

IMPLEMENTASI STAR SCHEMA PADA STUDI KASUS PERPUSTAKAAN BERSKALA UNIVERSITAS

Wildan Suharso¹, Abims Fardiansa², Yuda Munarko³, Hardianto Wibowo⁴

^{1,2,3,4} Informatika, Teknik, Universitas Muhammadiyah Malang
Jalan Bendungan Sutami No. 188, Kota Malang, Jawa Timur, Indonesia

e-mail: wsuharso@umm.ac.id², abims.fardi@gmail.com¹, yuda@umm.ac.id³, ardi@umm.ac.id

Received : October, 2019

Accepted : August, 2020

Published : April, 2021

Abstract

Libraries are service units with high storage complexity as evidenced by more data being stored for each year. The data that is not integrated makes the complex problem because every year the process that is carried out continues to increase, especially for the circulation of loans. As the number of books increases, the circulation of borrowing increases every year. On the other hand, the library must know exactly what collection of books they have and the transactions it has made. A lot of data is owned by the library cannot be utilized optimally, so that the managerial is unable to make full use of the data. In University scale libraries, this problem increases when the data is not fully integrated. In this study, the implementation of a star schema was carried out to solve problems related to data integration using a nine-step methodology, which includes selection, item selection, process dimensions, fact selection, fact storage, ensuring dimension tables, selecting database duration, changing dimensions, determining priorities, and query models. The results of this study indicate that the star schema can be implemented in the case of libraries, data warehouses and OLAP to support decision making for adding books, and produced 3 dimensions of the 4 grains found.

Keywords: data warehouse, etl, nine-step, library, star schema

Abstrak

Perpustakaan merupakan unit pelayanan dengan kompleksitas penyimpanan yang cukup tinggi yang dibuktikan dengan semakin besarnya data yang disimpan untuk setiap tahunnya. Data yang tidak terintegrasi menjadikan permasalahan semakin kompleks karena setiap tahunnya proses yang dilakukan terus bertambah terutama untuk sirkulasi peminjaman. Semakin bertambah jumlah buku maka makin bertambah pula sirkulasi peminjaman setiap tahunnya, di lain pihak Perpustakaan wajib mengetahui secara pasti koleksi buku yang dimiliki beserta transaksi-transaksi yang dilakukan. Banyak data yang dimiliki oleh perpustakaan masih tersebar sehingga tidak dapat dimanfaatkan secara maksimal, sehingga pihak manajerial tidak mampu memanfaatkan data secara maksimal. Pada perpustakaan berskala Universitas permasalahan ini semakin meningkat saat data belum terintegrasi seluruhnya. Pada penelitian ini dilakukan implementasi star schema untuk menyelesaikan permasalahan terkait integrasi data dengan menggunakan nine-step methodology, yang meliputi pemilihan proses, pemilihan grain, identifikasi dimensi, pemilihan fakta, penyimpanan fakta, memastikan tabel dimensi, memilih durasi database, menelusuri perubahan dimensi, penentuan prioritas dan model query. Hasil dari penelitian ini menunjukkan bahwa star schema dapat diimplementasikan pada kasus perpustakaan, data warehouse dan OLAP dapat mendukung pengambilan keputusan penambahan buku, serta dihasilkan 3 dimensi dari 4 grain yang ditemukan.

Kata Kunci: data warehouse, etl, nine-step, perpustakaan, star schema

1. PENDAHULUAN

Perpustakaan merupakan unit yang sangat kompleks dalam pengolahan data yang dibuktikan dengan semakin bertambahnya data koleksi buku dan transaksi peminjaman setiap tahunnya. Permasalahan semakin besar saat perpustakaan pada level Universitas dikarenakan transaksi tidak hanya dilakukan oleh anggota perpustakaan, melainkan juga berasal dari mahasiswa luar kampus dan stakeholder yang memiliki akses untuk perpustakaan. Permasalahan pada perpustakaan berskala Universitas juga bertambah saat semakin bertambah jurusan dari Universitas salah satunya adalah Universitas Muhammadiyah Malang yang merupakan kampus dengan koleksi buku yang cukup lengkap mendukung kegiatan akademik. Permasalahan juga semakin bertambah saat data masih belum terintegrasi sepenuhnya yang membuat hambatan-hambatan dalam pemrosesan data perpustakaan.

Sistem penambahan buku yang berjalan di Perpustakaan saat ini berdasarkan rekomendasi dari jurusan atau fakultas kemudian diteruskan melalui Biro Administrasi Akademik, dan juga berdasarkan informasi dari fitur kotak saran yang disediakan oleh pihak perpustakaan. Namun proses ini memberikan informasi yang terbatas bagi manajerial yang mengakibatkan rekomendasi penambahan buku yang kurang maksimal.

