Hortonworks Data Platform

Apache Solr Search Installation

Hortonworks Data Platform: Apache Solr Search Installation

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

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 consists of:

• Apache Solr 5.5.2

• Banana 1.6.12

• JARs for integration with Hadoop, Hive, HBase and Pig

• Software development kits for Storm and Spark

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 configuring indexes, ingesting documents, and searching indexed documents, see Getting Started with Solr.

• For detailed information about connecting to data sources and ingesting data on secure and non-secure clusters, see the Connector User Guide.

• For detailed information on using Banana with Solr, see the Guide to Banana Dashboards.

• For help troubleshooting issues with HDP Search, see Troubleshooting HDP Search.

2. Getting Ready

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

2.1. Meet Minimum System Requirements

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

2.1.1. Operating System Requirements

HDP Search 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 Debian 7

- 64-bit Ubuntu 12 and 14

2.1.2. JDK Requirements

HDP Search requires Oracle or OpenJDK Java 1.7 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
```

2.1.3. HDP Requirements

HDP Search is tested and certified with HDP 2.6 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	1.6.3

Component	Version
Apache Storm	1.1.0

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*.

2.2.1. Obtaining the HDP Search Repository

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

OS	Format	URL
RedHat 6, CentOS 6, or Oracle Linux 6	Repo File	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos6/hdp-solr.repo
	Tarball	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos6/HDP-SOLR-2.6-100-centos6.tar.gz (md5, asc)
RedHat 7, CentOS 7, or Oracle Linux 7	Repo File	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos7/hdp-solr.repo
	Tarball	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos7/HDP-SOLR-2.6-100-centos7.tar.gz (md5, asc)
SLES 11 SP3/SP4	Repo File	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/suse11sp3/hdp-solr.repo
	Tarball	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/suse11sp3/HDP-SOLR-2.6-100-suse11sp3.tar.gz (md5, asc)
Ubuntu 12	Repo File	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/ubuntu12/hdp-solr.list
	Tarball	http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/ubuntu12/HDP-SOLR-2.6-100-ubuntu12.tar.gz

OS	Format	URL
		(md5, asc)
Ubuntu 14	Repo File	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/ubuntu14/ hdp-solr.list
	Tarball	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/ubuntu14/ HDP-SOLR-2.6-100-ubuntu14.tar.gz (md5, asc)
Debian 6	Repo File	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/debian6/hdp-solr.list
	Tarball	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/debian6/ HDP-SOLR-2.6-100-debian6.tar.gz (md5, asc)
Debian 7	Repo File	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/debian7/hdp-solr.list
	Tarball	http://public-repo-1.hortonworks.com/ HDP-SOLR-2.6-100/repos/debian7/ HDP-SOLR-2.6-100-debian7.tar.gz (md5, asc)

2.2.2. Setting Up the HDP Search 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.

2.2.2.1. 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
```

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
  ```

2.2.3. Installing, Configuring, and Deploying HDP Search

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*.

3. Installing HDP Search Using Ambari

HDP Search can be installed using a Management Pack created by Lucidworks for Ambari 2.5. Complete the following steps to download and install the 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 http://public-repo-1.hortonworks.com/HDP-SOLR/hdp-solr-ambari-mp/solr-
   service-mpack-2.2.8.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-2.2.8.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. Add the Solr service, either during initial cluster installation using the Ambari installation wizard or after cluster deployment.

   ### Note

   It's important to note that during the Solr install, the Solr package has a set of prerequisite checks that are performed before the package can be completely installed. Specific checks such as disk space, installation directory access, solr users, and the Java setup are checked during this process. If any of these checks fail, they need to be manually remediated on each node where Solr will be installed before the installation task can be retried.

   In the case of the Java preinstall check failing, the easiest remediation is to login to each machine that Solr will be installed on, temporarily set the `JAVA_HOME` environmental variable, then then use **yum**/**zypper**/**apt-get** to install the package. For example on CentOS:

   ```
   export JAVA_HOME=/usr/jdk64/jdk1.8.0_77
   yum install lucidworks-hdpsearch
   ```

   Once all of the prerequisite checks have been satisfied and the package is installed, you can simply click "Retry" in the Ambari Web UI to move forward and complete the installation.

4. Restart Ambari server.

   ```
   # ambari-server restart
   ```

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

4. Installing HDP Search Manually

This chapter summarizes installation commands for installing HDP Search on an HDP 2.6 cluster that is not deployed and managed by Ambari. For detailed configuration instructions, see the HDP Search Manual Installation Guide.

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

To install HDP Search, 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 http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/
centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos6/
hdp-solr.repo
yum install lucidworks-hdpsearch
```

- CentOS/RHEL/Oracle Linux 7:

```
rpm --import http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/
centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/centos7/
hdp-solr.repo
yum install lucidworks-hdpsearch
```

- SUSE11SP3/SP4:

```
cd /etc/zypp/repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/suse11sp3/
hdp-solr.repo
zypper install lucidworks-hdpsearch
```

- Ubuntu12:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/ubuntu12/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Ubuntu14:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/ubuntu14/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Debian6:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/repos/debian6/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Debian7:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.6-100/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: http://public-repo-1.hortonworks.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
```