

Laporan Final Project

Business Intelligence



Disusun oleh:

Romy Mahardika Pangestu Lazuardi (2341720129)

TI_3A

**PROGRAM STUDI D4 TEKNIK INFORMATIKA
JURUSAN TEKNOLOGI INFORMASI
POLITEKNIK NEGERI MALANG
2025**

1. Studi Kasus Pengelolaan Data (Data Penjualan – Northwind)

Studi kasus yang digunakan adalah **pengelolaan data penjualan pada perusahaan distribusi produk** menggunakan **Northwind Dataset** (data dummy publik).

<https://www.kaggle.com/datasets/emmanueltugbeh/northwind-orders-and-order-details>

Sumber data dummy:

- Kaggle (Northwind Dataset – open source & legal)
- Data tersedia dalam format CSV

Tabel yang digunakan (≥ 4 tabel):

1. **customers**
customerID, companyName, contactName, contactTitle, city, country
2. **products**
productID, productName, quantityPerUnit, unitPrice, discontinued, categoryID
3. **orders**
orderID, customerID, employeeID, orderDate, requiredDate, shippedDate, shipperID, freight
4. **order_details**
orderID, productID, unitPrice, quantity, discount
5. **employees**
employeeID, employeeName, title, city, country, reportsTo
6. **categories**
categoryID, categoryName, description
7. **shippers**
shipperID, companyName

2. Studi Kasus Terpilih

Studi Kasus:

Analisis Penjualan dan Kinerja Produk pada Perusahaan Distribusi

Alasan pemilihan:

- Data lengkap (penjualan, pelanggan, produk, karyawan)
- Mudah diakses dan legal
- Cocok untuk perhitungan KPI (penjualan, produk terlaris, performa karyawan)
- Relasi data jelas untuk skema bintang

3. Pengambilan Data Dummy

Langkah:

1. Download dataset Northwind (CSV/SQL)
2. Identifikasi format data:
 - o CSV: customers.csv, products.csv, orders.csv, order_details.csv, dll
3. Validasi atribut:
 - o Tidak kosong
 - o Tipe data sesuai (tanggal, numerik, string)

4. Rancangan Skema Bintang (Star Schema)

Tabel Fakta

Fact_Penjualan

- order_id
- product_id
- customer_id
- employee_id
- shipper_id
- time_id
- quantity
- unit_price
- discount
- total_penjualan

Tabel Dimensi

1. Dim_Pelanggan

- customer_id
- company_name
- contact_name
- city
- country

2. Dim_Produk

- product_id
- product_name
- category_name
- unit_price
- discontinued

3. Dim_Karyawan

- employee_id
- employee_name
- title
- city
- country

4. Dim_Pengirim

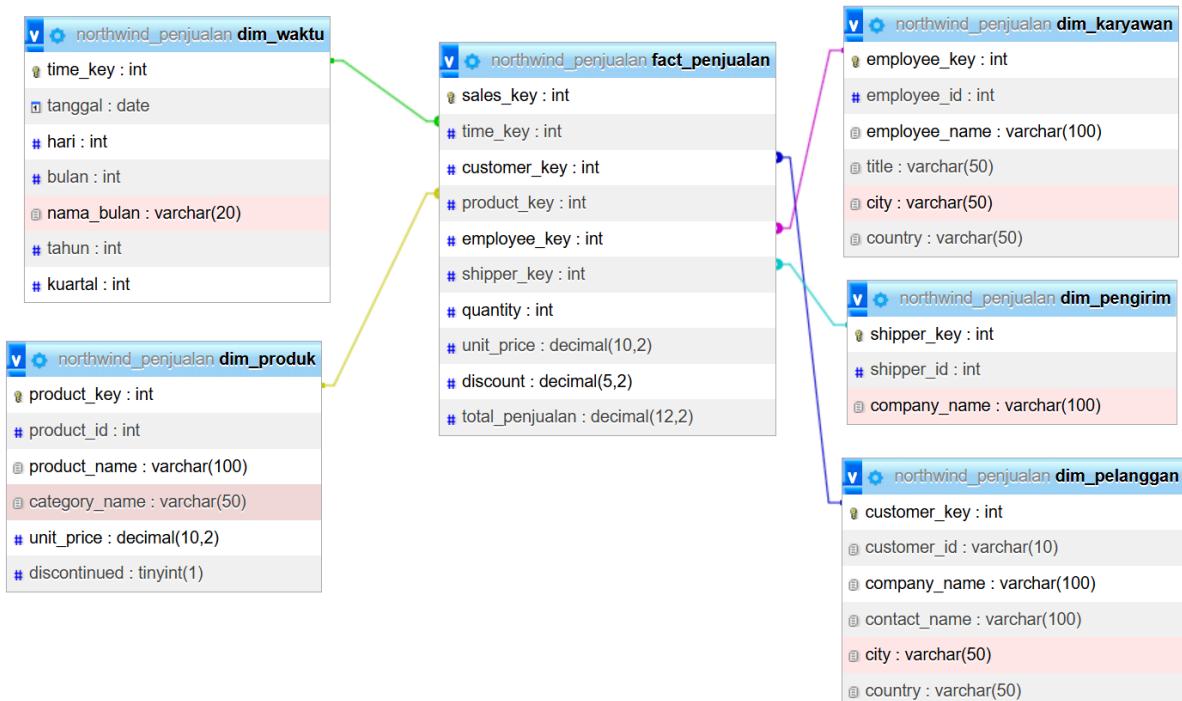
- shipper_id
- company_name

5. Dim_Waktu

- time_id
- tanggal
- bulan
- tahun
- kuartal

Relasi

- Fact_Penjualan memiliki **foreign key** ke seluruh tabel dimensi
- Relasi berbentuk **one-to-many**



