

Nama : Steven Yap
NIM : 2540125371

1. Kasus: Merancang Data Warehouse untuk Bank Sentral Mandiri Digital (BSMD)

1a. Buatlah list of requirements (business dan technical) yang bisa menjadi dasar untuk rancangan technical architecture.

Business Requirements (Kebutuhan Bisnis):

- **Data center (Pusat data)** = semua unit (risk, kredit, CX, BI, compliance) mengakses data konsisten dari satu repositori.
- **Integrasi data multi-sumber** = menggabungkan data dari CBS, LOS, Digital Banking, KYC/Onboarding, data eksternal OJK/BI, pasar modal dan terintegrasi satu sama lain.
- **Mendukung analisis lanjutan** = risk scoring, fraud detection, segmentasi nasabah, campaign analysis.
- **Self-Service BI** = user non-teknis (manajer, analis) dapat membuat laporan/grafik tanpa harus selalu bergantung pada tim IT.
- **Pelaporan otomatis** = otomatisasi laporan regulasi untuk OJK/BI sesuai jadwal.
- **Visualisasi dashboard** = untuk monitoring transaksi, performa produk, serta deteksi anomali cepat.
- **Mendukung keputusan strategis** = laporan profitabilitas, pertumbuhan nasabah, efektivitas produk digital.

Adapun bentuk yang terorganisir yang bisa diterapkan untuk mengintegrasikan kebutuhan tersebut yakni pembuatan bus matrix, sebagai berikut:

https://docs.google.com/spreadsheets/d/1du9_7XB14Sjg_1BAr-3VHs85o416WwAlZyiwrvjRzgs/edit?usp=sharing

Business Process	Atomic Granularity	Metrics	Nasabah	Produk (Tabungan, Pinj.)	Waktu (Tangal, Bulan)	Transaksi (CBS/LOS/DBP)	Lokasi (Geografi)	Channel (Mobile/Web/ATM)	Skor Kredit (BI Check)	Kampanye Produk
Risk Management	1 baris = 1 transaksi / account exposure per nasabah per hari	- Probability of Default (PD) - Loss Given Default (LGD)	X	X	X	X	X	X	X	

		- Non-Performing Loan (NPL) ratio - Exposure at Default (EAD)								
Kredit (Loan)	1 baris = 1 pengajuan pinjaman atau 1 cicilan pembayaran	- Jumlah pengajuan pinjaman - Jumlah pinjaman disetujui - Outstanding loan balance - Cicilan tertunggak	X	X	X	X	X	X	X	
Customer Experience	1 baris = 1 event aktivitas digital (login, transfer, top-up, bill payment)	- Jumlah login harian - Jumlah transaksi digital - Retention rate - ARPU (Average Revenue per User)	X	X	X	X	X	X		X
BI & Marketing	1 baris = 1 interaksi nasabah dengan kampanye (email, push notif, klik iklan, pembelian)	- Jumlah kampanye berjalan - Conversion rate (%) - ROI kampanye - CPA (Cost per Acquisition)	X	X	X	X	X	X		X
Compliance (OJK/BI)	1 baris = 1 transaksi yang masuk kategori regulasi (misalnya ≥ Rp500 juta atau mencurigakan)	- Jumlah transaksi di atas threshold - Jumlah	X	X	X	X	X		X	

		suspicious transaction flagged - SLA kepatuhan laporan (on-time vs late)								
--	--	--	--	--	--	--	--	--	--	--

Technical Requirements (Kebutuhan Teknis)

- **Data Integration** = koneksi ke berbagai sistem (API, batch file, streaming).
- **Staging Area** = untuk menampung data mentah sebelum diproses lebih lanjut.
- **Data Warehouse Centralized** = penyimpanan historis dengan struktur dimensional (star schema).
- **Data Mart** = untuk kebutuhan spesifik unit (risk, kredit, marketing).
- **ETL/ELT Pipeline** = proses otomatis untuk ekstrak, transformasi, dan load data.
- **Metadata Repository** = mendokumentasikan definisi data, lineage, kualitas data.
- **High Availability & Scalability** = mendukung jutaan transaksi harian dari 15 juta nasabah.
- **Security & Compliance** = enkripsi data, role-based access, audit trail, sesuai regulasi OJK/BI.
- **Real-Time & Batch Processing** = untuk mendukung fraud detection (real-time) dan pelaporan historis (batch).
- **BI Tools Integration** = kompatibel dengan Power BI, Tableau, Pentaho BI, atau sejenisnya, untuk merepresentasikan data ke dalam bentuk grafik visual.

1b. Rancang high-level technical architecture untuk solusi Data Warehouse dan Business

Intelligence BSMD dan sertakan komponen:

i. Sumber data internal dan eksternal

Sumber Data Internal

Core Banking System (CBS), dengan contoh data seperti rekening, saldo harian, mutasi tabungan/deposito, fee.

Loan Origination System (LOS), dengan contoh data seperti Aplikasi pinjaman, keputusan, limit, jadwal & pembayaran cicilan, status tunggakan.

Digital Banking Platform, dengan contoh data seperti event aplikasi (login, session), transaksi transfer/top-up/bill, device & channel.

KYC & Onboarding, dengan contoh data seperti Identitas, verifikasi liveness/face match, risk rating, dokumen.

Fraud Monitoring (internal), dengan contoh data seperti alert, rule hit, skor risiko transaksi.

Contact Center/CRM, dengan contoh data seperti ticket, kategori keluhan, SLA, outcome

Sumber Data Eksternal

BI Checking / SLIK, untuk data seperti skor kredit, riwayat kredit eksternal

OJK/BI Feeds, untuk data seperti kode referensi, ketentuan pelaporan, validasi schema

Jaringan Pembayaran (switch/payment gateway), untuk data seperti status settlement, fee, chargeback

ii. Area staging

Proses standarisasi data, terdiri dari:

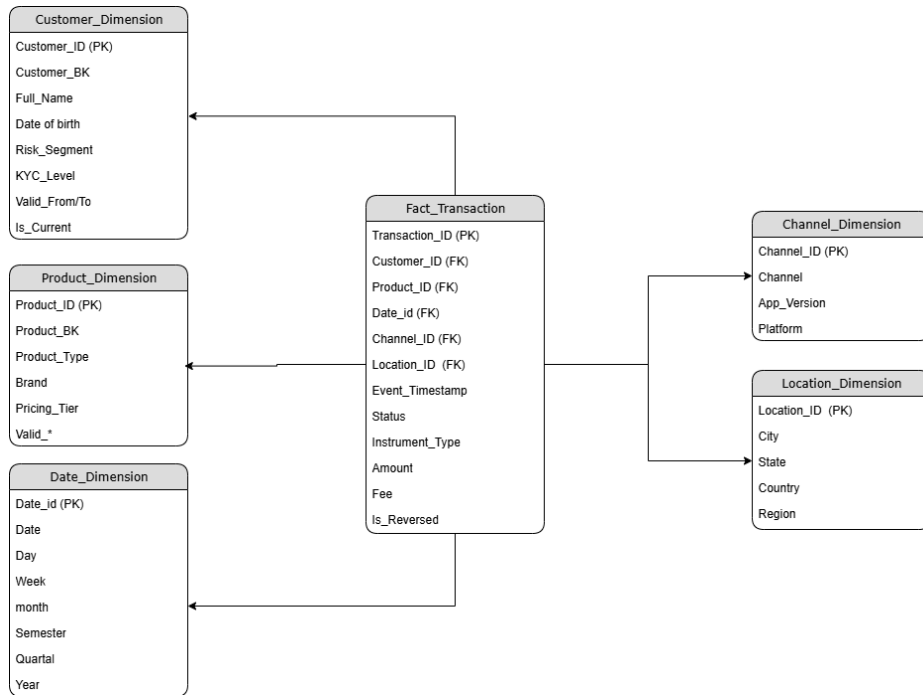
- **Landing zone**, proses load data awal, untuk mengecek fitur apa saja yang terdapat dalam data dan apakah ada sesuatu yang ganjil.
- **Raw / Bronze**, Proses pengelompokan fitur berdasarkan dimensi yang bersangkutan
- **Cleansed / Silver**, Proses pemberian ID, pengecekan setiap tipe fitur atau variabel yang dipake apakah integer, date atau string atau yang lainnya.
- **Curated / Gold**, Persiapan table per domain/dimension until dilanjutkan ke str schema

iii. Data warehouse dan data mart (jika ada)

