**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 Hutauruk
3. 12S19020 Imelda Siregar
4. 12S19051 Corri Hutahaean
5. 12S19052 Mulyani Gabe Sayoni Simanjuntak

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**2022**

# DAFTAR ISI

[DAFTAR ISI 1](#_Toc103813344)

[1. Pendahuluan 3](#_Toc103813345)

[1.1. Latar Belakang 3](#_Toc103813346)

[1.2. Ruang Lingkup 3](#_Toc103813347)

[1.3. Tim Pengembang 3](#_Toc103813348)

[2. Analisis 7](#_Toc103813349)

[2.1. Proses Bisnis 7](#_Toc103813350)

[2.2. Atribut dan Metrik 7](#_Toc103813351)

[2.3. Isu 9](#_Toc103813352)

[3. Desain 10](#_Toc103813353)

[3.1. Detail Bus Matrix 10](#_Toc103813354)

[3.2. Detailed Dimensional Modeling 10](#_Toc103813355)

[3.2.1. Dimensi Albums 10](#_Toc103813356)

[3.2.2. Dimensi Artists 11](#_Toc103813357)

[3.2.3. Dimensi Genres 11](#_Toc103813358)

[3.2.4. Fact Tracks 12](#_Toc103813359)

[3.3. Dimensional Hierarchies 13](#_Toc103813360)

[3.3.1. Dimensi Albums 13](#_Toc103813361)

[3.3.2. Dimensi Artists 13](#_Toc103813362)

[3.3.3. Dimensi Genres 13](#_Toc103813363)

[3.4. Dimensional Data Model 13](#_Toc103813364)

[3.5. High-Level Source to Target Map 14](#_Toc103813365)

[3.6. Detailed ETL Flow for Each Source to Target 14](#_Toc103813366)

[3.7. Business Intelligence Front End Mockup 15](#_Toc103813367)

[4. Implementasi 16](#_Toc103813368)

[4.1. ROLAP Schema 16](#_Toc103813373)

[4.2. ETL 16](#_Toc103813374)

[4.3. Analysis Services MOLAP Database 16](#_Toc103813375)

[4.4. Dashboard 17](#_Toc103813376)

[5. Penutup 18](#_Toc103813377)

[5.1. Kesimpulan 18](#_Toc103813378)

[5.2. Saran 18](#_Toc103813379)

[LAMPIRAN 19](#_Toc103813380)

# 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.

## 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.*

## Ruang Lingkup

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

1. Menggunakan dummy data 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.

## Tim Pengembang

Pada Tabel 1 disajikan susunan tim pengembang pada proyek ini.

**Tabel 1. Susunan Tim Pengembang**

| No. | Nama | Peran | Tanggung Jawab |
| --- | --- | --- | --- |
|  | Kartika Sianipar | Analyst | Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses pencatatan registrasi akun |
| BI Architect | Membangun BI dari data yang dimiliki untuk bisnis proses pencatatan registrasi akun |
| Data Architect | Membangun dimensional modelling dan implementasi |
|  | Agnes Abigael Hutauruk | Analyst | Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses pencatatan registrasi akun |
| BI Architect | Membangun BI dari data yang dimiliki untuk bisnis proses pencatatan registrasi akun |
| Data Architect | Membangun dimensional modelling dan implementasi |
|  | Imelda Siregar | Analyst | Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses pencatatan registrasi akun |
| BI Architect | Membangun BI dari data yang dimiliki untuk bisnis proses pencatatan registrasi akun |
| Data Architect | Membangun dimensional modelling dan implementasi |
|  | Corri Hutahaean | Analyst | Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses pencatatan registrasi akun |
| BI Architect | Membangun BI dari data yang dimiliki untuk bisnis proses pencatatan registrasi akun |
| Data Architect | Membangun dimensional modelling dan implementasi |
| 5 | Mulyani Simanjuntak | Analyst | Bertanggung jawab menganalisis dan mendesain kebutuhan BI untuk bisnis proses pencatatan registrasi akun |
| BI Architect | Membangun BI dari data yang dimiliki untuk bisnis proses pencatatan registrasi akun |
| Data Architect | Membangun dimensional modelling dan implementasi |

# 

# Analisis

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

## Proses Bisnis

* + 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 kepopularitasan lagu seperti judul track, album track, genre, dan jumlah streaming track tersebut.

