

Mata Kuliah :: Data Warehousing
Komposisi oleh Budi Prayitno
October 2012
Batam

Data Warehousing :: DWH Design

Data Warehouse Design

Desain Data Warehouse yang lazim digunakan adalah:

- Normalized Design
- Dimensional Design
- Hybrid Design

Normalized Design

Normalisasi database adalah teknik desain terstruktur untuk mengeliminasi anomali dan redundansi data

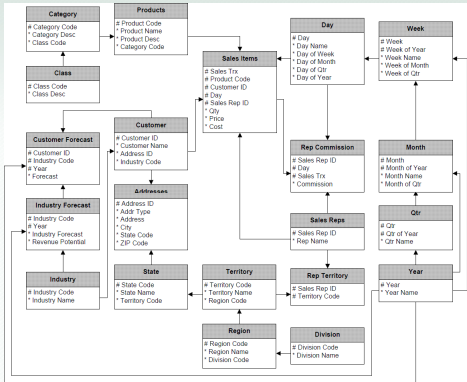
Diawali oleh E. F. Codd pada era 1970-an, normalisasi data menciptakan kebergantungan data serta mengurangi duplikasi dan inkonsistensi

Normalized Design...

Desain Normalized memberikan fleksibilitas dan efisiensi yang dibutuhkan pada Online Transactional Processing namun tidak menyediakan performa dan kemudahan untuk kebutuhan analisis data pada Business Intelligence

Normalized Design...

Pada awal era 1990-an, Bill Inmon membuat landasan untuk konstruksi Data Warehouse yang didasarkan pada bentuk normalisasi ketiga (3NF)



Dimensional Design

Desain Dimensional adalah teknik denormalisasi yang digunakan untuk menyediakan informasi historikal terkait dengan kebutuhan pengguna

Disebut juga sebagai Star Schema, desain Dimensional adalah struktur yang terdiri dari tabel Fact dan tabel Dimension

Dimensional Design...

Fact merepresentasikan proses, *event* atau aktifitas yang digunakan untuk mengukur sebuah bisnis

Sebuah tabel Fact terdiri dari dua jenis kolom:

- Kolom Foreign Key
Digunakan untuk melakukan JOIN dari tabel Fact ke tabel Dimension
- Kolom Metric/ukuran
Digunakan untuk menyimpan informasi mengenai event
Diklasifikasikan menjadi: *additive*, *semi-additive* dan *non-additive*

Dimensional Design...

- Ukuran *Additive*
Mendukung proses aritmatika terhadap keseluruhan Dimension
- Ukuran *Semi-Additive*
Mendukung proses aritmatika hanya pada sebagian Dimension-nya
- Ukuran *Non-Additive*
Merupakan nilai deskriptif dan tidak mendukung proses aritmatika

Dimensional Design...

Tabel Dimension berisi atribut dan hirarki yang mengelompokkan data secara logika

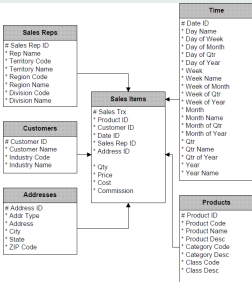
Atribut Dimension umumnya deskriptif dan memberikan detail mengenai Dimension tersebut

Hirarki pada sebuah Dimension secara natural memberikan summary dari level terendah hingga tertinggi

Contoh hirarki paling umum yang ditemukan pada Data Warehouse adalah waktu, yang menyediakan hirarki dengan level: hari, minggu, bulan dan tahun

Dimensional Design...

Dicetuskan oleh Ralph Kimball, model Dimensional sangat populer karena sederhana, mudah dikelola, kemampuan performa dan mudah dimengerti oleh pengguna



Hybrid Design

Desain Hybrid menggabungkan keunggulan dari teknik desain Normalized dan desain Dimensional, sehingga menyediakan solusi yang fleksibel dan berkesinambungan

Disebut juga sebagai Snowflake Design, desain ini menggunakan Fact dan Dimension namun juga melakukan normalisasi terhadap Dimension untuk mengurangi redundansi

Hybrid Design...

Desain Hybrid sangat berguna terutama jika Dimension mengandung banyak level yang terdiri dari banyak atribut

Normalisasi Dimension yang terdiri dari banyak atribut teks atau atribut dengan kardinalitas rendah

Desain Snowflake menyediakan fleksibilitas lebih terutama untuk membentuk hubungan *many-to-many* pada tabel Dimension dengan mencegah terlalu banyak level granularitas yang akan membingungkan pengguna bisnis ketika menggunakan atribut sebagai filter pada *query*

