Project Milestone-- Data Processing: Dataflow- apache beam

Faazil Shaikh - 100707829

6. Google Cloud has another processing service called DataProc. Name another processing service that is usually used in the cloud environment (not necessarily GCP). Compare between it and both Dataflow and DataProc. Your comparison may include but is not limited to the major differences, advantages, disadvantages, and limitations.

- DataPrep is the other processing service.
- Differences:
    - DataPrep - fully automates cluster setup, integrated BigTable and BigQuery
    - DataFlow - fully automates cluster setup, integrated with Apache Beam
    - DataProc - cluster setup is done manually, integrated with Apache Spark and Hadoop

- Advantages:
    - NoOps system that is fully managed
    - Easy to use
- Disadvantages
    - Not as easy as DataProc
    - Mainly useful for User Interface data only
- Limitations
    - Max number of datasets is 1000
    - The max file size exported has to be less than 100mb

7. Suggest a practical application using both stream and batch processing that can be applied to a given dataset. It's expected to use the dataset uploaded in the third milestone but you can use any other dataset. If you decided to use another dataset, It should maintain both variety and huge volume. Your report should include but not limited to

- **The application.**

  Long-term mapping, navigation and detection system (lidar) for changing environments

- **Its impact.**

There's a great deal of data for environmental change including things like:

- Vehicles
- Humans
- Lighting based on the day and time and seasons
- Overall change in environment (new buildings, plants, etc)

- **The used dataset (size, schema/structure).**

  - Uses the Long-Term Vision and Lidar Dataset or otherwise known as the NCLT dataset
  - Sample structure for some of the data includes, dates, time, snow, foliage, sky
- **List of other tools (AI, clustering,…) needed to implement that application.**
  - Machine Learning to learn the data from the dataset to observe environmental changes