

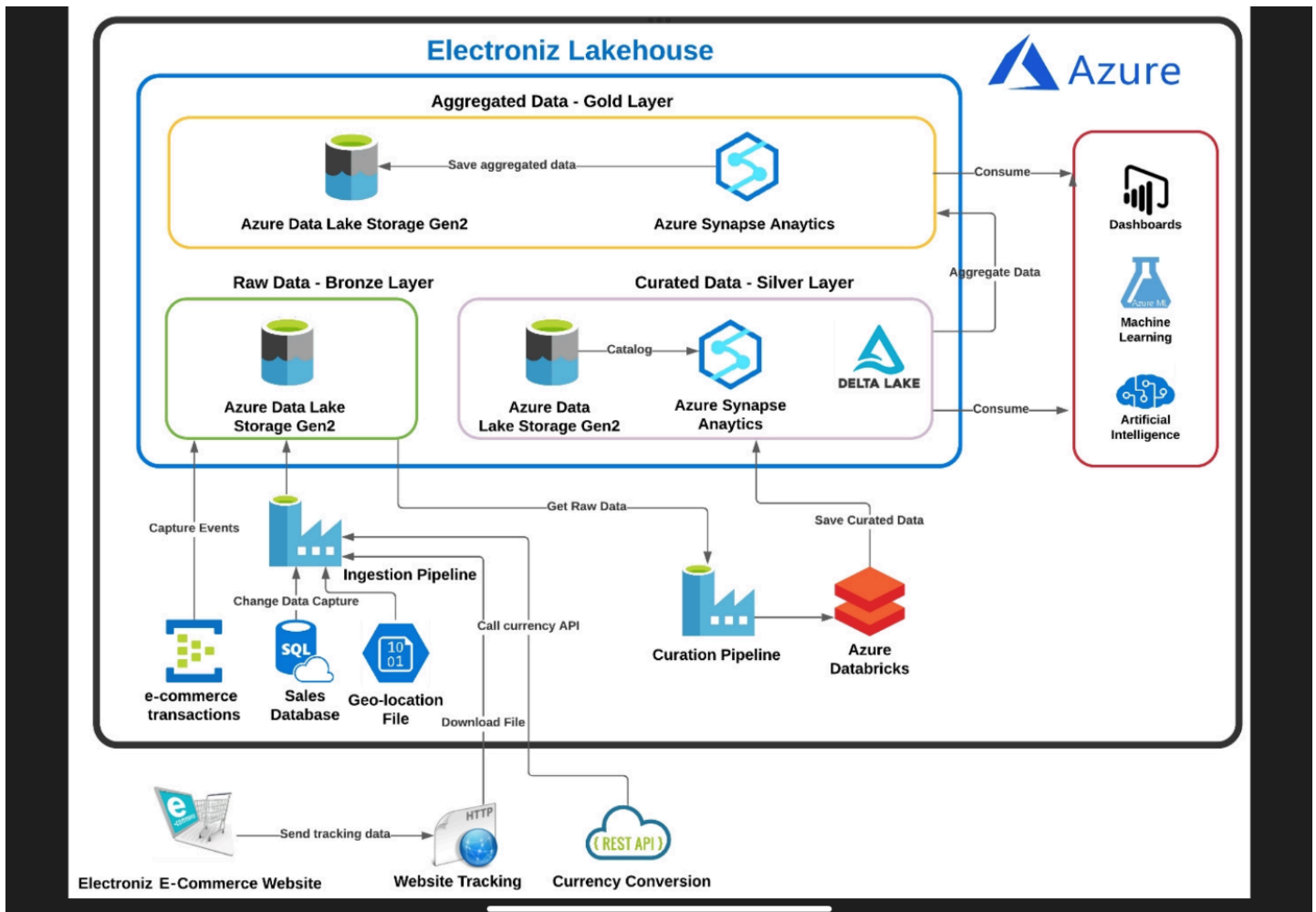
# Cloud Datawarehouse

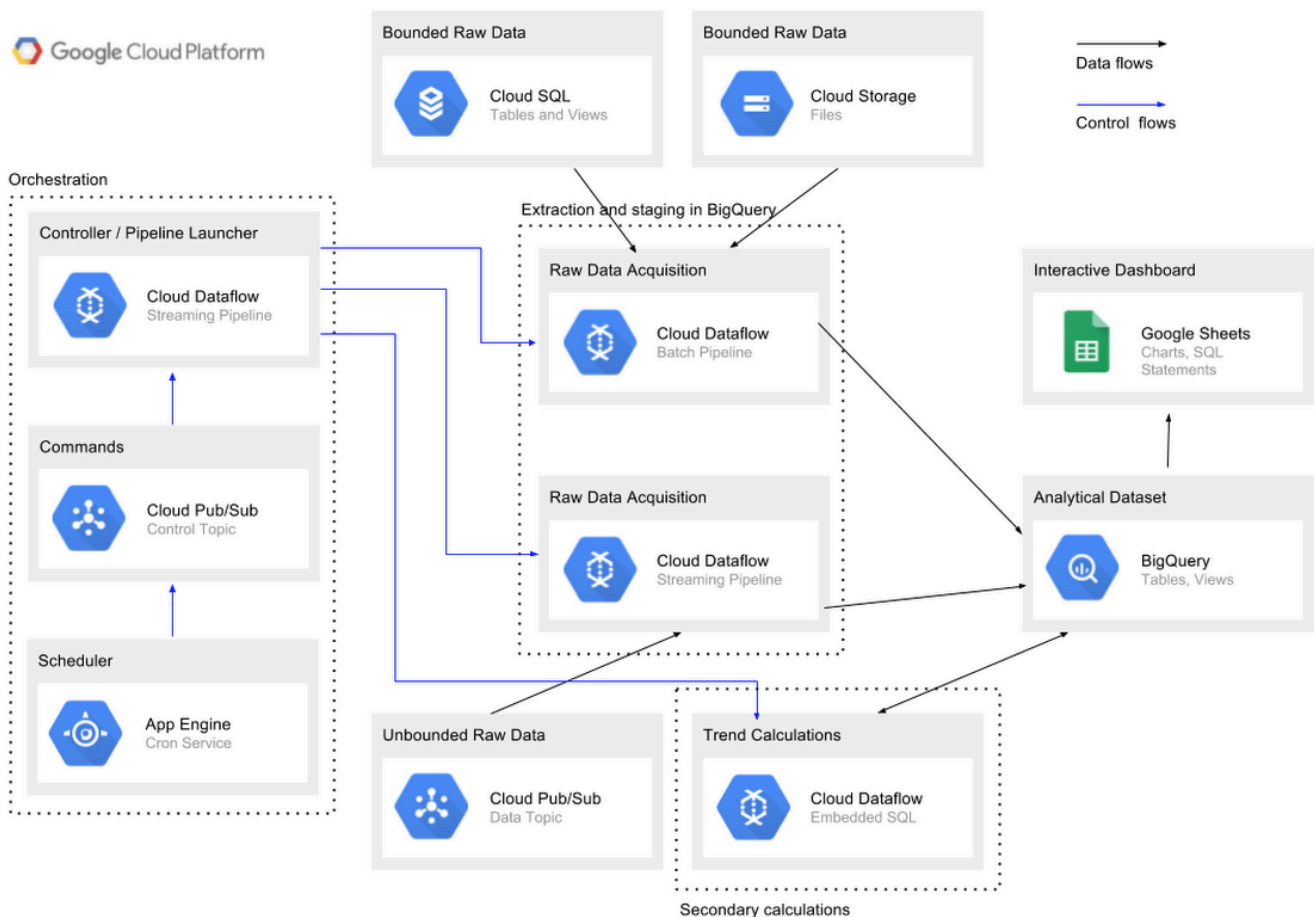