## Atribut dan Metrik

* + 1. Atribut dan Metrik [BP01]

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

|  |  |  |  |
| --- | --- | --- | --- |
| **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 |
| 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 |
| 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 | <https://p.scdn.co/mp3-preview/770e023eb0318270ecc5caa018d758e5e0844de9?cid=cde021ca5d3e42a8bd440f100...> |
| track\_number | merupakan nomor dari track yang sedang diputar | 13 |
| Genres | genreskey | Merupakan kode unik (*Primary Key*) yang menjadi identitas utama dari Genre | 84231 |
| id\_genres | Merupakan kode mendefinisikan identitas dari Genre tersebut | 6Wnd4kk021 |
| name\_genres | Merupakan nama atau tipe dari jenis genre | KPOP |

## Isu

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

|  |  |  |  |
| --- | --- | --- | --- |
| 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 |

# 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*.

## Detail Bus Matrix

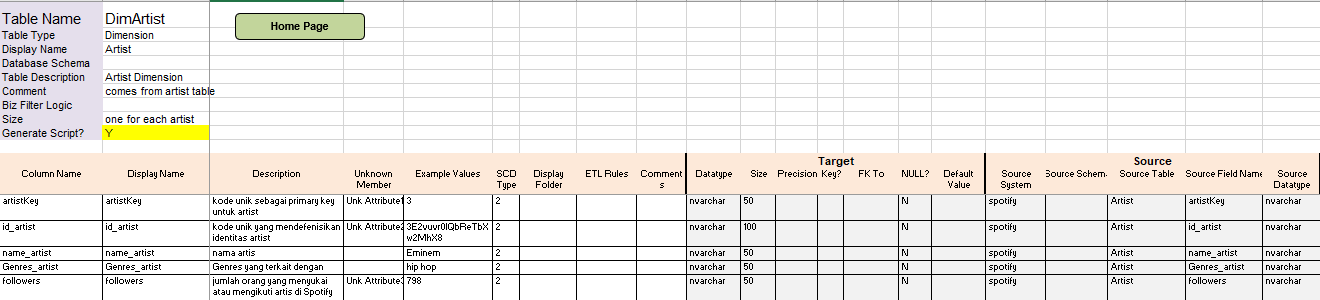
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Business Process Name | Fact Table | Fact Grain Type | Granularity | Facts | albums | artists | genres |
| Top track reporting | FactTracks | Periodic Snapshot | one row per popularity track | Most streaming track, album with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

