IBM Training

Exercise 1

Managing Hadoop clusters with Apache Ambari

Exercise 1: Managing Hadoop clusters with Apache Ambari

Exercise 1: Managing Hadoop clusters with Apache Ambari

Purpose:
You will explore the Apache Ambari web console and perform basic starting and stopping of services, giving you experience in using Apache Ambari to manage your Hadoop cluster.

VM Hostname: **http://ibmclass.localdomain**
User/Password: **biadmin / biadmin**
root/dalvm3

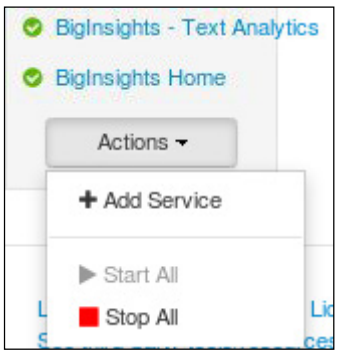
Task 1. Start the Apache Ambari web console and perform basic start/stop of services.

The major reference for Apache Ambari can be found at http://ambari.apache.org/.

There are various shells available for Ambari, but not all are fully developed or available at time of developing this course; those shells are not covered in this exercise. These shells include: Ambari shell and Python Shell & Client.

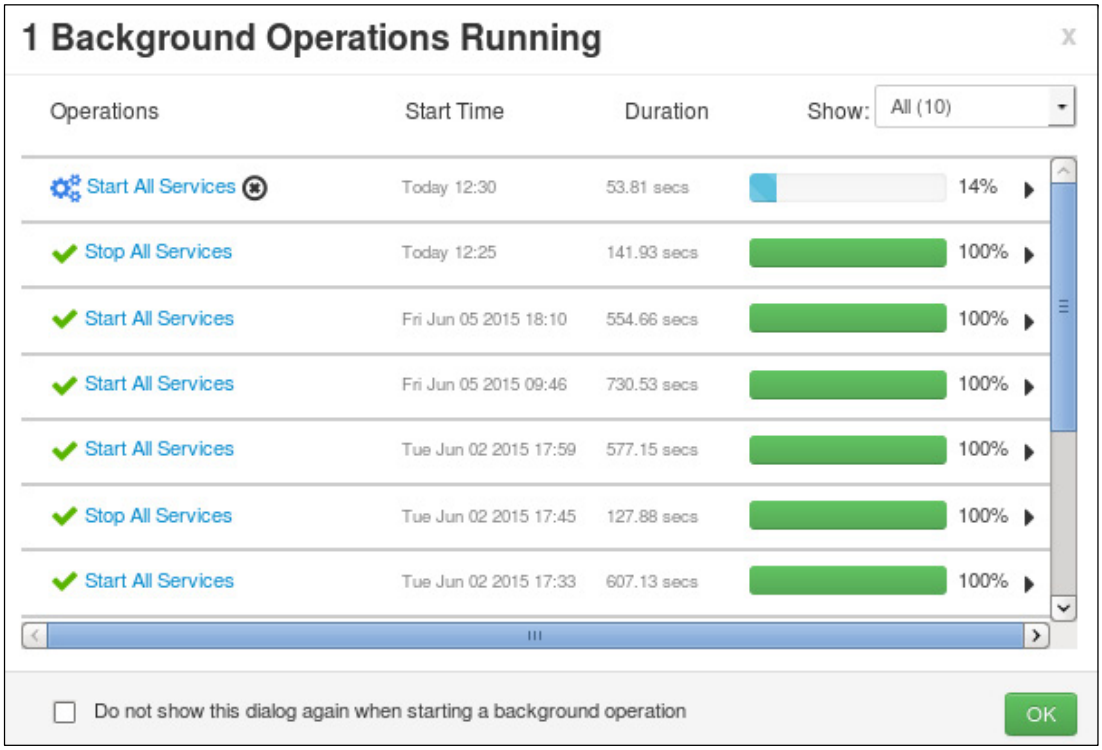
Note: Cluster administration is not normally done by end users, but knowledge of Apache Ambari is extremely useful for all users as it provides concepts needed for both administration and use of the cluster.

1. Connect to and login to your lab environment with the user **biadmin** and the password **biadmin**.
2. Launch **Firefox**, and navigate to the **Ambari** login page, **http://localhost:8080**, logging in as **admin/admin**.
   Note which services are currently running from the left pane.
3. At the bottom of the **Services** pane, click **Actions**, and then:
   - if the services are started, click **Stop All**.
   - If the services are stopped, click **Start All**.



4. When asked for confirmation, click **OK**.

Starting or stopping the services can sometimes take several minutes.



Operations which are in progress are shown in blue, complete and successful in green, and failed in red.

