository [12]

3.2.1. HDP Search 4.0 Minimum System Requirements

To use HDP Search, your system must meet the following minimum requirements.

3.2.1.1. HDP Search 4.0 Operating System Requirements

HDP Search 4.0 is supported on the following operating systems:

• 64-bit CentOS 7

• 64-bit Red Hat Enterprise Linux (RHEL) 7

3.2.1.2. JDK Requirements - HDP Search 4.0

HDP Search 4.0 requires Oracle or OpenJDK Java 1.8 or higher.

Make sure your $JAVA\_HOME$ and $PATH$ variables are set to the correct version; for example:

```
export JAVA_HOME=/usr/java/default
export PATH=$JAVA_HOME/bin:$PATH
```

3.2.1.3. HDP Search 4.0 HDP Requirements

HDP Search 4.0 is tested and certified with HDP 3.0.x and the following HDP components.

Component	Version
Apache Hadoop	3.1.0
Apache Hive*	3.0.0
Apache Pig	0.16.0
Apache Spark	2.3.1

* We have certified Hive with Tez only and not Map Reduce.

3.2.2. Using a Local Repository

Local repositories are frequently used in enterprise clusters that have limited outbound internet access. In these scenarios, local packages provide more governance and better installation performance. Local repositories are used during installation and for post-installation cluster operations such as service start and restart operations.

The following sections describe steps for setting up and using a local repository for HDP Search.

⚠️ **Important**

The following instructions assume that you have already obtained and prepared the appropriate Ambari repository and other HDP Stack repositories as described in Using a Local Repository in the *Ambari Installation Guide*.

3.2.2.1. Obtaining the HDP Search 4.0 Repository

Use the link appropriate for your OS family to download or reposync the HDP Search 4.0 repository.

3.2.2.2. Setting Up the HDP Search 4.0 Local Repository

The following instructions assume that you have obtained the appropriate HDP Search repository for your operating system, and that you have created an HTTP server and a web server directory as described in Getting Started Setting Up a Local Repository in the *Ambari Installation Guide*.

Based on your Internet access, choose one of the following options:

• **No Internet Access**: This option involves downloading the repository tarball, moving the tarball to the selected mirror server in your cluster, and extracting files to create the repository.

• **Temporary Internet Access**: This option involves using your temporary Internet access to synchronize (using reposync) the software package to your selected mirror server and creating the repository.

The following subsections describe how to set up each option.

3.2.2.2.1. Setting Up a Local Repository with No Internet Access

If you are setting up a local repository with no internet access, complete the following steps:

1. Copy the repository tarball to the web server directory, and untar the file.

   a. Browse to the web server directory you created.

      • **For RHEL/CentOS Linux**:

      ```
      cd /var/www/html/
      ```

      • **For SLES**:

      ```
      cd /srv/www/htdocs/rpms
      ```

      • **For Debian/Ubuntu**:

      ```
      cd /var/www/html/
      ```

   b. Untar the repository tarballs to the following location:

      ```
      <web.server.directory>/HDP-SOLR-<latest.version>/repos/<OS>
      ```

      where `<web.server.directory>`, `<OS>`, and `<latest.version>` represent the web server document root directory, the latest version of HDP-SOLR, the operating system type, and the release version, respectively.

2. Confirm that you can browse to the newly created local repository:

   ```
   http://<web.server>/HDP-SOLR-<latest.version>/<OS>/
   ```

   where `<web.server>` is the fully-qualified domain name of your web server host, and `<OS>` is centos6, centos7, sles11, ubuntu12, ubuntu14, debian6, or debian7.

Important: Be sure to record this URL. You will need it when installing HDP Search.

3. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all nodes in your cluster.

   a. Install the plug-in.

      For RHEL and CentOS 7:

      ```
      yum install yum-plugin-priorities
      ```

   b. Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

      ```
      [main]
      enabled=1
      gpgcheck=0
      ```

3.2.2.2.2. Setting up a Local Repository With Temporary Internet Access

If you are setting up a local repository with temporary internet access, complete the following steps:

1. Put the repository configuration files for HDP Search in place on the host.

2. Confirm availability of the repositories.

   For RHEL/CentOS 7:

   ```
   yum repolist
   ```

3. Synchronize the repository contents to your mirror server.

   a. Browse to the web server directory:

      For RHEL/CentOS 7:

      ```
      cd /var/www/html
      ```

   b. Create a directory for HDP Search, and then change to the new directory:

      ```
      mkdir -p hdp-solr/<OS>
      cd hdp-solr/<OS>
      ```

      `<OS>` is centos7 or RHEL 7

   c. `reposync` the directory:

      ```
      reposync -r HDP-SOLR-<latest.version>
      ```

4. Generate the repository metadata for HDP Search:

   ```
   createrepo <web.server.directory>/HDP-SOLR-<latest.version>/
   <OS>/
   ```

5. Using the following URL, confirm that you can browse to the newly created repository:

```
http://<web.server>/HDP-SOLR-<latest.version>/<OS>/
```

where `<web.server>` is the fully-qualified domain name of the web server host, and `<OS>` is centos7 or RHEL7.

Important: Be sure to record the Base URL. You will need it when installing HDP Search.

6. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all the nodes in your cluster.

   • Install the plug-in.

   ```
   yum install yum-plugin-priorities
   ```

   • Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

   ```
   [main]
   enabled=1
   gpgcheck=0
   ```

3.2.2.3. Installing, Configuring, and Deploying HDP Search 4.0

When finished with the preceding steps, run the Ambari Install Wizard to install, configure, and deploy HDP Search. For more information, refer to the HDP Search *Ambari Installation Guide*.

For information about the HDP Search Directory Layout, refer to the *HDP Search Directory Layout*.

3.3. Installing HDP Search 4.0 Management Pack

Prerequisites

Before installing HDP Search 4.0, you must disable the yum repo priorities plug-in on all hosts that will have the HDP Search Solr component installed, as follows:

```
# vi /etc/yum/pluginconf.d/priorities.conf
[main]
enabled = 0
```

Complete the following steps to download and install the HDP Search 4.0 Management Pack.

1. Download the Ambari management pack to the Ambari Server host.

   In this example, `/tmp` is a temporary directory that stores the management pack before it is installed.

   ```
   cd /tmp
   wget https://archive.cloudera.com/p/HDP-SOLR/hdp-solr-ambari-mp/solr-
   service-mpack-4.0.0.tar.gz
   ```

2. Install the management pack on the Ambari Server host, using the following command:

   ```
   # ambari-server install-mpack --mpack=/tmp/solr-service-mpack-4.0.0.tar.gz
   ```

You should see the following output:

```
Using python /usr/bin/python
Installing management pack
Ambari Server 'install-mpack' completed successfully.
```

The management pack has now been added to Ambari.

3. Restart Ambari server.

```
# ambari-server restart
```

4. Add the Solr service, either during initial cluster installation using the Ambari installation wizard or after cluster deployment.

For information about Solr configuration options, refer to the *Startup Option Reference* section of the *Lucidworks Ambari Installation Guide.*

3.4. Installing HDP Search 4.0 Manually

HDP Search 4.0 packages are located in the HDP-SOLR 4.0 repository.

To install HDP Search 4.0, run the appropriate commands for your operating system on all cluster nodes that will run Solr.

Note

RPM packages for CentOS/RHEL Linux are signed, so you will need to add the gpg key to your server. That step is included in the following instructions.

CentOS/RHEL Linux 7:

```
rpm --import https://archive.cloudera.com/p/HDP-SOLR/4.0.0-400/repos/centos7/
RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget https://archive.cloudera.com/p/HDP-SOLR/4.0.0-400/repos/centos7/hdp-solr.
repo
yum install lucidworks-hdpsearch
```

4. HDP Search 3.0

- HDP Search 3.0 Release Notes [17]

- HDP Search 3.0 - Getting Ready [18]

- Installing HDP Search 3.0 Management Pack [22]

- Upgrading HDP Search [26]

4.1. HDP Search 3.0 Release Notes

The HDP Search 3.0 Release Notes summarize and describe the following information released in HDP Search 3.0:

- HDP Search 3.0 New Features [17]

- HDP Search 3.0 Behavioral Changes [17]

- HDP Search 3.0 Apache Solr Version Information [18]

- HDP Search 3.0 Known Issues [18]

> **Note**
>
> **HDP Search is a separate product and does NOT come with the HDP platform.**

4.1.1. HDP Search 3.0 New Features

HDP 3.0 includes the following new feature:

Table 4.1. HDP Search

Feature	Description
Apache Solr 6.6.2	Support for Apache Solr 6.6.2 has been added to the Management Pack

4.1.2. HDP Search 3.0 Behavioral Changes

HDP Search 3.0 introduces the following change in behavior as compared to previous HDP Search versions:

Table 4.2. HDP Search

Description	Reference
Storm Solr Connector	The Storm Solr connector from Lucidworks is no longer supported, and has not been included in this release.

More Information

Ambari 2.6.0 Behavioral Changes

4.1.3. HDP Search 3.0 Apache Solr Version Information

HDP Search 3.0 is based on Apache Solr 6.6.2 (Apache Solr Release Notes).

