

**“MODEL STAR SCHEMA DAN SNOWFLAKE SCHEMA PADA DATA
WAREHOUSE”**

DOSEN PENGAMPU:

Gede Aditra Pradnyana, S.Kom.,M.Kom



Oleh:

I Gede Riyan Ardi Darmawan

(1815091037)

Sistem Informasi Kelas 5A

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNIK DAN KEJURUAN

UNIVERSITAS PENDIDIKAN GANESHA

TAHUN AKADEMIK 2019/2020

KATA PENGANTAR

Om Suastiastu,

Puji syukur kita panjatkan kehadiran Ida Sang Hyang Widhi Wasa, karena atas Asung Kerta Wara Nugrahanya sehingga penulis dapat menyelesaikan makalah mengenai “*Model Star Schema dan Snowflake Schema Pada Data Warehouse*” ini dengan tepat waktu dan tanpa masalah yang berarti.

Terimakasih penulis ucapkan kepada dosen pengampu mata kuliah Data Warehouse, Bapak Gede Aditra Pradnyana, S.Kom., M.Kom karena atas materi yang beliau berikan, sehingga membantu penulis didalam menyusun makalah ini. Tak lupa juga penulis berterimakasih kepada segenap pihak yang terlibat dalam penulisan makalah ini yang tidak bisa penulis sebutkan satu per satu.

Diluar itu, penulis sebagai manusia biasa menyadari sepenuhnya bahwa makalah yang penulis susun masih banyak kekurangan, baik dari segi tata bahasa, susunan kalimat maupun isi. Oleh sebab itu, dengan segala kerendahan hati, penulis makalah ini menerima segala kritik dan saran yang membangun dari pembaca.

Dengan makalah yang penulis susun, penulis berharap dapat membantu mahasiswa dan masyarakat umum di dalam mengetahui dan memahami mengenai model star schema dan snowflake schema sebagai model yang sering digunakan pada data warehouse. Demikianlah yang bisa penulis sampaikan, semoga dengan makalah ini dapat menambah khazanah ilmu pengetahuan dan memberikan manfaat nyata untuk mahasiswa dan masyarakat.

Om, Santih, Santih, Santih, Om

Singaraja, 02 Oktober 2020

Penulis

DAFTAR ISI

HALAMAN SAMPUL.....	i
KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
BAB 1. PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	1
1.3 Tujuan.....	1
1.4 Manfaat.....	2
BAB 2. PEMBAHASAN	
2.1 Star Schema dan Snowflake Schema.....	3
2.2 Contoh Penerapan Star Schema dan Snowflake Schema	8
2.3 Kelebihan dan Kekurangan Star Schema dan Snowflake Schema	9
BAB 3. PENUTUP	
3.1 Kesimpulan.....	11
3.2 Saran.....	11
DAFTAR PUSTAKA	

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam dunia komputer data warehouse adalah sebuah database untuk menyimpan data history yang sangat besar. Berdasarkan definisi (Inmon, 2005, pp. 29), Data Warehouse (DWH) adalah subject-oriented, integrated, time-variant dan non-volatile dari kumpulan data untuk membantu proses pengambilan keputusan oleh manajemen. Data warehouse membantu dalam mengkombinasikan informasi dengan meringkas (summarizing) dan mengelompokkan (aggregation). Informasi yang diperlukan ini didefinisikan oleh persetujuan para pengguna sesuai dengan informasi yang mereka butuhkan dalam pengambilan keputusan. Suatu data warehouse hanya berisi informasi yang relevan dengan kebutuhan user untuk mendukung pengambilan keputusan.

Data warehouse memisahkan beban kerja analisis dari beban kerja transaksi sehingga memungkinkan organisasi untuk menggabungkan dan mengkonsolidasi data dari berbagai sumber. Data warehouse membantu organisasi untuk menganalisa tren berdasarkan data repositori organisasi dengan jangka waktu tertentu. Fungsi utama dari data warehouse adalah untuk memfasilitasi organisasi dalam melakukan perencanaan strategis berdasarkan data jangka panjang yang tersimpan dan membuat keputusan yang baik dan cepat.