Salah satu cara untuk menyelesaikan permasalahan adalah dengan implementasi data warehouse pada perpustakaan, sehingga pihak pengambil kebijakan dapat dengan mudah membaca tren dan informasi penambahan buku perpustakaan. Data Warehouse memungkinkan untuk memanfaatkan basis data dalam membuat keputusan dengan informasi yang jelas dan mudah dimengerti. Analisis dilakukan dengan mengumpulkan data-data yang bersifat historis untuk menyediakan layanan yang bermanfaat dalam pengambilan keputusan [1][2]. Data warehouse merupakan kumpulan data *logic* yang terpisah dengan database operasional dan merupakan suatu ringkasan atau kesimpulan [3]. Data warehouse merupakan bagian dari Business Intelligence yang berupa basis data yang menyimpan data historis dari berbagai

sumber data dalam suatu organisasi. Data historis adalah data-data yang terkumpul sepanjang waktu dan sudah tidak mengalami perubahan.

Data warehouse bisa dirancang dan dikembangkan dengan menggunakan metode Nine-Step Methodology yang terdiri dari pemilihan proses, pemilihan *grain*, identifikasi dari penyampaian dimensi, pemilihan fakta, penyimpanan pre-kalkulasi di tabel fakta, memastikan tabel dimensi, pemilihan durasi database, melacak perubahan dari dimensi secara perlahan, penentuan prioritas dan model *query*.

Penelitian terkait data warehouse dan nine-step methodology telah banyak dilakukan beberapa diantaranya adalah implementasi data warehouse pada studi kasus call center [4], pada studi kasus perpustakaan [5], dan pada studi kasus lab [6], ataupun pada repository [7], dan data warehouse pada penilaian mahasiswa [8], namun belum ada penelitian yang fokus membahas mengenai permasalahan penambahan buku pada perpustakaan. Di lain pihak permasalahan integrasi dan kompleksitas dapat diselesaikan dengan menggunakan data warehouse secara bertahap [9][10], salah satunya menggunakan *nine-step methodology* [11]. Pada penelitian ini dilakukan implementasi data warehouse menggunakan *nine-step methodology* pada studi kasus perpustakaan berskala Universitas yang fokus pada penyelesaian permasalahan penambahan buku.

2. METODE PENELITIAN

2.1. Pengumpulan Data

a. Studi Pustaka

Pada tahap ini dilakukan studi pustaka dengan mencari dan mempelajari sumber, artikel yang berhubungan dengan objek yang diteliti. Memperoleh informasi dari penelitian terdahulu merupakan langkah yang penting dan perlu dilakukan dalam penelitian [12].

b. Wawancara

Melakukan tanya jawab dengan pegawai perpustakaan untuk menganalisa kebutuhan. Permasalahan yang sering dihadapi akan dijadikan bahan guna menentukan kebutuhan-kebutuhan apa saja yang diperlukan dalam pembuatan data warehouse.

2.2. Perancangan Data Warehouse

Metode perancangan data warehouse yang digunakan adalah Nine-step Methodology, yang terdiri dari Pemilihan proses (*choosing the process*), Pemilihan *grain* (*choosing the grain*), Mengidentifikasi dan penyesuaian tabel dimensi (*Identifying and conforming the dimension*), Pemilihan fakta (*choosing the facts*), Menyimpan perhitungan awal dalam tabel fakta (*storing pre-calculation in the fact table*), Memastikan tabel dimensi (*round out of the dimension table*), Memilih durasi *database* (*choosing the duration of the database*), Menelusuri perubahan dari dimensi (*tracking slowly changing dimensions*), Penentuan prioritas dan model *query* (*deciding the query priorities and the query model*).

a. Pemilihan proses (*choosing the process*)

Tahapan ini menentukan subjek utama. Subjek utama merupakan suatu kegiatan organisasi yang menjawab masalah-masalah bisnis yang penting. Pemilihan proses dilakukan untuk memperjelas batasan tentang data warehouse yang dibuat [13]. Adapun proses yang dipilih adalah proses peminjaman buku dan pengembalian buku yang dilakukan oleh anggota perpustakaan.

b. Pemilihan *grain* (*choosing the grain*)

Menentukan secara tepat apa yang akan ditampilkan oleh record pada tabel fakta dalam data warehouse. *Grain* sendiri adalah calon fakta yang dapat dianalisis. Beberapa data yang bisa menjadi calon tabel fakta untuk proses bisnis ialah analisa pada subjek utama penelitian yaitu proses transaksi peminjaman dan pengembalian. Calon fakta tersebut adalah rekomendasi pengadaan buku dari frekuensi subjek-subjek buku yang sering dipinjam, perbandingan jumlah peminjaman buku pada tiap jurusan, perbandingan persentase pengembalian buku tiap tahun.

c. Mengidentifikasi dan penyesuaian tabel dimensi (*Identifying and conforming the dimension*)

Tahap dilakukannya penyesuaian dimensi dan *grain* yang ditampilkan dalam bentuk *matrix*. Pada tahap ini membangun satu set dimensi untuk mengatur konteks pertanyaan apa saja yang nantinya akan diajukan tentang fakta-

fakta yang ada dalam tabel fakta, kemudian menentukan tabel dimensi apa saja yang harus ada. Penyesuaian dimensi dan *grain* yang ditampilkan dalam bentuk *matrix*, yang menjadi acuan dalam penyesuaian dimensi ialah terdapat relasi antara tabel fakta dan tabel dimensi untuk mengatur konteks pertanyaan apa saja yang nantinya akan diajukan tentang fakta-fakta yang ada dalam tabel fakta. Adanya relasi antar tabel menunjukkan bahwa yang berelasi dengan tabel fakta merupakan tabel dimensi pendukung tabel fakta.

d. Pemilihan fakta (*choosing the facts*)

Menentukan fakta mana saja yang bisa digunakan sesuai dengan *grain* yang telah ditentukan sebelumnya. Tabel *agresi* dan *foreign key* dari tabel dimensi.

