

***Legalyze: Perancangan Data Warehouse untuk
Mendukung Pengambilan Keputusan Firma Hukum***



Disusun Oleh Kelompok 21 :

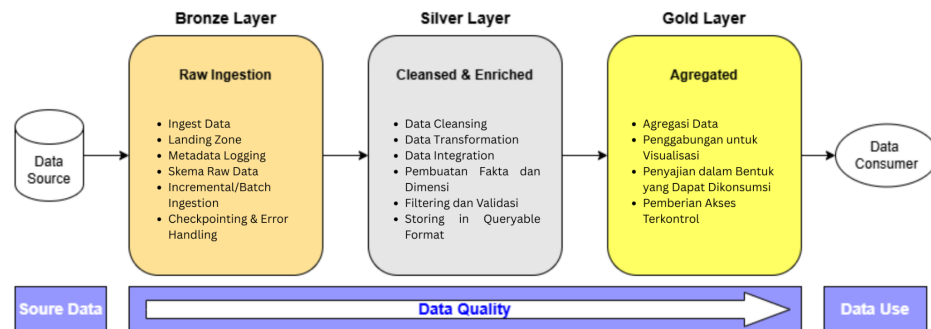
| | |
|---------------------------------|-------------|
| Gede Moena | (121450014) |
| Afifah Syaharani | (121450097) |
| Muhammad Regi Abdi Putra Amanta | (122450031) |
| Tria Yunanni | (122450062) |
| Ratu Keisha Jasmine Deanova | (122450106) |

Dosen Pengampu:
Ardika Satria, S.Si. M.Si

**PROGRAM STUDI SAINS DATA
FAKULTAS SAINS
INSTITUT TEKNOLOGI SUMATERA
LAMPUNG SELATAN
2024/2025**

A. Alur Aliran Data

Dalam pengelolaan data, *Legalyze* menerapkan alur aliran data yang sistematis dan bertahap, dimulai dari pengambilan data mentah hingga menghasilkan informasi yang siap digunakan untuk analisis hukum atau pengambilan keputusan strategis. Proses ini meliputi beberapa tahapan penting, seperti pengumpulan data dari berbagai sumber (*ingestion*), pembersihan serta transformasi data, penyimpanan dalam bentuk yang terstruktur, hingga penyajian data melalui agregasi atau visualisasi. Agar proses pengolahan data berlangsung secara efisien, terorganisir, dan mudah dikembangkan, *Legalyze* menerapkan konsep *Medallion Architecture*. Pendekatan ini mengelompokkan alur data ke dalam tiga tahapan utama yang saling berkesinambungan.



Gambar 1. Aliran Data

Medallion Architecture adalah sebuah pendekatan bertahap dalam pengelolaan data yang membagi proses berdasarkan tingkat kebersihan dan kegunaan data. Pada *Legalyze*, alur data dimulai dari *Bronze Layer*, tempat data mentah dikumpulkan dari berbagai sumber seperti file Excel, API eksternal, dan formulir online tanpa modifikasi. Data ini berisi informasi tentang klien, pengacara, kasus hukum, keuangan, dan *feedback*, yang disimpan dalam format seperti CSV atau JSON.

Selanjutnya, data bergerak ke *Silver Layer*, di mana pembersihan, integrasi, dan transformasi dilakukan. Pada tahap ini, data dari berbagai sumber digabungkan dan disusun menjadi format yang lebih terstruktur. Entitas seperti pengacara, klien, waktu, dan jenis kasus dipisahkan ke dalam tabel dimensi, sedangkan data terkait kasus disusun dalam tabel fakta yang lebih siap untuk dianalisis.

Terakhir, data dipindahkan ke *Gold Layer*, tempat data yang telah dibersihkan dan terstruktur diproses menjadi informasi agregat yang siap digunakan oleh pemangku kepentingan. Pada tahap ini, dilakukan perhitungan metrik seperti jumlah kasus, durasi penyelesaian rata-rata, dan skor kepuasan pengguna. Data disajikan dalam format yang mudah dianalisis dan digunakan dalam *dashboard*, laporan, atau visualisasi lainnya, dengan pengaturan akses yang sesuai dengan peran masing-masing *stakeholder*. Proses ini memastikan bahwa data berkembang dari format mentah menjadi informasi yang siap untuk pengambilan keputusan strategis.