Rancangan Proyek ETL

Extract

Data diekstrak dari file CSV Northwind menggunakan Pentaho dan dimuat ke staging area.

- Mengambil data dari:
 - File CSV (customers, orders, products, dll)
- Tools:
 - Pentaho Data Integration (Kettle)

Transform

Dilakukan proses:

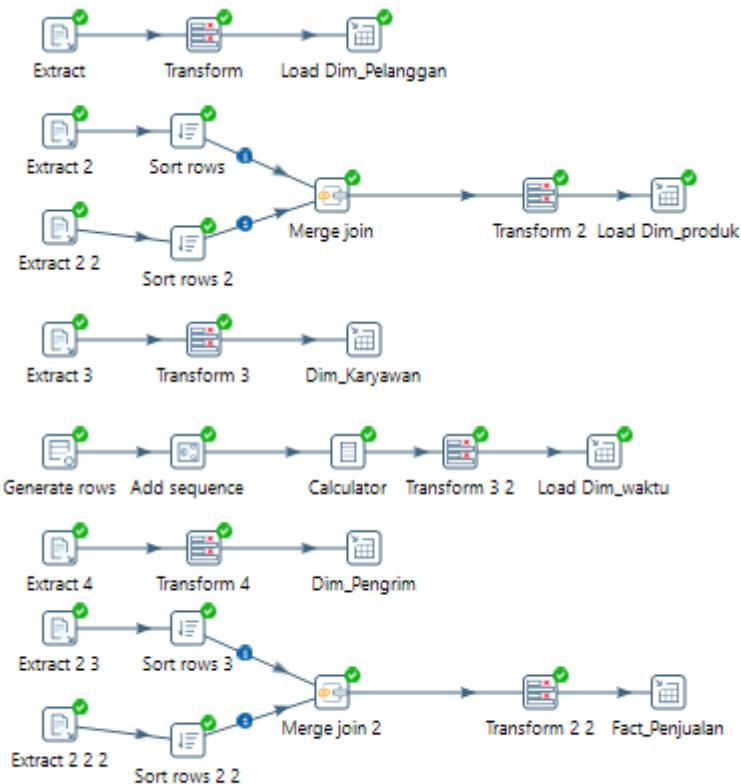
- Normalisasi data
- Penyesuaian struktur agar sesuai skema bintang
- Perhitungan nilai penjualan
- Validasi data kosong atau tidak konsisten
- Cleaning data:
 - Menghapus duplikat
 - Standarisasi nama negara & kota
 - Konversi format tanggal
- Transformasi:
 - Hitung `total_penjualan = quantity × unit_price × (1 - discount)`
 - Join tabel orders + order_details
 - Mapping kategori produk

Load

Data yang telah bersih dimasukkan ke data warehouse:

- Dimensi → Fakta
- Relasi dijaga dengan foreign key
 - Load ke:
 - Tabel dimensi terlebih dahulu
 - Dilanjutkan tabel fakta

- Perhatikan:
 - Primary key
 - Foreign key
 - Integritas data



Contoh Kasus

Total Penjualan

```

    ✓ Showing rows 0 - 0 (1 total, Query took 0.0044 seconds.)
    SELECT SUM(total_penjualan) AS total_penjualan FROM fact_penjualan;
  
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Show all | Number of rows: 25 ▾ Filter rows: Search this table

Extra options

total_penjualan

207306.10

Total Penjualan per Tahun

Showing rows 0 - 0 (1 total, Query took 0.0030 seconds.)

```
SELECT w.tahun, SUM(f.total_penjualan) AS total_penjualan FROM fact_penjualan f JOIN dim_waktu w ON f.time_key = w.time_key GROUP BY w.tahun;
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Show all | Number of rows: 25 Filter rows: Search this table

Extra options

tahun	total_penjualan
2018	8365.20

Top 5 Produk Terlaris

Showing rows 0 - 4 (5 total, Query took 0.0025 seconds.)

```
SELECT p.product_name, SUM(f.quantity) AS total_terjual FROM fact_penjualan f JOIN dim_produk p ON f.product_key = p.product_key GROUP BY p.product_name ORDER BY total_terjual DESC LIMIT 5;
```

Profiling [Edit inline] [Edit] [Explain SQL] [Create PHP code] [Refresh]

Extra options

product_name	total_terjual
Gula Malacca	20
Steeleye Stout	20
Chef Anton's Gumbo Mix	18
Louisiana Hot Spiced Okra	17
Aniseed Syrup	16

