**Q1: MegaSoft corporation is planning to store its data in a data center. It has two design choices: (a) install a regular desktop file system (like a Windows file system or Linux file system) on each server, or (b) install a cloud file system (such as HDFS or GFS) to store the data.**

**What are the advantages of choosing option b over option a? Please justify your answer.**

**Answer:** If MegaSoft corporation has choice to choose between the regular desktop file system and cloud file system then the corporation would definitely go for the cloud file system. The main advantages for going with the cloud file system are the cost management, storage management, security and the backup of the data.

The difference between the cloud file system and traditional file system is just a cloud. In cloud there is unlimited scale, very strong privacy concerns, data protections, no backup problems. Whereas with the traditional file systems there all are opposite, especially the backup of bundles of data.

In cloud file system the data can be accessed from anywhere you want to, but in traditional file system you can access the data where it is stored, but one disadvantage of the cloud file system is that in the traditional file system we don’t need the internet connection for the access but in cloud file system we need the internet connection.

**Q2:** **Captain Kirk has installed an HDFS-based file system to store Starship Enterprise's data. After a battle with Klingons, he found that the file system has been impacted. Will Kirk be able to get the data back in each case? If yes, then explain how. If no, then explain why Kirk can't get the data back.**

(a) The Klingon attack caused a datanode server hard drive to fail.

(b) The Klingon attack corrupted a Namenode hard drive that caused the FsImage and EditLog to be unusable.

**Answer a:** If Klingon attack caused a datanode server hard drive to fail, Captain Kirk will be still able to retrieve the data which is compromised. The reason behind that is that HDFS has the ability to replicate the data blocks. HDFS is designed in such a way that it tolerates the failures.

The data is basically distributed in many different datanodes in the cluster, with each small percent of data, so whenever the any of the datanode get failed the other datanode will be available for the lost data. When the systems detected that there is lost of any datanode so the other datanodes start replicating the data which they have.

**Answer b:** No, the data will be lost if Klingon attack corrupted the namenode hard drive. Namenode is very important part of the HDFS, which is responsible for keeping the track of metadata about their location and data blocks. If the namenode is corrupted it means that the fsimage and editlog is unusable now. And without the fsimage and editlog the clusters would lose the track of location of data blocks. By all these information we can get to know that getting back the data might be very difficult and challenging and we might lose some of the data too. But there is one way out there too which is that if we can install an extra NameNode for the backup. Whenever there is any problem or corruption in the main one, we can just switch it with the backup one. This would ensure that the HDFS cluster can recover from namenode failure without losing any data.