

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

SPOTIFY



Disusun oleh:

1. 12S19022 Nico Rajagukguk
2. 12S19027 Gilbert Hutajulu
3. 12S19029 Yosua Batubara
4. 12S19030 Jaime Christ Bonar Sirait

FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO
INSTITUT TEKNOLOGI DEL
2021/2022

DAFTAR ISI

DAFTAR ISI.....	i
1. Pendahuluan.....	1
2. Analisis	4
3. Desain	6
4. Implementasi	11
5. Penutup	14
LAMPIRAN.....	15

1. Pendahuluan

Bagian ini berisi latar belakang dan tujuan pengerjaan proyek.

1.1. Latar Belakang

Musik sepertinya sudah menjadi bagian dari keseharian banyak orang. Saat ini mendengarkan musik bisa dibilang menjadi salah satu sarana untuk menghilangkan kebosanan maupun menunjang aktivitas dalam kondisi tertentu. Smartphone menjadi salah satu sarana pilihan terbaik untuk mendengarkan musik karena selain sangat efektif juga sangat efisien untuk penggunaannya. Agar bisa mendengarkan lagu yang diinginkan beberapa orang mungkin akan membeli lagu tersebut secara digital, tidak sedikit juga yang mengunduhnya secara illegal downloading diberbagai situs website karena memiliki keuntungan yaitu gratis dan akses yang sangat mudah. Namun, sekarang sudah banyak aplikasi streaming musik di iOS/Android yang menyediakan banyak lagu untuk diputar secara gratis. Setiap harinya musik mengalami perkembangan, mulai dari semakin banyaknya genre (aliran musik), kemajuan teknologi alat musik hingga kualitas suara/rekaman yang semakin hari semakin baik. Sama halnya musik yang mengalami perkembangan, teknologi yang berhubungan dengan musik juga ikut berkembang, salah satunya aplikasi musik online.

Spotify merupakan aplikasi streaming musik yang ada di iOS/Android. Pemilik dari aplikasi Spotify ini berasal dari salah satu perusahaan yang ada di Swedia. Spotify sendiri diluncurkan pada tahun 2008 dan sudah mendukung lebih dari 70 bahasa. Selain itu, Spotify memiliki dua tipe streaming yaitu tipe Free dan Premium. Tentunya tipe Premium lebih banyak mendukung fitur seperti peningkatan kualitas audio dan juga fitur download untuk dimainkan secara offline. Selain bisa digunakan untuk perangkat iOS/Android, Spotify juga bisa digunakan untuk desktop seperti PC atau laptop.¹ Spotify menjadi salah satu “perpustakaan” musik terbesar yang menyediakan berbagai musik dari berbagai zaman, genre (aliran musik) dan artis. Jika kita berlangganan paket premium, kita dapat menikmati fasilitas ekstra yang tidak diberikan kepada pengguna biasa atau non-premium. Salah satu keuntungannya adalah kita dapat melewati (skip) lagu dan kita dapat menggunakannya pada waktu offline atau saat tidak adanya jaringan internet.

1.2. Ruang Lingkup

Ruang lingkup menjadi suatu hal yang dibutuhkan dalam pengembangan system sportify yang akan dibuat. Dalam pengembangan ini telah disediakan sumber data dari pihak ketiga

berupa SQLite. Pendekatan pengembangan system menggunakan Konsep Kimball Approach untuk data warehouse dan Konsep Inmon Approach untuk bisnis intelligent.

