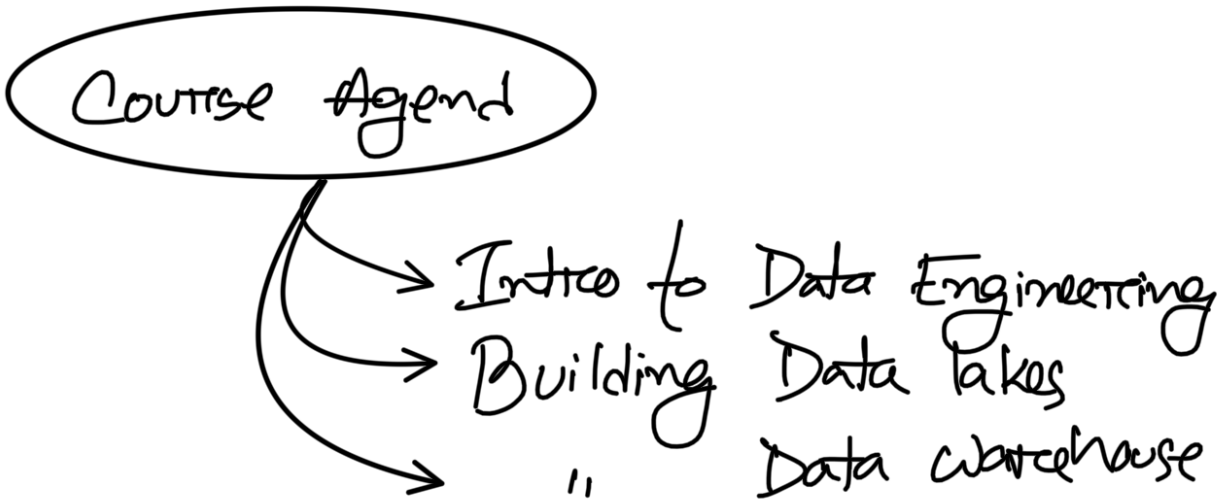
# Data lakes & Data Warehouse

1. Modernizing Data lakes & Data warehouse with Google Cloud
2. Building Batch Data pipelines on google cloud.
3. Building resilient streaming analytics system on GC
4. Smart Analytics, ML & AI on GC.




## Role of Data Engineer

01. Role
02. Challenges
03. Intro to Bigquery
04. Data lakes & Data warehouse

05
06

# Transactional Databases Vs Data Warehouse

Data Engineer → builds → data pipelines → data driven decision

raw data → not useful → data lakes

Consideration
- all types of data
- scale to meet demand
- support high throughput ingestion
- fine grained acces
- other tool connect easily

