

# Optimalisasi Performa Data Warehouse dengan Data Mart

<sup>1</sup>Agus Yulianto, <sup>2</sup>Firmansyah,  
<sup>1,2</sup>Universitas Bina Sarana Informatika  
Jakarta, Indonesia

<sup>1</sup>agus.aag@bsi.ac.id, <sup>2</sup>firmansyah.fmh@bsi.ac.id

\*Penulis Korespondensi

Diajukan : 01/10/2024

Diterima : 12/10/2024

Dipublikasi : 12/10/2024

## ABSTRAK

Meningkatkan efisiensi dalam pengambilan keputusan bisnis melalui optimalisasi kinerja data warehouse dengan implementasi data mart juga dapat membuat akurasi data yang digunakan untuk analisis bisnis dengan mengimplementasikan mekanisme pembersihan dan validasi data yang efektif pada data mart, kemudian menyederhanakan proses analisis data dengan menyediakan data yang relevan dan terstruktur dalam data mart dan menurunkan biaya operasional terkait dengan pengelolaan data warehouse melalui optimalisasi penggunaan sumber daya, Model star schema dan model snowflake dalam arsitektur data warehouse merupakan model yang tidak terpisahkan ketika membangun sebuah data warehouse. Performa masing-masing model perlu diukur untuk melihat model yang performanya lebih baik dan dapat diadopsi dalam pengembangan data warehouse. Dalam penelitian ini juga ditambahkan data mart untuk dibandingkan dengan star schema dan snowflake dikarenakan data mart merupakan konsep yang dapat diadopsi dalam data warehouse. Tiga model yaitu star schema, snowflake dan data mart diuji performanya dengan tiga query yang digunakan yaitu query performance, query complexity dan business intelligence performance, pengukuran metrik kinerja dilakukan dengan query performance, seberapa cepat query dapat dijalankan dan menghasilkan hasil. Query complexity melihat seberapa kompleks dapat dijalankan dengan efisien dan business intelligence performance untuk mendukung analisis bisnis dan pengambilan keputusan, namun, peningkatan kompleksitas query pada snowflake schema dapat mempengaruhi kinerja secara keseluruhan. Penelitian ini menyimpulkan bahwa pemilihan model data mart yang tepat sangat penting untuk mencapai optimalisasi kinerja data warehouse.

**Kata Kunci:** *Business Intelligence, Etl, Data, Database, Data Warehouse, Data Mart, Dimensional Modeling, Snowflake, Star Schema*

## I. PENDAHULUAN

Sejak diperkenalkan pertama kali oleh peneliti IBM yaitu *Devlin* dan *Murphy* pada tahun 1988 di dalam artikelnya yang berjudul *An Architecture for a Business Information System data warehouse* saat ini sudah berkembang sangat cepat seiring kebutuhan dengan kebutuhan bisnis. *Data warehouse* berevolusi seiring teknologi basis data yang lebih bervariasi serta mengintegrasikannya ke dalam satu gudang data seperti *data mart, big data, data lake* hingga *real-time data processing*. (Ponnusamy, 2023)

Ada beberapa hal yang mempengaruhi pentingnya data *warehouse* dalam penerapan ekosistem data seperti penyederhanaan dalam penyajian data, data yang berbasis analisis, besarnya data

yang dapat diproses hingga kecepatan performa dari akses data dibanding model data tradisional seperti DBMS (*Database Management System*). (Neamah, 2021)

Seperti namanya, system data warehouse mirip seperti gudang, data masuk ke dalam data warehouse dan keluar data warehouse. Namun, data warehouse memiliki kekurangan yaitu informasi dalam data warehouse belum menyesuaikan dengan kebutuhan setiap entitas dalam organisasi. Seperti misalnya divisi penjualan dan divisi keuangan pastinya memiliki kebutuhan data yang berbeda. Oleh karenanya maka dibutuhkan *data mart* agar informasi dapat disesuaikan dengan kebutuhan penggunaannya.

*Data mart* merupakan data yang diekstraksi dari *data warehouse* dan disesuaikan berdasarkan kebutuhan departemen di organisasi, namun data mart juga ada yang bersifat independen dan hibrida. *Independent* berarti *data mart* tidak harus ekstraksi dari *data warehouse* namun ekstrak dari OLTP (*Online Transactional Processing*). Hibrida berarti artinya *data mart* dapat melakukan ekstraksi dari data warehouse namun juga dapat berjalan tanpa data warehouse.

## II. STUDI LITERATUR

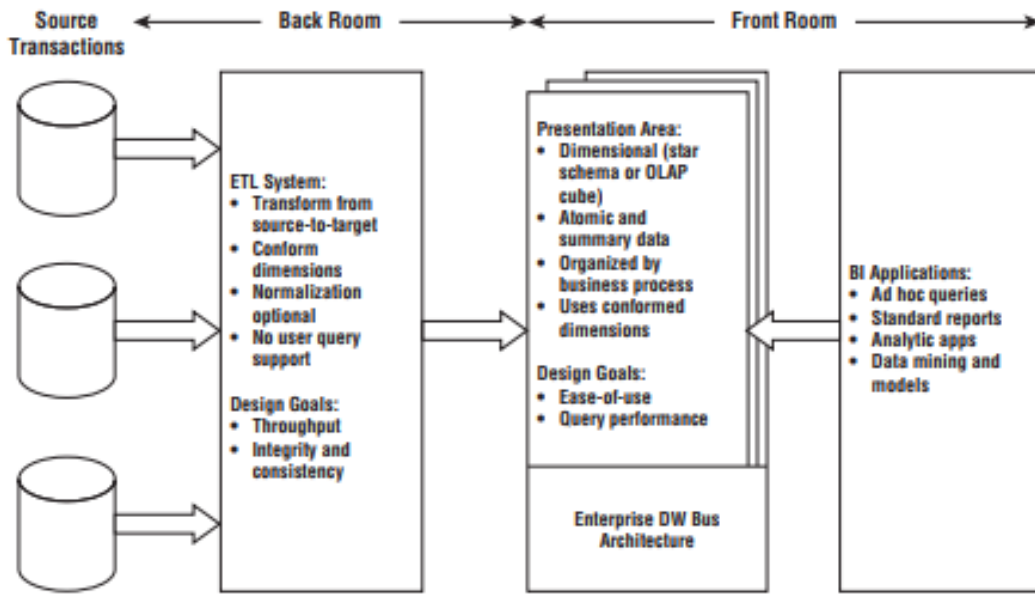
### 1. Penelitian Terkait

Integrasi *big data* dengan teknologi *data warehouse* pada perguruan tinggi dikembangkan untuk mengekstrak data perguruan tinggi menggunakan Hadoop. (Wijaya et al., 2024) Pada penelitian yang lain, peneliti membuat komparasi antara star schema dengan *snowflake* pada arsitektur data warehouse yang dibangunnya, disimpulkan bahwa star schema lebih unggul dibanding *snowflake* seperti kompleksitas query, performa query dan penggabungan antar tabel. (Mohammed, 2019)

Penelitian dilakukan di sebuah bank yaitu bank yaitu Bank Jordania, dimana peneliti mengukur efektifitas penerapan data warehouse pada bank Jordania. Konstruksinya terdiri dari kualitas data, manfaat individu, manfaat organisasi, kualitas sistem dan kepuasan pengguna (Okaily, 2022). Pada teknologi data warehouse selanjutnya, data warehouse akan dapat terintegrasi dengan ML (*Machine Learning*) dan AI (*Artificial Intelligence*), dimana fungsi ML dan AI untuk penyiapan data secara otomatis, optimasi query hingga pengaksesan data warehouse dengan NLP (*Natural Language Processing*). (Ponnusamy, 2023)

### 2. Data Warehouse

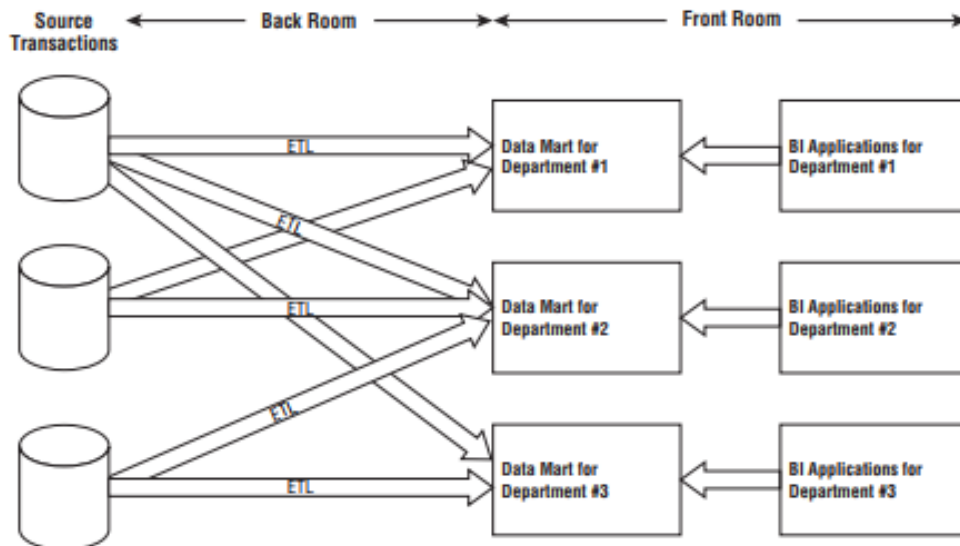
Data warehouse adalah repository informasi terpusat yang datanya berasal dari internal ataupun eksternal organisasi dengan tujuan analisis data dan dapat dikonsumsi oleh organisasi (*Business Intelligence* (Second Edition), David Loshin, 2013, Morgan Kauffman, Pages 105-118). Dalam perkembangannya, data warehouse kini terintegrasi dengan *big data* menggunakan *Hadoop* sebagai teknologinya. (Pticek & Vrdoljak, 2017) Dapat disimpulkan bahwa data warehouse adalah repositori informasi yang berasal dari berbagai sumber baik dari data terstruktur maupun data tidak terstruktur dengan tujuan menyediakan informasi yang dapat dianalisis oleh pengguna akhir menggunakan alat seperti *business intelligence*, *machine learning* ataupun *artificial intelligence*.



Gambar 1  
 Data Warehouse Concept (Kimball, 2013)

### 3. Data Mart

Data mart adalah sub data dari data warehouse yang dibuat untuk mendukung departemen atau divisi yang spesifik dengan menyajikan data sesuai dengan kebutuhan departemen atau divisi di dalam organisasi. Data mart menyediakan data dalam format yang disesuaikan dengan kebutuhan pengguna, grup pengguna tertentu atau kebutuhan untuk pendukung keputusan. (Khoshbakht et al., n.d.) Data mart merupakan ekstraksi informasi, dimana sumbernya adalah data warehouse atau OLTP (Online Transactional Processing) yang informasinya disegmentasikan berdasarkan entitas tertentu dalam organisasi.



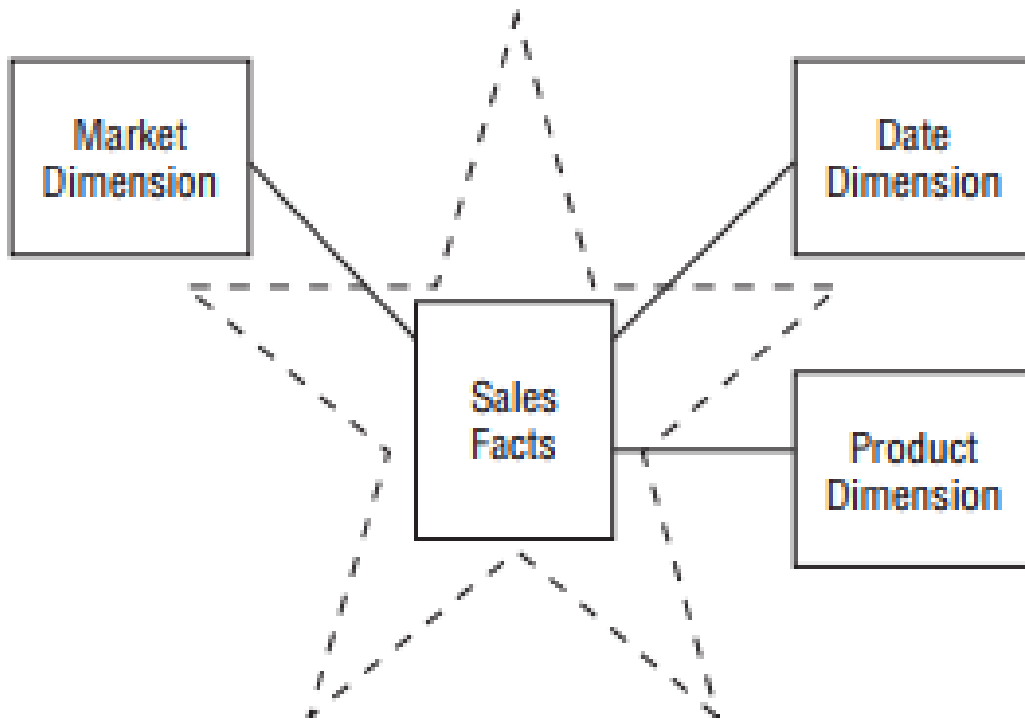
Gambar 2  
 Independent data mart architecture (Kimball, 2013)

#### 4. Data Real-Time Processing

*Data real-time processing* merupakan sejumlah data terbaru yang diambil secara langsung pada waktu tersebut dan disimpan pada penyimpanan tertentu. (KEKEVI & AYDIN, 2022) Teknologi data real-time processing saat ini seperti Apache Kafka, RabbitMQ, Apache Spark dan lain sebagainya.

#### 5. Star Schema

Star schema merupakan pemodelan data dimana data di pisahkan berdasarkan tabel dimensi dan tabel fakta yang arsitekturnya mirip seperti bintang sehingga dinamakan star schema. Tabel dimensi (dimension table) digunakan untuk menyimpan data rujukan (master), dalam hal data penjualan seperti data customer, produk atau wilayah. Tabel fakta (fact table) digunakan untuk menyimpan data yang terkait dengan proses atau transaksi seperti misalnya penjualan atau pembelian. (Amin et al., n.d.)



Gambar 3  
Star Schema (Kimball, 2013)

#### 6. Snowflake

Dalam model snowflake schema, dimensi yang terdapat dalam star schema dipecah lebih lanjut menjadi tabel-tabel tambahan yang lebih kecil. Setiap dimensi menjadi lebih ternormalisasi, dengan pengelompokan data ke dalam sub-tabel yang lebih spesifik, membentuk pola seperti kepingan salju (snowflake) serta model menjadi dimensi yang dinormalisasi dimana setiap atribut dimensi dipecah ke dalam tabel yang lebih kecil untuk menghindari redundansi dan mempermudah pengelolaan data. Relasi antar tabel dalam snowflake menggunakan banyak hubungan antar tabel untuk merepresentasikan data. (Novreza et al., 2020)

### III. METODE

#### A. Query Performance

Query performance merupakan cara untuk mengukur seberapa cepat query mulai dari mengeksekusi query sampai dengan menampilkan data. Proses eksekusi query dibagi menjadi 2, yaitu pada saat pertama kali query tersebut dieksekusi dan query yang sudah dieksekusi berikutnya setelah pertama kali dieksekusi. Perbedaan query yang dieksekusi pertama kali dan query yang dieksekusi setelahnya adalah pada kecepatannya, dimana query yang dieksekusi setelah eksekusi pertama adalah performanya lebih cepat karena sebelumnya sudah disimpan ke memori dan cache. (Triaji et al., 2022)

Dalam pengukuran kinerja query, peneliti menggunakan kecepatan eksekusi query berikutnya dan mengambil nilai rata-rata kecepatannya.