Pada data warehouse, terdapat istilah yang disebut dengan dimensional modelling, dimensional modelling adalah teknik logical design untuk menampilkan data dalam framework standard yang intuitif dan memungkinkan access data dengan performa yang tinggi. Setiap dimensional modelling terdiri atas satu tabel dengan banyak foreign key yang disebut facts table dan satu set tabel yang lebih kecil yang disebut dimension table, setiap dimension table mempunyai satu bagian primary key yang terhubung dengan tepat pada salah satu foreign key dari beberapa key pada tabel facts tersebut. Pada dimensional modelling terdapat beberapa istilah yakni Star Schema dan Snowflake Schema, pada makalah ini akan dijelaskan terkait dengan Star Schema dan Snowflake Schema.

1.2 Rumusan Masalah

- 1.2.1 Bagaimanakah yang dimaksud dengan Star Schema dan Snowflake Schema?
- 1.2.2 Bagaimanakah contoh penerapan Star Schema dan Snowflake Schema?
- 1.2.3 Bagaimakah kelemahan dan kelebihan dari Star Schema dan Snowflake Schema?

1.3 Tujuan

- 1.3.1 Mengetahui apa yang dimaksud dengan Star Schema dan Snowflake Schema dalam data warehouse
- 1.3.2 Mengetahui contoh penerapan Star Schema dan Snowflake Schema pada data warehouse
- 1.3.3 Mengetahui kelebihan dan kekurangan dari Star Schema dan Snowflake Schema

1.4 Manfaat

Makalah ini disusun dengan harapan memberikan kegunaan, baik secara teoritis maupun secara praktis bagi pembaca, penulis dan pihak lainnya. Secara teoritis makalah ini berguna untuk menambah khazanah ilmu pengetahuan terutama dalam bidang Teknologi Informasi yang membahas terkait dengan data warehouse khususnya permodelan data. Secara praktis, hasil dari makalah ini diharapkan bermanfaat bagi:

- Penulis, sebagai objek media penyampaian informasi hasil karya penulis yang ditujukan kepada para pembaca mengenai permodelan Star Schema dan Snowflake Schema pada data warehouse.
- Pembaca, sebagai media informasi di dalam mempelajari mengenai data warehouse, khususnya permodelan Star Schema dan Snowflake Schema.

BAB II

PEMBAHASAN

2.1 Star Schema dan Snowflake Schema

2.1.1 Star Schema

Menurut Connolly dan Begg (2010:1227), star schema adalah model data dimensional yang mempunyai fact table di bagian tengah, dikelilingi oleh tabel dimensi yang terdiri dari data reference (yang bisa di-denormalized). Star schema mengambil karakteristik dari factual data yang di-generate oleh event yang terjadi dimasa lampau.

Star Schema (SS) adalah salah satu implementasi desain logikal dari model multidimensional. Tujuan utama dari star schema adalah untuk membuat skema menjadi lebih simpel sehingga akan mempercepat proses query yang di dalamnya biasanya terdapat banyak perintah join. Star schema terdiri dari sebuah atau beberapa tabel fact yang berkorelasi dengan beberapa tabel dimensi. Star schema mempunyai keuntungan mempercepat proses query karena tabel dimensi telah di denormalized sehingga mengurangi waktu yang dibutuhkan untuk proses join. Ada beberapa karakteristik dari star schema, yakni sebagai berikut

- Setiap dimensi dalam skema bintang diwakili dengan hanya tabel satu dimensi.
- Tabel dimensi harus berisi kumpulan atribut.
- Tabel dimensi digabungkan ke tabel fakta menggunakan kunci asing
- Tabel dimensi tidak digabungkan satu sama lain
- Tabel fakta akan berisi kunci dan ukuran
- Skema Star mudah dipahami dan menyediakan penggunaan disk yang optimal.
- Tabel dimensi tidak dinormalisasi.
- Memiliki 1 tabel fakta dan beberapa tabel dimensi
- Tidak menggambarkan hirarki secara langsung
- Skema ini didukung secara luas oleh Alat BI

Berikut merupakan beberapa proses yang dapat dilakukan untuk membuat star schema secara umum.