4.1.4. HDP Search 3.0 Known Issues

Issue Description:

Please choose Map Reduce as the execution engine for Hive queries related to the Solr index. Tez will NOT work because Lucidwork's Hive SerDe does not yet support Hive 2 officially. It has been roadmapped for a future release. Currently, only Tez on Hive 1.x is supported.

Issue Description:

• On kerberized clusters, accessing the Banana UI gives a Kerberos Replay Error and fails to load.

Workaround:

• Disable the jvm kerberos replay cache for the solr process instance. Note this does not affect the global kerberos replay cache for the KDC, and therefore other services.

To accomplish this:

Go to Ambari UI -> Solr -> Configs -> Advanced solr-config-env -> solr.in.sh.template, then add:

```
SOLR_OPTS="$SOLR_OPTS -Dsun.security.krb5.rcache=none"
```

Then restart the Solr service.

4.2. HDP Search 3.0 - Getting Ready

This section describes information and materials that you should get ready before installing HDP Search 3.0.

• HDP Search 3.0 Minimum System Requirements [18]

• Using a Local Repository [19]

4.2.1. HDP Search 3.0 Minimum System Requirements

To use HDP Search, your system must meet the following minimum requirements.

4.2.1.1. HDP Search 3.0 Operating System Requirements

HDP Search 3.0 is supported on the following operating systems:

• 64-bit CentOS 6 and 7

• 64-bit Red Hat Enterprise Linux (RHEL) 6 and 7

- 64-bit Oracle Linux 6 and 7

- 64-bit SUSE Linux Enterprise Server (SLES) 11, SP3/SP4

- 64-bit SUSE Linux Enterprise Server (SLES) 12

- 64-bit Debian 7

- 64-bit Ubuntu 12 and 14

4.2.1.2. JDK Requirements - HDP Search 3.0

HDP Search 3.0 requires Oracle or OpenJDK Java 1.8 or higher.

Make sure your $JAVA_HOME and $PATH variables are set to the correct version; for example:

```
export JAVA_HOME=/usr/java/default
export PATH=$JAVA_HOME/bin:$PATH
```

4.2.1.3. HDP Search 3.0 HDP Requirements

HDP Search 3.0 is tested and certified with HDP 2.6.x and the following HDP components.

Component	Version
Apache Hadoop	2.7.3
Apache HBase	1.1.2
Apache Hive	1.2.1
Apache Pig	0.16.0
Apache Spark	2.2.0

4.2.2. Using a Local Repository

Local repositories are frequently used in enterprise clusters that have limited outbound internet access. In these scenarios, local packages provide more governance and better installation performance. Local repositories are used during installation and for post-installation cluster operations such as service start and restart operations.

The following sections describe steps for setting up and using a local repository for HDP Search.

Important

The following instructions assume that you have already obtained and prepared the appropriate Ambari repository and other HDP Stack repositories as described in Using a Local Repository in the *Ambari Installation Guide*.

4.2.2.1. Obtaining the HDP Search 3.0 Repository

Use the link appropriate for your OS family to download or reposync the HDP Search 3.0 repository.

4.2.2.2. Setting Up the HDP Search 3.0 Local Repository

The following instructions assume that you have obtained the appropriate HDP Search repository for your operating system, and that you have created an HTTP server and a web server directory as described in Getting Started Setting Up a Local Repository in the *Ambari Installation Guide*.

Based on your Internet access, choose one of the following options:

• **No Internet Access**: This option involves downloading the repository tarball, moving the tarball to the selected mirror server in your cluster, and extracting files to create the repository.

• **Temporary Internet Access**: This option involves using your temporary Internet access to synchronize (using reposync) the software package to your selected mirror server and creating the repository.

The following subsections describe how to set up each option.

4.2.2.2.1. Setting Up a Local Repository with No Internet Access

If you are setting up a local repository with no internet access, complete the following steps:

1. Copy the repository tarball to the web server directory, and untar the file.

   a. Browse to the web server directory you created.

      • **For RHEL/CentOS/Oracle Linux**:

      ```
      cd /var/www/html/
      ```

      • **For SLES**:

      ```
      cd /srv/www/htdocs/rpms
      ```

      • **For Debian/Ubuntu**:

      ```
      cd /var/www/html/
      ```

   b. Untar the repository tarballs to the following location:

      ```
      <web.server.directory>/HDP-SOLR-<latest.version>/repos/<OS>
      ```

      where `<web.server.directory>`, `<OS>`, and `<latest.version>` represent the web server document root directory, the latest version of HDP-SOLR, the operating system type, and the release version, respectively.

