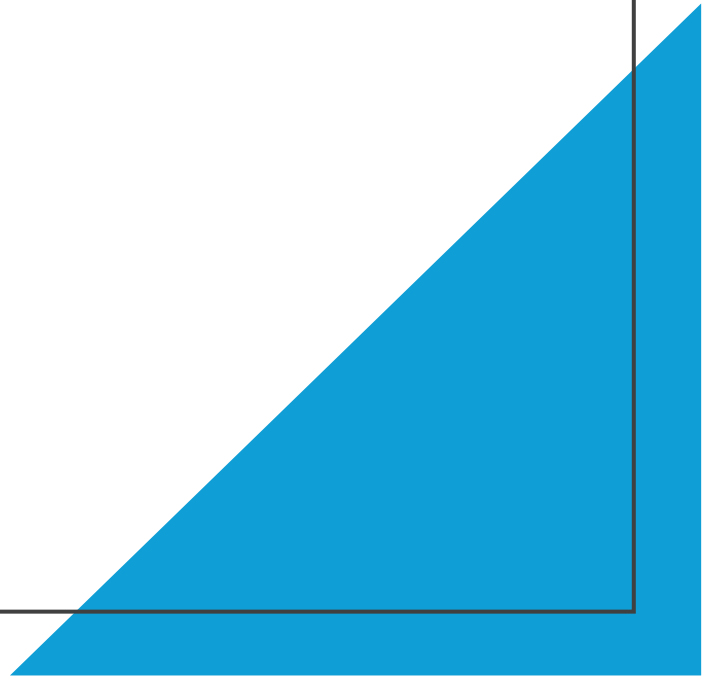# ETL Data Pipeline on Google Cloud

Jharna Kumari

Problem Statement

Data Pipeline to extract employee data and develop a dashboard to visualise the employee data securely

- **Data Extraction:** Multiple sources such as databases, APIs, csv files, etc.
- **Data Transformation:** Identify sensitive data
- **Data Loading:** Securely load data into BigQuery
- **Dashboard Visualisation:** Web-based dashboard on Power BI

Why ETL?

**EXTRACT**   **TRANSFORM**   **LOAD**
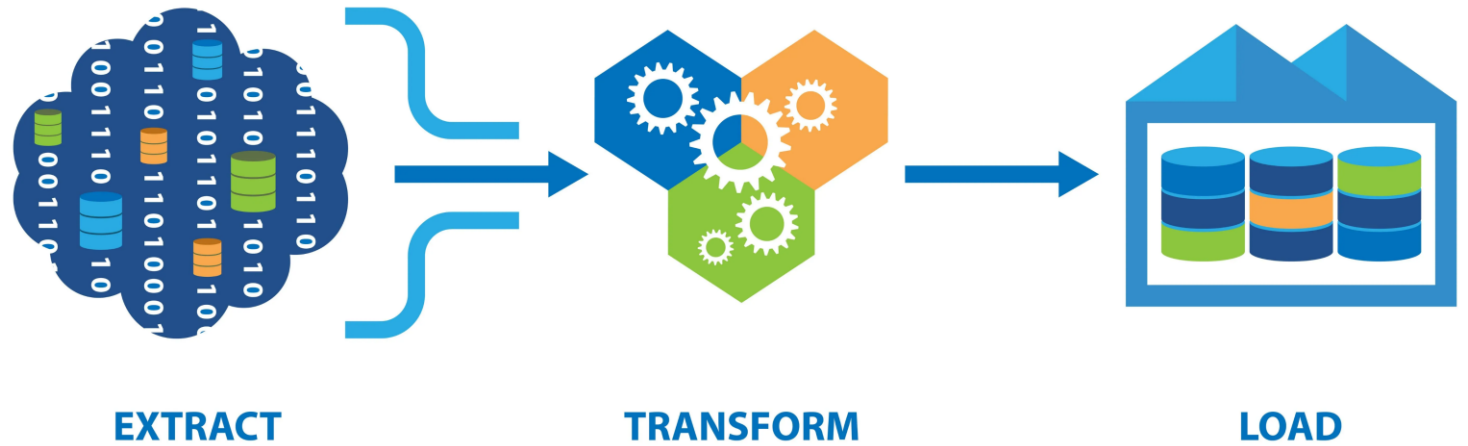
Architecture

1. **Data Extraction**: Extract data using python.

