

Mid way Report

On

Network and Service Automation



THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Faculty Mentor

Dr. Shashank Sheshar Singh
shashank.sheshar@thapar.edu

Industrial Mentor

Mr. Ashish Priydarshi
ashish.priydarshi@ericsson.com

Submitted By:

Name: Shubham Tiwari
Roll No: 101916126
Batch: 4CS12
stiwari_be19@thapar.edu

Computer Science and Engineering Department

Thapar Institute of Engineering and Technology, Patiala

1. Introduction/Background:

Automating data ingestion and refining architecture of company's internal platform which is based on google cloud platform. That will reduce user cost and provider cost. Basically, Light weighting that platform and automating data ingestion and aggregation work which comes from customer end. So ingested data can be used for monitoring etc. purposes easily and efficiently.

Area of study is optimal use of google cloud function, Big Query, cloud dataproc and composer. That help us to reduce cost, provide efficiency and light weighting the architecture.

2. Problem Statement/Domain:

Organizations are increasingly adopting cloud computing solutions **to drive digital transformation, reduce operational costs, and increase business agility**. Google Cloud Platform (GCP) is a popular cloud computing platform that offers a wide range of services, including compute, storage, networking, and data analytics. However, organizations face several challenges while using GCP, including security, compliance, cost optimization, and lack of expertise.

The domain for this problem statement is cloud computing and specifically the optimal use of Google Cloud Platform (GCP). This includes all aspects of cloud computing, such as infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS). The domain also covers the various services and tools provided by GCP, such as Google Compute Engine (specifically cloud dataproc), Google Cloud Storage, Google Cloud SQL, and Google Kubernetes Engine. Additionally, the domain includes areas such as cloud security, compliance, cost management, and governance. We studied cloud computing in college as well, that helped me a lot for diving in cloud computing infrastructure quickly and adopting new things easily.

The problems are exploratory that needs a lot of brain storming. We as a team are involved in solving these following problems for organization.

- i. **Overspending:** Organizations overprovision their resources, leading to a waste of resources and increased costs on cloud platform like GCP.
- ii. **Lack of Automation:** Manual resource management can be time-consuming and error-prone, making it difficult to optimize resources for better utilization. Although there is various automation needed i.e., Network automation, Service automation, Process automation, Test automation, Supply chain automation, Data automation, Security automation etc.
- iii. **Security and compliance:** Organizations need to ensure that resource utilization practices comply with security and compliance policies, which can limit their ability to optimize resource usage.
- iv. **Light weighting Architecture:** There are some little costumers in telecom industry of company they cannot pay that much amount. So, we are trying to rearchitected that platform for those costumers.

3. Technique/Tools/Technologies used:

For above problem statement we used following tools:

Google Cloud Console - To access and manage GCP services, deploy and manage applications, and monitor the performance of their infrastructure.

Google Cloud SDK - This is a set of command-line tools that allow users to interact with GCP services from their local development environment using APIs.

Cloud Deployment Manager - To create and manage cloud resources using templates. This allows for easier and more efficient management of cloud infrastructure.

Tools	Functionalities
Compute engine	VMs, GPUs, TPUs, disks
BigQuery	Data warehouse/analytics
Pub/Sub	Global real-time messaging
Dataflow	Streaming analytics service
Composer	Managed workflow orchestration service
Dataproc	Managed Apache Hadoop
Dataprep	Visual data wrangling
Cloud Run	Serverless for containerized Application
Cloud Function	Event-driven serverless functions
Bigtable	Petabyte-scale, low-latency, non-relational
Logging	Real-time log management and analysis

Google Cloud Training - To help users learn how to use GCP effectively.

Third-party tools - To manage and optimize their use of GCP, such as cloud management platforms, cost optimization tools, and security solutions.

4. Block diagram/Architecture/Methodology:

I and one of intern colleague developed a cloud function dataset_conf_fun for a specific project with the help of seniors (Product area manager, data engineer, solution architect, cloud data architect) using python as programming language and deployed on google cloud function. (See Fig. 1.)

This module has mainly two task first is read content from the file and second one is after some pre-processing, mapping, manipulation etc. data is ingested into table. There are mainly two BigQuery table one is Aggregation_mapping and other one is data_configuration. Obviously, we faced some challenges during implementation but here I'm explaining only outline of whole methodology.

