HDFS performance, tuning, and robustness

1.

Name the configuration file which holds HDFS tuning parameters

- mapred-site.xml
- core-site.xml
- hdfs-site.xml

2.

Name the parameter that controls the replication factor in HDFS

dfs.block.replication

dfs.replication.count

dfs.replication

replication.xml

3.

Check answers that apply when replication is lowered

- [x] HDFS is less robust
- [x] Less likely that data will be local to more workers
- [ ] Aggregate I/O rate will be worse
- [x] HDFS will have more space available

4.

Check answers that apply when NameNode fails to receive heartbeat from a DataNode

- [x] DataNode is marked dead
- [ ] NameNode will attempt to restart DataNode
- [x] No new I/O is sent to particular DataNode that missed heartbeat check
- [x] Blocks below replication factor are re-replicated on other DataNodes

5.

How is data corruption mitigated in HDFS

Data from all replicas is compared for correctness

checksums are computed on file creation and stored on clients

checksums are computed on file creation and stored in HDFS namespace for verification when data is retrieved.