

# Implementasi Data Warehouse Industri Film Menggunakan PostgreSQL

Stefanus Charles Sselvianto<sup>1</sup>, Septia Angelika Gettin Daely<sup>2</sup>, Andri Wijaya<sup>3\*</sup>

<sup>1,2,3</sup> Universitas katolik Musi Charitas

Email: <sup>1</sup>ajacharles79@gmail.com, <sup>2</sup>septiaangelika13@gmail.com, <sup>3</sup>andri\_wijaya@ukmc.ac.id

## Abstrak

Penelitian ini bertujuan untuk mengimplementasikan sebuah Data Warehouse film dengan memanfaatkan PostgreSQL sebagai sistem basis data utama. Data film diambil dari Kaggle dan berisi informasi seperti judul, sutradara, negara, tahun rilis, kategori, hingga durasi tayangan. Proses pengolahan data dilakukan melalui tahapan ETL, mulai dari *extract* data mentah, *transform* untuk membersihkan duplikasi, menangani nilai kosong, memperbaiki format tanggal, hingga menata ulang data agar lebih konsisten, lalu *load* ke dalam struktur Data Warehouse. Desain penyimpanan menggunakan model *star schema* yang terdiri dari tabel fakta dan beberapa tabel dimensi agar proses analisis berjalan lebih mudah dan terarah. Pada tahap analisis, konsep OLAP juga digunakan untuk menelusuri data secara multidimensi melalui proses seperti *drill-down*, *roll-up*, *slice*, dan *dice*. Hasil penelitian menunjukkan bahwa PostgreSQL tidak hanya efektif sebagai database penyimpanan, tetapi juga mampu mendukung analisis data dengan baik. Data Warehouse yang dibangun menjadi lebih rapi, terstruktur, dan siap digunakan untuk kebutuhan analisis di bidang film dan hiburan.

**Keyword:** Data Warehouse, PostgreSQL, ETL, Star Schema, OLAP, Film Dataset.

## I. PENDAHULUAN

Perkembangan teknologi digital saat ini membuat jumlah data yang dihasilkan semakin besar dari berbagai bidang, termasuk industri hiburan seperti film dan serial televisi. Platform streaming, salah satunya Netflix, menghasilkan dan menyimpan data dalam jumlah sangat besar yang mencakup informasi judul film, jenis tayangan, genre, negara produksi, durasi, hingga tahun rilis. Jika data ini tidak dikelola dengan baik, maka akan sulit untuk digunakan sebagai sumber informasi yang bermanfaat dalam proses analisis dan pengambilan keputusan [1].

Dalam mengatasi permasalahan tersebut, dibutuhkan sebuah sistem pengelolaan data yang mampu menyimpan data secara terstruktur dan mendukung proses analisis, yaitu Data Warehouse. Data Warehouse berfungsi sebagai tempat penyimpanan data terintegrasi yang dirancang khusus untuk keperluan pelaporan dan analisis. Dengan menggunakan Data Warehouse, data yang berasal dari berbagai sumber dapat diolah menjadi informasi yang lebih rapi, akurat, dan mudah dipahami [2].

Pada penelitian ini, data yang digunakan berasal dari dataset publik Netflix Titles yang tersedia di platform Kaggle. Dataset ini berisi kumpulan data film dan serial Netflix dari berbagai negara, lengkap dengan atribut seperti judul, tipe konten (film/serial), genre, durasi, rating usia, tahun rilis, serta deskripsi singkat. Dataset Kaggle sering digunakan dalam penelitian karena mudah diakses dan memungkinkan penelitian untuk dapat direplikasi oleh peneliti lain. Penggunaan dataset publik juga banyak direkomendasikan dalam penelitian berbasis data [3].

Dalam penelitian ini, proses ETL dan OLAP menjadi inti dalam mengimplementasikan Data Warehouse pada penelitian ini. Melalui tahap ETL, data *extract* lalu dibersihkan dari duplikasi, diperbaiki nilai kosongnya, serta diseragamkan format tanggalnya sebelum dimuat ke PostgreSQL. Tahap ini memastikan data konsisten dan siap dianalisis [4]. Setelah tersusun dalam *star schema*, OLAP digunakan untuk menganalisis data secara multidimensi melalui proses *drill-down*, *roll-up*, dan *slice* sehingga menghasilkan *insight* yang lebih mendalam dan mendukung pengambilan keputusan berbasis data [5].