B. Arsitektur Data

Untuk menjamin pengelolaan data yang terstruktur, scalable, dan mudah dikembangkan, sistem Legalyze menerapkan Medallion Architecture. Pendekatan ini membagi alur data menjadi tiga lapisan utama berdasarkan tingkat kebersihan dan kesiapan data, yaitu:

1. Bronze Layer (Raw Data)

Lapisan ini adalah tempat data pertama kali disimpan setelah proses ingestion. Data tersebut diperoleh dari berbagai sumber seperti dokumen keputusan pengadilan, hasil analisis hukum berbasis NLP, dan data operasional firma hukum (klien, pengacara, waktu, jenis kasus). Karakteristik data dalam lapisan ini:

- Mentah dan belum dimurnikan
- Mungkin terdapat duplikat, nilai nol, atau inkonsistensi
- Disimpan sebagai referensi historis

2. Silver Layer (Clean Layer)

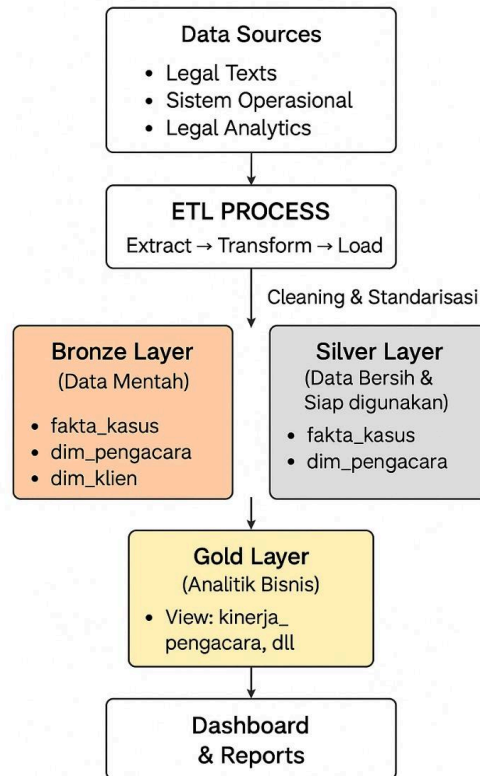
Selanjutnya data dari Bronze Layer kemudian diproses melalui tahapan transformasi dan pembersihan untuk menghasilkan data yang lebih terstruktur dan siap untuk analisis. Di Silver Layer, dilakukan beberapa proses, yakni:

- Normalisasi format data
- Penghapusan duplikasi
- Penggabungan data antar sumber
- Penyusunan ke dalam skema fakta dan dimensi

3. Gold Layer (Business-ready / Analytical Data)

Gold Layer biasanya berisi view atau agregasi data yang telah dirancang khusus untuk kebutuhan pelaporan manajemen atau analisis hukum lanjutan. Lapisan ini merupakan lapisan terakhir yang berisi data sudah sepenuhnya siap digunakan untuk:

- Analisis performa pengacara
- Statistik penyelesaian kasus
- Monitoring kepuasan klien
- Visualisasi dan Dashboard



Gambar 2. Arsitektur Data

C. ETL Pipeline

ETL(*Extract, Transform, Load*) Pada *Project Legalyze* dibangun dengan Pendekatan Modular dan Sistematis yang Mengacu Pada Prinsip *Medallion Architecture* yang Membagi Alur Data ke dalam tiga lapisan yaitu: *Bronze (raw data)*, *Silver (Cleaned and Standardized data)*, dan *Gold (Business-ready Data)*. Perancangan pipeline ini secara khusus terinspirasi dari proyek sumber terbuka GitHub dan juga Pencetus Arsitektur Databricks, dimana Arsitektur ini Mengimplementasikan Pendekatan dengan memberikan struktur logis yang memisahkan antara data mentah, data yang telah dibersihkan, dan data yang siap untuk dianalisis, sehingga memudahkan pengelolaan alur data secara sistematis dan *scalable*.

Extract

Tahapan Pertama dimulai dengan *Extract*, yang bertujuan untuk Mengambil data dari berbagai sumber operasional firma hukum. Pada *Project Legalyze*, data berasal dari open source yang sangat beragam baik yang terstruktur maupun tidak terstruktur . Pada tahap ini, data diambil tanpa modifikasi apa pun dan disimpan secara utuh di *Bronze Layer* sebagai arsip data mentah. Tujuan dari tahap ini adalah memastikan bahwa semua data hukum yang masuk ke sistem terdokumentasi secara lengkap dan dapat ditelusuri ulang jika terjadi kesalahan pada tahap selanjutnya.

