Hortonworks Data Platform

Search Guide

docs.cloudera.com

Hortonworks Data Platform: Search Guide

Copyright © 2012-2016 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

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 5.5.0.

• Banana 1.5 for analytical dashboards on Solr indexes.

• HDFS connector for indexing data from HDFS into Solr indexes.

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

1. Install and deploy HDP Search.

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.

2. Minimum System Requirements

To run 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

- 64-bit CentOS 7

- 64-bit Red Hat Enterprise Linux (RHEL) 6

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

- 64-bit Oracle Linux 6

- 64-bit Oracle Linux 7

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

- 64-bit Debian 6 (HDP support deprecated as of HDP 2.3.6)

- 64-bit Debian 7

- 64-bit Ubuntu 12

- 64-bit Ubuntu 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 requires deployment of HDP 2.3, HDP 2.4, or later version.

3. Installing HDP Search

HDP Search should be deployed on each node that runs HDFS. The HDP Search package is available as an RPM, located in the HDP-Utils repository.

This chapter summarizes installation commands for installing HDP Search on HDP 2.3 or HDP 2.4. For detailed installation and configuration instructions, see the HDP Search Installation Guide.

To install HDP Search, run the commands for your operating system on all cluster nodes:

> **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-2.3-100/repos/
centos6/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/centos6/
hdp-solr.repo
yum install lucidworks-hdpsearch
```

- CentOS/RHEL/Oracle Linux 7:

```
rpm --import http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/
centos7/RPM-GPG-KEY/RPM-GPG-KEY-Jenkins
cd /etc/yum.repos.d/
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/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-2.3-100/repos/suse11sp3/
hdp-solr.repo
zypper install lucidworks-hdpsearch
```

- Debian 6:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/debian6/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Debian 7:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/debian7/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Ubuntu 12:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/ubuntu12/
hdp-solr.list
apt-get update
apt-get install lucidworks-hdpsearch
```

- Ubuntu 14:

```
cd /etc/apt/sources.list.d
wget http://public-repo-1.hortonworks.com/HDP-SOLR-2.3-100/repos/ubuntu14/
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
```