**Assignment 1 – Hoang Duy Vu - 986170**

1. **Can you think of a use case of Big Data?  Explain it briefly. (Do not repeat the ones from the slides!)**

Big data can be used as a Fraud prevention system. Fraud prevention systems are orders of magnitude better at detecting criminal activity and preventing false positives.

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

A key advantage of using Hadoop and HDFS are fault tolerance, cost effective, scalable, and fast by using commodity hardware to build a cluster to solve big problem instead of using HPC (High performance computing).

1. **Explain the term block abstraction in Hadoop.**

Files in HDFS are broken into block-sized chunks 64MB-128MB (default 128 MB), which are stored as independent units. Unlike a files system 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.

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

Fault tolerance in HDFS refers to the working strength of a system in unfavorable conditions and how that system can handle such situation. HDFS is highly fault tolerant. It handles faults by the process of replica creation.

On multiple rack cluster, block replications are maintained with the following policy: (when the replication factor is 3)

* Only one replica is placed on one node.
* And no more than 2 replicas are placed in the same 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**?

Cannot store 560 TB on HDFS if default replication factor is 3.

100 Data nodes capacity 15 TB = 15\*100 = 1500 TB

By default, replication factor is 3, meaning 560 TB \* 3 = 1680 TB, hence data is bigger than the storage HDFS

**But we able to store 560 TB to HDFS if we change replication factor to 2**