

E-commerce analytical and monitoring platform

Malik Fajar Hidayatullah





Malik Fajar Hidayatullah

Education

*S1 Teknik Informatika Binus University
(2020–2024)*

Working

Dibimbing Data engineering students

Overview Project

E-commerce analytical and monitoring platform

Platform data terpusat secara batch dan stream dan visualisasi analisis dengan metabase


On the left side of the slide, there are three overlapping geometric shapes: a large black parallelogram at the top, a medium-sized light orange parallelogram in the middle, and a smaller dark orange parallelogram at the bottom. All shapes are slanted to the right.

Project Background

Project Background

Platform ini dirancang untuk mengintegrasikan kumpulan data format CSV dengan menggabungkan pemrosesan batch dan streaming untuk menciptakan satu single source of truth.

Tujuan platform ini menyediakan insight tentang transaksi pelanggan, performa produk, pertumbuhan bisnis, serta aktivitas pelanggan agar mendukung analisis data yang lebih akurat dan pengambilan keputusan yang lebih strategis.

A large black parallelogram is positioned on the left side of the slide. Overlapping its bottom edge are two orange parallelograms, one in a lighter shade and one in a darker shade, both pointing towards the right.

Problem Statement

Problem Statement

Masalah

- Volume data besar dalam format csv
- Tidak adanya sumber data terpusat yang terintegrasi
- Tantangan dalam menganalisis transaksi dan aktivitas pelanggan

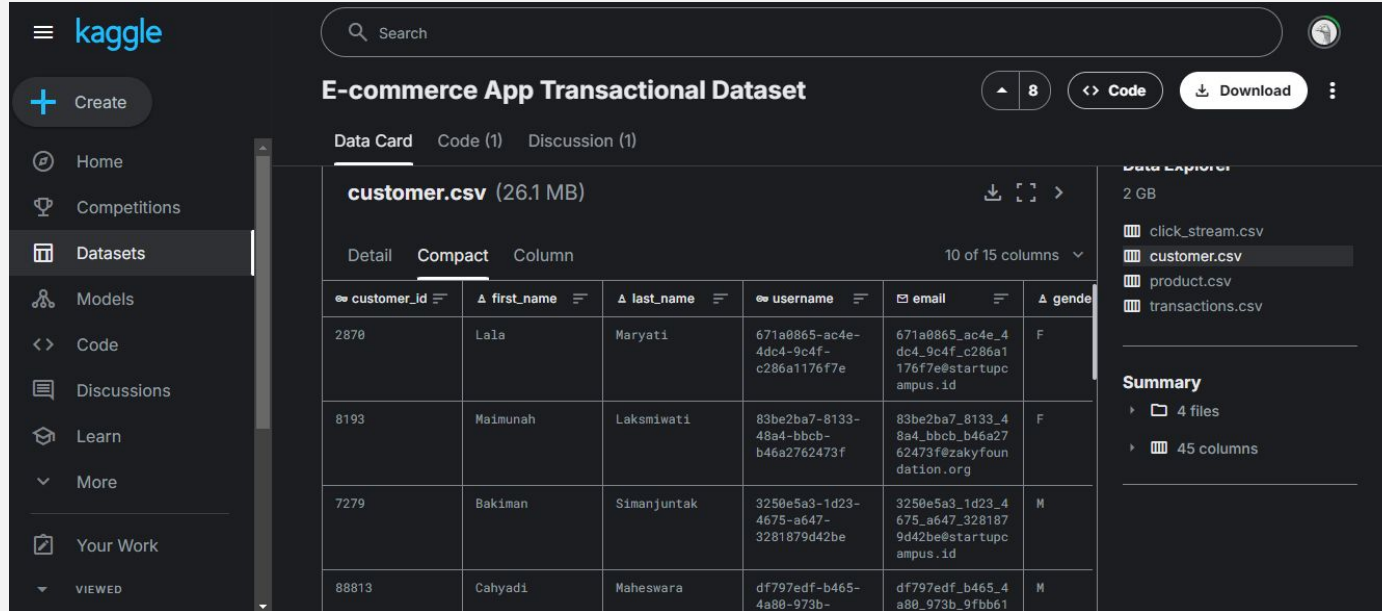
Tujuan

- Meningkatkan visibilitas pola transaksi dan perilaku pelanggan dengan dashboard
- Mengidentifikasi pertumbuhan bisnis berdasarkan performa periode sebelumnya
- Membantu pengambilan keputusan bisnis berdasarkan data yang real

A large black parallelogram is positioned on the left side of the slide. Below it, two overlapping parallelograms in shades of orange and yellow are also positioned on the left, creating a layered, geometric effect.

Data Understanding

Data Understanding



The screenshot shows the Kaggle interface for the 'E-commerce App Transactional Dataset'. The 'customer.csv' file (26.1 MB) is selected. The interface includes a sidebar with navigation options like Home, Competitions, Datasets, Models, Code, Discussions, Learn, More, and Your Work. The main area displays the file details and a preview of the data in a table format.

customer_id	first_name	last_name	username	email	gender
2870	Lala	Maryati	671a0865-ac4e-4dc4-9c4f-c286a1176f7e	671a0865-ac4e-4dc4-9c4f-c286a1176f7e@startupcampus.id	F
8193	Maimunah	Laksmiwati	83be2ba7-8133-48a4-bbcb-b46a2762473f	83be2ba7_8133_48a4_bbcb_b46a2762473f@zakyfoundation.org	F
7279	Bakiman	Simanjuntak	3250e5a3-1d23-4675-a647-3281879d42be	3250e5a3_1d23_4675_a647_3281879d42be@startupcampus.id	M
88813	Cahyadi	Maheswara	df797edf-b465-4a80-973b-	df797edf_b465_4a80_973b_9fbb61	M

Data diperoleh dari kaggle berbentuk csv, file tersebut yang akan diproses nantinya secara batch dan streaming

<https://www.kaggle.com/datasets/bytadit/transactional-ecommerce?select=customer.csv>