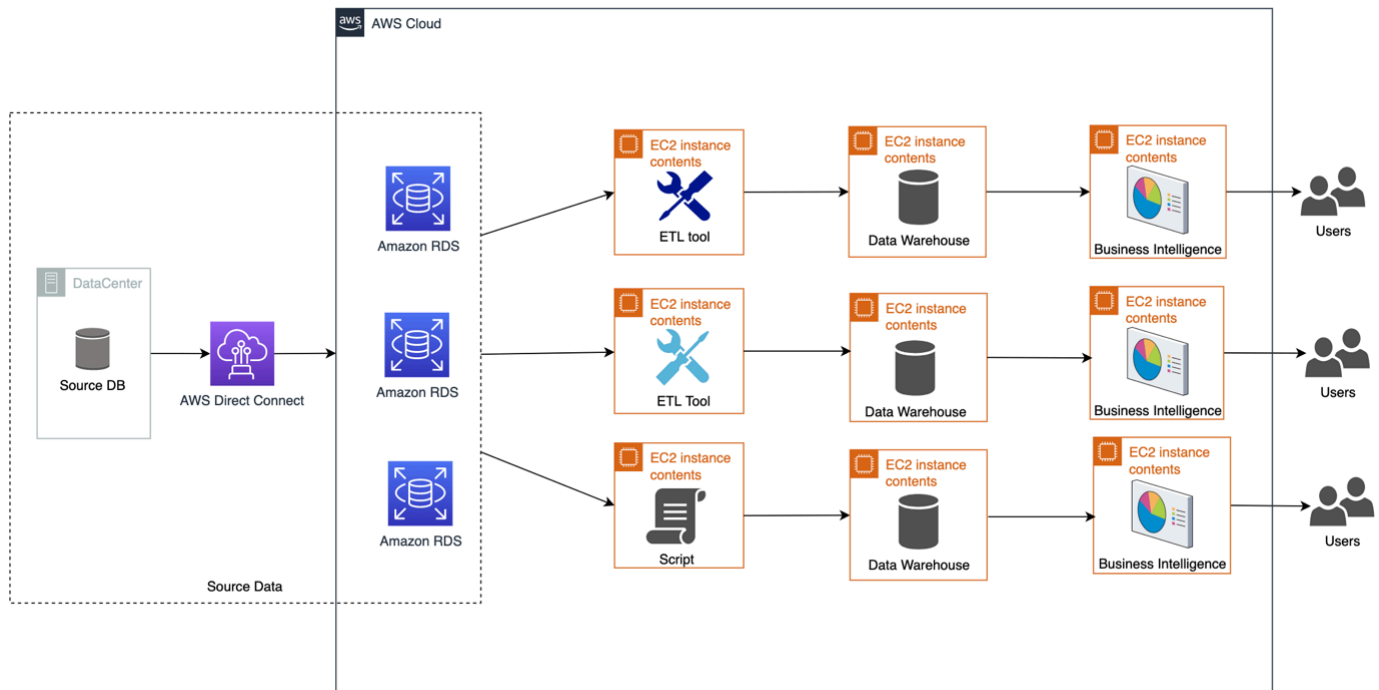
## Cloud Data Warehouse (Cloud DWH)