# Backup
Replace/decommission
Analytics & ML

ETL → Extract, Transform, Load

Data process
↓
Dataproc
Dataflow

Streaming

Data process

pub/sub  Dataflow  Bigquery

---

Challenges

Access of Data

Data accuracy & quality

Availability of computational resources

Query performance

Silod Operational Systems

Refresh process

Ad hoc & Reporting Queries


Web

⇨ Data accuracy & Quality

# raw data

↳ cleaned → transformed

↓

Data Warehouse — Stored

↓

queries

↓

structured

⇒ Computational resources

⇒ Query performance → Query Engine

↓

Software

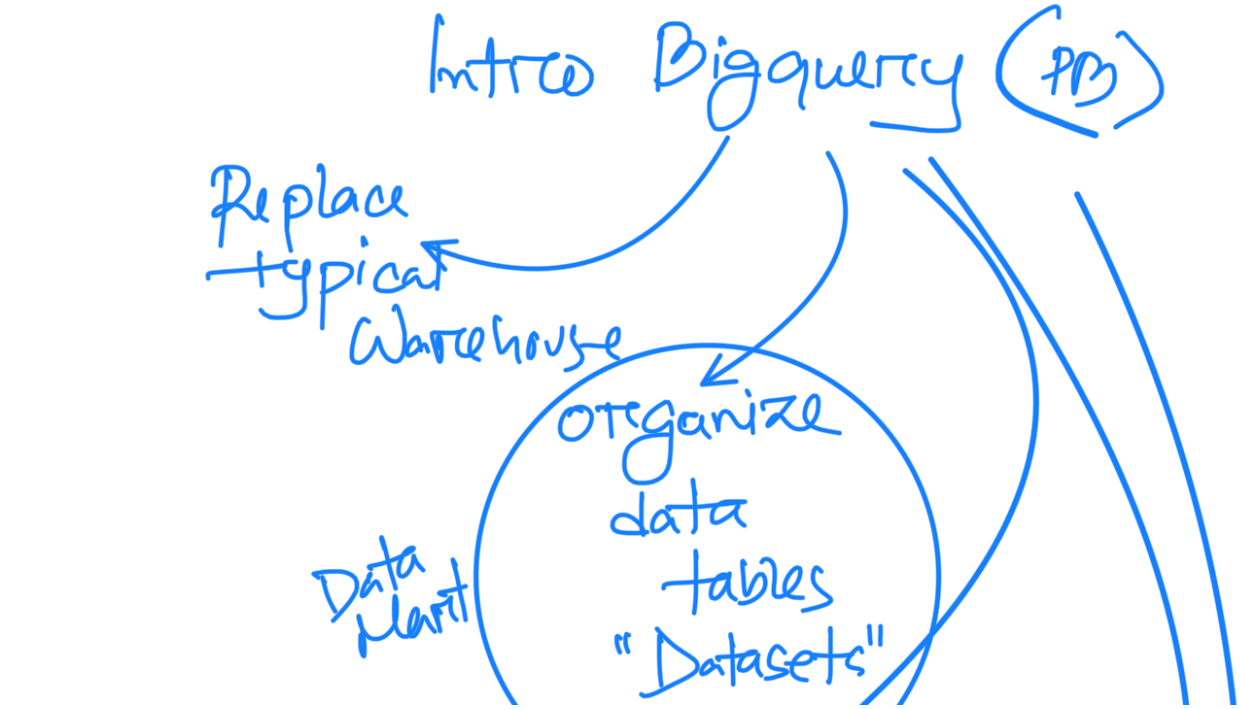
↓

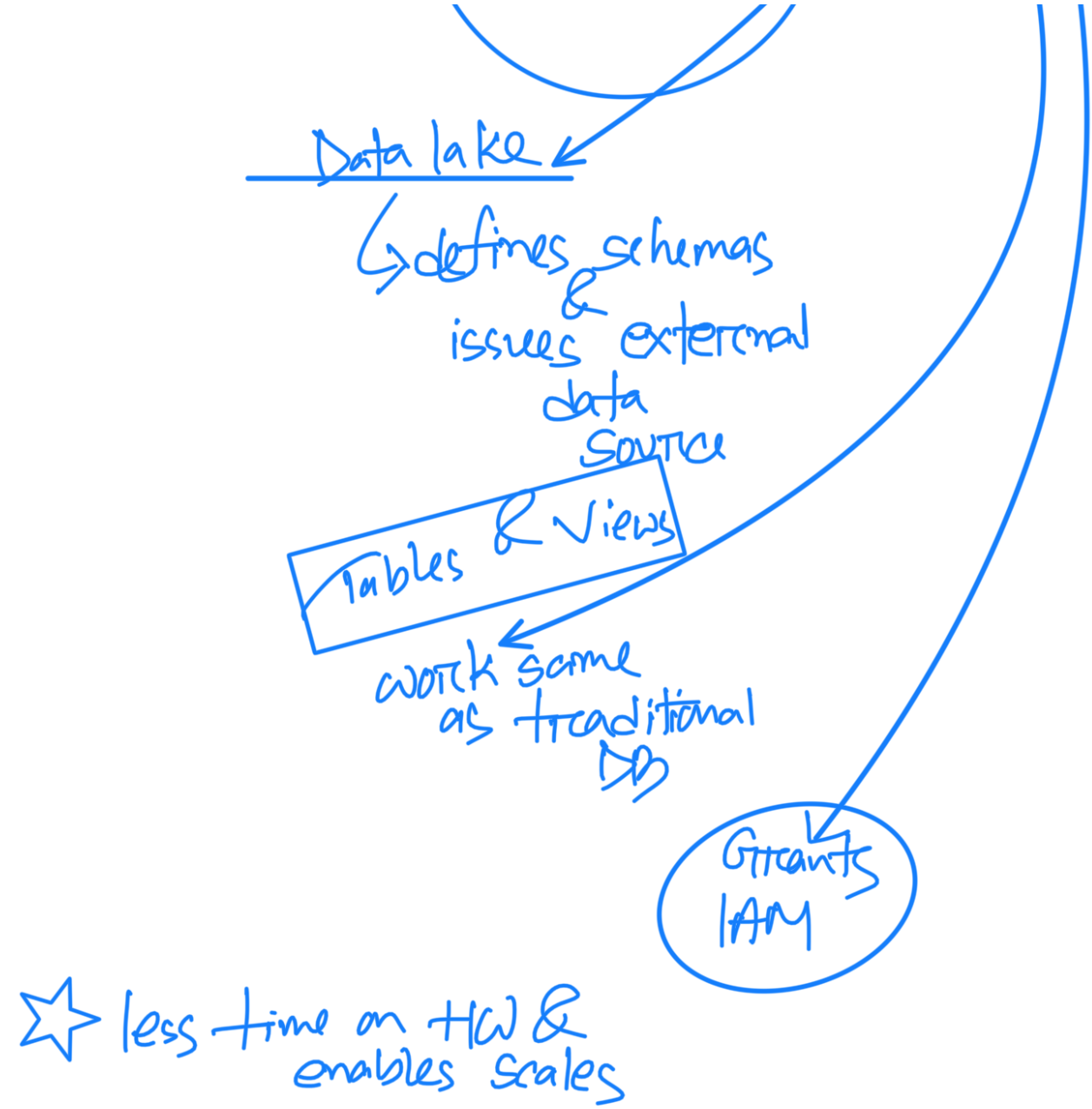
Managing cluster

↓

optimize

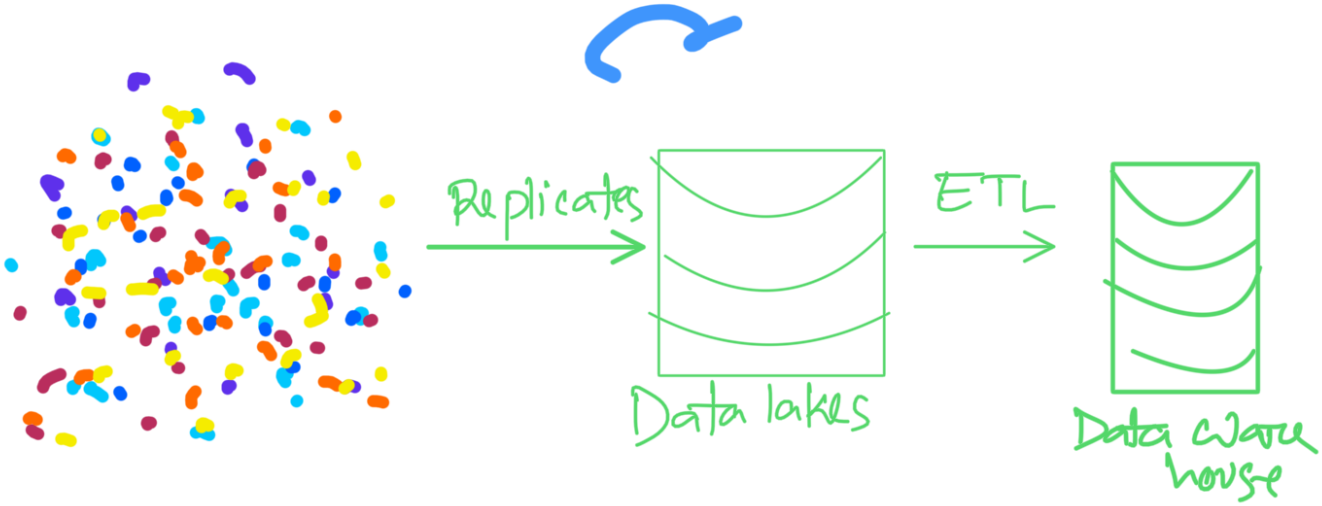