Data Understanding

customer.csv

- Identitas customer
- Tanggal sign up
- Alamat tinggal

product.csv

- Kategori
- Atribut
- Brand

transactions.csv

- Product
- Qty
- Payment
- Shipping

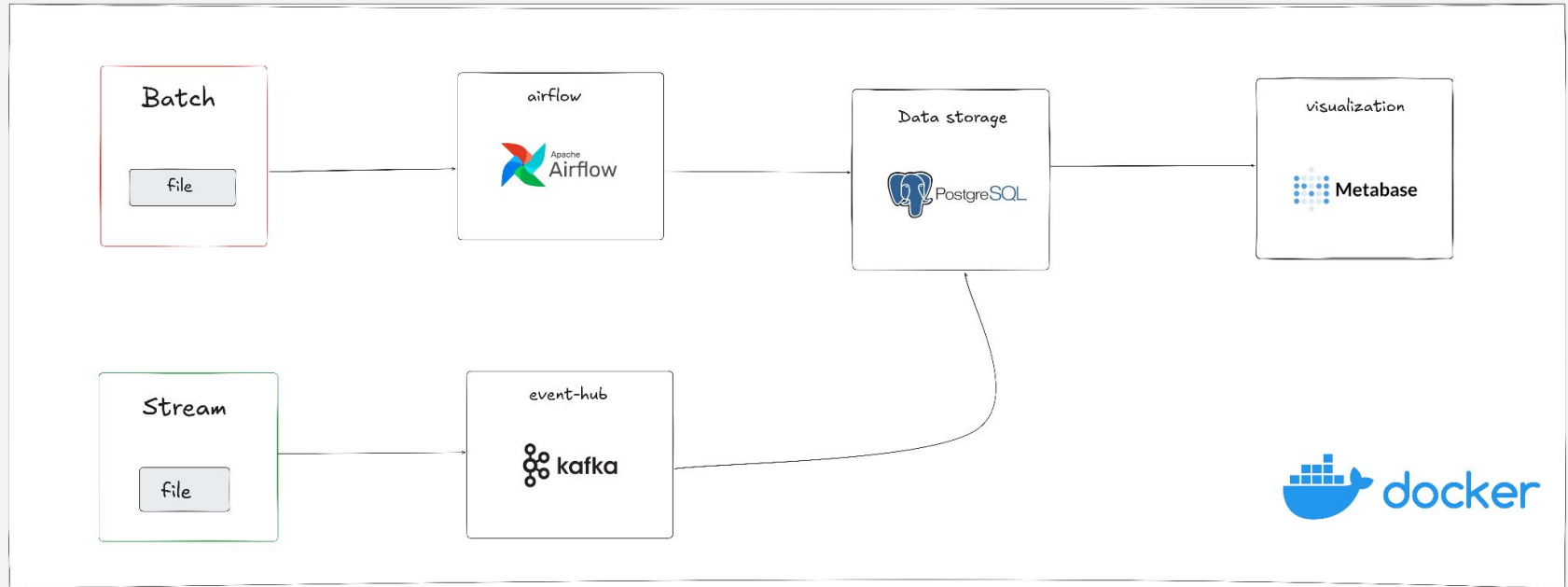
click_activity.csv

- Aktivitas customer
- Event: click, scroll, cart
- Sumber traffic: web, mobile

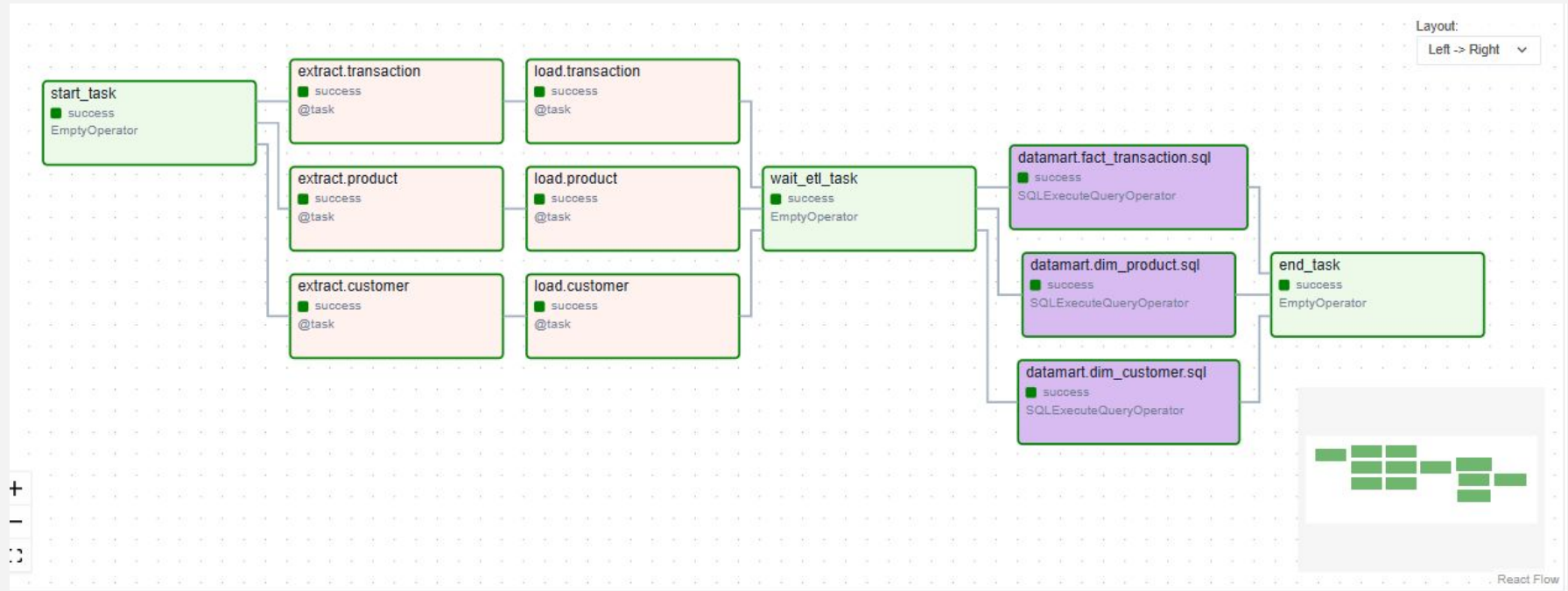
A large black parallelogram on the left side of the slide, with two overlapping orange parallelograms positioned below it, creating a layered effect.