2. **Data Masking**: Apply data masking & encoding techniques to sensitive information in Cloud Data Fusion before loading it into BigQuery.

3. **Data Loading**: Load transformed data into Google BigQuery tables.

4. **Orchestration**: Automate complete Data pipeline using Airflow (Cloud Composer)
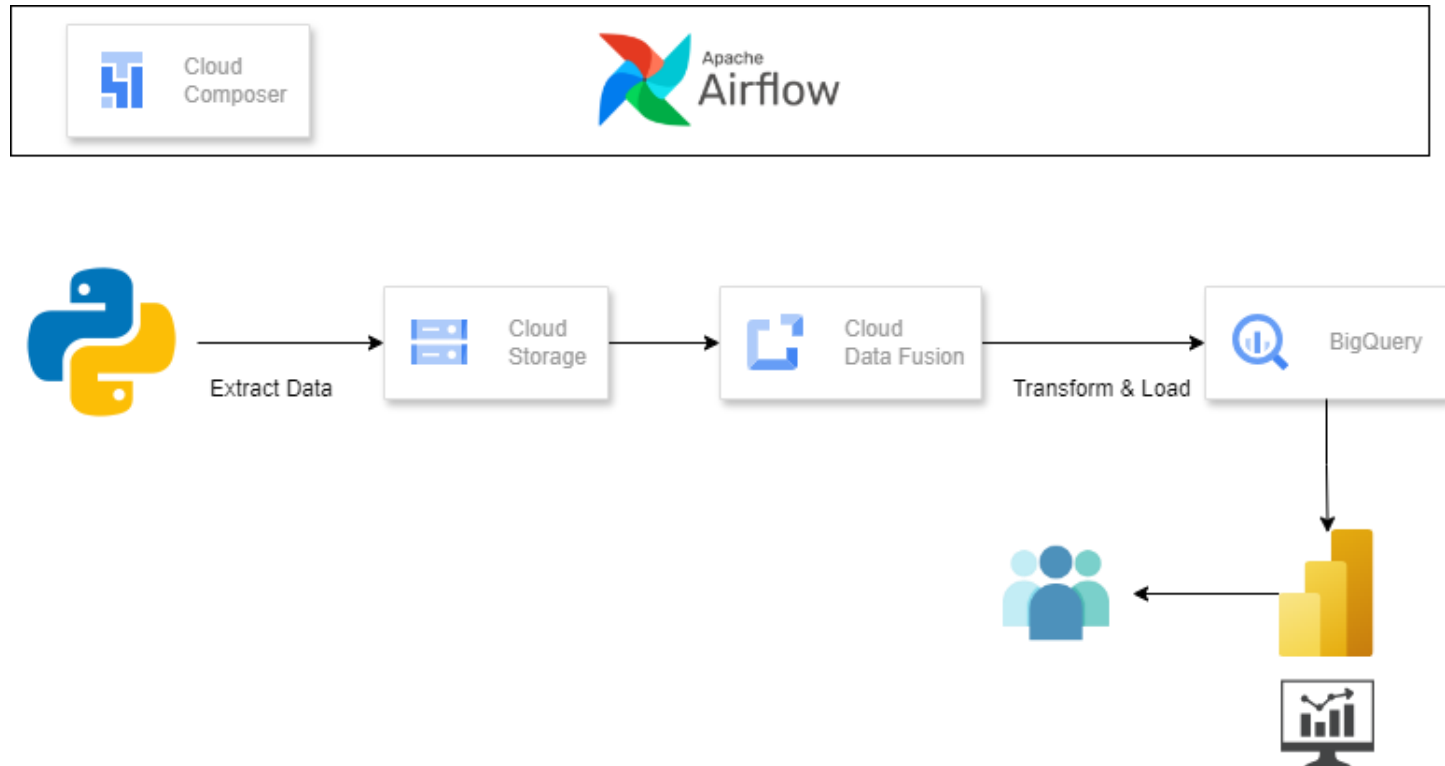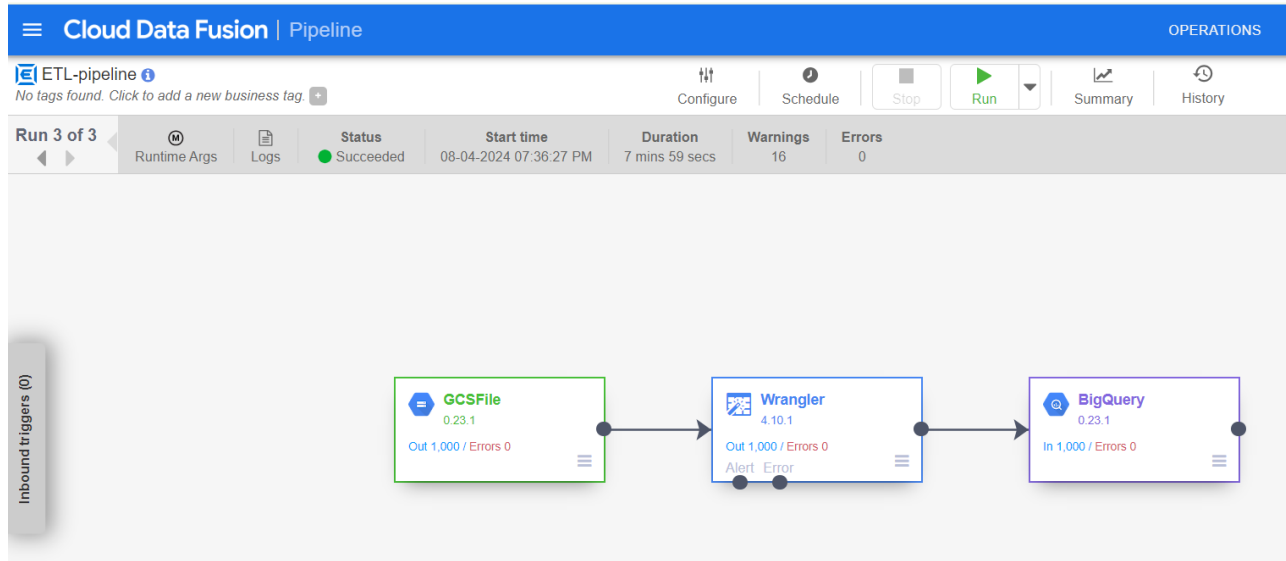
Data Fusion