Tahap ditampilkannya isi data yang terdapat pada tabel fakta. Isi data tersebut adalah *agresi* dari *foreign key* tabel dimensi. Berikut adalah isi dari tabel fakta yaitu *kode_sirkulasi*, *kode_anggota*, *tgl_pinjam*, *tgl_kembali*, *status_sirkulasi*.

e. Menyimpan perhitungan awal dalam tabel fakta (*storing pre-calculation in the fact table*)

Menyimpan perhitungan awal dalam tabel fakta, sesuai dengan *grain* yang telah ditentukan sebelumnya yang merupakan calon-calon fakta. Pada saat fakta-fakta sudah dipilih, setiap fakta harus diperiksa ulang guna menentukan apakah ada fakta-fakta yang dapat diterapkan untuk perhitungan awal. Hasil dari kalkulasi awal ini disimpan dalam tabel fakta. Fakta peminjam dan pengembalian : kalkulasi fakta peminjam adalah jumlah buku yang merupakan jumlah dari proses peminjaman dan pengembalian buku yang bernilai 1 (satu) untuk setiap record pada tabel fakta, penunjang data rekomendasi pengadaan buku berdasarkan subjek-subjek buku yang sering, perbandingan jumlah peminjaman buku pada tiap jurusan perbandingan persentase pengembalian buku tiap tahun.

f. Memastikan tabel dimensi (*round out of the dimension table*)

Tahapan untuk menambahkan keterangan pada tabel dimensi yang telah diidentifikasi. Keterangan yang ditambahkan harus jelas dan mudah untuk dimengerti oleh pengguna.

Tabel 1 : Identifikasi Tabel Dimensi

Grain	Dimesi	Time	Sirkulasi	Anggota
Jumlah peminjam berdasarkan subjek buku		X	X	
Jumlah peminjam berdasarkan jurusan		X		X
Jumlah peminjam berdasarkan pengembalian tepat waktu		X	X	
Total peminjam		X	X	X

Dimensi data_sirkulasi (tabel ini berisikan tentang data-data sirkulasi buku yang ada pada Perpustakaan Pusat UMM dan sesuai dengan kebutuhan penelitian saat ini), dimensi data_anggota (tabel ini berisikan tentang data-data anggota perpustakaan yang melakukan transaksi peminjaman buku pada Perpustakaan Pusat UMM), dimensi waktu (tabel dimensi waktu adalah suatu tabel dimensi yang harus ada dalam sebuah data warehouse karena setiap data warehouse adalah time series. Isi tabel ini berhubungan langsung dengan tanggal peminjaman buku dan pengembalian buku).

Untuk mengetahui asal tabel dimensi maka menggunakan model schema star. Gambar 1 dibawah merupakan gambaran dari model star schema yang digunakan pada data warehouse Perpustakaan Pusat UMM.

g. Memilih durasi database (choosing the duration of the database)

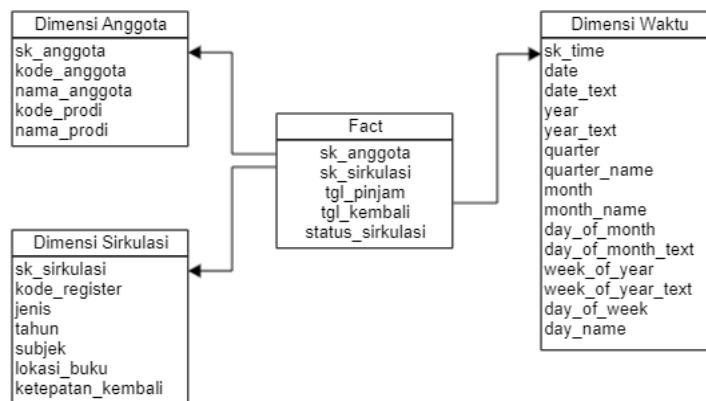
Menentukan batas waktu pada data yang diambil dan memindahkannya kedalam tabel fakta. Tahap ini bertujuan untuk menentukan batas waktu pada data yang akan diambil dan memindahkannya ke dalam tabel fakta. Durasi database yang digunakan dengan rentang waktu antara tahun 2015 sampai 2017.

h. Menelusuri perubahan dari dimensi (tracking slowly changing dimensions)

Perubahan pada tabel dimensi dapat dilakukan dengan tiga cara, yaitu mengganti secara langsung pada tabel dimensi, membentuk baris baru untuk setiap perubahan baru dan perubahan data yang membentuk kolom baru yang berbeda[4].

i. Penentuan prioritas dan model query (deciding the query priorities and the query model)

Mempertimbangkan pengaruh dari rancangan fisik, seperti penyortiran urutan tabel fakta pada disk dan keberadaan dari penyimpanan awal ringkasan (*summaries*) atau penjumlahan (*aggregate*). Pada tahap ini difokuskan terhadap rancangan fisik pada data warehouse. Permasalahan rancangan fisik yang paling kritis adalah mengenai proses ETL (Extract, Transform, Loading) dan memperkirakan kapasitas media penyimpanan. Penentuan prioritas query dan proses untuk membangun data warehouse meliputi ekstraksi data dilakukan pada aplikasi Pentaho Data Integration Kettle, MySQL sebagai pengolah dan pemroses data yang berlangsung, Saiku untuk menampilkan hasil analisa berupa tabel dan chart.



