COMP 6231

Group: PJMY_DSD

40197003 Jay Patel | 40193721- Mahek Master | 40195362- Poornaa Bhattacharya | 40189803- Yash Trivedi
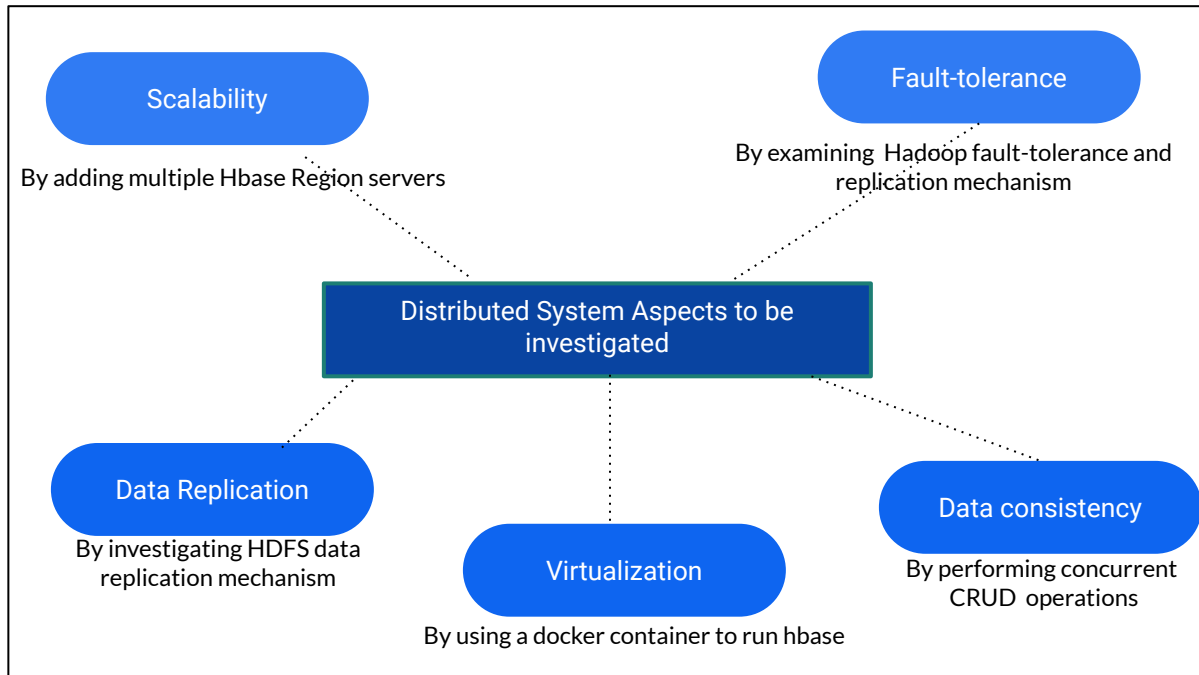
APACHE HBASE

Motivation:

Why Hbase ?

- Open Source noSQL database written in Java
- Scalable, Faster and Fault Tolerant with Random Access on the top of HDFS
- Scalable and Efficient Hashmap
- Flexible Schema
- Comprehensive Java and REST APIs
- Providing effective solutions to Facebook Messenger

Scalability

By adding multiple Hbase Region servers

Fault-tolerance

By examining Hadoop fault-tolerance and replication mechanism

Distributed System Aspects to be investigated

Data Replication

By investigating HDFS data replication mechanism

Virtualization

By using a docker container to run hbase

Data consistency

By performing concurrent CRUD operations

Dataset: 1) Amazon Video Games Dataset: https://s3.amazonaws.com/amazon-reviews-pds/tsv/amazon_reviews_us_Video_Games_v1_00.tsv.gz **(1.2 GB)**

2) IMDB Dataset: https://datasets.imdbws.com/ **(1.5 GB)**