Analyzing Data using Big Data System Techniques

By:
Ali Hussain Alaswad
Abdullah Khadem Aljame

Supervisor:
Dr.Eng.Mouhib Alnoukari
Eng.Anas Abdulaziz

2023/2024

Outlines

01 ABOUT THE PROJECT

02 Project Goals

03 Reference Study

04 System Technologies

05 System workflow Chart

06 Performed Works

07 About Data!

08 Unfinished Works

ABOUT THE PROJECT

What is **Big Data**?

Overview

Big data refers to the large, diverse sets of information that grow at ever-increasing rates. It encompasses the volume of information, the velocity at which it is created and collected, and the variety of the data points being covered

Aim

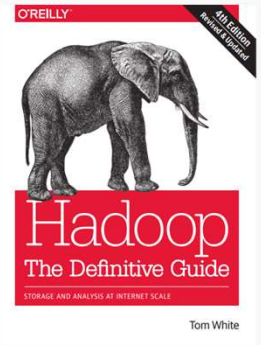
In this project, we are supposed to build a working environment for a big data system and perform data analysis using big data technology

Project Goals

1. Build A Real Big Data system In SPU
2. Analyzing Data using Big Data Techniques
3. Make the environment ready to receive future big data projects

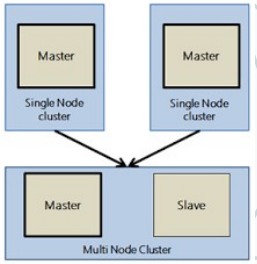
Reference Study

1. Hadoop The Definitive Guide : Tom White



We worked using the reference to build the system and work on the best possible performance

2. Hadoop Multinode Cluster Setup



We chose to work with Multinode Cluster In order to achieve the best performance with the available servers

https://www.projectpro.io/hadoop-tutorial/hadoop-multinode-cluster-setup

System Technologies

Apache Hadoop

Map Reduce

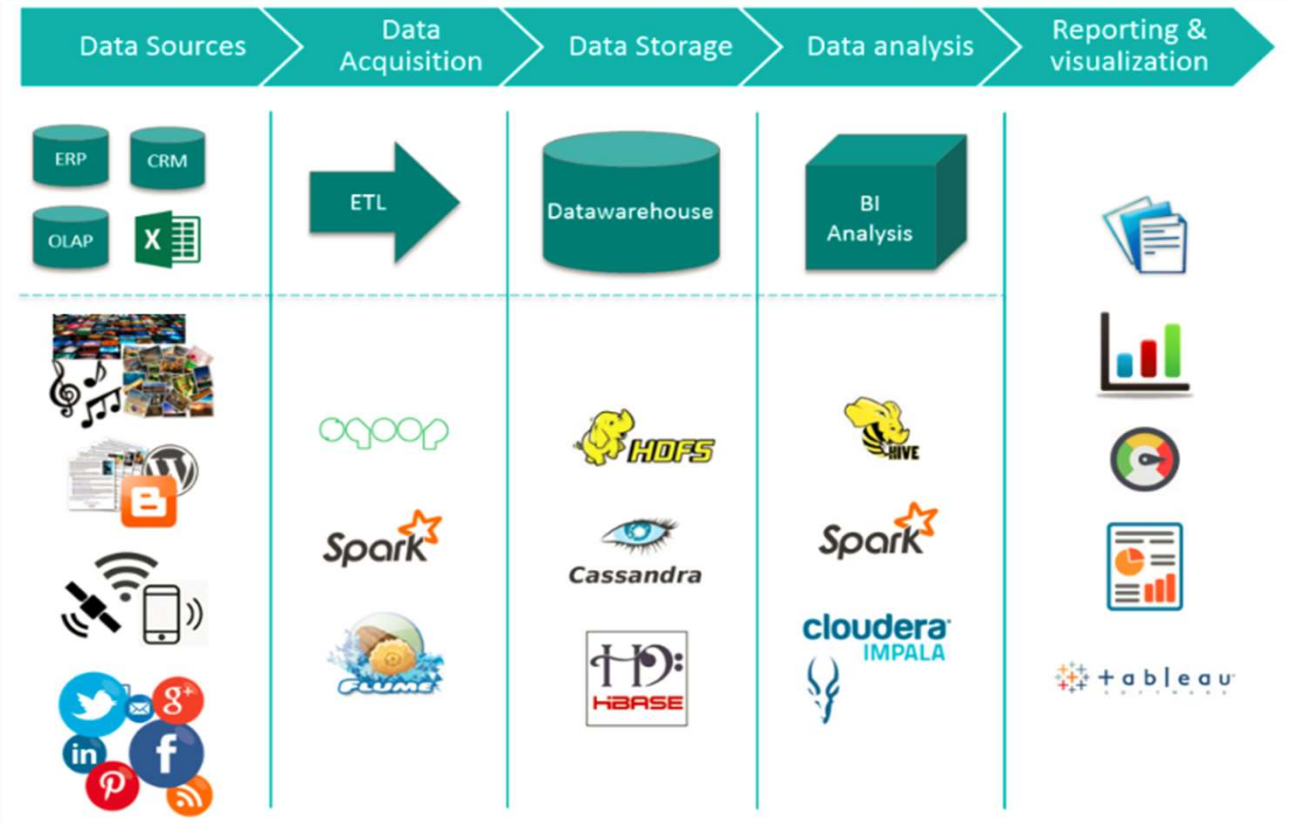
Apache Hive

Python

Apache Pig

Java

System workflow Chart



Performed works

1. Get an overview of the available servers
2. Start working on the servers and install the appropriate operating system
3. Make a plan to build the system in more than one step within 4 weeks
4. Find data to test on the system with metadata

About Data!

- **Forex tick data huge database since april 2020**
- **Below, you'll find the major currency pairs with tick-by-tick historical rates. This data is top-of-book, tick-by-tick market data, with fractional pip spreads in millisecond detail. All timestamps are based on GMT.**
- **Source: https://www.kaggle.com/datasets/joseserrat/forex-april-2020-to-june-2021-tick-data?rvi=1**

Part Of The Data Used

Currencies traded between two countries

Date and time of currency trading

Purchasing price

selling price

USD/MXN	20200401 00:00:...	23.7537	23.7935
1 unique value	2175679 unique values	23.2 – 25.8	23.3 – 25.8 (24.85 - 24.90 Count: 52,956)
USD/MXN	20200401 00:00:50.157	23.7537	23.792
USD/MXN	20200401 00:01:04.399	23.7537	23.7914
USD/MXN	20200401 00:01:09.038	23.7537	23.792
USD/MXN	20200401 00:01:17.469	23.7537	23.788
USD/MXN	20200401 00:01:46.162	23.7452	23.802
USD/MXN	20200401 00:02:03.158	23.7538	23.7919

Unfinished Works

1. Make a plan to analyze more than one step within 4 weeks
2. Analyzing Data using Big Data Techniques
3. Make the environment ready to receive future big data projects

THANK YOU

Do you have any questions?