

**LAPORAN PROYEK MATA KULIAH**  
**12S3202 – GUDANG DATA DAN KECERDASAN BISNIS**

**PERANCANGAN SISTEM DATA MART DAN**  
**DASHBOARD SPOTIFY**



**Disusun oleh:**

1. 12S19014 Kartika Lidya Rotua Sianipar
2. 12S19018 Agnes Abigael Hutaauruk
3. 12S19020 Imelda Siregar
4. 12S19051 Corri Hutahaeon
5. 12S19052 Mulyani Gabe Sayoni Simanjuntak

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**  
**INSTITUT TEKNOLOGI DEL**

**2022**

# DAFTAR ISI

DAFTAR ISI	1
DAFTAR TABEL	3
DAFTAR GAMBAR	4
1. Pendahuluan	5
1.1. Latar Belakang	5
1.2. Ruang Lingkup	5
1.3. Tim Pengembang	6
2. Analisis	9
2.1. Proses Bisnis	9
2.2. Atribut dan Metrik	9
2.3. Isu	12
3. Desain	13
3.1. Detail Bus Matrix	13
3.2. Detailed Dimensional Modeling	14
3.2.1. Dimensi Albums	14
3.2.2. Dimensi Artists	15
3.2.3. Dimensi Genres	15
3.2.4. Fact Tracks	16
3.3. Dimensional Hierarchies	17
3.3.1. Dimensi Albums	17
3.3.2. Dimensi Artists	17
3.3.3. Dimensi Genres	17
3.4. Dimensional Data Model	18
3.5. High-Level Source to Target Map	18
3.6. Detailed ETL Flow for Each Source to Target	19
3.7. Business Intelligence Front End Mockup	19
4. Implementasi	20
4.1 ROLAP Schema	20
4.2 ETL	20
4.3 Analysis Services MOLAP Database	23
4.4 Dashboard	24

5. Penutup	25
5.1. Kesimpulan	25
5.2. Saran	25
LAMPIRAN	26

## DAFTAR TABEL

Tabel 1 Susunan Tim Pengembang.....	6
Tabel 2 Daftar Atribut dan Metrik.....	9
Tabel 3 Isu list attribut dan Metric.....	12
Tabel 4 Detail Bus Matrix.....	13

## DAFTAR GAMBAR

Gambar 1Detail Dimensi Modeling Albums .....	14
Gambar 2 Detail Dimensi Modeling Artists .....	15
Gambar 3Detail Dimensi Modeling Genres .....	15
Gambar 4 Detail Dimensi Modeling Fact Tracks .....	16
Gambar 5 Hirarki Dimensi Albums.....	17
Gambar 6 Hirarki Dimensi Artists.....	17
Gambar 7 Hirarki Dimensi Genres .....	17
Gambar 8 Dimensional Data Model Spotify .....	18
Gambar 9 High level Source to Target Spotify .....	18
Gambar 10 Detail ETL Spotify.....	19
Gambar 11 Mockup .....	19
Gambar 12 Rolap .....	20
Gambar 13 ETL Dim Albums.....	21
Gambar 14 ETL Dim Artist .....	21
Gambar 15 ETL Dim Genre .....	21
Gambar 16 ETL DIm Fact Tracks .....	22
Gambar 17 ETL Molap Database .....	23
Gambar 18 Dashboard Top Track Reporting.....	24

## 1. Pendahuluan

Bab pendahuluan berisi penjelasan mengenai latar belakang pengerjaan proyek, ruang lingkup proyek, tim pengembang yang akan mengerjakan proyek, serta pemangku kepentingan utama dari proyek Perancangan Sistem Data Mart dan Dashboard Spotify.

### 1.1. Latar Belakang

Spotify memproses banyak data karena berbagai alasan, termasuk pelaporan bisnis, rekomendasi musik, penayangan iklan, dan wawasan artis. Miliaran aliran disajikan di 61 pasar yang berbeda dan ribuan trek baru ditambahkan ke katalog setiap hari. Untuk menangani aliran data yang sangat besar ini, Spotify memiliki ~2500 node cluster Apache Hadoop di lokasi, salah satu penerapan terbesar di Eropa, yang menjalankan lebih dari 20 ribu pekerjaan sehari. Spotify menyediakan banyak fitur yang dapat dinikmati pengguna baik pengguna gratis maupun premium seperti membuat playlist lagu sendiri, pilihan lagu berdasarkan genre, mendengarkan lagu secara offline, dan tangga lagu berdasarkan kepopuleran lagu tersebut (*Top Track Reporting*). Untuk fitur tangga lagu, spotify selalu melakukan update pada *top track* yang populer saat ini. Untuk mengidentifikasi data pada fitur tangga lagu tersebut maka kami melakukan Perancangan Sistem Data Mart dan Dashboard Spotify dengan Proses Bisnis *Top Track Reporting*.

### 1.2. Ruang Lingkup

Ruang lingkup dari pengerjaan proyek perancangan sistem data mart dan dashboard pada Spotify adalah sebagai berikut:

1. Menggunakan dataset dari <https://www.kaggle.com/datasets/maltegrosse/8-m-spotify-tracks-genre-audio-features> sebagai Data Source untuk pembangunan sistem.
2. Menggunakan Kimball Approach sebagai pendekatan sistem yang akan digunakan dalam perancangan proyek ini.
3. Proyek ini menyediakan laporan pelaksanaan kegiatan proyek.

### 1.3. Tim Pengembang

Pada Tabel 1 disajikan susunan tim pengembang pada proyek ini.

