Milestone IV: Sarthak Sharma 100604428
Report/Summary:

Q6> **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.**

- Service: Trifacta under google cloud
- Differences: DATAFLOW- It's a very simple streaming service (used in both stream and batch) that focuses on cost efficiency and keeping latency and processing time to be very close to a minimum. DATAPROC- uses a scaling technique that is optimal for cluster data and processes.
- Advantages: DATAFLOW- easy to implement and very efficient in terms of direct streaming and data processing. Better for real-time. DATAPROC- the fact that multiple processes can be carried out at the same time makes it a more flexible system.
- Disadvantages: DATAFLOW- the service is still new and premature given the fact that it lacks many features and also the community that uses this service is fairly small. DATAPROC- you can still be left dealing with latency issues when processing big data sets compared to dataflow and dataprep.
- Limitations:
    - DATAFLOW- input of 20k shards only; dataflow job 100; messages interval- 150000 per 30 seconds; element size upto80Mb
    - DATAPROC- workflow request time 400 per 60 seconds; scaling request 400; cluster data operation requests 200 per minute; Get/pull requests 7500 per minute.