

LAPORAN TUGAS PERGUDANGAN DATA

Perancangan dan Implementasi *Data Warehouse* Pada Industri Pariwisata dan *Hospitality*

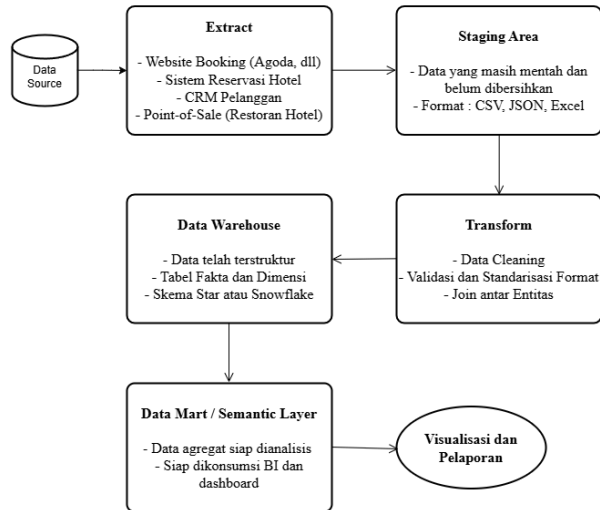


**Disusun oleh:
Kelompok 12**

- | | |
|------------------------------------|------------------|
| 1. Muhammad Hanif Faros | 121450048 |
| 2. Jaclin Alcavella | 122450015 |
| 3. Elok Fiola | 122450051 |
| 4. Febiya Jomy Pratiwi | 122450074 |
| 5. Rut Junita Sari Siburian | 122450103 |
| 6. Elisabeth Claudia S. | 122450123 |

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2025**

1. Alur Aliran Data



Gambar 1. Alur Aliran Data

Proses aliran data dalam *Data Warehouse* di sektor pariwisata dan perhotelan dimulai dari berbagai sumber seperti platform pemesanan daring, sistem reservasi hotel, basis data manajemen hubungan pelanggan (CRM), dan catatan restoran hotel. SSIS digunakan untuk mengekstrak data dari berbagai sumber lalu menyimpannya di *Bronze Tier (Staging Area)* dalam bentuk mentah sebagai file CSV, JSON, atau Excel. Pada tahap ini, data belum melalui proses pembersihan dan hanya berfungsi sebagai lokasi penyimpanan sementara.

Setelah itu, data tersebut kemudian menjalani proses transformasi di *Silver Layer* (gudang data) yang meliputi validasi, normalisasi, dan integrasi data. Setelah dipersiapkan dengan struktur yang rapi, data tersebut dipindahkan ke *Gold Layer* (lapisan pelabelan) untuk analisis lebih lanjut menggunakan model skema bintang atau kepingan salju. Data yang siap dianalisis kemudian divisualisasikan menggunakan Power BI atau SSRS, dan jika perlu, diproses lebih lanjut menggunakan SSAS dalam bentuk kubus OLAP. Diagram alir menunjukkan urutan sistematis dari entri data mentah hingga tahap akhir dalam bentuk laporan dan dasbor interaktif.

2. Arsitektur Data Medallion

Arsitektur data yang digunakan untuk proyek ini adalah arsitektur data medallion yang memiliki tiga lapisan utama yaitu Bronze untuk tempat penyimpanan data mentah, Silver sebagai tempat penyimpanan data yang telah melalui tahap pembersihan sehingga data yang masuk hanya data bersih dan terstruktur dan terakhir lapisan Gold sebagai tempat penyimpanan data akhir. Berikut penjelasan lebih mendalam mengenai lapisan utama Medallion.

1. Bronze Layer

Layer ini sebagai tempat penyimpanan awal bagi semua data yang baru dikumpulkan atau data mentah yang di extract dari berbagai sumber, tetapi

pada proyek kali ini menggunakan data yang telah tersedia dari satu sumber saja.

Cara kerja lapisan ini menerima data langsung dari sumber tanpa perubahan lalu data disimpan apa adanya, lengkap dengan kesalahan, missing value, format yang belum konsisten.

