

Project Milestone 4: Data Processing

Owais Quadri 100697281

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.

Another processing service used in the cloud environment is Apache Spark. DataProc provides a cluster on GCP, where you gain access to processing tools. Dataflow gives a place to run Apache Beam jobs and you don't need to worry about cluster issues such as load balancing. Apache spark can be run on Dataproc to provide efficient processing on the hadoop clusters provided by dataproc.