Data film kemudian diolah menggunakan konsep data Warehouse dengan bantuan PostgreSQL sebagai sistem manajemen basis datanya. Melalui proses ini, data yang awalnya acak dan belum rapi dibersihkan, ditata ulang, dan diorganisasi sehingga menjadi informasi yang lebih terstruktur dan mudah dipahami. Pada akhirnya, proses ini bertujuan untuk mengimplementasikan sebuah Data Warehouse yang mampu menghasilkan *insight* yang lebih bernilai dan dapat dimanfaatkan sebagai dasar analisis dalam industri hiburan berbasis data.

## II. TINJAUAN PUSTAKA

### A. Implementasi

Implementasi dapat diartikan sebagai proses menerapkan suatu rancangan atau konsep menjadi sebuah sistem yang benar-benar dapat dijalankan [6]. Dalam konteks penelitian, implementasi bukan hanya soal membuat desain, tetapi juga bagaimana desain tersebut diwujudkan dalam bentuk nyata dan dapat digunakan. Proses implementasi mencakup kegiatan membangun, menguji, dan mengevaluasi sistem agar dapat berfungsi sesuai dengan tujuan yang telah ditetapkan [7].

Konsep ini banyak digunakan dalam penelitian terapan yang berfokus pada bagaimana suatu metode atau teknologi dapat digunakan secara efektif di dunia nyata [8].

### B. Data Warehouse

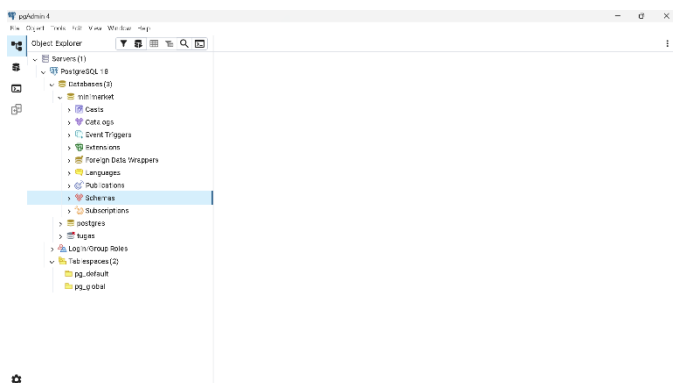
Data Warehouse adalah sistem penyimpanan data yang dibuat untuk memenuhi kebutuhan analisis dan pengambilan keputusan. Berbeda dengan database operasional yang digunakan untuk transaksi harian, Data Warehouse menyimpan data historis dalam format yang sudah terstruktur dan terintegrasi. Dengan demikian, data dari berbagai sumber dapat dikumpulkan di satu tempat, yang memudahkan proses analisis seperti melihat tren, pola, dan perbandingan data dari berbagai wak. Studi sebelumnya menunjukkan bahwa penggunaan Data Warehouse sangat bermanfaat bagi perusahaan karena meningkatkan kualitas informasi yang digunakan untuk pengambilan keputusan. [9].

### C. Film

Film dan serial merupakan bentuk karya audio-visual yang dibuat untuk tujuan hiburan, edukasi, maupun penyampaian pesan sosial. Dalam dunia penelitian berbasis data, film dan serial tidak hanya dipandang sebagai karya seni, tetapi juga sebagai kumpulan data yang memiliki banyak atribut, seperti judul, genre, tahun rilis, negara produksi, durasi, dan rating. Data-data ini dapat dianalisis untuk melihat pola perkembangan industri hiburan, tren genre yang populer, serta distribusi konten berdasarkan wilayah. Pendekatan ini telah banyak digunakan dalam penelitian yang menggabungkan studi media dan analisis data.

