Fundamentals of Hadoop.

Assignment –2.

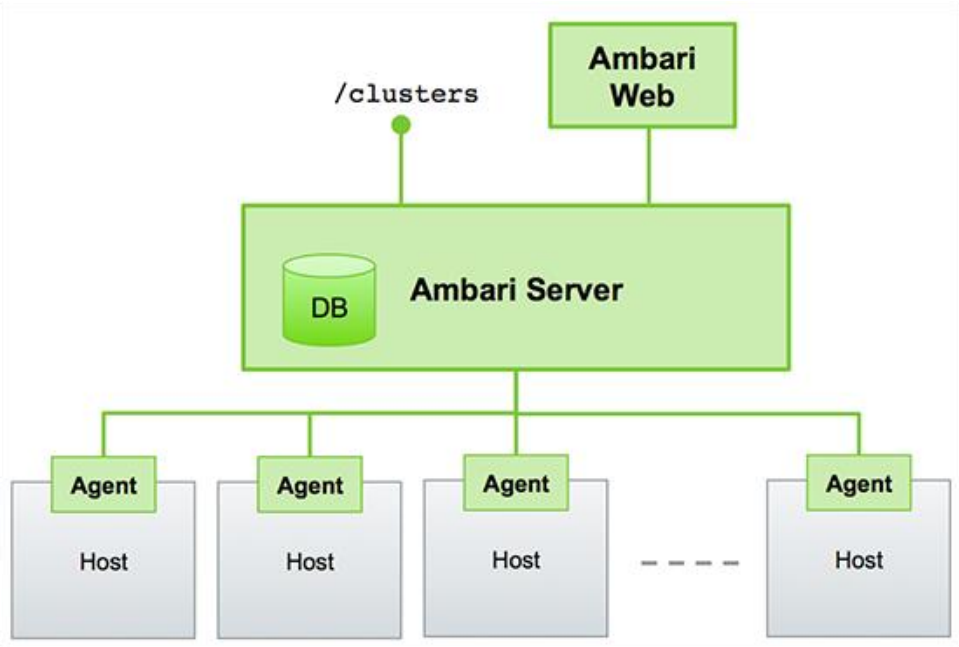
Hasnain Ullah Khan

2000102643

DSAI

1. Overall Architecture of Ambari.

The Ambari Server collects data from across your cluster. Each host has a copy of the Ambari Agent, which allows the Ambari Server to control each host.



Ambari Web is a client-side, JavaScript application that calls the Ambari REST API (accessible from the Ambari Server) to access cluster information and perform cluster operations. After authenticating to Ambari Web, the application authenticates to the Ambari Server. Communication between the browser and server occurs asynchronously using the REST API.

Further, by using the REST API, communication between the browser and server occurs asynchronously.

In addition, there is a REST API in Ambari which is accessed by Web UI, that resets the session timeout. Hence, we can say Ambari Web sessions do not timeout automatically. And, after a period of inactivity, we can configure Ambari to timeout.

2. Mention main components of Ambari.

1. Ambari Server

The entry point for all administrative activities on the master server is known as Ambari Server. It is a shell script. Internally, this script uses Python code, **ambari-server.py**, and routes all requests to it.

Ambari Server consists of several entry points that are available when passed different parameters to the Ambari Server program. They are:

- Daemon management
- Software upgrade
- Software setup
- LDAP (Lightweight Direct Access Protocol)/PAM (Pluggable Authentication Module) /Kerberos management
- Ambari backup and restore
- Miscellaneous options

2. Ambari Agent

Ambari Agent runs on all the nodes that you want to manage with Ambari. This program periodically sends heartbeats to the master node. By using Ambari Agent, Ambari Server executes many tasks on the servers.

3. Ambari Web User Interface

Ambari Web UI is one of the powerful features of Apache Ambari. The web application is deployed through the server of Ambari program which is running on the master host exposed on port 8080. This application is protected by authentication. You can access and then control and view all aspects of your Hadoop cluster, once you log in to the web portal.

4. Database

Ambari supports multiple RDBMS (Relational Database Management Systems) to keep track of the state of the entire Hadoop infrastructure. You can choose the database you want to use during the setup of Ambari. Ambari supports these following databases at the time of writing:

- PostgreSQL
- Oracle
- MySQL or MariaDB
- Embedded PostgreSQL
- Microsoft SQL Server
- SQL Anywhere
- Berkeley DB

3. Initiating start services and stop services from Ambari Web Console.

Starting the services

- Click **Actions** > **Start All** from the Ambari web interface.
- Optional: Follow this sequence to start the services on Ambari web interface:
- The order in which to start the services:

    a. ZooKeeper
    b. Ambari Metrics
    c. KAFKA
    d. HDFS
    e. YARN
    f. MapReduce2
    g. Network Performance Insight

Stopping the services

- Click **Actions** > **Stop All** from the Ambari web interface.
- Then, wait for all of the services to stop.
- Optional: Follow this sequence to stop the services on Ambari web interface:
- The order in which to stop the services:

    h. Network Performance Insight
    i. MapReduce2
    j. YARN
    k. HDFS
    l. KAFKA
    m. Ambari Metrics
    n. ZooKeeper