Nama	Keterangan
Perangkat Lunak Basis Data	Microsoft SQL Server
Perangkat Lunak ETL	Pentaho
Perangkat Lunak Dashboard	Tableau

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	Nico Rajagukguk	Team Leader	<ul style="list-style-type: none"> • Mengecek Setiap progress yang telah dikerjakan. • Membantu jalannya proses. • Mengambil Keputusan.
2	Gilbert Hutajulu	Implementasi	<ul style="list-style-type: none"> • melakukan implementasi pada ROLAP schema, kode program pada ETL, basis data MOLAP, dan Business Intelligence Front End.
3	Yosua Batubara	Analyst	<ul style="list-style-type: none"> • bisnis proses, bus matrix, attributes and metric list, issues list.
4	Jaime Sirait	Desain	<ul style="list-style-type: none"> • Melakukan Design detailed bus matrix, design 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.
--	--	--	--

2. Analisis

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

2.1. Bisnis Proses

Pada sistem yang akan dibangun, salah satu bisnis proses yang ada pada spotify adalah “**Top Song**”, Fact Grain Type yang cocok untuk proses bisnis ini adalah **Periodic snapshot fact**, dengan granularity one row per song details and one row per track details. Bisnis proses ini akan mengumpulkan, menyaring, mengolah, menggabungkan dan mencatat data Top Song dalam pembentukan data warehouse.

Dimension/Fact Table	Atribute/ Fact Name	Description
Album	<i>Id</i>	Kode unik dari sebuah album
	<i>Name</i>	Judul album dari artist track list
	<i>Album_type</i>	Total lagu pada album yang akan dikelompokkan pada parameter tertentu
	<i>Release_date</i>	Dimana lagu atau single pertama kali di publis
	<i>Popularity</i>	Lagu yang paling banyak di dengar dari sebuah lagu
Artist	<i>Id</i>	Kode unik dari nama artis
	<i>Name</i>	Nama dari artis yang menyanyikan lagu
	<i>Popularity</i>	Popularitas dihitung dari jumlah total pemutaran track dari sebuah lagu
	<i>Followers</i>	Pengikut atau penggemar dari artis
Genre	<i>Id</i>	Kode unik dari sebuah genre
Tracks	<i>Id</i>	Kode unik dari sebuah track
	<i>Duration</i>	Lamanya track diputar
	<i>Popularity</i>	Nama dari sebuah track yang diputar
FactSong	<i>Artistsid</i>	Kode unik dari sebuah artis
	<i>Albumsid</i>	Kode unik dari sebuah album
	<i>Genresid</i>	Kode unik dari sebuah genre
	<i>Trackside</i>	Kode unik dari sebuah track

2.2. Atribut dan Metrik

<Tuliskan daftar atribut dan metrik (*fact*) yang Anda gunakan dalam sistem yang Anda bangun.>

2.3. Isu

<Tuliskan masalah-masalah yang ditemui pada sistem sumber data yang perlu dipertimbangkan dalam tahap desain *data mart* secara detail dan tahap desain sistem *extract, transform, load*. Misalnya, bahasa>

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. Detailed Bus Matrix

[illegible]

Gambar Detailed Bus Matrix

Instructions!	Attribute / Fact Name	Description	Alternate Names	Sample Values
artists	release_date	dimana lagu atau single pertama kali di publis		954633600000
	popularity	lagu yang paling banyak didengar dari sebuah album		0
	id	kode unik dari nama artis		4tujQJicOnuZRLBFdp3Ou
	name	nama dari artis yang menyanyikan lagu		Xzibt
	popularity	popularitas di hitung berdasarkan jumlah total pemutaran track dari sebuah lagu.		69
genres	followers	pengikut atau penggemar dari artis		1193665
	id	kode unik dari sebuah genre		detroit hip hop
tracks	id	kode unik dari sebuah tracks		
	duration	Lamanya track diputar atau berlangsung		275893
	name	Nama dari sebuah track yang diputar		Gz And Hustlas (feat. Nancy Fletcher)
FactSong	popularity	ingkat popularitas sebuah track		0
	artistsid	kode unik dari nama artis		4tujQJicOnuZRLBFdp3Ou
	albums	kode unik dari sebuah album		2jKoVIU7VAmExKJ1h3w9P
	genresid	kode unik dari sebuah genre		detroit hip hop
	tracksid	kode unik dari sebuah tracks		1dizvxctg9dHEyaYTFufVi