Gambar 1 : Star Schema Data Warehouse Perpustakaan

Pada gambar 1 ditunjukkan star schema yang bersumber dari database saling terkait pada perpustakaan, ditunjukkan terdapat 1 tabel fakta dan 3 tabel dimensi.

3. HASIL DAN PEMBAHASAN

3.1 Persiapan Implementasi Program

Persiapan dalam pembangunan data warehouse diantaranya menyiapkan sumber data yang akan diolah dalam data warehouse, aplikasi Pentaho Data Integration Kettle, aplikasi MySQL dan aplikasi penunjang lainnya. PDI Kettle tentunya sebagai aplikasi utama untuk melakukan ETL dalam data warehouse, hasil dari ETL tersebut adalah sebuah data yang akan disimpan dalam database MySQL.

3.2 Data Source

Sumber data yang digunakan dalam penelitian berupa *file* Excel (.xls/.xlsx), berisi data peminjaman dan pengembalian buku dari tahun 2015-2017 di Perpustakaan Pusat UMM dengan jumlah data 277.845 baris. Data excel tersebut tidak langsung digunakan sebagai sumber data dalam data warehouse, melainkan masih diolah didalam excel untuk mendapatkan jumlah hari peminjaman dan ketepatan pengembalian.

Jumlah hari peminjaman tersebut didapat dari perbandingan tanggal_kembali dengan tanggal_pinjam yaitu dengan rumus (=DAYS(tanggal_kembali;tanggal_pinjam)). Jika sudah mendapatkan jumlah hari peminjaman maka bisa didapatkan status ketepatan pengembalian, data didapatkan dengan menggunakan rumus =if(lama_pinjam>14; "Lambat";"Tepat"). Maksud dari rumus kondisi

tersebut ialah, jika jumlah hari peminjaman lebih dari 14 hari maka berstatus lambat, jika kurang dari atau sama dengan 14 hari maka statusnya tepat. 14 hari tersebut adalah maksimal peminjaman buku

Jumlah hari peminjaman tersebut didapat dari perbandingan tanggal_kembali dengan tanggal_pinjam yaitu dengan rumus (=DAYS(tanggal_kembali;tanggal_pinjam)). Jika sudah mendapatkan jumlah hari peminjaman maka bisa didapatkan status ketepatan pengembalian, data didapatkan dengan menggunakan rumus =if(lama_pinjam>14;"Lambat";"Tepat").

Maksud dari rumus kondisi tersebut ialah, jika jumlah hari peminjaman lebih dari 14 hari maka berstatus lambat, jika kurang dari atau sama dengan 14 hari maka statusnya tepat. 14 hari tersebut adalah maksimal peminjaman buku

3.3 Implementasi ETL

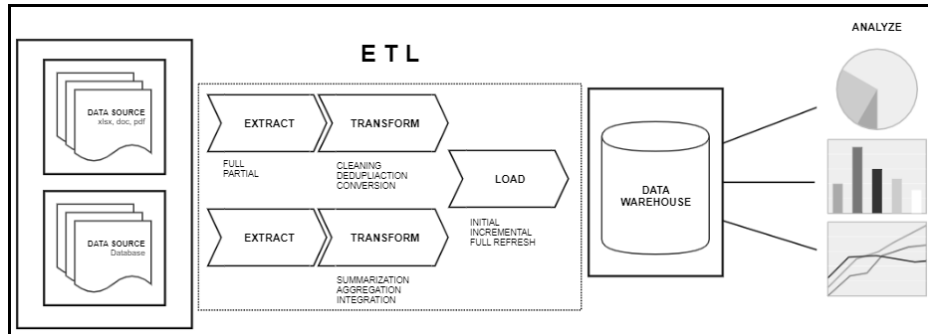
Tujuan dari ETL sendiri adalah mengekstrak, menyaring, menggabungkan dan menyimpan data ke dalam data warehouse.

Hasil dari proses ETL adalah sebuah data yang akan disimpan dalam database MySQL. Maka dari itu perlu dilakukan pembuatan sebuah koneksi tersendiri untuk menghubungkan antara aplikasi penyimpanan yaitu MySQL dengan Pentaho sebagai aplikasi untuk melakukan proses ETL. Koneksi tersebut bertujuan untuk penyimpanan dalam database yang lebih spesifik. ETL berlangsung paada 3 tabel dimensi dan tabel fakta, diantaranya adalah tabel dimensi anggota, dimensi sirkulasi, dimensi waktu, dan tabel fakta.

