Google I/O Extended

GDG Sonargaon

Google I/O Extended Bangladesh

From Chaos to Clarity:

Data Engineering Magic

Wednesday 30th August, 2023

@ Daffodil International University

Google I/O Extended

Google I/O Extended
Bangladesh

Sanower Abedin Tamjit

Sr. Data Engineer (Team Lead)

Mi-C3 International Ltd. | Affectli

linkedin.com/in/sanower-tamjit/

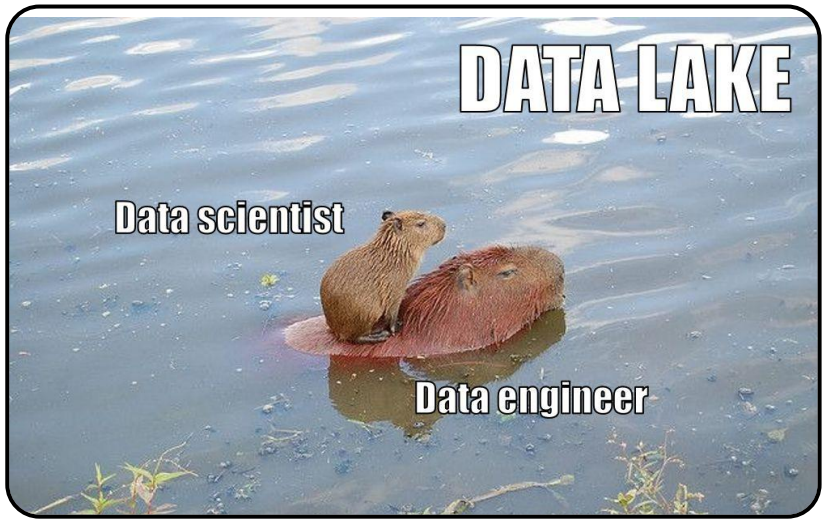
Agenda

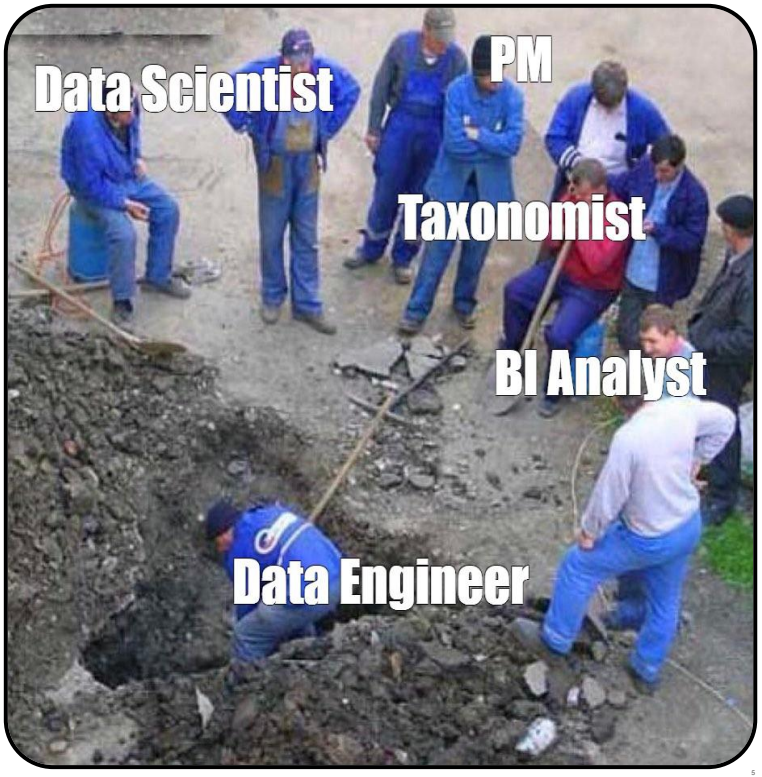
Google I/O Extended Bangladesh

- Chaos to Clarity: Data Engineering Way
- How Data Engineering is evolving
- Use Cases
- Data Platform on GCP
- GCP DE Tools & Learning Path

So, What Actually Data Engineers Do?