➤ Membuat proses bisnis dan memilih proses

Proses bisnis adalah kegiatan operasional yang dilakukan dalam sebuah organisasi atau perusahaan, di dalam pembuatan starschema ataupun cube maka proses bisnis menghasilkan metric kinerja yang diterjemahkan ke dalam fakta-fakta yang ditampung pada tabel fakta. Tabel fakta berfokus pada hasil dari proses bisnis, sehingga proses bisnis yang sesuai dengan studi kasus FIT-WORLD GYM yaitu mendapatkan data pendapatan atau revenue.

➤ Menentukan Grain atau detail

Dalam memodelkan sebuah dimensi, grain atau kedetailan menetapkan sebuah representasi baris dalam tabel fakta. Kedetailan atau grain dipakai dalam melakukan analisis.

➤ Mengidentifikasi Dimensi

Dimensi memberikan konteks apa yang ingin dilihat dalam proses bisnis seperti siapa, apa, kapan dan sebagainya. Tabel dimensi berisi atribut deskriptif yang digunakan untuk menyaring dan pengelompokan fakta-fakta.

➤ Mengidentifikasi Fakta

Fakta berisi informasi data utama yang digunakan dalam proses analisa, di dalamnya juga terdapat data kunci yang berkorelasi dengan tabel dimensi. Data pada tabel fact haruslah numerik sehingga dengan mudah untuk diagregasi

➤ Membuat Star Schema dan Metadata

Setelah semuanya selesai, selanjutnya adalah tahap pembuatan star schema, bentuknya yang tidak terlalu rumit memudahkan dalam melakukan query pada data yang dibuat.

Ada beberapa komponen yang terdapat pada Star Schema, yakni adalah sebagai berikut:

➤ Fact

Fact adalah suatu angka dari pengukuran yang menunjukkan aspek tertentu dari suatu bisnis atau suatu aktivitas . Fact Table berisi beberapa fakta yang terhubung dengan masing-masing Dimension nya. Fact dapat berupa nilai yang telah ada atau baru diturunkan pada saat Run-time.

➤ Dimensions

Dimension adalah karakteristik suatu ukuran yang menyediakan tambahan cara melihat suatu fakta yang telah diberikan pada Fact Table. Dimension disimpan pada Dimension Table.

➤ Attributes

Setiap tabel dimensi mempunyai Attributes. Attributes sering dipakai pada operasi Search, Filter, atau Grouping dari suatu Fact. Dimensions menyediakan karakteristik deskriptif (uraian) tentang Fact lewat Attributes.

➤ Attributes Hierarchie

Attributes pada suatu Dimension dapat diurutkan dengan definisi yang baik dalam suatu Attribute Hierarchies. Attribute Hierarchies menyediakan Data dengan organisasi Top-Down yang terutama berguna untuk Aggregation dan Drill-Down/Roll-up Data Analysis.

➤ Granularity

Granularity adalah salah satu aspek terpenting dalam desain Data Warehouse karena menentukan volume data yang akan disimpan dalam Data Warehouse dan menentukan kedalaman detail Query yang bisa dijalankan. Secara ekstrem ada Lowest Grain (Grain terendah) dan Highest Grain (Grain tertinggi). Lowest Grain menyimpan transaksi di level detail (Atomic Transaction) sedangkan Highest Grain menyimpan data hanya di level Enterprise atau level Perusahaan (Summary Transaction) Level dari Granularity disimpan pada Hierarchy suatu Dimension.

2.1.2 Snowflake Schema

Snowflake schema adalah salah satu metode pemodelan data untuk data warehouse yang merupakan pengembangan dari star schema, yaitu metode normalisasi untuk tabel dimensi yang ada dalam star schema (Ponniiah, 2001). Menurut Inmon (2005) snowflake schema adalah gabungan dari beberapa pemodelan data star schema yang dinormalisasi.

Secara umum, ada beberapa karakteristik dari Snowflake, pada Snowflake schema setiap tabel dimensi dapat memiliki sub-tabel dimensi lagi. Hal ini bertujuan untuk meminimalkan data yang berlebihan (redudancy

data)(Rainardi, 2008). Dimensi data inilah yang menjadi subjek informasi untuk menjadi bahan dalam pengambilan keputusan, karena pada setiap dimensi data dimungkinkan untuk dilakukan pemecahan lebih detail lagi. Dengan demikian sumber data yang bisa diolah untuk menjadi informasi bisa menjadi lebih banyak dan detail.

