

Data Lake (GCS)

Data Fellowship : Data Engineer
Muhammad Hanif Fajari

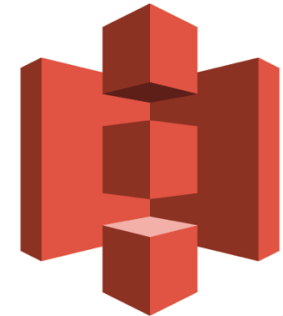
Data Lake

File system untuk menyimpan **data berstruktur** dan **tidak berstruktur**. Kenapa dibuat karena lebih **murah, simple dan fleksible** dibanding **data warehouse**.

Contoh service : Google Cloud, S3 Data Lake (AWS)



Google Cloud



Amazon S3

Data Warehouse

Database untuk menyimpan **data berstruktur** yang sudah siap untuk dilakukan **analisis**.

Contoh service : Google Big Query, AWS Athena



Google
BigQuery



amazon
ATHENA

Tipe Data Warehouse

ETL (*Extract Load Transform*)

Konsep data warehouse yang melakukan **transformasi** sebelum dimasukan ke data warehouse. Sehingga data menjadi **lebih aman** karena sudah dilakukan masking/transkripsi akan tetapi memiliki *cost yang lebih mahal*

ELT (*Extract Transform Load*)

Konsep data warehouse terbaru dengan **langsung load diawal** tanpa melakukan transformasi. Sehingga bisa **menghemat budget** akan tetapi *kurang aman* karena data tidak terenkripsi atau masking

