Big Data Administration

We began by choosing a theme for our cluster. Each group of virtual machines was named after either cats, birds, dogs or reindeer and assigned unique IP addresses. Once I had configured my vagrant files I installed vagrant on my computers to create four virtual machines.

To begin with, we installed Hadoop in Pseudo-Distributed Mode, setting up the required six core Hadoop daemons (listed below) on a single virtual machine before expanding out to others.

* NameNode
* SecondaryNameNode
* DataNode
* ResourceManager
* NodeManager
* JobHistoryServer

As a simple test, we stored a text file in the HDFS with specific file permissions.

Quorum of machines for majority vote. (zookeeper)