Gambar Attributes & Metrics

3.2. Detailed Dimensional Modeling

[illegible]

Table Name	DimTrack																
Table Type	Dimension	<div>Home Page</div>															
Display Name	Track																
Database Schema																	
Table Description	Track Dimension																
Comment	comes from tracks table in spotify.sqlite																
Biz Filter Logic																	
Size	one for each track																
Generate Script?	N																
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	refcso						
TrackKey	TrackKey	Surrogate primary key	-1	1, 2, 3...	key				int		PK ID						
TrackID	TrackID	Business key from source system (aka natural key)		0TXU/bK6BSRsJSeBsfX8	key				nvarchar 30								
TrackName	TrackName	Name of the track		Bouge	2				nvarchar 50								
TrackDuration	TrackDuration	Track's Duration		209475	2				int								
TrackPopularity	TrackPopularity	Track popularity count		0	2				int								
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2		nchar 1								
RowStartDate	Row Start Date	When did this row become valid for this member?	01/01/1900	24/01/2011	n/a	Exclude from cube	Standard SCD-2		datetime								
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	31/12/9999	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2		datetime								
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2		nvarchar 200								
InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dim		int		FK						
UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dim		int		FK						

Table Name	DimAlbum										
Table Type	Dimension										
Display Name	Album										
Database Schema											
Table Description	Album Dimension										
Comment	comes from albums table in spotify.sqlite										
Biz Filter Logic											
Size	one for each album										
Generate Script?	N										
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments	Datatype	Size	reclaid
AlbumKey	AlbumKey	Surrogate primary key	-1	1, 2, 3...	key				int		PK ID
AlbumID	AlbumID	Business key from source system (aka natural key)		7fL6o9tox1zghpKUfh9vuc	key				nvarchar	30	
AlbumName	AlbumName	Name of the album		In Da Club	2				nvarchar	50	
AlbumType	AlbumType	Type of the Album		album	2				int		
AlbumReleaseDate	AlbumReleaseDate	Album's Release Date		1,04449E+12	2				int		
AlbumPopularity	AlbumPopularity	Album popularity count		0	2				int		
RowsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2		nchar	1	
RowStartDate	Row Start Date	When did this row become valid for this member?	01/01/1900	24/01/2011	n/a	Exclude from cube	Standard SCD-2		datetime		
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	31/12/9999	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2		datetime		
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2		nvarchar	200	
InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dm		int		FK
UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dm		int		FK

Table Name	DimGenre								
Table Type	Dimension								
Display Name	Genre								
Database Schema									
Table Description	Genre Dimension								
Comment	comes from genre table in spotify.sqlite								
Biz Filter Logic									
Size	one for each genre								
Generate Script?	N								

Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments
GenreKey	GenreKey	Surrogate primary key	-1	1, 2, 3...	key			
GenreID	GenreID	Business key from source system (aka natural key)		jazz	key			
GenreName	GenreName	Name of the Genre		jazz	2			
GenreArtistName	GenreArtistName	Name of the artist in Genre table		Xzibit	2			
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2	
RowStartDate	Row Start Date	When did this row become valid for this member?	01/01/1900	24/01/2011	n/a	Exclude from cube	Standard SCD-2	
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	31/12/9999	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2	
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2	
InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dim	
UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dim	

Gambar detail dimensional tabel genre

Table Name	DimGenre								
Table Type	Dimension								
Display Name	Artists								
Database Schema									
Table Description	Artists Dimension								
Comment	comes from genre table in spotify.sqlite								
Biz Filter Logic									
Size	one for each genre								
Generate Script?	N								

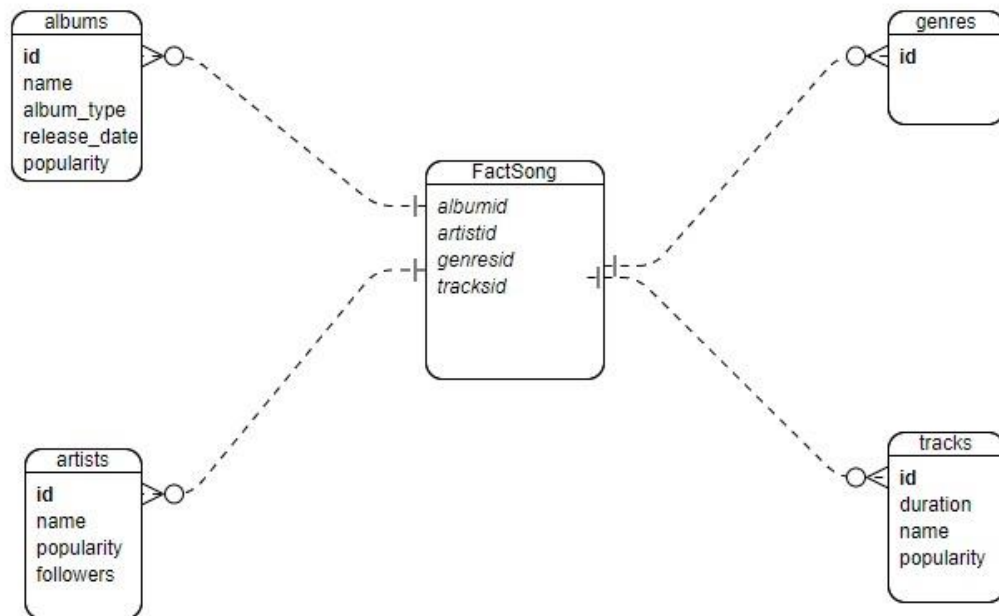
Column Name	Display Name	Description	Unknown Member	Example Values	SCD Type	Display Folder	ETL Rules	Comments
ArtistKey	ArtistKey	Surrogate primary key	-1	1, 2, 3...	key			
ArtistID	ArtistID	Business key from source system (aka natural key)		4tuQJicOnuZRLBFdp3Ou	key			
ArtistName	ArtistName	Name of the artist		Xzibit	2			
ArtistPopularity	ArtistPopularity	Artist popularity count.		69	2			
ArtistFollowers	ArtistFollowers	Fans of artist		1193665				
RowIsCurrent	Row Is Current	Is this the current row for this member (Y/N)?	Y	Y, N	n/a	Exclude from cube	Standard SCD-2	
RowStartDate	Row Start Date	When did this row become valid for this member?	01/01/1900	24/01/2011	n/a	Exclude from cube	Standard SCD-2	
RowEndDate	Row End Date	When did this row become invalid? (12/31/9999 if current row)	31/12/9999	1/14/1998, 12/31/9999	n/a	Exclude from cube	Standard SCD-2	
RowChangeReason	Row Change Reason	Why did the row change last?	N/A		n/a	Exclude from cube	Standard SCD-2	
InsertAuditKey	InsertAuditKey	What process loaded this row?	-1		n/a	Exclude from cube	Standard Audit dim	
UpdateAuditKey	UpdateAuditKey	What process most recently updated this row?	-1		n/a	Exclude from cube	Standard Audit dim	