Data Platform Understanding

Data Platform Understanding



Batch processing



Stream processing





Transformation & Consideration

Transformation & Consideration

1. Processing Method

- Batch processing secara daily dengan mode full dan incremental
- Streaming processing untuk data click_activity customer
- Chunked processing untuk file besar streaming

2. Tools Selection

- Pandas: Untuk transformasi data
- sqlalchemy: connect postgresql dan mengeksekusi query
- numpy: mengatasi struktur data

3. Key Transformations

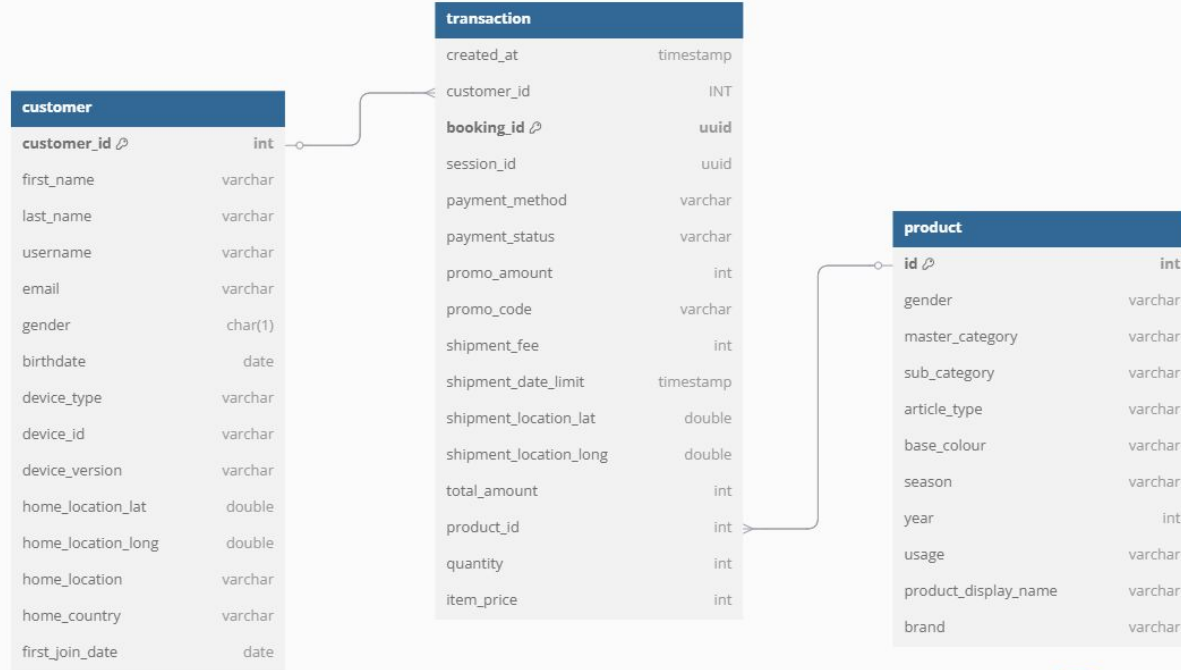
- Normalisasi zona waktu
- Data cleaning & standardization: ekstrak nama brand product
- Feature engineering: memecah data metadata dari json ke kolom baru

A large black parallelogram on the left side of the slide, with two overlapping orange parallelograms positioned below it, creating a modern, abstract design.

