Ashwin Shanmugam
100700236

1) The other data processing service provided by the Google Cloud Platform is **DataPrep**.

   The differences between Dataproc, Dataflow and DataPrep are :

   - It depends on how reliant the system is. If a system is heavily reliant on the usage of hadoop clusters, then, it is viable to use Dataproc over Dataflow or Dataprep.

   - Dataflow and Dataprep provide provision to completely automate the management of clusters.

   - The three mainly differ on how they are used. Dataproc is best for Machine Learning and Data Science environments. Dataflow is used for batch and stream processing of data. And DataPrep uses a user interface driven approach for the processing of data.

   - DataProc and DataPrep are easy to use while DataFlow is quite hard to get it set up.

Advantages of DataPrep :

- Since it follows a user interface data driven approach. It provides the user with an easy to use UI and displays data in a neat and tabular format.

- DataPrep allows users to virtually examine data by converting it to various formats such as CSV, JSON, etc.

- DataPrep can process data in the range of several Terabytes.

DisAdvantages of DataPrep :

- It does not perform well with Hadoop clusters.
- Mainly used for systems focussed on user interface.

2)

Application : The LIDAR dataset is used for various purposes such as Land Surveillance, Forestry and Farming,