**LAPORAN PROYEK MATA KULIAH**

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

**SPOTIFY**



**Disusun oleh:**

| 12S19002 | Yoshua Marganda Alexander Silitonga |
| --- | --- |
| 12S19003 | Dian Filia Matondang |
| 12S19005 | Amelia Jane Audri Lumbanraja |
| 12S19006 | Cintya Yossy Silvana Nainggolan |
|  |  |

**FAKULTAS INFORMATIKA DAN TEKNIK ELEKTRO**

**INSTITUT TEKNOLOGI DEL**

**2021/2022**

# DAFTAR ISI

[DAFTAR ISI i](#_gjdgxs)

[1.](#_30j0zll) Pendahuluan 1

[2.](#_3znysh7) Analisis 2

[3.](#_tyjcwt) Desain 3

[4.](#_4d34og8) Implementasi 5

[5.](#_3rdcrjn) Penutup 7

[LAMPIRAN 8](#_26in1rg)

# Pendahuluan

Bagian ini berisi latar belakang dan tujuan pengerjaan proyek.

* 1. **Latar Belakang**

Perkembangan teknologi informasi saat ini menyediakan berbagai kebutuhan dalam banyak aspek kehidupan yang semakin lama semakin kompleks. Salah satunya pada aspek bisnis. Dimana ini dapat membantu para pihak manajemen untuk proses pengambilan keputusan. Dalam dunia bisnis, manajemen perusahaan harus membuat sistem untuk dapat membantu proses pengelolaan basis data, mendapatkan informasi dan laporan, untuk digunakan dalam proses pengambilan keputusan. Akan tetapi, sistem operasional yang digunakan perusahaan untuk yang mengolah database operasional, yaitu OLTP (Online Transaction Processing) kurang maksimal dalam proses penyimpanan data yang akan semakin bertambah dan semakin menumpuk. Padahal terdapat banyak sekali transaksi yang terjadi pada kurun waktu tertentu, misalnya transaksi setiap bulan. Ini menghasilkan akumulasi data dengan jumlah besar. Data tersebut tidak dapat digunakan secara langsung untuk proses pengambilan keputusan. Sehingga ini memerlukan analisis terhadap data yang dihasilkan tersebut, untuk mengelola informasi-informasi pada OLAP (Online Analytical Processing).

Konsep dari data warehouse mampu mendukung pembuatan keputusan dan menerima data dari berbagai sumber data operasional. Secara definisi, data warehouse adalah suatu koleksi data yang bisa digunakan untuk menunjang pengambilan keputusan manajemen, yang berorientasi subjek (topik), terpadu, time-variant, dan tidak mudah berubah. Data warehouse merupakan sebuah 3 relational database yang didesain dengan lebih menekankan pada query dan analisis dibandingkan untuk proses transaksi. Data warehouse digunakan untuk mendukung pengekstrakan, pemrosesan, dan penyajian yang efisien untuk analisa dan tujuan-tujuan pembuatan keputusan. Sedangkan data mart merupakan bagian dari data warehouse yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu.

* 1. **Ruang Lingkup**

Ruang lingkup menjadi suatu hal yang dibutuhkan dalam pengembangan sistem spotify yang akan dibuat. Dalam pengembangan ini telah disediakan sumber data dari pihak ketiga berupa SQLite. Pendekatan pengembangan sistem 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. **Tim Pengembang**

Pada Tabel 1 disajikan susunan tim pengembang pada proyek ini.

**Tabel 1. Susunan Tim Pengembang**

| No. | Nama | Peran | Tanggung Jawab |
| --- | --- | --- | --- |
| 1 | Yoshua Marganda Alexander Silitonga | Team Leader | * Memeriksa apakah setiap progress yang telah dikerjakan sesuai dengan apa yang telah dirancang sebelumnya. * Membantu menyelesaikan proses pekerjaan. * Mampu mengambil keputusan yang tepat. |
| 2 | Amelia Jane Audri Lumbanraja | ETL Developer | * Mengimplementasi ROLAP schema dan juga kode program pada ETL. * Mengimplementasikan basis data MOLAP. |
| 3. | Dian Filia Matondang | Designer | * Membuat Design detailed bus matrix dan design detailed dimensional modeling. * Membuat dimensional hierarchies dan dimensional data model. * Membuat spesifikasi dari extract, transform, load (ETL) dan detailed ETL flow dari source ke target * Mendesain user interface dashboard mockup. |
| 4. | Cintya Yossy Silvana Nainggolan | Analyst | * Menganalisis bisnis proses yang akan diimplementasikan. * Melakukan analisis pada bus matrix, attributes and metric list, issues list. |

# Analisis

Analisis merupakan suatu bagian pada dokumen ini yang berisi terkait pengamatan dan juga pendeskripsian komponen-komponen yang akan dikaji. Pada bagian ini kita dapat menemukan beberapa hasil pengamatan seperti bisnis proses, *bus matrix*, *attributes and metric list* dan juga *issues list*.