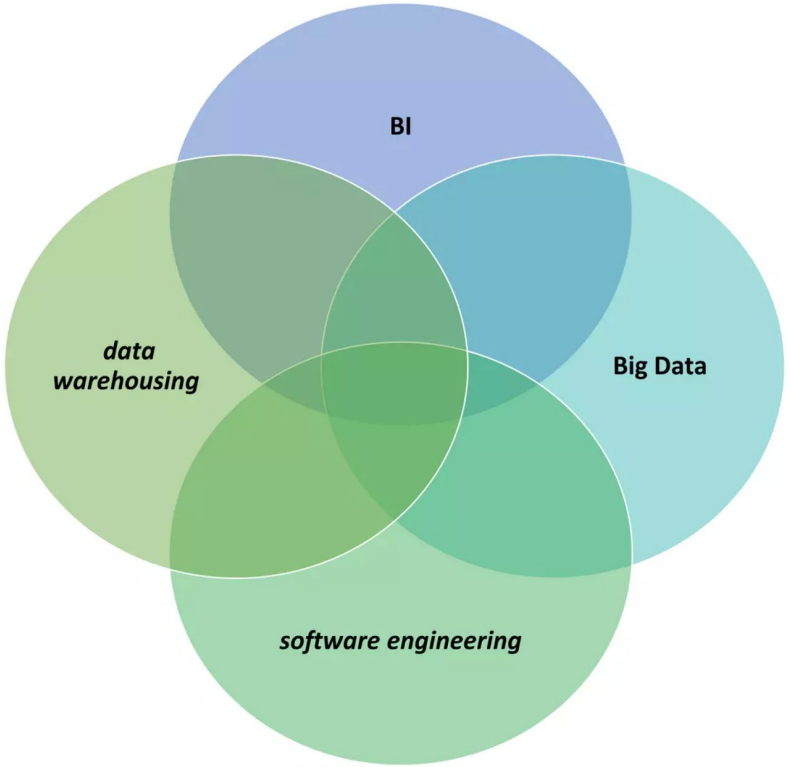
We are doing this





Data Engineer is

Who designs and maintains data pipelines for efficient data storage and processing.



Let’s Solve Jigsaw Puzzle

Chaos (Raw Data):

Google I/O Extended Bangladesh

At the beginning of the data engineering process,

- We encounter raw and unprocessed data
- Data can come from various sources, such as sensors, databases, social media, and more.
- It's often messy, unstructured,

Data Ingestion:

Google I/O Extended Bangladesh

The first step is to gather data from various sources.

- From Databases
- APIs
- IoT/IIoT Devices
- Streams
- Files and more...

Just like gathering puzzle pieces from different places.

Data Transformation:

Google I/O Extended Bangladesh

Data engineers then clean and transform the raw data. This step involves

- removing inconsistencies, errors, and duplicates,
- restructuring the data into a unified format,
- validating the data

Think of it as arranging puzzle pieces in a way that they fit together perfectly..