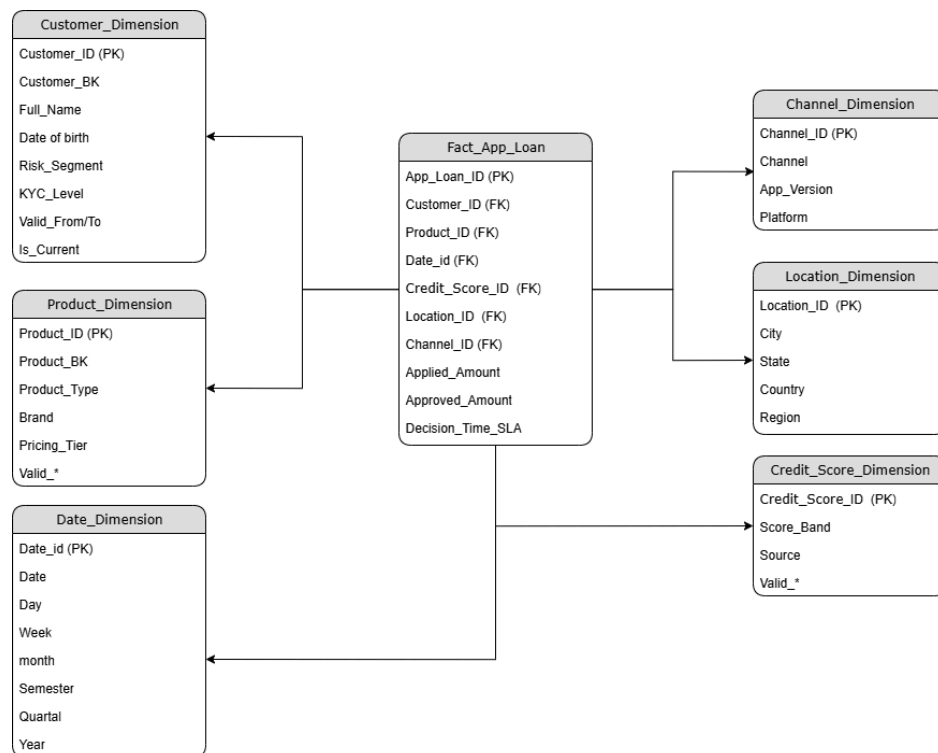
Untuk data warehouse akan direpresentasikan dalam star schem dengan pembagian fact sebagai berikut:

<https://drive.google.com/file/d/1KsSmfwvDYYDnBlpXtnM5tiwxpvEAXhfb/view?usp=sharing>

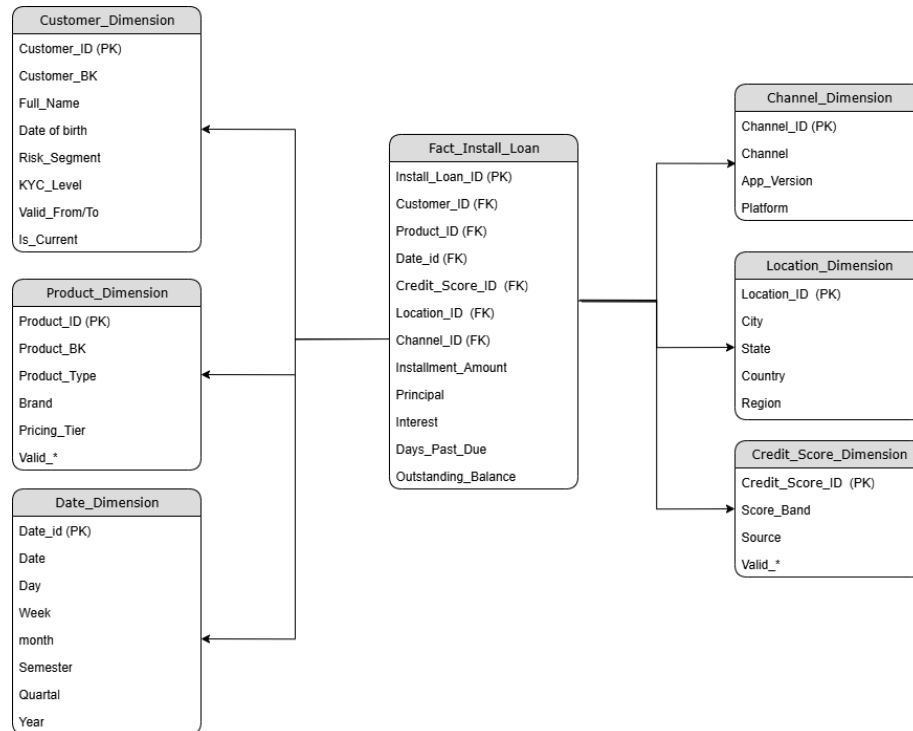
- **Fact_Transaction**



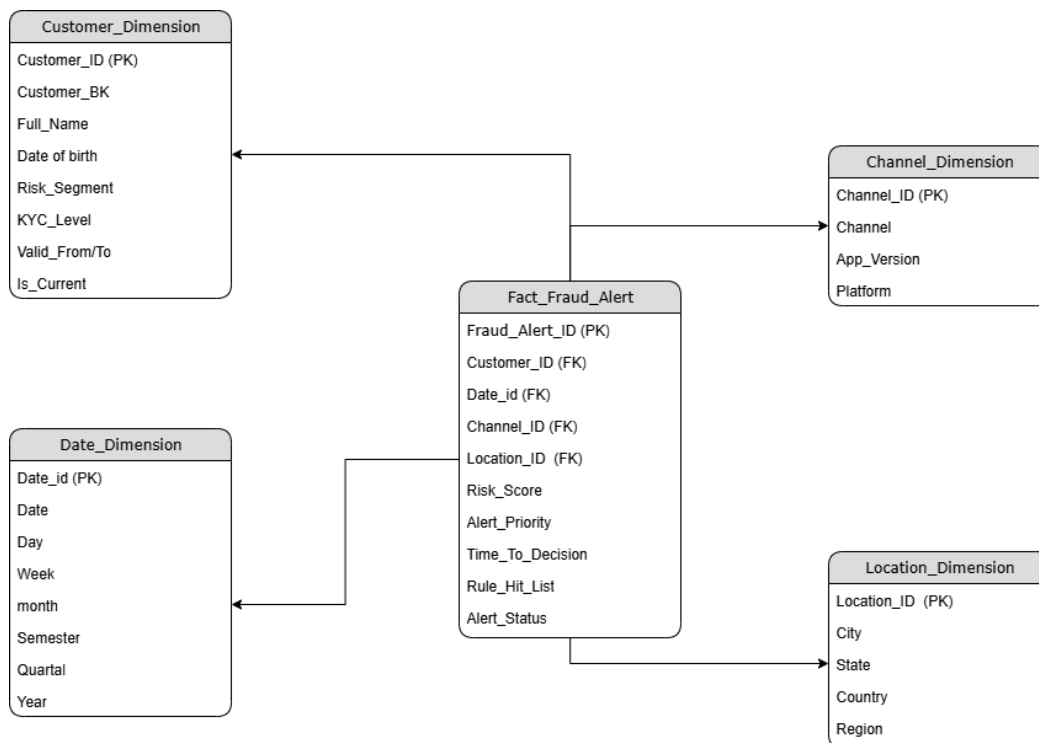
- **Fact_Loan** yang terbagi menjadi 2 fact table untuk:
 - **Application:** keputusan, limit