Transform

Setelah data berada di *Bronze Layer*, tahap selanjutnya adalah *Transform*, yang dilakukan di *Silver Layer*. Di sini, data dibersihkan dari nilai yang hilang, data duplikat dihapus, dan format standar (seperti format tanggal, penamaan jenis kasus, dan konsistensi nama pengacara) diterapkan. Selain itu, dilakukan pula proses *data enrichment*, seperti klasifikasi otomatis jenis kasus berdasarkan isi dokumen menggunakan pendekatan *rule-based* atau *NLP* sederhana. Misalnya, apabila teks laporan memuat kata “penipuan”, maka sistem dapat mengklasifikasikannya sebagai kasus “Pidana”. Hasil transformasi ini menghasilkan data yang bersih dan terstruktur, yang siap dimasukkan ke dalam skema analitik.

Load

Tahap terakhir adalah *Load*, di mana data yang telah dibersihkan dan diperkaya dimuat ke *Gold Layer*, yaitu tempat penyimpanan data yang telah terstruktur secara analitis. Di lapisan ini, data dimuat ke dalam skema multidimensi berbentuk *Star Schema* sesuai dengan rancangan pada Misi Kedua. Tabel fakta (*Fakta_Kasus*) menyimpan metrik penting seperti durasi penyelesaian, biaya kasus, dan skor kepuasan, yang saling terhubung dengan tabel dimensi (*dim_Pengacara*, *dim_Klien*, *dim_Jenis_Kasus*, dan *dim_waktu*). Setelah berada di *Gold Layer*, data siap diakses oleh pengguna akhir melalui dashboard analitik seperti Power BI, query SQL ad-hoc, atau bahkan untuk pelatihan model machine learning ke depannya.

Dengan membagi proses ETL ke dalam tahap yang terstruktur melalui pendekatan Medallion Architecture, proyek *Legalyze* akan dapat memastikan bahwa data hukum yang kompleks dan tidak seragam dapat dikelola secara akurat, transparan, dan efisien. Pipeline ini juga memungkinkan pelacakan proses transformasi di setiap lapisan, sehingga mendukung integritas data dan auditabilitas sistem secara keseluruhan.

D. Alat ETL

1. Alat Pemodelan dan Desain Skema

1.1 SQL Server Management Studio (SSMS)

SQL Server Management Studio (SSMS) adalah alat utama yang digunakan dalam pengelolaan struktur dan konten database SQL Server untuk proyek *Legalyze*. Sebagai platform berbasis GUI, SSMS menyediakan antarmuka yang intuitif untuk membuat dan memodifikasi objek-objek database seperti tabel, indeks, view, dan stored procedure.

Dalam konteks arsitektur Medallion yang diadopsi *Legalyze*, SSMS digunakan untuk menyusun skema database pada masing-masing lapisan, mulai dari *Bronze* (penyimpanan data mentah), *Silver* (data yang telah dibersihkan), hingga *Gold Layer* (data agregat yang siap dianalisis). Penggunaan fitur “Database Diagrams” sangat membantu dalam memvisualisasikan hubungan antar tabel, terutama dalam penerapan star schema yang digunakan pada *Gold Layer*. Dengan dukungan fitur query editor, SSMS juga memfasilitasi tim untuk menulis, menguji, dan mengoptimasi query SQL dalam proses transformasi dan pelaporan data hukum.

1.2 SQL Server Data Tools (SSDT)

SQL Server Data Tools (SSDT) merupakan ekstensi Visual Studio yang digunakan dalam proses pengembangan skema database secara terstruktur dan modular. SSDT memungkinkan tim untuk membangun proyek basis data sebagai kode (Database as Code), yang sangat berguna dalam implementasi dan manajemen skema di lingkungan tim pengembang Legalyze.

Dengan SSDT, tim dapat mendesain tabel dimensi dan fakta dalam bentuk template, melakukan validasi skema sebelum deployment, serta membandingkan perubahan antar versi skema menggunakan fitur schema comparison. Hal ini sejalan dengan kebutuhan proyek yang menuntut akurasi dan keterlacakan tinggi pada skema data yang digunakan dalam analisis hukum dan pengambilan keputusan. SSDT juga sangat terintegrasi dengan pipeline ETL yang dibangun menggunakan SSIS, sehingga memudahkan manajemen keseluruhan arsitektur data dalam satu lingkungan pengembangan yang konsisten.

2. Alat Integrasi dan Proses ETL

2.1 SQL Server Integration Services (SSIS)

SQL Server Integration Services (SSIS) adalah alat utama dalam membangun pipeline Extract, Transform, Load (ETL) pada proyek Legalyze. SSIS menyediakan lingkungan grafis yang memungkinkan pengembang merancang alur ETL secara modular, sesuai dengan pembagian lapisan dalam Medallion Architecture: Bronze, Silver, dan Gold.

