

Rancang Bangun *Data Warehouse* untuk Status Mahasiswa Menggunakan *Snowflake Schema*

Gabby Febrian^{#1}, Tiur Gantini^{*2}

[#]Program Studi Sistem Informasi, Universitas Kristen Maranatha

Jln. Prof. Drg. Surya Sumantri No. 65, Sukawarna, Bandung, Indonesia.

¹1873035@maranatha.ac.id

³tiur.gantini@maranatha.ac.id

Abstract — A data warehouse is an information aggregation system that stores a collection of data from various current and previous sources. The data warehouse is able to produce important information in a nutshell. This can help stakeholders make decisions easier. The university in Bandung has an academic board of directors which is responsible for regulating and managing student status. Currently, universities in Bandung need an information system to obtain reliable information to support operational activities. Universities in Bandung still apply manual transactions to manage student status. Universities in Bandung need a data warehouse to assist in managing student status data. Therefore, it is necessary to study the data warehouse design to process student status data.

Keywords— data warehouse, student status data.

I. PENDAHULUAN

Saat ini Universitas di Bandung masih menerapkan transaksi manual untuk mengelola status mahasiswa. Status mahasiswa di universitas ini terbagi menjadi 2 yaitu, mahasiswa aktif dan mahasiswa tidak aktif. Dikarenakan Universitas di Bandung masih menerapkan transaksi manual, maka Universitas di Bandung memiliki kendala dalam pengolahan data. Universitas di Bandung memerlukan data warehouse guna menunjang pengelolaan data status. Pengelolaan data status mahasiswa ini guna membantu kebutuhan informasi berupa jumlah status mahasiswa. Maka dari itu, diperlukan penelitian untuk membuat rancangan data warehouse guna menghasilkan laporan untuk Universitas di Bandung yang dapat menjadi dasar untuk pengambilan keputusan strategis.

Penelitian ini akan membahas mengenai bagaimana merancang *data warehouse* untuk mengelola status mahasiswa. Status mahasiswa terdiri dari status aktif dan tidak aktif. Penelitian ini akan menggunakan metode Kimball *Nine Step* dan *tools* Pentaho. Adapun kontribusi penelitian ini guna untuk meningkatkan informasi dalam membantu pengambilan keputusan.

II. KAJIAN TEORI

A. Data Warehouse

Data warehouse merupakan sistem penggabungan informasi yang menyimpan kumpulan data dari berbagai sumber masa kini dan masa lalu. Data warehouse harus mampu menyesuaikan perubahan yang mungkin terjadi di sumber data yang mendasarinya. Data warehouse dapat mengolah data yang tidak terstruktur menjadi data yang terstruktur dengan berbagai informasi [1]. Data warehouse dirancang guna menyediakan data untuk menganalisis bisnis. Data warehouse juga disebut sebagai database OLAP dikarenakan fungsinya sebagai database untuk menjawab kebutuhan analisis [2]. Informasi mengenai performa dan pengambilan keputusan sebuah organisasi secara keseluruhan merupakan tujuan pembuatan data warehouse. Maka itu, struktur data warehouse berbeda dengan struktur database operasional biasa. Dalam data warehouse, terdapat dua komponen penting yaitu dimensi dan fakta. Dimensi merupakan perspektif dan fakta merupakan hubungan antara kunci atribut dengan atribut yang nilainya numerik [3].

B. Karakteristik Data Warehouse

Laporan dan analisis dari data warehouse bisa digunakan sebagai kebutuhan organisasi guna menunjang pengambilan keputusan. Data warehouse memiliki karakteristiknya tersendiri yaitu [4] [5]:

1. Berorientasi subjek
Data warehouse dirancang guna kebutuhan laporan dan analisis data berdasarkan subjek yang sama.
2. Terintegrasi
Data terintegrasi dari beberapa sumber dan dibuat menjadi sebuah subjek proses bisnis.
3. *Non-volatile*
Data yang sudah dimasukkan ke dalam data warehouse tidak bisa diperbaharui atau dihapus. Data akan terus tersimpan dalam data warehouse.
4. *Time variant* (rentang waktu)
Data warehouse mampu menyimpan data dari awal terbentuknya data hingga yang terbaru. Semua data tersebut akan dijadikan bahan dasar analisis masa kini maupun sebagai prediksi di masa depan.