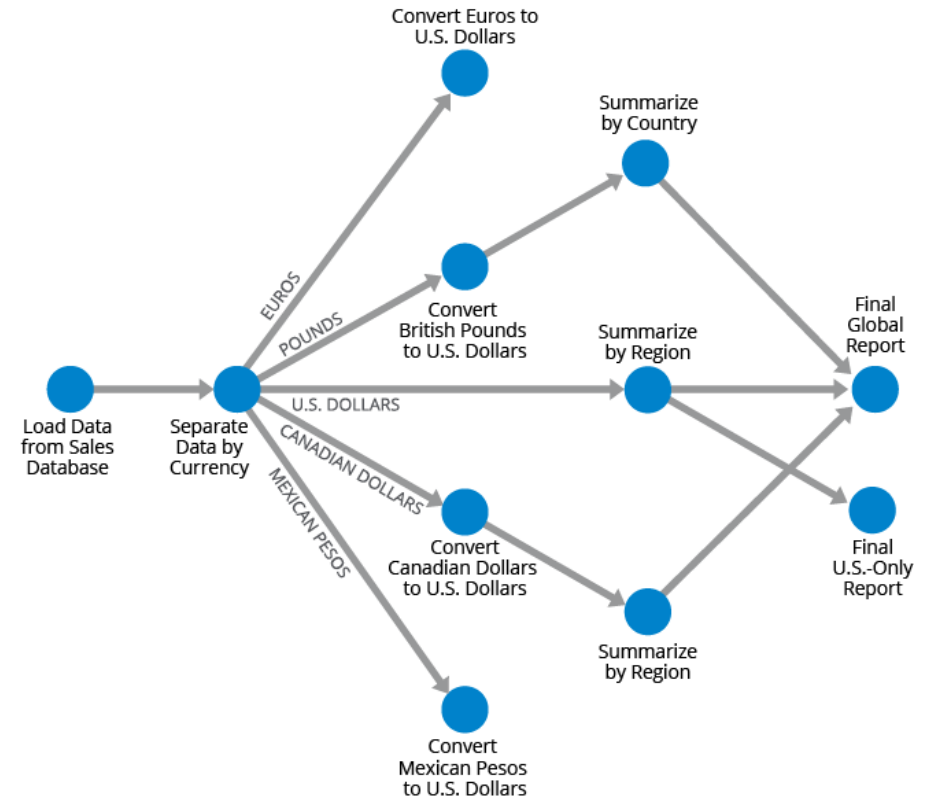
Orchestration Pipeline

Suatu proses untuk mengatur dan menjadwalkan pekerjaan

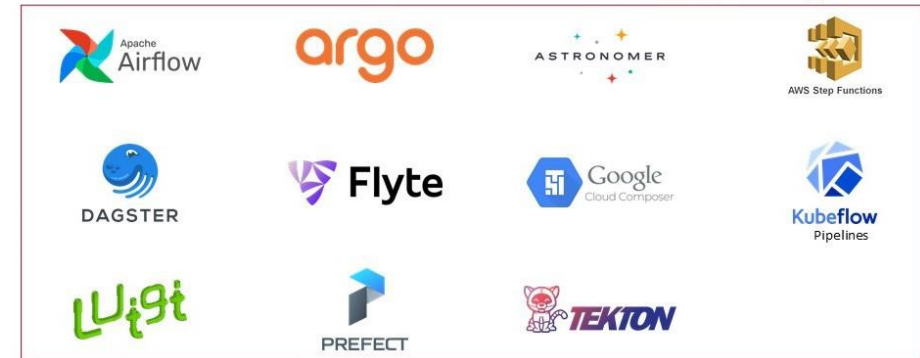
Directed Acyclic Graph (DAG)

Visualisasi dari urutan aktivitas. DAG dapat merepresentasikan banyak tipe dari suatu aliran, termasuk aliran *data processing*

Banyak dari *data processing environments*, melakukan komputasi untuk **satu atau lebih tujuan**



Data Orchestration Landscape

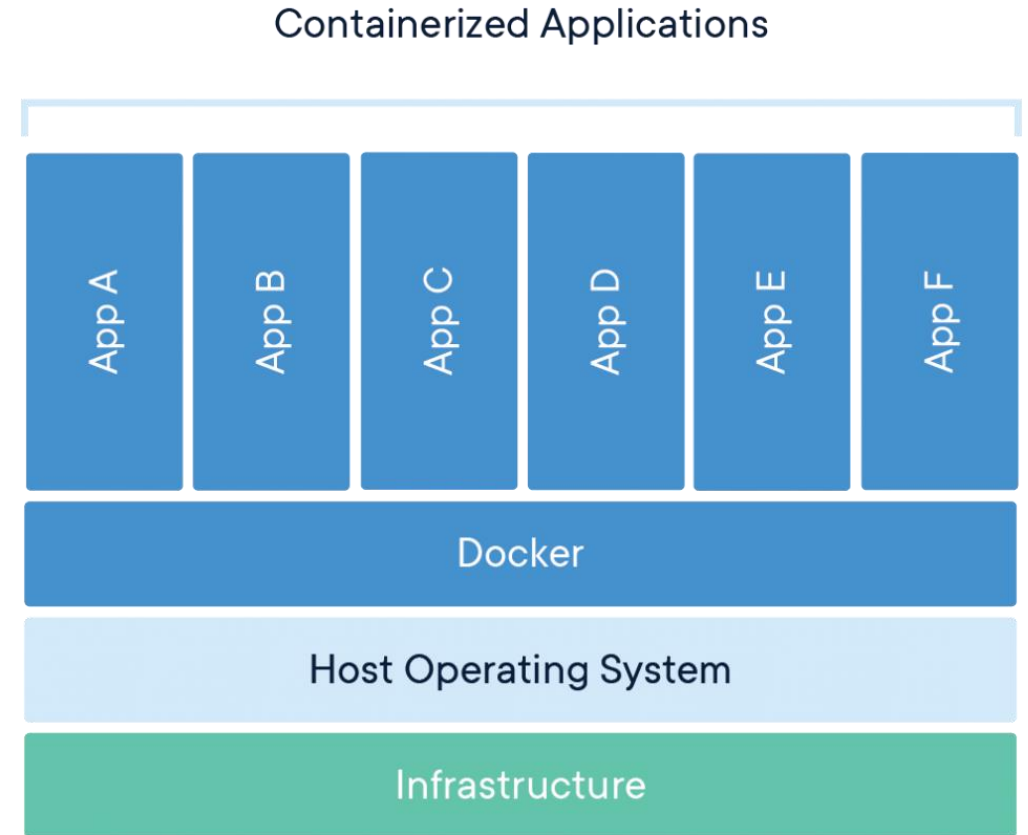


Docker

Open platform yang dapat digunakan untuk *developing, shipping* dan menjalankan aplikasi

Docker dapat **memisahkan aplikasi** dari *infrastruktur* sehingga dapat berjalan **lebih cepat**

Docker compose dapat membantu membagikan multi-container aplikasi



THANK YOU



[hanifajari](#)



[mhaniffajari](#)