1. Explain the importance of below 4 demons in job execution with minimum of 5 points

** Name node**

* + - Name node is the Master daemon process which runs on a single machine. Usually it’s one name node per cluster
    - Major functionality of namenode is maintenance of namespace. Namenode has details about all the data and where its stored and also information about backup copies. Data node keep the name node updated about its status by pinging(heartbeats). If name node doesn’t get any heartbeat from data node for 10 mins then it assumes this data node is down and looks for alternatives to store the data.
    - Namenode maintains information about the namespace using FSImage and its latest changes (i.e insert/delete/rename files or directories) are tracked in edit\_log file. These two files are merged frequently using check pointing technique. The frequency of this checkpoint process can be controlled through settings.

 Data node

* + - Data node is the Slave Daemon process which runs on multiple machines. Data node is where the actual data is stored and tracked by namenode. Data node keeps pinging the name node (heartbeats). Through these heartbeat DN communicates to NN about the available space in the node on which its running.
    - By default, heartbeat is sent every 3 sec and if no heartbeats is sent for 10 mins then NN assumes DN is down and arranges for alternatives.
    - DN sends block report every 6 hrs to NN. This checks for data corruption
    - Read/Write data always happen in DN

 Resource Manager

* + Resource Manager(RM) is a concept introduced in Hadoop 2.x. RM is a Master daemon and its kind of replacement of Job Tracker
  + There is 1RM/Cluster. You can have standby RM
* Client submits jar+input file + outputfile +config to RM. RM checks these files, on successful check returns app\_id to client. On receiving the app\_id client copies all these files to distributed cache.
  + RM looks for machine, node, where more ram is available. Then asks that respective NM to create container and within which Application Master is created
  + If for some reason job fails, then RM creates new Application Master.

 Node manager

* + NM is a slave daemon process and its introduced in Hadoop 2.x replacing Task Tracker. Usually we have 1NN per node and therefore there would be multiple NNs
  + The primary job of NN is Task execution and can take any kind of tasks
  + NM sends heartbeat to RM communicating the resources available in the node
  + When the job gets started on RM instruction NM create container and within which Application Master (AM) is created. There is only 1AM/Application and it closes when jobs are done
  + AM instructs the NM to start the container and execute task. NM copies files from DC to the node jvm