

Welcome to the unit on Hadoop Administration.

The agenda covers adding nodes to a cluster, verifying the health of a cluster, and stopping / starting components. Then the unit covers configuring Hadoop and setting up rack topology.

Let's begin with adding and removing nodes from a cluster.

Adding nodes can be performed from the BigInsights Console or the command line. To do so requires either the ip address or hostname of the node to be added. The node to be added must also be reachable. And as a matter of fact, it works both ways. The master and child nodes must all be able to communicate with each other. In this case, a child node refers to the node that is being added. It may not have BigInsights already installed on it. When the node is added to a cluster, the BigInsights code is transferred to the new node and installed.

From the BigInsights Console you use the Cluster Status tab and, on the left side, select Nodes. You then click the Add nodes pushbutton.

You are then presented with a dialog that allows you to specify one or more nodes to be added. You may type in ip address, hostnames or any combination thereof. You can even specify an ipaddress range or a regular expression with your hostname. In the example, child5.ibm.com, child6.ibm.com, etc would be added.

Once the nodes have been added, you define which services are to be hosted on those nodes. You can select multiple services for one or more nodes.

Services can also be removed from a node. As a matter of fact, if you are using the BigInsights Console to remove a node, you must first remove all services from that node. Depending on which services are running on a node, you select which are to be removed.

When there are no services running on a node, it can be removed using the BigInsights Console.

All that we just said about adding and removing nodes can also be done from the command line. You use the addnode.sh script to add one or more nodes. Once again, either ipaddresses or hostname can be specified. You also specify which services are to run on the new nodes. In the example, the hadoop service was specified. This means that the node will be used for a DataNode and TaskTracker.

Listnode.sh displays all of the nodes in the cluster and which services are running on those nodes.

You can remove a service from a node using the removenode.sh script and you can remove a node using that same script. In this case, you can remove a node even though it has services running. Maybe not the best idea, but it is doable. At least you are prompted whether to continue.

Next let us discuss verifying the health of your cluster. You are able to view all of the nodes in the cluster, see the status of each node and which services are running on each node.

From the command line you can run the DFS Disk Check report . This lets you see how much space is still available on each DataNode.

Next let us look at start and stopping components and services.

You may in order to save some resources, you only want to start a subset of the Hadoop components . From the command line you can run the start.sh script and specify one or more components that are to be started. If you want to start all components, then you can execute start-all.sh. Likewise, you can stop a subset of running components and if you want stop all components, execute stop-all.sh.

You also have the option of starting and stopping components using the BigInsights Console.

Please continue to the next part of this presentation.