1. In security intelligence like FBI collects many datas for everywhere for security reason in a day. It collects data from many different sources like growing number of high- tech crimes based terrorism, major cyber fraud and other things. To meet the security challenge enhance cyber security and intelligence analysis platforms with big data technologies to process and analyze track and try to find the data exchange in the day.

2. The advantages of using Hadoop and HDFS are

a. Highly scalable storage platform

b. Highly reliable

c. Ability to store big date with no limit on storage

d. WORM( write one real more)

e. very powerful

f. it is open source software

g. it is faults tolerance

3.Block abstraction is that data node does not know anything about the d blocks that are stored. For instance, file is stored in one or more split block. The default data size of block is 128MB and also based on the block size, it helps us to figure out the number of blocks are needed for a file.

4. Fault tolerance is used in order to work system functions properly without any data loss . The main achievement is that if some functions has failed, instead of losing all the date , let us have a backup date (duplication date) in different set of date nodes.

5.The main principle in which Hadoop is based are

a. scale “out”, not “up”

b. Assume failures are common and find remedy

c. Move processing to the data

d. Process data sequentially, avoid random access

e. Hide system-level details from the application developer

f. seamless scalability

6.

a. Hadoop cluster is the set of machines running HDFS and MapReduce. Individual machine is called node. Cluster can have as few as one node, as many as thousands of nodes. It is preferable to have more nodes for better performance.

b. The necessity of re-replicate may arise due to

1. A DataNode may become unavailable

2. A replica may become corrupted.

3. A hard disc on datanode may fail

4.Replication factor on the block may be increased