Data Modeling (Business)

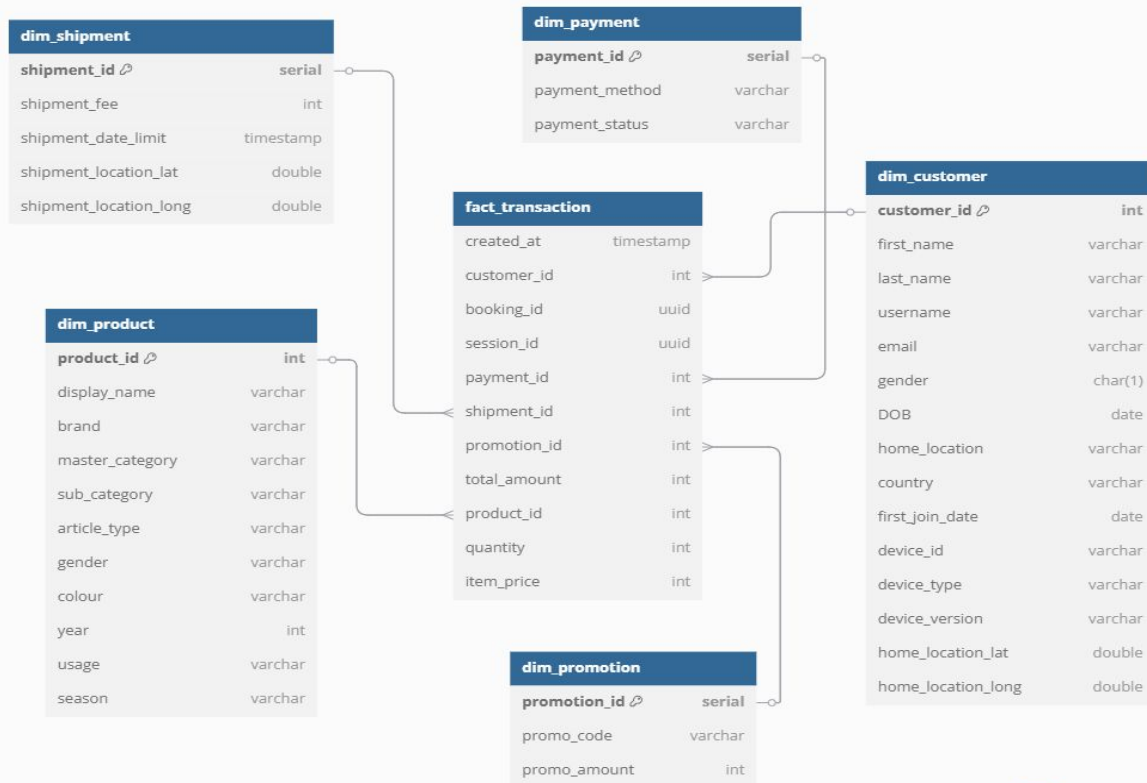
Data Modeling (Business)

Bronze



Data Modeling (Business)

Silver



Data Modeling (Business)



Stream

click_activity	
session_id	uuid
event_name	varchar
event_time	timestamp
event_id	uuid
traffic_source	varchar
product_id	int
quantity	int
item_price	int
payment_status	varchar
search_keywords	text
promo_code	varchar
promo_amount	int

Abstract geometric shapes in the bottom-left corner: a large black parallelogram, a medium yellow parallelogram, and a smaller orange parallelogram, all slanted to the right.