## Detailed Dimensional Modeling

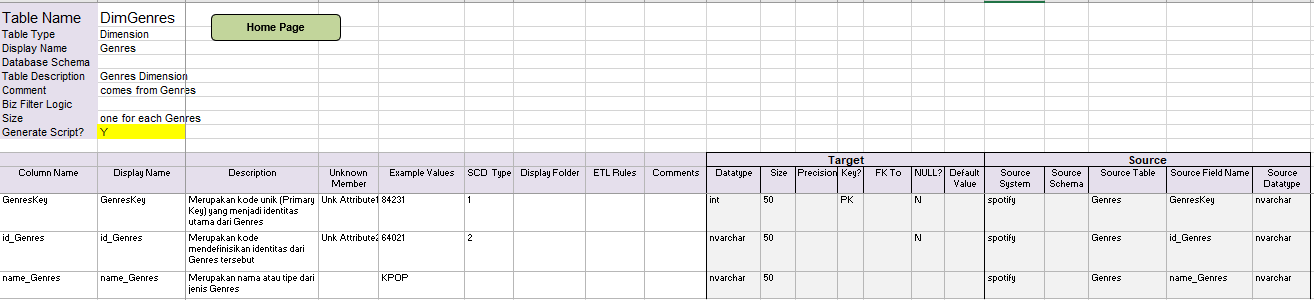
### Dimensi Albums



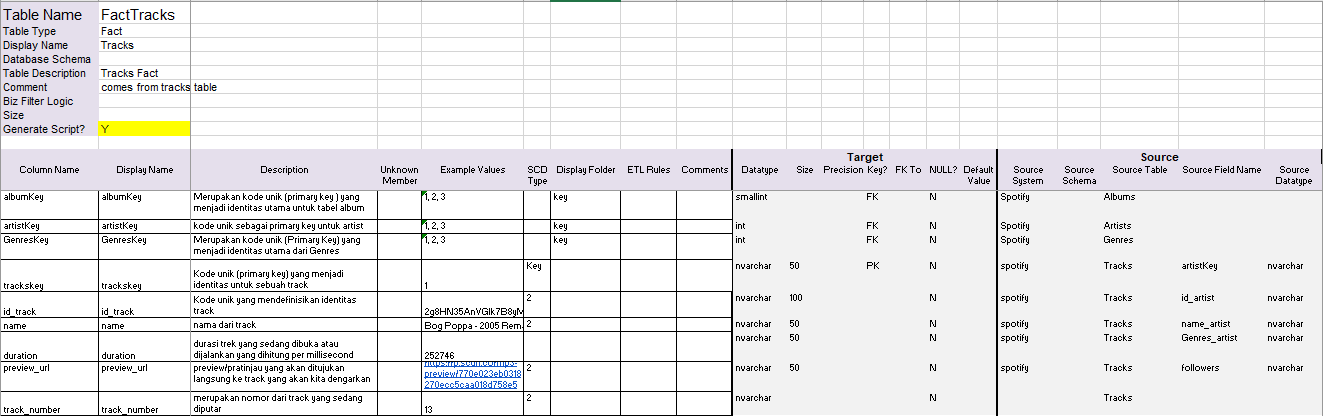
### Dimensi Artists



### Dimensi Genres

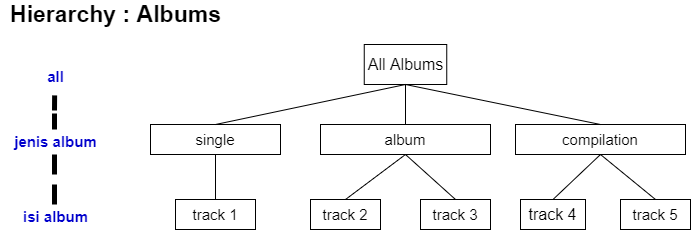


### Fact Tracks

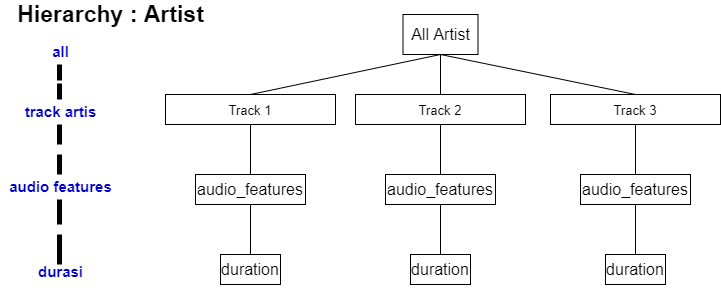


## Dimensional Hierarchies

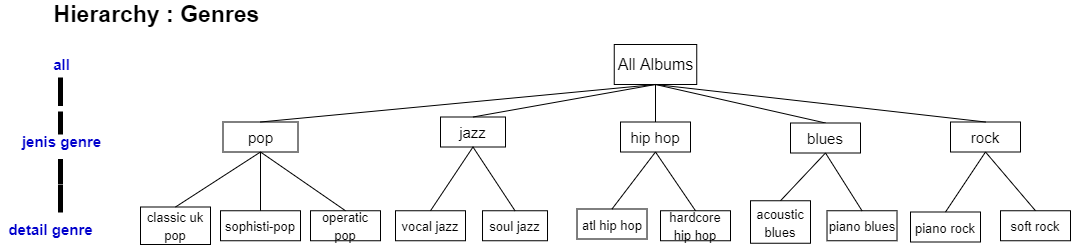
### Dimensi Albums



### Dimensi Artists



### Dimensi Genres

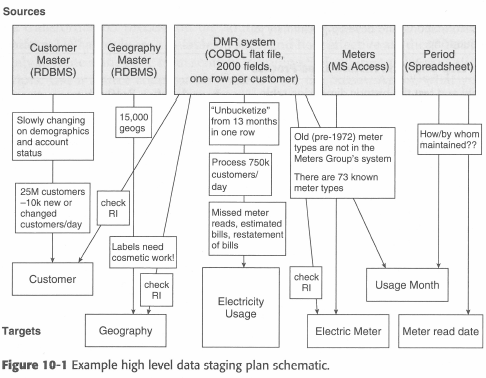


## Dimensional Data Model

<Gambarkan *schema* model dimensional.>

## High-Level Source to Target Map

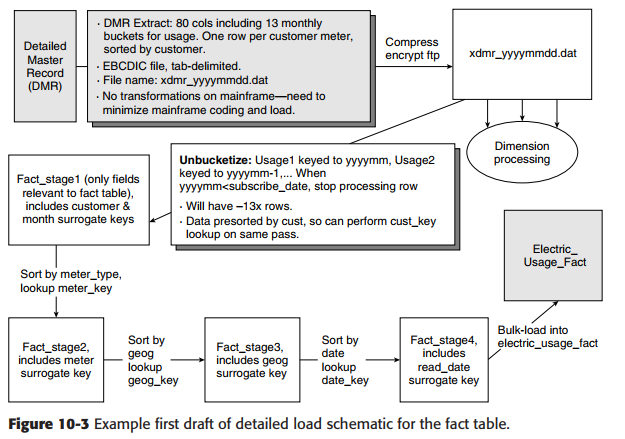
<Gambarkan *high-level source to target map*. Contoh dapat dilihat pada Gambar 1.>



**Gambar 1. Diagram Tingkat Tinggi Sumber ke Target**

## Detailed ETL Flow for Each Source to Target

<Tuliskan aliran ETL secara rinci untuk setiap aliran sumber ke target. Contoh dapat dilihat pada Gambar 2.>



