

# 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 tanggallnya 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 bernalih 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*

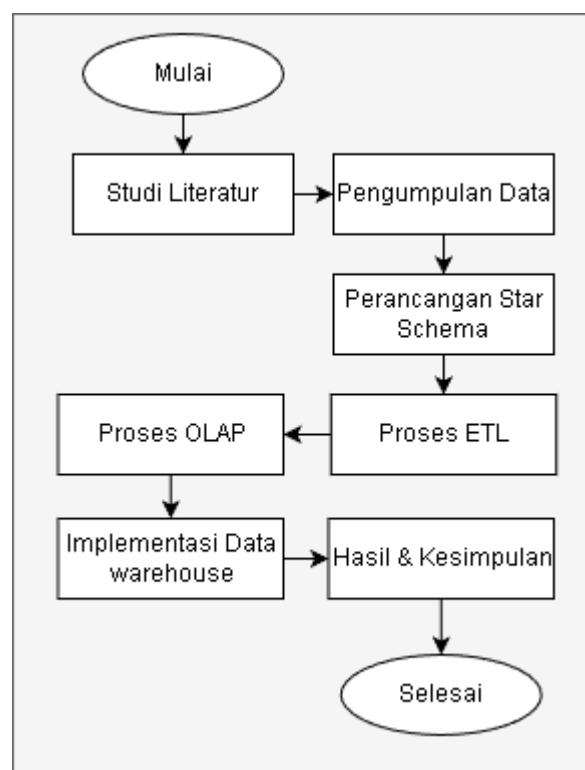
Data *Warehouse* adalah sistem penyimpanan data yang dibuat untuk memenuhi kebutuhan analisis dan pengambilan keputusan. Berbeda dengan *database* operasional yang digunakan untuk transaksi harian, Data *Warehouse* menyimpan data historis dalam format yang sudah terstruktur dan terintegrasi. Dengan demikian, 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 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**

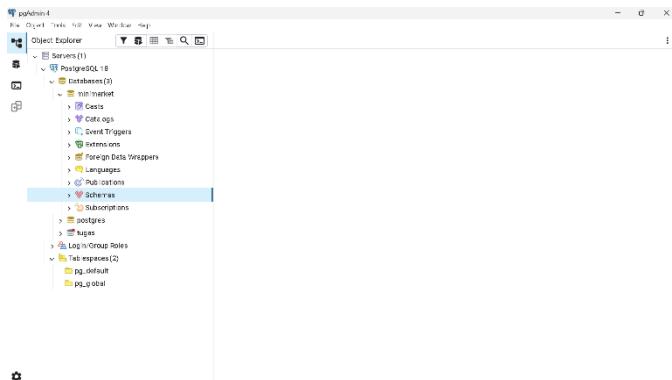
Pengumpulan data dilakukan dengan mengambil dataset publik film dan serial Netflix dari platform Kaggle dalam bentuk file CSV sebagai sumber data utama yang siap diolah tanpa perlu pengumpulan data lapangan.

### **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



Gambar 1. PostgreSQL

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

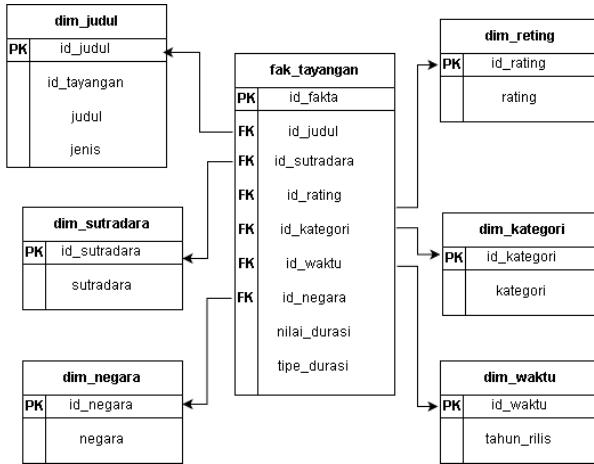
Tahap implementasi Data Warehouse dilakukan dengan mewujudkan rancangan yang sudah dibuat menjadi sistem database yang benar-benar berjalan di PostgreSQL, sehingga data dapat tersimpan secara rapi di dalam tabel-tabel dan dapat digunakan secara langsung untuk proses analisis lebih lanjut [13].