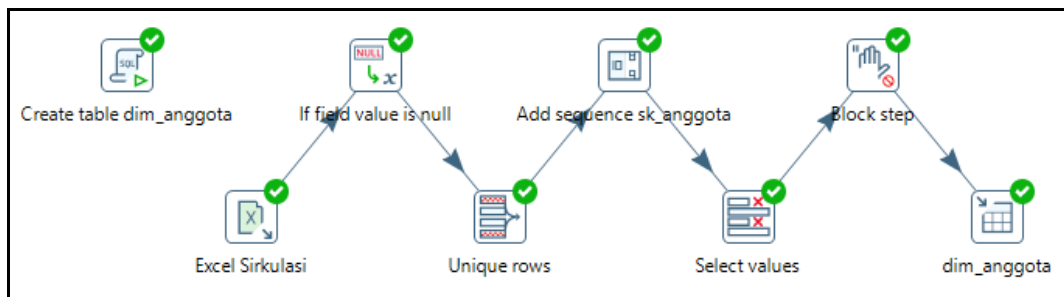
3.3.1 Dimensi Anggota

Pada gambar 3 diatas menggambarkan proses yang dimulai dari extract (input Excel Sirkulasi),

transform (unique rows, add sequence, select values), dan loading (output dim_anggota).



Gambar 2 : Alur Extract, Tranform, Loading (ETL)

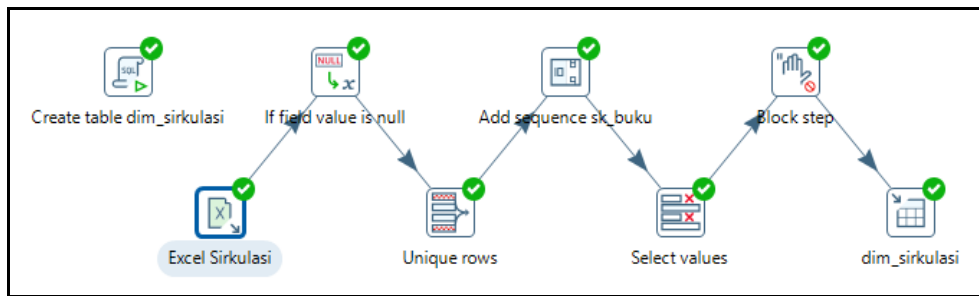


Gambar 3 : Proses ETL Dimensi Anggota

- a. Microsoft Excel Input
Microsoft Excel Input adalah proses untuk mengambil keseluruhan data berupa *file* Excel (.xls/.xlsx) yang digunakan sebagai sumber data yang akan diolah pada proses ETL
- b. If Field Value Is Null
Menstandarisasi penulisan pada data yang kosong pada tiap kolom. Jika terdapat kolom yang kosong (<null>) maka tahapan ini akan melakukan penulisan otomatis pada kolom tersebut dengan *value* yang telah ditentukan.
- c. Unique Rows
menentukan kolom-kolom yang akan masuk dalam proses transformasi dimensi anggota
- d. Add Sequence
Add sequence digunakan membuat field baru yang berisi data integer. Tujuan dari field ini untuk membuat suatu Surrogate Key yang di increment dan yang digunakan pada dimensi anggota
- e. Select Value
Select Values bertujuan untuk memilih dan mengambil atribut/field yang tersedia dan yang akan dimasukkan ke dalam database target melalui table output
- f. Block Until Step Finish
Proses ini bertujuan untuk menunggu segala proses sampai langkah pembuatan tabel dimensi anggota selesai yaitu blocking step Execute SQL Script.
- g. Table Output
Proses Table Output berguna untuk loading data ke dalam tabel database OLAP sesuai dengan target tabel yaitu tabel dim_anggota.
- h. Execute SQL Script
Execute SQL Script adalah langkah dalam membuat sebuah tabel dimensi pada database OLAP. SQL Script dibawah *query* yang dibuat untuk tabel dim_anggota

3.3.2 Dimensi Sirkulasi

Gambar 4 dibawah adalah langkah-langkah ETL pada dimensi sirkulasi.



Gambar 4 : Proses ETL Dimensi Sirkulasi

Proses yang dimulai dari Extract (input Excel Sirkulasi), Transform (unique rows, add sequence, select values), dan Loading (output dim_sirkulasi).

- a. Microsoft Excel Input
Microsoft Excel Input adalah proses untuk mengambil keseluruhan data berupa *file* Excel (.xls/.xlsx) yang digunakan sebagai sumber data yang akan diolah pada proses ETL
- b. If Field Value Is Null
Menstandarisasi penulisan pada data yang kosong pada tiap kolom. Jika terdapat kolom yang kosong (<null>) maka tahapan ini akan melakukan penulisan otomatis pada kolom tersebut dengan *value* yang telah ditentukan.
- c. Unique Rows
menentukan kolom-kolom yang akan masuk dalam proses transformasi dimensi sirkulasi
- d. Add Sequence
Add sequence digunakan membuat field baru yang berisi data integer. Tujuan dari field ini untuk membuat suatu Surrogate Key yang di increment dan yang digunakan pada dimensi sirkulasi
- e. Select Value
Select Values bertujuan untuk memilih dan mengambil atribut/*field* yang tersedia dan yang akan dimasukkan ke dalam database target melalui table output
- f. Block Until Step Finish
Proses ini bertujuan untuk menunggu segala proses sampai langkah pembuatan tabel dimensi sirkulasi selesai yaitu blocking step Execute SQL Script.
- g. Table Output

