1. [2] **Can you think of a use case of Big Data?  Explain it briefly.  
   (Do not repeat the ones from the slides!)**
   * **Stream analysis in revolution industry 4.0 to improve machines performance**
   * **Recommendation Engines**
   * **Social Media Analysis and Response**
   * **Internet of Things (IOT)**
2. [2] **What are the advantages of using Hadoop and HDFS?**

**Advantages of Hadoop and HDFS**

1. **Scalability**
2. **Cost effective**
3. **Flexible**
4. **Fast**
5. **Resilient to failure**
6. **Fault tolerance**
7. **Open source**
8. **write once and read many, very useful for streaming**
9. [2] **Explain the term block abstraction in Hadoop and state it's advantages.**

**block abstraction in Hadoop**

* **Like in a filesystem for a single disk, files in HDFS are broken into block-sized chunks, which are stored as independent units**
* **A typical block size is 128 MB (was 64 MB in Hadoop 1.x).**
* **Unlike a filesystem for a single disk, a file in HDFS that is smaller than a single block does not occupy a full block’s worth of underlying storage.**
* **Blocks are themselves stored on standard single-machine file systems, so HDFS lies on top of the standard OS stack**

1. [2] **What is the meaning of fault tolerance in HDFS and how is it achieved?**

* **In HDFS each file is a sequence of blocks.**
* **All blocks in the file except the last one is of the same size.**
* **Blocks are replicated for fault tolerance.**
* **Block size and number of replicas are configurable for files.**
* **Replication factor is 3 by default.**
* **NameNode tries to place replicas of blocks on multiple racks for improved fault tolerance.**

1. [2] **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**?

**Total DataNodes available = 100 \* 15 TB = 1500 TB, we can’t store text file in this cluster reason (cluster size less than text file).**