### 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

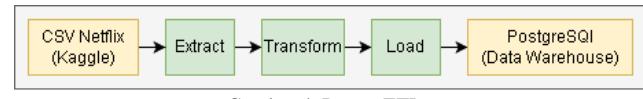
### A. Star Schema



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)



### 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.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	show_id	Type	Title	Director	Cast	Country	Date_Aired	Release_Year	Rating							
2	x1	Movie	Dick Johnson, Kristen Johnston	Un Reel States	15-Sep-21	2020	PG-13	90 min	Docuumentary	As her father nears the end of his life, filmmaker Kris Johnson sets out to make a final film about his life. She follows him around as he goes about his daily routine, including his work as a landscaper, his hobbies, his friends, and his family.						
3	x2	TV Show	Blood & Water	Amy Gummie South Africa	24-Sep-21	2020	TV-MA	2 Seasons	International	After crossing paths at a party, a Cape Town set designer and a former police officer team up to solve a cold case.						
4	x3	TV Show	Chernobyl	Alfred Molina, Emily Watson, David Hemmings	15-Jun-21	2019	TV-MA	1 Season	Documentary	Based on the Chernobyl nuclear disaster, this series follows the cleanup operation and its aftermath.						
5	x4	TV Show	Jaime's New Orleans	Jaime Oliver	United Kingdom	24-Sep-21	2020	TV-MA	1 Season	Documentaries, Food, Nutrition and Cooking	Jaime Oliver goes on a road trip across the United Kingdom to learn more about food and cooking.					
6	x5	TV Show	Kate Peacock	Maxine Peacock	United Kingdom	24-Sep-21	2021	TV-MA	2 Seasons	Documentaries, People	Intimate portraits of a city of colour, characters known to their residents.					
7	x6	TV Show	Land of Milk and Honey	Stephanie D'Abruzzo, H	15-Jun-21	2021	PG-13	120 min	Documentaries, War	Documentary about the life of a woman who returns to her home country after 20 years in America.						
8	x7	Movie	Mix Little Carter	Vivienne, Kiersten, Kristin, C	24-Sep-21	2021	PG-13	93 min	Documentaries & TV Documentaries	But a high-energy teen leaves California and moves to China, an American model school.						
9	x8	Movie	Sandora	Halle Berry	United States	24-Sep-21	1998	TV-MA	125 min	Documentaries, Drama	Based on the true story of a woman who becomes a single mother and tries to raise her son alone.					
10	x9	TV Show	The Great British Baking Show	Paul Hollywood, Matt Tebbutt	15-Jun-21	2019	TV-MA	90 min	Documentaries, Food, Nutrition and Cooking	It's time to get baking! This year, the show is off to new locations, including Scotland, Northern Ireland, and the Isle of Wight.						
11	x10	Movie	The Queen's Gambit	Anyella Neely, Ethan Phillips	24-Sep-21	2021	PG-13	100 min	Documentaries, Drama	A woman adjusting to life after a love affair with a chess prodigy.						
12	x11	TV Show	Vendetta, Truth, Lies and The Devil	Leigh Francis, Sophie	24-Sep-21	2021	TV-MA	1 Season	Documentaries, Reality-TV	Critically acclaimed, this show is a love letter to women in the UK.						
13	x12	Movie	Barbados Breakthrough	Chris O'Dowd, Sophie	24-Sep-21	2021	TV-MA	120 min	Documentaries, Drama	Struggles to live in Barbados, a man who has never been to Barbados, falls in love with a woman who has.						
14	x13	Movie	Confessions of a Blue-Green Hand Catcher	Lucy	22-Sep-21	2021	TV-14	127 min	Documentaries, Drama	Children & Families	When the clever but socially awkward Lucy joins a British TV drama, following Bengaluru police to offer a new life, she finds herself in a bind.					
15	x14	TV Show	Curious India: India Detective	Suraj Singh	22-Sep-21	2021	TV-14	61 min	Documentaries, Mystery	British TV detective following Bengaluru police to offer a new life.						
16	x15	TV Show	Dark River	David Morrissey, Gillian Anderson, M	24-Sep-21	2021	TV-14	81 min	Documentaries, Mystery	British TV detective following Bengaluru police to offer a new life.						
17	x16	Movie	Declassified: Declassified	George Clooney, Julia Roberts	22-Sep-21	2020	TV-MA	1 Season	Documentaries, Mystery	British TV detective following Bengaluru police to offer a new life.						
18	x17	Movie	Carnage	Colin Firth, Michael Stipe, Edward Norton, Penelope	22-Sep-21	2020	TV-MA	67 min	Documentaries, Mystery	Declassified: Declassified documents reveal the pre-9/11 life of George W. Bush.						
19	x18	Movie	Father Identified	Paula Hawkins	15-Jun-21	2020	TV-MA	2 Seasons	Documentaries, Mystery	Crime Thriller	After a deadly home invasion, a convicted killer returns to the scene of the crime, threatening his wife and daughter.					
20	x19	Movie	Goodbye, Darkness	Adrienne Shelly, Helen Hunt, Helen Hunt	23-Sep-21	2021	TV-14	54 min	Documentaries, Mystery	Crime Thriller	Struggles to live in Barbados, a man who has never been to Barbados, falls in love with a woman who has.					
21	x20	TV Show	Heartbreak High	Hannah Hart, Leah Remini, Leah Remini	24-Sep-21	2021	TV-14	100 min	Documentaries, Reality-TV	Heartbreak High: A look into the lives of teenagers in the 1980s.						