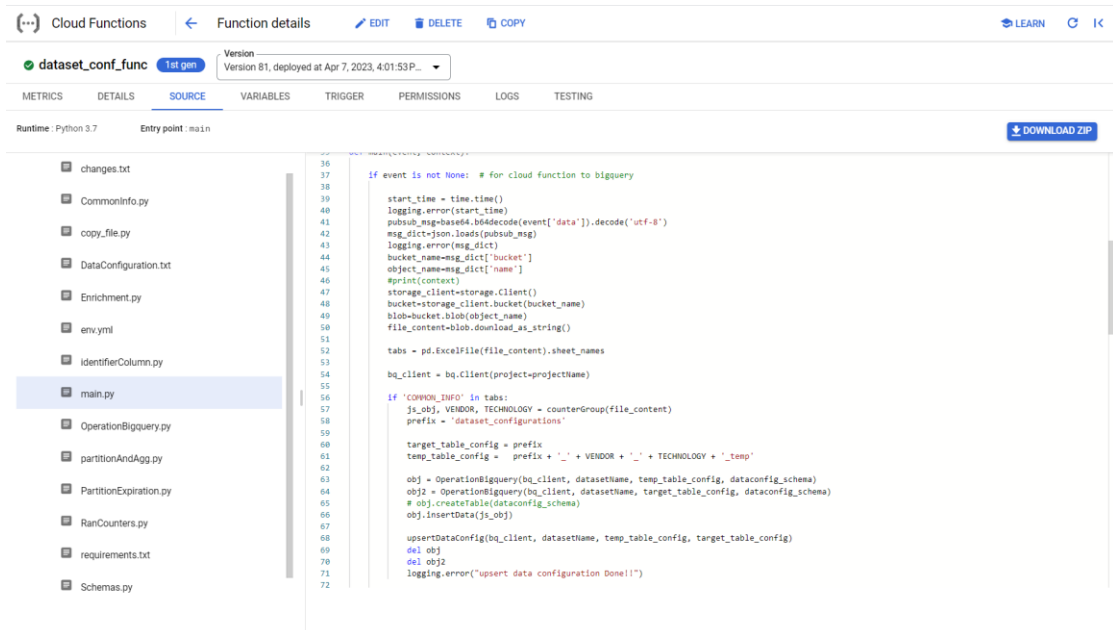


Fig. 1. Deployed Cloud Function Screenshot

The working of above cloud function is –

1. Ingest: When new data file (excel file) is ingested to google cloud bucket which is a cloud storage container. We set up pub/sub between services or application by defining topics Cloud pub/sub publish message and then subscriptions which allow services (our cloud function) to receive the messages published on those topics.
2. Process: Once new file arrives in bucket message will be published and we subscribed our function on that topic then function will start processing the file the cloud function fall in process part of architecture. (Fig. 2.)
3. Store: Once all processing is done then cloud function ingest the data to BigQuery table for further processing.
4. Analyse: Then using BigQuery analytics further processing become easy and the cloud dataproc start its job.

Excel files comes from customer end that have different set of information about key performance indicator (KPI). Information of data is against company policies.

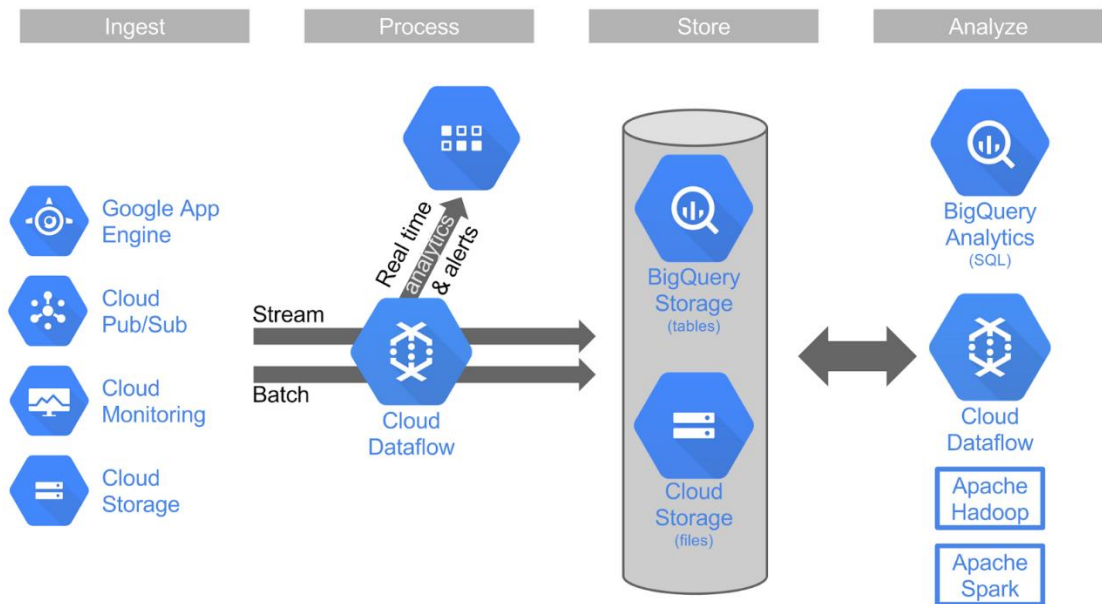


Fig. 2. High level Architecture

The above explanation is all about automation data ingestion pipeline by using GCP functionalities. So, till now we are able to automate about 98% task and 2% task comes in CI/CD.

Once data reach in BigQuery table then our new task starts because the above-mentioned architecture is for big customer but there is small customer too and they need our services on small level, but our architecture is too expensive.

So, in upcoming next few months we will work on light weighting the architecture and one more point the above-mentioned architecture is just for reference the original architecture is company's internal architecture which somehow similar to this one in terms of ideology.

(Companies resist for sharing their specific detail due to non-disclosure agreements)