Cloudera Streaming Analytics 1.4.1

Deployment

Date published: 2019-12-17
Date modified: 2021-07-20

https://docs.cloudera.com/

Legal Notice

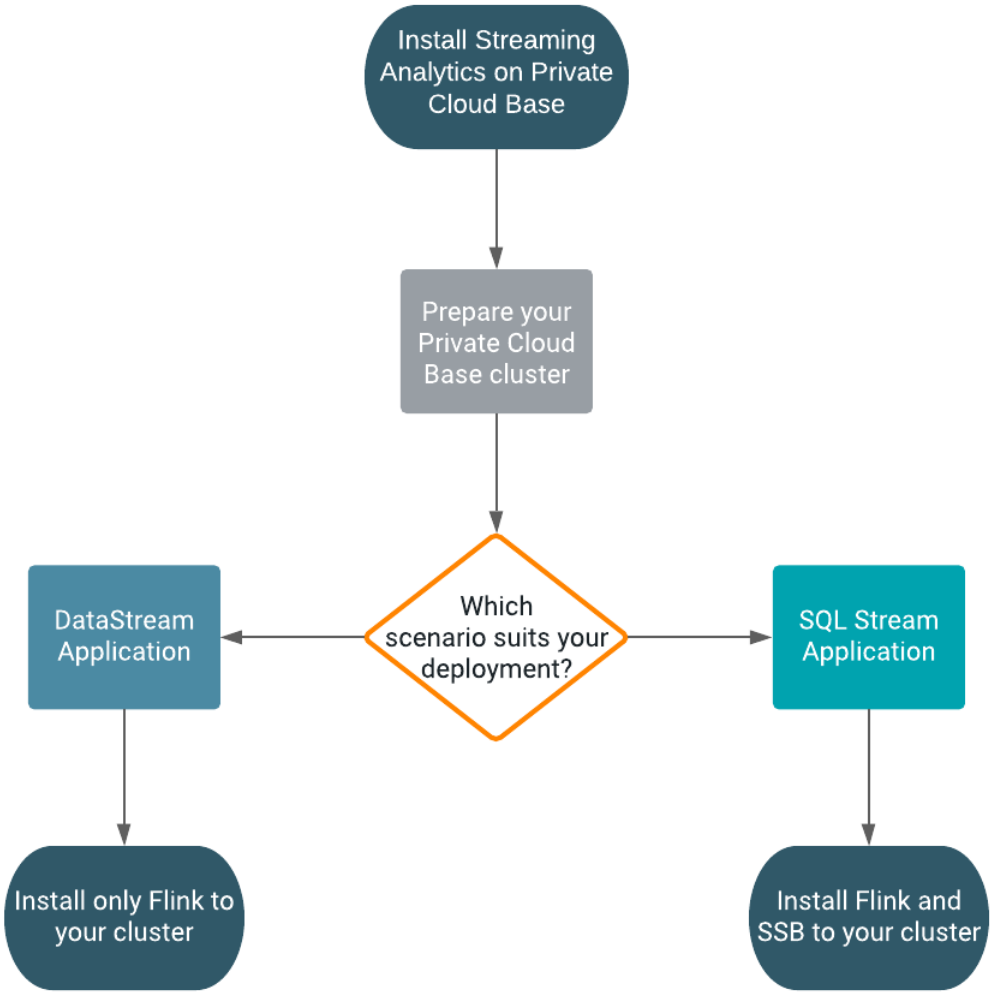
Contents

Deployment scenarios

You can deploy Streaming Analytics on Private Cloud Base depending on the application you want to build.

- DataStream application using only Flink. In this case, you need to create a Flink application cluster.
- SQL Streaming application using Flink with SQL Stream Builder. In this case, you need to create a Streaming SQL cluster.

You can use the following workflow to have an understanding of the deployment process:



Cluster service layout with Flink

In Cloudera Streaming Analytics (CSA), Flink has mandatory dependencies with HDFS, YARN, and Zookeeper. You need to assign the Flink Gateway and HistoryServer roles to the host, based on the mandatory dependencies.

Flink jobs are executed as YARN applications. HDFS is used to store recovery and log data, while ZooKeeper is used for high availability coordination for jobs. In a standard layout, an Apache Kafka cluster is often located close to the YARN cluster executing the Flink cluster.

The Flink Gateway is collocated with YARN and HDFS Gateways. The Flink HistoryServer is collocated with an HDFS role, which can be either an active role or a Gateway. Use the following general service layout when collocating Flink roles and dependencies.

Cluster service layout with SSB

In Cloudera Streaming Analytics (CSA), SQL Stream Builder (SSB) has mandatory dependencies with Flink and Kafka. But due to its dependency with Flink, you also need to add YARN, HDFS and Zookeeper as a mandatory service on your cluster. You need to assign the SSB roles in the same way as you assign the Flink roles.


Cluster: CDP Streaming
cdp1.acme.com
KB
SR
ZK
SMM
RNGR
SRM
CM AGNT
cdp2.acme.com
KB
ZK
RNGR
CM AGNT
cdp3.acme.com
KB
ZK
CM AGNT
cdp4.acme.com
KB
CM AGNT
Cluster: CDP
edge1.acme.com
Flink GW
YARN GW
HDFS GW
CM AGNT
utility2.acme.com
CM
CMS
NAV
CM AGNT
master1.acme.com
HDFS NN
HDFS JN
HDFS FC
YARN RM
YARN HS
Flink HS
ZK
CM AGNT
SSB CNS
master2.acme.com
HDFS NN
HDFS JN
HDFS FC
YARN RM
ZK
CM AGNT
SSB ENG
MV ENG
master3.acme.com
HDFS JN
ZK
CM AGNT
worker1.acme.com
HDFS DN
YARN NM
CM AGNT
worker2.acme.com
HDFS DN
YARN NM
CM AGNT
worker3.acme.com
HDFS DN
YARN NM
CM AGNT
SSB CNS SQL Stream Builder Console
SSB ENV SQL Stream Builder Engine
MV ENG Materialized View Engine
Flink GW Flink Gateway
Flink HS Flink History Server
KB Kafka Broker
SR Schema Registry
SMM Streams Messaging Manager
SRM Streams Replication Manager
CM Cloudera Manager Server
CM AGNT Cloudera Manager Agent
CMS Cloudera Management Service
NAV Cloudera Navigator
YARN GW YARN Gateway
HDFS GW HDFS Gateway
HDFS NN HDFS NameNode
HDFS JN HDFS JournalNode
HDFS FC HDFS Failover Controller
HDFS DN HDFS DataNode
YARN RM YARN ResourceManager
YARN NM YARN NodeManager
YARN HS YARN History Server
RNGR Ranger Server
ZK ZooKeeper Server