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

**(Do not repeat the ones from the slides!)**

-> Analysis user behavior on Social Network(Facebook, Google, Amazon,...) to run suitable Ads.

**2/ What are the advantages of using Hadoop and HDFS?**

-> Hadoop is a highly scalable storage platform. It can store and distribute very large data sets across hundreds of inexpensive servers that operate in parallel. HDFS is open source software, so that if an organization chooses, it can be used with zero licensing and support costs.

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

-> File ine HDFS are broken into block-sized chunks, which are stored as independent units. A typical block size is 128MB. A file in HDFS is smaller than a single block does not occupy a full block’s worth of underlying storage. Therefore we can save the memory.

**4/ What is the meaning of fault tolerance in HDFS and how is it achieved?**

-> Each file is a sequence of blocks in HDFS. Each file has the same size expect the last one. All Block are replicated so if one block is corrupted, we will not lost the data.

**5/ 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?**

-> The HDFS Cluster capacity is 15 TB\*100 = 1500 TB

By default, replicator facotor is 3. To store the text file, We need 560 TB \* 3=1680 TB.

So It is not possible to store that file in this HDFS cluster.