**LAPORAN PROYEK MATA KULIAH**

**12S3202 – GUDANG DATA DAN KECERDASAN BISNIS**

**Pembuatan Sistem Data Mart dan Dashboard**

**pada Spotify**



**Disusun oleh:**

| **12S18001** | **Cindy Angelia Siregar** |
| --- | --- |
| **12S18014** | **Giovanni Situmorang** |
| **12S18017** | **Putri Yohana Panjaitan** |
| **12S18049** | **Natasya Sitorus** |
| **12S18061** | **Angela Friscilia Simamora** |

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**APRIL 2022**

# DAFTAR ISI

[DAFTAR ISI i](#_heading=h.gjdgxs)

[1.](#_heading=h.30j0zll) Pendahuluan 1

[2.](#_heading=h.3znysh7) Analisis 2

[3.](#_heading=h.2et92p0) Desain 3

[4.](#_heading=h.1t3h5sf) Implementasi 5

[5.](#_heading=h.17dp8vu) Penutup 7

[LAMPIRAN 8](#_heading=h.3rdcrjn)

# Pendahuluan

Bab pendahuluan berisi penjelasan terkait latar belakang pemilihan topik proyek, ruang lingkup proyek dan tim pengembang dalam proyek.

* 1. **Latar Belakang**

Meningkatnya jumlah interaksi sosial publik terjadi melalui komputer, dimana semuanya dibagikan mulai dari lelucon, cerita, berita dan terutama dunia industri musik. Salah satu layanan aplikasi media sosial yang menawarkan musik kepada pengguna adalah *spotify. Spotify* adalah salah satu layanan musik digital. podcast, dan video yang memberikan akses jutaan lagu dari kreator di seluruh dunia. *Spotify* menawarkan musik kepada pengguna, dimana pengguna dapat mencari musik berdasarkan artis, album, atau genre serta pengguna dapat mengedit dan berbagi daftar putar.

*Spotify* juga memiliki beberapa fitur salah satunya adalah dapat menambah lagu sebanyak mungkin dan juga dari lagu-lagu yang bertambah, maka artis atau penyanyi juga akan bertambah. Akan tetapi seiring perkembangan lagu, maka data mengenai lagu juga semakin banyak dan biasanya *spotify* akan menyimpan lagu dan artis tersebut dalam waktu yang lama. Kumpulan data yang begitu banyak mengandung informasi mengenai lagu dan juga artis atau penyanyi yang dapat dianalisis. Saat ini data yang digunakan untuk merekapitulasi data *spotify* sudah ada, akan tetapi karena banyaknya fitur yang terdapat pada *spotify*, maka diperlukan sistem yang akan mengintegrasi dan dapat merangkum data-data tersebut.

Salah satu pendekatan yang dapat digunakan dalam proses mengintegrasikan data adalah dengan membuat *data mart*. Dalam mengintegrasikan data, proses yang akan dilakukan adalah mengumpulkan data dari sumber dan diolah melalui proses ETL, dan setelah itu akan data akan digabungkan dalam satu penyimpanan. Data yang sudah diintegrasikan ke dalam satu penyimpanan akan ditampilkan dalam bentuk *dashboard* yang bertujuan untuk mempermudah proses pemantauan. *Dashboard* merupakan salah satu penerapan kecerdasan bisnis, yang digunakan sebagai sistem yang menampilkan informasi visual yang mudah dipahami oleh pengguna berdasarkan data-data yang sudah dikumpulkan dan diolah.

Pada pengerjaan proyek ini berfokus dalam pengerjaan menyaring informasi popularitas artis atau penyanyi pada setiap lagu di *Spotify* yang dapat difilter setiap minggu ataupun setiap bulan. Banyaknya jumlah musik yang dipublikasikan oleh para musisi dunia juga membuat pengelolaan terhadap basis data sangat diperlukan dalam menentukan kepopuleran artisnya. Oleh karena itu, pada pengerjaan proyek ini dilakukan analisis dengan membuat sistem data mart dan dashboard untuk *Spotify* supaya setiap informasi dapat diintegrasikan dengan baik.