Pada tahap Extract, SSIS digunakan untuk mengambil data dari berbagai sumber seperti file CSV hasil ekstraksi dokumen hukum, API eksternal dari sistem peradilan, serta formulir input online dari klien. Data ini kemudian dimasukkan ke Bronze Layer tanpa modifikasi, sebagai arsip data mentah.

Transformasi dilakukan di Silver Layer dengan memanfaatkan komponen-komponen dalam SSIS seperti Derived Column, Conditional Split, dan Lookup. Proses ini mencakup penghapusan duplikasi, konversi format data, dan enrichment data hukum menggunakan logika rule-based (misalnya pengelompokan kasus berdasarkan isi teks). Hasil akhir dari proses ini adalah data terstruktur dalam bentuk tabel dimensi dan fakta yang siap dianalisis.

Tahap Load mengarahkan data yang telah dibersihkan ke Gold Layer, di mana data dimuat ke dalam struktur star schema. SSIS memungkinkan integrasi penuh dengan SQL Server Agent untuk penjadwalan proses ETL secara otomatis, memastikan data selalu terbaru dan siap digunakan oleh stakeholder hukum dan manajemen firma.

3. Alat Pemantauan dan Optimasi

3.1 SQL Server Profiler

SQL Server Profiler digunakan oleh tim Legalyze untuk memantau performa query, aktivitas transaksi, dan beban sistem SQL Server secara real-time. Alat ini sangat berguna dalam fase produksi, ketika banyak pengguna akhir mulai mengakses sistem dan menjalankan query analitik pada Gold Layer.

Profiler membantu tim database administrator (DBA) dalam mengidentifikasi bottleneck performa, seperti query yang lambat atau indeks yang tidak efisien. Selain itu, alat ini juga digunakan untuk mendeteksi konflik transaksi atau deadlock, sehingga sistem dapat tetap stabil dan responsif meskipun menangani beban data hukum yang besar.

3.2 Dynamic Management Views (DMVs)

Dynamic Management Views (DMVs) adalah kumpulan view internal di SQL Server yang memberikan informasi penting mengenai kondisi dan performa sistem database. DMVs digunakan dalam proyek Legalyze untuk menganalisis pola akses data, penggunaan CPU dan memori, serta efektivitas indeks yang telah dibuat.

Dengan memanfaatkan DMV, tim dapat melakukan tuning dan optimasi terhadap struktur tabel dan query. Sebagai contoh, dari DMV dapat diketahui bahwa indeks tertentu tidak pernah digunakan dalam eksekusi query, sehingga dapat dipertimbangkan untuk dihapus atau disesuaikan. Hal ini berperan penting dalam menjaga efisiensi sistem seiring pertambahan volume dan keragaman data hukum yang dikelola oleh Legalyze.

4. Alat Visualisasi dan Pelaporan

4.1 SQL Server Reporting Services (SSRS)

SSRS digunakan dalam proyek Legalyze untuk membuat laporan reguler dan mendistribusikannya ke stakeholder seperti manajemen firma, pengacara senior, atau analis hukum. Laporan ini mencakup berbagai metrik seperti jumlah kasus per pengacara, rata-rata durasi penyelesaian kasus, serta skor kepuasan klien.

Dengan dukungan untuk grafik, tabel pivot, dan peta geografis, SSRS mampu menyajikan data hukum dalam bentuk visual yang mudah dicerna. Fitur langganan otomatis juga memungkinkan laporan dikirim ke email pengguna sesuai jadwal, tanpa perlu mengakses sistem secara langsung.

4.2 Power BI (Integrasi Tambahan)

Meskipun Power BI bukan bagian langsung dari SQL Server, alat ini digunakan sebagai pelengkap visualisasi yang lebih interaktif dan fleksibel. Power BI terhubung langsung ke Gold Layer di SQL Server dan menampilkan dashboard real-time yang menggambarkan performa firma hukum secara menyeluruh.

Dashboard Power BI digunakan oleh tim manajemen untuk mengevaluasi beban kerja pengacara, efektivitas strategi litigasi, dan tren jenis kasus yang paling sering ditangani. Alat ini juga memungkinkan pengguna untuk melakukan eksplorasi data secara mandiri (self-service analytics) dan mengakses insight hukum tanpa memerlukan keahlian teknis khusus.