**Tabel 1 Susunan Tim Pengembang**

No	Nama	Peran	Tanggung Jawab
1	Kartika Sianipar	Analyst	Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses <i>Top Track Reporting</i>
		BI Architect	Membangun BI dari data yang dimiliki untuk bisnis proses <i>Top Track Reporting</i>
		Data Architect	Membangun dimensional modelling dan implementasi
2	Agnes Abigail Hutaeruk	Analyst	Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses <i>Top Track Reporting</i>
		BI Architect	Membangun BI dari data yang dimiliki untuk bisnis proses <i>Top Track Reporting</i>
		Data Architect	Membangun dimensional modelling dan implementasi

No	Nama	Peran	Tanggung Jawab
3	Imelda Siregar	Analyst	Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses <i>Top Track Reporting</i>
		BI Architect	Membangun BI dari data yang dimiliki untuk bisnis proses <i>Top Track Reporting</i>
		Data Architect	Membangun dimensional modelling dan implementasi
4	Corri Hutahaeen	Analyst	Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses <i>Top Track Reporting</i>
		BI Architect	Membangun BI dari data yang dimiliki untuk bisnis proses <i>Top Track Reporting</i>
		Data Architect	Membangun dimensional modelling dan implementasi
5	Mulyani Simanjuntak	Analyst	Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses <i>Top Track Reporting</i>
		BI Architect	Membangun BI dari data yang dimiliki untuk bisnis proses <i>Top Track Reporting</i>



No	Nama	Peran	Tanggung Jawab
.			
		Data Architect	Membangun dimensional modelling dan implementasi

## 2. Analisis

Bagian ini berisi bisnis proses, *bus matrix*, *attributes and metric list*, *issues list*.

### 2.1. Proses Bisnis

#### 2.1.1. [BP-01]-Proses Bisnis *Top Track Reporting*

*Top Track Reporting* adalah salah satu fitur yang disediakan oleh spotify berupa tangga lagu yang dimana akan diupdate secara berkala berdasarkan popularitas lagu, genre, atau mood. *Top Track* spotify dapat dilihat dengan masuk ke spotify kemudian memilih *playlist top track*. Kemudian pengguna akan memutar lagu pada *playlist Top Track* dan mendengarkan lagu baik secara *shuffle* maupun secara berurutan sesuai dengan tingkat kepopuleran lagu. Pada setiap *track* akan ditunjukkan data yang menunjukkan kepopuleran lagu seperti judul track, album track, genre, dan jumlah streaming track tersebut.

### 2.2. Atribut dan Metrik

#### 2.2.1. Atribut dan Metrik [BP01]

Berikut adalah daftar atribut dan metrik yang akan digunakan dalam Perancangan Sistem Data Mart dan Dashboard Spotify.

**Tabel 2 Daftar Atribut dan Metrik**

Dimension/Fact Table	Attribute/Fact Name	Description	Sample Values
Albums	albumkey	Merupakan kode unik (primary key ) yang menjadi identitas utama untuk tabel album	1
	id_album	Merupakan kode unik yang mendefinisikan album	5U6viBMb17ayuEad6Inpg
	name	Nama dari lagu	I Wish You Love

	album_type	Merupakan jenis album, ada 3 jenis album yaitu single, compilation dan album. Single jika album hanya terdiri dari satu lagu. Album jika album lebih dari satu lagu	single
	release_date	Merupakan tanggal rilis dari album	1591833600000
	Popularity	tingkat popularitas album dari rentang 0-100	0
Artists	artistkey	kode unik sebagai primary key untuk artist	3
	id_artist	kode unik yang mendefinisikan identitas artist	3E2vuvr0IQbReTbXw2MhX8
	name_artist	nama artis	Justin Bieber
	genre_artist	genre yang terkait dengan artis	hip hop
	popularity	tingkat popularitas artis dari rentang 0-100	94
	followers	jumlah orang yang menyukai atau mengikuti artis di Spotify	798
Tracks	trackskey	Kode unik (primary key) yang menjadi identitas untuk sebuah track	1

	id_track	Kode unik yang mendefinisikan identitas track	2g8HN35AnVGIk7B8yMucww
	name	nama dari track	Bog Poppa - 2005 Remaster
	duration	durasi trek yang sedang dibuka atau dijalankan yang dihitung per millisecond	252746
	preview_url	preview/pratinjau yang akan ditujukan langsung ke track yang akan kita dengarkan	<a href="https://p.scdn.co/mp3-preview/770e023eb0318270ecc5caa018d758e5e0844de9?cid=cde021ca5d3e42a8bd440f100...">https://p.scdn.co/mp3-preview/770e023eb0318270ecc5caa018d758e5e0844de9?cid=cde021ca5d3e42a8bd440f100...</a>
	track_number	merupakan nomor dari track yang sedang diputar	13
	popularity	tingkat popularitas lagu dari rentang 0-100	77
Genres	genreskey	Merupakan kode unik ( <i>Primary Key</i> ) yang menjadi identitas utama dari Genre	84231
	id_genres	Merupakan kode mendefinisikan identitas dari Genre tersebut	Hip hop

### 2.3. Isu

Berikut ini adalah issue list attribute and metric yang dianalisis dalam pengerjaan Perancangan Sistem Data Mart dan Dashboard Spotify.

**Tabel 3 Isu list attribut dan Metric**

Issue#	Attribute	Issue	Format
1.	Tanggal	Untuk penulisan tanggal pada data sangat bervariasi, oleh karena itu diperlukan pemilihan format penulisan yang tetap.	dd/mm/yy

### 3. Desain

Bagian ini berisi *detailed bus matrix*, *detailed dimensional modeling*, *dimensional hierarchies*, *dimensional data model*, spesifikasi *extract*, *transform*, *load* (ETL), *detailed ETL flow for each source to target*, dan *user interface dashboard mockup*.

#### 3.1. Detail Bus Matrix

**Tabel 4 Detail Bus Matrix**

Business Process Name	Fact Table	Fact Grain Type	Granularity	Facts	albums	artists	genres
Top track reporting	FactTracks	Periodic Snapshot	one row per popularity track	Most popular track, album with the most popular track, artist with the most popular track	v	v	v

## 3.2. Detailed Dimensional Modeling

### 3.2.1. Dimensi Albums

Table Name	DimAlbum																			
Table Type	Dimension																			
Display Name	Album																			
Database Schema																				
Table Description	Album Dimension																			
Comment	comes from album																			
Biz Filter Logic																				
Size																				
Generate Script?	Y																			
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
albumKey	albumKey	Merupakan kode unik (primary key) yang menjadi identitas utama untuk tabel album	-1	1	key				int	50		PK ID		N		spotify			albumKey	int
id_album	id_album	Merupakan kode unik yang mendefinisikan album	-1	5U6vBMb17ayuEad6lnpg	key				nvarchar	100				N		spotify		Album	id_album	nvarchar
name	name	Nama dari album		I Wish You Love	1				nvarchar	100				N		spotify		Album	name	nvarchar
album_type	album_type	Merupakan jenis album, ada 2 jenis album yaitu single dan album. Single jika album hanya terdiri dari satu lagu. Album jika album lebih dari satu lagu		single	2				nvarchar	50				N		spotify		Album	album_type	nvarchar
release_date	release_date	Merupakan tanggal rilis dari album		1,59183E+12					datetime	100				N		spotify		Album	release_date	datetime
popularity	popularity	tingkat popularitas album dari rentang 0-100		0	2				int	100				N		spotify		Album	popularity	int

Gambar 1Detail Dimensi Modeling Albums

### 3.2.2. Dimensi Artists

Table Name	DimArtist																			
Table Type	Dimension																			
Display Name	Artist																			
Database Schema																				
Table Description	Artist Dimension																			
Comment	comes from artist table																			
Biz Filter Logic																				
Size	one for each artist																			
Generate Script?	Y																			
									Target							Source				
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
artistKey	artistKey	kode unik sebagai primary key untuk artist		3	2				nvarchar	50				N		spotify		Artist	artistKey	nvarchar
id_artist	id_artist	kode unik yang mendefinisikan identitas artist		3E2vuvr0IQbReTbXw2MhX8	2				nvarchar	100				N		spotify		Artist	id_artist	nvarchar
name_artist	name_artist	nama artis		Eminem	2				nvarchar	50				N		spotify		Artist	name_artist	nvarchar
Genres_artist	Genres_artist	Genres yang terkait dengan artis		hip hop	2				nvarchar	50				N		spotify		Artist	Genres_artist	nvarchar
followers	followers	jumlah orang yang menyukai atau mengikuti artis di Spotify		798	2				nvarchar	50				N		spotify		Artist	followers	nvarchar
popularity	popularity	tingkat popularitas artis dari rentang 0-100		94	2				nvarchar	50				N		spotify		Artist	followers	nvarchar

Gambar 2 Detail Dimensi Modeling Artists

### 3.2.3. Dimensi Genres

Table Name	DimGenres																			
Table Type	Dimension																			
Display Name	Genres																			
Database Schema																				
Table Description	Genres Dimension																			
Comment	comes from Genres																			
Biz Filter Logic																				
Size	one for each Genres																			
Generate Script?	Y																			
									Target							Source				
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	Precision	Key?	FK To	NULL?	Default Value	Source System	Source Schema	Source Table	Source Field Name	Source Datatype
GenresKey	GenresKey	Merupakan kode unik (Primary Key) yang menjadi identitas utama dari Genres		84231	1				int	50		PK		N		spotify		Genres	GenresKey	nvarchar
id_Genres	id_Genres	Merupakan kode mendefinisikan identitas dari Genres tersebut		64021	2				nvarchar	50				N		spotify		Genres	id_Genres	nvarchar
name_Genres	name_Genres	Merupakan nama atau tipe dari jenis Genres		KPOP					nvarchar	50						spotify		Genres	name_Genres	nvarchar

Gambar 3Detail Dimensi Modeling Genres



### 3.2.4. Fact Tracks

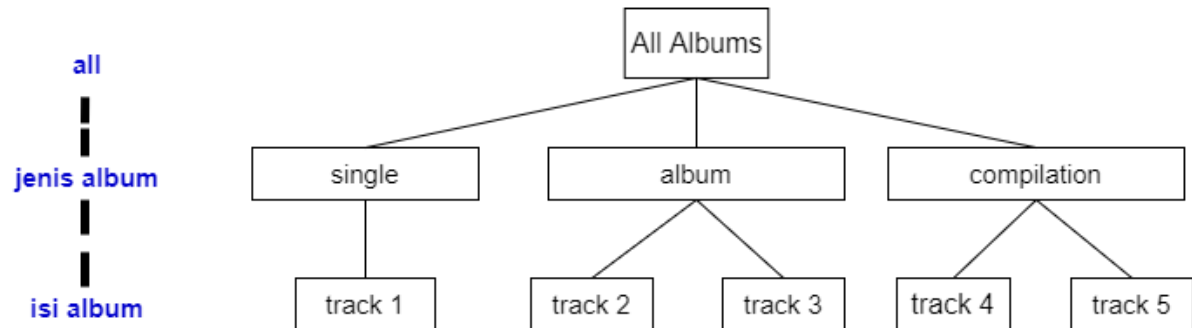
[illegible]

*Gambar 4 Detail Dimensi Modeling Fact Tracks*

### 3.3. Dimensional Hierarchies

#### 3.3.1. Dimensi Albums

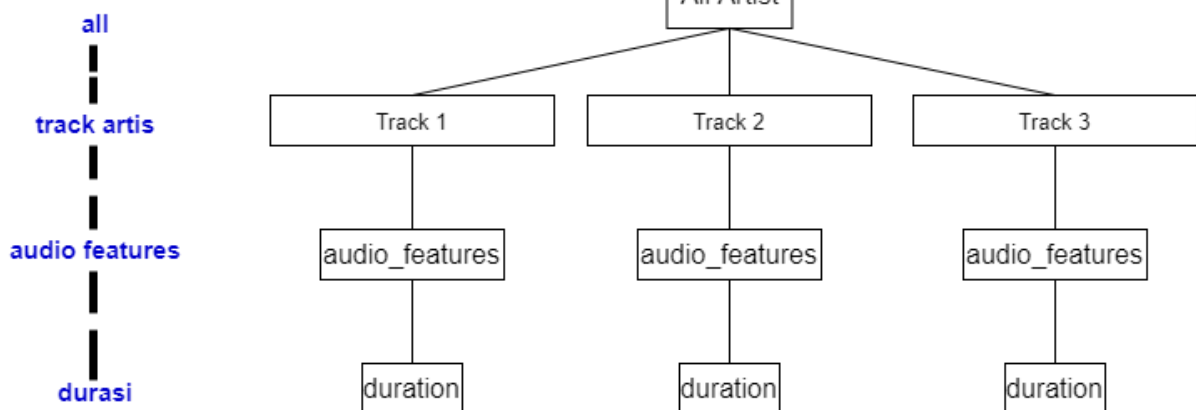
##### Hierarchy : Albums



Gambar 5 Hirarki Dimensi Albums

#### 3.3.2. Dimensi Artists

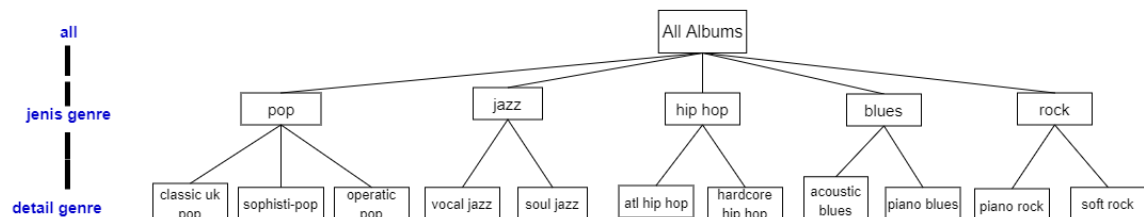
##### Hierarchy : Artist



Gambar 6 Hirarki Dimensi Artists

#### 3.3.3. Dimensi Genres

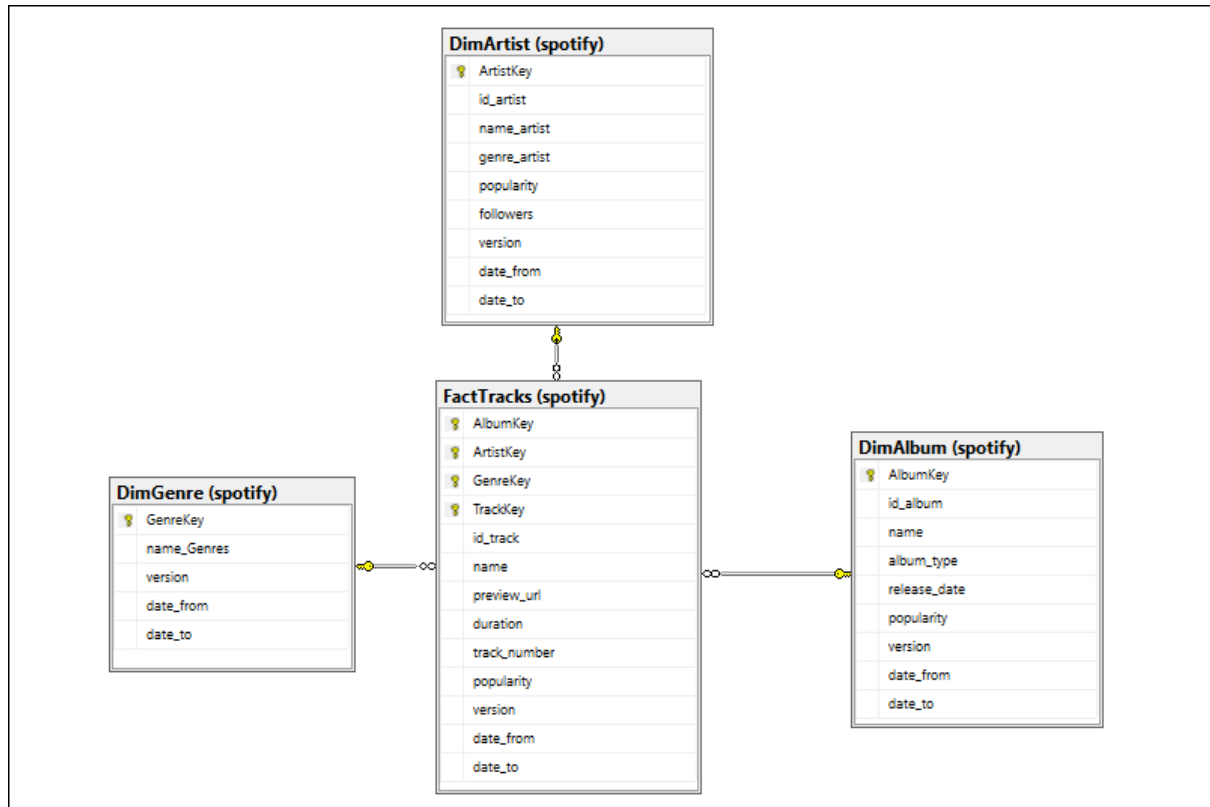
##### Hierarchy : Genres



Gambar 7 Hirarki Dimensi Genres

### 3.4. Dimensional Data Model

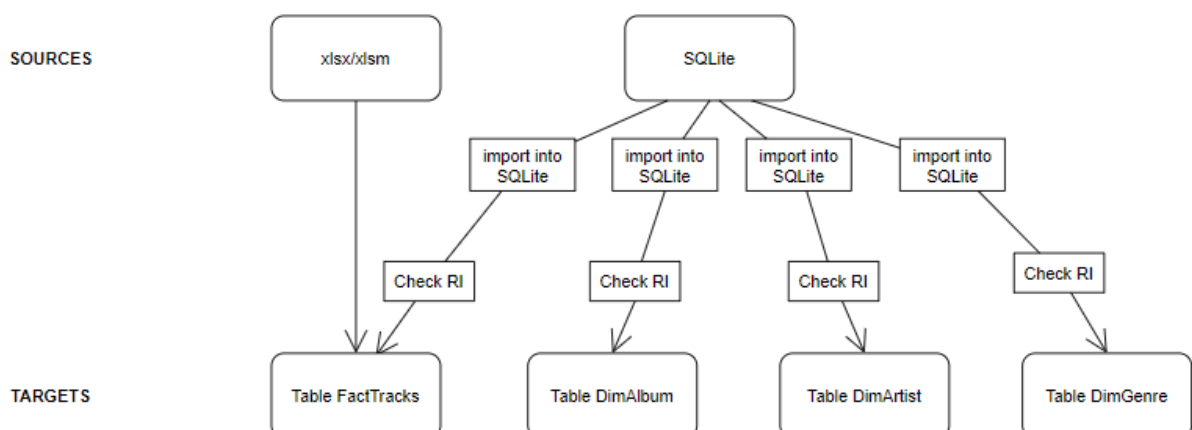
Berikut merupakan Dimensional Data Model dari Spotify.