* 1. **Ruang Lingkup**

Adapun ruang lingkup pada proyek ini diantaranya adalah

1. Dataset yang digunakan adalah dataset spotify.sqlite
2. Alat ETL yang digunakan pada proyek ini adalah pentaho(spoon) untuk melakukan proses ETL dan PostgreSQL untuk DBMS.
   1. **Tim Pengembang**

Pada Tabel 1 disajikan susunan tim pengembang pada proyek ini.

**Tabel 1. Susunan Tim Pengembang**

| No. | Nama | Peran | Tanggung Jawab |
| --- | --- | --- | --- |
|  | Cindy Angelia Siregar | Ketua | Desain |
|  | Giovanni Situmorang | Anggota | Implementasi |
|  | Putri Yohana Panjaitan | Anggota | Implementasi |
|  | Natasya Sitorus | Anggota | Desain |
| *5* | Angela Friscilia Simamora | Anggota | Desain |

# Analisis

Bab analisis berisi penjelasan terkait bisnis proses pada proyek, atribut dan metrik dan isu pada proyek

* 1. **Bisnis Proses**

Pada proyek ini proses bisnis yang akan dibangun adalah pelaporan artis terpopuler. Dimana proses bisnis ini bertujuan untuk menyaring informasi popularitas artis atau penyanyi pada setiap lagu di *Spotify* yang dapat difilter setiap minggu ataupun setiap bulan.

* 1. **Atribut dan Metrik**

Pada bagian ini akan ditampilkan daftar dimensi dan atribut yang dipakai pada proyek ini yaitu yang berhubungan dengan pelaporan artis terpopuler. Dimensi yang digunakan adalah *albums, artist, audio\_features, genres. tracks*.

| **Dimension Table** | **Attribute/Fact Name** | **Description** |
| --- | --- | --- |
| albums | id | id dari album |
|  | name | nama album yang digunakan |
|  | album\_group | group dari album |
|  | album\_type | tipe album yang digunakan |
|  | release\_date | tanggal rilis album |
|  | popularity | popularitas dari album |
| artists | name | nama dari artis |
|  | id | id artis |
|  | popularity | popularitas artis |
|  | followers | pengikut artis |
| audio\_features | id | id dari fitur audio |
|  | acousticness | fitur akustik dalam audio |
|  | analysis\_url | analisis link url untuk fitur audio |
|  | danceability | genre musik dance untuk fitur audio |
|  | duration | durasi atau lama waktu audio |
|  | energy | kekuatan audio fitur |
|  | instrumentalness | instrumental yang ada pada audio |
|  | key | kunci fitur audio |
|  | liveness | pendeteksi hidup atau mati dari fitur audio |
|  | loudness | kuat dari audio |
|  | mode | mode yang digunakan dalam audio |
|  | speechiness | speechiness dari fitur audio |
|  | tempo | tempo dari audio |
|  | time\_signature | waktu untuk menandakan beat audio |
|  | valence | valensi dalam fitur audi |
| genres | id | id dari genre |
| tracks | id | id dari track |
|  | disc\_number | nomor dari disc yang diputar |
|  | duration | durasi dari track |
|  | explicit | kejelasan track yang berjalan |
|  | audio\_feature\_id | id fitur audio |
|  | name | nama track |
|  | preview\_url | link url yang sebelumnya diputar |
|  | track\_number | nomor dari track |
|  | popularity | popularitas track |
|  | is\_playable | keterangan bahwa track dapat dimainkan |

* 1. **Isu**

| No | *Task/Topic* | *Issue* | *Reported By* | *Responsible* | Status |
| --- | --- | --- | --- | --- | --- |
| 1. | Menentukan Topik Proyek | Menentukan bisnis proses apa saja yang dibuat dalam proyek. | Semua anggota | Semua Anggota | Solved |
| 2. | *Detailed Dimensional Modeling* | Menentukan *Detailed Dimensional Modeling* | Semua anggota | Semua Anggota | Solved |
| 3. | ETL process | Menentukan proses ETL setiap dimensional yang ada | Semua anggota | Semua Anggota |  |
| 4 . |  |  |  |  |  |

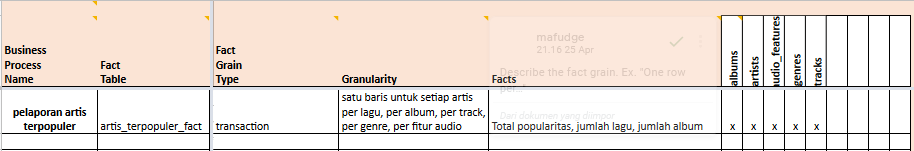
# 

# Desain

Bab 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*.

* 1. ***Detailed Bus Matrix***

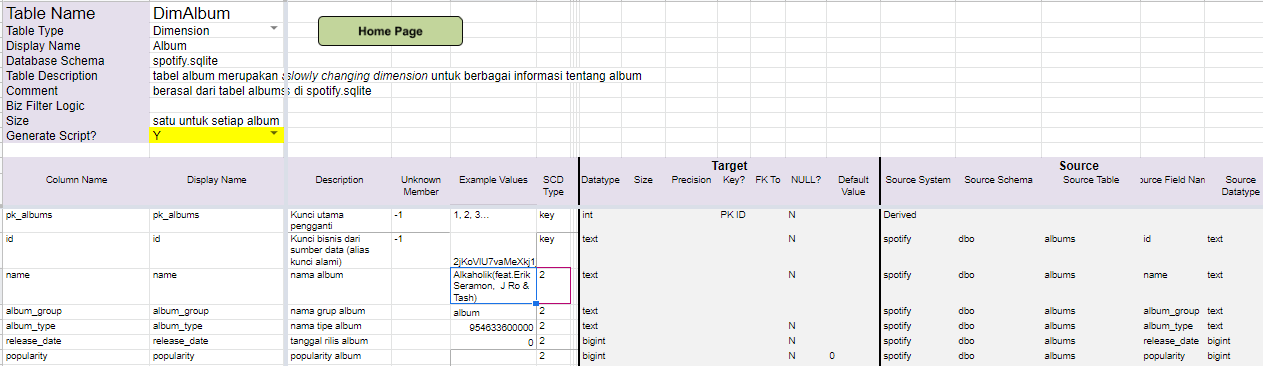
Pada bagian ini adalah menampilkan *detail bus matrix* yang ada pada proyek, dimana bisnis proses yang digunakan adalah pelaporan artis terpopuler, tabel faktanya adalah artis\_terpopuler, *fact grain type* adalah transaksi, kemudian *granularity* pada proses bisnis ini adalah satu baris untuk setiap artis per lagu, per album, per track, per genre, per fitur audio dan faktanya adalah total popularitas, jumlah lagu, jumlah artist dan jumlah album.

****

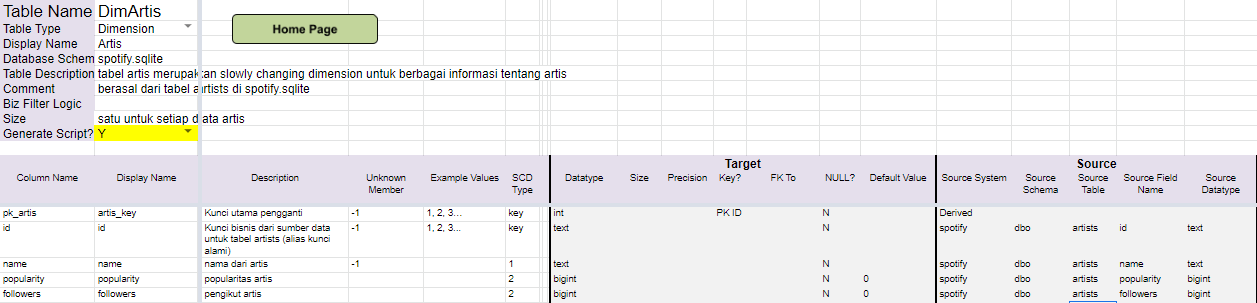
* 1. ***Detailed Dimensional Modeling***

Pada bagian ini adalah menampilkan *dimensional modeling* dari proses bisnis pelaporan artis terpopuler, dimana terdapat lima dimensi diantaranya adalah *albums, artist, audio\_features, genres, tracks*.

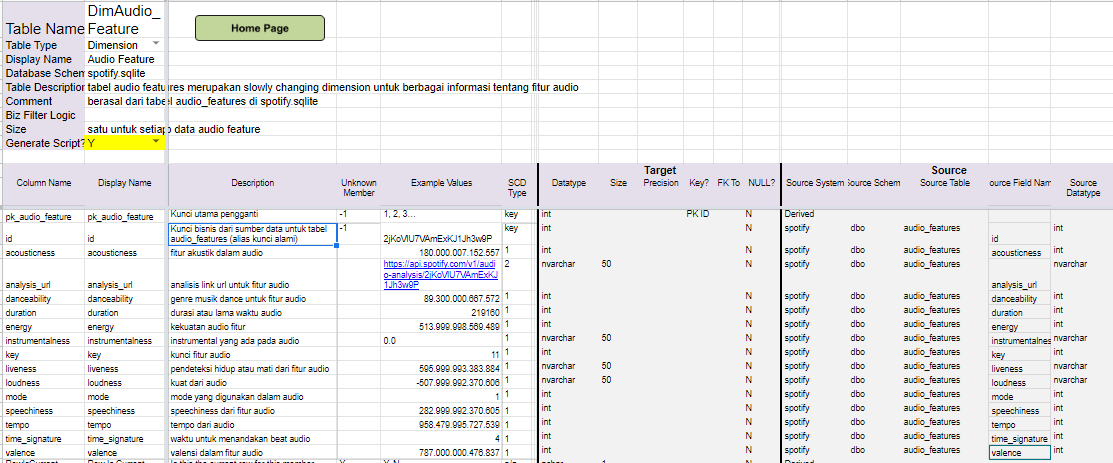
**3.2.1. Dimensi Albums**

****

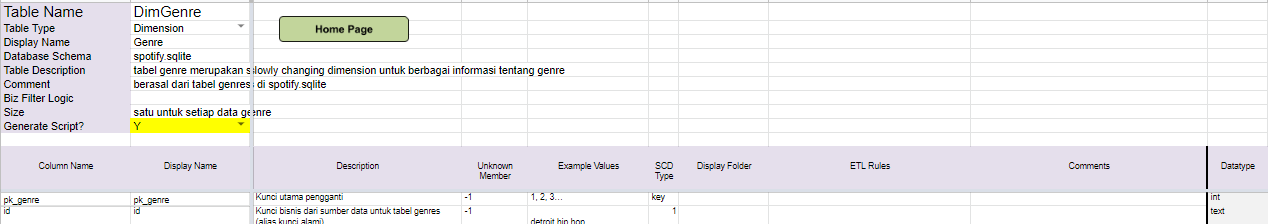
**3.2.2. Dimensi Artist**

****

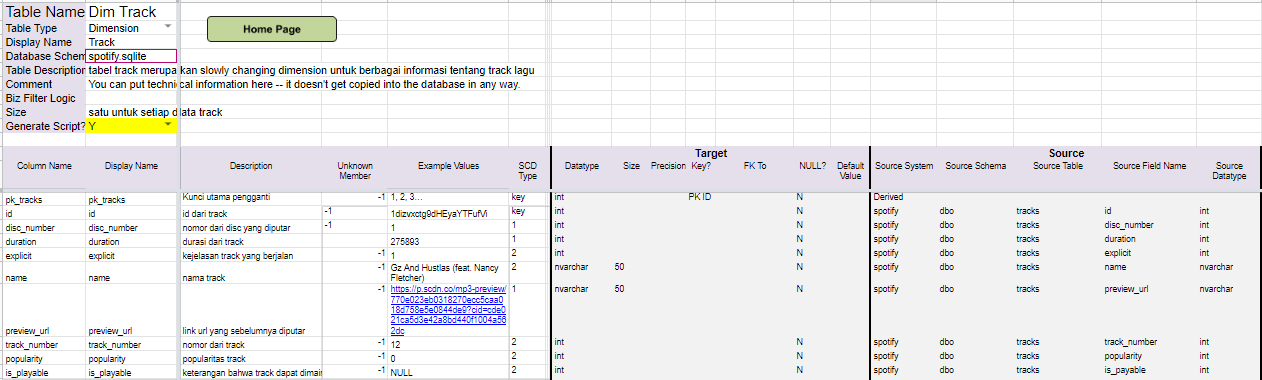
**3.2.3. Dimensi Audio Features**

****

**3.2.4. Dimensi Genres**

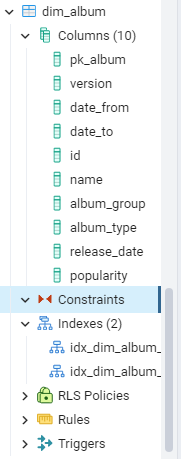
****

**3.2.5. Dimensi Tracks**

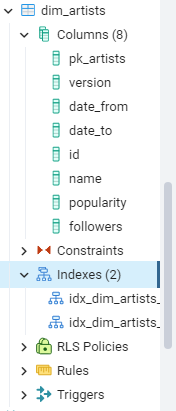
****

* 1. ***Dimensional Hierarchies***

**3.2.1. Dimensi Albums**

****

**3.2.2. Dimensi Artist**

****

**3.2.3. Dimensi Audio Features**

**3.2.4. Dimensi Genres**

**3.2.5. Dimensi Tracks**

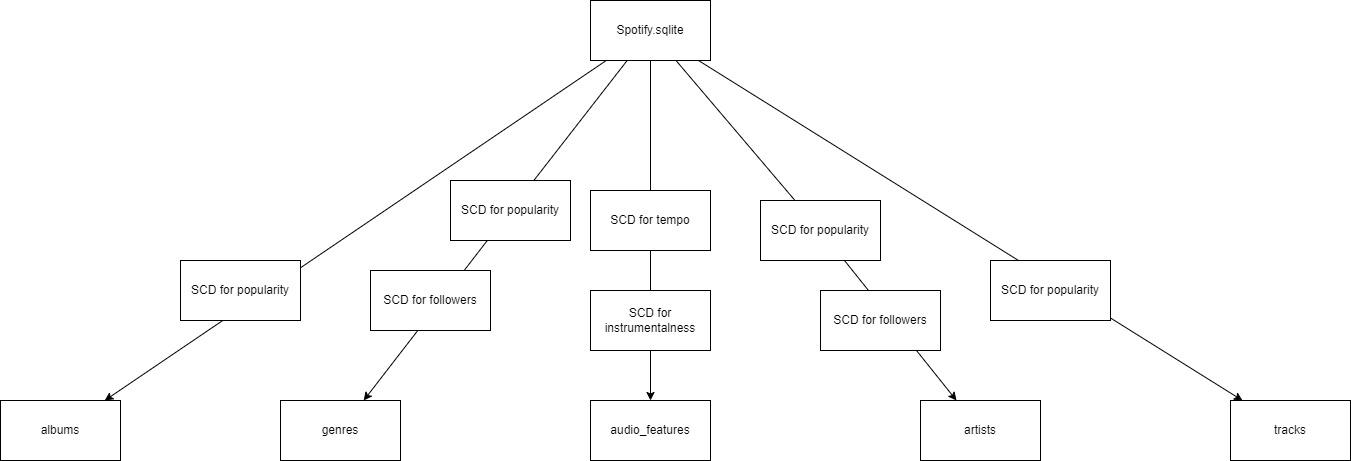