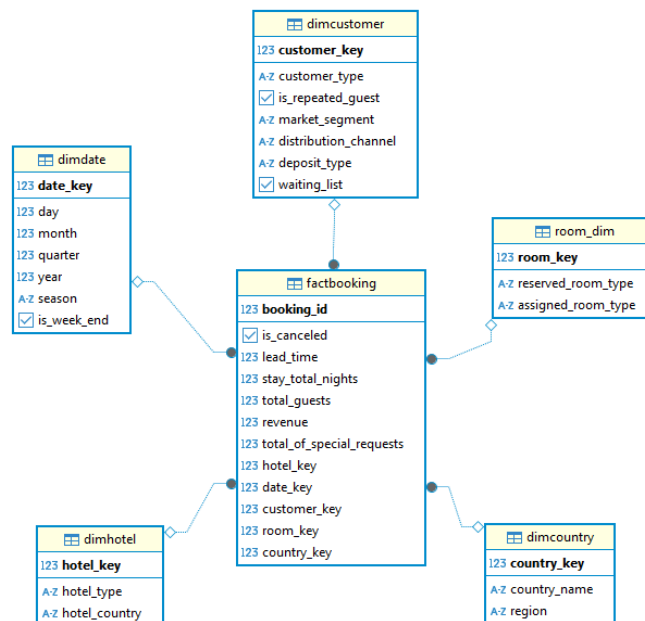
2. Silver Layer

Silver layer menjadi tempat pemberhentian data yang telah dibersihkan dan divalidasi, format data nya akan menjadi parquet berbeda dengan layer Bronze yang masih berbentuk CSV. Data pada layer ini merupakan hasil dari staging melalui proses pembersihan, validasi dan integrasi yang selanjutnya disusun ke dalam bentuk tabel fakta dan tabel dimensi.

3. Gold Layer

Lapisan terakhir merupakan tempat kumpulan data yang siap dianalisis atau diinput machine learning, pada lapisan ini akan mengagregasi, pemodelan, dan menyajikan data agar siap digunakan untuk mendukung pengambilan keputusan

2.1 Design Konseptual



Gambar 2. Star Schema

Design skema yang digunakan pada proyek ini adalah skema bintang *star schema* dimana tabel fakta diisi oleh booking fakta yang berisi data transaksi reservasi dan lima tabel dimensi: customer, date, hotel, room_dim, dan country_dim. Karena struktur relasinya yang sederhana dan performa query yang optimal untuk keperluan OLAP, skema ini mempermudah proses analisis.

3. ETL Pipeline

Proses ETL (*Extract, Transform, Load*) merupakan inti dari aliran data dalam sistem Data Warehouse yang dirancang untuk industri pariwisata dan *hospitality*. ETL berfungsi untuk memindahkan data dari berbagai sumber ke dalam gudang data melalui tahapan yang terstruktur dan otomatis, dengan tujuan menghasilkan data yang konsisten, bersih, dan siap dianalisis. Proses ini dijalankan menggunakan *SQL Server Integration Services (SSIS)*, yang memungkinkan desain pipeline yang efisien dan dapat dijadwalkan secara periodik.

Extract

Tahap ini bertujuan untuk mengambil data dari berbagai sumber, baik internal maupun eksternal. Sumber internal dapat berupa CRM hotel, sistem manajemen reservasi kamar, atau transaksi internal restoran hotel, sementara sumber eksternal meliputi platform seperti TripAdvisor, Booking.com, atau penyedia data wisata lainnya. Proses ekstraksi harus memastikan bahwa data yang diambil lengkap dan mencerminkan kondisi aktual operasional.

Transform

Setelah data dikumpulkan, proses transformasi dilakukan untuk membersihkan dan menyelaraskan data. Tahapan ini meliputi validasi format (misalnya tanggal dan mata uang), penghapusan duplikasi, normalisasi nilai (misalnya penyamaan kode negara), serta penggabungan dari berbagai tabel atau sumber menjadi satu struktur yang konsisten. Transformasi juga mencakup perhitungan derivatif seperti total biaya pemesanan atau durasi inap.

Load

Data yang telah dibersihkan dan ditransformasi kemudian dimuat terlebih dahulu ke *Staging Area (Bronze Layer)*. Setelah diverifikasi, data tersebut dipindahkan ke *Data Warehouse (Silver Layer)* untuk disimpan dalam struktur relasional seperti tabel fakta dan dimensi. Akhirnya, data siap dikonsumsi dipindahkan ke *Data Mart (Gold Layer)*, yang digunakan untuk keperluan analitik dan pelaporan.