Gambar 8 Dimensional Data Model Spotify

### 3.5. High-Level Source to Target Map

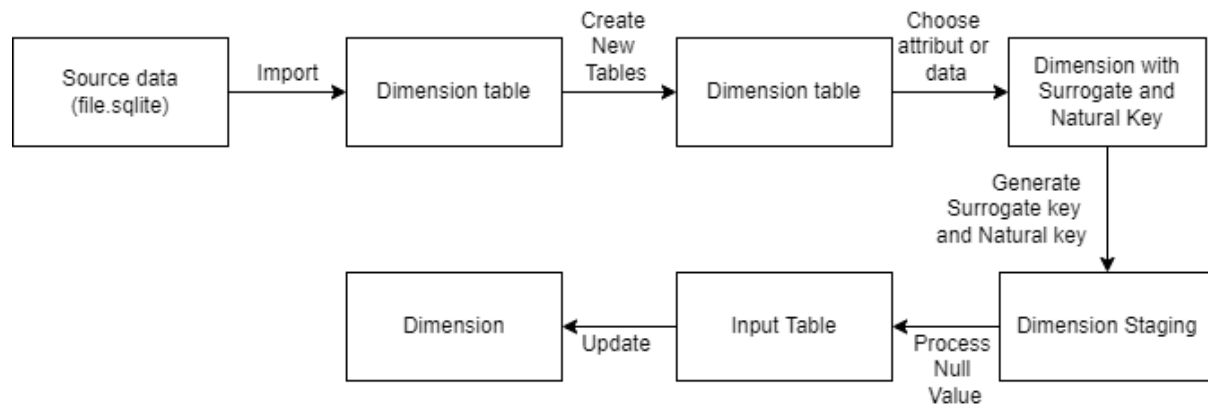
Berikut adalah High-Level Source to Target Map di Spotify.



Gambar 9 High level Source to Target Spotify

### 3.6. Detailed ETL Flow for Each Source to Target

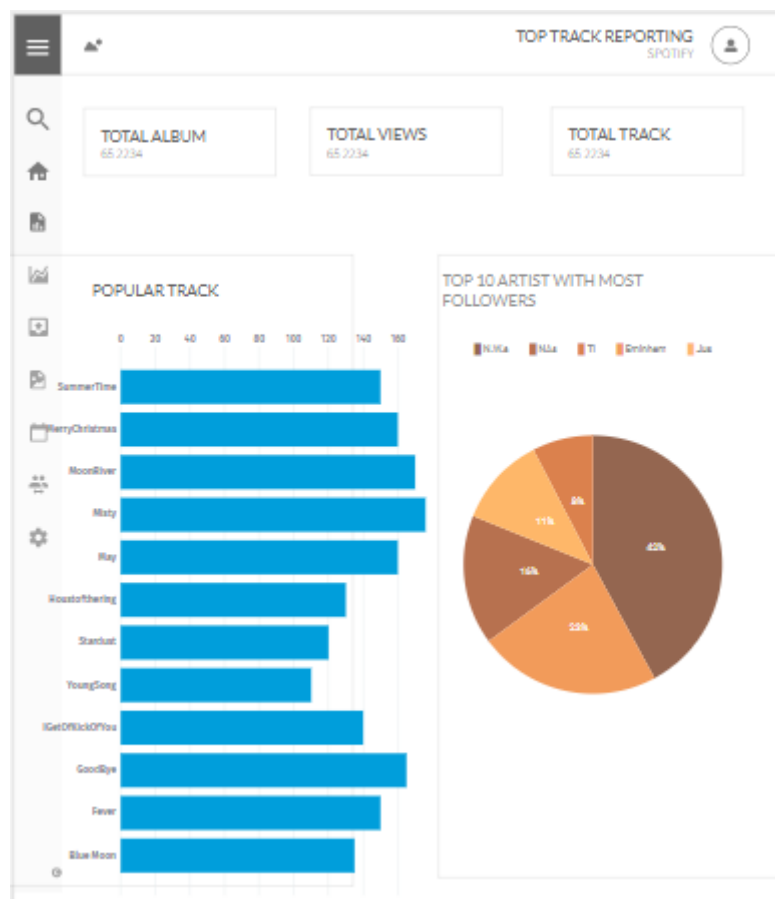
Berikut adalah aliran ETL secara rinci untuk setiap aliran sumber ke target dari spotify.



Gambar 10 Detail ETL Spotify

### 3.7. Business Intelligence Front End Mockup

Berikut adalah mockup dari proses bisnis Top Track Reporting menggunakan moqups.com.