Gambar detail dimensional tabel artists

3.3. Dimensional Hierarchies

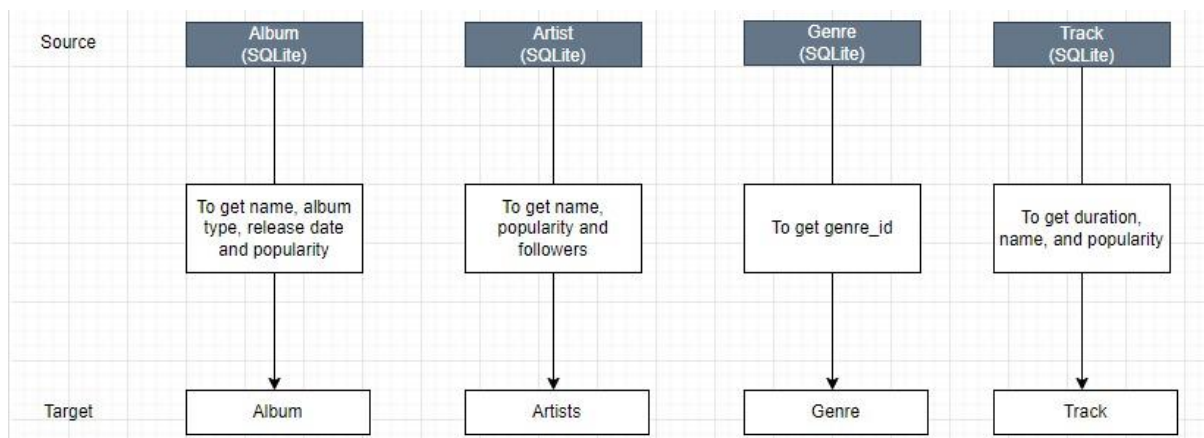
<Gambarkan hirarki pada setiap tabel dimensi.>

3.4. Dimensional Data Model



3.5. 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

3.6. 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.>

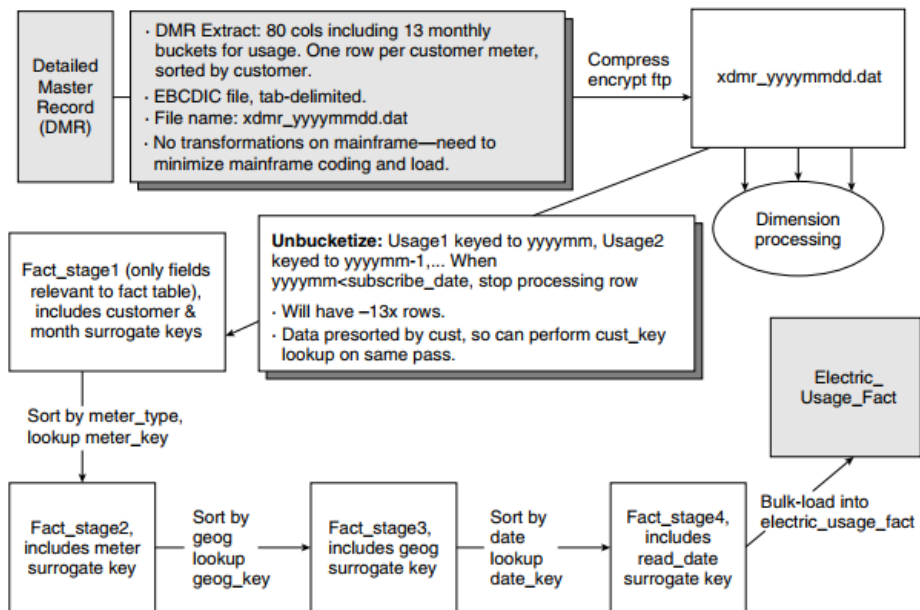


Figure 10-3 Example first draft of detailed load schematic for the fact table.