Pipeline ETL ini dapat dijadwalkan berjalan secara otomatis harian atau real-time, tergantung kebutuhan bisnis, sehingga informasi yang tersedia selalu terkini dan akurat untuk mendukung pengambilan keputusan berbasis data.

4. Alat ETL

Dalam proses pembangunan sebuah sistem data warehouse, proses ETL (*Extract, Transform, Load*) tidak bisa berjalan dengan sendirinya. Diperlukan rangkaian alat yang saling terintegrasi satu sama lain, mulai dari perancangan skema data hingga visualisasi data. Adapun alat yang digunakan dalam proses tersebut yaitu :

1. Alat Pemodelan dan Desain Skema

Tahapan awal dalam membangun data warehouse adalah merancang sebuah struktur datanya. Hal ini mencakup pembuatan tabel fakta dan tabel dimensi yang saling terhubung satu sama lain, biasanya dalam

bentuk *star schema* atau *snowflake*. Pada tahapan ini, digunakan dua alat utama yang digunakan, yaitu :

1.1 SQL Server Management Studio (SSMS)

SSMS adalah alat untuk mengelola database SQL Server yang digunakan untuk membuat tabel fakta dan dimensi, menjalankan query untuk uji coba struktur skema, dan mengelola database seperti insert, update, atau delete data. SSMS juga menjalankan fungsi utama yaitu berinteraksi langsung dengan database setelah data dimuat melalui proses ETL.

1.2 SQL Server Data Tools (SSDT)

SSDT digunakan untuk merancang dan mengelola struktur database secara visual yang digunakan untuk membuat diagram relasi antar tabel, mendesain paket ETL secara grafis, dan menyusun struktur data warehouse tanpa harus menulis semua kode SQL manual.

2. Alat Integrasi dan Proses ETL

Setelah ditentukan skema data, selanjutnya dilakukan pengambilan data dari berbagai sumber. Data yang dimiliki dilakukan pembersihan dan disimpan ke dalam *data warehouse*. Alat utama yang digunakan dalam proses ETL adalah :

2.1 SQL Server Integration Services (SSIS)

SQL Server Integration Services (SSIS) merupakan alat ETL (Extract, Transform, Load) dari Microsoft yang berfungsi untuk menggabungkan, mengubah, dan memuat data. SSIS dapat mengambil data dari berbagai sumber seperti file Excel, CSV, maupun database lainnya, kemudian memprosesnya dengan cara membersihkan dan mentransformasi data sebelum disimpan ke dalam gudang data. Proses ETL dalam SSIS mencakup dua komponen utama, yaitu Control Flow untuk mengatur urutan proses dan Data Flow untuk memindahkan serta mengolah data dari sumber menuju tujuan secara optimal.

3. Alat Pemantauan dan Optimasi

Supaya sistem yang dimiliki berjalan lancar dan cepat, diperlukan proses pemantauan performa dan mengoptimalkan proses yang berjalan di dalam SQL Server. Hal tersebut dilakukan dengan dua alat berikut.

3.1 SQL Server Profiler

SQL Server Profiler adalah alat Microsoft SQL Server yang digunakan untuk melacak dan merekam aktivitas dalam server database secara real-time. Alat ini memungkinkan pengguna untuk melacak berbagai peristiwa penting, seperti eksekusi query, login pengguna, kesalahan sistem, dan kinerja query. Selain itu, SQL Server Profiler sangat berguna dalam diagnosis masalah performa, mengaudit aktivitas pengguna, dan mengoptimalkan query atau prosedur tersimpan berdasarkan data yang relevan. Profiler menjadi alat penting bagi pengembang dan administrator database untuk menilai dan meningkatkan efisiensi sistem database dengan kemampuan ini.

3.2 Dynamic Management Views (DMVs)

Administrator database dapat melakukan diagnosis masalah, menganalisis beban kerja, dan mengoptimalkan sistem database secara real-time dengan menggunakan Dynamic Management Views (DMVs), sekumpulan tampilan sistem pada SQL Server yang digunakan untuk memantau kondisi, performa, dan aktivitas internal database secara real-time. DMVs juga memberikan informasi penting seperti penggunaan sumber daya, eksekusi query, indeks yang tidak optimal, dan statistik performa server.