<Gambarkan hirarki pada setiap tabel dimensi.>

* 1. ***Dimensional Data Model***

<Gambarkan *schema* model dimensional.>

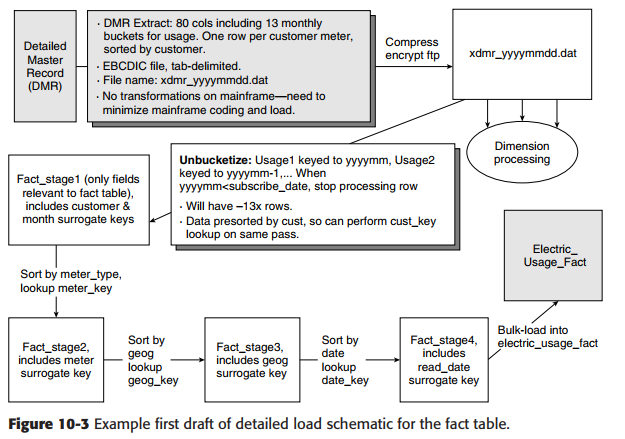
* 1. ***High-Level Source to Target Map***

Pada bagian ini adalah menampilkan *high level plan* yang ada pada bisnis proses, yaitu terdapat source to target. Yang menjadi *source* pada proyek ini adalah spotify.sqlite dan untuk target terdapat lima tabel, diantaranya adalah *albums, artists, audio\_features, genres, dan tracks*.

