

# Virtual Internship Experience

**Data Pipeline**



***kimia farma***

# Disclaimer

“Dokumen ini memiliki hak cipta. Barang siapa yang menyebarluaskan atau menduplikasi tanpa izin dari instansi terkait dapat diproses sesuai dengan ketentuan hukum yang berlaku.”

# Outline

- 1 Data Pipeline
- 2 Data Processing
- 3 Data Pipeline Framework
- 4 Implementasi Data Pipeline di Kimia Farma



# Data Pipeline

## Definisi

Data pipeline adalah cara untuk memindahkan data dari satu tempat (sumber database tertentu) ke tujuan (seperti data lake). Kemudian, data tersebut dapat diubah dan dioptimalkan, sehingga dapat dianalisis dan digunakan untuk keperluan analisis.

## Keuntungan

- Memungkinkan untuk melakukan *real-time analysis*
- Menyimpan data di cloud
- Menyatukan data dari berbagai sumber dalam satu tempat





# Elemen Data Pipeline

## 1. Sources

Sumber data yang akan dialirkan. Misalnya, dari ERP system, mobile apps, websites, web apps, microservices, IoT devices, dst.

## 2. Proses

Proses dalam mengalirkan data. Akan kita bahas di slide selanjutnya

## 3. Tujuan

Dari proses yang dilakukan, datanya akan disimpan dimana? Biasanya akan masuk ke staging, kemudian disimpan di data lake sampai data warehouse.

# Data Processing

## (1) Collection

1. Data diambil dari sumber database yang tersedia datanya
2. Tentukan table apa saja yang akan diambil / ditarik datanya

## (2) Ingestion

1. Perintah untuk menarik data dari sumber ke tujuan
2. Bisa dilakukan dalam 2 cara: batching dan streaming
  - Batching: terjadwal pada waktu tertentu, contoh: 1 kali sehari di jam 02:30 AM
  - Streaming: setiap ada perubahan atau event, datanya akan dialirkan
3. Tujuan: dimasukkan ke dalam Data Lake

## (3) Preparation

1. Operasi extract, transform, load (ETL), yaitu membersihkan, membentuk, dan menggabungkan data
2. Tujuan: dimasukkan ke dalam data warehouse



# Data Processing

## (4) Computation

1. Tim Data Analyst, Data Scientist, Machine Learning mengolah data pada proses ini
2. Output dari proses ini bisa disimpan di Data Mart