### 1. What is a Data Warehouse?

A **Data Warehouse (DWH)** is a **centralised repository** designed for **storing, integrating, and analysing large volumes of structured and semi-structured data** from multiple sources.

- Optimised for **analytical processing (OLAP)**, not transactional (OLTP).
- Helps in **reporting, BI (Business Intelligence), and decision-making**.





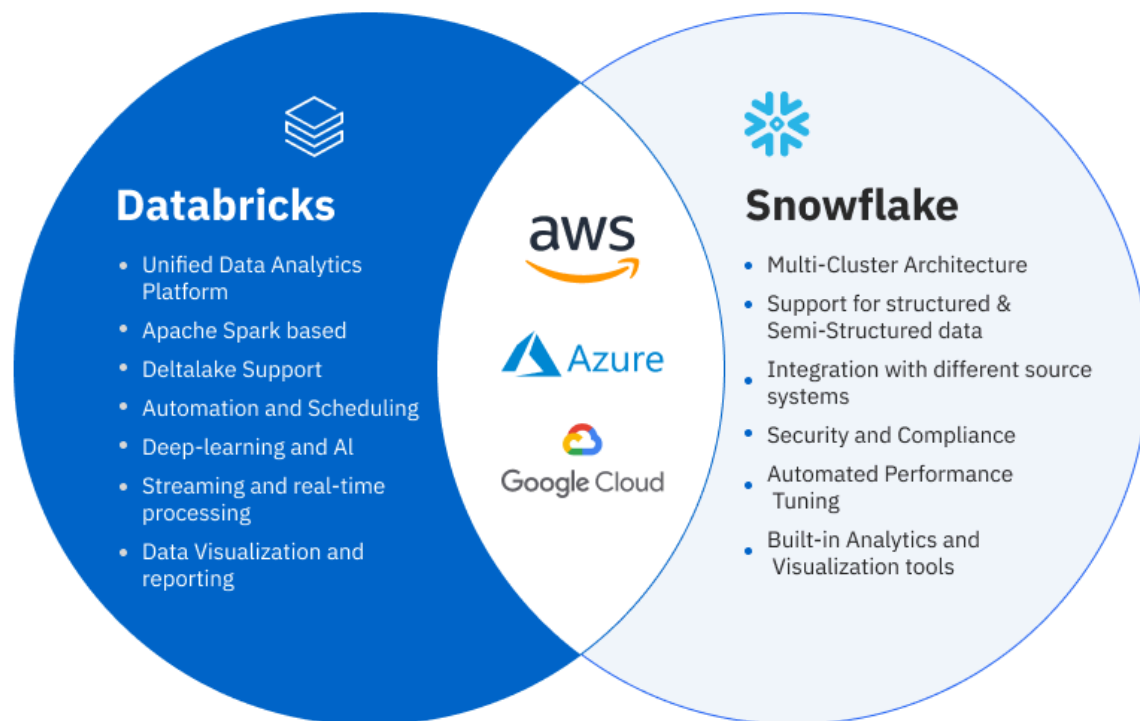
## 2. What is a Cloud Data Warehouse?