### **B. Query Complexity**

Kompleksitas kueri merupakan metode pengukuran untuk mengukur kompleksitas suatu kueri. Kompleksitas kueri dapat dilihat dari beberapa aspek seperti seberapa banyaknya join tabel, condition dan grouping di dalam perintah kueri.

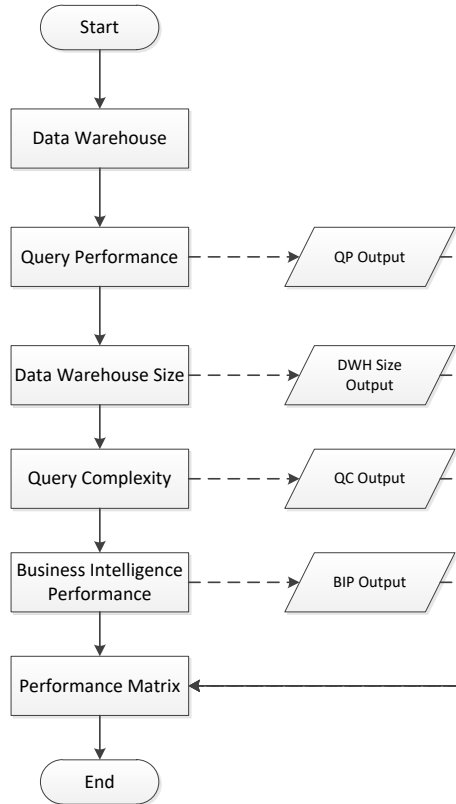
### **C. Business Intelligence Performance**

Business intelligence (BI) merupakan alat untuk menyajikan informasi dalam bentuk dashboard dimana di dalamnya terdapat informasi dalam tabel, grafik dan label. Produk BI saat ini bervariasi seperti PowerBI, Tableau, Qlikview, Looker dan lain sebagainya. Setiap produk BI pada umumnya memiliki fitur untuk mengukur performa data yang disajikan, dalam hal ini kami menggunakan fitur performance dalam PowerBI. Alasan mengukur performa di dalam BI adalah dikarenakan BI merupakan alat yang menyajikan informasi langsung melalui antar muka kepada pengguna.

## **IV. HASIL DAN PEMBAHASAN**

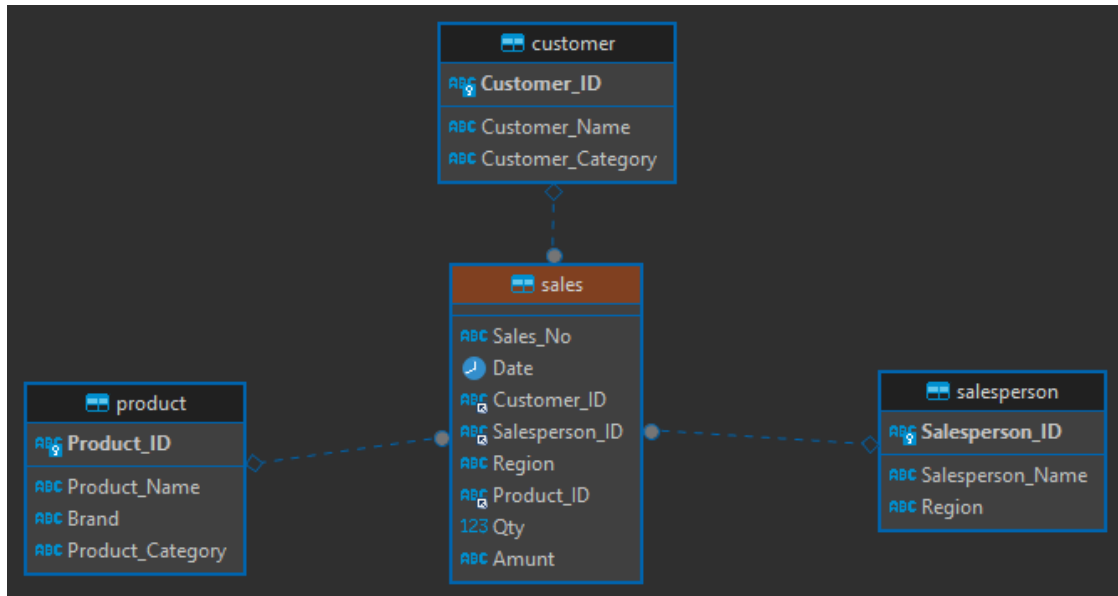
Dalam penelitian ini akan diukur terlebih dahulu performa data warehouse tanpa data mart. Model yang digunakan dalam data warehouse yaitu model star schema dan model snowflake. Data yang digunakan untuk pembuatan data warehouse adalah data transaksi penjualan periode tahun 2018 sampai dengan tahun 2023 yang telah dibuat dalam dua model di atas.

Alur proses pengukuran performa dalam penelitian ini dimulai dari proses Extract Transform Load (ETL) data penjualan ke dalam data warehouse. Data warehouse yang sudah didesain dengan star schema dan snowflake kemudian diukur performanya mulai dari query performance, data warehouse size, query complexity dan business intelligence performance. Masing-masing proses pengukuran performa menghasilkan keluaran seperti kecepatan kueri, ukuran data warehouse, kompleksitas kueri dan kecepatan akses pada BI. Alur proses untuk mengukur performa data warehouse seperti di bawah ini :



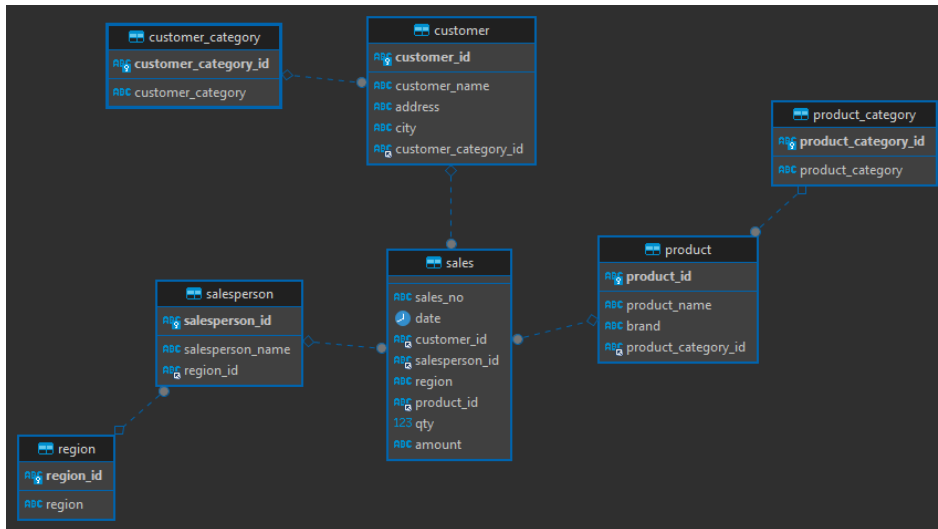
Gambar 4  
 Flowchart Penelitian Uji Performa

Data warehouse dengan model star schema dimana di dalamnya terdapat 3 tabel dimensi dan 1 tabel fakta seperti di bawah ini :



Gambar 5  
 Star Schema Model

Sedangkan untuk model snowflake terdiri dari 3 tabel dimensi, 3 tabel sub dimensi dan 1 tabel fakta seperti di bawah ini :



Gambar 6  
 Snowflake Model

Dari data warehouse yang sudah dimodelkan, dilakukan kueri untuk mengekstrak data menjadi data penjualan dimana kolomnya adalah customer, year, month dan sales qty. Kueri diukur menggunakan query performance (QP) dan query complexity (QC). Datawarehouse size (DS) digunakan untuk mengukur besaran data. Business intelligence performance (BIP) digunakan untuk mengukur performa jika kueri tersebut dijalankan menggunakan PowerBI. Hasil pengujian terhadap 2 model di bawah ini :

Table 1  
 Hasil uji komparasi star schema vs snowflake

Model	QP	QC	BIP
Star Schema	1411.466 ms	4 table	175.8 ms
Snowflake	1358.358 ms	7 table	159.3 ms

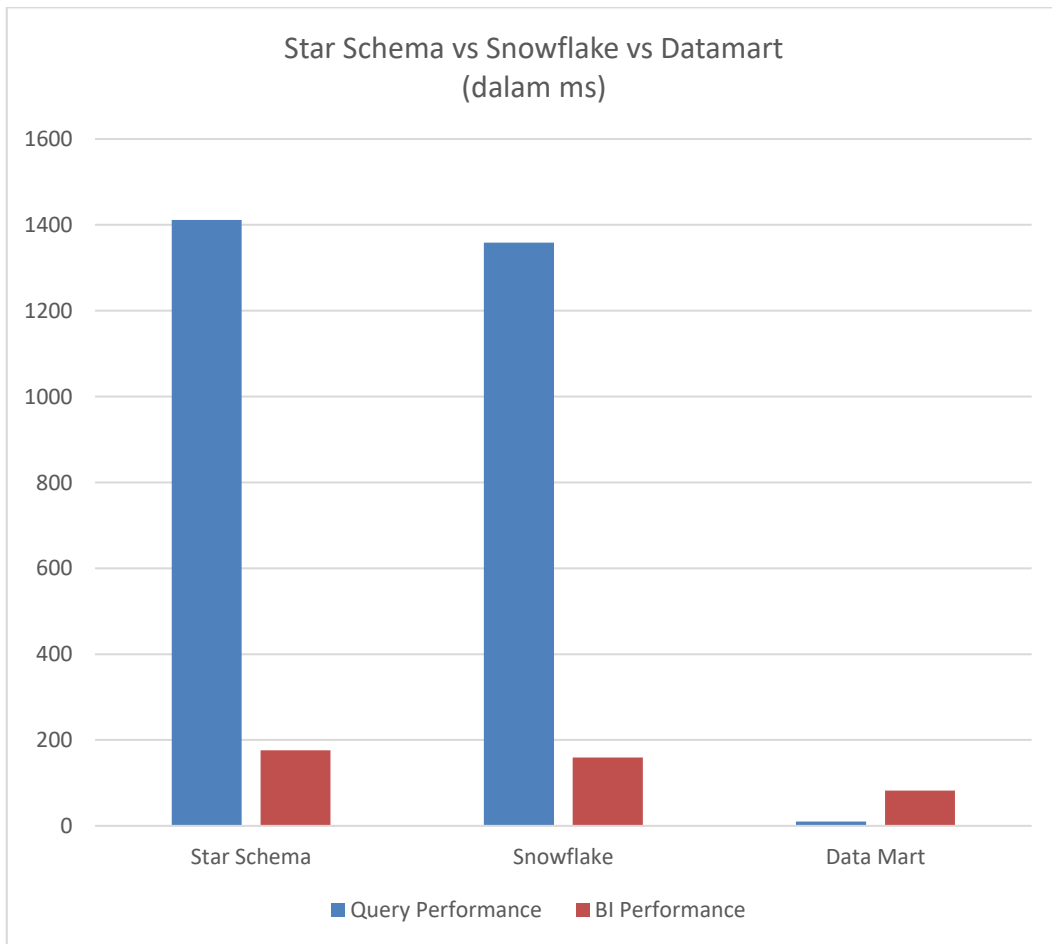
Dari hasil pada tabel di atas kita dapat melihat bahwa model snowflake lebih cepat berdasarkan query performance dan business intelligence performance. Sedangkan untuk query complexity, star schema lebih sedikit menggunakan relasi antar tabel ini wajar dikarenakan dibandingkan dengan snowflake, di dalam model star schema tidak ada sub dimension. Pengukuran performa menggunakan PowerBI terlihat bahwa kecepatan visualisasi secara rata-rata, model snowflake lebih cepat dibanding model star schema.