**Gambar 2. Diagram Rinci ETL DMR ke Tabel Electric\_Usage\_Fact**

## Business Intelligence Front End Mockup

<Gambarkan *mockup BI front end*, misalnya *dashboard* beserta panel-panel yang sesuai dengan bisnis proses yang telah didefinisikan.>

# Implementasi

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

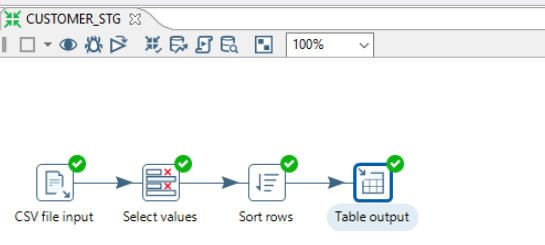


## ROLAP Schema

<Tampilkan cuplikan ROLAP *schema* yang menunjukkan *fact tables*, *dimension tables*, *star/snowflake/galaxy schema diagram*, dan PK/FK.>

## ETL

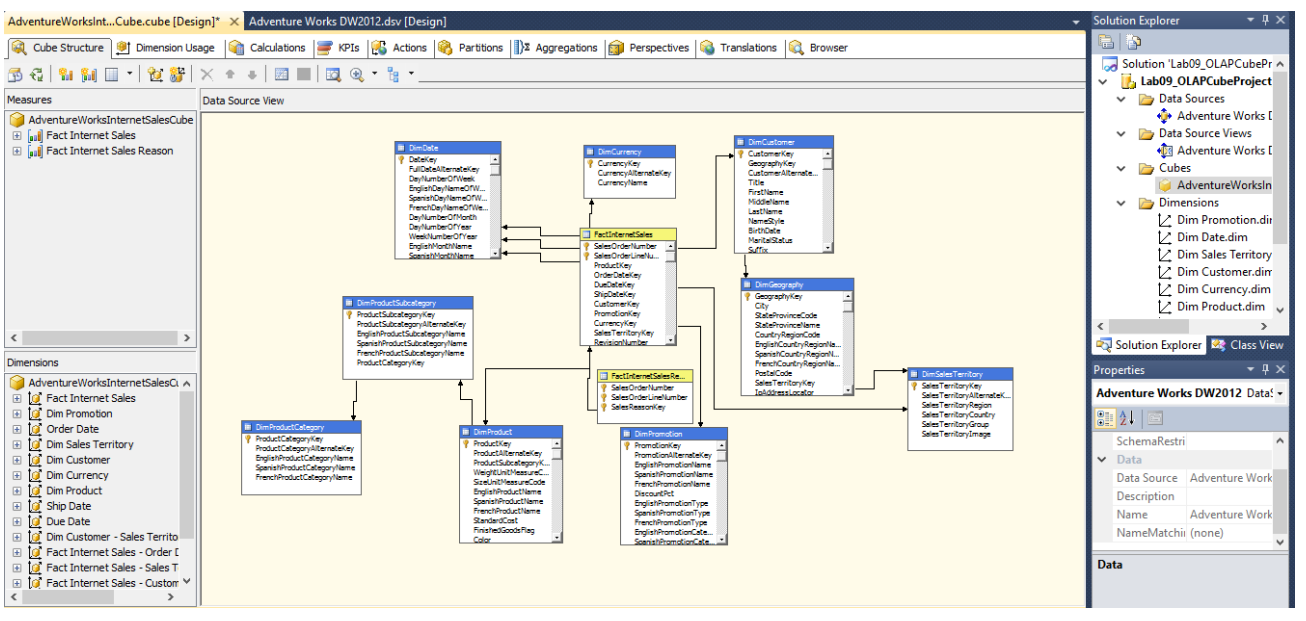
<Tuliskan kode program (atau diagram proses apabila menggunakan **SQL Server Integration Services** ataupun **Spoon**) yang terlibat dalam proses ETL. Misalnya, kode program untuk *populates ROLAP schema from sources*, kode program untuk *initial load*, dan kode program untuk *subsequent loads*. Contoh dapat dilihat pada Gambar 3.>



**Gambar 3. Cuplikan ETL dari Sumber ke Tabel Staging Customer**

## Analysis Services MOLAP Database

<*Cubes to match star schemas / business processes, dimensional hierarchies configured, facts / KPI’s as appropriate.* Contoh dapat dilihat pada Gambar 4.>



**Gambar 4. Cuplikan Internet Sales Cube**

## Dashboard

<Tampilkan cuplikan *dashboard* serta beri penjelasan penggunaannya seperti apa.>

# Penutup

## Kesimpulan

<Tuliskan apakah sistem selesai dibangun dan apakah dapat menyelesaikan masalah yang ditulis pada latar belakang.>

## Saran

Tuliskan saran pengembangan.>

# LAMPIRAN

<Opsional.>