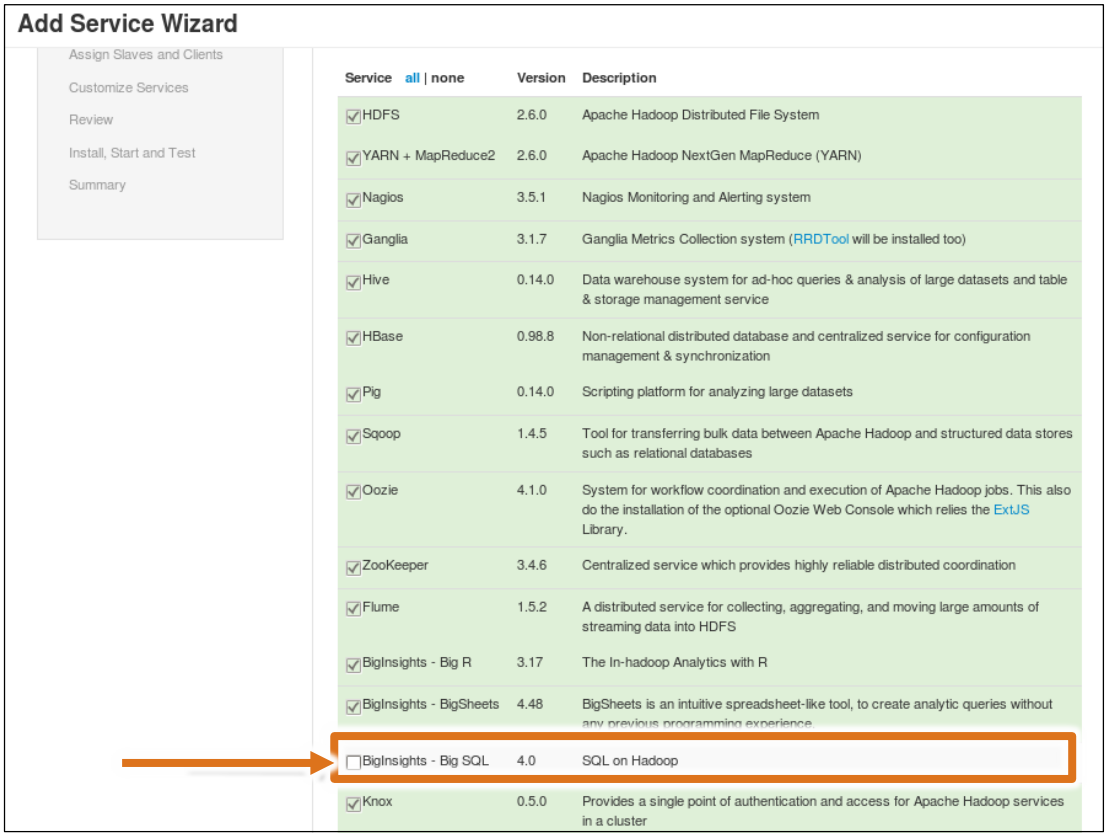
If you stopped the services in step 3, from the **Actions** menu, click **Start All**, and then click **OK** when prompted for confirmation.

5. To see what services can be added, from the **Actions** menu, click **Add Service**.

   The services shown will include some or all of the following; your listing will probably differ if you have a later version of some or all of these components.

Service	Version	Description
HDFS	2.6.0	Apache Hadoop Distributed File System
YARN + MapReduce2	2.6.0	Apache Hadoop NextGen MapReduce (YARN)
Nagios	3.5.1	Nagios Monitoring and Alerting system
Ganglia	3.1.7	Ganglia Metrics Collection system (RRDTool will be installed too)
Hive	0.14.0	Data warehouse system for ad-hoc queries & analysis of large datasets and table & storage management service
HBase	0.98.8	Non-relational distributed database and centralized service for configuration management & synchronization
Pig	0.14.0	Scripting platform for analyzing large datasets
Sqoop	1.4.5	Tool for transferring bulk data between Apache Hadoop and structured data stores such as relational databases
Oozie	4.1.0	System for workflow coordination and execution of Apache Hadoop jobs. This also do the installation of the optional Oozie Web Console which relies the ExtJS Library
ZooKeeper	3.4.6	Centralized service which provides highly reliable distributed coordination
Knox	0.5.0	Provides a single point of authentication and access for Apache Hadoop services in a cluster
Slider	0.60.0	A framework for deploying, managing and monitoring existing distributed applications on YARN
Solr	4.10.3	Solr is the popular, blazing fast open source enterprise search platform from the Apache Lucene project
Spark	1.2.1	Apache Spark is a fast and general engine for large-scale data processing
Flume	1.5.2	A distributed service for collecting, aggregating, and moving large amounts of streaming data into HDFS

