1. [4 points] 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.

Ans. There are various benefits to using a cloud file system (option b) for data storage in a data centre as opposed to a standard desktop file system (option a). These arguments support this decision:-

- **Scalability: Because they are made to extend horizontally, cloud file systems can manage increased workloads and volumes of data by dividing the data across several servers. A cloud file system can readily expand to meet MegaSoft Corporation's growing data storage needs without requiring significant hardware upgrades or reconfigurations. A standard desktop file system, on the other hand, would find it difficult to manage the expanding data volume and might necessitate frequent hardware upgrades or server replacements.**

- **Fault Tolerance and Data Reliability: There are built-in mechanisms for failure tolerance in cloud file systems. In the event of hardware problems or server breakdowns, they ensure high availability and data durability by replicating data across numerous servers. Regular desktop file systems, in comparison, frequently lack reliable built-in fault tolerance capabilities, making them more prone to data loss or outages.**

- **Data Accessibility and Distribution: Data can be distributed and accessed easily across numerous servers and locations thanks to cloud file systems. MegaSoft's use of a cloud file system enables effective data replication, backup, and disaster recovery by distributing its data across many data centres or regions. Faster data access and better performance are also made possible by this distributed nature, especially when working with big datasets. The accessibility and spread of data may be constrained in a normal desktop file system, which could result in slower data access and less geographic redundancy.**

- **Elasticity and Resource Optimization: Due to the elasticity of cloud file systems, MegaSoft is able to dynamically modify its storage capacity in response to demand. Additional servers can be quickly added or withdrawn using a cloud file system as needed, optimising resource utilisation and lowering expenses. On the other hand, regular desktop file systems may need user intervention and hardware provisioning to account for changes in storage needs.**

- **Advanced Features and Integration: Additional tools and connectors that improve data management and analytical capabilities are frequently included with cloud file systems. For instance, they might have built-in support for analytics tools, distributed computing, and data processing frameworks. With the use of these capabilities, MegaSoft may be able to gain insightful information from its data and boost overall data processing effectiveness.**

2. [2x2=4 points] 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.

Ans. If a Klingon attack causes a failure in the hard drive of a datanode server, Captain Kirk will still be able to retrieve the data stored in the HDFS-based file system. The HDFS architecture replicates data across multiple datanodes, ensuring fault tolerance. Even with a failed datanode server, the data remains accessible on other datanodes that hold replicas of the same data block. The Namenode, which manages metadata and data block locations, redirects read requests to healthy datanodes. Hence, Captain Kirk can retrieve the data despite a datanode hard drive failure.

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

Ans. In the event of a Klingon attack corrupting the hard drive of the Namenode, rendering the FsImage and EditLog unusable, it becomes more challenging for Captain Kirk to recover the data. The Namenode is crucial for data accessibility as it maintains metadata. However, HDFS incorporates a secondary Namenode that periodically checkpoints and merges the FsImage and EditLog, enabling faster recovery from primary Namenode failures. To recover from a corrupted Namenode hard drive, Captain Kirk must restore the last reliable backups of FsImage and EditLog, replacing the corrupted drive and recovering the Namenode. This process allows access to the data stored in the HDFS-based file system.