E. 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*

```
IF OBJECT_ID('bronze.fakta_kasus', 'U') IS NOT NULL
    DROP TABLE bronze.fakta_kasus;
GO

CREATE TABLE bronze.fakta_kasus (
    kasus_id          INT,
    pengacara_id      INT,
    klien_id          INT,
    jenis_kasus_id    INT,
    waktu_id          INT,
    lokasi_id         INT,

    durasi_penyelesaian INT,
    biaya_kasus        DECIMAL(12,2),
    skor_kepuasan_pengguna DECIMAL(3,2),
    jumlah_kasus       INT,
    jumlah_kasus_per_tipe INT,
    rata-rata_durasi_per_tipe DECIMAL(6,2)
);
GO

IF OBJECT_ID('bronze.dim_pengacara', 'U') IS NOT NULL
    DROP TABLE bronze.dim_pengacara;
GO

CREATE TABLE bronze.dim_pengacara (
    pengacara_id      INT PRIMARY KEY,
    nama              INT,
    spesialisasi_hukum VARCHAR(100),
    lama_pengalaman   INT
);
GO

IF OBJECT_ID('bronze.dim_klien', 'U') IS NOT NULL
```



```

DROP TABLE bronze.dim_klien;
GO

CREATE TABLE bronze.dim_klien (
    klien_id      INT PRIMARY KEY,
    nama          VARCHAR(100),
    usia          INT,
    jenis_kelamin VARCHAR(15),
    kota          VARCHAR(100)
);
GO

IF OBJECT_ID('bronze.dim_jenis_kasus', 'U') IS NOT NULL
    DROP TABLE bronze.dim_jenis_kasus;
GO

CREATE TABLE bronze.dim_jenis_kasus (
    jenis_kasus_id INT PRIMARY KEY,
    tipe_kasus     VARCHAR(100),
    sub_kategori   VARCHAR(100)
);
GO

IF OBJECT_ID('bronze.dim_waktu', 'U') IS NOT NULL
    DROP TABLE bronze.dim_waktu;
GO

CREATE TABLE bronze.dim_waktu (
    waktu_id      INT PRIMARY KEY,
    tanggal       DATE,
    bulan         VARCHAR(25),
    tahun         INT
);
GO

IF OBJECT_ID('bronze.dim_lokasi', 'U') IS NOT NULL
    DROP TABLE bronze.dim_lokasi;
GO

CREATE TABLE bronze.dim_lokasi (
    lokasi_id     INT PRIMARY KEY,
    kota          VARCHAR(100),
    provinsi      VARCHAR(100)
);
GO

```

2. *Silver Layer*

```
CREATE TABLE silver.dim_pengacara AS
SELECT DISTINCT
    nama,
    INITCAP(spesialisasi_hukum) AS spesialisasi_hukum,
    lama_pengalaman
FROM bronze.dim_pengacara
WHERE nama IS NOT NULL;
```

```
CREATE TABLE silver.dim_klien AS
SELECT DISTINCT
    nama,
    usia,
    jenis_kelamin,
    kota
FROM bronze.dim_klien
WHERE nama IS NOT NULL;
```

```
CREATE TABLE silver.dim_jenis_kasus AS
SELECT DISTINCT
    tipe_kasus,
    sub_kategori
FROM bronze.dim_jenis_kasus;
```

```
CREATE TABLE silver.dim_waktu AS
SELECT DISTINCT
    tanggal,
    bulan,
    tahun
FROM bronze.dim_waktu;
```

```
CREATE TABLE silver.dim_lokasi AS
SELECT DISTINCT
    kota,
    provinsi
FROM bronze.dim_lokasi;
```

```
CREATE TABLE silver.fakta_kasus AS
SELECT
    durasi_penyelesaian,
    biaya_kasus,
    skor_kepuasan_pengguna,
    jumlah_kasus,
    jumlah_kasus_per_tipe,
    rata-rata_durasi_per_tipe
FROM bronze.fakta_kasus
WHERE durasi_penyelesaian IS NOT NULL;
```

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.

```
CREATE VIEW gold.kinerja_pengacara AS
SELECT
  p.nama AS nama_pengacara,
  p.spesialisasi_hukum,
  k.tipe_kasus,
  w.tahun,
  COUNT(f.jumlah_kasus) AS total_kasus,
  AVG(f.durasi_penyelesaian) AS rata_durasi,
  AVG(f.biaya_kasus) AS rata_biaya,
  AVG(f.skor_kepuasan_pengguna) AS skor_rata
FROM silver.fakta_kasus f
JOIN silver.dim_pengacara p ON f.pengacara_id = p.id
JOIN silver.dim_jenis_kasus k ON f.kasus_id = k.id
JOIN silver.dim_waktu w ON f.waktu_id = w.id
GROUP BY
  p.nama, p.spesialisasi_hukum, k.tipe_kasus, w.tahun;
```