**Hadoop MapReduce Intro**

**Question 1: Select the type of failure when a server in a cluster gets out of service because of the power supply has burnt out**

Fail-Stop

Correct: Yes, because the server requires somebody to fix the problem

**Question 2: Select the type of failure when a server in a cluster reboots because of a momentary power failure**

Fail-Recovery

Correct: That's right, the server recovers after a failure by itself

**Question 3: Select the type of failure when a server in a cluster outputs unexpected results of the calculations**

Byzantine failure

Correct: Yes, it is a Byzantine failure, because it looks like the server doesn't behave according to the protocol

**Question 4: Select the facts about Fair-Loss Link**

Some packets are lost regardless the contents of the message

Correct: Yes, when packets are lost regardless their contents, this is the definition of a Fair-Loss Link

**Question 5: Select the facts about Byzantine Link:**

Some packets are created out of nowhere

Correct: Yes, in case of Byzantine Link packets can be created

Some packets are modified

Correct: Yes, in case of Byzantine Link packets can be modified

**Question 6: Select a mechanism for capturing events in chronological order in the distributed system**

Both ways are possible, logical clocks hide inaccuracies of clocks synchronization

Correct: That's right, use a logical clocks, but don’t forget about clocks synchronization

**Question 7: Select the failure types specific for distributed computing, for example for Hadoop stack products**

Fail-Recovery, Fair-Loss Link and Asynchronous model

Correct: That's right, these types of failures are inherent in Hadoop

**Question 8: What phase in MapReduce paradigm is better to filtering input records without any additional processing?**

Map

Correct: Yes, mappers process records independently and in parallel, so they are suitable for filtering

**Question 9: Select a MapReduce phase which input records are sorted by key:**

Reduce

Correct: That’s right, reducers process records are sorted by key

**Question 10: What map and reduce functions should be used (in terms of Unix utilities) to select the only unique input records?**

map=’cat’, reduce=’uniq’

Correct: Yes, because 'uniq' on the sorted input records gives the required result

map=’uniq’, reduce=’uniq’

Correct: Yes, 'uniq' on the Map phase in some cases reduces the amount of records, 'uniq' on the Reduce phase gets the sorted records and solves the task

**Question 11: What map and reduce functions should be used (in terms of Unix utilities) to select only the repeated input records?**

map=’cat’, reduce=’uniq -d’

Correct: Yes, mappers pass all the records to reducers and then 'uniq -d' on the sorted records solves the task

**Question 12: What service in Hadoop MapReduce v1 is responsible for running map and reduce tasks:**

TaskTracker

Correct: That’s right, TaskTracker is responsible for running tasks

**Question 13: In YARN Resource Manager is...**

A service which processes requests for cluster resources

Correct: Yes, Resource Manager manages cluster resources and processes requests from Node Managers