1. Explain what is a cluster and what is a Hadoop cluster

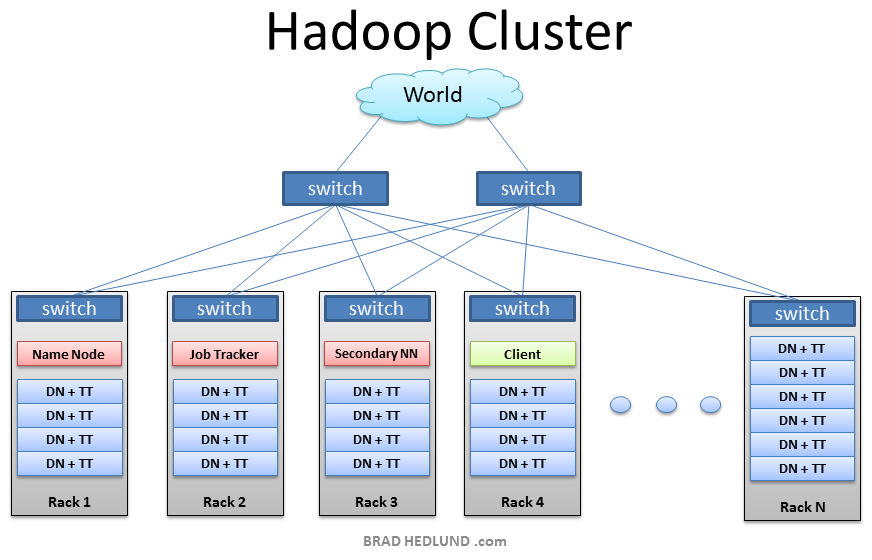
**Cluster:** A cluster consists of a set of loosely or tightly connected [computers](https://en.wikipedia.org/wiki/Computers) that work together so that, in many respects, they can be viewed as a single system. Unlike [grid computers](https://en.wikipedia.org/wiki/Grid_computing), computer clusters have each node set to perform the same task, controlled and scheduled by software.

**Hadoop Cluster:** In simple words, A Cluster used for Hadoop is called a Hadoop cluster. **Hadoop cluster is a special type of computational cluster designed for storing and analyzing vast amount of unstructured data in a distributed computing environment. These clusters run on low cost commodity computers.**

Hadoop clusters are often referred to as "shared nothing" systems because the only thing that is shared between nodes is the network that connects them.   
Large Hadoop Clusters are arranged in several racks. Network traffic between different nodes in the same rack is much more desirable than network traffic across the racks.

1. What is meant by a Rack and explain the rack arrangement in a Hadoop cluster.

**Rack:** A rack is a collection of 30 or 40 nodes that are physically stored close together and are all connected to the same network switch. Network bandwidth between any two nodes in rack is greater than bandwidth between two nodes on different racks. A Hadoop Cluster is a collection of racks.



This is the typical architecture of a Hadoop cluster. You will have rack servers (not blades) populated in racks connected to a top of rack switch usually with 1 or 2 GE boned links. 10GE nodes are uncommon but gaining interest as machines continue to get more dense with CPU cores and disk drives. The rack switch has uplinks connected to another tier of switches connecting all the other racks with uniform bandwidth, forming the cluster. The majority of the servers will be Slave nodes with lots of local disk storage and moderate amounts of CPU and DRAM. Some of the machines will be Master nodes that might have a slightly different configuration favoring more DRAM and CPU, less local storage.