---






☆ less time on HW & enables scales

---



⇨ Serve as a sink for both Batch & Stream data pipelines

⇨ Warehouse scale to meet needs

☆ organized, categorized & access control

☆ designed for performance

⇨ level of maintenance

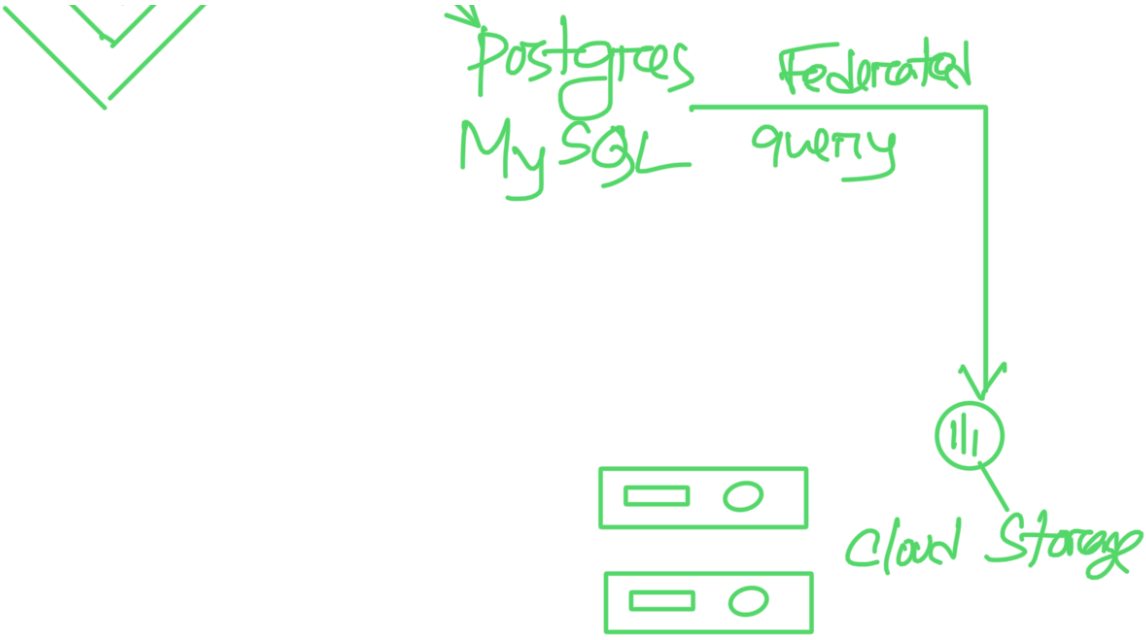
Traditional DW → Complex → only few user → optimized for BI