C. Tujuan Data Warehouse

Data warehouse memiliki tujuan sebagai berikut [6]:

A. OLAP

OLAP merupakan perangkat yang mendeskripsikan teknologi menggunakan visualisasi multidimensi dari sejumlah data yang kemudian akan dibuat laporannya sesuai dengan permintaan user. Metode OLAP secara konseptual digambarkan sebagai cube yang terdiri dari kategori dimensi dan ukuran [7]. OLAP juga memiliki operator berupa [8]:

1. *Roll-up* = kemampuan untuk melihat data dari tingkat perincian yang lebih rendah.
2. *Drill-down* = kemampuan untuk melihat data dari tingkat perincian yang lebih tinggi.
3. *Slice and dice* = kemampuan untuk melihat subview/subcube tertentu dari sebuah data.

B. Pembuatan laporan

Pembuatan laporan merupakan kegunaan data warehouse yang paling umum. Penggunaan query sederhana dalam data warehouse mampu menghasilkan laporan informasi.

C. *Data mining*

Penggalian informasi dari data yang berjumlah banyak di data warehouse dapat menggunakan data mining dengan menggunakan kecerdasan buatan, statistika, dan matematika.

D. Proses informasi eksekutif

Data warehouse mampu membuat informasi penting secara ringkas. Hal ini dapat mempermudah para pimpinan tingkat atas untuk mengambil keputusan.

D. Prinsip Data Warehouse

Prinsip pada data warehouse adalah mengumpulkan data dan informasi dari sumber internal dan sumber eksternal. Pengelolaan data warehouse dengan mengelompokkan dan mengolah terlebih dahulu untuk disimpan ke data warehouse pusat. Pada prinsipnya data warehouse dapat digunakan guna melengkapi kebutuhan laporan dan pengembangan aplikasi, OLAP, dan *data mining* [9].

E. Manfaat Data Warehouse

Berikut adalah manfaat yang bisa didapatkan dari penggunaan data warehouse [10]:

1. Kemampuan penyimpanan informasi historis.
2. Mengurangi biaya operasi.
3. Mempermudah proses akses data.
4. Kemampuan menganalisis yang cepat.
5. Kemampuan untuk meningkatkan kualitas dan kelengkapan data.
6. Mempermudah pengambilan keputusan.

F. Tabel Fakta

Tabel fakta merupakan tabel utama dalam model dimensi numerik untuk pengukuran kinerja yang umumnya memiliki primary key [11].

G. Tabel Dimensi

Tabel dimensi merupakan tabel yang memiliki banyak kolom dan atribut. Atribut ini mempresentasikan suatu tabel dimensi yang dimana setiap dimensi didefinisikan oleh primary key [11].

H. Snowflake Schema

Snowflake schema merupakan pemodelan dari star schema yang sudah dinormalisasikan. Hal ini dapat membantu untuk meminimalisir terjadinya redundansi data. Kelebihan menggunakan snowflake schema ini yaitu, dapat melakukan maintenance lebih mudah dan ukuran data yang lebih kecil. Kekurangan menggunakan snowflake schema ini yaitu, cenderung lebih sulit dipahami karena tingkat kompleksitasnya dan sulit mencari data dikarenakan strukturnya yang bercabang [12].

I. Online Transaction Processing (OLTP)

OLTP merupakan sistem yang dirancang guna menangani proses transaksi tingkat tinggi pada data operasional organisasi yang dibutuhkan oleh organisasi untuk menangani operasi harian [13].

J. Extraction, Transformation, Load (ETL)

ETL merupakan instrumen khusus untuk mengelola masalah homogenitas, pembersihan, dan penumpukan pusat informasi. ETL memiliki panduan sumber informasi ke dalam data warehouse [14].