### D. PostgreSQL

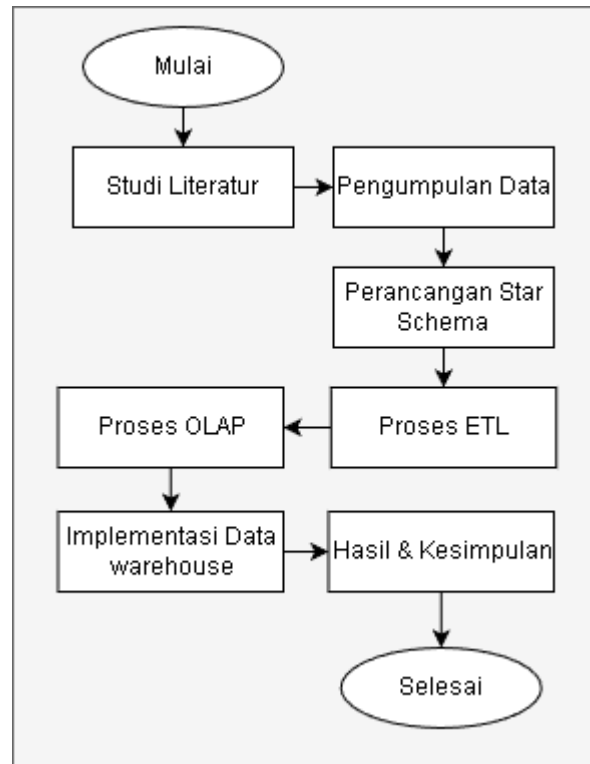
PostgreSQL adalah sistem manajemen basis data relasional bersifat open-source yang dikenal stabil, aman, dan memiliki performa yang baik. PostgreSQL banyak digunakan dalam pengolahan data skala besar, termasuk untuk keperluan Data Warehouse. Keunggulan PostgreSQL terletak pada kemampuannya mengelola data dalam jumlah besar, mendukung standar SQL secara lengkap, serta memiliki fitur optimasi query yang membantu proses analisis data menjadi lebih cepat dan efisien. Berbagai penelitian menyebutkan bahwa PostgreSQL merupakan pilihan yang tepat untuk membangun sistem pengolahan data yang andal [10].



Gambar 1. PostgreSQL

## III. METODE PENELITIAN

Berikut adalah metode penelitian atau langkah-langkah yang akan di lakukan:



Gambar 2. Metode Penelitian

#### A. Studi Literatur

Studi literatur dilakukan dengan membaca dan mempelajari berbagai jurnal, buku, dan sumber terpercaya yang membahas Data Warehouse, pengolahan data, dan visualisasi data agar penelitian ini memiliki dasar teori yang kuat dan terarah.

#### B. Pengumpulan Data

Pengumpulan data dilakukan dengan mengambil dataset publik film dan serial Netflix dari platform Kaggle dalam bentuk file CSV sebagai sumber data utama yang siap diolah tanpa perlu pengumpulan data lapangan.

#### C. Perancangan Star schema

Tahap perancangan Data Warehouse dilakukan dengan menyusun struktur database yang akan digunakan, meliputi penentuan tabel fakta, tabel dimensi, serta hubungan antar tabel dengan pendekatan model star schema, sehingga data dapat tersusun secara rapi dan mendukung proses analisis yang lebih cepat dan terstruktur sebelum diimplementasikan ke dalam PostgreSQL.

#### D. Proses ETL

Proses ETL (Extract, Transform, Load) merupakan tahap utama dalam penelitian ini yang dilakukan dengan mengambil data yang telah dibersihkan, mengubah struktur dan format data agar sesuai dengan desain Data Warehouse, kemudian

memuat data tersebut ke dalam database PostgreSQL sehingga data tersimpan secara terstruktur dan siap digunakan untuk analisis lebih lanjut, dimana proses ini sangat penting karena memungkinkan integrasi data dari sumber mentah menjadi data yang rapi dan historis seperti yang juga banyak diterapkan dalam penelitian Data Warehouse sebelumnya [11].

### E. Proses OLAP

Online Analytical Processing (OLAP) adalah teknologi yang memungkinkan analisis data secara multidimensional pada data warehouse sehingga dapat mendukung pengambilan keputusan yang lebih cepat dan akurat[12]. OLAP biasa digunakan untuk mengekstrak dan menyusun data dalam struktur yang fleksibel seperti kubus multidimensi sehingga mempermudah eksplorasi seperti *drill-down*, *roll-up*, *slice*, dan *dice*, yang membantu dalam penelusuran data berdasarkan banyak dimensi seperti waktu, kategori, lokasi. Implementasi OLAP juga telah banyak diterapkan di berbagai penelitian untuk mendukung pembuatan laporan dan analisis komprehensif pada berbagai bidang seperti pendidikan, perizinan, dan pelayanan kesehatan [12].

