Hortonworks Data Platform

Apache Solr Search Installation

(Aug 29, 2016)

docs.hortonworks.com

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.

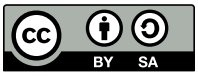

Table of Contents

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:

- Solr 5.5.1
- Banana 1.6.0
- 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 Banana User Guide.
- For help troubleshooting issues with HDP Search, see the Guide to Troubleshooting HDP Search Issues.

2. Minimum System Requirements

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

2.1. 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

2.2. JDK Requirements

HDP Search requires one of the following JDK versions:

- OracleJDK 1.7.0_67+
- OracleJDK 1.8.0_51+
- OpenJDK 1.7.0_67+
- OpenJDK 1.8.0_51+

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
```

2.3. HDP Requirements

HDP Search is tested and certified with HDP 2.5 and the following HDP components.

Component	Version
Apache Hadoop	2.7.1
Apache HBase	1.1.2
Apache Hive	1.2.1
Apache Pig	0.15.0
Apache Solr	5.5.1
Apache Spark	1.6.1
Apache Storm	0.9.4

3. Installing HDP Search Using Ambari

HDP Search can be installed using a new Management Pack created by Lucidworks for Ambari 2.4. 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-5.5.2.2.5.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-5.5.2.2.5.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.

Before using the management pack, you must create a definition for the HDP Search repository. To populate details for the repository, execute the following commands from the same Ambari Server host:

1. Using your preferred editor, edit the `/var/lib/ambari-server/resources/stacks/HDP/2.5/repos/repoinfo.xml` file.

   For the specific OS on which you will be installing HDP Search, add the HDP-SOLR-2.5-100 `<repo/>` definition (the third definition in the following example) to the existing `<os/>` section. The following example illustrates this for `redhat6`:

   ```
   <os family="redhat6">
       <repo>
         <baseurl>http://public-repo-1.hortonworks.com/HDP/centos6/2.x/updates/2.5.0.0</baseurl>
         <repoid>HDP-2.5</repoid>
         <reponame>HDP</reponame>
       </repo>
       <repo>
         <baseurl>http://public-repo-1.hortonworks.com/HDP-UTILS-1.1.0.21/repos/centos6</baseurl>
         <repoid>HDP-UTILS-1.1.0.21</repoid>
         <reponame>HDP-UTILS</reponame>
       </repo>
       <repo>
         <baseurl>http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/centos6/</baseurl>
         <repoid>HDP-SOLR-2.5-100</repoid>
         <reponame>HDP-SOLR</reponame>
       </repo>
   ```

```
</os>
```

Specify the appropriate <baseurl> for the operating system that you plan to use, as follows:

CentOS/RHEL/Oracle Linux 6	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/centos6
CentOS/RHEL/Oracle Linux 7	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/centos7
SUSE11SP3/SP4	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/suse11sp3
Ubuntu12	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/ubuntu12
Ubuntu14	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/ubuntu14
Debian6	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/debian6
Debian7	http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/debian7

2. Restart the Ambari Server, to recognize the new repository and the management pack:

```
ambari-server restart
```

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

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 simple click "Retry" in the Ambari Web UI to move forward and complete the installation.

For information about Solr configuration options, refer to the Startup Option Reference section of the Lucidworks Solr Service documentation.

4. Installing HDP Search Manually

This chapter summarizes installation commands for installing HDP Search on an HDP 2.5 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-2.5-100/repos/
centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/centos6/
hdp-solr.repo
yum install lucidworks-hdpsearch
```

- CentOS/RHEL/Oracle Linux 7:

```
rpm --import http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/repos/
centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.5-100/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-2.5-100/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-2.5-100/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-2.5-100/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-2.5-100/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-2.5-100/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
```