K. Online Analytical Processing (OLAP)

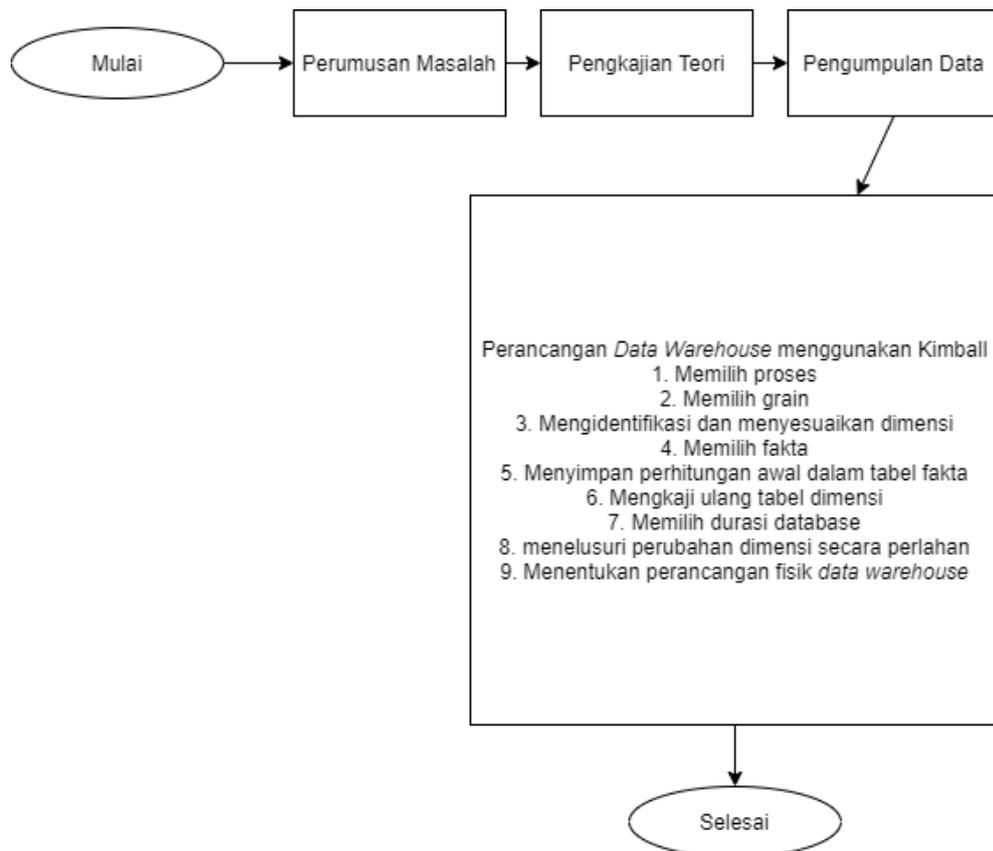
OLAP merupakan perangkat yang mendeskripsikan teknologi menggunakan visualisasi data multidimensi yang bertujuan untuk mempercepat proses analisis [15].

L. Metode Kimball Nine Step

Dalam data warehouse ada banyak metodologi untuk membangun sebuah data warehouse. Metode Kimball merupakan salah satu metode yang dikenal untuk merancang suatu data warehouse. Metode Kimball ini memiliki sembilan Langkah untuk membangun data warehouse yang baik. Berikut adalah sembilan Langkah untuk membangun data warehouse [16]:

1. Memilih proses (*choose the process*)
Pemilihan proses yang merujuk pada kegiatan proses bisnis.
2. Memilih grain (*choose the grain*)
Memilih grain berarti menentukan perwakilan dari sebuah tabel fakta. Dari ini kita dapat menentukan tingkat spesifikasi data yang didapatkan dari model dimensional.
3. Mengidentifikasi dan menyesuaikan dimensi (*identify and conform the dimension*)
Langkah ini akan mengidentifikasikan dimensi yang berhubungan dengan tabel fakta. Dimensi juga merupakan tabel deskripsi yang digunakan untuk menyaring, mengelompokkan, dan mendeskripsikan tabel fakta.
4. Memilih fakta (*choose the fact*)
Pemilihan fakta menetapkan tabel fakta yang dapat mengimplementasikan semua grain yang akan digunakan. Pada tahap ini harus menentukan ukuran yang dibutuhkan tabel fakta.
5. Menyimpan perhitungan awal dalam tabel fakta (*store precalculations in the fact table*)
Pada tahap ini, proses kalkulasi suatu atribut akan dipertimbangkan guna disimpan dalam database. Hal ini guna mengurangi resiko kesalahan pada program.
6. Mengkaji ulang tabel dimensi (*round out the dimension tables*)
Tahap ini akan melengkapi tabel dimensi dengan atribut beserta keterangannya setelah tabel dimensi tersebut diidentifikasi sehingga memuat informasi terstruktur.
7. Memilih durasi database (*choose the duration of the database*)
Tahap ini akan menyimpan durasi data yang dikumpulkan pada beberapa tahun terakhir.
8. Menelusuri perubahan dimensi secara perlahan (*determine the need to track slowly changing dimensions*)
Menentukan perubahan dimensi pada tabel dimensi.
9. Menentukan perancangan fisik data warehouse (*decide the physical design*)
Tahap ini melakukan perancangan fisik pada data warehouse.