2. Confirm that you can browse to the newly created local repository:

   ```
   http://<web.server>/HDP-SOLR-<latest.version>/<OS>/
   ```

   where `<web.server>` is the fully-qualified domain name of your web server host, and `<OS>` is centos6, centos7, sles11, ubuntu12, ubuntu14, debian6, or debian7.

   Important: Be sure to record this URL. You will need it when installing HDP Search.

3. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all nodes in your cluster.

    a. Install the plug-in.

- **For RHEL and CentOS 7**:

```
yum install yum-plugin-priorities
```

- **For RHEL and CentOS 6**:

```
yum install yum-plugin-priorities
```

    b. Edit the `/etc/yum/pluginconf.d/priorities.conf` file to add the following:

```
[main]
enabled=1
gpgcheck=0
```

4.2.2.2.2. Setting up a Local Repository With Temporary Internet Access

If you are setting up a local repository with temporary internet access, complete the following steps:

1. Put the repository configuration files for HDP Search in place on the host.

2. Confirm availability of the repositories.

- **For RHEL/CentOS/Oracle Linux**:

```
yum repolist
```

- **For SLES**:

```
zypper repos
```

- **For Debian/Ubuntu**:

```
Dpkg-list
```

3. Synchronize the repository contents to your mirror server.

    a. Browse to the web server directory:

- **For RHEL/CentOS/Oracle Linux**:

```
cd /var/www/html
```

- **For SLES**:

```
cd /srv/www/htdocs/rpms
```

- **For Debian/Ubuntu**:

```
cd /var/www/html
```

    b. Create a directory for HDP Search, and then change to the new directory:

```
mkdir -p hdp-solr/<OS>
cd hdp-solr/<OS>
```

<OS> is centos6, centos7, sles11, sles12, ubuntu12, ubuntu14, or debian7.

c. reposync the directory:

```
reposync -r HDP-SOLR-<latest.version>
```

4. Generate the repository metadata for HDP Search:

```
createrepo <web.server.directory>/HDP-SOLR-<latest.version>/
<OS>/
```

5. Using the following URL, confirm that you can browse to the newly created repository:

```
http://<web.server>/HDP-SOLR-<latest.version>/<OS>/
```

where <web.server> is the fully-qualified domain name of the web server host, and <OS> is centos6, centos7, sles11, sles12, ubuntu12, ubuntu14, or debian7.

Important: Be sure to record the Base URL. You will need it when installing HDP Search.

6. Optional: If you have multiple repositories configured in your environment, deploy the following plug-in on all the nodes in your cluster.

• Install the plug-in.

• **For RHEL and CentOS 7**:

```
yum install yum-plugin-priorities
```

• **For RHEL and CentOS 6**:

```
yum install yum-plugin-priorities
```

• Edit the /etc/yum/pluginconf.d/priorities.conf file to add the following:

```
[main]
enabled=1
gpgcheck=0
```

4.2.2.3. Installing, Configuring, and Deploying HDP Search 3.0

When finished with the preceding steps, run the Ambari Install Wizard to install, configure, and deploy HDP Search. For more information, refer to the HDP Search *Ambari Installation Guide*.

For information about Solr configuration options, refer to the *Startup Option Reference* section of the HDP Search *Ambari Installation Guide*.

4.3. Installing HDP Search 3.0 Management Pack

Prerequisites

Before installing HDP Search 3.0, you must disable the yum repo priorities plug-in on all hosts that will have the HDP Search Solr component installed, as follows:

```
# vi /etc/yum/pluginconf.d/priorities.conf
[main]
enabled = 0
```

Complete the following steps to download and install the HDP Search 3.0 Management Pack.

1. Download the Ambari management pack to the Ambari Server host.

   In this example, `/tmp` is a temporary directory that stores the management pack before it is installed.

   ```
   cd /tmp
   wget https://archive.cloudera.com/p/HDP-SOLR/hdp-solr-ambari-mp/solr-
   service-mpack-3.0.0.tar.gz
   ```

2. Install the management pack on the Ambari Server host, using the following command:

   ```
   # ambari-server install-mpack --mpack=/tmp/solr-service-mpack-3.0.0.tar.gz
   ```

   You should see the following output:

   ```
   Using python /usr/bin/python
   Installing management pack
   Ambari Server 'install-mpack' completed successfully.
   ```

   The management pack has now been added to Ambari.

3. Restart Ambari server.

   ```
   # ambari-server restart
   ```

4. Add the Solr service, either during initial cluster installation using the Ambari installation wizard or after cluster deployment.