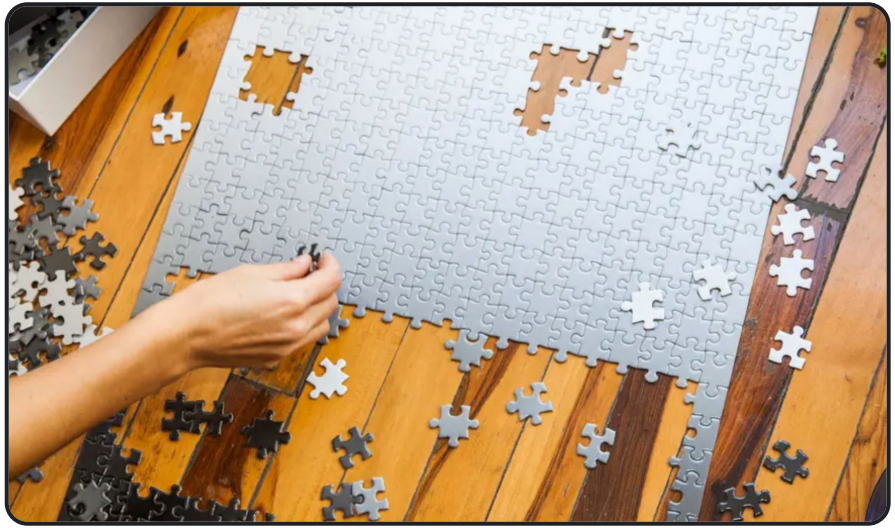
Data Storage:

Google I/O Extended Bangladesh

Once cleaned and transformed, data is stored in

- Databases,
- Data lakes, or warehouses.

This is similar to putting the pieces of a jigsaw puzzle together in a specific order. Once the pieces are in the correct order, they form a complete picture.

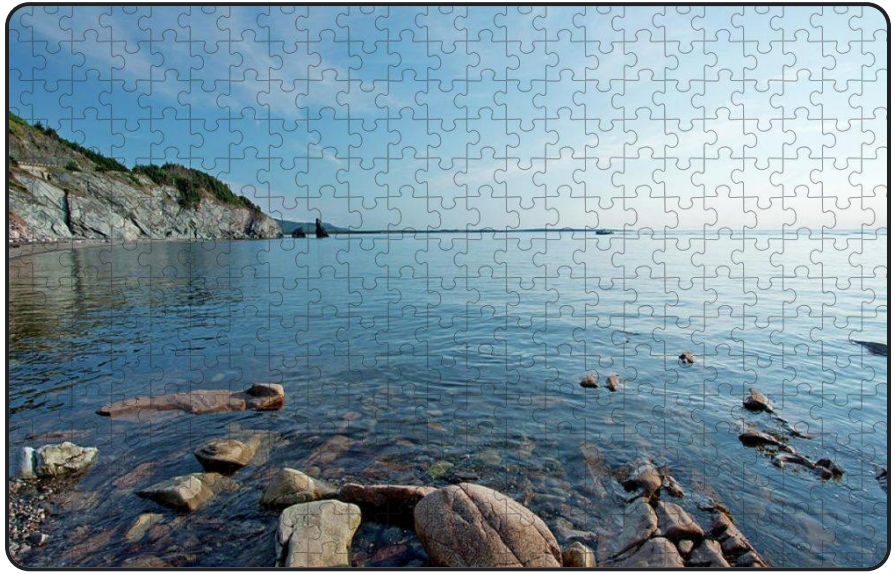


Clarity (Actionable Insights):

Google I/O Extended Bangladesh

The ultimate goal of the data engineering journey is to achieve clarity - Through

