**Hadoop terms and notes**

MapReduce, Yarn, HDFS are part of Hadoop

HDFS: Hadoop distributed file system, allows distribute storage of big data, maintains redundant copies of the data, can recover backup data automatically

HDFS is responsible for storing large data sets of structured or unstructured data across various nodes and thereby maintain the metadata in the form of log files.

Hadoop uses HDFS (Hadoop distributed file system). HDFS is the component of Hadoop that lets access to data across computer clusters or group of computers that work together. HDFS does things like copy data multiple times and distributes copies to individual nodes.

YARN: yet another resource negotiator, data processing comes into play, manages resources on computing cluster, decides what gets to run tasks when,

MapReduce: programming model, uses mappers and reducers(scripts or functions), mappers transform data in parallel, reducers aggregate(combine) that data together, both mapper and reducer help solve more complicated problems

Pig: high level programming API that allows simple scripts to be written without writing java, python ,etc., sits on top of MapReduce

Hive: