

- Getting Access to the **bigdata** Cluster
 1. Interacting with the Cluster
 2. The Hardware
 3. The Software
 1. The Software

1 Getting Access to the **bigdata** Cluster

The **bigdata** cluster is available for use by the Vanderbilt community. Users should contact ACCRE to get access to the cluster.

2 Interacting with the Cluster

Approved Vanderbilt users will be able to connect to the hostname **bigdata.accre.vanderbilt.edu** with **ssh**. Users can submit jobs from the command line using the tools maintained by Cloudera (e.g. MapReduce, Spark, HBase, etc.).

Cloudera provides the Hadoop User Experience (Hue) tool to coordinate constructing and executing jobs (not to mention browsing and managing HDFS data) through the interface **bigdata.accre.vanderbilt.edu:8888**.

3 The Hardware

The current **bigdata** cluster is a test cluster comprised of commercial hardware. To conform to Cloudera recommended configuration, three nodes are dedicated to the management of the cluster:

- abd740
- abd741
- abd742

The computational *guts* of **bigdata** are the datanodes:

- abd743
- abd744
- abd745
- abd746
- abd747
- abd748

where MapReduce jobs actually take place. Each of these nodes have one 250 GB drive reserved mainly for the OS and software, but for mounting the Hadoop Distributed File System (HDFS), each node has 2 x 4 TB hard disk drives.

4 The Software

Cloudera Manager is deployed on the cluster to coordinate job submission, with the following services running on each node:

- abd740
 - Cloudera Manager
 - Alert Publisher
 - Event Server
 - Host Monitor
 - Service Monitor
 - HBase Master
 - Hive
 - Hiveserver2
 - Hive Gateway
 - Hue Server
 - Oozie Server
 - Spark Gateway
 - YARN (MR2 Included)
 - Resource Manager
 - Job History Server
 - Zookeeper Server
-

4.1 The Software

- abd741
- HDFS NameNode
- Hive
 - Hive Metastore Server
 - Hive Gateway
- Spark Gateway
- YARN (MR2 Included) Resource Manager
- ZooKeeper Server
- abd742
- HDFS SecondaryNameNode
- Hive Gateway
- Impala
 - Impala StateStore
 - Impala Catalog Server

- Spark Gateway
- ZooKeeper Server