ETL Pipeline

Airflow Tasks
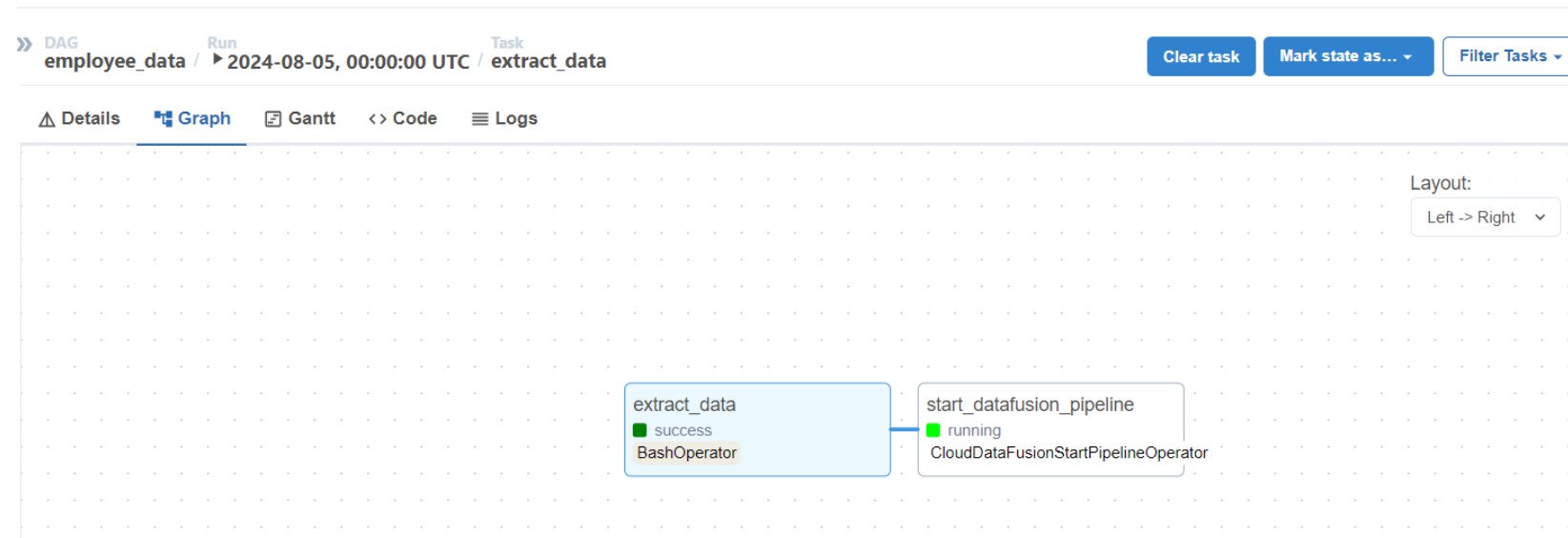
Employee_data

▶ RUN · 💾 SAVE QUERY ▾ · ⬇ DOWNLOAD · 👤 SHARE ▾ · 🕑 SCHEDULE · ⚙ MORE ▾ · ✅ This

```
1  SELECT Employee_ID,Full_Name,Job_Title,Department,Salary,Password FROM `weighty-bounty-431300-p2.employee.empo_dat` LIMIT
```

Query results

⬇ SAVE RESULTS ▾

JOB INFORMATION · **RESULTS** · CHART · JSON · EXECUTION DETAILS · EXECUTION GRAPH

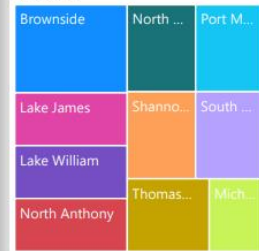| Row	Employee_ID	Full_Name	Job_Title	Department	Salary	Password
37	5140	Leslie,Baker	IT Specialist	HR	xxxxx	Ciyb8akHxYy4aNaUJ/r0+w==
38	9820	Catherine,Mcneil	IT Specialist	HR	xxxxx	gv6M7CwHVICy85HgBbdmoA==
39	5695	Jonathan,Phillips	IT Specialist	HR	xxxxx	gnk16a6pMHuwMQ7YFva1Jw==
40	6489	Sabrina,Collins	IT Specialist	HR	xxxxx	cF42UaqV51l18tPnWf0vHQ==
41	4056	Ashley,Watson	IT Specialist	HR	xxxxx	yGJWObEy45BoJt7feNESlQ==
42	2647	Daniel,Rodgers	IT Specialist	HR	xxxxx	wFnaY4L/IcweCadajwwxTA==
43	4676	Ethan,Lopez	Data Scientist	HR	xxxxx	h0T94OkK1mriAvmcJlFkGw==
44	9289	Jason,Snow	Data Scientist	HR	xxxxx	OvibaplZ+KUVqrFyjp0RUQ==
45	6925	Anthony,Williams	Data Scientist	HR	xxxxx	EB0LCynfsuTZUkNkWpBzVg==
46	7739	Rachel,Butler	Data Scientist	HR	xxxxx	9tr9WUgPSiOPgQlh5UbjIw==
47	6244	Amy,Marks	Data Scientist	HR	xxxxx	Qs1TraEVtKxloTi0YIPxAg==
48	7421	Deborah,Rios	Data Scientist	HR	xxxxx	5f0k8WGr0I2pay+WP5tq4A==
49	4747	Linda,Hall	Data Scientist	HR	xxxxx	gfNkpA+is+aeSGVmSSazSA==
50	4523	Tammie,Zamora	Data Scientist	HR	xxxxx	z7Gi0Y2+EQnuhzXK5Kd0xQ==

Results per page: 50 ▾   1 – 5

Table in BigQuery

Power BI Dashboard

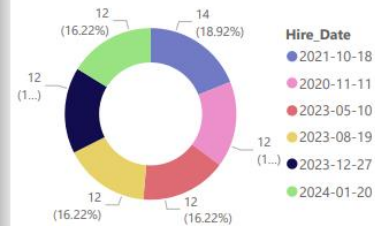# Thank you