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.

DataProc

* Doesn’t have stream processing
* Unable to preprocess for machine learning
* Contains machine learning with Spark ML
* Used for data mining and analysis in datasets of known size

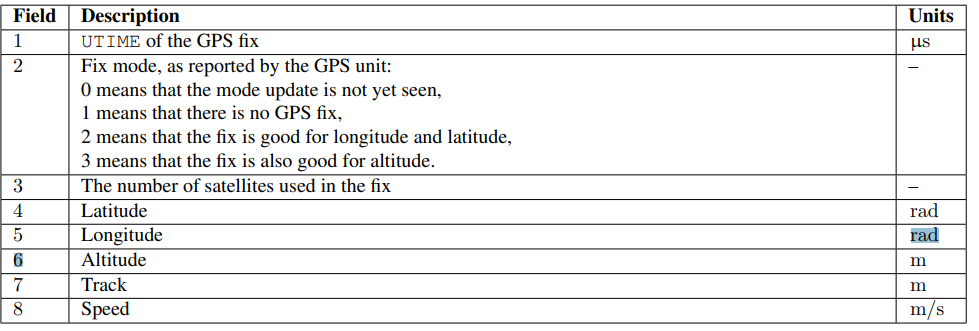
Cloud Dataflow

* Can preprocess for machine learning
* Doesn’t have iterative processing and notebooks
* Can handle datasets of unpredictable size

Databricks Lakehouse Platform

The application reads the dataset and converts the ‘UTIME’, which is a 16 digit integer timestamp, into a readable timestamp. This application will also detect if at that time the robot is moving or not based on the speed provided in the dataset.

The used dataset is the gps data of the NCLT dataset. It contains 8 columns with each column description stated down below:



Proposed Pipelines:

