

BUSSINESS INTELLIGENCE

Laporan Project UAS BI

Dosen Pembimbing:

Endah Septa Sintiya, S.Pd., M.Kom.



Disusun oleh:

Achmad Maulana Hamzah	2341720172
Gwido Putra Wijaya	2341720103
Stevan Zaky Setyanto	2341720101
Vemas Bagus Fermanda	2341720137

JURUSAN TEKNOLOGI INFORMASI
PROGRAM STUDI D-IV TEKNIK INFORMATIKA
POLITEKNIK NEGERI MALANG

2025

DAFTAR ISI

BAB 1 - PENDAHULUAN.....	3
1.1 Latar Belakang.....	3
1.2 Tujuan.....	3
1.3 Ruang Lingkup.....	3
BAB 2 - TINJAUAN PUSTAKA.....	5
2.1 Data Warehouse.....	5
2.2 ETL (Extract, Transform, Load).....	5
2.3 Star Schema.....	5
2.4 Key Performance Indicator (KPI).....	5
2.5 Tools yang Digunakan.....	5
BAB 3 - STUDI KASUS DAN PENGAMBILAN DATA.....	6
3.1 Deskripsi Studi Kasus: Superstore Sales.....	6
3.2 Sumber Dataset.....	7
3.3 Struktur Data.....	7
3.4 Tabel Yang di-Bentuk.....	7
BAB 4 - PERANCANGAN DATA WAREHOUSE.....	8
4.1 Identifikasi Tabel Fakta dan Dimensi.....	8
4.2 Star Schema.....	9
4.3 Penjelasan Relasi Antar Tabel.....	9
BAB 5 - PERANCANGAN ETL.....	10
5.1 Struktur Perancangan ETL.....	11
5.2 Membuat Database.....	11
5.3 Membuat ETL Untuk Setiap Dimensi.....	12
5.3.1 Buat Transformasi ETL untuk Dim_Customer.....	12
5.3.2 Buat Transformasi ETL untuk Dim_Product.....	12
5.3.3 Buat Transformasi ETL untuk Dim_Order.....	13
5.3.4 Buat Transformasi ETL untuk Dim_Date.....	13
5.4 Membuat ETL Untuk Tabel Fakta.....	14
5.5 Membuat Job Untuk Menjalankan Semua Proses.....	14
BAB 6 - HASIL DAN ANALISIS.....	16
6.1 Query Testing.....	16
6.2 Analisis KPI dan Dashboard.....	19
BAB 7 - KESIMPULAN DAN SARAN.....	21
7.1 Kesimpulan.....	21
7.2 Saran.....	21

BAB 1 - PENDAHULUAN

1.1 Latar Belakang

Dalam era digital saat ini, perusahaan dan organisasi menghasilkan data dalam jumlah yang sangat besar dari berbagai proses operasional, seperti transaksi penjualan, pengelolaan inventaris, data pelanggan, serta aktivitas karyawan. Data yang melimpah tersebut tidak akan memiliki nilai apabila tidak dikelola dan dianalisis dengan baik. Oleh karena itu, dibutuhkan suatu sistem yang mampu mengintegrasikan data dari berbagai sumber, membersihkan, serta menyimpannya dalam bentuk yang terstruktur sehingga dapat digunakan untuk mendukung pengambilan keputusan.

Business Intelligence (BI) merupakan salah satu pendekatan penting dalam pengelolaan data yang berfokus pada proses transformasi data mentah menjadi informasi yang bermanfaat. Salah satu komponen utama dalam BI adalah Data Warehouse, yaitu sistem penyimpanan data terintegrasi yang dirancang untuk analisis dan pelaporan. Untuk membangun sebuah data warehouse, proses ETL (Extract, Transform, Load) menjadi tahapan utama yang bertujuan untuk mengekstrak data dari berbagai sumber, melakukan transformasi agar data bersih dan konsisten, serta memuatnya ke dalam struktur data yang telah dirancang.

Melalui tugas proyek ini, mahasiswa diminta untuk memahami secara langsung bagaimana merancang data warehouse menggunakan model star schema, membangun pipeline ETL sesuai kebutuhan analisis, serta menghasilkan visualisasi KPI (Key Performance Indicators) sebagai bagian dari proses analisis Business Intelligence. Dengan demikian, mahasiswa tidak hanya mempelajari konsep teoretis, tetapi juga mengimplementasikan secara nyata proses pengelolaan data menggunakan dataset dummy dari sumber publik.

1.2 Tujuan

Tujuan dari proyek ini adalah:

1. Mengambil dan memahami dataset dummy dari sumber publik.
2. Merancang data warehouse menggunakan star schema.
3. Membangun proses ETL untuk mengekstrak, membersihkan, dan memuat data.
4. Menyediakan data terstruktur untuk analisis Business Intelligence.
5. Menampilkan visualisasi KPI sebagai hasil analisis.

1.3 Ruang Lingkup

Ruang lingkup proyek ini mencakup seluruh proses pembangunan sistem Business Intelligence menggunakan dataset *Superstore Sales*. Proyek ini difokuskan pada kegiatan sebagai berikut:

1. Pemilihan dan pengambilan dataset publik yang berisi data transaksi penjualan, pelanggan, produk, pesanan, dan informasi terkait lainnya.
2. Perancangan data warehouse menggunakan model *Star Schema* yang terdiri dari satu tabel fakta dan beberapa tabel dimensi.
3. Pembuatan proses ETL yang meliputi tahap Extract dari file sumber, Transform untuk pembersihan serta penyesuaian struktur data, dan Load ke dalam tabel fakta dan dimensi.
4. Pengolahan dan penyimpanan data terstruktur di dalam data warehouse agar dapat digunakan untuk analisis lebih lanjut.
5. Penyusunan visualisasi KPI, seperti analisis penjualan, performa produk, segmentasi pelanggan, dan tren waktu.
6. Pembuatan dokumentasi proyek, meliputi desain, implementasi proses ETL, hasil data warehouse, dan visualisasi.

Ruang lingkup difokuskan pada proses integrasi data, transformasi, dan analisis, tanpa mencakup pengembangan aplikasi front-end atau backend di luar kebutuhan visualisasi BI.

BAB 2 - TINJAUAN PUSTAKA

2.1 Data Warehouse

Data warehouse adalah sistem penyimpanan data yang digunakan untuk mendukung proses analisis dan pelaporan. Karakteristiknya meliputi: subject-oriented, integrated, time-variant, dan non-volatile.

2.2 ETL (Extract, Transform, Load)

ETL adalah proses mengambil data dari sumber (extract), membersihkan dan mengubahnya (transform), lalu memasukkannya ke target data warehouse (load).

2.3 Star Schema

Star schema adalah model data yang terdiri dari satu tabel fakta dan beberapa tabel dimensi yang terhubung melalui foreign key.

2.4 Key Performance Indicator (KPI)

KPI adalah metrik terukur yang digunakan untuk mengevaluasi performa bisnis.

2.5 Tools yang Digunakan

→ Google Looker Studio

Digunakan sebagai alat visualisasi akhir untuk menampilkan dashboard KPI berdasarkan data warehouse.

1. Mengambil data dari database data warehouse (MySQL/PostgreSQL).
2. Membuat dashboard visual untuk menunjukkan KPI seperti:
 - Total penjualan per periode,
 - Produk terlaris,
 - Penjualan per wilayah,
 - Segment pelanggan terbesar.
3. Menyajikan hasil analisis sesuai permintaan tugas pada poin:
 - Hasil query,
 - Visualisasi dashboard,
 - Analisis KPI.

→ Pentaho Data Integration

Digunakan sebagai alat utama untuk membangun proses ETL sesuai kebutuhan tugas:

1. Extract
Mengambil data dummy (CSV) dari studi kasus seperti data penjualan, pelanggan, produk, dan lokasi.
2. Transform

- Membersihkan data (hilangkan duplikasi, perbaikan format tanggal, normalisasi teks).
 - Memisahkan data menjadi tabel dimensi (Dim_Produk, Dim_Pelanggan, dll).
 - Menggabungkan data untuk membuat tabel fakta (Fact_Penjualan).
 - Menyesuaikan struktur agar sesuai dengan *star schema*.
3. Load
- Memuat data ke database data warehouse (MySQL/PostgreSQL) sesuai skema bintang yang telah dirancang.

BAB 3 - STUDI KASUS DAN PENGAMBILAN DATA

3.1 Deskripsi Studi Kasus: Superstore Sales

Dataset *Superstore Sales* merupakan salah satu dataset publik yang banyak digunakan dalam analisis Business Intelligence. Dataset ini berisi data transaksi penjualan sebuah toko retail yang beroperasi di berbagai wilayah di Amerika Serikat. Di dalamnya terdapat informasi lengkap mengenai pelanggan, produk, lokasi pengiriman, metode pengiriman, hingga nilai penjualan dan profit.

Dataset ini cocok digunakan sebagai studi kasus karena memiliki struktur data yang kaya, mencakup transaksi, pelanggan, produk, kategori, dan variabel lain yang memungkinkan dilakukan analisis performa penjualan secara komprehensif. Selain itu, data sudah tersedia dalam format tabular sehingga sangat sesuai untuk proses ETL dan pembangunan data warehouse.

3.2 Sumber Dataset

Dataset ini biasanya tersedia di platform publik yaitu kagle:

<https://www.kaggle.com/datasets/rohithsahoo/sales-forecasting>

Format dataset: CSV

3.3 Struktur Data

Dataset Superstore berisi beberapa atribut utama, antara lain:

1. Data Penjualan (Transaksi)
2. Data Pelanggan
3. Data Produk
4. Data Waktu

3.4 Tabel Yang di-Bentuk

1. Tabel Fakta: Fact_Sales
 - Sales
 - Quantity
 - Discount
 - Profit
 - Foreign Key → Customer, Product, Date, Location, Ship Mode
2. Tabel Dimensi:
 - Dim_Customer
 - Dim_Product
 - Dim_Date

- Dim_Location
- Dim_ShipMode
- (Optional) Dim_Category

BAB 4 - PERANCANGAN DATA WAREHOUSE

4.1 Identifikasi Tabel Fakta dan Dimensi

Berdasarkan analisis terhadap struktur data pada dataset *Superstore Sales*, maka data warehouse dirancang menggunakan model Star Schema. Model ini dipilih karena sederhana, mudah diimplementasikan, serta sesuai untuk kebutuhan analisis penjualan.

Dalam skema ini, terdapat satu tabel fakta yang berisi data transaksi penjualan, serta beberapa tabel dimensi yang menyimpan informasi deskriptif terkait pelanggan, produk, pesanan, dan tanggal. Hubungan antar tabel menggunakan mekanisme foreign key yang menghubungkan tabel fakta dengan masing-masing tabel dimensi.

1. Tabel Fakta: fact_sales
2. Tabel Dimensi
 - a. Dimensi Pesanan – dim_order
 - b. Dimensi Pelanggan – dim_customer
 - c. Dimensi Produk – dim_product
 - d. Dimensi Tanggal – dim_date

4.2 Star Schema



Star Schema yang dihasilkan terdiri dari satu tabel fakta yaitu fact_sales yang berada di pusat, dan empat tabel dimensi yang terhubung langsung dengannya. Struktur ini memungkinkan proses query menjadi lebih cepat dan efisien untuk analisis penjualan berdasarkan:

- Waktu transaksi
- Detail pelanggan

- Detail produk
- Detail pesanan/pengiriman

Diagram Star Schema yang telah dibuat menggambarkan hubungan *one-to-many* antara tabel dimensi dengan tabel fakta melalui foreign key pada masing-masing dimensi.

4.3 Penjelasan Relasi Antar Tabel

- dim_customer → fact_sales: Satu pelanggan dapat melakukan banyak transaksi.
- dim_product → fact_sales: Satu produk dapat muncul dalam banyak transaksi.
- dim_order → fact_sales: Satu order dapat berisi beberapa item (multi-row).
- dim_date → fact_sales: Setiap transaksi memiliki satu tanggal pemesanan.

Relasi-relasi ini memastikan bahwa data warehouse mampu menjawab berbagai pertanyaan analitis, seperti:

- Penjualan per produk
- Penjualan per kategori
- Penjualan berdasarkan wilayah
- Penjualan per segmen pelanggan
- Tren penjualan per bulan/tahun

BAB 5 - PERANCANGAN ETL

5.1 Struktur Perancangan ETL

```
ETL_PROJECT/
|
|— dim_customer.ktr
|— dim_product.ktr
|— dim_order.ktr
|— dim_date.ktr
|— fact_sales.ktr
|
|— job_master.kjb (job untuk menjalankan semua ETL sekaligus)
```

5.2 Membuat Database

Buat database di phpmyadmin

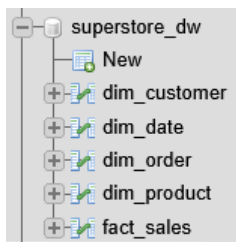


Table	Action	Rows	Type	Collation	Size	Overhead
<input type="checkbox"/> dim_customer	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
<input type="checkbox"/> dim_date	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
<input type="checkbox"/> dim_order	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
<input type="checkbox"/> dim_product	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	16.0 KiB	-
<input type="checkbox"/> fact_sales	★ Browse Structure Search Insert Empty Drop	0	InnoDB	utf8mb4_0900_ai_ci	80.0 KiB	-
5 tables	Sum	0	InnoDB	utf8mb4_0900_ai_ci	144.0 KiB	0 B

Buat koneksi di pentaho

Connection name:
superstore_dw

Connection type:

KingbaseES

LucidDB

MS Access

MS SQL Server

MS SQL Server (Native)

MariaDB

MaxDB (SAP DB)

MonetDB

MySQL

Native Mondrian

Neoview

Netezza

...

Settings

Host Name:
localhost

Database Name:
superstore_dw

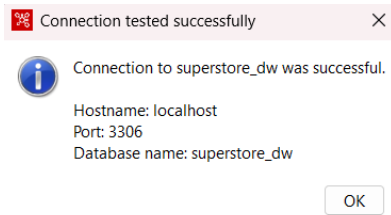
Port Number:
3306

Username:
root

Password:

☒ Use Result Streaming Cursor

Access:
Native (JDBC)
ODBC
JNDI



5.3 Membuat ETL Untuk Setiap Dimensi

5.3.1 Buat Transformasi ETL untuk Dim_Customer

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

2025-12-05 08:40:38.743 - Spoon - Save file as...
 2025-12-05 08:40:38.843 - Spoon - Transformation opened.
 2025-12-05 08:40:38.843 - Spoon - Launching transformation [dim_customer]...
 2025-12-05 08:40:38.843 - Spoon - Started the transformation execution.
 2025-12-05 08:40:39.159 - dim_customer - Dispatching started for transformation [dim_customer]
 2025-12-05 08:40:39.180 - Carte - Installing timer to purge stale objects after 1440 minutes.
 2025-12-05 08:40:39.190 - Table output.0 - Connected to database [superstore_dw] (commit=1000)
 2025-12-05 08:40:39.217 - CSV file input.0 - Header row skipped in file 'D:\Materi Kuliah\SEMESTER 5\Business Intelligence\PBL\dataset\train.csv'
 2025-12-05 08:40:39.300 - CSV file input.0 - Finished processing (I=9801, O=0, R=0, W=9800, U=0, E=0)
 2025-12-05 08:40:39.305 - Select values.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)
 2025-12-05 08:40:39.341 - Sort rows.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)
 2025-12-05 08:40:39.346 - Unique rows.0 - Finished processing (I=0, O=0, R=9800, W=793, U=0, E=0)
 2025-12-05 08:40:39.490 - Table output.0 - Finished processing (I=0, O=793, R=793, W=793, U=0, E=0)
 2025-12-05 08:40:39.492 - Spoon - The transformation has finished!!

	customer_id	customer_name	segment	country	city	state	postal_code	region
<input type="checkbox"/>	AR-10825	Anthony Rawles	Corporate	United States	Vancouver	Washington	98661	West
<input type="checkbox"/>	AS-10045	Aaron Smayling	Corporate	United States	Jacksonville	North Carolina	28540	South
<input type="checkbox"/>	AS-10090	Adam Shillingsburg	Consumer	United States	Charlottesville	Virginia	22901	South
<input type="checkbox"/>	AS-10135	Adrian Shami	Home Office	United States	New York City	New York	10035	East
<input type="checkbox"/>	AS-10225	Alan Schoenberger	Corporate	United States	New York City	New York	10024	East
<input type="checkbox"/>	AS-10240	Alan Shonely	Consumer	United States	Tampa	Florida	33614	South
<input type="checkbox"/>	AS-10285	Alejandro Savely	Corporate	United States	San Francisco	California	94109	West
<input type="checkbox"/>	AS-10630	Ann Steele	Home Office	United States	Pasadena	Texas	77506	Central
<input type="checkbox"/>	AT-10135	Alvesa Tata	Home Office	United States	Los Angeles	California	90004	West

5.3.2 Buat Transformasi ETL untuk Dim_Product

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

2025-12-05 08:49:14.587 - Spoon - Save file as...
 2025-12-05 08:49:14.609 - Spoon - Transformation opened.
 2025-12-05 08:49:14.609 - Spoon - Launching transformation [dim_product]...
 2025-12-05 08:49:14.609 - Spoon - Started the transformation execution.
 2025-12-05 08:49:14.889 - dim_product - Dispatching started for transformation [dim_product]
 2025-12-05 08:49:14.892 - Table output.0 - Connected to database [superstore_dw] (commit=1000)
 2025-12-05 08:49:14.896 - CSV file input.0 - Header row skipped in file 'D:\Materi Kuliah\SEMESTER 5\Business Intelligence\PBL\dataset\train.csv'
 2025-12-05 08:49:14.987 - CSV file input.0 - Finished processing (I=9801, O=0, R=0, W=9800, U=0, E=0)
 2025-12-05 08:49:14.994 - Select values.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)
 2025-12-05 08:49:15.042 - Sort rows.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)
 2025-12-05 08:49:15.070 - Unique rows.0 - Finished processing (I=0, O=0, R=9800, W=1861, U=0, E=0)
 2025-12-05 08:49:15.301 - Table output.0 - Finished processing (I=0, O=1861, R=1861, W=1861, U=0, E=0)
 2025-12-05 08:49:15.302 - Spoon - The transformation has finished!!

		product_id	category	sub_category	product_name
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000112	Furniture	Bookcases	Bush Birmingham Collection Bookcase, Dark Cherry
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000330	Furniture	Bookcases	Sauder Camden County Barrister Bookcase, Planked C...
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000362	Furniture	Bookcases	Sauder Inglewood Library Bookcases
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000468	Furniture	Bookcases	O'Sullivan 2-Shelf Heavy-Duty Bookcases
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000711	Furniture	Bookcases	Hon Metal Bookcases, Gray
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10000780	Furniture	Bookcases	O'Sullivan Plantations 2-Door Library in Landvery ...
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10001337	Furniture	Bookcases	O'Sullivan Living Dimensions 2-Shelf Bookcases
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10001519	Furniture	Bookcases	O'Sullivan 3-Shelf Heavy-Duty Bookcases
<input type="checkbox"/>	Edit Copy Delete	FUR-BO-10001567	Furniture	Bookcases	Rush Westfield Collection Bookcases, Dark Cherry F

5.3.3 Buat Transformasi ETL untuk Dim_Order

dim_customer dim_product dim_order

CSV file input → Select values → Sort rows → Unique rows → Table output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-12-05 08:54:03.594 - Spoon - Save file as...

2025-12-05 08:54:03.614 - Spoon - Transformation opened.

2025-12-05 08:54:03.614 - Spoon - Launching transformation [dim_order]...

2025-12-05 08:54:03.614 - Spoon - Started the transformation execution.

2025-12-05 08:54:03.865 - dim_order - Dispatching started for transformation [dim_order]

2025-12-05 08:54:03.867 - Table output.0 - Connected to database [superstore_dw] (commit=1000)

2025-12-05 08:54:03.870 - CSV file input.0 - Header row skipped in file 'D:\Materi Kuliah\SEMESTER 5\Business Intelligence\PBL\dataset\train.csv'

2025-12-05 08:54:03.932 - CSV file input.0 - Finished processing (I=0, O=0, R=0, W=9800, U=0, E=0)

2025-12-05 08:54:03.938 - Select values.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 08:54:03.981 - Sort rows.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 08:54:03.999 - Unique rows.0 - Finished processing (I=0, O=0, R=9800, W=4922, U=0, E=0)

2025-12-05 08:54:04.556 - Table output.0 - Finished processing (I=0, O=4922, R=4922, W=4922, U=0, E=0)

2025-12-05 08:54:04.557 - Spoon - The transformation has finished!!

		order_id	ship_date	ship_mode
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100006	2015-09-13	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100090	2015-07-12	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100293	2015-03-18	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100328	2015-02-03	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100363	2015-04-15	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100391	2015-05-29	Standard Class
<input type="checkbox"/>	Edit Copy Delete	CA-2015-100678	2015-04-22	Standard Class

5.3.4 Buat Transformasi ETL untuk Dim_Date

dim_customer dim_product dim_order dim_date

CSV file input → Select values → Select values 2 → Sort rows → Unique rows → User defined Java expression → Calculator → Select values 3 → Table output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-12-05 09:23:06.678 - Spoon - Transformation opened.

2025-12-05 09:23:06.678 - Spoon - Launching transformation [dim_date]...

2025-12-05 09:23:06.678 - Spoon - Started the transformation execution.

2025-12-05 09:23:06.702 - dim_date - Dispatching started for transformation [dim_date]

2025-12-05 09:23:06.705 - Table output.0 - Connected to database [superstore_dw] (commit=1000)

2025-12-05 09:23:06.727 - CSV file input.0 - Header row skipped in file 'D:\Materi Kuliah\SEMESTER 5\Business Intelligence\PBL\dataset\train.csv'

2025-12-05 09:23:06.737 - Select values.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:23:06.752 - Select values 2.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:23:06.793 - Sort rows.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:23:06.814 - Unique rows.0 - Finished processing (I=0, O=0, R=9800, W=1230, U=0, E=0)

2025-12-05 09:23:06.817 - User defined Java expression.0 - Finished processing (I=0, O=0, R=1230, W=1230, U=0, E=0)

2025-12-05 09:23:06.822 - Calculator.0 - Finished processing (I=0, O=0, R=1230, W=1230, U=0, E=0)

2025-12-05 09:23:06.844 - Select values 3.0 - Finished processing (I=0, O=0, R=1230, W=1230, U=0, E=0)

2025-12-05 09:23:06.962 - Table output.0 - Finished processing (I=0, O=1230, R=1230, W=1230, U=0, E=0)

2025-12-05 09:23:06.962 - Spoon - The transformation has finished!!

				date_id	full_date	day	month	month_name	year	quarter
<input type="checkbox"/>	Edit	Copy	Delete	20150103	2015-01-03	3	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150104	2015-01-04	4	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150105	2015-01-05	5	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150106	2015-01-06	6	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150107	2015-01-07	7	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150109	2015-01-09	9	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150110	2015-01-10	10	1	1	2015	1
<input type="checkbox"/>	Edit	Copy	Delete	20150111	2015-01-11	11	1	1	2015	1

5.4 Membuat ETL Untuk Tabel Fakta

dim_customer dim_product dim_order dim_date fact_sales

CSV file input Select values Lookup Customer Lookup Product Lookup Order Lookup Date Select values 2 Table output

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2025-12-05 09:46:33.490 - Spoon - Save file as...

2025-12-05 09:46:33.512 - Spoon - Transformation opened.

2025-12-05 09:46:33.512 - Spoon - Launching transformation (fact_sales)...

2025-12-05 09:46:33.512 - Spoon - Started the transformation execution.

2025-12-05 09:46:33.798 - fact_sales - Dispatching started for transformation [fact_sales]

2025-12-05 09:46:33.806 - Table output.0 - Connected to database [superstore_dw] (commit=1000)

2025-12-05 09:46:33.852 - CSV file input.0 - Header row skipped in file 'D:\Materi Kuliah\SEMESTER 5\Business Intelligence\PBL\dataset\train.csv'

2025-12-05 09:46:33.871 - CSV file input.0 - Finished processing (I=9801, O=0, R=0, W=9800, U=0, E=0)

2025-12-05 09:46:33.892 - Select values.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:35.919 - Lookup Customer.0 - Finished processing (I=9800, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:35.949 - Lookup Product.0 - Finished processing (I=9800, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:35.962 - Lookup Order.0 - Finished processing (I=9800, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:36.908 - Lookup Date.0 - Finished processing (I=9800, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:36.913 - Select values 2.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

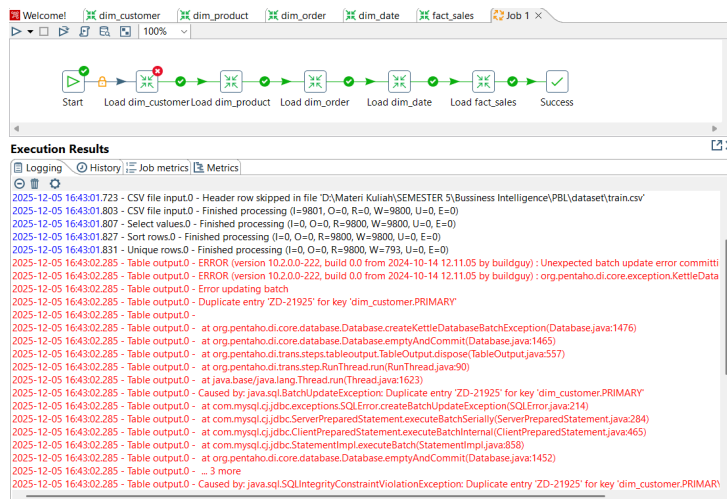
2025-12-05 09:46:37.021 - Table output.0 - Finished processing (I=0, O=0, R=9800, W=9800, U=0, E=0)

2025-12-05 09:46:37.021 - Spoon - The transformation has finished!!

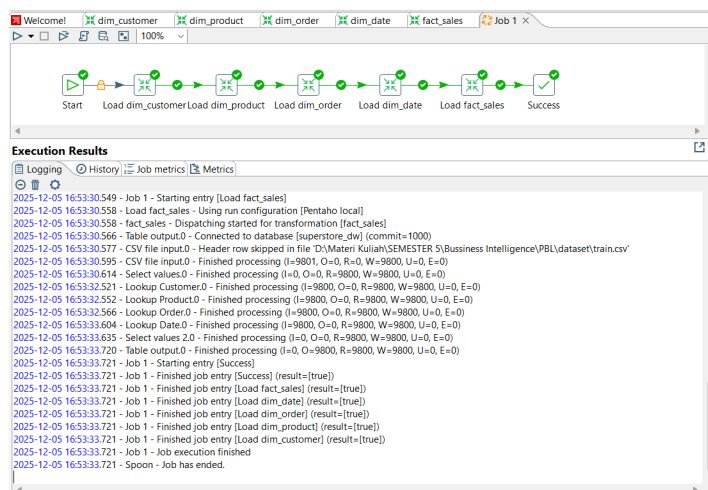
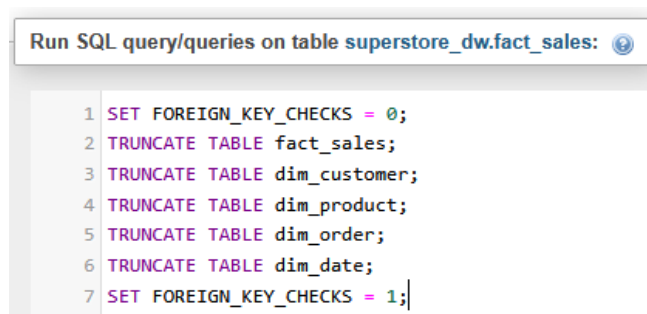
				fact_id	order_id	customer_id	product_id	date_id	sales
<input type="checkbox"/>	Edit	Copy	Delete	1	CA-2017-152156	CG-12520	FUR-BO-10001798	20171108	261.96
<input type="checkbox"/>	Edit	Copy	Delete	2	CA-2017-152156	CG-12520	FUR-CH-10000454	20171108	731.94
<input type="checkbox"/>	Edit	Copy	Delete	3	CA-2017-138688	DV-13045	OFF-LA-10000240	20170612	14.62
<input type="checkbox"/>	Edit	Copy	Delete	4	US-2016-108966	SO-20335	FUR-TA-10000577	20161011	957.58
<input type="checkbox"/>	Edit	Copy	Delete	5	US-2016-108966	SO-20335	OFF-ST-10000760	20161011	22.37
<input type="checkbox"/>	Edit	Copy	Delete	6	CA-2015-115812	BH-11710	FUR-FU-10001487	20150609	48.86
<input type="checkbox"/>	Edit	Copy	Delete	7	CA-2015-115812	BH-11710	OFF-AR-10002833	20150609	7.28
<input type="checkbox"/>	Edit	Copy	Delete	8	CA-2015-115812	BH-11710	TEC-PH-10002275	20150609	907.15

5.5 Membuat Job Untuk Menjalankan Semua Proses

Karena semua tabel telah terisi data, sehingga ketika kita menjalankan Job Master, maka otomatis akan error karena ada duplikasi data di MySQL



Maka dari itu, kita perlu lakukan truncate dulu di MySql untuk menghilangkan data, lalu jalankan Job Master untuk mengisi data kembali



Cek ulang di database superstore_dw, data otomatis akan terisi lagi, jadi fungsi dari Job Master yaitu sebagai orkestrator yang menjalankan semua proses ETL secara otomatis dan berurutan. Dengan Job Master, proses loading data warehouse dapat dilakukan dengan satu kali eksekusi, memastikan urutan yang benar (dimensi dahulu, fakta kemudian), serta memudahkan monitoring dan error handling.

BAB 6 - HASIL DAN ANALISIS

6.1 Query Testing

- Query 1: Total Penjualan Keseluruhan

```
1 SELECT
2     SUM(sales) AS total_sales,
3     COUNT(*) AS total_transactions
4 FROM fact_sales;
```

total_sales	total_transactions
2261536.97	9800

- Query 2: Top 5 Produk Terlaris (berdasarkan jumlah penjualan)

```
1 SELECT
2     p.product_name,
3     p.category,
4     COUNT(f.fact_id) AS total_transactions,
5     SUM(f.sales) AS total_sales
6 FROM fact_sales f
7 JOIN dim_product p ON f.product_id = p.product_id
8 GROUP BY p.product_id, p.product_name, p.category
9 ORDER BY total_sales DESC
10 LIMIT 5;
```

product_name	category	total_transactions	total_sales
Canon imageCLASS 2200 Advanced Copier	Technology	5	61599.83
Fellowes PB500 Electric Punch Plastic Comb Binding...	Office Supplies	10	27453.38
Cisco TelePresence System EX90 Videoconferencing U...	Technology	1	22638.48
HON 5400 Series Task Chairs for Big and Tall	Furniture	8	21870.57
GBC DocuBind TL300 Electric Binding System	Office Supplies	11	19823.48

- Query 3: Penjualan per Kategori Produk


```

1 SELECT
2     p.category,
3     COUNT(f.fact_id) AS total_transactions,
4     SUM(f.sales) AS total_sales,
5     AVG(f.sales) AS avg_sales
6 FROM fact_sales f
7 JOIN dim_product p ON f.product_id = p.product_id
8 GROUP BY p.category
9 ORDER BY total_sales DESC;

```

category	total_transactions	total_sales ▾ 1	avg_sales
Technology	1813	827455.94	456.401511
Furniture	2078	728658.75	350.653874
Office Supplies	5909	705422.28	119.380992

- Query 4: Penjualan per Segment Pelanggan

```

1 SELECT
2     c.segment,
3     COUNT(f.fact_id) AS total_transactions,
4     SUM(f.sales) AS total_sales
5 FROM fact_sales f
6 JOIN dim_customer c ON f.customer_id = c.customer_id
7 GROUP BY c.segment
8 ORDER BY total_sales DESC;