* 1. **Bisnis Proses**

Bisnis Proses yang kami pakai untuk diimplementasikan adalah most listened artist atau artis dengan pendengar terbanyak. Terdapat banyak artis yang tersedia pada database Spotify yang dapat diakses oleh pengguna. Artis - artis tersebut memiliki lagu - lagu yang dapat didengar oleh pengguna sehingga kita dapat memperoleh data dari artis mana yang paling banyak didengar oleh pengguna Spotify. Kita juga dapat membuat batasan waktu dari artis yang paling banyak didengar, misalnya artis yang paling banyak didengar pada bulan April tahun 2022, artis yang paling banyak didengar pada tahun 2021, artis yang paling banyak didengar bulan Februari 2022, dan lainnya.

* 1. **Atribut dan Metrik**

Berikut adalah atribut dan metrik yang akan digunakan pada sistem :

* **Data**

Untuk proyek ini kami menggunakan The Million Playlist Dataset (MPD) yang dirilis oleh Spotify. Sesuai dengan namanya, dataset terdiri dari satu juta playlist dan setiap playlist berisi jumlah lagu dan juga metadata tambahan seperti judul playlist, durasi, jumlah lagu, jumlah artis, dll.

Kumpulan data ini dibuat dengan mengambil sampel daftar putar dari miliaran daftar putar yang telah dibuat pengguna Spotify selama bertahun-tahun.

* **Enhancing the data**

Karena kumpulan data ini dirilis oleh Spotify, kumpulan data ini menyertakan track\_id yang dapat digunakan untuk menghasilkan panggilan API guna mengakses berbagai informasi yang disediakan dari Spotify untuk lagu, artis, atau pengguna tertentu. Untuk melakukan semua panggilan API, kami menggunakan Spotify , yang merupakan pustaka Python ringan untuk Spotify Web API .

* **Shaping the data**

20.000 daftar putar diurai melalui API Spotify untuk mengambil fitur audio untuk setiap lagu di setiap daftar putar, ini sendiri merupakan operasi yang sangat memakan waktu karena latensi panggilan API. Sebagai gambaran, 40 jam diperlukan untuk mengumpulkan fitur untuk semua lagu di setiap daftar putar dalam kumpulan data sampel ini.

Setelah fitur dikumpulkan, rata-rata diambil untuk memiliki fitur audio untuk playlist. Dengan ini, kumpulan data dikurangi menjadi satu baris per daftar putar yang mewakili fitur audio untuk daftar putar.

* **Modeling**

Menggunakan pendekatan tanpa pengawasan di mana akan mengelompokkan data dan kemudian memprediksi klaster untuk pengguna tertentu dan menggunakan klaster itu hanya untuk rekomendasi. Dari sini, pengelompokan akan dikembangkan dan algoritma yang berbeda akan digunakan dan dievaluasi untuk memutuskan mana yang melakukan yang terbaik dari data yang ada.

Tujuh algoritma pengelompokan yang berbeda dipilih, mereka dapat dibagi menjadi 2 kelompok; satu kelompok yang tidak memiliki sejumlah klaster sebagai parameter input (model kepadatan) dan kelompok lainnya yang menggunakan sejumlah k klaster sebagai parameter input (model berbasis centroid).

* 1. **Isu**

Masalah yang ditemui pada sistem data yang perlu di pertimbangkan dalam tahap desain data mart secara detail dan tahap desain sistem extract, transform, load :

# 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*.

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

<Bagian ini mengacu pada [High-Level-Dimensional-Modeling-Workbook.xlsx](https://apipuro.del.ac.id/v1/file/7a0717c21c07474d85483228d3b15a20).>

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

<Bagian ini mengacu pada [Detailed-Dimensional-Modeling-Workbook-KimballU.xlsm](https://apipuro.del.ac.id/v1/file/e11033a04e21c267f7babf6be10e4504).>

* 1. ***Dimensional Hierarchies***

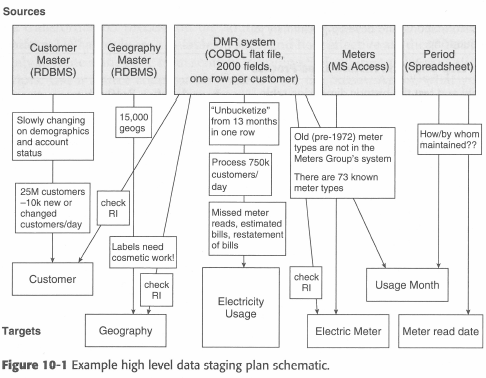
<Gambarkan hirarki pada setiap tabel dimensi.>