Bigquery → Automate → accessible → foundation AI

optimized for Batch → Continuous patching → Among of DB

Tee up real time → Fully manage → Simplify Data operation

Cloud SQL



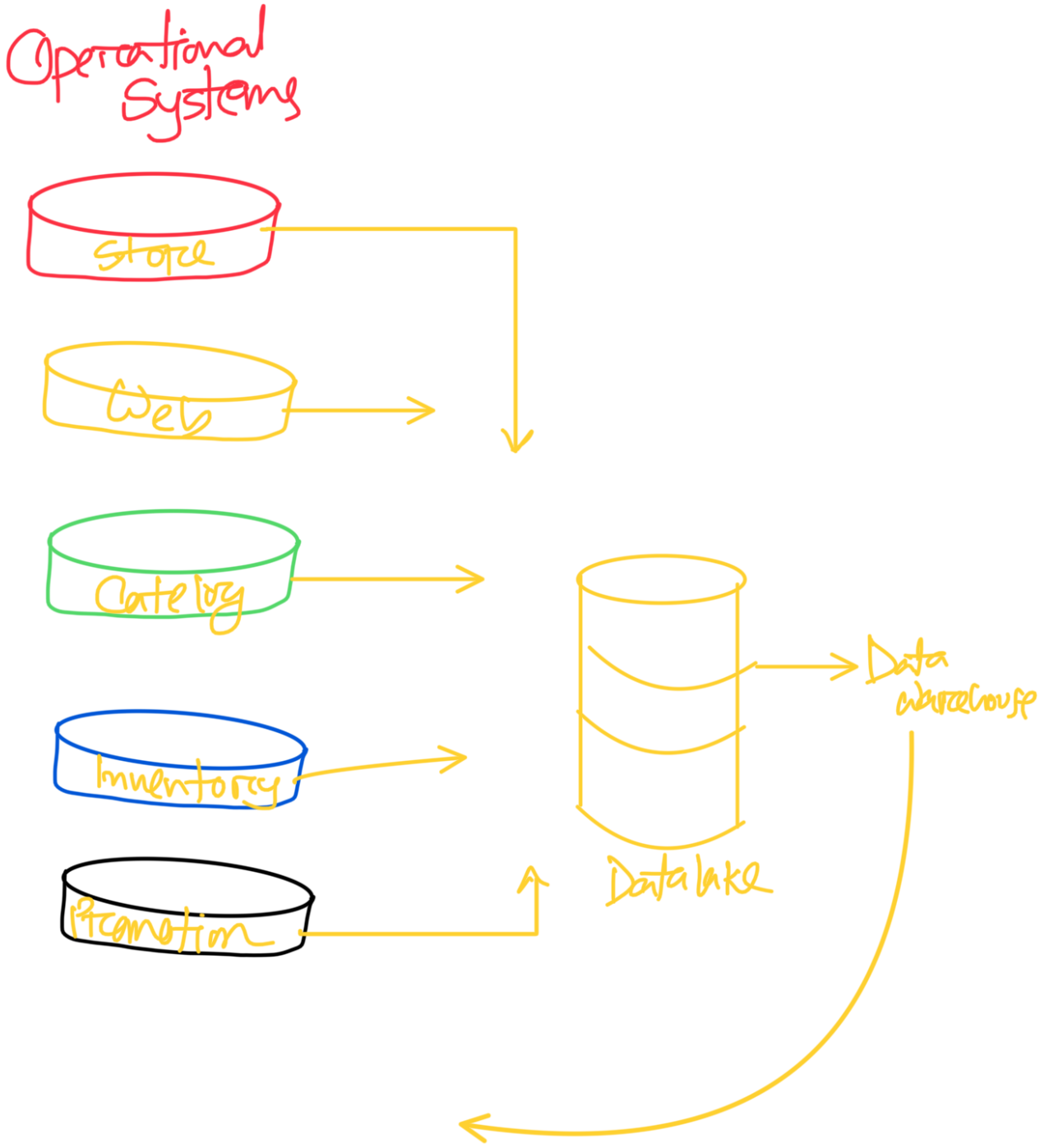

---

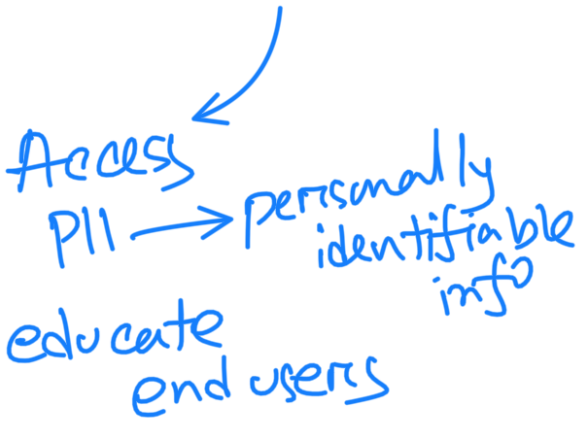
CloudSQL → manage → SQL servers
Postgres
MySQL
RDBMS

CloudSQL
- GB/TB
- Backend DB apps
- Record Based Storage

Bigquery
- PB
- connect external Data source
- Column Based Storage

RDBMS → optimized → Writes to Records


Operational Systems
Store
Web
Catelog
Inventory
Promotion
Data lake
Data warehouse




Simplify data discovery at any scale
Unified view of all Datasets
Data governance foundation.