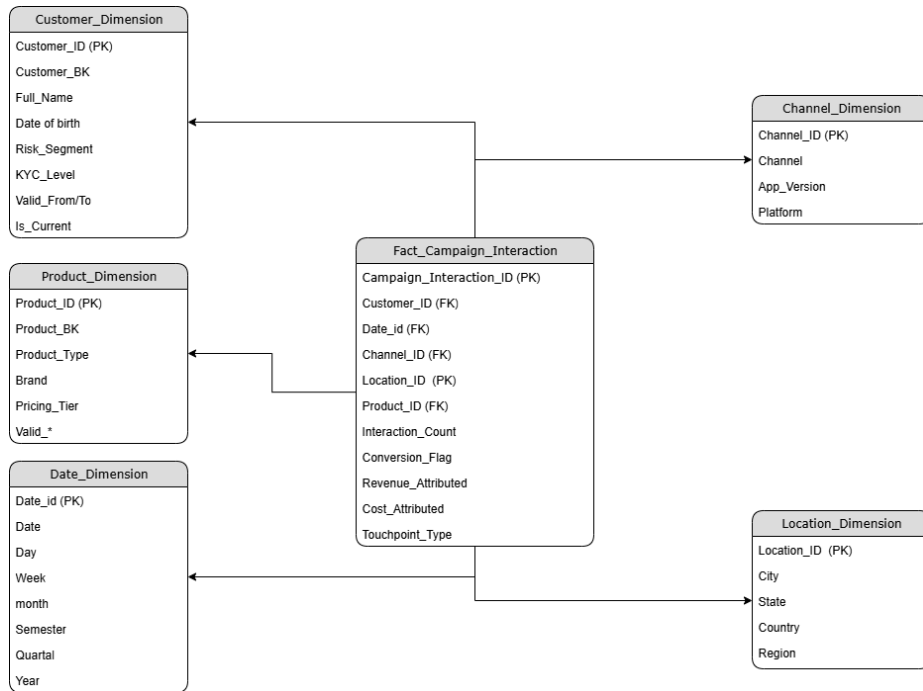
- **Installment:** pembayaran cicilan



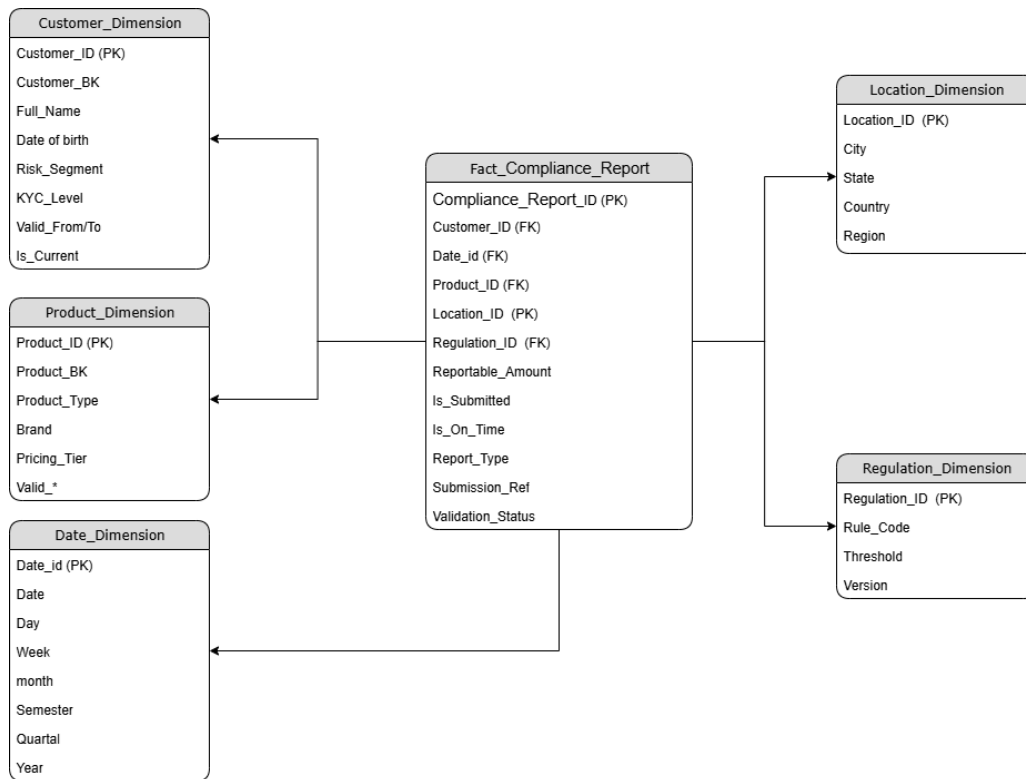
- **Fact_Fraud_Alert**



- **Fact_Campaign_Interaction**



- **Fact_Compliance_Report**



Data Mart, dimana beberapa fact tersebut dapat mendefinisikan proses bisnis sebagai berikut:

- Risk Management, dapat dianalisis berdasarkan Fact_Transaction, Fact_Loan(Installment), Fact_Fraud_Alert
- Loan, dapat dilihat dari Fact_Loan_Application, Fact_Loan_Installment
- Customer Experience, yang didapatkan dari Fact_Transaction
- BI & Marketing, yang direpresentasikan oleh Fact_Campaign_Interaction
- Compliance Mart, dengan melihat Fact_Compliance_Report

iv. Proses ETL/ELT

ETL terdiri dari 3 langkah Extract Transform dan Load:

Extract, proses pengambilan data, terdiri dari:

- **Sumber internal:** Core Banking System (CBS), Loan Origination System (LOS), CRM, Digital Banking apps.
- **Sumber eksternal:** BI Checking, OJK Reporting, 3rd party KYC/fraud detection.
- **Jenis data:** transaksi, profil nasabah, status pinjaman, fraud alerts, log aktivitas, dsb.
- **Tools** (contoh Pentaho Data Integration): menggunakan Table Input untuk load data sumber(csv), Database Connection (untuk target data yg dicocokkan berdasarkan dimension yang telah dibentuk), File Input.

Transform, Proses pengolahan data, terdiri dari:

- **Cleansing:** hapus duplikat, isi missing values, validasi format (tipe data, banyaknya digit).
- **Standardisasi:** format tanggal, currency, kode produk → seragam.
- **Business rules:**
 - Hitung Risk Score = function(loan history, fraud flags, income).
 - Tentukan kategori produk (Retail Loan, Corporate Loan, KPR).
 - Map data transaksi = fact table (misal, jumlah pinjaman masuk ke FactLoan).
- **Surrogate key:** buat key di dimension tables (misal Customer_ID, Product_ID).
- **SCD** (Slowly Changing Dimension): simpan histori perubahan, contoh perubahan alamat customer.

Load, Proses memasukkan value atau isi dari data ke variabelnya masing masing secara bertahap

1. Value dimasukkan sesuai dengan variabel per dimensi
2. Lanjut dimasukkan ke dalam Fact yang bersangkutan dan membutuhkan dimensi tersebut

v. Metadata repository

Tahap ini akan menjadi tahap yang menyimpan log atau dokumentasi beberapa tahap sebelumnya seperti bagaimana suatu variabel dapat menghasilkan atau memiliki value tersebut

formula yang digunakan serta bagaimana valuenya diolah ditransformasi dan tipe data yang dimiliki.

Contoh metadata berdasarkan jenisnya adalah:

1. Technical Metadata:
 - Struktur database: nama tabel, kolom, tipe data, indeks.
 - Lineage data: sumber → staging → DW → mart → dashboard.
 - ETL rules: transformasi, mapping, business rules.
 - Jadwal proses: job ETL harian, mingguan, bulanan.
2. Business Metadata
 - Definisi metrik bisnis (contoh: Loan Default Rate = jumlah pinjaman macet ÷ total pinjaman).
 - Hierarki dimensi: (Nasabah → Segmen → Region → Nasional).
 - KPI: NPL, Risk Score, Fraud Rate, Loan Growth, Customer Churn Rate.
 - Owner data: siapa PIC (misalnya Risk Mgmt Dept untuk FraudAlert).
3. Operational Metadata
 - Status job ETL (success/failure, timestamp, error log).
 - Volume data tiap batch.
 - History perubahan (SCD – Slowly Changing Dimension).

vi. BI tools / dashboard layer

Komponen BI Layer di BSMD

Reporting:

- Laporan rutin ke OJK/BI (compliance).
- Laporan operasional: pinjaman harian, risk scoring, fraud detection.
- Ad-hoc reporting untuk manajemen.

Dashboard

- Visualisasi KPI utama → Loan Growth, NPL, Fraud Alerts, Customer Satisfaction.
- Interaktif: filter per region, segmen nasabah, produk.
- Drill-down: dari level nasional → region → cabang → customer.

Analytics

- Data mining & prediktif: credit scoring, churn prediction.
- Trend analysis: loan disbursement per bulan, seasonal pattern.
- What-if simulation: dampak kenaikan suku bunga terhadap NPL.

Tools, terdapat beberapa tools yang bisa digunakan seperti Power BI, Tableau, Qlik Sense

Dashboard Layer, pemilihan grafik perlu disesuaikan dengan tujuan, variabel yang bersangkutan dan interpretasi suatu data. Contoh grafik berdasarkan business prosesnya adalah:

1. Risk Management Dashboard

- **Header**

- a. Judul: "Risk Management Dashboard"
- b. Filter Global:
 - i. Periode: [Dropdown Bulanan / Tahunan]
 - ii. Region: [Nasional / Regional / Cabang]
 - iii. Produk Kredit: [KPR, Kredit UMKM, Kredit Konsumtif, dll]

- **Ringkasan KPI (Top Cards)**

Ditampilkan dalam bentuk **4–5 kartu indikator** dengan warna hijau/kuning/merah:

- Total Loan Portfolio (Rp)
- Non-Performing Loan (NPL) Ratio (%)
- Fraud Alerts (Active Cases) (jumlah)
- Avg. Loan Approval Time (hari)
- High Risk Customers (%)

- **Grafik Utama**

- 1. **Trend NPL per Bulan**

- Line chart: menampilkan tren NPL 12 bulan terakhir.
 - Bisa dibandingkan target OJK.