### F. Implementasi Data Warehouse

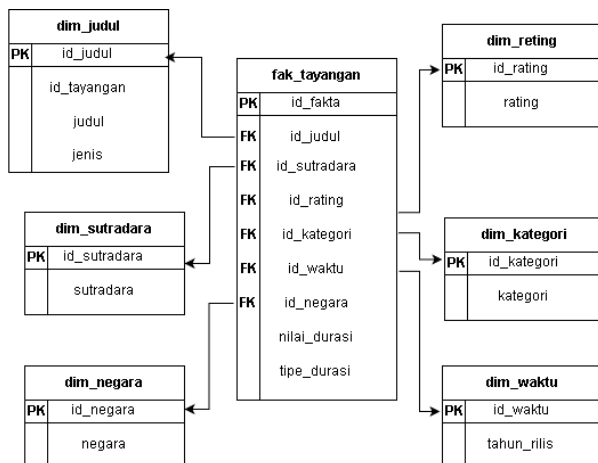
Tahap implementasi Data Warehouse dilakukan dengan mewujudkan rancangan yang sudah dibuat menjadi sistem database yang benar-benar berjalan di PostgreSQL, sehingga data dapat tersimpan secara rapi di dalam tabel-tabel dan dapat digunakan secara langsung untuk proses analisis lebih lanjut [13].

### G. Hasil & Kesimpulan

Tahap analisis hasil dilakukan dengan mengkaji data dan visualisasi yang telah dihasilkan dari Data Warehouse untuk melihat pola, tren, serta insight yang muncul, sehingga dapat diketahui apakah sistem yang dibangun sudah mampu menjawab tujuan penelitian dan menghasilkan informasi yang bermanfaat untuk pengambilan kesimpulan.

## IV. HASIL DAN PEMBAHASAN

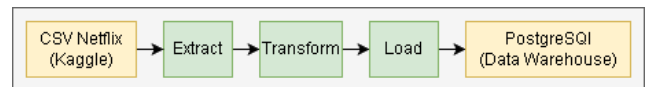
### A. Star Schema



Gambar 3. Star Schema

Star schema menunjukkan bagaimana data diatur dengan menempatkan tabel fakta fak\_tayangan sebagai pusat yang berisi data utama seperti durasi tayangan dan ID referensi ke tabel-tabel dimensi. Setiap tabel dimensi seperti dim\_judul, dim\_sutradara, dim\_rating, dim\_kategori, dim\_waktu, dan dim\_negara menyediakan informasi detail misalnya judul, negara asal, kategori, rating, sampai tahun rilisnya. Pola seperti ini memudahkan kita saat melakukan analisis karena tabel fakta hanya menyimpan data yang benar-benar dibutuhkan untuk perhitungan, sementara tabel dimensi menyimpan detail untuk keperluan filter dan pengelompokan.

### B. Proses ETL (Extract → Transform → Load)



Gambar 4. Proses ETL

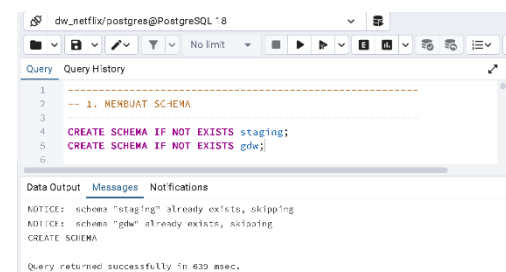
#### 1. Extract

Tahap *Extract* adalah proses awal dalam ETL yang bertujuan untuk mengambil data mentah dari sumber aslinya agar bisa diproses pada tahap berikutnya. Pada tahap ini, data belum diolah, belum dibersihkan, dan belum ditransformasi masih dalam kondisi asli seperti saat diperoleh dari sumber data.

Gambar 5. Dataset mentah *netflix\_titles.csv* (Sumber.Kaggle)

Berikut keterangan isi dari dataset:

- Dataset berjudul “Netflix Movies and TV Shows”
- Format file: CSV (Comma Separated Values)
- Sumber dataset: *Kaggle.com*
- File utama: *netflix\_titles.csv*
- Dataset ini terdiri 8.807 baris data



Gambar 6. Query membuat schema

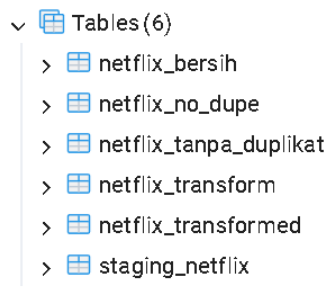
Pada tahap awal proses ETL, dibuat dua schema utama yaitu *staging* dan *gdw*. Schema *staging* digunakan sebagai

tempat penampungan awal untuk menyimpan data mentah hasil import sebelum dibersihkan atau diproses lebih lanjut, sedangkan schema *gdw* berfungsi sebagai lokasi penyimpanan data akhir yang sudah bersih, terstruktur, dan siap digunakan dalam analisis data warehouse.