III. METODOLOGI PENELITIAN



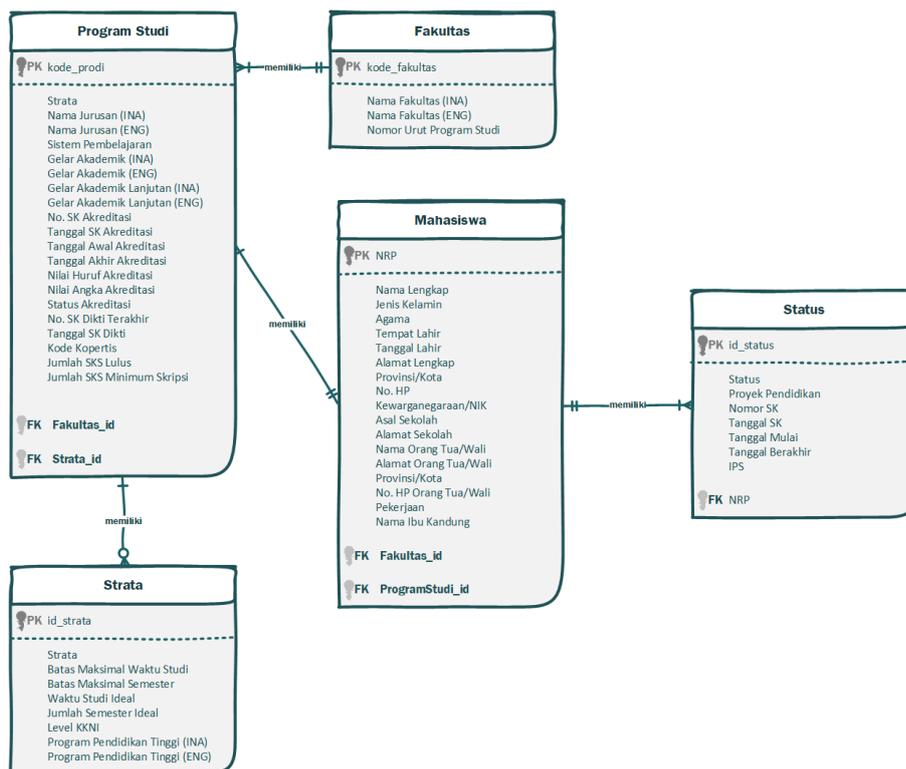
Gambar 1. Alur Penelitian

Berikut adalah alur penelitian yang digambarkan pada gambar 1 :

- A. Perumusan masalah
Pada tahap ini, menentukan terlebih dahulu latar belakang masalah agar dapat menghasilkan sebuah rumusan masalah.
- B. Pengkajian teori
Pada tahap ini, melakukan proses mengkaji teori yang dibutuhkan dalam penelitian, seperti teori mengenai data warehouse dan penelitian terkait.
- C. Pengumpulan data
Pada tahap ini, proses mengumpulkan data dengan metode wawancara dan observasi langsung dari data status mahasiswa di Direktorat Akademik Universitas di Bandung.
- D. Perancangan data warehouse
Pada tahap ini, melakukan perancangan data warehouse menggunakan metode nine step methodology Ralph Kimball.

IV. HASIL PENELITIAN

A. Sistem Saat Ini



Gambar 2. ERD

Saat ini, data status mahasiswa tersimpan disuatu sistem. Dalam system tersebut data masih berupa tabel dan hanya pihak yang bersangkutan yang dapat merubah data tersebut. Dari sistem yang ada, dapat menciptakan ERD dari data tabel yang ada agar dapat melanjutkan pembuatan data warehouse. Hasil dari ERD ini akan dijadikan referensi untuk pembuatan snowflake schema. Tabel-tabel pada ERD ini akan dijadikan untuk referensi pembuatan tabel dimensi.

B. Proses Pembuatan Data Warehouse

Pada gambar 3 tertera data warehouse yang dibuat menggunakan MariaDB. Awalnya membuat database terlebih dahulu yang terdiri dari beberapa tabel dimensi dan fakta yang dibutuhkan. Tabel-tabel tersebut masih kosong belum terisi data. Data akan terisi jika sudah diproses menggunakan ETL.



Gambar 3. Data Warehouse

C. Rancangan Kimball Nine Step

Berikut merupakan rancangan data warehouse sistem akademik menggunakan metode Kimball nine step :

1. Memilih Proses (choose the process)
 Proses dari penelitian ini adalah menghitung status mahasiswa aktif dan status mahasiswa tidak aktif pada Universitas di Bandung.
2. Memilih Grain (choose the grain)
 Pemilihan grain dari penelitian ini berdasarkan jumlah mahasiswa aktif dan tidak aktif yang dilihat berdasarkan proyek pendidikan dan strata.
3. Mengidentifikasi dan Menyesuaikan Dimensi (identify and conform the dimension)
 Rancangan matriks dimensi status mahasiswa pada gambar 4 disesuaikan dengan hasil analisis kebutuhan untuk dasar pembuatan data warehouse sebagai berikut :

Grain	Dimensi						
	Mahasiswa	Prodi	Status	Fakultas	Strata	Perwalian	Proyek_Pendidikan
Mahasiswa aktif dan tidak aktif	✓	✓	✓	✓	✓	✓	✓
Rata-rata IPK	✓	✓	✓	✓	✓	✓	✓