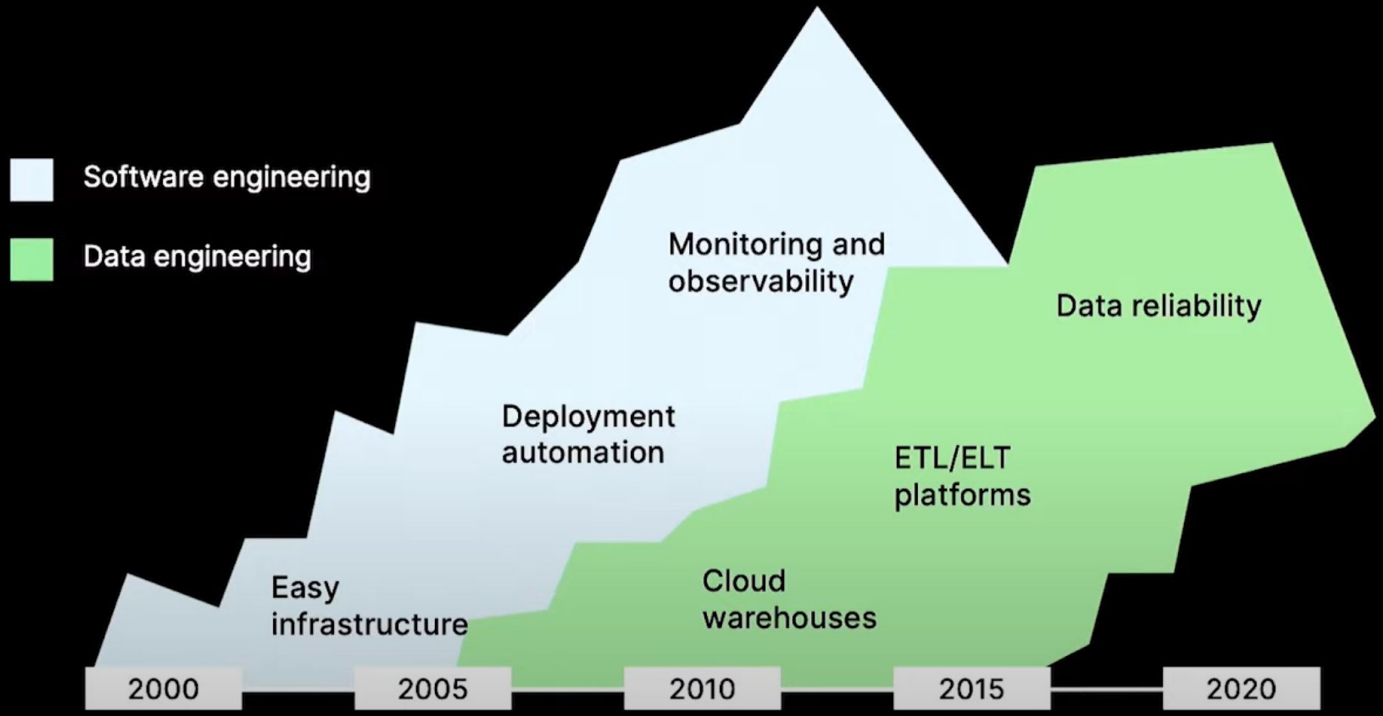
- BI Analytics
- Predictive and Prescriptive Analysis



Just like a fully assembled puzzle reveals a clear image, processed and visualized data unveils valuable insights.