Snowflake schema merupakan bentuk perluasan dari star schema dimana terjadi proses normalisasi dari beberapa atau seluruh dimension table. Snowflake schema biasanya digunakan pada tabel yang sangat besar dan jika star schema tidak mampu menggambarkan kompleksitas dari database tersebut. Karakteristik lainnya adalah pada model ini mempunyai satu atau lebih dimensi yang sama seperti di Star scheme. Hanya saja pada Snowflake, tabel yang berelasi pada fact table hanya tabel dimensi utama, sedangkan tabel yang lain dihubungkan pada tabel dimensi utama. Model snowflake ini hampir sama seperti teknik normalisasi

Oleh karena Snowflake schema merupakan perkembangan dari Star schema, maka beberapa komponen yang terdapat pada Star schema juga ada pada Snowflake schema, komponen tersebut diantaranya adalah sebagai berikut:

➤ Fact

Fact adalah suatu angka dari pengukuran yang menunjukkan aspek tertentu dari suatu bisnis atau suatu aktivitas . Fact Table berisi beberapa fakta yang terhubung dengan masing-masing Dimension nya. Fact dapat berupa nilai yang telah ada atau baru diturunkan pada saat Run-time.

➤ Dimensions

Dimension adalah karakteristik suatu ukuran yang menyediakan tambahan cara melihat suatu fakta yang telah diberikan pada Fact Table. Dimension disimpan pada Dimension Table.

➤ Attributes

Setiap tabel dimensi mempunyai Attributes. Attributes sering dipakai pada operasi Search, Filter, atau Grouping dari suatu Fact. Dimensions menyediakan karakteristik deskriptif (uraian) tentang Fact lewat Attribut nya.

➤ Attributes Hierarchie

Attributes pada suatu Dimension dapat diurutkan dengan definisi yang baik dalam suatu Attribute Hierarchies. Attribute Hierarchies menyediakan Data dengan organisasi Top-Down yang terutama berguna untuk Aggregation dan Drill-Down/Roll-up Data Analysis.

➤ Granularity

Granularity adalah salah satu aspek terpenting dalam desain Data Warehouse karena menentukan volume data yang akan disimpan dalam Data Warehouse dan menentukan kedalaman detail Query yang bisa dijalankan. Secara ekstrem ada Lowest Grain (Grain terendah) dan Highest Grain (Grain tertinggi). Lowest Grain menyimpan transaksi di level detail (Atomic Transaction) sedangkan Highest Grain menyimpan data hanya di level Enterprise atau level Perusahaan (Summary Transaction) Level dari Granularity disimpan pada Hierarchy suatu Dimension.

Untuk membuat Snowflake, menurut saya secara umum cara-cara yang dilakukan hampir sama seperti pada Star scheme, yang membedakan nantinya adalah pada Snowflake akan dilakukan normalisasi terlebih dahulu untuk menentukan relasi-relasi antar tabelnya. Berikut merupakan langkah-langkah yang dilakukan secara umum:

➤ Membuat proses bisnis dan memilih proses

Proses bisnis adalah kegiatan operasional yang dilakukan dalam sebuah organisasi atau perusahaan, di dalam pembuatan starschema ataupun cube maka proses bisnis menghasilkan metric kinerja yang diterjemahkan ke dalam fakta-fakta yang ditampung pada tabel fakta. Tabel fakta berfokus pada hasil dari proses bisnis, sehingga proses bisnis yang sesuai dengan studi kasus FIT-WORLD GYM yaitu mendapatkan data pendapatan atau revenue.

➤ Menentukan ER-Model

Perancangan data warehouse dengan menggunakan konsep ER model sebagai dasar perancangan yaitu dengan melakukan identifikasi entitas dasar dan melakukan relasi pada setiap entitas yang berhubungan.

➤ Menentukan Grain atau detail

Dalam memodelkan sebuah dimensi, grain atau kedetailan menetapkan sebuah representasi baris dalam tabel fakta. Kedetailan atau grain dipakai dalam melakukan analisis.

