



DATA WAREHOUSE PERTEMUAN 23

OUTLINE

AGGREGATE FACT TABLES

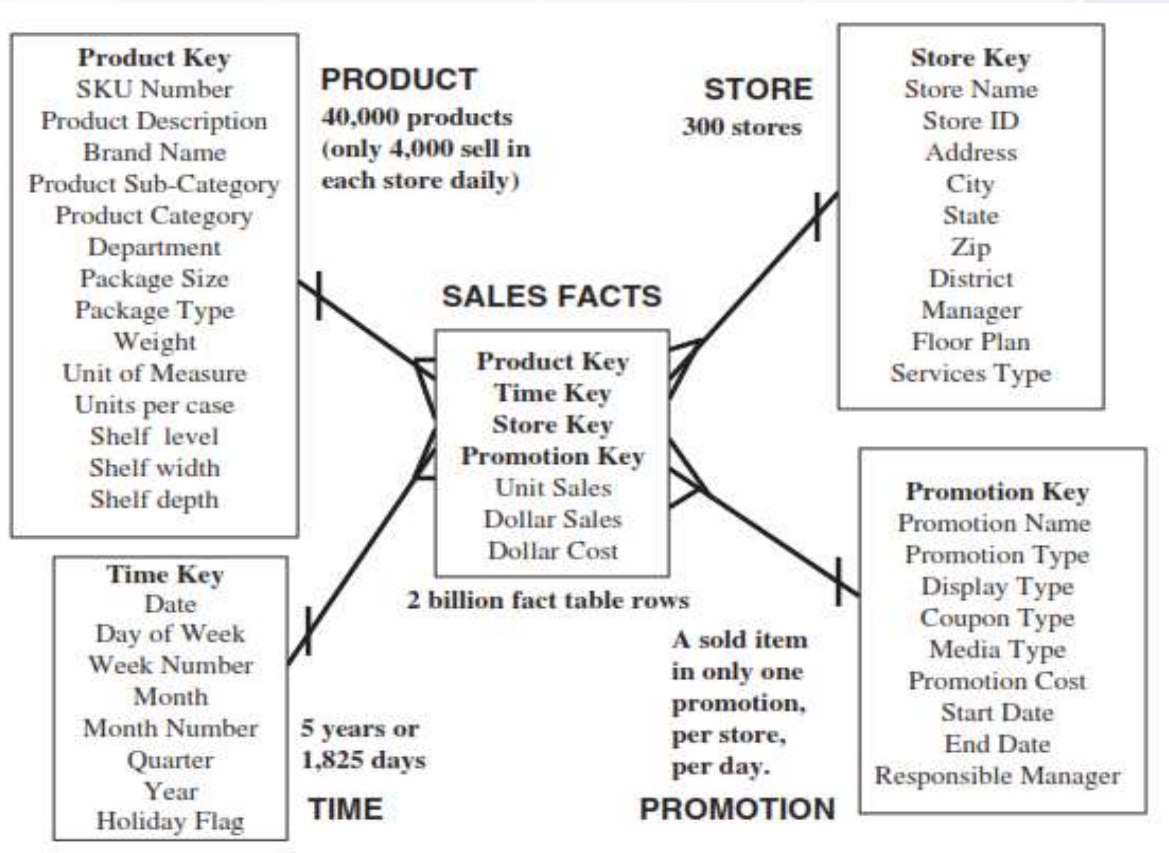
10.1 Fact Table Sizes

10.2 Need for Aggregates

10.3 Aggregating Fact Tables

10.4 Aggregation Options

10.1 Fact Table Sizes



Star Skema grocery chain

10.1 Fact Table Sizes (Lanjutan)

Ada sekitar dua miliar baris tabel fakta dasar dengan tingkat perincian terendah. Perhitungannya ditunjukkan di bawah ini:

- Dimensi waktu: 5 tahun x 365 hari = 1825
- Dimensi toko: 300 toko yang melaporkan penjualan harian
- Dimensi produk: 40.000 produk di setiap toko (4.000 penjualan di setiap toko setiap hari)
- Dimensi promosi: barang yang dijual mungkin hanya satu promosi di toko pada hari tertentu

Jumlah maksimum catatan tabel fakta dasar: $1825 \times 300 \times 4000 \times 1 = 2$ miliar

10.2 Need for Aggregates

Di 300 toko, asumsikan ada 500 produk per merek. Dari 40.000 produk, asumsikan bahwa ada di setidaknya satu penjualan per produk per toko per minggu. Mari kita perkirakan jumlah tabel fakta, baris yang akan diambil dan dirangkum untuk jenis query adalah sebagai berikut:

- Query melibatkan 1 produk, 1 toko, 1 minggu — ambil/rangkum hanya 1 baris tabel fakta
- Query melibatkan 1 produk, semua toko, 1 minggu — ambil/rangkum 300 baris tabel fakta
- Query melibatkan 1 merek, 1 toko, 1 minggu — ambil / rangkum 500 baris tabel fakta
- Query melibatkan 1 merek, semua toko, 1 tahun — ambil / rangkum 7.800.000 baris tabel fakta

10.3 Aggregating Fact Tables

- a. Multi-Way Aggregate Fact Tables
- b. Effect of Sparsity on Aggregation



Multi-Way Aggregate Fact Tables

- One-Way Aggregates
- Two-Way Aggregates
- Three-Way Aggregates



Effect of Sparsity on Aggregation

Dari kasus di atas (Need for Aggregates)

- Jika sparsity tetap sebagai 10% berlaku untuk tabel dasar, perkiraan nyata jumlah baris dalam tabel agregat akan jauh lebih sedikit.
- Ketika naik ke tingkat agregat yang lebih tinggi, persentase sparsitas naik dan bahkan mencapai 100%. Karena kegagalan sparsity akan tetap rendah.

10.4 Aggregation Options

Dari referensi ke kegagalan sparsity dalam tabel agregat, bahwa proses agregasi tidak mengurangi jumlah baris secara proporsional. Dengan kata lain, jika sparsity dari tabel fakta dasar adalah 10%, sparsity dari tabel agregat tingkat yang lebih tinggi tidak tetap pada 10%. Persentase sparsity meningkat lebih banyak dan ketika tabel agregat naik lebih tinggi dan dalam tingkat summarization lebih tinggi.

Goals for Aggregation Strategy

- Jangan terjebak dengan terlalu banyak agregat. Ingat, harus membuat dimensi turunan tambahan juga untuk mendukung agregat.
- Cobalah untuk melayani berbagai kelompok pengguna. Dalam hal apapun, sediakan untuk penggunaan listrik.
- Untuk agregat yang tidak terlalu meningkatkan penggunaan penyimpanan secara keseluruhan. Perhatikan dengan seksama ke dalam agregat yang lebih besar dengan persentase sparsity yang rendah.
- Menyimpan agregat yang disembunyikan dari pengguna akhir tetapi transparan ke permintaan pengguna akhir. Artinya, ini adalah alat query yang perlu menyadari agregat untuk mengarahkan pertanyaan untuk akses yang tepat.
- Berusahalah untuk menjaga dampak pada proses menampilkan data seminimal mungkin.

Review dan Latihan

☺ **END** ☺