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

<Gambarkan *schema* model dimensional.>

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

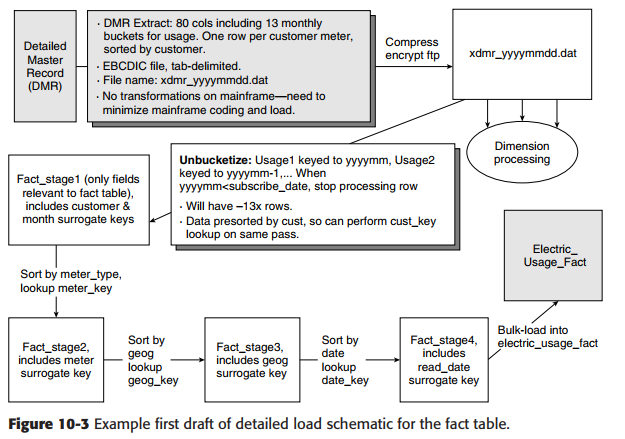
<Gambarkan *high-level source to target map*. Contoh dapat dilihat pada Gambar 1.>



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

* 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

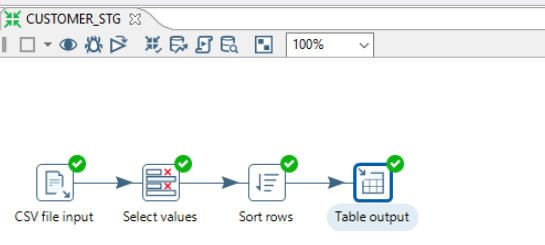
Bagian ini berisi ROLAP *schema*, kode program pada ETL, basis data MOLAP, dan *Business Intelligence Front End*.

* 1. ***ROLAP Schema***

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

* 1. ***ETL***

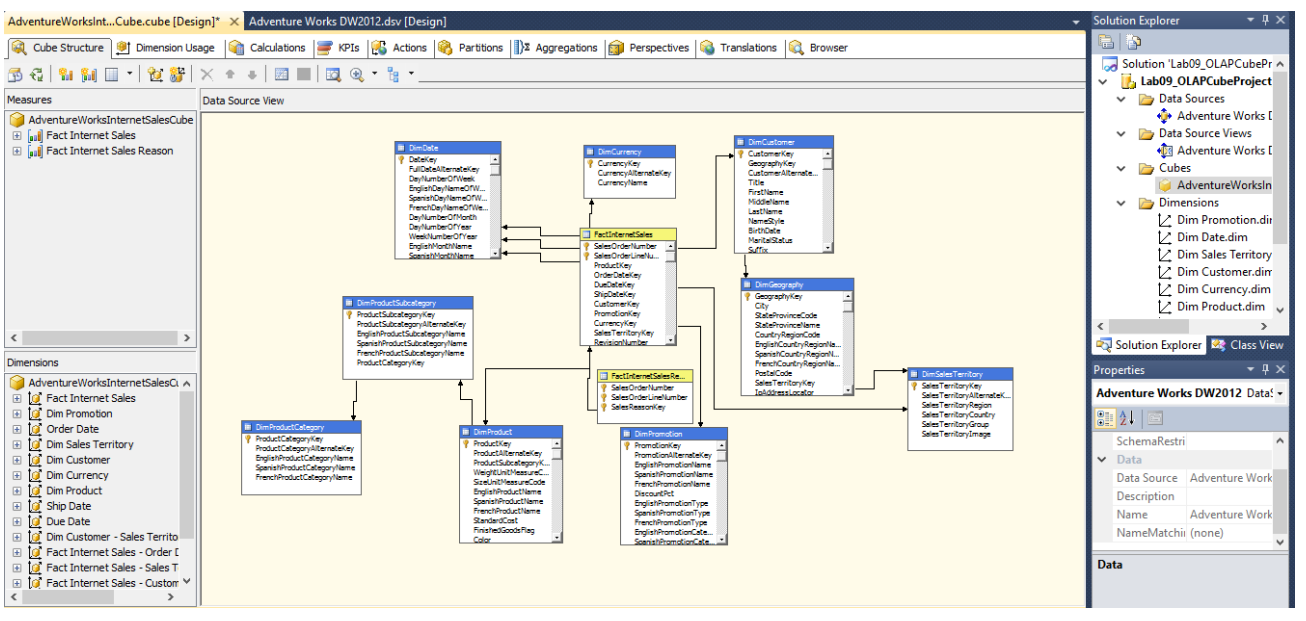
<Tuliskan kode program (atau diagram proses apabila menggunakan **SQL Server Integration Services** ataupun **Spoon**) yang terlibat dalam proses ETL. Misalnya, kode program untuk *populates ROLAP schema from sources*, kode program untuk *initial load*, dan kode program untuk *subsequent loads*. Contoh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