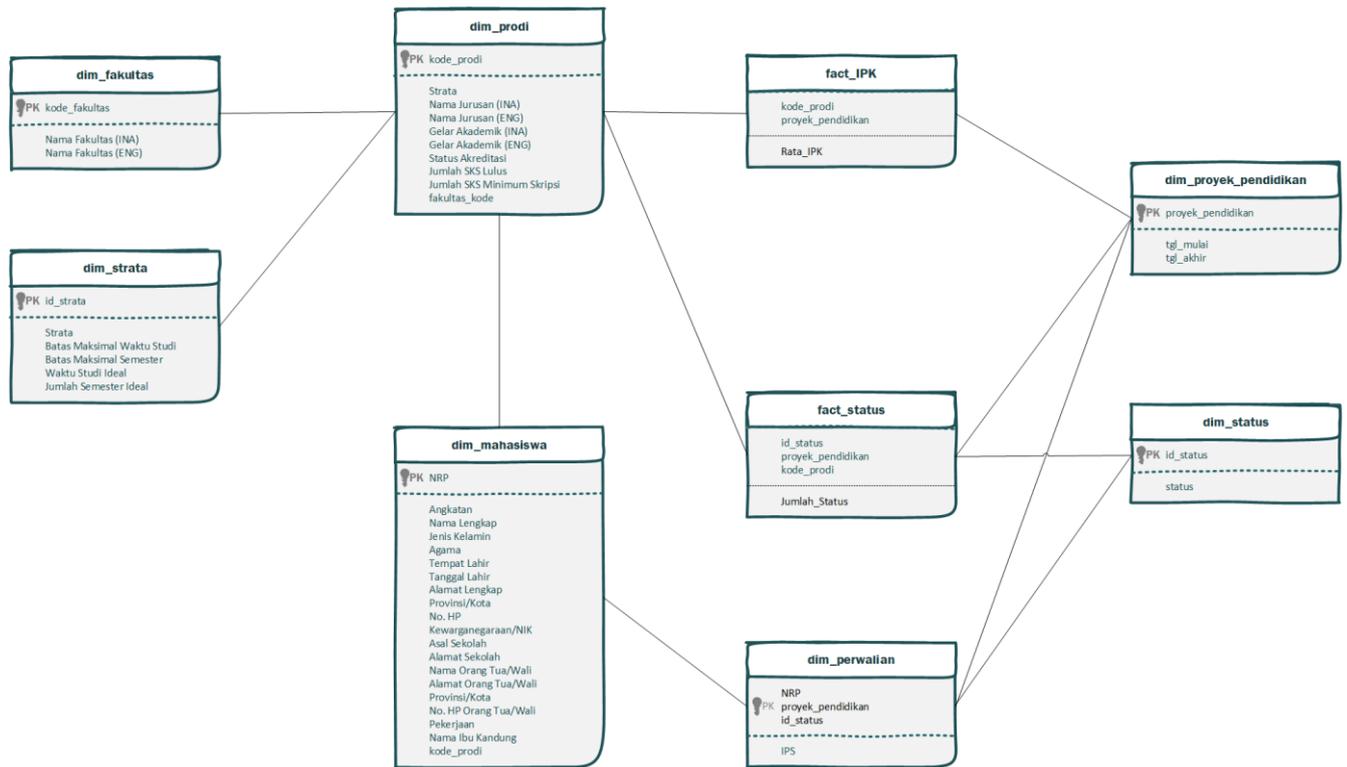
Gambar 4. Rancangan Matriks Dimensi

4. Memilih Fakta (choose the fact)
 Pemilihan fakta dari penelitian ini adalah jumlah status mahasiswa yang dilihat berdasarkan tingkat strata dan proyek pendidikan.
5. Menyimpan Perhitungan Awal dalam Tabel Fakta (store precalculations in the fact table)
 Perhitungan awal pada tabel fakta adalah jumlah status mahasiswa berdasarkan program studi. Jumlah status mahasiswa dihitung berdasarkan berapa banyak status aktif dan tidak aktif setiap proyek pendidikan.
6. Mengkaji Ulang Tabel Dimensi (round out the dimension tables)
 Pemberian deskripsi untuk setiap tabel dimensi seperti pada gambar 5 agar dapat dimengerti pengguna.