Data Engineering is evolving

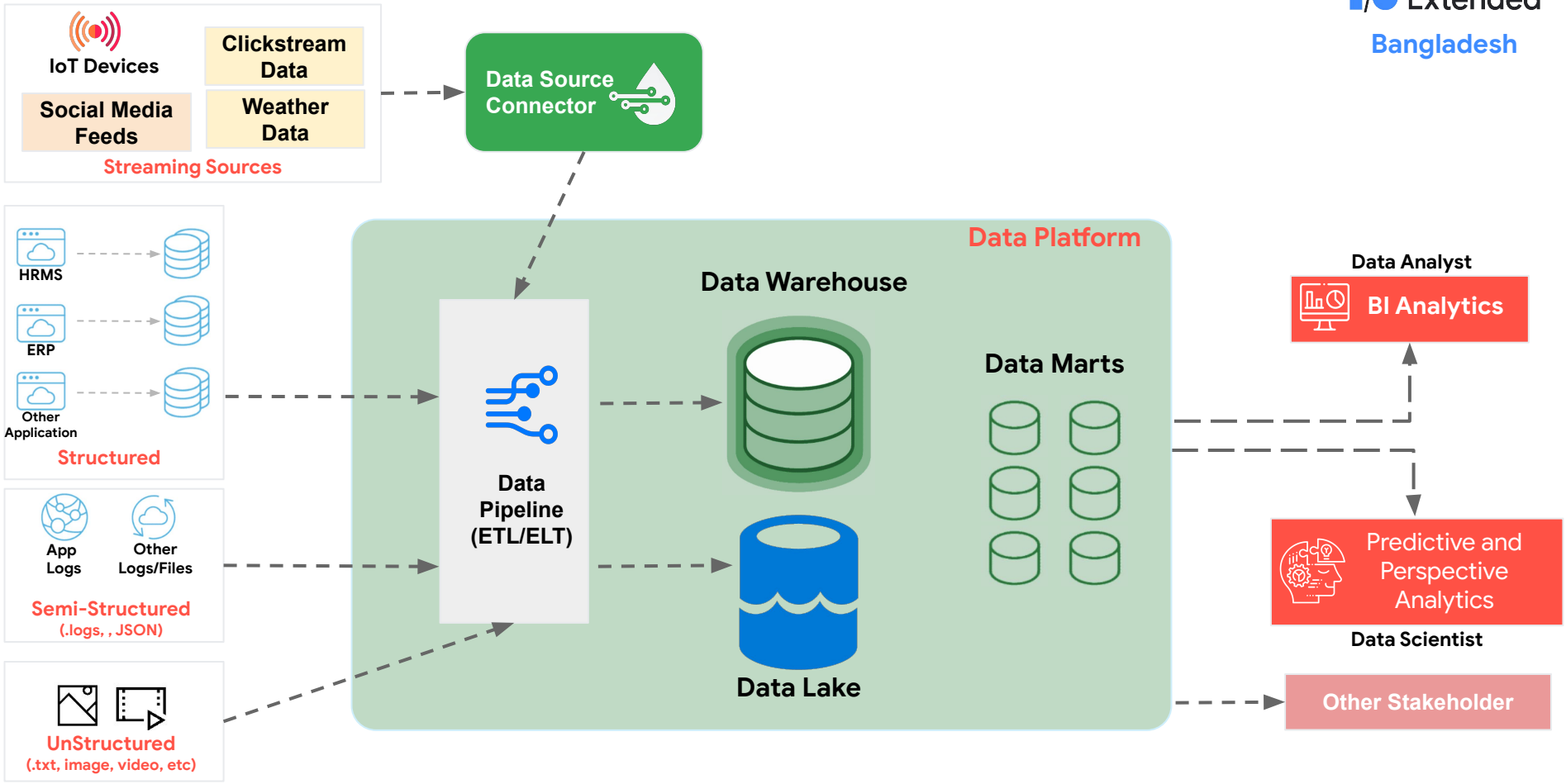
Google I/O Extended Bangladesh



Sources: https://www.datacouncil.ai/talks/data-reliability-engineering-a-new-approach-to-data-quality?