➤ Mengidentifikasi Dimensi

Dimensi memberikan konteks apa yang ingin dilihat dalam proses bisnis seperti siapa, apa, kapan dan sebagainya. Tabel dimensi berisi atribut deskriptif yang digunakan untuk menyaring dan pengelompokan fakta-fakta.

➤ Mengidentifikasi Fakta

Fakta berisi informasi data utama yang digunakan dalam proses analisa, di dalamnya juga terdapat data kunci yang berkorelasi dengan tabel dimensi. Data pada tabel fact haruslah numerik sehingga dengan mudah untuk diagregasi

➤ Membuat Physical Dimension Representations

Tahapan ini adalah transformasi dimension model ke dalam Relational Database Management System (RDBMS).

➤ Structure Query Language (SQL)

Setelah itu, jika semua sudah beres selanjutnya tahapan akan dilakukan penyusunan SQL yang akan digunakan pada data warehouse.

2.2 Contoh Penerapan Star Schema dan Snowflake Schema

2.2.1 Star Schema

Salah satu contoh penerapan dari Star Schema adalah pada kasus perancangan data warehouse penjualan pada Cemerlang Skin Care. Pada contoh kasus ini, Star Schema digunakan untuk membantu mengintegrasikan data perusahaan yang tersebar diberbagai cabang dalam bentuk yang lebih ringkas dan menunjang informasi yang dibutuhkan oleh para eksekutif. Selain itu, dengan penggunaan Star Schema, informasi menjadi lebih ringkas dan mempermudah serta mempercepat para eksekutif dalam melakukan analisis informasi sehingga pengambilan keputusan dapat terlaksana dengan cepat dan dapat mengakses data tanpa harus melakukan banyak query sehingga tidak mengurangi beban server utama.

2.2.2 Snowflake Schema

Salah satu contoh penerapan dari Snowflake Schema adalah pada kasus perancangan data warehouse penjualan pada Cemerlang Skin Care. Pada kasus ini, Snowflake schema digunakan ketika Cemerlang Skink Care mulai memiliki data yang sangat banyak sehingga membutuhkan tempat penyimpanan yang lebih hemat. Ketika tabel-tabel dimensi memerlukan ruang penyimpanan yang besar. Dalam sebagian besar kasus, umumnya tabel fakta memerlukan ruang penyimpanan yang lebih besar daripada tabel dimensi. Tabel fakta juga berkembang lebih banyak dibanding tabel dimensi. Akan tetapi ada beberapa kondisi dimana ukuran tabel dimensi lebih besar dibandingkan dengan tabel fakta. Contohnya, tabel-tabel dimensi dapat berisi atribut-atribut redundan, namun diperlukan di dalam proses query laporan.

2.3 Kelebihan dan Kekurangan Star Schema serta Snowflake Schema

2.3.1 Star Schema

➤ Kelebihan

- a. Efisiensi, struktur database konsisten sehingga efisien dalam mengakses data dengan menggunakan tool untuk menampilkan data termasuk laporan tertulis dan query.
- b. Kemampuan untuk mengatasi perubahan kebutuhan, skema bintang dapat beradaptasi terhadap perubahan kebutuhan pengguna karena semua tabel dimensi memiliki kesamaan dalam hal menyediakan akses ke tabel fakta.
- c. Extensibilitas, model dimensional dapat dikembangkan. Seperti menambah tabel fakta selama data masih konsisten, menambah tabel dimensi selama ada nilai tunggal di tabel dimensi tersebut yang mendefinisikan setiap record tabel fakta yang ada, menambahkan attribute tabel dimensi, dan memecah record tabel dimensi yang ada menjadi level yang lebih rendah daripada level sebelumnya.
- d. Kemampuan untuk menggambarkan situasi bisnis pada umumnya, pendekatan standar untuk menangani situasi umum di dunia bisnis yang terus bertambah.
- e. Proses query yang bisa diprediksi, aplikasi data warehouse yang mencari data dari level yang di bawahnya akan mudah

menambahkan jumlah attribute pada tabel dimensi dari sebuah skema bintang. Aplikasi yang mencari data dari level yang setara akan menghubungkan tabel fakta yang terpisah melalui tabel dimensi yang dapat diakses bersama.

f. Proses query lebih cepat pada saat proses OLAP

➤ **Kekurangan**

- a. Ukuran penyimpanan relatif lebih besar. Karena ada data yang berulang sehingga disk space yang digunakan lebih banyak.
- b. Maintenance dan update lebih sulit. Karena tabel yang tidak normal.