Pengujian berikutnya adalah menggunakan data mart dimana di dalam data mart data digabungkan dalam 1 tabel yang sudah digrup dan diagregasi. Hasil pengujian menggunakan query performance, query complexity dan business intelligence performance terlihat lebih cepat dibanding model star schema dan snowflake seperti terlihat tabel di bawah ini :

Table 2  
 Hasil uji performa data mart

QP	QC	BIP
10.14 ms	1 table	82 ms

Data mart signifikan mempercepat proses penyajian data baik dari sisi kueri di PostgreSQL maupun pada business intelligence PowerBI.



Gambar 7  
 Hasil Uji Komparasi

Pada grafik di atas terlihat antara star schema, snowflake dan data mart perbedaan performanya sangat signifikan sehingga data mart dapat menjadi solusi dalam pengembangan business intelligence.

**V. KESIMPULAN**

Model data mart terbukti lebih cepat dalam penyajian data dibanding model data warehouse star schema dan snowflake. Secara teori, data mart memang bukanlah data model seperti halnya star schema atau snowflake namun data mart merupakan tabel yang diekstrak dan diaggregasi menjadi lebih sederhana dan disesuaikan dengan kebutuhan penggunaannya. Namun data mart juga berdampak pada ukuran data yang besar dikarenakan isi dari data mart disesuaikan dengan

kebutuhan divisi di dalam organisasi. Analisa performa yang digunakan dalam penelitian ini menggunakan Power BI, namun penelitian ke depan perlu juga dibandingkan dengan BI lain seperti Looker, Tableau, Qlikview dan sebagainya.

## VI. REFERENSI

- Amin, M. M., Sutrisman, A., & Dwitayanti, Y. (n.d.). Development of Star-Schema Model for Lecturer Performance in Research Activities. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 12, Issue 9). [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- KEKEVI, U., & AYDIN, A. A. (2022). Real-Time Big Data Processing and Analytics: Concepts, Technologies, and Domains. *Computer Science*. <https://doi.org/10.53070/bbd.1204112>
- Khoshbakht, F., Shiranzaei, A., & Quadri, S. M. K. (n.d.). International Journal of INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING Design & Develop: Data Warehouse & Data Mart for Business Organization. In *International Journal of Intelligent Systems and Applications in Engineering IJISAE* (Vol. 2023, Issue 3s). [www.ijisae.org](http://www.ijisae.org)
- Mohammed, K. I. (2019). Data Warehouse Design and Implementation Based on Star Schema vs. Snowflake Schema. *International Journal of Academic Research in Business and Social Sciences*, 9(14). <https://doi.org/10.6007/ijarbss/v9-i14/6502>
- Neamah, A. F. (2021). Adoption of Data Warehouse in University Management: Wasit University Case Study. *Journal of Physics: Conference Series*, 1860(1). <https://doi.org/10.1088/1742-6596/1860/1/012027>
- Novreza, V., Munarko, Y., & Husniah, L. (2020). Data Warehouse Menggunakan Snowflake Schema Pada Virtual Shop. *REPOSITOR*, 2(1), 67–78.
- Ponnusamy, S. (2023). Evolution of Enterprise Data Warehouse: Past Trends and Future Prospects. *International Journal of Computer Trends and Technology*, 71(9), 1–6. <https://doi.org/10.14445/22312803/ijctt-v71i9p101>
- Pticek, M., & Vrdoljak, B. (2017). Big data and new data warehousing approaches. *ACM International Conference Proceeding Series*, 6–10. <https://doi.org/10.1145/3141128.3141139>
- Triaji, B., Andriyani, W., Suprawoto, T., Nugroho, M. A., & Kartadie, R. (2022). Query Execution Performance Analysis of Column-Oriented Database in Dashboard. *Journal of Intelligent Software Systems*, 1(2), 122. <https://doi.org/10.26798/jiss.v1i2.768>
- Wijaya, W., Wiratama, J., & Wijaya, S. F. (2024). Implementation of Data Warehouse and Star Schema for Optimizing Property Business Decision Making. *G-Tech: Jurnal Teknologi Terapan*, 8(2), 1242–1250. <https://doi.org/10.33379/gtech.v8i2.4091>