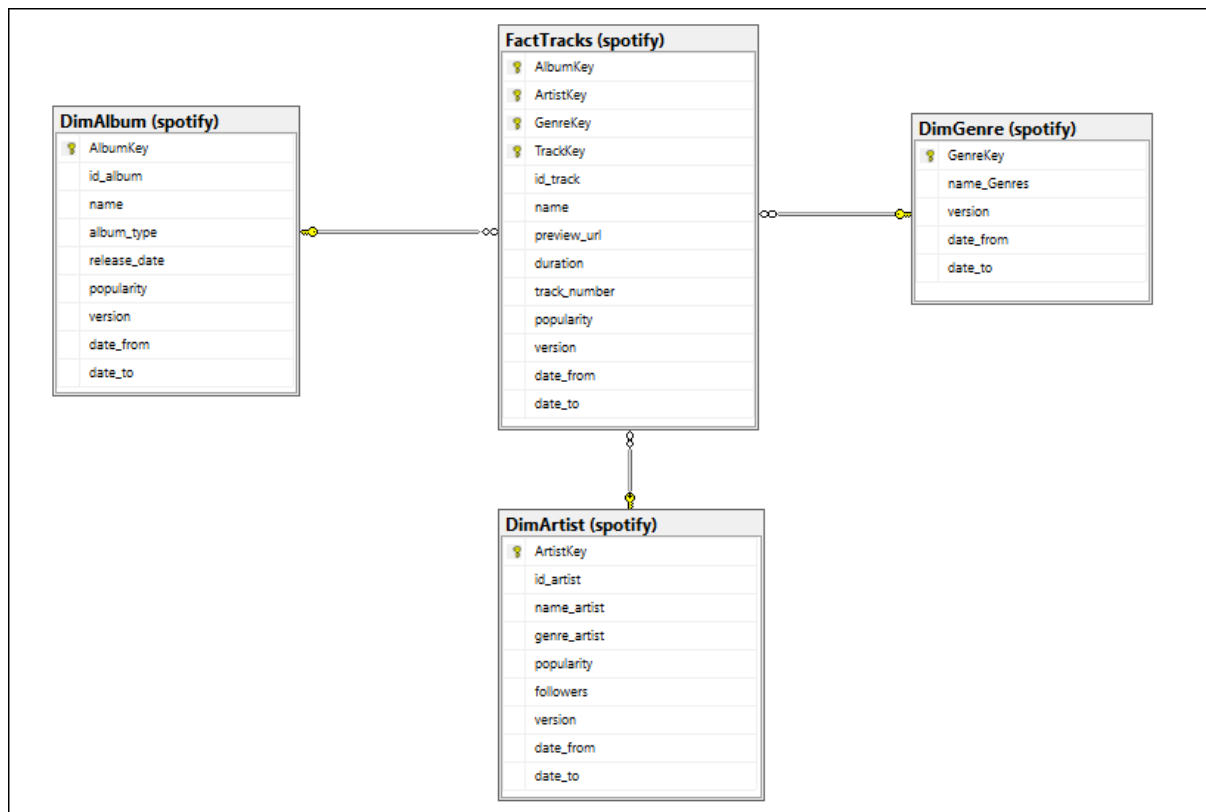
Gambar 11 Mockup

## 4. Implementasi

Bagian ini berisi ROLAP *schema*, kode program pada ETL, basis data MOLAP, dan *Business Intelligence Front End*.

### 4.1 ROLAP Schema

Berikut merupakan ROLAP Schema dari spotify.

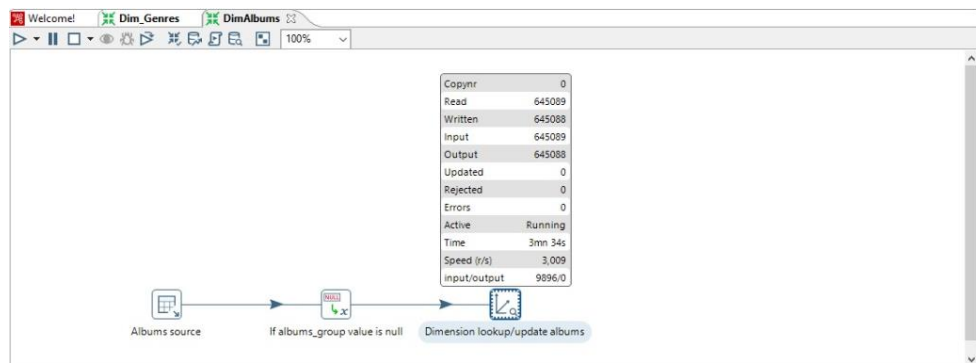


Gambar 12 Rolap

### 4.2 ETL

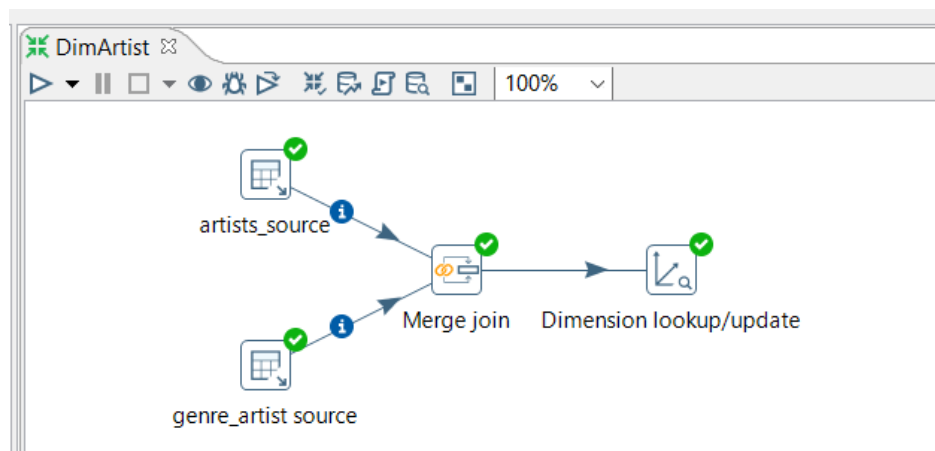
Untuk pembuatan ETL menggunakan Spoon (Pentaho Data Integration). Adapun diagram proses ETL spotify dapat dilihat sebagai berikut.