Gambar 7. Query Extract dataset

Tahap berikutnya memasukkan data mentah dari file CSV ke dalam tabel staging menggunakan perintah COPY, yang berfungsi untuk membaca seluruh isi file *netflix\_titles.csv*.



Gambar 8. Struktur tabel di chema *staging*

## 2. Transform

Pada tahap transformasi, data yang sudah masuk ke *staging* akan dibersihkan dan diolah agar lebih konsisten, rapi, dan siap dimodelkan ke dalam skema gudang data. Proses transformasi mencakup beberapa langkah:

### a) Menghapus Duplikasi

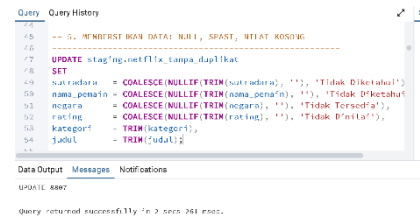


Gambar 9. Query untuk menghapus duplikasi

Pada tahap ini, query digunakan untuk menyingkirkan baris data yang duplikat dari tabel *staging*. Untuk melakukannya, perintah *Select Distinct* digunakan untuk membuat tabel baru bernama *Netflix\_tanpa\_duplikat*. Ini akan memastikan bahwa hanya baris yang benar-benar unik yang disimpan. Untuk memastikan bahwa data yang masuk ke proses transformasi dan data

gudang bebas dari pengulangan yang dapat memengaruhi hasil analisis.

### b) Membersihkan Data



Gambar 10. Query untuk membersihkan data

Pada tahap ini, query digunakan untuk membersihkan data dari nilai yang tidak terisi, spasi berlebih, atau kosong. *TRIM* berfungsi untuk menghilangkan spasi, *NULLIF* mengubah string kosong menjadi *NULL*, dan *COALESCE* mengganti *string* tersebut dengan nilai standar, seperti "Tidak Diketahui" atau "Tidak Dinilai." Dengan demikian, setiap atribut penting memiliki nilai yang sama, dan tidak ada data yang tidak relevan yang dapat mengganggu proses analisis berikutnya.

### c) Memperbaiki Format Tanggal



Gambar 11. Query untuk memperbaiki format tanggal

Query ini digunakan untuk menyeragamkan format tanggal pada kolom *tanggal\_tambah*. Dalam mengidentifikasi berbagai format tanggal, query menggunakan ekspresi reguler (*Regex*). Selanjutnya, *To\_Date* digunakan untuk mengonversi menjadi tipe Date PostgreSQL yang umum. Nilai tanggal dikosongkan menjadi *Null* jika format tidak memenuhi pola apa pun. Kolom tanggal menjadi bersih, terstruktur, dan siap untuk digunakan dalam analisis.

### d) Tabel dimensi



Gambar 12. Query untuk membuat tabel dimensi

Pada tahap pembuatan tabel dimensi, seluruh data seperti judul, negara, sutradara, rating, kategori, dan tahun rilis dipisahkan ke dalam tabel dimensi agar lebih terstruktur dan mudah dianalisis. Masing-masing tabel dimensi dibuat menggunakan perintah *Create Table* dengan atribut utama berupa *primary key* berjenis *Serial* yang berfungsi sebagai *Id* unik, sedangkan nilai asli seperti nama negara, judul tayangan, atau nama sutradara disimpan sebagai atribut unik untuk mencegah data ganda. Pemisahan data ke dalam tabel dimensi ini bertujuan untuk mendukung desain *Star Schema*, sehingga proses analisis di tabel fakta menjadi lebih cepat, efisien, dan tidak menimbulkan redundansi data.

#### e) Tabel fakta

```
Query Query History
197
198 -- 11. MEMBUAT TABEL FAKTA
199
200 DROP TABLE IF EXISTS gdw.fak_tayangan CASCADE;
201
202 CREATE TABLE gdw.fak_tayangan (
203   id_fakta SERIAL PRIMARY KEY,
204   id_judul INT REFERENCES gdw.dim_judul(id_judul),
205   id_negara INT REFERENCES gdw.dim_negara(id_negara),
206   id_sutradara INT REFERENCES gdw.dim_sutradara(id_sutradara),
207   id_rating INT REFERENCES gdw.dim_rating(id_rating),
208   id_kategori INT REFERENCES gdw.dim_kategori(id_kategori),
209   id_waktu INT REFERENCES gdw.dim_waktu(id_waktu),
210   nilai_duras INT,
211   tipe_duras TEXT
212 );
213
Data Output Messages Notifications
CREATE TABLE
Query returned successfully in 174 msec.
```