- 2. **Fraud Alerts Distribution**

- Bar chart: fraud case berdasarkan region/cabang.
 - Bisa drill-down ke level customer.

- 3. **Loan Risk Heatmap**

- Heatmap / matrix: Risiko vs Produk Kredit.
 - Warna merah → risiko tinggi, hijau → rendah.

- **Detail Analitik**

- **Tabel Interaktif Fraud Cases**

- Kolom: ID Nasabah, Cabang, Jenis Fraud, Status Investigasi..

- **Risk Scoring Segmentation**

- Pie chart: kategori customer (Low, Medium, High Risk).
 - Bisa klik untuk filter.

2. Loan / Kredit Dashboard

- **Header**

- a. **Judul:** "Loan Performance Dashboard"
- b. **Filter Global:**
 - i. Periode: [Dropdown Bulanan / Tahunan]

- ii. Region: [Nasional / Regional / Cabang]
- iii. Produk Kredit: [KPR, Kredit UMKM, Kredit Konsumtif, dll]
- iv. Status Kredit: [Disetujui, Ditolak, Dalam Proses]

- **Ringkasan KPI (Top Cards)**

Ditampilkan dalam bentuk kartu indikator:

- Total Outstanding Loan (Rp)
- Jumlah Aplikasi Kredit (dalam periode terpilih)
- Approval Rate (%)
- Rata-rata Waktu Proses (Hari)
- Jumlah Kredit Bermasalah (NPL Cases)

- **Grafik Utama**

1. **Loan Disbursement Trend**

- Line chart: menampilkan tren penyaluran kredit per bulan.
- Bisa filter per produk kredit (KPR, UMKM, dll).

2. **Approval vs Rejection Rate**

- Stacked bar chart: perbandingan jumlah kredit disetujui vs ditolak.

3. **Loan by Product Type**

- Donut / Pie chart: distribusi portfolio berdasarkan jenis produk kredit.

- **Detail Analitik**

- **Loan Pipeline Funnel**

- Funnel chart: Tahapan aplikasi (Pengajuan → Analisa → Persetujuan → Pencairan).
- Bisa menunjukkan bottleneck di proses approval.

- **Regional Loan Distribution**

- Map atau bar chart: jumlah dan nilai kredit per region/cabang.

- **Top 10 Borrowers**

- Tabel interaktif berisi nasabah dengan nilai pinjaman terbesar.

3. Customer Experience Dashboard

- **Header**

- a. **Judul:** *"Customer Experience & Engagement Dashboard"*

- b. **Filter Global:**

- i. Periode: [Bulanan / Kuartal / Tahunan]
- ii. Segmen Nasabah: [Retail, UMKM, Korporasi]
- iii. Channel Layanan: [Mobile Banking, Internet Banking, Cabang, Call Center]

- **Ringkasan KPI (Top Cards)**

KPI utama untuk memonitor pengalaman nasabah:

- Customer Satisfaction Score (CSAT)
- Net Promoter Score (NPS)
- First Call Resolution (%)
- Average Handling Time (menit)
- Digital Adoption Rate (%)
- **Grafik Utama**
 - **Customer Satisfaction Trend**
 - Line chart: tren skor kepuasan (CSAT/NPS) per bulan.
 - **Channel Usage Distribution**
 - Donut chart: porsi transaksi via mobile banking, internet banking, cabang, call center.
 - **Complaint Volume & Resolution Rate**
 - Stacked bar chart: jumlah keluhan vs jumlah keluhan terselesaikan.
- **Sentiment & Feedback Analysis**
 - Word cloud atau bar chart dari hasil survei feedback nasabah.

4. BI & Marketing Dashboard

- **Header**
 - a. **Judul:** *"Business Intelligence & Marketing Dashboard"*
 - b. **Filter Global:**
 - i. Periode: [Bulanan / Kuartal / Tahunan]
 - ii. Segmen Nasabah: [Retail, UMKM, Korporasi]
 - iii. Produk: [Tabungan, Kredit, Kartu Kredit, Investasi, Asuransi]
 - iv. Channel Campaign: [Email, SMS, Mobile App, Branch, Call Center]

- **Ringkasan KPI (Top Cards)**

KPI utama yang dipantau:

- Total Campaigns Active
- Conversion Rate (%)
- Customer Reach (%)
- Campaign ROI (%)
- Cross-sell / Up-sell Rate
- **Grafik Utama**
 1. **Campaign Performance Over Time**
 - Line chart: tren jumlah campaign & conversion rate per bulan.
 2. **Channel Effectiveness**
 - Bar chart: perbandingan konversi per channel (email, SMS, mobile app, dll).
 3. **Customer Segmentation**
 - Donut chart: distribusi pelanggan berdasarkan segmen (retail, UMKM, korporasi).

- **Analitik Tambahan**
 - **Top Performing Campaigns**
 - Tabel: daftar campaign dengan conversion & ROI tertinggi.
 - **Product Sales by Campaign**
 - Heatmap: produk vs campaign.
 - **Customer Lifetime Value (CLV)**
 - KPI + tren: rata-rata revenue per customer dari campaign marketing.
- **Predictive & Targeting Insights**
 - **Lookalike Audience Prediction**
 - Bar chart: estimasi jumlah nasabah potensial berdasarkan model prediksi.
 - **Churn Risk vs Retention Campaign**
 - Scatter plot: nasabah berisiko churn dan hasil intervensi campaign retensi.

5. Compliance Dashboard

- **Header**
 - **Judul: "Regulatory Compliance Dashboard"**
 - **Filter Global:**
 - Periode: [Bulanan / Kuartal / Tahunan]
 - Jenis Regulasi: [OJK, BI, AML, Basel III]
 - Unit Bisnis: [Retail, Kredit, Treasury, Operasional]
 - Risiko: [Likuiditas, Kredit, Operasional, Pasar]
- **Ringkasan KPI (Top Cards)**
 - Regulatory Reporting Status
 - Compliance Breach Count
 - Average Reporting Delay (days)
 - Capital Adequacy Ratio (CAR %)
 - Liquidity Coverage Ratio (LCR %).
- **Grafik Utama**
 1. **Regulatory Breach Trend**
 - Line chart: tren jumlah pelanggaran per bulan.
 2. **Compliance by Business Unit**
 - Bar chart: jumlah pelanggaran per unit bisnis (Retail, Kredit, Treasury).
 3. **Capital & Liquidity Ratios**
 - Dual axis line chart: CAR & LCR dibanding threshold regulasi.
- **Analitik Tambahan**
 - **Late Reports to OJK/BI**
 - Tabel: daftar laporan yang telat, dengan status (pending, submitted, overdue).
 - **AML & Fraud Alerts**
 - Pie chart: distribusi suspicious transaction reports (STR) per kategori.
 - **Stress Test Results**
 - Heatmap: dampak simulasi risiko terhadap CAR & LCR.

vii. Security and compliance layer

Tahapan ini merupakan tahapan pengaturan keamanan, seperti:

1. Authentication & Authorization

- **Single Sign-On (SSO)**: integrasi dengan Active Directory / LDAP supaya user cukup sekali login.
- **Role-Based Access Control (RBAC)**: akses dashboard & data dibatasi berdasarkan role (misalnya: *Risk Analyst*, *Compliance Officer*, *Branch Manager*).
- **Multi-Factor Authentication (MFA)** untuk akses ke data sensitif.

2. Data Security

- **Data Encryption**:
 - *At rest*: data warehouse & data mart disimpan dengan enkripsi AES-256.
 - *In transit*: komunikasi antar layer (ETL, DWH, BI tools) via SSL/TLS.
- **Data Masking / Anonymization**: informasi pribadi (mis. KTP, nomor rekening) dimasking sesuai regulasi OJK/BI.
- **Row-Level Security (RLS)**: misalnya cabang hanya bisa lihat data customer di cabangnya sendiri.