For information about Solr configuration options, refer to the *Startup Option Reference* section of the *Lucidworks Ambari Installation Guide.*

4.4. Installing HDP Search 3.0 Manually

HDP Search 3.0 packages are located in the HDP-SOLR 3.0 repository.

To install HDP Search 3.0, run the appropriate commands for your operating system on all cluster nodes that will run Solr.

> **Note**
>
> RPM packages for CentOS/RHEL/Oracle Linux are signed, so you will need to add the gpg key to your server. That step is included in the following instructions.

• CentOS/RHEL/Oracle Linux 6:

```
rpm --import https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/
centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/centos6/hdp-
solr.repo
yum install lucidworks-hdpsearch
```

- CentOS/RHEL/Oracle Linux 7:

```
rpm --import https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/
centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/centos7/hdp-
solr.repo
yum install lucidworks-hdpsearch
```

- SUSE11SP3/SP4:

```
cd /etc/zypp/repos.d/
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/suse11sp3/hdp-
solr.repo
zypper install lucidworks-hdpsearch
```

- Ubuntu12:

```
cd /etc/apt/sources.list.d
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/ubuntu12/hdp-
solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Ubuntu14:

```
cd /etc/apt/sources.list.d
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/ubuntu14/hdp-
solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Debian6:

```
cd /etc/apt/sources.list.d
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/debian6/hdp-
solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Debian7:

```
cd /etc/apt/sources.list.d
wget https://archive.cloudera.com/p/HDP-SOLR/3.0.0-100/repos/debian7/hdp-
solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

⚠️ Important

For Debian or Ubuntu, if you see the following error during `apt-get update`:

```
W: GPG error: https://archive.cloudera.com HDP-SOLR Release: The
 following signatures couldn't be verified because the public key
 is not available: NO_PUBKEY B9733A7A07513CAD
```

Run the following commands:

```
apt-key adv --keyserver keyserver.ubuntu.com --recv-keys
 B9733A7A07513CAD
apt-get update
```

5. Upgrading HDP Search

This chapter describes the upgrade process from HDP Search 4.0 to HDP Search 5.0.

> ⚠️ **Important**
>
> This procedure is only applicable to HDP Search 4.0 to HDP Search 5.0 upgrades. Do not attempt to upgrade a cluster with HDP Search 3.0 or earlier using this method.

Prerequisites

Before upgrading HDP Search, you should ensure that all indexing has stopped for Solr and all connectors, including Hadoop, Hive, and via Pig.

1. Stop Solr service.

2. Remove Solr service from Ambari.

3. Upgrade Ambari to 2.7.3 and HDP 3.1.4.0.

4. Verify that Ambari and HDP upgrade suceeded.

5. Upgrade Ambari Metrics, Smartsense, and Infra Solr.

6. Upgrade the Stack by removing the HDP Search 4.0.0 mpack and installing the Cloudera HDP Search 5.0.0 mpack.

7. Add the Solr service to your cluster, using the Ambari Web UI. Make sure that ZooKeeper, HDFS, Ranger and memory settings are consistent with the old installation.

6. Applying Minor Upgrades to HDP Search

This chapter describes the procedure of applying update patches to HDP Search 5.0.

Prerequisites

Before updating HDP Search, you should ensure that all indexing has stopped for Solr and all connectors, including Hadoop, Hive, and via Pig.

1. Download the latest HDP Search Package:

```
cd /tmp
wget https://archive.cloudera.com/p/HDP-SOLR/5.0.0.7-401/repos/centos7/
mpack/cloudera-hdp-solr-mpack-5.0.0.7.tar.gz
```

2. Upgrade the Management Pack with the following command in Ambari:

```
ambari-server upgrade-mpack --mpack=cloudera-hdp-solr-mpack-mpack-5.0.0.7.
tar.gz --verbose
```

3. Restart Ambari to make the changes effective:

```
ambari-server restart
```

4. Log in to Ambari Web UI.

5. Go to **Stack and Versions** > **Versions** > **Manage Versions** and click on the actual stack version link.

6. Update the `CLOUDERAHDPSOLR` base URL.

   Set it to `https://archive.cloudera.com/p/HDP-SOLR/5.0.0.7-401/repos/centos7`

7. Click **Save** and confirm changes.

8. Return to Ambari Web UI and stop the Solr service.

9. Remove Solr service from Ambari.

10. Log in to the hosts that were running the Solr service and remove the Solr packages.

```
yum remove cloudera-hdp-solr
```

11. Add the Solr service to your cluster, using the Ambari Web UI.