Visualization

Visualization

Overall business health

24,356

Transaction Q3 2022

↓ 71.06% • was 84,166 last quarter

Rp 493,933,841

Sales August 2022

↓ 98.83% • was Rp 42,176,442,295 last month

2,202

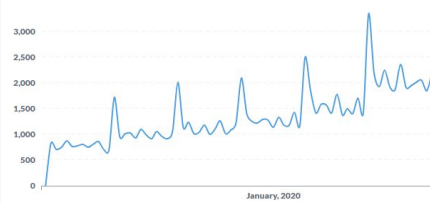
Customer acquisition August 2022

↑ 18.96% • was 1,851 last month

Transaction growth



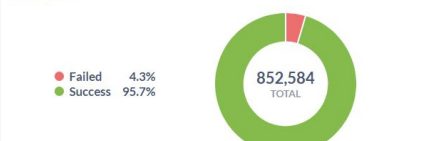
user acquisition



payment method by revenue and transaction

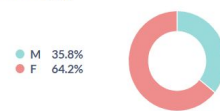


payment status



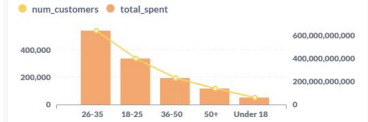
Customers profiles

gender demography

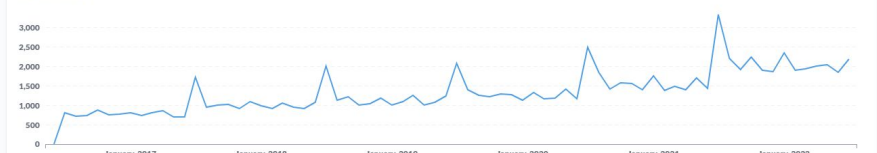


100,000
user

age demography by spending



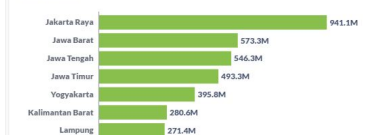
user acquisition



transaction by day



home location



Visualization

Product analysis

38.35%

promotion impact to sales

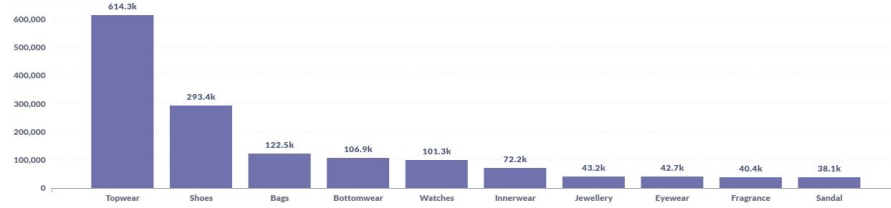
44,446

total products

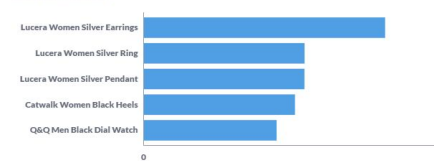
84.42

Average delivery time ⓘ

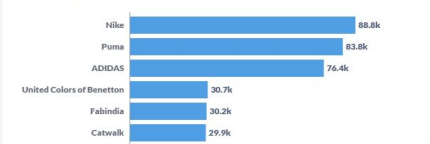
top selling category



most selling products



top 5 most selling brand



Monitoring the click activity

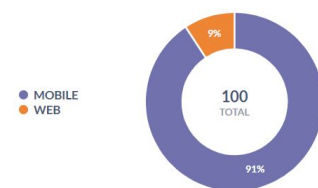
9

sessions

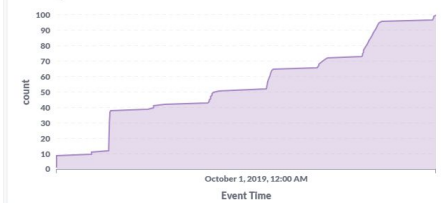
100

total activity

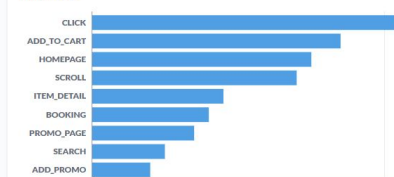
traffic source



session by time



session event



search keyword

Search Keywords	Count
Celana Panjang	1
Dress Kondangan	1
Kaos	1
Sepatu	1
Tas Wanita	1
Grand totals	5

A large black parallelogram on the left side of the slide, with two overlapping orange parallelograms positioned below it, creating a modern, abstract geometric design.

Conclusion & Recommendation

Conclusion & Recommendation

Kesimpulan

- Mengintegrasikan data csv ke postgres dengan batch dan streaming
- Platform mampu mengolah 12 juta baris data secara stream dan 1 juta baris data secara batch
- Mampu mengolah dan memvisualisasi data dengan metabase

Rekomendasi

- Optimasi dengan pandas map dan apply untuk read file csv yang besar
- Implementasi data monitoring seperti grafana atau prometheus
- Analisis lebih mendalam seperti user retention, market share

A large, stylized graphic on the left side of the slide. It consists of a blue outline of a person's head and shoulders. Inside the head is a series of concentric circles: a small light orange circle, a medium orange circle, and a large orange circle. The body is a large blue circle, and inside it is a large orange circle, which contains a smaller orange circle in the center.

**Terima
Kasih**