#### 1.1.1. Dimension Album



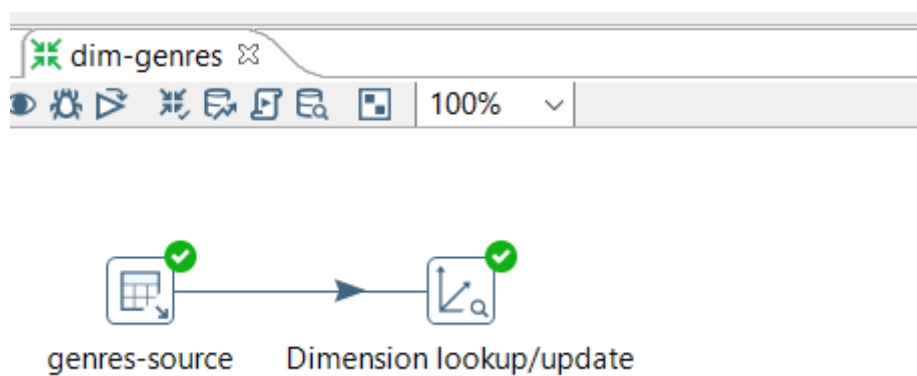
Gambar 13 ETL Dim Albums

### 1.1.2. Dimension Artist



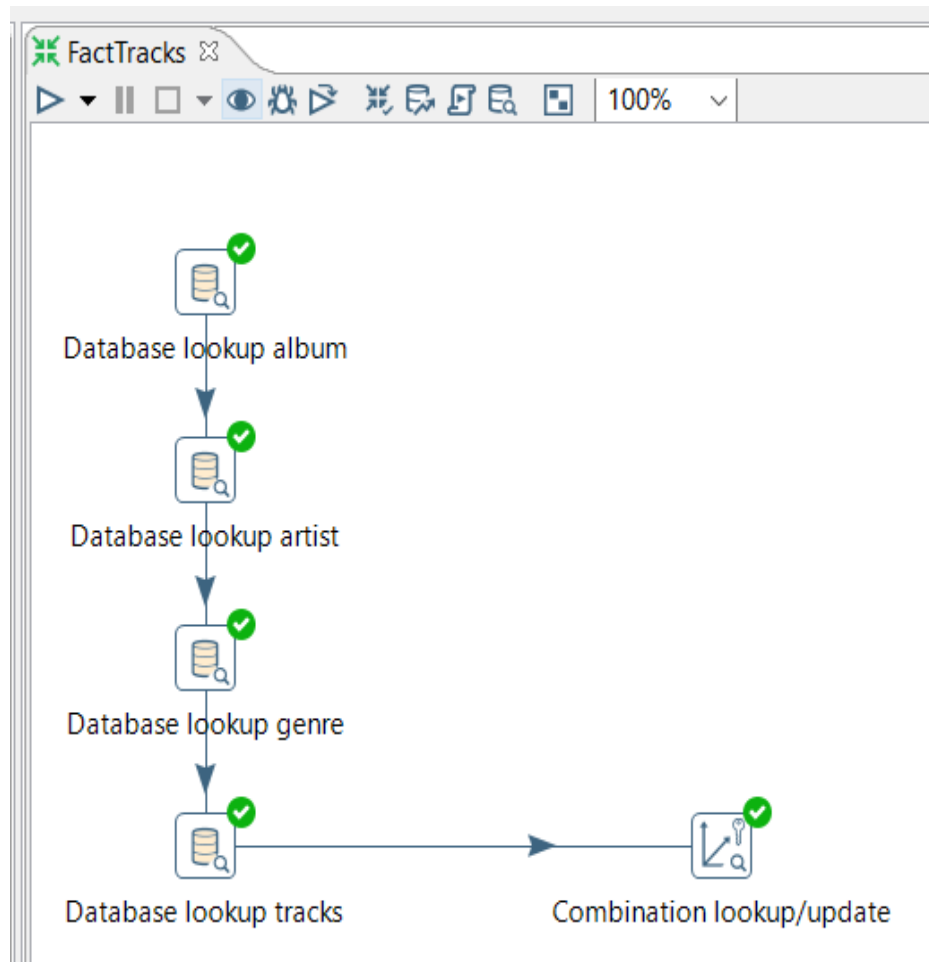
Gambar 14 ETL Dim Artist

### 1.1.3. Dimension Genre



Gambar 15 ETL Dim Genre

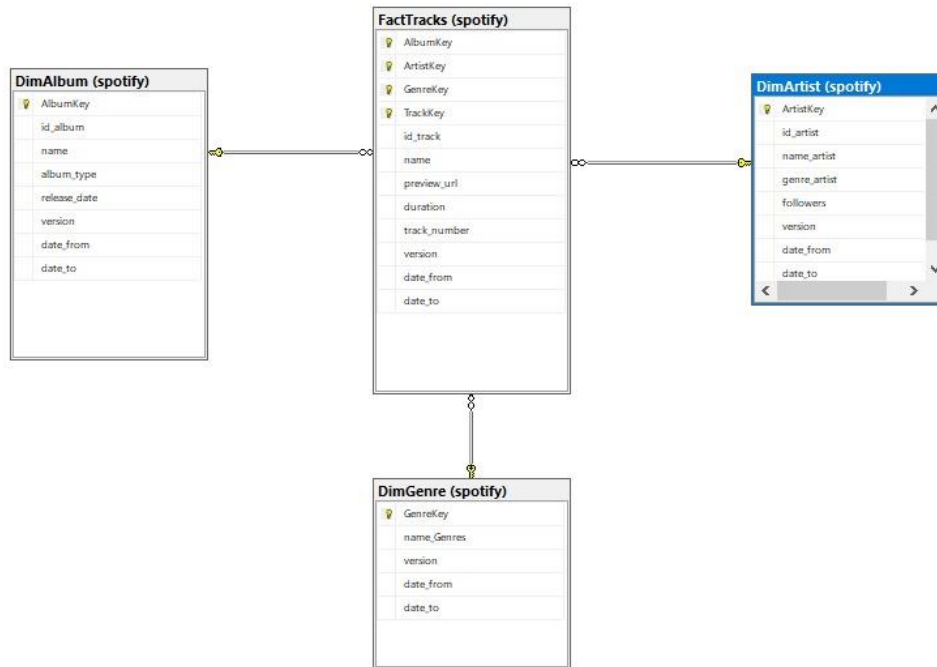
#### 1.1.4. Fact Track



Gambar 16 ETL DIm Fact Tracks

### 4.3 Analysis Services MOLAP Database

Berikut adalah tampilan MOLAP dari proses bisnis Top Track Reporting.

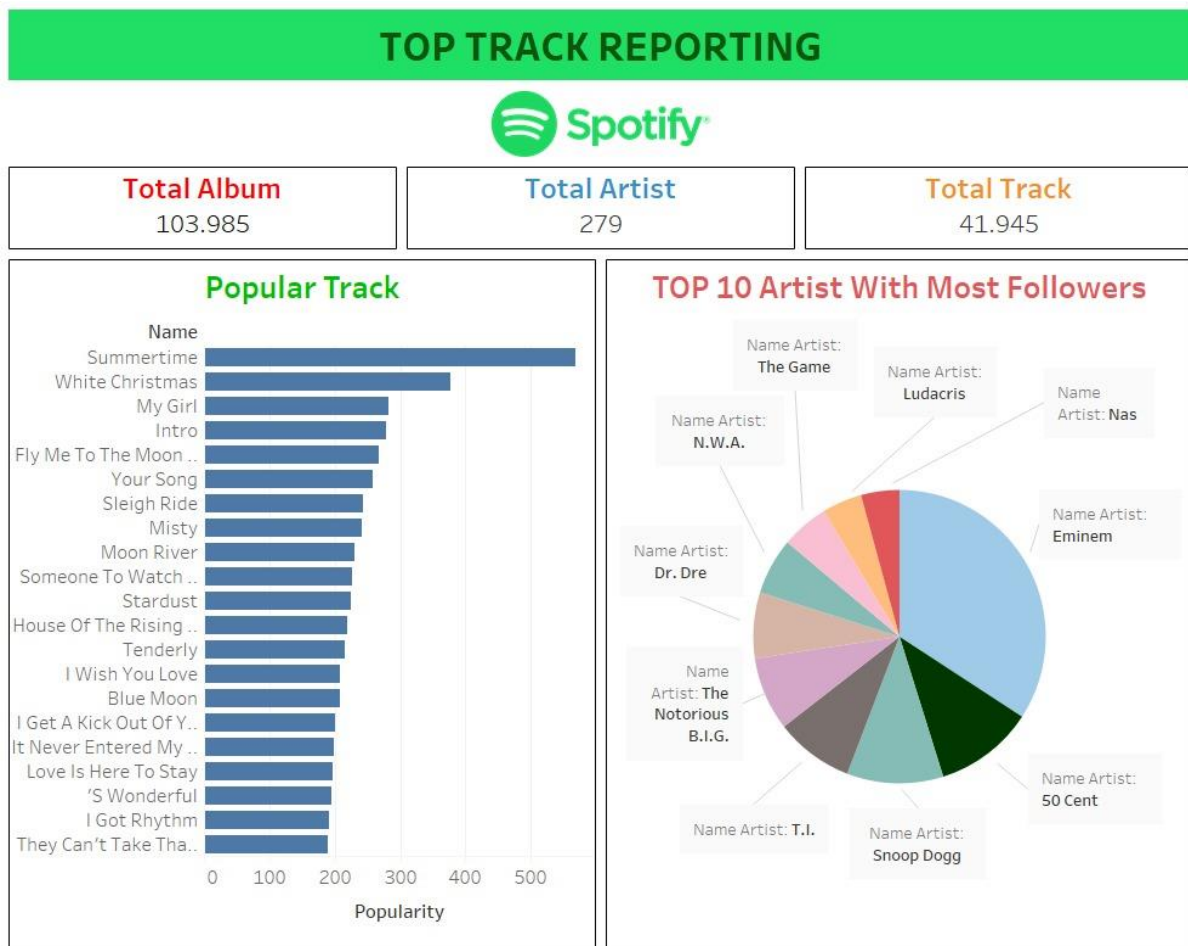


Gambar 17 ETL Molap Database



#### 4.4 Dashboard

Berikut adalah tampilan dashboard dari proses bisnis Top Track Reporting menggunakan Tableau. Pada dashboard ini ditampilkan Total Album, Total Artist, Total Track, Popular Track, dan Top 10 Artist With Most Followers.



Gambar 18 Dashboard Top Track Reporting

## 5. Penutup

### 5.1. Kesimpulan

Perancangan sistem data mart dan dashboard pada Spotify memiliki banyak proses bisnis yang dapat dibangun. Salah satu proses bisnis yang kami pilih adalah *Top Track Reporting* yang merupakan pengindentifikasian data lagu terpopuler yang disatukan pada tangga lagu. Pada proses bisnis ini, data pada *Top Track* akan selalu berubah oleh karena itu kami memakai *Periodic Snapshot* sebagai *Fact Grain Type*. Proses bisnis ini sudah selesai kami bangun dengan tampilan dashboard yang telah kami dokumentasikan sehingga masalah pada latar belakang ini sudah tersolusikan.

### 5.2. Saran

Saran dari tim kami yaitu agar menggunakan database yang lebih sederhana untuk memudahkan dalam pengolahan data dan pembuatan dashboard.

## **LAMPIRAN**