4. Alat Visualisasi dan Pelaporan

Data yang sudah rapi dan terstruktur perlu ditampilkan dalam bentuk yang mudah dipahami untuk pengguna bisnis. Sehingga, perlu digunakan alat pelaporan dan visualisasi sebagai berikut :

4.1 SQL Server Reporting Services (SSRS)

Salah satu platform pelaporan Microsoft, SQL Server Reporting Services (SSRS) memungkinkan pengguna membuat, mengelola, dan mengakses laporan berbasis web dari data warehouse atau sumber data lainnya. SSRS mendukung berbagai jenis laporan, termasuk laporan tabular, grafik, dan interaktif, yang dapat disesuaikan untuk memenuhi kebutuhan bisnis. Untuk keperluan analisis dan pengambilan keputusan, laporan dapat dijadwalkan dan dibagikan melalui email atau portal web.

4.2 Power BI (Integrasi Tambahan)

Power BI, alat visualisasi data interaktif milik Microsoft, memungkinkan pengguna menghubungkan berbagai sumber data, termasuk data gudang, melalui dashboard dan laporan yang menarik dan mudah dipahami. Power BI memiliki fitur drag-and-drop, integrasi dengan berbagai layanan cloud, dan kemampuan untuk berbagi visualisasi kepada tim atau manajemen secara online. Power BI memungkinkan pengguna yang tidak memiliki keahlian teknis untuk melihat data dan mendapatkan pemahaman yang mendalam yang akan membantu membuat keputusan strategis perusahaan.

5. Desain Index dalam Data Warehouse

Index merupakan struktur data yang sangat membantu untuk mempercepat pencarian data dan meningkatkan kinerja query, serta efisiensi dalam proses pengambilan data khususnya pada Data Warehouse. Pada proyek ini akan menggunakan 2 jenis utama index yaitu :

5.1 Clustered Index

Jenis index ini mengatur urutan fisik data dalam tabel sesuai dengan urutan indeks. Hanya satu indeks clustered yang dapat dibuat per tabel. Pada proyek ini akan membuat kolom `booking_id` sebagai penanda unik tiap transaksi untuk mengurutkan data secara fisik.

5.2 Non-Clustered Index

Indexing ini tidak mengubah urutan fisik data dalam tabel, tetapi membuat struktur terpisah yang menyimpan referensi ke baris data. Pada proyek ini akan menerapkan pengelompokan, dan penggabungan beberapa kolom penting yang diindeks secara non-clustered seperti `customer_key` untuk menghubungkan dengan `dimcustomer`, `date_key` untuk analisis berdasarkan waktu, dan `country_key` untuk agregasi berdasarkan wilayah

Secara keseluruhan, penerapan indexing dalam data warehouse bertujuan untuk meningkatkan efisiensi sistem dalam menangani permintaan data yang kompleks dan berukuran besar. Dengan adanya index, proses seperti pencarian data, penggabungan tabel, serta pengelompokan dapat dijalankan dengan lebih cepat dan hemat sumber daya. Dengan strategi indexing yang tepat, data warehouse menjadi lebih responsif dan siap digunakan sebagai dasar pengambilan keputusan berbasis data.

6. Skrip Query

Pada perancangan gudang data ini digunakan arsitektur data Medallion yang terdiri atas tiga lapisan, yaitu *Bronze Layer*, *Silver Layer*, dan *Gold Layer*. Berikut adalah query script SQL dari tiap lapisan.

1. *Bronze Layer*

```
SELECT
    CAST([ReservasiID] AS INT) AS reservasi_id,
    CAST([CustomerID] AS INT) AS customer_id,
    TRY_CAST([TanggalCheckIn] AS DATE) AS tanggal_checkin,
    TRY_CAST([TanggalCheckOut] AS DATE) AS tanggal_checkout,
    RTRIM(LTRIM([TipeKamar])) AS tipe_kamar,
    TRY_CAST([JumlahTamu] AS INT) AS jumlah_tamu,
    TRY_CAST([TotalHarga] AS MONEY) AS total_harga,
    UPPER(RTRIM(LTRIM([Status]))) AS status_reservasi,
    GETDATE() AS import_timestamp
INTO staging_reservasi
FROM OPENROWSET(
    BULK 'C:\data\reservasi.csv',
    FORMAT = 'CSV',
    FIRSTROW = 2,
    FIELDTERMINATOR = ',',
    ROWTERMINATOR = '\n'
) AS rows
WITH (
    [ReservasiID] VARCHAR(20),
    [CustomerID] VARCHAR(20),
    [TanggalCheckIn] VARCHAR(50),
    [TanggalCheckOut] VARCHAR(50),
    [TipeKamar] VARCHAR(100),
    [JumlahTamu] VARCHAR(10),
```

```
[TotalHarga] VARCHAR(50),  
[Status] VARCHAR(20)  
);
```

