

DAS: Core Concepts

Date of Publish: 2020-03-27



Contents

Data Analytics Studio overview.....	3
DAS architecture.....	3
Understanding and obtaining DAS binaries.....	4
Installation roadmap.....	6

Data Analytics Studio overview

Data Analytics Studio (DAS) is an application that provides diagnostic tools and intelligent recommendations to make the business analysts self-sufficient and productive with Hive.

DAS helps you to perform operations on Hive tables and provides recommendations for optimizing the performance of your queries. With DAS, you can search for queries, compose and execute queries, manage databases to create and edit tables, view reports and recommendations to optimize the query performance.

About DAS and DAS-Lite

DAS has two variants: DAS and DAS-Lite. DAS-Lite is a subset of the full-featured version of DAS. The following features that are available in DAS are not available in DAS-Lite:

- Query optimization recommendation
- Query comparison
- Table optimization recommendation
- Read and write report
- Join report

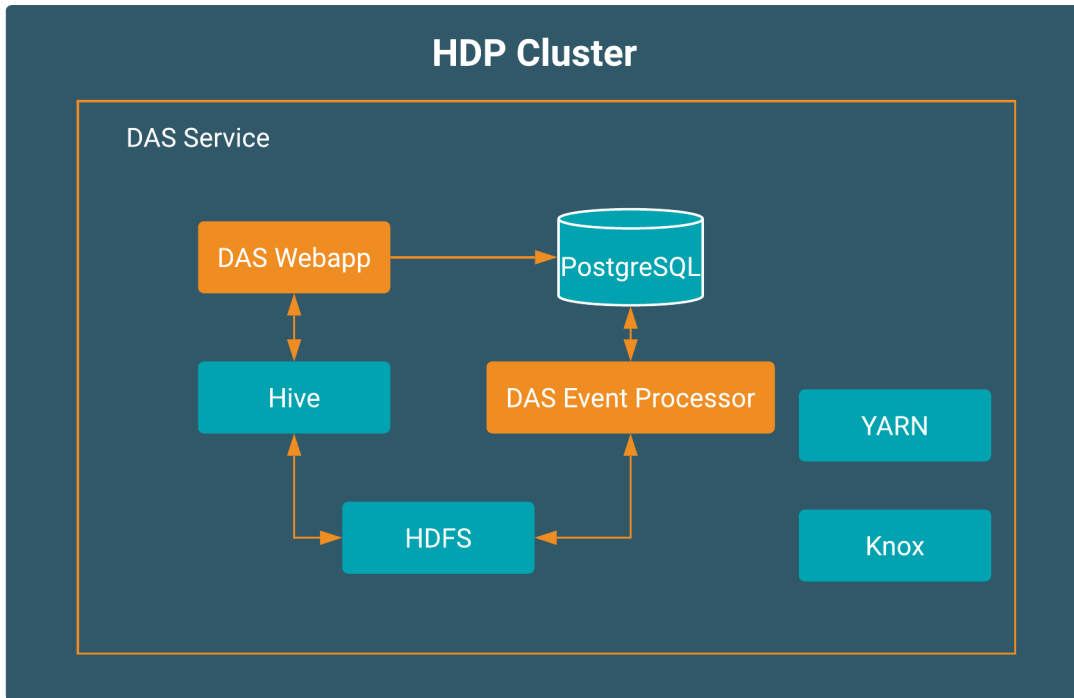


Note: DAS and DAS-Lite have the same release versions. However, the binaries for DAS and DAS-Lite are different for a given version. You must download the appropriate binary (DAS or DAS-Lite) depending on your purchase from Hortonworks before starting the installation.

DAS architecture

DAS cluster agent comprises of a DAS Webapp and an Event Processor. These components are installed on the Ambari cluster to use with DAS.

Following is a block diagram showing the various components in a typical DAS deployment:



The DAS Webapp serves the DAS UI, and it should be on a machine that is accessible to your end users, or you should have a proxy setup to access the DAS UI.

Understanding and obtaining DAS binaries

Access to Data Analytics Studio application for production purposes requires authentication. To access this application, you must have an active subscription agreement along with the required authentication credentials (namely, the username and password).

The authentication credentials are provided in an email sent to the customer account from Cloudera when a new license is issued. If you have an existing license with a Data Analytics Studio entitlement, you might not have received an email.



Note: If you do not have the authentication credentials, contact your account representative to receive the same.

Make sure that you download the DAS/DAS-Lite binaries according to the HDP version and operating system as given in the following table.