A **Cloud Data Warehouse** is a **data warehouse hosted in the cloud**.

👉 Instead of buying on-perm servers and storage, organisations use cloud providers to store and process analytical data.

Examples:

- **Amazon Redshift**
- **Google BigQuery**
- **Snowflake**
- **Azure Synapse Analytics**



### f3. Key Characteristics

- **Elastic Scalability** – Scale up/down compute & storage independently.
- **Pay-as-you-go** – No upfront CapEx, only OpEx.
- **Separation of Storage & Compute** – Optimise cost/performance.
- **High Availability** – Built-in redundancy across regions.
- **Multi-format Support** – Handles structured + semi-structured data (JSON, Avro, Parquet).
- **Serverless / Managed** – Little to no infrastructure management.

### 4. Cloud Data Warehouse Architecture

Typical components:

1. **Data Sources** – ERP, CRM, IoT, flat files, APIs, social media, transactional DBs.

2. **ETL/ELT Layer** – Data pipelines (Extract, Transform, Load) with tools like AWS Glue, Apache Spark, Talend.
3. **Cloud DWH Storage & Compute** – Redshift, BigQuery, Snowflake, Synapse.
4. **Analytics & BI Layer** – Tableau, Power BI, Looker, AWS QuickSight.
5. **Users** – Analysts, Data Scientists, Business Teams.

## 5. Benefits of Cloud Data Warehouse

- **Faster Analytics** – Query petabytes in seconds.
- **Cost Efficiency** – Pay per query or per resource usage.
- **Performance** – Distributed, parallel query execution.
- **Global Accessibility** – Access anywhere via internet.
- **Security & Compliance** – IAM, encryption, GDPR, HIPAA support.
- **Automatic Scaling & Maintenance** – No DBA headaches.

## 6. Challenges / Risks

- **Data Security & Privacy** (sensitive info in the cloud).
- **Vendor Lock-in** (switching providers can be costly).
- **Cost Management** (unexpected bills if queries/storage not optimized).
- **Data Latency** (network dependency).

## 7. Cloud Data Warehouse vs Traditional (On-Prem) DWH

Feature	On-Premises DWH	Cloud DWH
<b>Cost Model</b>	CapEx (hardware, setup)	OpEx (pay-as-you-go)
<b>Scalability</b>	Hardware-limited	Elastic, on-demand
<b>Deployment Time</b>	Weeks/Months	Minutes/Hours
<b>Maintenance</b>	DBA responsibility	Provider-managed
<b>Performance</b>	Limited parallelism	Massively parallel processing (MPP)
<b>Accessibility</b>	Local/limited VPN	Global internet access

## 8. Popular Cloud Data Warehouses

- **Amazon Redshift** – Scalable, part of AWS ecosystem.
- **Google BigQuery** – Serverless, pay-per-query.
- **Snowflake** – Multi-cloud (AWS, Azure, GCP), flexible compute-storage separation.
- **Azure Synapse Analytics** – Deep integration with Microsoft stack.

**In short:**

A **Cloud Data Warehouse** is a **scalable, cost-efficient, and managed solution** for storing and analyzing large datasets in the cloud—powering **BI, analytics, and data-driven decision-making**.