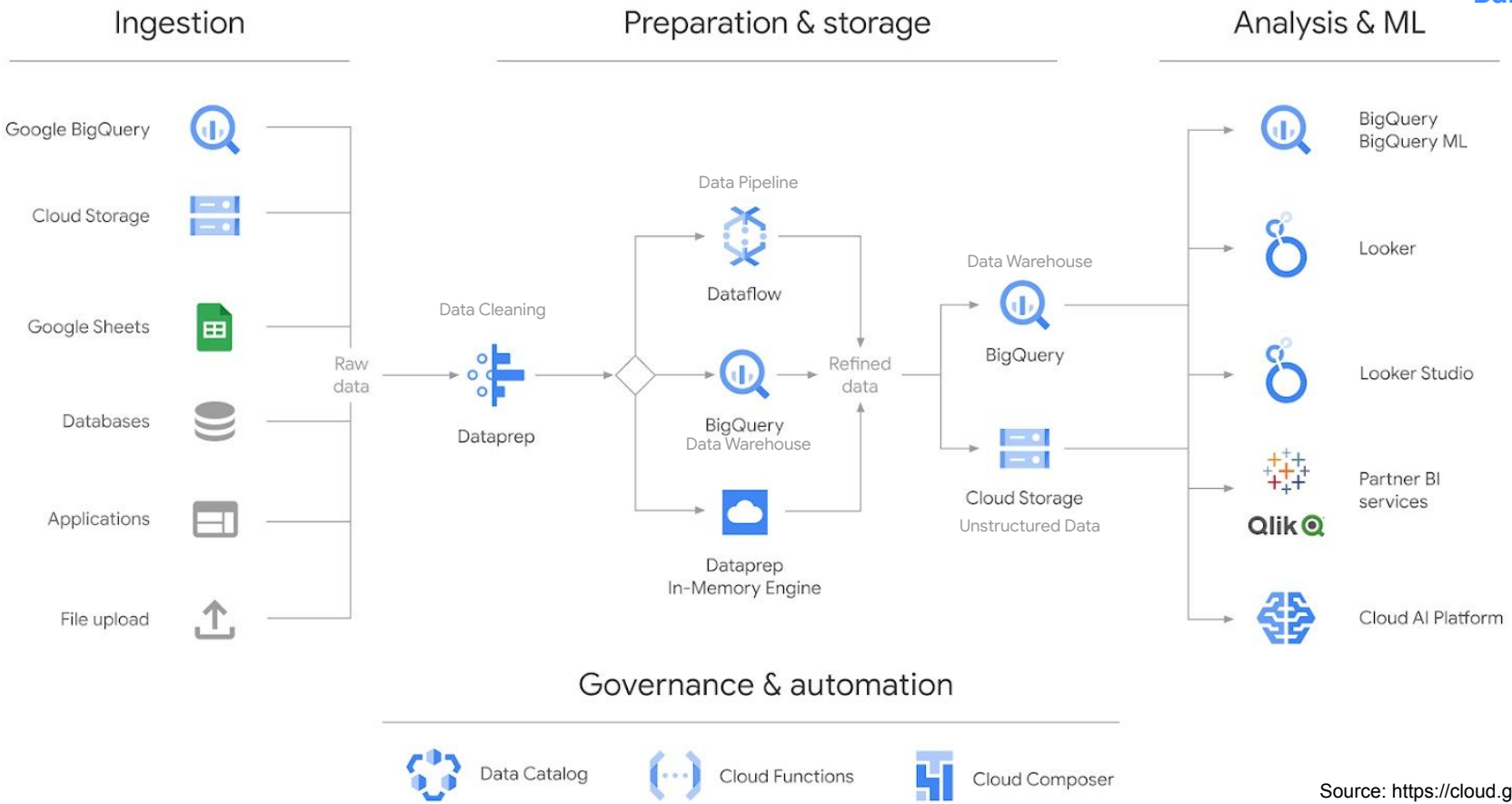
Data Platform Architecture

Google Extended Bangladesh



A Typical Data Platform on GCP

Google I/O Extended Bangladesh



Source: https://cloud.google.com/dataprep

GCP Data Engineering Tools

Google I/O Extended Bangladesh

Database Management


Cloud SQL


AlloyDB for PostgreSQL


Cloud Bigtable


Cloud DataStore


Cloud Spanner

Data Pipeline


Dataflow


Data Fusion


Dataproc


Composer

Data Warehousing & Governance


BigQuery


Cloud Storage


Data Plex

Streaming Data Analytics


Pub/Sub


Dataflow