2.3.2 Snowflake Schema

➤ **Kelebihan**

- a. Mampu mengatasi kekurangan dari star schema, yakni pada masalah ukuran penyimpanan data yang lebih kecil
- b. Maintenance dan update jadi lebih mudah,
- c. Proses query lebih cepat pada saat proses ETL

➤ **Kekurangan**

- a. Penggunaan akhir terhambat oleh kompleksitas
- b. Sulit mencari isi, karena sifatnya yang terlalu kompleks
- c. Performa menurun, hal ini karena query yang digunakan semakin banyak join table antar dimensi

BAB III

PENUTUP

3.1 Kesimpulan

Menurut Connolly dan Begg (2010:1227), star schema adalah model data dimensional yang mempunyai fact table di bagian tengah, dikelilingi oleh tabel dimensi yang terdiri dari data reference (yang bisa di-denormalized). Star schema mengambil karakteristik dari factual data yang di-generate oleh event yang terjadi dimasa lampau. Snowflake schema adalah salah satu metode pemodelan data untuk data warehouse yang merupakan pengembangan dari star schema, yaitu metode normalisasi untuk tabel dimensi yang ada dalam star schema (Ponniah, 2001). Menurut Inmon (2005) snowflake schema adalah gabungan dari beberapa pemodelan data star schema yang dinormalisasi.

Star Schema adalah salah satu implementasi desain logikal dari model multidimensional. Tujuan utama dari star schema adalah untuk membuat skema menjadi lebih simpel sehingga akan mempercepat proses query yang di dalamnya biasanya terdapat banyak perintah join. Snowflake schema merupakan bentuk perluasan dari star schema dimana terjadi proses normalisasi dari beberapa atau seluruh dimension table. Snowflake schema biasanya digunakan pada tabel yang sangat besar dan jika star schema tidak mampu menggambarkan kompleksitas dari database tersebut. Terkait komponen, karakteristik, serta proses pengerjaan masing-masing model memiliki kemiripan yang hampir sama. Secara umum, Star Schema unggul dalam hal performa sementara itu Snowflake unggul dalam hal kapasitas data.

3.2 Saran

Secara umum materi terkait dengan data warehouse ini merupakan materi yang sangat menarik, namun pembaca harus benar-benar pandai mencari informasi terkaitnya, hal ini karena beberapa penjelasan pada website tidak terlalu mudah dimengerti dan juga pembahasannya cukup sulit dicari.

DAFTAR PUSTAKA

Data Warehouse.(2014).”*Star Schema Snowflake Schema dan Starflake Schema*”.Dimuat pada <http://datawarehouse12.blogspot.com/2014/12/star-schema-snowflake-schemastarflake.html>. Diakses pada 2 Oktober 2020.

Levinbosz.(02 Januari 2014). “*Star Schema dan Snowflake Scehma*”. Dimuat pada <https://levinbosz.wordpress.com/2014/01/02/star-schema-dan-snowflake-schema-baru/#more-245>. Diakses pada 2 Oktober 2020.

Library Binus. (2018). “*Data Warehouse*”. Dimuat pada <https://library.binus.ac.id/eColls/eThesisdoc/Bab2/TSA-2012-0088%202.pdf>. Diakses pada 2 Oktober 2020.

Nanang.(2010). “*Skema Star dan Snowflake*”. Dimuat pada <http://nanang.lecture.ub.ac.id/2010/04/17/skema-star-dan-snowflake/> . 2 Oktober 2020.

Pujanggakos.(2017). “*Belajar Star Schema*”. Dimuat pada <https://pujanggakos.blogspot.com/2017/10/belajar-star-schema.html>. Diakses pada 2 Oktober 2020.

Yoyonb.(2009).”*Apakah Dimensional Modelling Dm*”. Dimuat pada <https://yoyonb.wordpress.com/2009/11/24/apakah-dimensioanal-modeling-dm/>. Diakses pada 2 Oktober 2020.