Gambar 13. Query untuk membuat tabel fakta

Pada tahap ini, tabel fakta dibuat sebagai pusat dari star schema, yaitu tempat semua informasi penting dari berbagai dimensi akhirnya dikumpulkan. Tabel fakta\_tayangan berisi hubungan ke dimensi seperti judul, negara, sutradara, rating, kategori, dan waktu, sehingga setiap tayangan bisa dilacak detailnya dari berbagai sisi. Selain itu, tabel fakta juga menyimpan data numerik seperti nilai durasi dan tipe durasinya. Intinya, tabel ini berfungsi sebagai tempat kumpulnya semua data yang sudah dibersihkan dan dipisah ke dimensi, sehingga proses analisis nanti jadi lebih mudah dan cepat karena semua data penting berada dalam satu titik pusat.

#### f) Pemetaan data dimensi

Tables(7)
dim_judul
dim_kategori
dim_negara
dim_rating
dim_sutradara
dim_waktu
fak_tayangan

Gambar 14. Tabel-tabel dimensi

Semua proses ETL yang telah dilakukan, mulai dari import data mentah, pembersihan nilai

kosong, normalisasi format, dan transformasi data, membentuk tabel dimensi pada data warehouse ini. Setiap dimensi, *dim\_judul*, *dim\_kategori*, *dim\_negara*, *dim\_rating*, *dim\_sutradara*, dan *dim\_waktu*, dibangun untuk menyimpan atribut deskriptif yang tidak berubah. Nilai unik yang dihasilkan dari tabel *staging* yang sudah dibersihkan digunakan untuk membentuk tabel dimensi. Nilai-nilai ini kemudian dimasukkan secara selektif untuk mencegah duplikasi. Dengan dimensi ini, struktur star schema dapat bekerja optimal karena tabel dimensi menyediakan informasi konteks yang lengkap untuk analisis, sementara tabel fakta menyimpan data numerik yang akan dihubungkan dengan masing-masing dimensi.

### 3. Load

	id_fakta [PK] integer	id_judul integer	id_negara integer	id_sutradara integer	id_rating integer	id_kategori integer	id_waktu integer	nilai_duras integer	tipe_duras text
1	1	1108	213	2796	8	25	18	74	Menit
2	2	1377	422	690	8	401	37	73	Menit
3	3	466	408	2427	15	122	37	95	Menit
4	4	6594	213	4138	1	193	33	103	Menit
5	5	7403	248	3200	10	77	33	1	Musim
6	6	4849	213	2050	17	193	18	114	Menit
7	7	6117	106	3854	8	81	45	99	Menit
8	8	Koran	104	4916	18	487	95	81	Musim

Gambar 15. Tabel fakta

Pada tahap Load, semua data yang sudah dibersihkan dan ditransformasikan akhirnya dimasukkan ke dalam Data Warehouse sesuai struktur yang sudah dirancang sebelumnya. Di tahap ini, data dari staging dimasukkan ke tabel-tabel dimensi untuk menyimpan informasi unik seperti judul, negara, sutradara, rating, kategori, dan waktu rilis. Setelah itu, tabel fakta diisi dengan menghubungkan setiap baris data ke dimensi-dimensi tersebut, lengkap dengan nilai durasinya. Proses ini memastikan seluruh data sudah tersimpan dengan rapi, konsisten, dan siap dipakai untuk analisis lebih lanjut, seperti pembuatan laporan atau dashboard.

### C. OLAP (Online Analytical Processing)

Konsep *OLAP (Online Analytical Processing)* untuk membantu menelusuri data secara multidimensi. Tujuannya supaya data di dalam Data Warehouse tidak hanya disimpan begitu saja, tetapi bisa dianalisis dari berbagai sisi sesuai kebutuhan. *OLAP* bisa melakukan eksplorasi data seperti melihat detail yang lebih dalam (*drill-down*), merangkum data ke level yang lebih tinggi (*roll-up*), atau memotong data berdasarkan kategori tertentu (*slice dan dice*). Pendekatan ini membuat proses analisis terhadap data film *Netflix* menjadi lebih fleksibel, mudah dipahami.