BigQuery Streaming


Apache Kafka® on Confluent Cloud™

Batch Ingestion

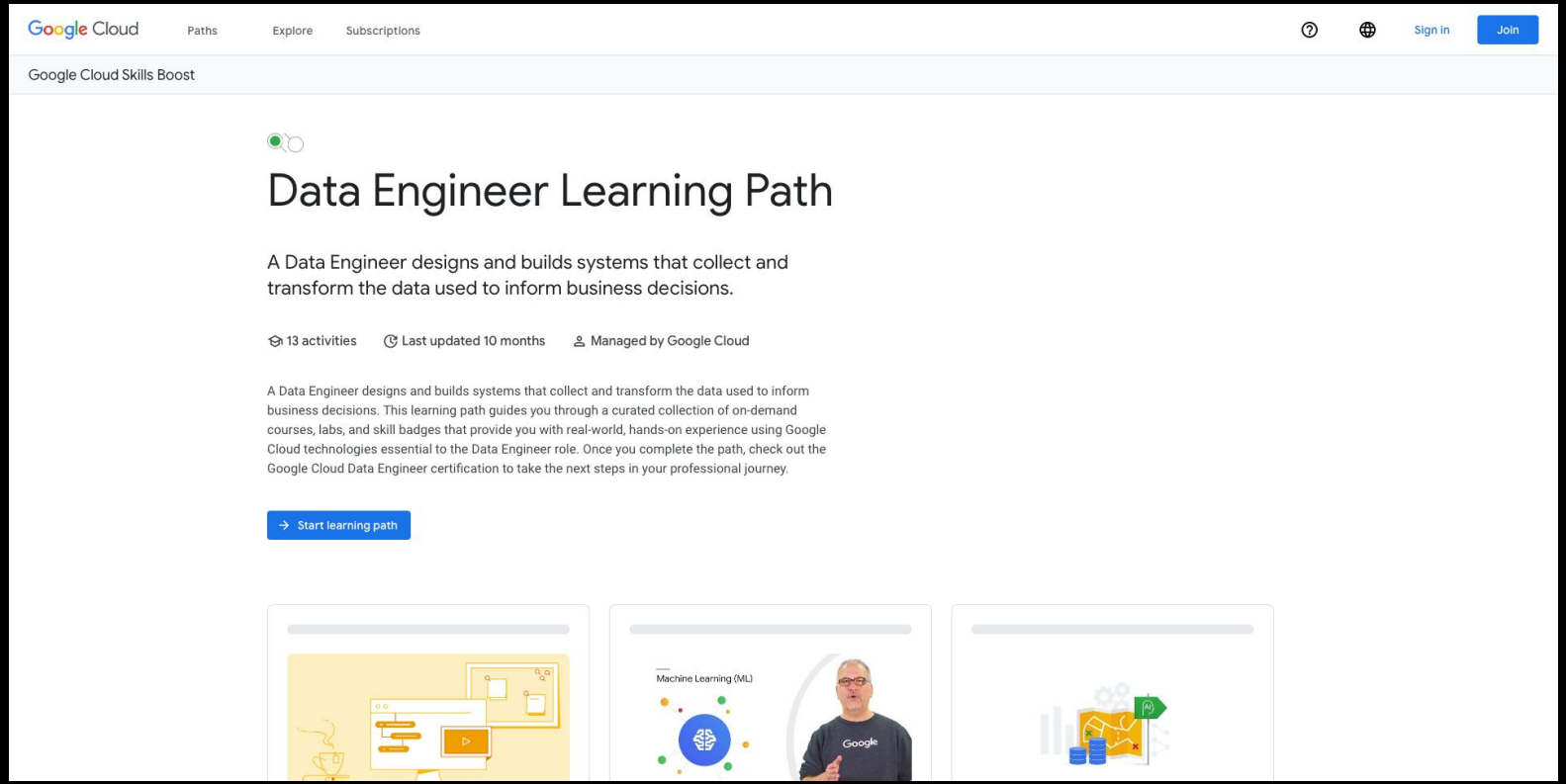

BigQuery Transfer Service


Dataflow (Batch Mode)


DataPrep

Data Engineering Learning Path

Google I/O Extended Bangladesh



https://www.cloudskillsboost.google/journeys/16

Google I/O Extended

GDG Sonargaon

Google I/O Extended Bangladesh

Thank You

Sanower Abedin Tamjit

Sr. Data Engineer (Team Lead)
Mi-C3 International Ltd. | Affectli

linkedin.com/in/sanower-tamjit/

github.com/SanowerTamjit

facebook.com/sanower.tamjit.1