****

* 1. ***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**

* 1. ***Business Intelligence Front End Mockup***

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

# Implementasi

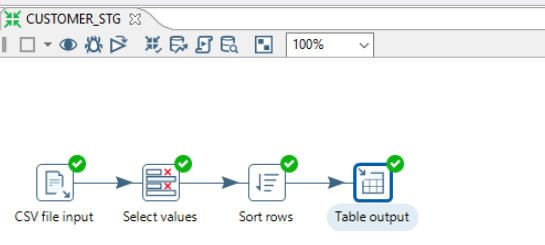
Bab ini berisi ROLAP *schema*, kode program pada ETL, basis data MOLAP, dan *Business Intelligence Front End*.

* 1. ***ROLAP Schema***

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

* 1. ***ETL***

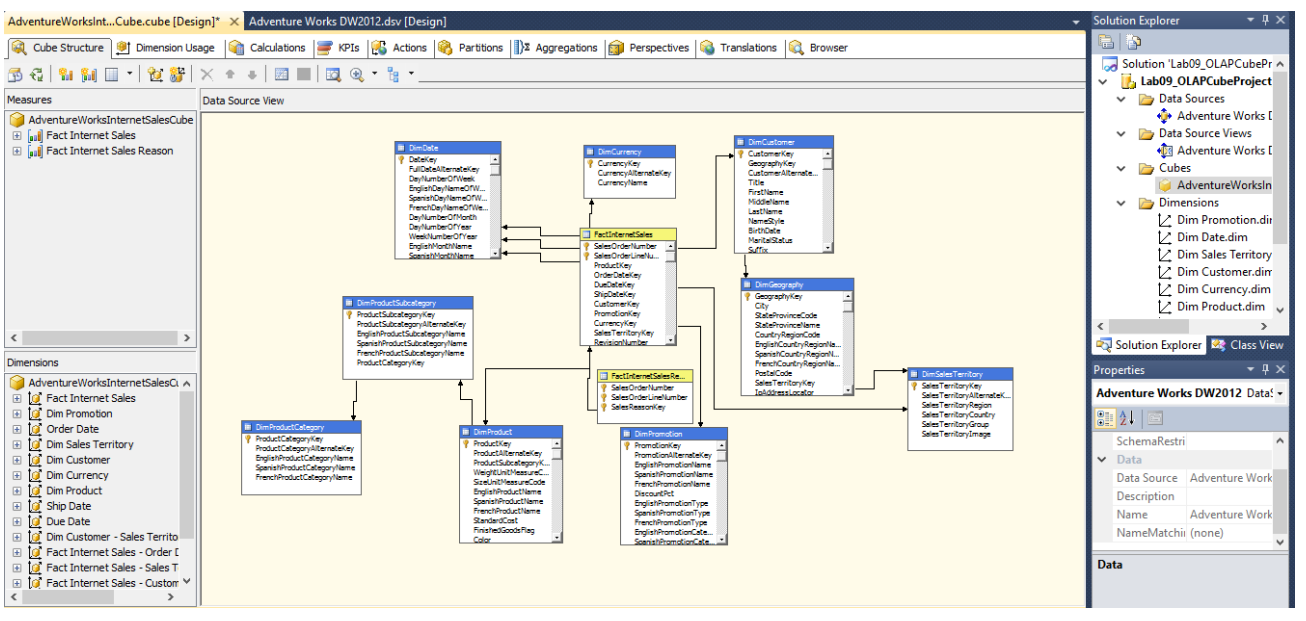
<Tuliskan kode program (atau diagram proses apabila menggunakan **SQL Server Integration Services** ataupun **Spoon**) yang terlibat dalam proses ETL. Misalnya, kode program untuk *populates ROLAP schema from sources*, kode program untuk *initial load*, dan kode program untuk *subsequent loads*. Contoh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



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

* 1. ***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**

* 1. ***Dashboard***

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

# Penutup

* 1. **Kesimpulan**

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

* 1. **Saran**

Tuliskan saran pengembangan.>

# LAMPIRAN

<Opsional.>