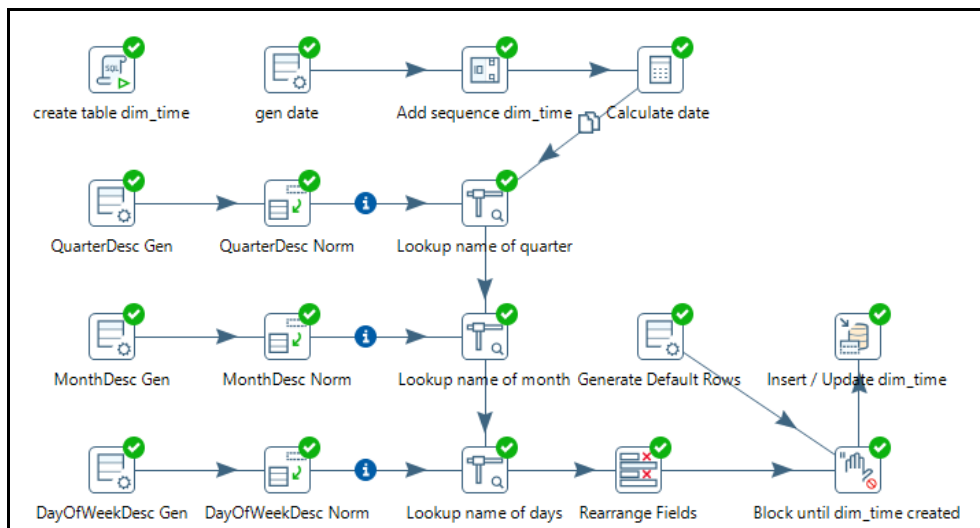
Proses Table Output berguna untuk loading data ke dalam tabel database OLAP sesuai dengan target tabel yaitu tabel dim_sirkulasi.

- h. Execute SQL Script
Execute SQL Script adalah langkah dalam membuat sebuah tabel dimensi pada database OLAP. SQL Script dibawah *query* yang dibuat untuk tabel dim_anggota

3.3.3 Dimensi Waktu

Pada pembuatan tabel dimensi waktu, pembatasan untuk *generate* tanggal sebanyak 3000 hari terhitung dari tanggal 01-01-2014.

- a. Generate rows
Generate rows pada tahap ini berisikan pembuatan tanggal awal yang akan dimasukkan dalam database dim_time. Jumlah tanggal yang akan dihasilkan dengan batas 3000 hari dengan mengatur pada filed limit.
Start_date diperlukan guna menghasilkan nilai awal yang akan digunakan pada tahap Add Sequence yaitu proses pemberian nilai Surrogate Key pada setiap tanggal.
- b. Add sequence
Add sequence digunakan membuat field baru yang berisi data integer. Tujuan dari field ini untuk membuat suatu Surrogate Key dengan perhitungan Surrogate Key dimulai dari 0 dengan perulangan (increment) 1.
- c. Calculator
Pembuatan calculator dalam tahapan pembuatan dimensi waktu menjadi sangat penting, karena ini merupakan tahapan utama dalam mengkalkulasi tanggal, bulan, tahun, dan semester.



Gambar 5 : Proses ETL Dimensi Waktu

- d. Stream Lookup (Lookup name of quarter)
Stream lookup memungkinkan untuk menggunakan informasi dari langkah-langkah lain dalam transformasi.
- e. Generate Rows (QuarterDesc Gen)
Generate row (QuarterDesc Gen) berisikan baris statis yang akan digunakan pada tahap QuarterDesc Norm (Row Normaliser) untuk dinormalkan
- f. Row Normaliser (QuarterDesc Norm)
Row normaliser disini bertujuan untuk normalisasi data yang telah ditentukan oleh generate row (QuarterDesc Gen)
- g. Stream Lookup (Lookup name of month)
Lookup name of month hampir sama dengan dengan lookup name of quarter sebelumnya, hanya saja yang membedakan adalah informasi awal dari calculate date dan informasi dari aliran selanjutnya, yaitu dari MonthDesc Norm
- h. Generate Rows (MonthDesc Gen)
MonthDesc Gen berisikan baris statis yang akan digunakan pada tahap MonthDesc
- i. Row Normaliser (MonthDesc Norm)
Tujuan dari MonthDesc Norm adalah untuk normalisasi data yang telah ditentukan oleh generate row (MonthDesc Gen) yaitu field 'month'
- j. Stream Lookup (Lookup name of days)
Lookup name of days hampir sama dengan dengan step Stream Lookup Value sebelumnya, yang membedakan adalah informasi awal dari calculate date dan informasi dari aliran selanjutnya
- k. Generate Rows (DayOfWeekDesc Gen) berisikan baris statis yang akan digunakan pada tahap DayOfWeekDesc Norm (Row Normaliser) untuk dinormalkan.
- l. Row Normaliser (DayOfWeekDesc Norm)
Tujuan dari DayOfWeekDesc Norm adalah untuk normalisasi data yang telah ditentukan oleh generate row
- m. Select Values (Rearrange Fields)
Select Value bertujuan untuk memilih dan mengambil atribut/*field* yang tersedia dan yang akan dimasukkan kedalam database target melalui output
- n. Block Until Step Finish
Proses ini bertujuan untuk menunggu segala proses sampai langkah pembuatan tabel dim_time (Execute SQL Script) selesai
- o. Insert / Update
Inset / Update berguna untuk loading data ke dalam tabel database OLAP sesuai dengan target tabel yaitu tabel dim_time
- p. Execute SQL Script
Execute SQL Script adalah langkah dalam membuat sebuah tabel dimensi pada database OLAP.