Kode ini berfungsi untuk mengimpor data mentah reservasi hotel dari file CSV ke tabel staging_reservasi di Bronze Layer menggunakan OPENROWSET. Setiap kolom diformat ulang dengan TRY_CAST untuk menghindari kesalahan data, serta dibersihkan dari spasi menggunakan LTRIM/RTRIM, dan UPPER untuk menyamakan format teks. Kolom tambahan import_timestamp dicatat dengan GETDATE(), dan bagian WITH (...) digunakan untuk mendefinisikan struktur data awal secara eksplisit.

2. Silver Layer

```
SELECT  
    r.reservasi_id,  
    r.customer_id,  
    c.nama_pelanggan,  
    c.jenis_kelamin,  
    c.negara_asal,  
    r.tanggal_checkin,  
    r.tanggal_checkout,  
    DATEDIFF(DAY, r.tanggal_checkin, r.tanggal_checkout) AS  
durasi_menginap,  
    r.tipe_kamar,  
    r.jumlah_tamu,  
    r.total_harga,  
    r.status_reservasi,  
    DATEPART(YEAR, r.tanggal_checkin) AS tahun_reservasi,  
    DATEPART(MONTH, r.tanggal_checkin) AS bulan_reservasi,  
    GETDATE() AS last_updated  
INTO dw_reservasi  
FROM staging_reservasi r  
JOIN staging_customer c ON r.customer_id = c.customer_id  
WHERE  
    r.status_reservasi = 'CONFIRMED'  
    AND r.tanggal_checkin IS NOT NULL  
    AND r.tanggal_checkout IS NOT NULL  
    AND c.nama_pelanggan IS NOT NULL;
```

Query ini melakukan transformasi data dari tabel staging dengan menggabungkan informasi reservasi dan pelanggan untuk membentuk tabel dw_reservasi pada Silver Layer. Selain hanya mengambil data berstatus confirmed, query ini juga menghitung durasi menginap, menambahkan informasi waktu (bulan dan tahun check-in), serta menyaring data yang tidak lengkap. Hasilnya berupa data yang telah dibersihkan dan terstruktur untuk mendukung analisis pada lapisan selanjutnya.

3. *Gold Layer*

Pada lapisan ini, data sudah siap digunakan untuk keperluan analitik, pelaporan bisnis, atau visualisasi tingkat lanjut. Berikut adalah contoh *query* yang digunakan pada lapisan *Gold* untuk keperluan analisis kinerja pengacara per jenis kasus dan per tahun.

```
SELECT
    negara_asal,
    COUNT(*) AS total_kunjungan,
    CONCAT(bulan_reservasi, '-', tahun_reservasi) AS periode_reservasi
FROM dw_reservasi
WHERE
    tahun_reservasi >= 2023 -- bisa disesuaikan untuk filter analisis tahunan
    AND negara_asal IS NOT NULL
GROUP BY
    negara_asal,
    bulan_reservasi,
    tahun_reservasi
ORDER BY
    tahun_reservasi DESC,
    bulan_reservasi DESC,
    total_kunjungan DESC;
```

Query ini merupakan bagian dari Gold Layer yang berfungsi untuk menyajikan data agregat siap analisis, yaitu menghitung total kunjungan wisatawan berdasarkan negara asal per bulan dan tahun. Data yang diambil telah melalui proses validasi di Silver Layer dan disaring hanya untuk data sejak tahun 2023. Hasilnya digunakan untuk visualisasi tren wisatawan dalam dashboard BI, memudahkan analisis performa pariwisata berdasarkan wilayah asal wisatawan dari waktu ke waktu.

GitHub : [Kelompok 12 DW RB](#)