Things to Review for this Past QC (Week 5)

1. Components of HDFS (Hadoop Distributed File System)
   1. NameNode
      1. A master server that manages the file system namespace and regulates access to files by clients
      2. Manages all of the DataNodes on a HDFS cluster
      3. Only one per cluster, but there can be any number of backup NameNodes
   2. DataNodes
      1. Manages storage attached to the node it is running on.
      2. Data is stored in blocks (128 MB by default) which are replicated across other DataNodes in the cluster (3 replications by default
2. Components of YARN (Yet Another Resource Negotiator)
   1. Resource Manager
      1. Manage resources across all the nodes working in Hadoop
      2. Manage the Node Managers
      3. SCHEDULER
         1. Fair Scheduler
            1. Schedule tasks based on a FIFO system
         2. Capacity Scheduler
            1. Schedule tasks based on size of resources
      4. APPLICATIONS MANAGER
         1. Manages running of Application Master (inside of Node Manager) and helps to restart Application Master when it fails
   2. Node Manager
      1. - Responsible for the execution of the task in each DataNode.
      2. - Manages workflow and user jobs on a specific node
      3. - Send a heartbeat to the Resource Manager with information of the resources in each container.
      4. Container
         1. Set of resources (RAM, CPU, Storage) on a single node.
         2. Resources are allocated by the Resource Manager and monitored by the Node Manager
      5. Application Master
         1. Monitor the execution of tasks running on each node in the cluster.
         2. Main responsibility is to negotiate resources from the Resource Manager
3. Description of Unix
   1. Unix is an operating system which allows us to write commands in a terminal that are interpreted into readable instructions for our processor. Linux is a family of open-source Unix-like OSs. On Windows, we run a distribution of Linux (Ubuntu) on a Virtual Machine (VM) (WSL). We write instructions using the shell (terminal), which communicates with the kernel to execute those instructions.