



Google Cloud Platform — Big Data Analytics Solution: A Case Study



Blessing Adesiji · [Follow](#)

3 min read · May 10, 2020



53



1



To show my understanding of Google Cloud Platform (GCP), I have decided to write a short summary on how its big data analytics solution can be leveraged by a small data-driven company that relies on getting streams of data, process and analyse for insights.

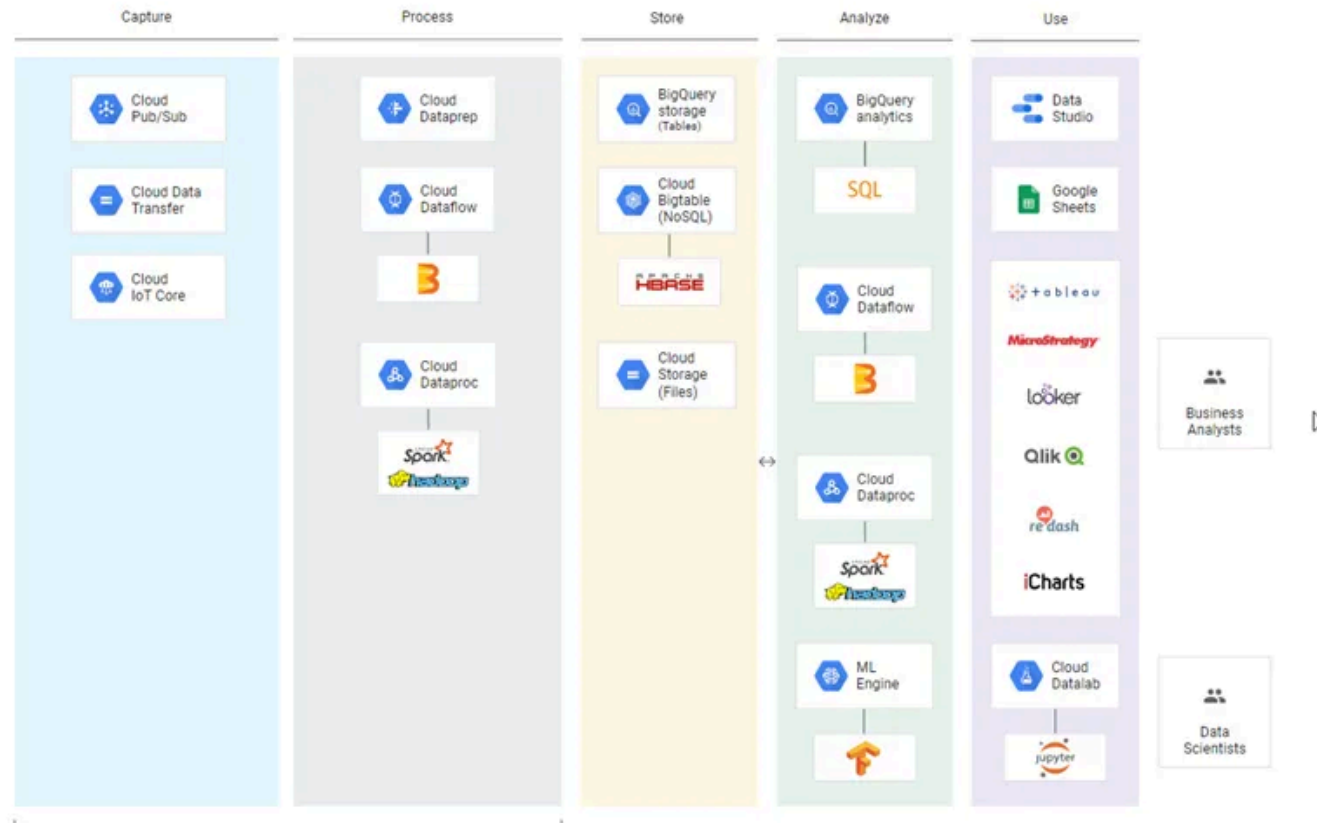
This is to express my interest in applying for the role of Cloud Data Engineer Intern at Cloud 7. I do hope that this short summary will portray my

motivation to work and learn as a cloud data engineer in the firm.

Why GCP for Big Data Analytics

Google Cloud Platform is a suite of cloud computing services that runs on the same infrastructure that Google uses internally for its end-user products. It lets you build, deploy, and scale applications, websites, and services on the same infrastructure.

Based on the distributed services that Google offers, complexities in data analytics can be easily handled on the cloud. Insights generation processes are offered on serverless, integrated, and end-to-end data analytics services. The platform handles scalability, security, compliance and performance needs with cost-effective approaches.



The Big Picture — Google Cloud Platform

Products that can be leveraged upon by the data-driven company.

Google cloud platform offers products that enable an end-to-end data analytics process.

Assumption — The company ingests large data from distributed sensors or IOT devices.

1. Capture Data- Cloud Pub/sub or Cloud IoT Core

Cloud Pub/sub lets you ingest millions of events per second from anywhere (and lets you publish it anywhere) in the world via an open API. Thus, making data ingestion, processing, and storage easy for data engineers.

Cloud IoT Core is a fully managed service that allows you to easily and securely connect, manage, and ingest data from millions of globally dispersed devices.

2. Process Data — Cloud Dataflow

Cloud Dataflow enables faster streaming and batch data pipeline development with lower data latency. The team can focus on programming as Google manages the resources and removes complexities in data engineering workloads.

3. Store & Analyse Data

BigQuery

This product enables gigabytes to petabytes data analytics using SQL at a fast speed. Making it easy to query streaming data in real-time and get up-to-date information on business processes. It also offers robust security as data is encrypted with encryption keys.

There are sub-products of BigQuery that are of benefits, namely, BigQuery ML and BigQuery BI Engine.

BigQuery ML supports analytics team to build fast machine learning models with structured and semi-structured data directly inside BigQuery using simple SQL. The models can be exported to the Cloud AI Platform for integration into applications and the company's products. Also, GCP has an AutoML tool that can be used by developers with no machine learning coding experience.

BigQuery BI Engine supports the analytics team to analyze large datasets with fast query response time and high concurrency. It can be connected to Google's visualization tool (Google Data Studio) and other products (Tableau, Qlik Looker), for analysis and report generation.

Other Google products that can be leveraged upon for a seamless analytics process are: Cloud Data Storage for tables and files, Cloud Dataproc

4. Use — Data Studio for Visualization and Reporting

Data Studio is a visualization and reporting tool. It enables the creation of informative and interactive reports. From other GCP analytics tools, processed and analyzed data are imported to create custom visualizations tailored to specific needs.

The platform also supports building report template, embedding reports, and publishing reports as template and connector.

In conclusion, the data-driven company can simply host its data processes on Google Cloud Platform for a seamless end-to-end operation.



Written by Blessing Adesiji

114 Followers · 50 Following

Follow