SOFE 4630U: Cloud Computing
February 15, 2022
Group 11 - Group Report
Project Milestone - Data Storage Implementation: KV + relational

Fajer Zayed (100672347), Ireni Ruthirakuhan (100657302), Raveenth Maheswaran (100704540), Yale Wang (100673933)

GitHub Link: https://github.com/fzayed/Project-Milestone-Group-11.git

Google Cloud has another processing service called DataProc. Name another processing service that is usually used in the cloud environment (not necessarily GCP).

- Amazon EMR

Compare between it and both Dataflow and DataProc. Your comparison may include but is not limited to the major differences, advantages, disadvantages, and limitations.

- Similarities
    - Both give you ETL data warehousing
- Major Differences
    - Data Proc gives you access to Hadoop
    - Dataflow gives a platform to use Apache Beam
- Advantages of Data Proc
    - Simplifies operations
    - Easy install/resize
- Disadvantages of Data Proc
    - No choice of version choice for Hadoop/hive/spark stack.
    - Cannot pause/stop a cluster.
- Advantages of Dataflow
    - Don't need to manually balance
    - Automatically scales
- Disadvantages of Dataflow
    - Not as flexible
    - Can't customize implementation

Suggest a practical application using both stream and batch processing that can be applied to a given dataset.

- A social media application that uses
    - Batch processing to track logins
    - Stream processing to track social media interactions

Its impact.

- This application can have an incredible scalability to number of users
- Social media applications are good for maintaining many connections

The used dataset (size, schema/structure).

The used dataset was the same images and csv files from Lab 3.

List of other tools (AI, clustering,…) needed to implement that application

The other tools used in the videos include:

- Analyzing data with BigQuery
    - To ingest user activity
- Vision API
    - To analyze content of images using pre-trained neural networks
- Cloud translation
    - To provide the service in many languages