```

segment	total_transactions	total_sales ▾ 1
Consumer	5101	1148060.51
Corporate	2953	688494.14
Home Office	1746	424982.32

- Query 5: Penjualan per Region

```

1 SELECT
2     c.region,
3     COUNT(f.fact_id) AS total_transactions,
4     SUM(f.sales) AS total_sales
5 FROM fact_sales f
6 JOIN dim_customer c ON f.customer_id = c.customer_id
7 GROUP BY c.region
8 ORDER BY total_sales DESC;

```

region	total_transactions	total_sales ▾ 1
West	3160	744293.54
East	2709	606351.33
Central	2334	514251.40
South	1597	396640.70

- Query 6: Penjualan per Bulan

```

1 SELECT
2     d.year,
3     d.month,
4     d.month_name,
5     COUNT(f.fact_id) AS total_transactions,
6     SUM(f.sales) AS total_sales
7 FROM fact_sales f
8 JOIN dim_date d ON f.date_id = d.date_id
9 GROUP BY d.year, d.month, d.month_name
10 ORDER BY d.year, d.month;

```

year ▲ 1	month	month_name	total_transactions	total_sales
2015	1	1	77	14205.71
2015	2	2	46	4519.92
2015	3	3	154	55205.83
2015	4	4	130	27906.86
2015	5	5	121	23644.30
2015	6	6	131	34322.94
2015	7	7	142	33781.52
2015	8	8	146	37447.53

- Query 7: Penjualan per Quarter

```

1 SELECT
2     d.year,
3     d.quarter,
4     COUNT(f.fact_id) AS total_transactions,
5     SUM(f.sales) AS total_sales
6 FROM fact_sales f
7 JOIN dim_date d ON f.date_id = d.date_id
8 GROUP BY d.year, d.quarter
9 ORDER BY d.year, d.quarter;

```

year ▲ 1	quarter	total_transactions	total_sales
2015	1	277	73931.46
2015	2	382	85874.10
2015	3	555	142522.57
2015	4	739	177528.14
2016	1	249	62357.68
2016	2	431	87713.46
2016	3	570	129560.14

- Query 8: Top 5 Customer dengan Pembelian Terbanyak

```

1 SELECT
2     c.customer_name,
3     c.segment,
4     c.region,
5     COUNT(f.fact_id) AS total_transactions,
6     SUM(f.sales) AS total_sales
7 FROM fact_sales f
8 JOIN dim_customer c ON f.customer_id = c.customer_id
9 GROUP BY c.customer_id, c.customer_name, c.segment, c.region
10 ORDER BY total_sales DESC
11 LIMIT 5;

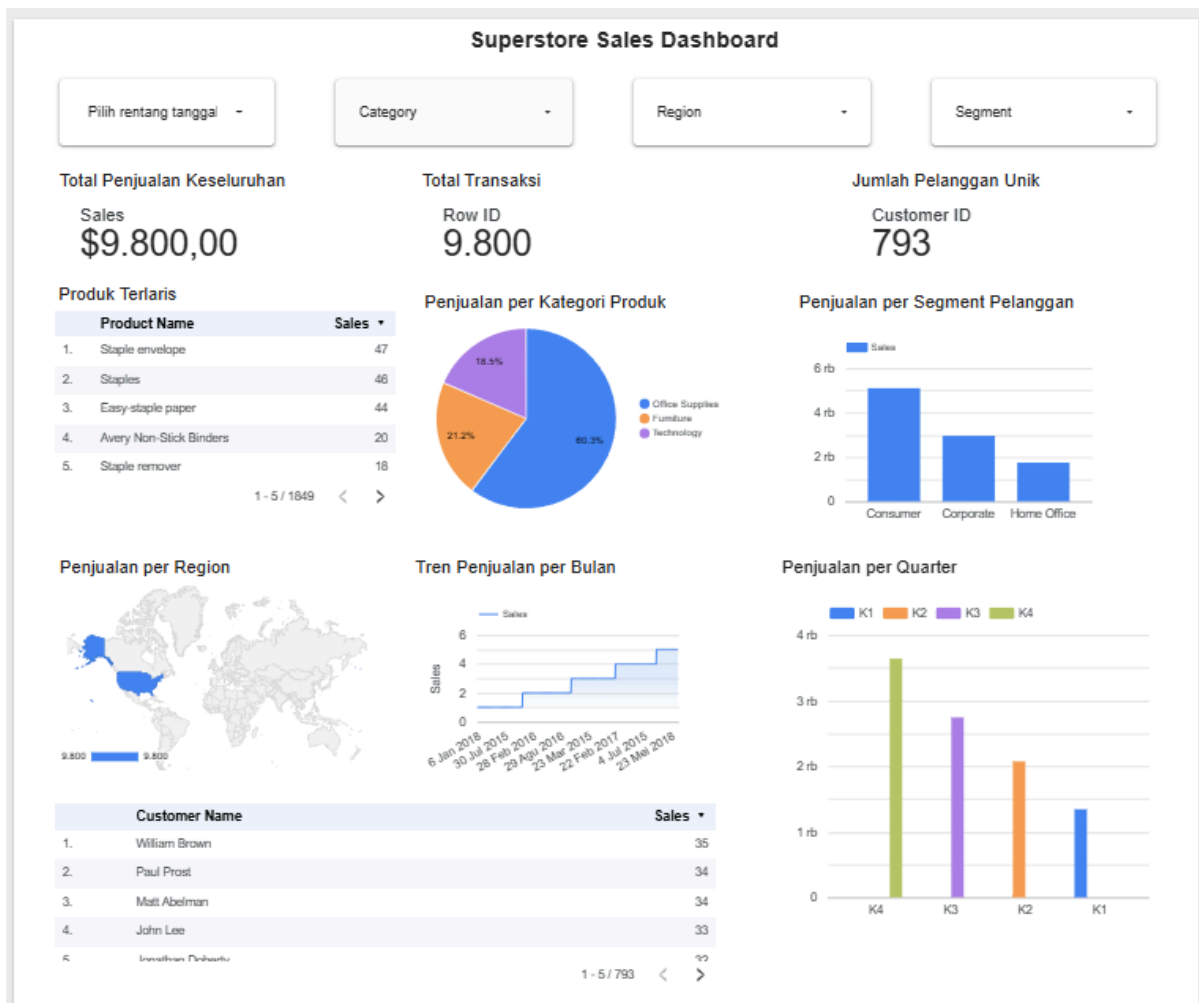
```

