1. **Can you think of a use case of Big Data?  Explain it briefly.**

* The internal Revenue Service (IRS), Where millions of Americans and Non-Americans file for taxes. The IRS uses Big Data to stop fraud, Identity theft and detecting who is not paying taxes.
* The Center for Disease Control and Prevention uses historical data to protect people’s health, control and prevent outbreaks.

1. **What are the advantages of using Hadoop and HDFS?**

Hadoop solved the problem of dealing with Big Data because Hadoop runs on a cluster of machines, clusters scale horizontally, handle unstructured data, provider storage & computing, provide storage at reasonable cost and captures new or new data.

HDFS is one of the three main components of Hadoop, it is the storage layer where data are stored across multiple nodes and divides the data into bocks and stores them on multiple machines. DataNode and NameNode are the master/slave architectures for managing and distribute the date. Where master stores meta data and slaves manage the data blocks.

1. **Explain the term block abstraction in Hadoop and state it's advantages.**

Where the files are divided block, and each block is 128MB. Blocks are stored on standard machine file system and HDFS lies on top of the OS stack, and we should not care about HOW Hadoop operate internally.

The advantages

1. File sizes can exceed the capacity of the HDD so splitting the files into equal sizes is good way to maintain the big files.
2. Easy to configure the Block Size and Replication factor
3. **What is the meaning of fault tolerance in HDFS and how is it achieved?**

How Hadoop will response in case or hardware failure. Hadoop handles this issue through replicating the blocks and by default the replication factor is 3 and can be configured so in case of failure to read specific block there are another 2 copies on another rack.

1. **Consider a 560 TB of text file which needs to be stored in HDFS. The block size has been set to be 128 MB with a replication factor of 3. The cluster has 100 DataNodes each with a capacity of 15 TB.  Will it be possible to store this text file in this HDFS cluster? Why or why not**? It’s not possible because we need at least 1680TB which is 560 \* 3 Replication factor. If we decreased the replication factor to 2 then we can store it.