Table 1: Obtaining binaries: DAS deployed in a standalone cluster

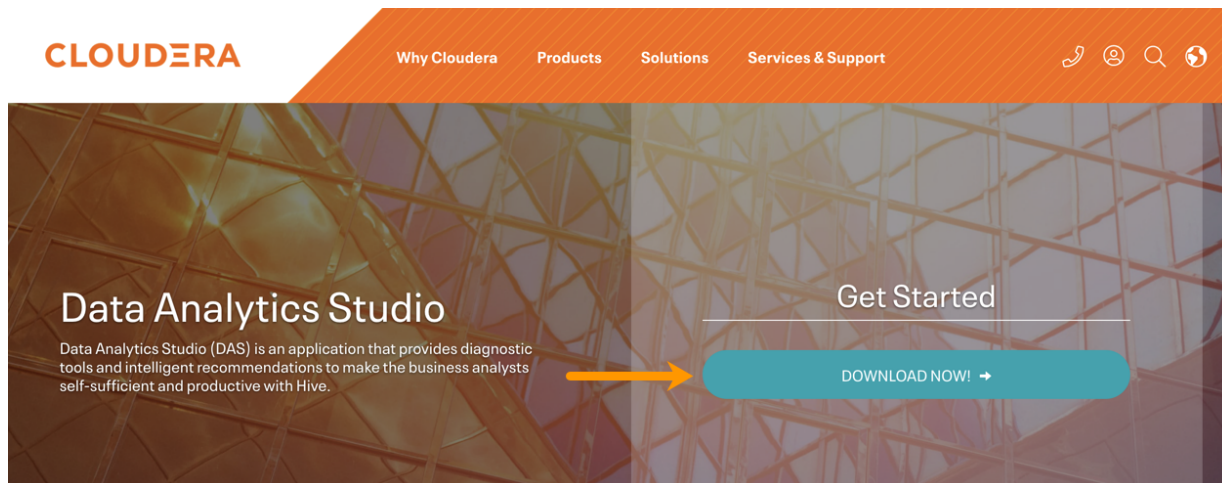
Versions and Operating Systems	CentOS 7	Ubuntu	Debian
HDP 3.x	<ul style="list-style-type: none"> HDP 3 MPack for CentOS HDP 3 DAS/DAS Lite tarball for CentOS 	<ul style="list-style-type: none"> HDP 3 MPack for Ubuntu HDP 3 DAS/DAS Lite tarball for Ubuntu 	<ul style="list-style-type: none"> HDP 3 MPack for Debian HDP 3 DAS/DAS Lite tarball for Debian

DAS is provided as RPMs in tarball repositories and as an Ambari Management Pack (MPack).

The DAS cluster components are installed in to your cluster using the Ambari MPack. The DAS cluster components are available for RHEL/CentOS/OEL 7, Debian 9 (HDP 3.x), and Ubuntu 16 platforms.

To download the Data Analytics Studio application, follow these steps:

1. Log into the cloudera.com account associated with the Data Analytics Studio license agreement.
2. On the Data Analytics Studio [download page](#), click **DOWNLOAD NOW!**.



A table listing the release packages with the associated release numbers, supported operating system, and downloadable link is displayed.

DAS			
1.0.0.0-89	centos7	https://archive.cloudera.com/web/DAS/1.0.0.0-89/centos7/DAS-1.0.0.1.0.0.0-89-centos7-rpm.tar.gz	rpm
1.0.0.0-89	centos7	https://archive.cloudera.com/web/DAS/1.0.0.0-89/centos7/DAS-1.0.0.1.0.0.0-89-centos7-tars-tarball.tar.gz	tarball
1.0.0.0-89	centos7	https://archive.cloudera.com/web/DAS/1.0.0.0-89/centos7/data-analytics-studio-mpack-1.0.0.1.0.0.0-89.tar.gz	mpack
1.0.0.0-89	centos7	https://archive.cloudera.com/web/DAS/1.0.0.0-89/centos7/data_analytics_studio-1.0.0.1.0.0.0-89-release_source.tar.gz	tarball
1.0.2.0-6	centos7	https://archive.cloudera.com/web/DAS/1.0.2.0-6/oE7aK/centos7/DAS-1.0.2.0-6-centos7-rpm.tar.gz	rpm
1.0.2.0-6	centos7	https://archive.cloudera.com/web/DAS/1.0.2.0-6/oE7aK/centos7/data-analytics-studio-mpack-1.0.2.1.0.2.0-6.tar.gz	mpack
1.0.2.0-6	centos7	https://archive.cloudera.com/web/DAS/1.0.2.0-6/oE7aK/centos7/data_analytics_studio-1.0.2.1.0.2.0-6-source.tar.gz	tarball
1.0.2.0-6	debian9	https://archive.cloudera.com/web/DAS/1.0.2.0-6/oE7aK/debian9/DAS-1.0.2.0-6-debian9-deb.tar.gz	deb
1.0.2.0-6	ubuntu16	https://archive.cloudera.com/web/DAS/1.0.2.0-6/oE7aK/ubuntu16/DAS-1.0.2.0-6-ubuntu16-deb.tar.gz	deb

3. Download the DAS MPack and the DAS RPM corresponding to your operating system and the DAS version that you want to install on your computer by clicking the downloadable link.