No	Nama Tabel Dimensi	Keterangan
1	Tabel dimensi mahasiswa	Tabel dimensi mahasiswa mendeskripsikan informasi mengenai data mahasiswa.
2	Tabel dimensi program studi	Tabel dimensi program studi mendeskripsikan informasi mengenai data dari setiap program studi yang ada.
3	Tabel dimensi status	Tabel dimensi status mendeskripsikan informasi mengenai status mahasiswa.
4	Tabel dimensi fakultas	Tabel dimensi fakultas mendeskripsikan informasi mengenai data fakultas.
5	Tabel dimensi strata	Tabel dimensi strata mendeskripsikan informasi mengenai strata yang ada dari setiap program studi.
6	Tabel dimensi perwalian	Tabel dimensi perwalian mendeskripsikan informasi mengenai mahasiswa, proyek Pendidikan, status, dan IPS.
7	Tabel dimensi proyek pendidikan	Tabel dimensi proyek pendidikan mendeskripsikan informasi mengenai proyek pendidikan(semester).

Gambar 5. Keterangan Tabel Dimensi

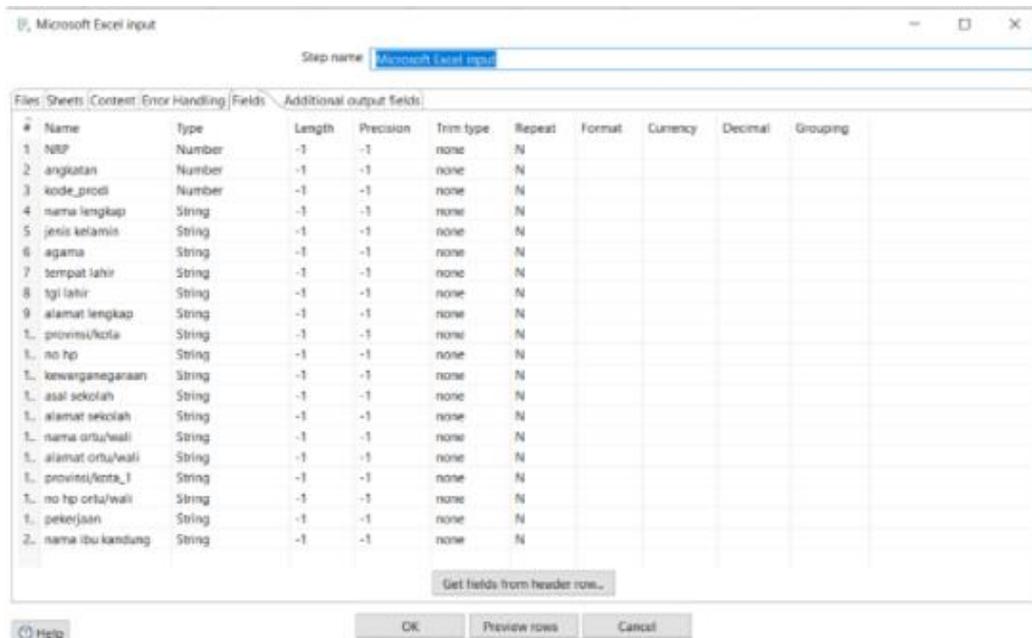
7. Memilih Durasi Database (choose the duration of the database)
 Durasi data yang digunakan pada penelitian ini adalah dari tahun ajaran Ganjil 2018/2019 sampai Ganjil 2019/2020.
8. Menelusuri Perubahan Dimensi Secara Perlahan (determine the need to track slowly changing dimensions)
 Pada tahap ini menentukan potensi perubahan yang terjadi pada atribut di tabel dimensi. Jika terjadi perubahan atribut maka data yang lama akan diganti dengan data baru. Tetapi tahap ini tidak dilakukan karena tidak terjadi perubahan pada atribut di tabel dimensi.
9. Menentukan Perancangan Fisik Data Warehouse (decide the physical design)
 Pada tahap ini, menentukan perancangan fisik data warehouse menggunakan snowflake schema seperti pada gambar 6. Snowflake schema lebih mudah untuk melakukan maintenance dan proses query yang cepat untuk proses ETL.