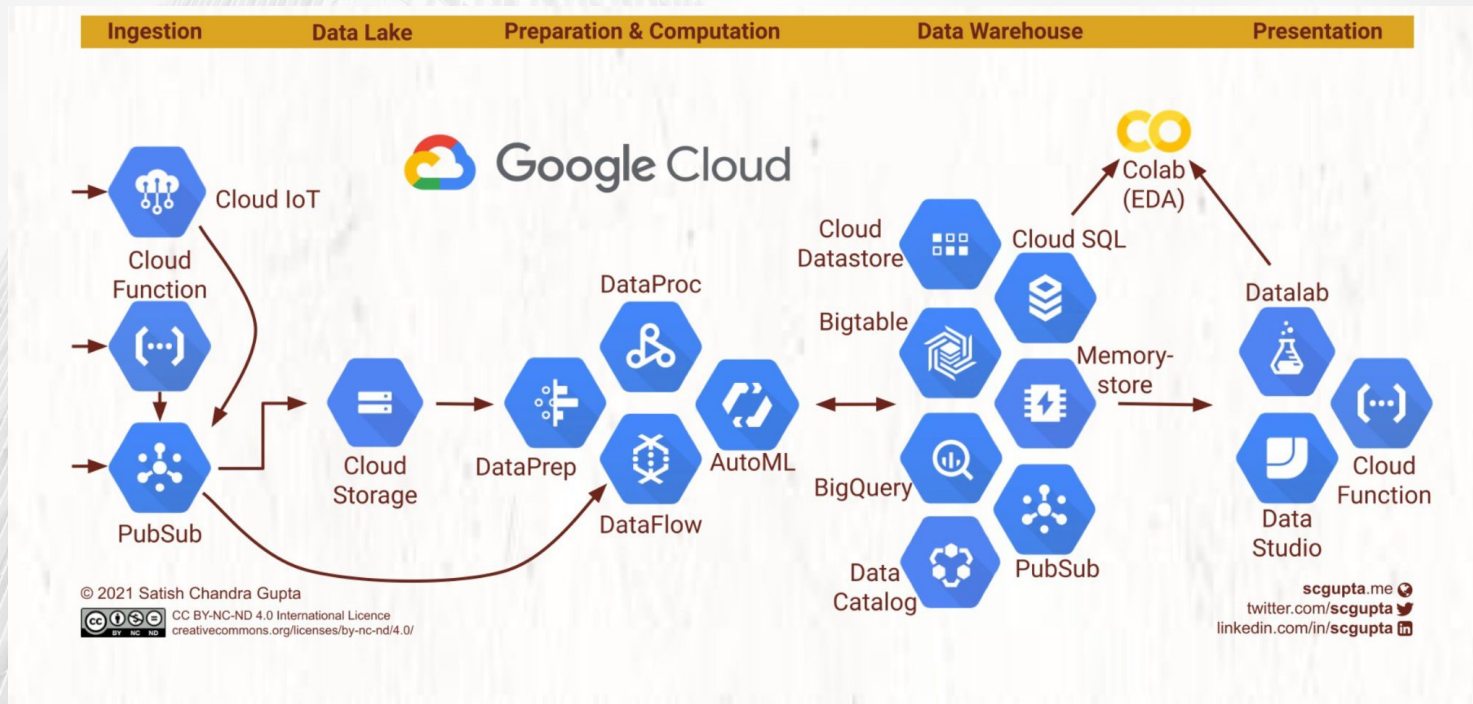
## (5) Presentation

1. Menampilkan hasil computation di dashboard, email, push notifications, report, dst.



# Data Pipeline Framework

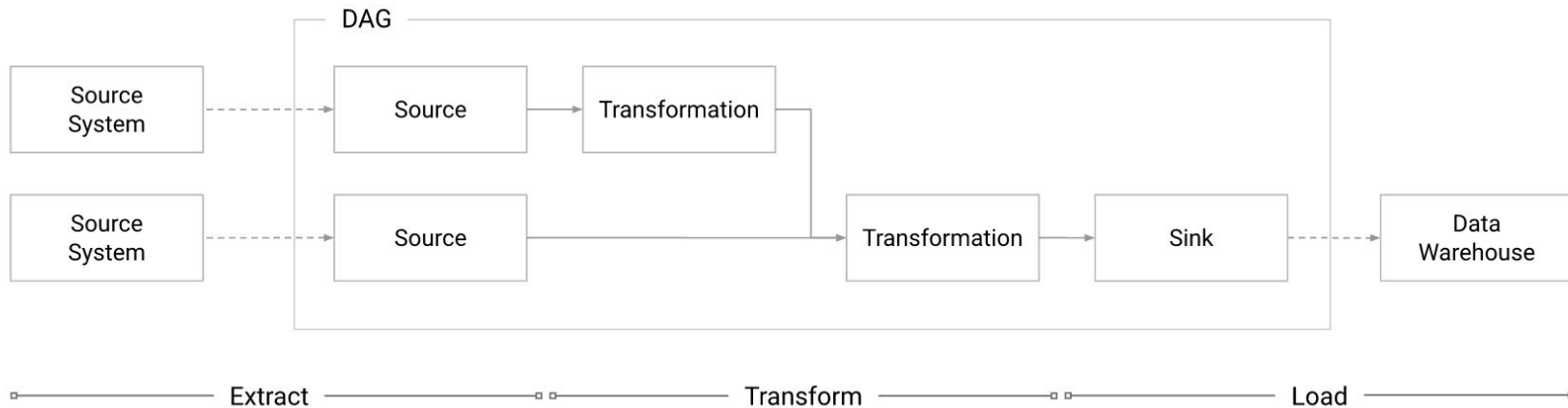
Contoh: Google Cloud Platform - Data Pipeline Framework





# Konsep Pengembangan Data Pipeline

Extract, Transform, Load (ETL)





*kimia farma*

# **Implementasi Data Pipeline di Kimia Farma**



# Thank You!