customer_name	segment	region	total_transactions	total_sales ▼ 1
Sean Miller	Home Office	South	15	25043.07
Tamara Chand	Corporate	West	12	19052.22
Raymond Buch	Consumer	East	18	15117.35
Tom Ashbrook	Home Office	East	10	14595.62
Adrian Barton	Consumer	West	20	14473.57

6.2 Analisis KPI dan Dashboard

Dashboard Looker Studio

link : <https://lookerstudio.google.com/reporting/b1ae5597-7a38-43a1-acb2-a391576d7597>



Analisis insight dari data:

- Kategori apa yang paling laris? Berdasarkan analisis data, kategori Office Supplies merupakan kategori yang paling laris dengan penjualan sebesar 60.3% dari total penjualan. Kategori ini diikuti oleh Furniture (21.25) dan Technology (18.3%).
- Region mana yang paling profitable? Berdasarkan analisis data menunjukkan region West memiliki penjualan tertinggi sebesar \$725,458 (31.6%), diikuti oleh EAST (\$678,781 - 29.5%).
- Segment customer mana yang paling banyak belanja? Berdasarkan analisis menunjukkan Consumer yang paling banyak belanja, dengan kontribusi 51.5% dari total penjualan.
- Trend penjualan naik/turun? Berdasarkan analisis tren penjualan menunjukkan kenaikan (terutama dari 22 Mei 2016 hingga 4 Juli 2018). Penjualan tertinggi secara periodik selalu terjadi pada Quarter 4 (K4), yang merupakan puncak penjualan tahunan.

BAB 7 - KESIMPULAN DAN SARAN

7.1 Kesimpulan

Berdasarkan hasil perancangan dan implementasi sistem Business Intelligence pada dataset *Superstore Sales*, dapat disimpulkan bahwa proses pembangunan data warehouse dengan model star schema mampu mengintegrasikan data transaksi penjualan secara terstruktur dan mudah dianalisis. Data yang sebelumnya berada dalam format mentah berhasil diolah melalui proses ETL (Extract, Transform, Load) sehingga menjadi data yang bersih, konsisten, dan siap digunakan untuk analisis.

Proses ETL yang dibangun menggunakan Pentaho Data Integration berhasil memisahkan data ke dalam tabel fakta dan tabel dimensi sesuai dengan kebutuhan analisis. Dengan adanya tabel fakta penjualan dan tabel-tabel dimensi seperti pelanggan, produk, pesanan, dan waktu, proses query analitis dapat dilakukan secara lebih cepat dan efisien.

Hasil analisis KPI yang divisualisasikan melalui Google Looker Studio menunjukkan bahwa kategori *Office Supplies* merupakan kategori dengan penjualan tertinggi, region *West* menjadi wilayah dengan kontribusi penjualan terbesar, serta segmen *Consumer* sebagai segmen pelanggan dengan nilai belanja paling tinggi. Selain itu, tren penjualan menunjukkan peningkatan dari waktu ke waktu dengan puncak penjualan yang konsisten terjadi pada kuartal keempat setiap tahunnya. Hal ini membuktikan bahwa sistem Business Intelligence yang dibangun mampu memberikan insight yang bernilai dalam mendukung pengambilan keputusan bisnis.

7.2 Saran

Berdasarkan proyek yang telah dikerjakan, terdapat beberapa saran untuk pengembangan lebih lanjut, antara lain:

1. Data warehouse dapat dikembangkan dengan menambahkan dimensi tambahan, seperti dimensi kategori atau dimensi supplier, agar analisis dapat dilakukan dengan sudut pandang yang lebih luas.
2. Proses ETL dapat ditingkatkan dengan menambahkan mekanisme incremental load, sehingga tidak perlu melakukan truncate data setiap kali proses ETL dijalankan.
3. Analisis KPI dapat diperluas dengan menambahkan metrik lain, seperti *customer lifetime value*, *profit margin per produk*, atau analisis prediksi penjualan untuk mendukung perencanaan bisnis jangka panjang.
4. Dashboard Business Intelligence dapat dikembangkan dengan fitur interaktif yang lebih kompleks, seperti filter dinamis dan drill-down analysis, agar pengguna dapat melakukan eksplorasi data secara lebih mendalam.
5. Pada pengembangan selanjutnya, sistem Business Intelligence dapat diintegrasikan dengan data real-time agar informasi yang dihasilkan selalu bersifat aktual dan relevan.