3. Compliance Controls

- **Audit Trail & Logging**: semua query, akses data, export, atau perubahan konfigurasi tercatat = untuk kepatuhan OJK/BI.
- **Data Retention Policy**: pengaturan untuk menentukan berapa lama data transaksi disimpan (mis. 5–10 tahun).
- **Regulatory Alignment**: Penyesuaian beberapa ketentuan bisnis sesuai dengan standar yang diperlukan dari beberapa pihak luar

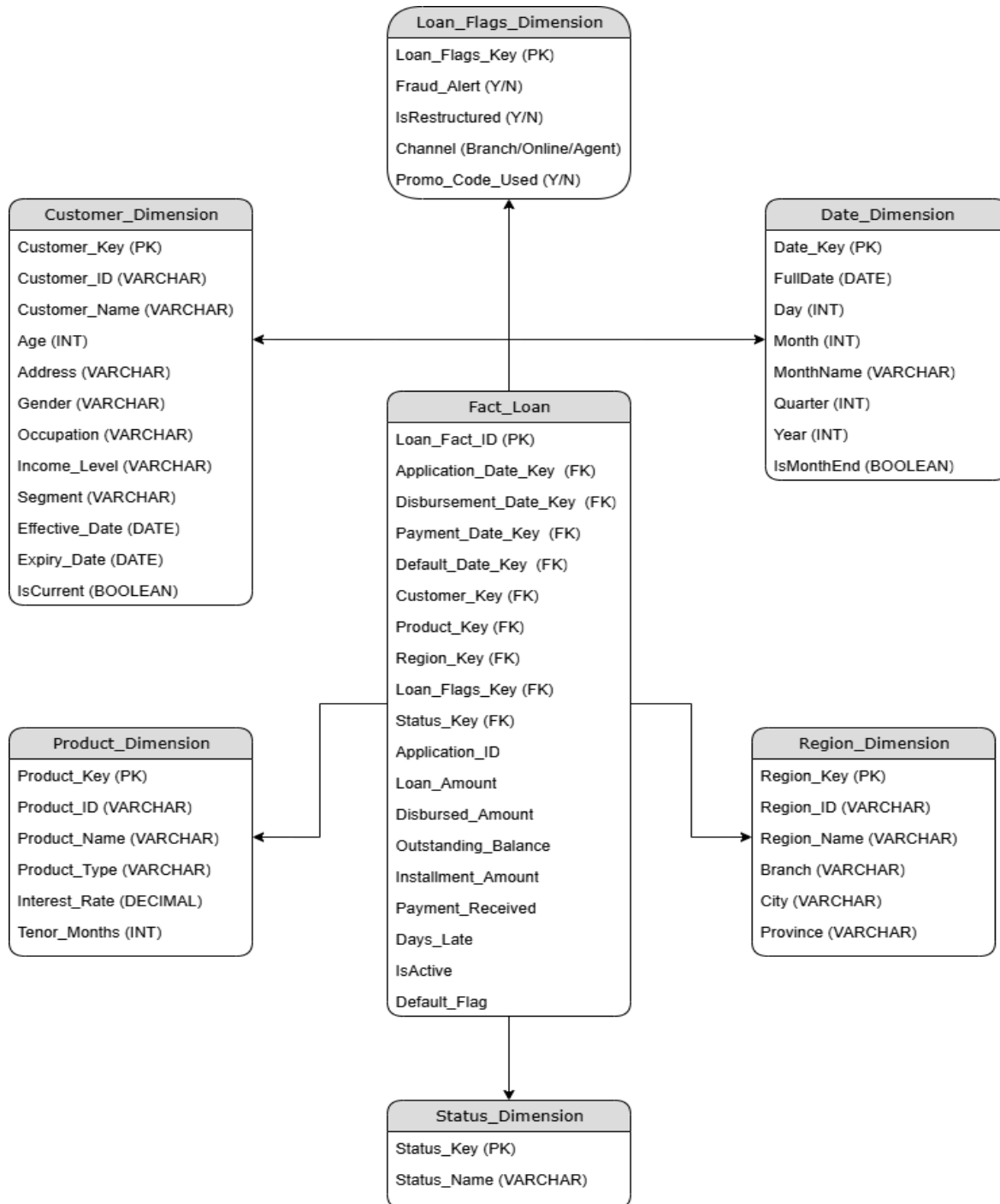
4. Monitoring & Alerting

- Merupakan tahapan untuk mengawasi, dan pemberian alert atau notifikasi bila sewaktu-waktu terdapat regulasi yang salah

2. Kasus: Logical Design dan Dimensional Model untuk BSMD

2a. Rancang skema dimensional (star schema) untuk modul pinjaman digital tersebut. Skema Anda harus mencakup fact dan dimension table yang relevan.

https://drive.google.com/file/d/11zCTQOx49k3pm_2NISWI8I-kkII66Pal/view?usp=sharing



Notes: di setiap dimension dibuatkan 2 id (key & ID) hal ini diterapkan jika terdapat perbedaan input di setiap sistem yang mengupload value pada dimensi tersebut, contohnya pada dimensi customer, customer key akan berlaku sebagai primary atau surrogate key sementara customer_id akan menampung ID nasabah dari sistem operasional

Star Schema tersebut mencakup, beberapa kebutuhan yang dicantumkan dalam soal seperti:

- Melacak semua transaksi pengajuan dan pencairan pinjaman digital nasabah.
 - Hal ini dapat dilakukan dengan menggunakan Loan_Amount dari fact_loan dengan keterangan customer pada customer dimension dan juga di filter atau diurutkan berdasarkan Date dimension atau dimensi lain sesuai kebutuhan analisis
- Mengukur keterlambatan pembayaran, status angsuran aktif/tidak aktif.
 - Perhitungan keterlambatan dapat dilakukan dengan menerapkan formula rata-rata / AVG pada fact (Days_Late)
 - Kemudian untuk status dapat dilakukan dengan menghubungkan status name dari status dimension
- Menganalisis performa pinjaman berdasarkan tipe produk, wilayah, dan segmen nasabah.
 - Untuk segmen dapat dilakukan dengan menghubungkan segment dari customer dimension dan juga Loan_Amount dari fact_loan
 - Untuk Wilayah dapat dilakukan dengan menghubungkan province dari region dimension dengan Disbursed_Amount dari fact_loan
 - Untuk segmen dapat dilakukan dengan menghubungkan Product_Type dari Product dimension dan juga Loan_Amount dari fact_loan
 - Atau bisa juga menggunakan ketiga dimensi terkait untuk analisis tersebut
- Mendukung waktu pelaporan harian dan bulanan.
 - dilakukan dengan menghubungkan OutstandingBalance dari fact_loan dan dimensi date dapat berupa fulldate untuk rincian daily dan juga month untuk rincian bulanan

2b. Buatlah contoh beberapa dimensi khusus yang diperlukan, seperti slowly changing dimension, role-playing, dan junk dimension.

1. **Slowly Changing Dimension**, contoh kasusnya adalah disaat atribut seperti alamat, pekerjaan, status perkawinan bisa saja sewaktu waktu berubah, 1 customer yang sama bisa saja berpindah tempat, semisal nasabah bertempat tinggal di Jakarta di tahun 2023 akan tetapi pindah di tahun berikutnya. Solusinya adalah menggunakan SCD tipe 2 dimana diberikan penambahan dimension row.

Customer_Dimension
Customer_Key (PK)
Customer_ID (VARCHAR)
Customer_Name (VARCHAR)
Age (INT)
Address (VARCHAR)
Gender (VARCHAR)
Occupation (VARCHAR)
Income_Level (VARCHAR)
Segment (VARCHAR)
Effective_Date (DATE)
Expiry_Date (DATE)
IsCurrent (BOOLEAN)

nantinya akan ada 2 baris dimensi yang mengarah ke customer id yang sama akan tetapi dibedakan dari Customer_Key, Effective_Date, Expiry_Date dan IsCurrent

2. **Role-Playing Dimension**, dimana 1 dimensi memiliki beberapa peran contohnya,

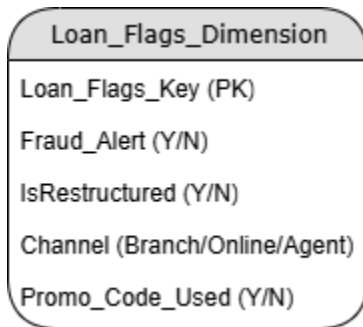
Fact_Loan
Loan_Fact_ID (PK)
Application_Date_Key (FK)
Disbursement_Date_Key (FK)
Payment_Date_Key (FK)
Default_Date_Key (FK)
Customer_Key (FK)
Product_Key (FK)
Region_Key (FK)
Loan_Flags_Key (FK)
Status_Key (FK)
Application_ID
Loan_Amount
Disbursed_Amount
Outstanding_Balance
Installment_Amount
Payment_Received
Days_Late
IsActive
Default_Flag

beberapa variable seperti:

- Application_Date_Key = Melihat pinjaman berdasarkan tanggal pengajuan.
- Disbursement_Date_Key = Melihat pencairan berdasarkan tanggal pencairan.
- Payment_Date_Key = Mengukur keterlambatan berdasarkan tanggal pembayaran
- Default_Date_Key = kapan gagal bayar terjadi

Beberapa variable diatas merupakan variabel yang berasal dari date dimension akan tetapi memiliki peranan yang berbeda satu sama lain

3. **junk dimension**, dimana suatu value kurang bermakna sehingga hanya memakan kapasitas pada fact table dan value tersebut juga tidak dapat dikategorikan ke dimension lainnya, oleh karena itu diperlukan suatu dimension untuk menampungnya disebut dengan junk dimension, contohnya

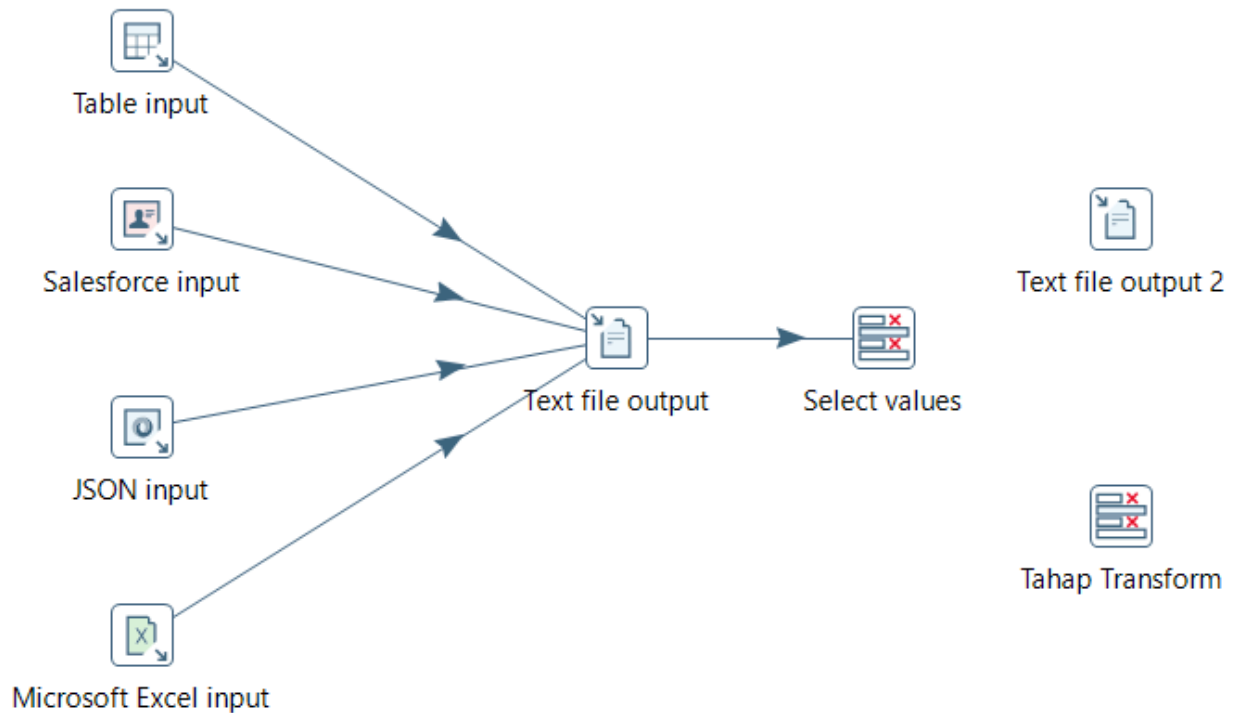


dimana variable tersebut berperan sebagai:

- Fraud_Alert = apakah transaksi pernah di-flag fraud (Y/N).
- IsRestructured = apakah pinjaman pernah restruktur (Y/N).
- Channel = pengajuan lewat *branch*, *online*, *agent*.
- Promo_Code_Used = apakah pakai promo/kode referral (Y/N).

3. Kasus: Strategi ETL pada Perusahaan Ritel Online

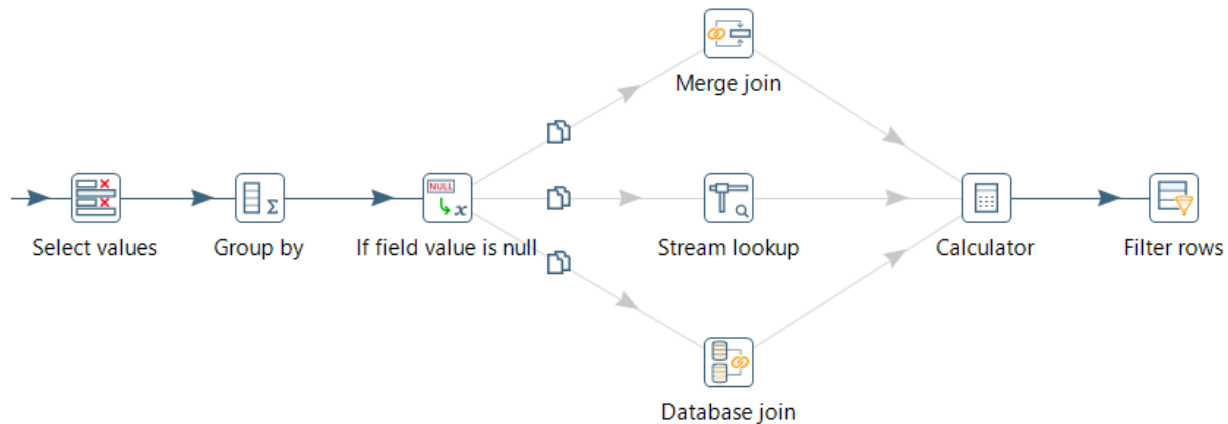
3a. Jelaskan langkah-langkah ETL yang Anda usulkan untuk proyek BI RetailNow, mulai dari proses extract dari berbagai sumber, transform, hingga load ke data warehouse.



EXTRACT

1. Tahapan ekstraksi dimulai dengan mengambil data dari beberapa sumber data, yakni:
 - Table Input dengan menetapkan connection ke MySQL, maka proses pengambilan data “Transaksi harian” dapat dilakukan
 - Salesforce Input, dengan menggunakan Salesforce API untuk menarik “Data pelanggan”
 - JSON input, tools ini akan Baca file JSON secara batch dari folder sistem logistik dan melakukan parsing terhadap data “ log pengiriman dan pengembalian barang” menjadi bentuk tabular
 - Microsoft Excel Input, digunakan untuk membaca dan mengambil data dengan format excel sehingga “data keuangan historis” dapat diekstrak
2. Tahap berikutnya adalah melakukan output (text file output), hal ini berguna untuk mengekstraksi “raw data” yang sudah diambil dari beberapa sumber sehingga proses *re-run* / dapat dilakukan tanpa menarik ulang dari sumber. Tahapan ini dapat merupakan tahapan staging area
3. Tahapan selanjutnya adalah pre-processing, dimana pada tahapan ini akan dilakukan:
 - Pengaturan tipe data(misalkan data tanggal dari string ke date dengan format “YYYY-MM-DD”)
 - Rename field yang sama (cust_id di MySQL = CustomerID di Salesforce)
 - Metadata Alignment
4. Tahapan terakhir dari extract dapat menuju ke “text file output” sebagai suatu checkpoint sebelum ke tahapan transform, atau bisa langsung ke tahapan transform

TRANSFORM



Terdapat beberapa tahapan pada transform yakni:

1. **Data Cleansing**, tahapan untuk memfilter data, memastikan apakah data memiliki duplikat atau bersifat redundant atau tidak, dalam tahapan ini kita dapat memakai fitur group by untuk melakukan hal tersebut.
2. **Data Integration**, Integrasi data, hal ini dilakukan karena kita melakukan penggabungan dari beberapa sumber data yang bervariasi, untuk itu kita perlu menetapkan 1 sumber data sebagai acuan atau lookup yang memastikan keabsahan suatu data, beberapa teknik yang bisa dilakukan adalah:
 - **Merge Join**, dipakai kalau mau menggabungkan 2 stream yang sudah disortir berdasarkan key.
 - **Stream Lookup**, dipakai untuk lookup nilai dari stream lain berdasarkan key tertentu.
 - **Database Join**, digunakan untuk menjalankan query SQL untuk memperkaya data yang sedang mengalir di pipeline.
3. **Standardization & Business Rules**, tahapan pemberian value tambahan untuk memenuhi business rule, tools yang bisa digunakan adalah, Calculator yang digunakan untuk menghitung variable measure dimana valuenya tercipta dari perhitungan seperti formula, contohnya
 - $\text{Net Sales} = \text{Sales} - \text{Returns}$
 - Dan beberapa variabel lainnya yang valuenya didapatkan berdasarkan hasil operasi perhitungan
4. **Data Quality Checks**, Proses evaluasi atau validasi akhir, contoh tools yang dapat digunakan adalah filter rows, yang dapat menyaring data yang tidak memenuhi kondisi tertentu.