Gambar 2. Diagram Rinci ETL DMR ke Tabel Electric_Usage_Fact

3.7. Business Intelligence Front End Mockup

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

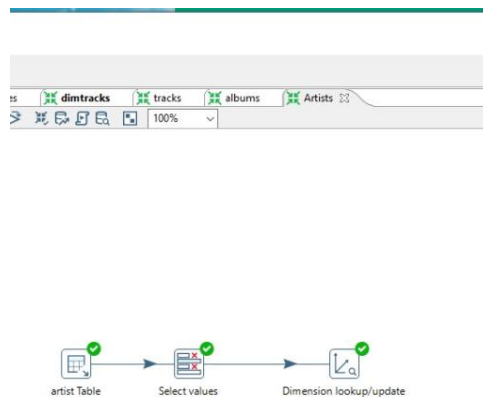
4. Implementasi

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

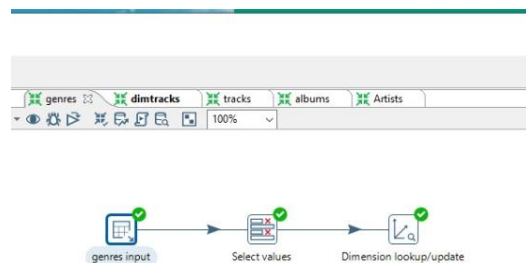
4.1. ROLAP Schema

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

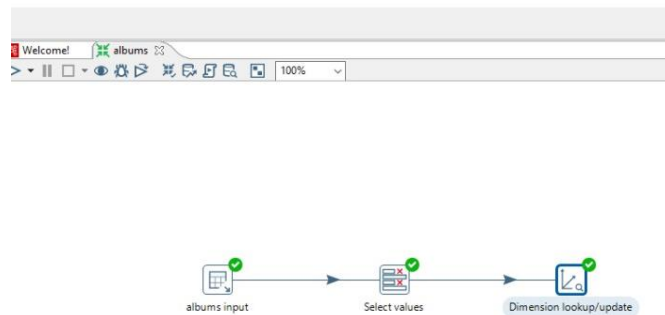
4.2. ETL



Gambar 3. Cuplikan ETL dari Sumber ke Tabel artist



Gambar 4. Cuplikan ETL dari Sumber ke Tabel genre



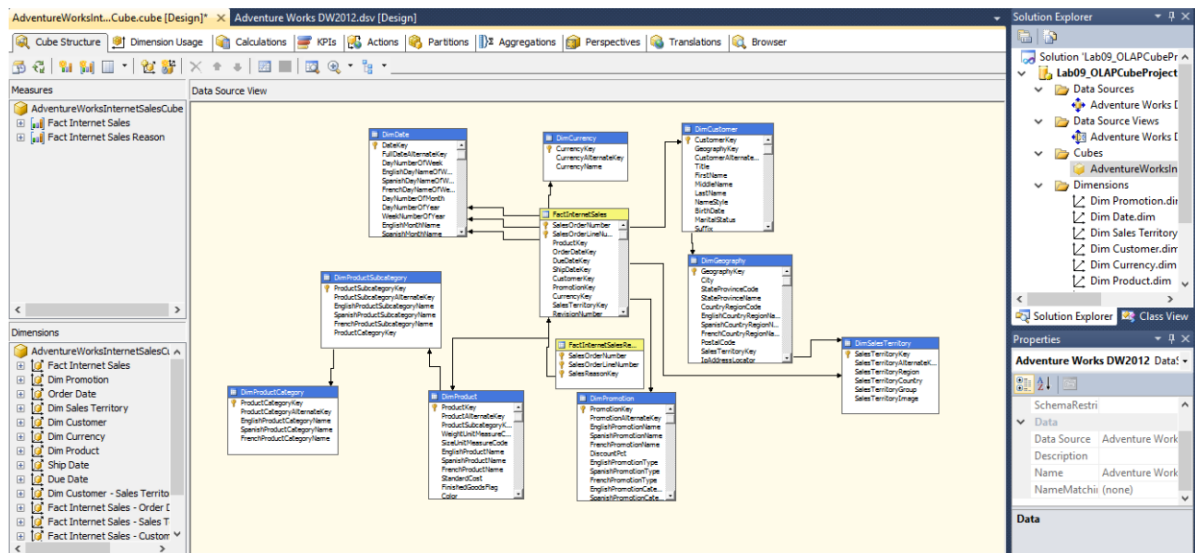
Gambar 5. Cuplikan ETL dari Sumber ke Tabel albums



Gambar 6. Cuplikan ETL dari Sumber ke Tabel tracks

4.3. 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

4.4. Dashboard

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

5. Penutup

5.1. Kesimpulan

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

5.2. Saran

Tuliskan saran pengembangan.>

LAMPIRAN

<Optional.>