#### 1. Jumlah film per tahun rilis

```
Query Query History
1 SELECT
2   dw.tahun_rilis,
3   COUNT(ft.id_fakta) AS jumlah_tayangan
4 FROM gdw.fak_tayangan ft
5 JOIN gdw.dim_waktu dw
6 ON ft.id_waktu = dw.id_waktu
7 GROUP BY dw.tahun_rilis
8 ORDER BY dw.tahun_rilis ASC;
```



Gambar 16. Query untuk jumlah film per tahun rilis

tahun_rilis	jumlah_tayangan
1925	1
1942	2
1943	3
1944	3
1945	4
1946	2
1947	1

Gambar 17. Hasil jumlah film pertahun rilis

Query pada gambar 17 menghitung berapa banyak tayangan yang dirilis pada setiap tahun dengan mengelompokkan data berdasarkan tahun rilis dan gambar di atas menampilkan hasil secara berurutan dari tahun paling tua hingga paling muda.

## 2. Dua negara dengan produksi film terbanyak

```

1 SELECT
2   dn.negara,
3   COUNT(ft.id_fakta) AS total_tayangan
4 FROM gdw.fak_tayangan ft
5 JOIN gdw.dim_negara dn
6 ON ft.id_negara = dn.id_negara
7 GROUP BY dn.negara
8 ORDER BY total_tayangan DESC
9 LIMIT 2;

```

Gambar 18. Query melihat 2 negara produksi film terbanyak

negara	total_tayangan
United Stat...	2818
India	972

Gambar 19. Hasil 2 negara produksi film terbanyak

Gambar di atas menampilkan 2 negara dengan jumlah produksi tayangan terbanyak dengan cara menghitung total tayangan per negara dan mengurutkannya dari yang paling tinggi.

## 3. Perbandingan jumlah Film vs TV Show

```

1 SELECT
2   dj.jenis,
3   COUNT(ft.id_fakta) AS total
4 FROM gdw.fak_tayangan ft
5 JOIN gdw.dim_judul dj
6 ON ft.id_judul = dj.id_judul
7 GROUP BY dj.jenis
8 ORDER BY total DESC;

```

Gambar 20. Query untuk jumlah film vs tv show

jenis	total
Movie	6131
TV Sho...	2676

Gambar 21. Hasil query dari jumlah film vs tv show

Query pada gambar 21 membandingkan jumlah tayangan berdasarkan jenisnya (Film vs TV Show) dengan menghitung total masing-masing jenis dan mengurutkannya dari yang terbanyak.

## 4. Identifikasi kategori terbanyak

```

1 SELECT
2   dk.kategori,
3   COUNT(ft.id_fakta) AS jumlah
4 FROM gdw.fak_tayangan ft
5 JOIN gdw.dim_kategori dk
6 ON ft.id_kategori = dk.id_kategori
7 GROUP BY dk.kategori
8 ORDER BY jumlah DESC
9 LIMIT 5;

```

Gambar 22. Query untuk kategori terbanyak

kategori	jumlah
Dramas, International Movies	362
Documentaries	359
Stand-Up Comedy	334
Comedies, Dramas, International Movies	274
Dramas, Independent Movies, International Movies	252

Gambar 23. Hasil kategori terbanyak

Query pada gambar 23 tersebut mencari kategori yang paling banyak muncul dengan menghitung jumlah tayangan per kategori lalu memilih yang jumlahnya paling tinggi.

Secara keseluruhan, analisis ini memberi gambaran yang jelas tentang bagaimana tren jumlah tayangan dari tahun ke tahun, negara mana yang paling banyak memproduksi konten, kategori apa yang paling sering muncul, rating yang umum digunakan, perbandingan antara Film dan TV Show, hingga genre yang paling dominan, sehingga membantu kita memahami pola dan arah produksi tayangan Netflix secara lebih menyeluruh.

## V. KESIMPULAN

Penelitian ini berhasil mengimplementasikan sebuah Data Warehouse untuk data film menggunakan PostgreSQL sebagai sistem pengelola basis data. Seluruh proses ETL mulai dari mengambil data dari kaggle dengan format file CSV, membersihkan duplikasi, mengisi nilai kosong, memperbaiki format tanggal, hingga menjadi bentuk yang lebih terstruktur dan menghasilkan data yang lebih rapi serta konsisten. Data kemudian dimodelkan menggunakan star schema yang memisahkan tabel fakta dan dimensi, sehingga struktur data menjadi jauh lebih mudah dipahami dan diolah. Dengan desain ini, data yang sebelumnya acak dan tidak seragam berhasil diubah menjadi dataset yang siap dianalisis.

