

# Assignment 3 of CISC 3018

ZHANG HUA KANG

May 7, 2021

## 1

- File-based Virtual Storage provides the clients the data through directory trees, folders and individual files which is easy to lead the single-path issue.
- Block-based Virtual Storage will break up data into blocks and storing the blocks into separate pieces, each with a unique identifier. When a client or application requests data from a block storage system, the underlying storage system reassembles the data blocks and presents the data to the client or application.
- Object-based Virtual Storage will break data file into pieces called objects, and the storing those objects in a single repository. Each object has a unique ID, and comparing with Block-based Virtual Storage it stores metadata about the file.

## 2

- Volume: The amount of data to be stored, processed or analyzed.
- Velocity: The data throughput to be stored.
- Variety: Different data types and formats to be stored, processed or analyzed.

## 3

### 3.1

- Replication mechanism: Break up data into blocks and store one block on different machines simultaneously.

- Erasure Coding: Break up data into blocks and use XOR operation to generate the parity check block. And stor this three blocks in different machines.

### 3.2

$$\begin{array}{rcl}
 & A1= & 01101110 \\
 \text{XOR} & A2= & 10111001 \\
 \hline
 & A_p= & 11010111
 \end{array}$$

### 3.3

$$\begin{array}{rcl}
 & A_p= & 11010111 \\
 \text{XOR} & A2= & 10111001 \\
 \hline
 & A1= & 01101110
 \end{array}$$

### 3.4

Assume the size of each bolck is  $M$  With the replication mechanism, we will store two pieces of A1 and A2, the total we use is  $4M$ . With the Erasure Coding, we will store one piece of A1 A2 and  $A_p$ , the total we use is  $3M$ . Only used 75% of the space we use with replication mechanism.

## 4

### 4.1

- NameNode: Control the clients' access file, manage the Meta-data of blocks of the clients' file, and manage the file system operations.
- DataNode: Executing the clients' detailed tasks and performing block operations according to the instructions from the NameNode.

### 4.2

- HMaster manages all RegionServers, and stores the Metadata to RegionServers.
- RegionServer: Large logical tables are separated into multiple blocks and stores them in different regionserver.

### 4.3

They both use Master/Slave architecture.

## 5

### 5.1

HMaster is easy to be attack. If the HMaster have some problems, the whole system will not work.

### 5.2

**Similarity** They both use NoSQL database to store data.

**Difference**

- HBase uses Master/Salve architecture
- Cassandra uses Masterlees architecture.