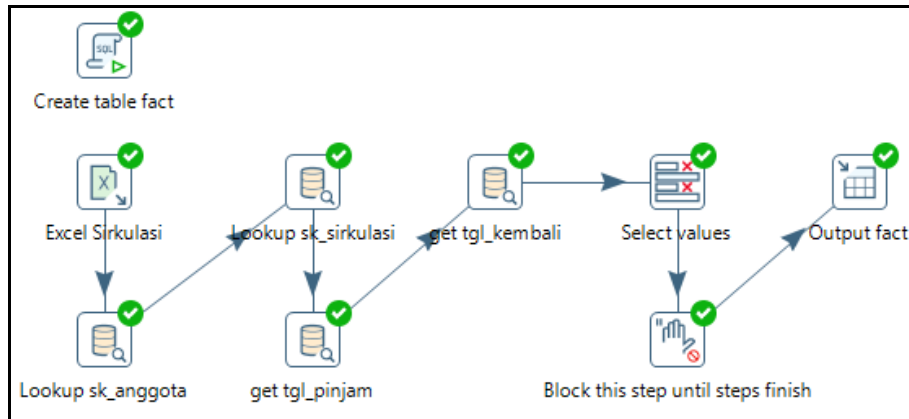
3.3.4 Tabel Fakta

Gambar 6 diatas menunjukkan bahwa proses transformasi pada tabel fakta berhasil dilakukan dan menghasilkan sebuah data warehouse dan gambar 7 merupakan contoh dari hasil ETL pada tabel fakta.

- a. Microsoft excel input
Dilakukan untuk mengambil keseluruhan data berupa *file* Excel (.xls/.xlsx).

- b. Database lookup (Lookup sk_anggota)
Database lookup pada dimensi anggota untuk komparasi isi dari tabel dimensi

anggota dengan data source yang hasilnya nanti akan dimasukkan ke tabel fakta.



Gambar 6 : Proses ETL Tabel Fakta

- c. Database Lookup (Lookup sk_sirkulasi)
Database lookup pada dimensi sirkulasi untuk komparasi isi dari tabel dimensi sirkulasi dengan data source yang hasilnya nanti akan dimasukkan ke tabel fakta.
- d. Database Lookup (get tgl_pinjam)
Database Lookup untuk melihat value dari tabel di database. Value tersebut yang nantinya akan dibuatkan field baru dalam tabel database. Database lookup pada dimensi time untuk komparasi isi dari tabel dimensi time dengan data source yang hasilnya nanti akan dimasukkan ke tabel fakta.
- e. Database Lookup (get tgl_kembali)
Database lookup pada dimensi time untuk komparasi isi dari tabel dimensi time dengan data source yang hasilnya nanti akan dimasukkan ke tabel fakta.

- f. Select Values
Select Values bertujuan untuk memilih dan mengambil atribut/*field* yang tersedia dan yang akan dimasukkan kedalam database target melalui output. Atribut yang digunakan untuk tabel fakta adalah sk_anggota, sk_sirkulasi, tgl_pinjam, tgl_kembali, ketepatan_kembali, status_sirkulasi
- g. Block this step until steps finish
Proses ini bertujuan untuk menunggu segala proses sampai langkah pembuatan tabel (Execute SQL Script / Create table fact) selesai
- h. Output
Proses Table Output berguna untuk loading data ke dalam tabel fakta pada database OLAP
- i. Execute SQL Script
Execute SQL Script adalah langkah dalam membuat sebuah tabel dimensi pada database OLAP

sk_anggota	sk_sirkulasi	tgl_pinjam	tgl_kembali	status_sirkulasi
1	1	367	422	K
2	2	367	374	K
3	3	367	375	K
4	4	367	578	K
5	5	367	380	K
6	6	367	379	K
7	7	367	644	K
8	8	367	381	K
9	9	367	540	K
10	10	367	374	K

Gambar 7 : Hasil ETL Tabel Fakta

DATA EXCEL SIRKULASI 2015-2017

Kode register : 080003801
 Jenis : Book
 Tahun : 2005
 Subjek : 1. POLITICAL CONDITIONS
 Bahasa : INDONESIA
 Kode anggota : 201410050311125
 Nama anggota : Sulung Satrio Irkham P
 Kode prodi : 005
 Nama prodi : Ilmu Pemerintahan
 Tgl pinjam : 2015-01-02
 Tgl kembali : 2015-10-05
 Status S : K
 Tempat : Perpustakaan Pusat (Kampus III)
 Ketepatan : Lambat

Dimensi Anggota

7	201410050311125	Sulung Satrio Irkham P	005
---	-----------------	------------------------	-----