Setelah proses pemodelan selesai, Data Warehouse yang dibuat mampu mendukung analisis multidimensi menggunakan konsep OLAP, seperti drill-down, roll-up, slice, dan dice. Hal ini menunjukkan bahwa PostgreSQL tidak hanya bisa digunakan untuk kebutuhan transaksi harian, tetapi juga efektif sebagai basis analisis data berskala besar [14].

Secara keseluruhan, penelitian ini memberikan hasil berupa sistem penyimpanan dan pengolahan data film yang lebih terstruktur, informatif, dan mudah dieksplorasi untuk kebutuhan analisis maupun pengambilan keputusan berbasis data.

## VI. DAFTAR PUSTAKA

- [1] J. M. Azri Saputra, L. M. Huizen, And D. B. Arianto, 'Sistem Rekomendasi Film Pada Platform Streaming Menggunakan Metode Content-Based Filtering', *Jurnal Transformatika*, Vol. 22, No. 1, Pp. 10–21, Jul. 2024,
- [2] C. Budihartanti *Et Al.*, 'Pengelompokan Film Pada Platform Netflix Menggunakan Metode K-Means Clustering Sebagai Rekomendasi Film', *Journal Of Information System Research (Josh)*, Vol. 5, No. 4, Pp. 1392–1402, Jul. 2024.
- [3] N. Nirpataka, 'Penerapan Content-Based Filtering Untuk Rekomendasi Film Netflix Berdasarkan Genre Dan Klasifikasi Usia Penonton', *Jurnal Ilmiah Teknik Informatika Dan Komunikasi*, Vol. 5, 2025.
- [4] W. S. Fana, R. Sovia, R. Permana, And M. A. Islam, 'Data Warehouse Design With Etl Method (Extract, Transform, And Load) For Company Information Centre', *International Journal Of Artificial Intelligence Research*, Vol. 5, No. 2, Jun. 2021.
- [5] A. A. Setyawan, E. Dudi Darmawan, W. H. Adji, B. Digital, P. P. Ganesha, And K. Akuntansi, 'Dampak Pemanfaatan Teknologi Business Intelligence Untuk Optimalisasi Strategi Penjualan Dengan Metode Olap Di Warmino X', 2025.
- [6] M. Yusuf, R. Akbar, And Imilda, 'Perancangan E-Perpustakaan Berbasis Web Untuk Madrasah Ibtidaiyah Negeri 45 Di Kabupaten Aceh Besar', *Jurnal Sistem Komputer (Siskom)*, Vol. 4, No. 1, Pp. 22–31, Feb. 2024.
- [7] H. Tehuayo, 'Rancang Bangun Sistem Monitoring Aktivitas Santri Berbasis Web Pada Pesantren Rahmatan Lil'alam', Apr. 2024.
- [8] R. Mash, J. Nyasulu, Z. Malan, And L. Hirschhorn, 'Understanding Implementation Research', *Afr J Prm Health Care Fam Med*, Vol. 17, No. 2, P. 4934, 2025,
- [9] A. Dibouliya And W. Bank, 'Review On: Modern Data Warehouse & How Is It Accelerating Digital Transformation', 2023.
- [10] S. V. Salunke And A. Ouda, 'A Performance Benchmark For The Postgresql And Mysql Databases', Oct. 01, 2024, *Multidisciplinary Digital Publishing Institute (Mdpi)*.
- [11] J. Sains *Et Al.*, 'Optimasi Proses Etl Dengan Metode Heuristik Untuk Membangun Data Warehouse', 2019.
- [12] J. Purwanto And Renny, 'Perancangan Data Warehouse Rumah Sakit Berbasis Online Analytical Processing (Olap)', *Jurnal Teknologi Informasi Dan Ilmu Komputer*, Vol. 5, No. 8, Pp. 1077–1088, Oct. 2021.
- [13] L. Setiyani, E. Tjandra, P. Studi Sistem Informasi, And S. Rosma Karawang, 'Perancangan Dan Implementasi Data Warehouse Untuk Perpustakaan Kampus (Studi Kasus: Stmik Rosma Karawang) Design And Implementation Data Warehouse For Campus (Case Study: Stmik Rosma Karawang)', 2020.
- [14] S. V. Salunke And A. Ouda, 'A Performance Benchmark For The Postgresql And Mysql Databases', Oct. 01, 2024, *Multidisciplinary Digital Publishing Institute (Mdpi)*.