LOAD



Tahapan terakhir dimana data yang sudah di extract dan Transform akan di load, pada tahapan ini kita menggunakan table output untuk melakukan migrasi data berdasarkan mapping yang sudah ditetapkan di dalam table output, yang dimaksud mapping adalah proses penyatuan variabel dari sumber data dengan dimension atau fact table yang disesuaikan dengan star schema yang kita bentuk pada database kita.

3b. Apa tantangan teknis utama yang mungkin timbul pada setiap tahap ETL dalam kasus ini?

Terdapat beberapa tantangan teknis yang mungkin saja terjadi pada thapan-tahapan ETL, yakni:

EXTRACT

- Stabilitas koneksi, setiap sumber memiliki limit atau batasan koneksi yang bisa ditampung masing - masing
- Volume data yang besar, tingkat besaran data akan mempengaruhi performa proses ETL
- Konektivitas & API Limit, API Salesforce bisa lambat/timeout

TRANSFORM

- Data cleansing, banyaknya perbedaan tipe data antara data 1 dengan yang lainnya dapat menjadi suatu tantangan
- Metadata alignment, proses penyesuaian metadata yang salah dapat merusak proses ETL
- Integrasi data, Terdapat beberapa option untuk melakukan integrasi data oleh karena itu diperlukan pemahaman yang baik atas pengambilan opsi integrasi tersebut
- Standardization, penetapan format, format date data yang 1 bisa saja berbeda dengan data yang lain

LOAD

- Initial load, volume data yang besar akan memepengaruhi kinerja proses ETL
- Data warehouse design, Skema star harus konsisten, foreign key fact table harus match ke dimension.

3c. Jelaskan bagaimana strategi ETL Anda akan memastikan kualitas data dan konsistensi waktu untuk pelaporan BI.

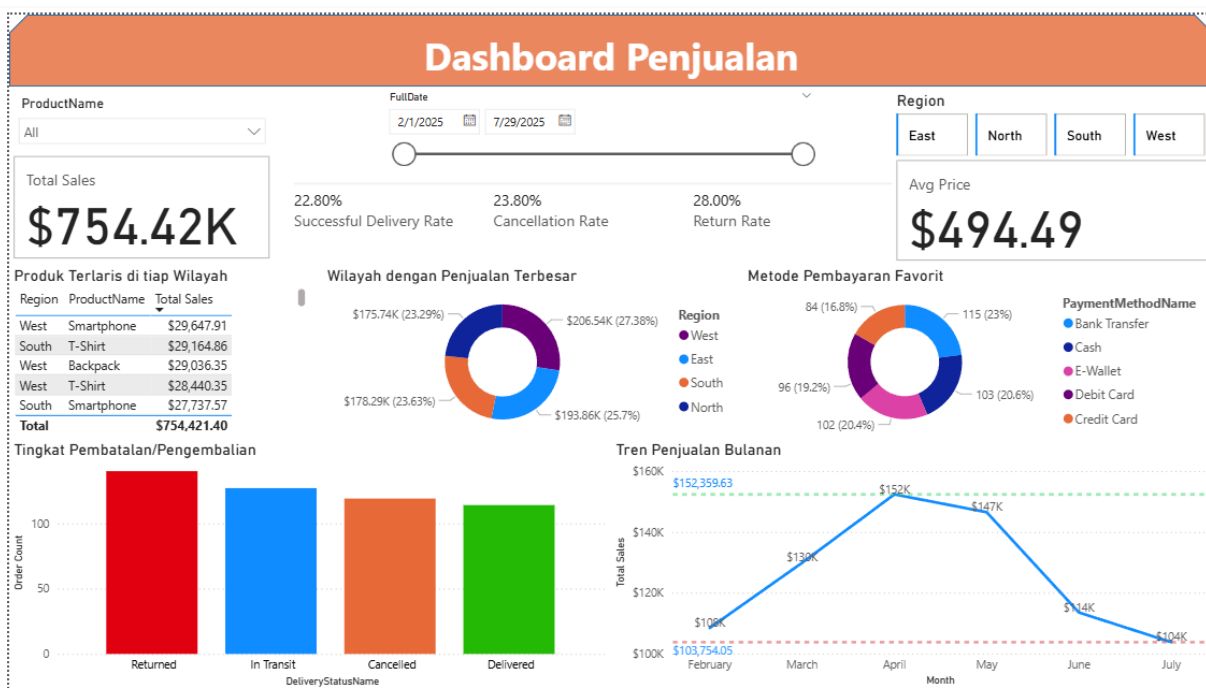
Terdapat beberapa hal yang perlu dipastikan untuk menjamin **kualitas data**, yakni:

- **Staging Layer**, memastikan semua data masuk dulu ke staging sebelum transformasi digunakan guna mencegah pencampuran data mentah & hasil olahan.
- **Data Cleansing**, menangani data yang redundant, null
- **Data Standardization**, penetapan format seperti date atau currency dari setiap sumber agar menggunakan satuan yang sama

Sedangkan untuk menjamin tingkat konsistensi waktu, beberapa hal yang bisa dilakukan seperti:

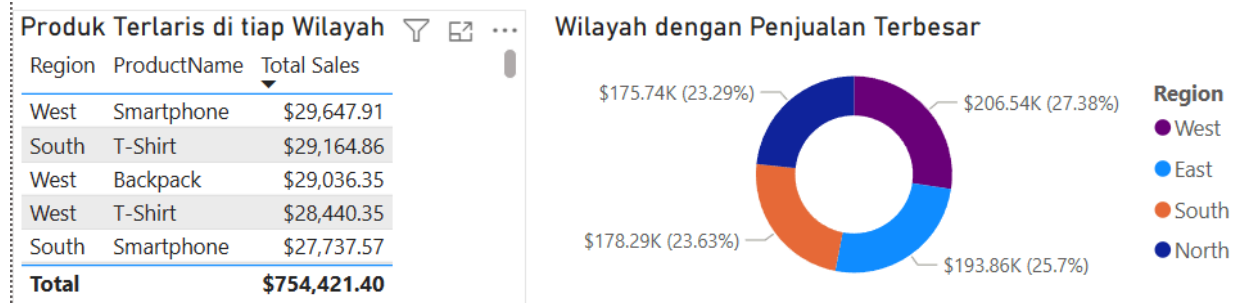
- **Incremental Load**, yakni proses pengambilan data hanya dilakukan untuk data baru di lihat dari acuan date paling baru
- **Data Warehouse Optimizations**, dilakukan dengan menerapkan Indexing & partitioning di fact table

4. Kasus: Pengembangan Dashboard BI



<https://app.powerbi.com/view?r=eyJrljoiYjIzODBhNWJkMS00NDRjLWE1NWItNGI0NGI1YjIwYTJkIiwidCI6IjM0ODViOTYzLTgyYmEtNGE2Zi04MTBmLWI1Y2MyMjZmZjg5OCIsImMiOiEwfQ%3D%3D>

4a. Mengetahui wilayah mana yang memberikan kontribusi terbesar terhadap penjualan, dan produk apa yang paling laku di masing-masing wilayah.



Visualisasi yang digunakan:

- **Donut Chart** = “Wilayah dengan Penjualan Terbesar” menampilkan proporsi penjualan tiap region (West, East, South, North).
- **Table Chart** = “Produk Terlaris di tiap Wilayah” memperlihatkan produk dengan total sales tertinggi di masing-masing wilayah.