Table 2: Supported platforms - DAS cluster components

DAS/DAS-Lite	CentOS 7	Debian 9	Ubuntu 16
DAS Cluster Components (DAS Event Processor and DAS Webapp)	Yes	Yes (HDP 3.1.5)	Yes

Installation roadmap

The DAS cluster components comprise of the DAS Webapp and a DAS Event Processor. These components are installed on the Ambari cluster to use with DAS.

To install DAS/DAS-Lite, review the installation roadmap and follow the steps. We strongly encourage you to read the support requirements and the pre-installation tasks before you start the installation.

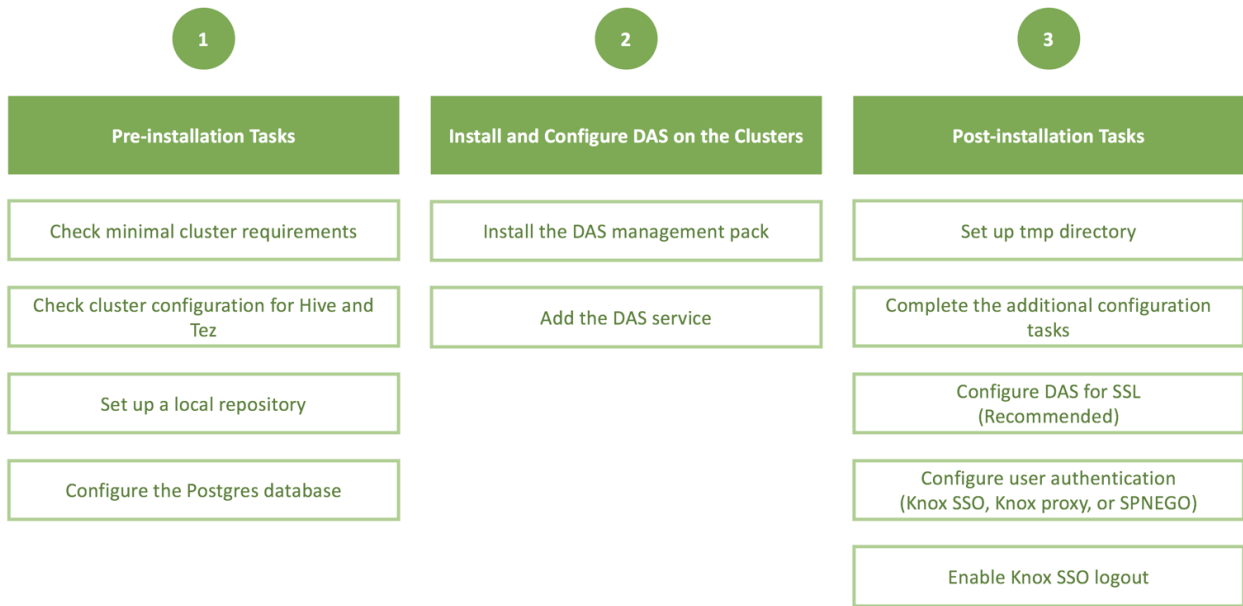


Table 3: DAS/DAS-Lite installation roadmap

No.	Task	Description
Pre-installation tasks		
1	Check the minimal cluster requirements.	To install and set up the necessary components and services, see Install the prerequisite components and services .
2	Check the cluster configuration for Hive and Tez.	Check the configuration settings for Hive and Tez in Ambari. See Check cluster configuration for Hive and Tez .
3	Create and set up a local repository.	Hortonworks does not host any public repository for DAS. Therefore, you need to setup a local repository to install the binaries. See Create and set up a local repository .
4	Configure the PostgreSQL database.	DAS requires a PostgreSQL database for storing query event information. See Configure PostgreSQL database .
Installing and configuring DAS on the HDP clusters		
5	Install the DAS management pack.	Install DAS cluster components (the DAS Webapp and the DAS Event Processor) on the same machine on which you have installed the Ambari server, using an Ambari management pack (MPack). See Installing the DAS cluster components .
6	Add the DAS service.	After installing the DAS MPack through the Ambari console, add the DAS service through the Ambari UI. See Adding the DAS Service through the Ambari UI .
Post-installation tasks		
7	Set up tmp directory.	You need write permission on the /tmp directory to write logs for secure clusters. See Setting up the tmp directory .
8	Complete the additional configuration tasks.	You can increase or decrease the time for which the audit and event logs are retained in the system, after which they are auto-purged. You can also make DAS work with HiveServer2 in case your Hive deployment is not LLAP-enabled and does not have Hive Server Interactive. See Additional configuration tasks .
9	Configure DAS for SSL.	It is recommended that you configure DAS for SSL. You can configure SSL if your HDP cluster is SSL-enabled. See (Recommended) Configuring DAS for SSL/TLS .
10	Configure user authentication.	To authenticate users using either Knox SSO, Knox proxy, or SPNEGO, configuring Knox SSO for HA clusters, and to enable Knox SSO logout option, see Configuring user authentication .