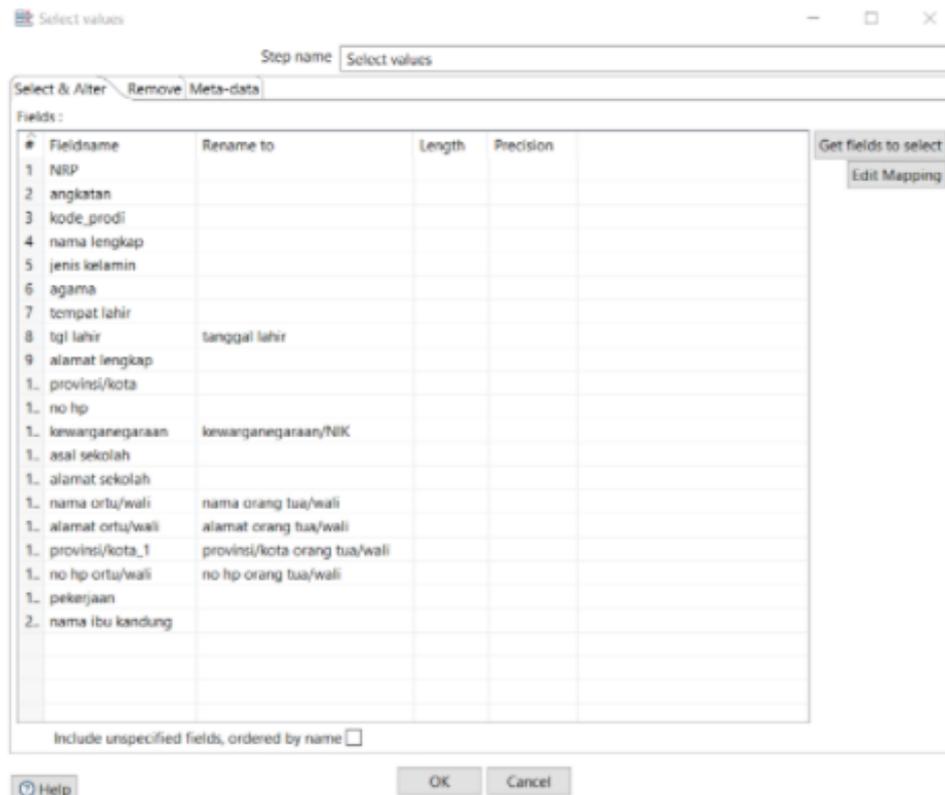
Gambar 6. Snowflake Schema

D. Proses ETL

Berikut merupakan contoh proses ETL untuk dimensi mahasiswa. Proses ETL ini menggunakan *tools* Pentaho. Pada awalnya, dilakukan proses *extract* dari file Excel. Proses ini mengambil semua data mahasiswa yang berada di file Excel tersebut. Pada gambar 7 merupakan data yang berada di dimensi mahasiswa.

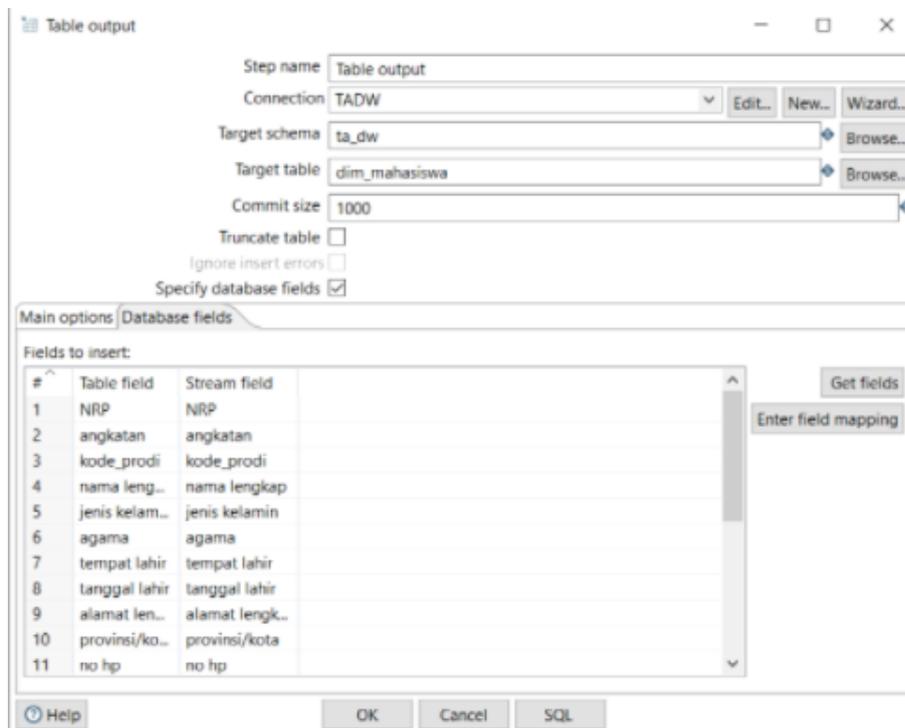


Gambar 7. Proses Extract untuk Dimensi Mahasiswa



Gambar 8. Proses *Transformation* untuk Dimensi Mahasiswa

Setelah proses *extract*, dilakukan proses transformasi seperti pada gambar 8 Proses ini untuk mencocokkan fieldname di dalam Excel dan atribut pada MariaDB. Jika penamaan dalam Excel dan MariaDB berbeda, maka nama penulisan harus diubah sesuai dengan nama atribut di MariaDB pada kolom Rename to.