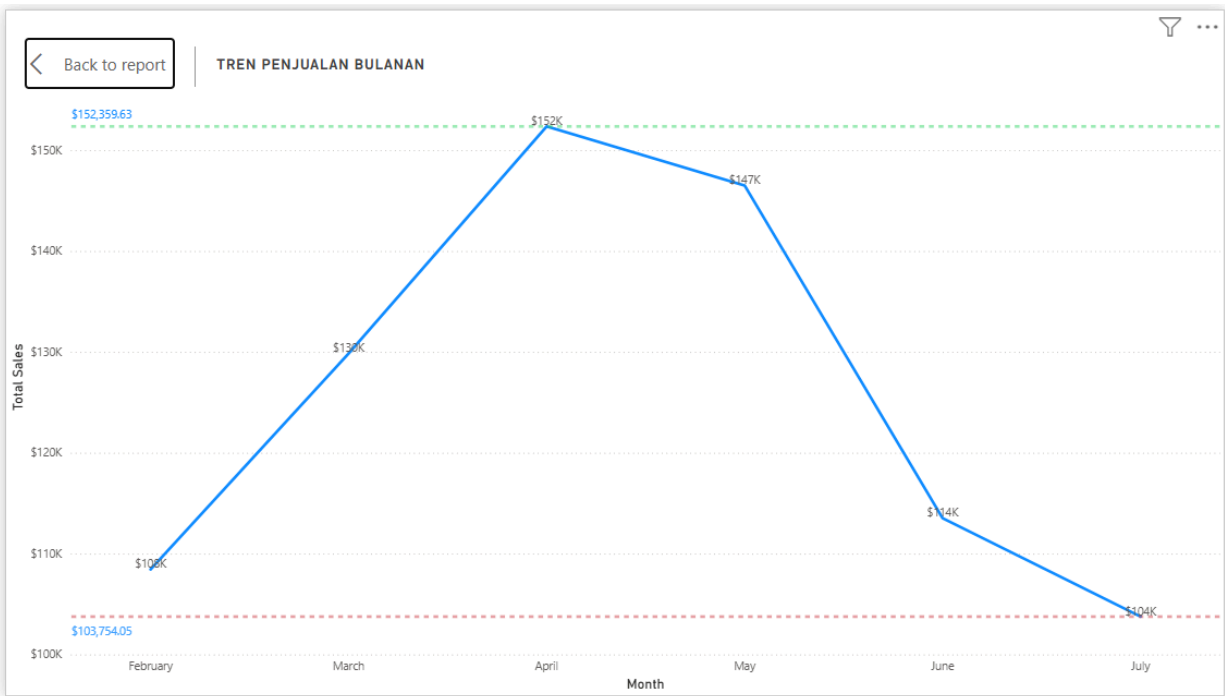
Alasan pemilihan chart:

- Donut Chart mudah menunjukkan perbandingan kontribusi antar wilayah dalam bentuk persentase. (dikarenakan jumlah wilayah hanya 4, memungkinkan penggunaan donut chart)
- Table chart memberi detail produk spesifik, sehingga manajemen dapat melihat produk unggulan per region.

Insight:

- Wilayah **West** memiliki kontribusi penjualan terbesar \$206.54K (sekitar 27.38%), disusul East dan South.
- Produk terlaris berbeda di tiap wilayah, misalnya **Smartphone** di West dan **T-Shirt** di South.
- Penjualan terbanyak masih didominasi oleh Region West dapat dilihat dari top penjualan dimana terdapat setidaknya 3 produk west yang menempati top 5 penjualan, yakni Smartphone, backpack, T-shirt, dengan mengetahui hal tersebut tiap pembagian wilayah bisa mengatur jumlah stok yang diperlukan berdasarkan tingkat demand dari barang tersebut

4b. Menganalisis fluktuasi penjualan dalam periode waktu (misalnya bulanan atau mingguan) untuk mengidentifikasi musim puncak atau penurunan permintaan.



Visualisasi yang digunakan:

- **Line Chart** = “Tren Penjualan Bulanan” memperlihatkan tren penjualan bulanan dari Februari–Juli.

Alasan pemilihan chart:

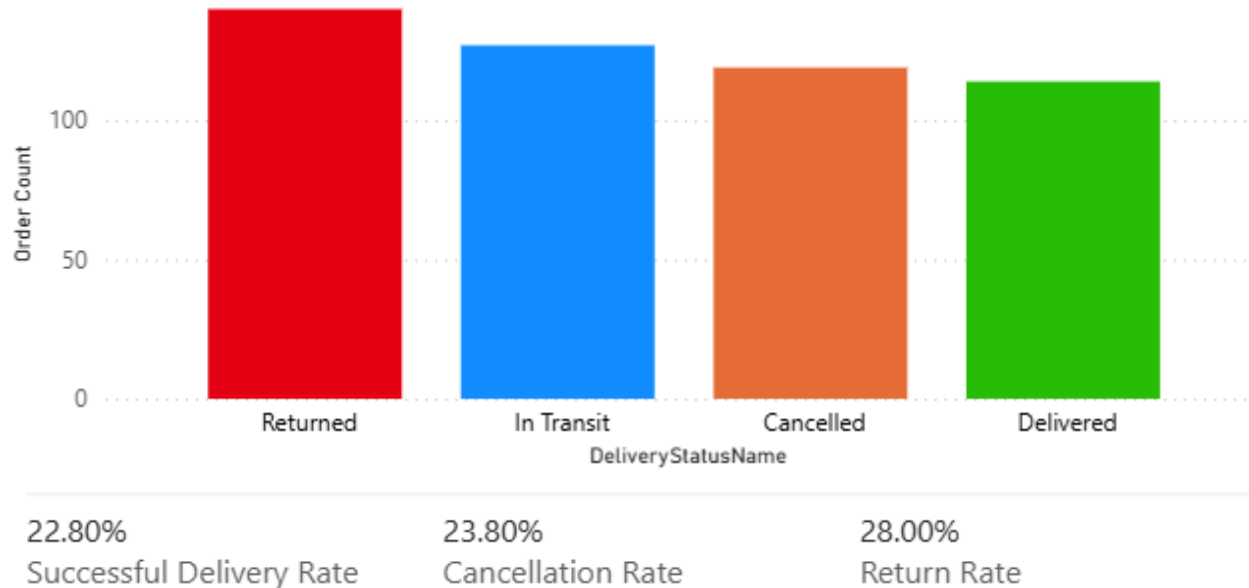
- Line Chart sangat cocok untuk analisis time series karena memperlihatkan pola kenaikan/penurunan secara jelas.

Insight:

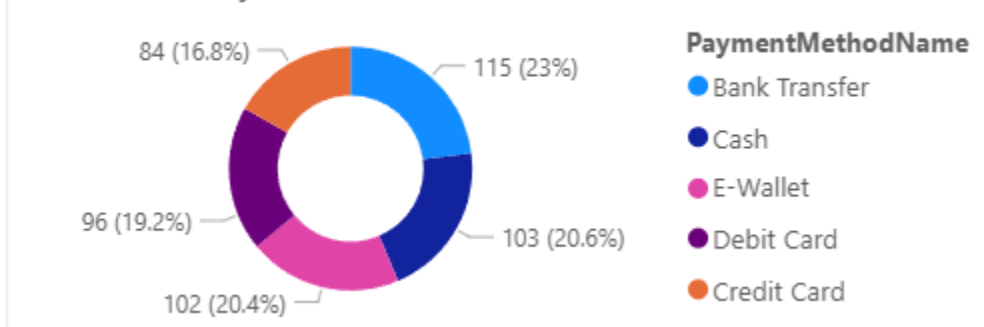
- Penjualan mengalami puncak pada **April** (+/~ \$152K), lalu menurun setelah Mei hingga Juli.
- Penjualan terendah berada pada **Juli** (+/~ \$103K)
- Dengan informasi tersebut maka kita dapat mengatur stok dan kampanye promosi dengan lebih tepat.

4c. Mengetahui metode pembayaran yang paling sering digunakan serta efektivitas pengiriman produk (misalnya tingkat pembatalan dan pengembalian).

Tingkat Pembatalan/Pengembalian



Metode Pembayaran Favorit



Visualisasi yang digunakan:

- **Donut Chart** = “Metode Pembayaran Favorit” memperlihatkan distribusi jumlah transaksi per metode pembayaran.
- **Bar Chart** = “Tingkat Pembatalan/Pengembalian” membandingkan jumlah order berdasarkan status pengiriman (Delivered, In Transit, Cancelled, Returned).
- **KPI Cards** = Menampilkan persentase **Successful Delivery Rate**, **Cancellation Rate**, **Return Rate**.

Alasan pemilihan chart:

- Donut chart memudahkan melihat metode pembayaran paling populer.
- Bar chart efektif untuk membandingkan kategori status pengiriman secara kuantitatif.

- KPI cards menyorot angka kunci sehingga mudah dilihat manajemen.

Insight:

- Metode pembayaran yang paling sering digunakan adalah **Bank Transfer** (23%), disusul **Cash** dan **E-Wallet**.
- Tingkat pembatalan (23.8%) dan pengembalian (28%) cukup tinggi dibanding pengiriman sukses (22.8%), menandakan ada **masalah dalam proses delivery** yang perlu ditangani.
- Keputusan yang bisa diambil: evaluasi mitra logistik, perbaikan proses retur, dan promosi metode pembayaran yang paling disukai pelanggan.