Services that are installed and available are shown with check-marks. Services without a check-mark (such as Big SQL in this instance) are not yet installed or available:

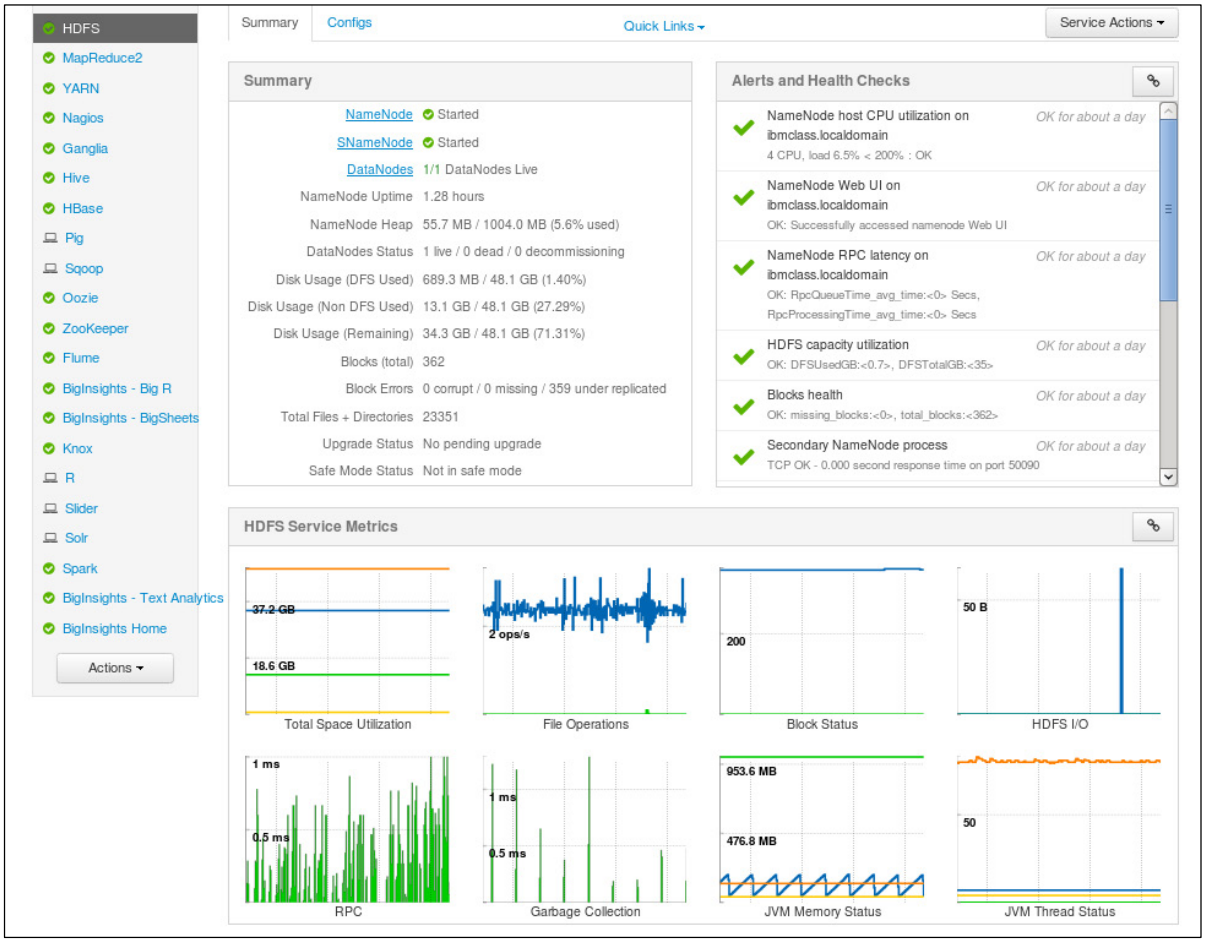


Important: this listing shows the version of each component, whether Open Data Platform or the IBM BigInsights Add-in that you have and whether it is an installed service. You may need this information later for IBM service support.

6. Close the **Add Services** window.
7. When all of the services have been started, look at center of the page to see the high level information available there.

   You will explore the various individual services that are listed, to better understand what is available.

8. Click the **HDFS** service to expose the relevant information and metrics:



9. Explore other services on the left panel of the main web page.

Task 2. Explore other aspects of the Ambari web server.

1. Navigate and explore other aspects of the Ambari web server interface using the earlier content pages and slides in this unit to guide you.
2. Close all open windows.

Results:
You explored the Apache Ambari web console and performed basic starting and stopping of services, gaining experience in using Apache Ambari to manage your Hadoop cluster.