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

```

    pgAdmin4 - dw_netflix/postgres@PostgreSQL 8
    CREATE SCHEMA IF NOT EXISTS staging;
    CREATE SCHEMA IF NOT EXISTS gdw;
  
```

The screenshot shows the pgAdmin interface with a query window. The query history contains the commands to create the 'staging' and 'gdw' schemas. The message area indicates that the schema 'staging' already exists and is being skipped. The status bar at the bottom shows the query was returned successfully in 639 msec.

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.

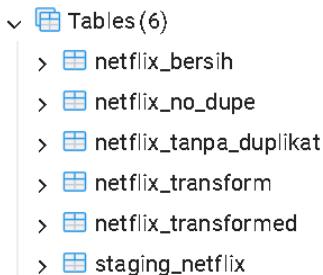
```

Query  Query History
26
27 -- 3. IMPORT CSV KE STAGING
28
29 COPY staging.netflix_titles
30 FROM 'C:/path/netflix_titles.csv'
31 DELIMITER ','
32 CSV HEADER
33 QUOTE '"';
34
Data Output  Messages  Notifications
COPY 8867
Query returned successfully in 373 msec.

```

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 schema *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

```

Query  Query History
35
36 -- 4. MEMBUAT TABEL TANPA DUPLIKAT
37
38 DROP TABLE IF EXISTS staging.netflix_tanpa_duplicat;
39
40 CREATE TABLE staging.netflix_tanpa_duplicat AS
41 SELECT DISTINCT *
42 FROM staging.netflix_titles;
43
Data Output  Messages  Notifications
SELECT 8867
Query returned successfully in 321 msec.

```

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\_duplicat*. 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

```

Query  Query History
45 -- 5. MEMBERSIHKAN DATA: NULL, SPASI, MELAH KOSONG
46
47 UPDATE staging.netflix_tanpa_duplicat
48 SET
49   judul_drama = COALESCE(NULLIF(TRIM(judul_drama), ''), 'Tidak Diketahui')
50   nama_penulis = COALESCE(NULLIF(TRIM(nama_penulis), ''), 'Tidak Diketahui')
51   negara = COALESCE(NULLIF(TRIM(negara), ''), 'Tidak Tersebut')
52   rating = COALESCE(NULLIF(TRIM(rating), ''), 'Tidak D'Inilai')
53   kategori = TRIM(kategori)
54   judul = TRIM(judul);
55
Data Output  Messages  Notifications
UPDATE 8867
Query returned successfully in 2 secs 261 msec.