Dimensi Sirkulasi

7	080003801	Book	2005	1. POLITICAL CONDITIONS	Perpustakaan Pusat (Kampus III)	Lambat
---	-----------	------	------	-------------------------	---------------------------------	--------

Tabel Fakta

sk_anggota	sk_sirkulasi	tgl_pinjam	tgl_kembali	status_sirkulasi
1	1	367	422	K
2	2	367	374	K
3	3	367	375	K
4	4	367	578	K
5	5	367	380	K
6	6	367	379	K
7	7	367	644	K
8	8	367	381	K
9	9	367	540	K
10	10	367	374	K

Dimensi Waktu

664	2015-10-26	26-10-2015	2015	2015	4
-----	------------	------------	------	------	---

Dimensi Waktu

367	2015-01-02	02-01-2015	2015	2015	1
-----	------------	------------	------	------	---

Gambar 8 : Validasi Data Hasil ETL

Pada gambar 7 ditunjukkan hasil dari ETL tabel fakta, dan pada gambar 8 ditunjukkan hasil dari validasi data hasil ETL dan diperoleh data adalah valid karena sesuai dengan tabel dimensi.

3.4 Validasi Data

Pengecekan validasi terhadap data sebelum dan sesudah proses ETL perlu dilakukan. Pengecekan validasi dilakukan sebanyak 4 kali pada data Excel Sirkulasi 2015-2017 dan database OLAP yaitu tabel fakta, dimensi sirkulasi dan dimensi anggota. Berdasar dari validasi yang dilakukan, data hasil ETL pada tabel fakta, dimensi anggota, dimensi sirkulasi dikatakan valid.

4. KESIMPULAN

Hasil dari penelitian menunjukkan bahwa implementasi nine-step methodology dapat membantu pengerjaan data warehouse menjadi lebih terstruktur, tertata dan terdokumentasi dengan baik. Data warehouse

berhasil dibuat dengan menggunakan Star Schema. Data warehouse dan OLAP mendukung pengambilan keputusan penambahan buku telah berhasil dibuat dengan menggunakan PDI Kettle untuk proses integrasinya dan dihasilkan 3 dimensi dari 4 grain yang ditemukan. Pengujian validasi memberikan hasil bahwa data pada semua tabel fakta, sirkulasi, dan anggota adalah valid.

DAFTAR PUSTAKA

- [1] A. A. T Ole, "The Importance of Data Warehouses in the Development of Computerized Decision Support Solutions. A Comparison between Data Warehouses and Data Marts," *Database Syst. J.*, vol. VI, no. 4, pp. 18–26, 2016.
- [2] S. H. A. El-Sappagh, A. M. A. Hendawi, and A. H. El Bastawissy, "A proposed model for data warehouse ETL processes," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 2, pp. 91–104, 2011.
- [3] A. Amborowati, "Perancangan Data Warehouse Pada Perpustakaan STMKI

- AMIKOM YOGYAKARTA," *Data Manaj. dan Teknol. Inf.*, vol. 11, no. 1, pp. 43–55, 2010.
- [4] A. SUDRADJAT, "Pengembangan Data Warehouse Call Center Dengan Metodologi Kimball Nine-Step Pada Bina Sarana Informatika," *J. Inform.*, vol. 2, no. 2, pp. 345–361, 2016.
- [5] Andri and B. Tujni, "Analisis Dan Perancangan Data Warehouse Perpustakaan (Studi Kasus: Perpustakaan Universitas Binadarma Palembang)," *Semin. Nas. Inform.*, vol. 2015, no. November, pp. 28–34, 2015.
- [6] A. Chandra, "Perancangan Data Warehouse Pada Software Laboratory Center," *ComTech Comput. Math. Eng. Appl.*, vol. 1, no. 2, p. 585, 2010.
- [7] H. B. Seta, T. Wati, and I. N. Isnainiyah, "Data Warehouse Development for UPN 'Veteran' Jakarta Library (Perancangan Data Warehouse pada Perpustakaan UPN 'Veteran' Jakarta)," *J. Pekommas*, vol. 2, no. 2, p. 161, 2017.
- [8] G. Wijaya, "Perancangan Data Warehouse Nilai Mahasiswa dengan Kimball Nine-Step Methodology," *J. Inform. UBSI*, vol. 4, no. 1, pp. 1–11, 2017.
- [9] Inmon, W. H., *Building The Data Warehouse, Third Edition*. New York : John Wiley & Sons, Inc. 2002.
- [10]. Ponniah, P., *Data Warehousing Fundamentals : A Comprehensive Guide for IT Professionals*. New York : John Wiley & Sons, Inc. 2010.
- [11]. Connolly, T., Begg, C.. *Database Systems : A Practical Approach to Design, Implementation, ans Management, Fourth Edition*. USA : Addison-Wesley. 2005.
- [12] Ardiansyah, D., Suharso, W., & Marthasari, G. I., Analisis Penerima Bantuan Sosial menggunakan Bayesian Belief Network. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 2(2), 506-513. 2018.
- [13] Darman, R., ANALISIS DATA KEJADIAN BENCANA ANGIN PUTING BELIUNG DENGAN METODE ONLINE ANALYTICAL PROCESSING (OLAP). *SINTECH (Science and Information Technology) Journal*, 2(1), 18-23. 2019.