Gambar 9. Proses *Load* untuk Dimensi Mahasiswa

Setelah proses transformasi, dilakukan proses *load* seperti gambar 9 Proses ini untuk me-load data dari hasil *extract* masuk kedalam data warehouse. Jika proses ini berhasil, maka data berhasil masuk ke MariaDB. Proses ETL untuk dimensi lainnya sama seperti proses ETL pada dimensi mahasiswa.

V. SIMPULAN DAN SARAN

A. Simpulan

Kesimpulan yang didapatkan dari hasil penelitian ini adalah pembuatan data warehouse dibuat dengan menggunakan *extract, transformation, load* (ETL) untuk mengolah data status mahasiswa dari file Excel agar dapat digunakan untuk membuat dashboard laporan jumlah status mahasiswa.

B. Saran

Saran yang dapat diberikan untuk pengembangan sistem di masa depan adalah :

1. Data sudah terorganisir sehingga diharapkan dapat diperluas.
2. Menambah fitur agar dapat menghitung presentasi kelulusan dan masa studi.

DAFTAR PUSTAKA

- [1] E. Kau Suni dan W. Ridwan, "Analisis dan Perancangan Data Warehouse untuk Mendukung Keputusan Redaksi Televisi Menggunakan Metode Nine-Step Kimball (Studi Kasus pada Redaksi Kompas TV Jakarta)," 2018.
- [2] N. Hendrakusma Wardani, N. Yudi Setiawan dan S. Agung Wicaksono, Buku Ajar Data Warehouse, UB Press, 2019.
- [3] A. Filiana, A. G Prabawati, M. N A Rini, G. Virginia dan B. Susanto, "Perancangan Data Warehouse Perguruan Tinggi untuk Kinerja Penelitian dan Pengabdian kepada Masyarakat," Jurnal Teknik Informatika dan Sistem Informasi, vol. 6, 2020.
- [4] A. Harsa K, H. dan N. Yanti, "Perancangan Data Warehouse Penjualan untuk Perhitungan Laba Rugi," vol. 6, p. 120, 2011.
- [5] I. G. Wira Darma, K. Sri Utami dan N. W. Sri Aryani, "Data Warehouse Analysis to Support UMKM Decisions using the Nine-step Kimball Method," International Journal of Engineering and Emerging Technology, vol. 4, 2019.
- [6] G. Wijaya, "Perancangan Data Warehouse Nilai Mahasiswa Dengan Kimball Nine-Step Methodology," 2017.
- [7] C. Hadis dan Rianto, "Perancangan Data Warehouse pada Rumah Sakit (STUDI KASUS: BLUD RSU KOTA BANJAR)," vol. 3, 2017.
- [8] M. Febriady dan B. Adhi Tama, "Rancang Bangun Data Warehouse Untuk Menunjang Evaluasi Akademik Di Fakultas," 2011.
- [9] M. Pentaho: Solusi Open Source Untuk Membangun Data Warehouse, Yogyakarta: C.V. Andi Offset, 2014.
- [10] I. G. Ngurah Wira Partha, P. Arya Mertasana dan P. Novenando.M.W, "Data Center Data Warehouse Development at Z Bali Clinic Using the Kimball Nine-Step Method," International Journal of Engineering and Emerging Technology, vol. 4, 2019.
- [11] M. Ali Suradi, "Pembangunan Perangkat Lunak Data Warehouse di CV. Karya Anugerah Tritunggal," vol. 1, 2016.
- [12] Statiswaty, N. Ransi dan Ardiyanti, "Membangun Data Warehouse Menggunakan Metode Snowflake Schema untuk Data Kecelakaan Lalu Lintas (Studi Kasus : Polres Kota Kendari)," vol. 3, 2017.
- [13] Maskur, W. Soeharso dan M. Agung, "Analisis Dan Perancangan Data Warehouse Evaluasi Mahasiswa Studi Kasus Pada Jurusan Teknik Informatika UMM," vol. 1, p. 61, 2019.
- [14] R. Subash, "Designing a Data Warehouse System for Sales and Distribution Company," 2021.
- [15] I. Bagus, I. G. N. A. S. Mahendra dan Y. P. Sudarmojo, "Design General Hospital Data Warehouse Base on Nine Step Methodology," International Journal of Engineering and Emerging Technology, vol. 4, 2019.
- [16] A. Suganda Girsang, "NINE STEPS DESIGN METHODOLOGY FOR DEVELOPING DATA WAREHOUSE," 22 Desember 2017. [Online]. Available: <https://mti.binus.ac.id/2017/12/22/nine-steps-design-methodology-for-developing-data-warehouse/>.