```

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

```

Query  Query History
51
52 -- 6. MEMPERBAIKI TANGGAL (tambah) YEN ADD DATE
53
54 UPDATE staging.netflix_tanpa_duplicat
55 SET tanggal_tambah =
56 CASE
57 WHEN tanggal_tambah ~ '^([0-9]{1,2})-(0[1-9])-(0[1-9])$' THEN TO_DATE(tanggal_tambah, 'MM Month, YYYY')
58 WHEN tanggal_tambah ~ '^([0-9]{1,2})/([0-9]{1,2})/([0-9]{1,2})$' THEN TO_DATE(tanggal_tambah, 'DD Month, YYYY')
59 WHEN tanggal_tambah ~ '^([0-9]{1,2})/([0-9]{1,2})/([0-9]{1,2})$' THEN TO_DATE(tanggal_tambah, 'DD Month Year')
60 WHEN tanggal_tambah ~ '^([0-9]{1,2})/([0-9]{1,2})/([0-9]{1,2})$' THEN TO_DATE(tanggal_tambah, 'DD Month, YYYY')
61 WHEN tanggal_tambah ~ '^([0-9]{1,2})/([0-9]{1,2})/([0-9]{1,2})$' THEN TO_DATE(tanggal_tambah, 'DD Month, YYYY, Year')
62 ELSE NULL END
63 WHERE tanggal_tambah IS NOT NULL AND TRIM(tanggal_tambah) <> '';
64
Data Output  Messages  Notifications
UPDATE 8867
Total Rows: 0 (0.000 sec) 0.000 sec
Query returned successfully in 533 msec.

```

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

```

Query  Query History
116
117 -- 9. MEMBUAT TABEL DIMENSI DI SCHEMA gdw
118
119
120 -- DIMENSI JUDUL
121 DROP TABLE IF EXISTS gdw.dim_judul CASCADE;
122 CREATE TABLE gdw.dim_judul (
123   id_judul SERIAL PRIMARY KEY,
124   judul_tayangan TEXT UNIQUE,
125   judul TEXT,
126   jenis TEXT
127 );
128
129 -- DIMENSI NEGARA
130 DROP TABLE IF EXISTS gdw.dim_negara CASCADE;
131 CREATE TABLE gdw.dim_negara (
Data Output  Messages  Notifications
CREATE TABLE
Query returned successfully in 313 msec.

```

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

```

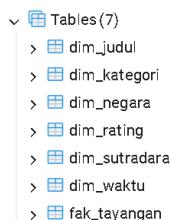
Query   Query History
197
198 -- 11. MEMBUAT TABEL FAKTA
199
200 DROP TABLE IF EXISTS gdw.fak_tayangan CASCADE;
201
202 CREATE TABLE gdw.fak_tayangan (
203     id_fakta SERIAL PRIMARY KEY,
204     id_judul INT REFERENCES gdw.dim_judul(id_judul),
205     id_negara INT REFERENCES gdw.dim_negara(id_negara),
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_durası INT,
211     tipe_durası 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



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_durası integer	tipe_durası text
1	1	1168	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	Muslim
6	6	4049	213	2050	17	193	18	114	Menit
7	7	6117	106	3854	8	81	45	99	Menit
8	8	4049	104	4050	10	489	99	89	Menit

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

```

Gambar 16. Query untuk jumlah film per tahun rilis

Data Output		Messages	Graph Visualiser
tahun_rilis	integer	jumlah_tayangan	bigint
1	1925	1	
2	1942	2	
3	1943	3	
4	1944	3	
5	1945	4	
6	1946	2	
7	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 menampilkannya hasil secara berurutan dari tahun paling tua hingga paling muda.

## 2. Dua negara dengan produksi film terbanyak

```
Query Query History
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 meliat 2 negara produksi film terbanyak

Data Output		Messages	Graph Visualiser
negara	text	total_tayangan	bigint
1	United Stat...	2818	
2	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

```
Query Query History
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

Data Output		Messages
jenis	text	total
1	Movie	6131
2	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

```
Query Query History
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

Data Output		Messages	Graph Visualiser	Notifications
kategori	text	jumlah	bigint	
1	Dramas, International Movies	362		
2	Documentaries	359		
3	Stand-Up Comedy	334		
4	Comedies, Dramas, International Movies	274		
5	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.*, ‘Pengelompokkan 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 Warmindo X’, 2025.
- [6] M. Yusuf, R. Akbar, And Imilda, ‘Perancangan E-Perpustakaan Berbasis Web Untuk Madrasah Ibtidaiyah Negeri 45 Di Kabupaten Aceh Besar’,
- [7] H. Tehuayo, ‘Rancang Bangun Sistem Monitoring Aktivitas Santri Berbasis Web Pada Pesantren Rahmatan Lil’alamin’, 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: Stmk Rosma Karawang) Design And Implementation Data Warehouse For Campus (Case Study: Stmk 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)*.