Please be ready with the answers for these questions. While we will be meeting in the next class.

* What are the differences between regular FileSystem and HDFS?

### Why is HDFS fault-tolerant?

### Explain the architecture of HDFS.

### What is the difference between a federation and high availability?

### If you have an input file of 350 MB, how many input splits would HDFS create and what would be the size of each input split?

### How does rack awareness work in HDFS?

### What would happen if you store too many small files in a cluster on HDFS?

### How do you copy data from the local system onto HDFS?

### What is the distributed cache in MapReduce?

### What role do RecordReader, Combiner, and Partitioner play in a MapReduce operation?

### Why is MapReduce slower in processing data in comparison to other processing frameworks?

### Is it possible to change the number of mappers to be created in a MapReduce job?

### What happens when a node running a map task fails before sending the output to the reducer?

### Yarn main components

### What are the different schedulers available in YARN?