

Assignment 03

Lab Penambahan Data pada Data Relasional



Data Relasional

Penyimpanan data yang efektif dan efisien adalah fondasi utama bagi organisasi untuk memperoleh wawasan berharga yang dapat mendukung pengambilan keputusan yang tepat waktu dan berbasis bukti. Saat ini, dalam industri yang terus berkembang, penggunaan skema relasional data menjadi semakin penting dan krusial. Skema relasional data memberikan kerangka kerja yang terstruktur untuk menyimpan, mengorganisir, dan mengelola data dengan cara yang teratur dan efisien.

Dalam konteks ini, penguasaan terhadap skema relasional data menjadi suatu hal yang esensial. Memahami konsep dasar data relasional, termasuk tabel, kolom, dan kunci-kunci hubungan antar entitas, menjadi pondasi bagi keberhasilan pengelolaan data dalam suatu organisasi. Dengan menguasai skema relasional data, para profesional dapat mengoptimalkan penggunaan database, meningkatkan efisiensi operasional, dan menghasilkan analisis yang lebih akurat dan bermakna pada data terstruktur.

Oleh karena itu, dalam sesi ini, Anda diminta untuk dapat meninjau kembali penggunaan skema data relasional yang sesuai untuk kebutuhan analisis dan pengambilan keputusan dalam konteks industri yang beragam, dengan fokus pada penggunaan bahasa SQL (Structured Query Language).

Review SQL

Tipe Data

1. **Integer**
Tipe data untuk menyimpan angka bulat (tidak memiliki desimal).
2. **Varchar**
Tipe data untuk menyimpan angka desimal dengan jumlah digit maksimum p, dan jumlah digit di belakang koma maksimum s.
3. **Date**
Tipe data untuk menyimpan tanggal (tahun, bulan, dan hari).

Untuk lebih lengkapnya dapat dilihat pada dokumentasi SQL yang ada di dokumentasi [ini](#).

Table Join



Tautan: https://documentation.mindsphere.io/resources/html/predictive-learning/en-US/Images/Joins_Diagram.png

Table Key

- Primary Key: key utama yang bersifat sebagai id dari tabel spesifik.
- Foreign Key: key dari suatu tabel lain yang bersifat sebagai referensi menuju primary key di tabel lain.

Data Terstruktur: Pengenalan Konsep *Data Warehousing*

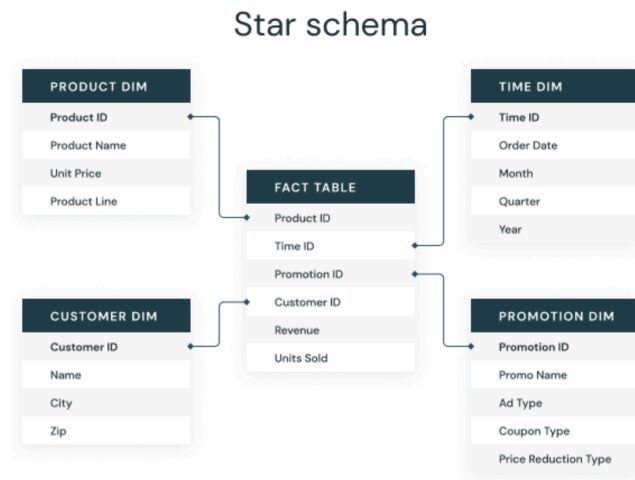
Data warehousing adalah pendekatan yang sangat penting dalam pengelolaan data bagi perusahaan modern. Dalam ekosistem data warehousing, terdapat dua konsep utama yang perlu dipahami: OLAP (Online Analytical Processing) dan OLTP (Online Transactional Processing). OLTP merujuk pada sistem yang bertanggung jawab atas transaksi sehari-hari, seperti pemesanan, pembelian, dan penjualan. Sebaliknya, OLAP difokuskan pada analisis data yang dihasilkan dari sistem OLTP untuk mendapatkan wawasan yang lebih mendalam.

Di dalam struktur data warehouse, data disusun dengan cara yang mendukung analisis OLAP. Salah satu elemen penting dalam struktur ini adalah tabel fakta (fact table) dan tabel dimensi (dimension table). Fact table berisi data yang berkaitan langsung dengan fakta bisnis yang ingin dianalisis, seperti volume penjualan atau jumlah produksi. Sementara itu, tabel dimensi menyediakan konteks tambahan untuk menganalisis data fakta, seperti waktu, lokasi, atau atribut produk. Dengan memahami konsep OLAP, OLTP, dan struktur data warehouse, perusahaan dapat mengoptimalkan penggunaan data mereka untuk pengambilan keputusan yang lebih cerdas dan strategis, meningkatkan kinerja operasional, dan mendapatkan keunggulan kompetitif di pasar yang semakin kompetitif.

Skema *Data Warehousing*

Snowflake schema dan star schema adalah dua jenis skema yang digunakan dalam perancangan data warehousing untuk menyusun hubungan antara tabel fakta dan dimensi.

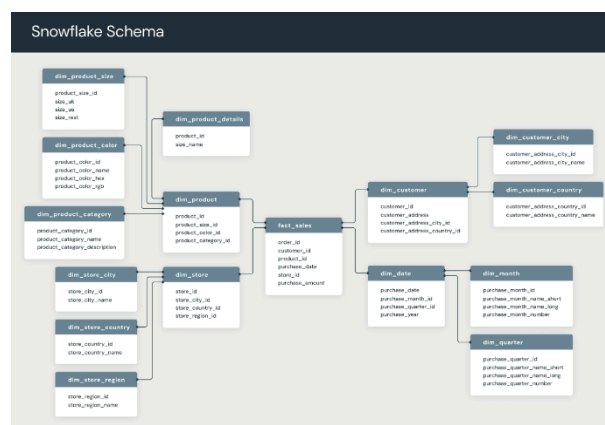
Star Schema:



Tautan: <https://www.databricks.com/wp-content/uploads/2022/04/star-schema-erd.png>

Star schema adalah skema yang paling sederhana dan populer dalam data warehousing. Dalam skema ini, terdapat satu tabel fakta yang berada di tengah dan dikelilingi oleh beberapa tabel dimensi. Tabel fakta berisi data numerik atau kuantitatif yang merepresentasikan fakta bisnis, seperti penjualan atau produksi. Sedangkan tabel dimensi berisi atribut-atribut deskriptif yang memperluas informasi dalam tabel fakta, seperti waktu, lokasi, produk, atau pelanggan. Skema ini memiliki struktur yang mirip dengan bintang, dimana tabel fakta berfungsi sebagai pusat (center) dan tabel dimensi sebagai lengan-lengan (spokes) yang menyertainya. Star schema relatif mudah dipahami, diimplementasikan, dan dijalankan, sehingga sering menjadi pilihan utama dalam desain data warehousing.

Snowflake Schema:



Tautan:

https://cms.databricks.com/sites/default/files/inline-images/snowflake-schema-120723_0.png

Snowflake schema adalah variasi dari star schema yang lebih kompleks. Dalam snowflake schema, tabel dimensi terbagi lagi menjadi sub-dimensi, yang kemudian terhubung ke tabel

dimensi utama. Dengan kata lain, tabel dimensi tidak hanya memiliki atribut tunggal, tetapi dapat mempunyai struktur hirarki yang lebih dalam. Misalnya, tabel dimensi waktu dapat terbagi menjadi tahun, bulan, dan hari. Struktur snowflake schema seringkali tampak seperti kristal salju, dengan cabang-cabang tambahan yang menyebabkannya lebih rumit daripada star schema. Meskipun snowflake schema dapat menyediakan tingkat normalisasi yang lebih tinggi dan menghemat ruang penyimpanan, namun dapat mempersulit query dan analisis data karena jumlah join yang lebih besar.

Kedua skema ini memiliki kelebihan dan kelemahan masing-masing tergantung pada kebutuhan bisnis dan kompleksitas data. Pemilihan antara star schema dan snowflake schema akan bergantung pada pertimbangan seperti performa, kemudahan pemeliharaan, dan kebutuhan analisis data yang spesifik.

Tugas: Implementasi SQL Query dan Menjawab Pertanyaan

Tugas untuk implementasi SQL query dengan tiga tabel **customers**, **orders**, dan **products** dirancang untuk memungkinkan Anda untuk mereview kembali pengetahuan dasar tentang basis data dan SQL query. Melalui tugas ini, Anda dapat memperdalam pemahaman Anda tentang konsep-konsep dasar SQL serta praktek-praktek terbaik dalam merancang dan menjalankan query SQL dalam konteks penambangan data.

Tujuan utama dari tugas ini adalah untuk:

- Menguatkan konsep-konsep dasar dalam SQL, termasuk SELECT, FROM, WHERE, GROUP BY, JOIN, dan fungsi agregasi.
- Berlatih dalam menggunakan perintah JOIN untuk menggabungkan data dari beberapa tabel, sehingga memperdalam pemahaman tentang hubungan antara tabel dalam database relasional.
- Meningkatkan kemampuan dalam menganalisis masalah dan merancang query yang tepat untuk mengekstrak data yang diperlukan.
- Memperdalam pemahaman tentang penggunaan fungsi agregasi seperti COUNT, SUM, dan AVG untuk merangkum data.
- Memahami prinsip-prinsip dasar pengoptimalan kinerja query SQL, termasuk penggunaan indeks dan strategi JOIN yang efisien.
- Melalui tugas ini, Anda memiliki kesempatan untuk mereshuffle dan memperdalam pengetahuan serta keterampilan Anda dalam SQL query, yang merupakan keterampilan yang sangat berharga dalam berbagai bidang industri dan pekerjaan yang melibatkan analisis data.

Tugas kedua adalah untuk menjawab beberapa pertanyaan yang berkaitan dengan konsep dasar dan pengenalan dari *data warehousing* yang ada pada notebook yang tersedia.

Bobot penilaian:

- Menjawab research question (60%)
- Mengimplementasikan initial table creation query (15%)
- Menjawab pertanyaan teori (25%)

Pengumpulan Tugas

Kumpulkan berkas dengan format penamaan seperti berikut:

Assignment3_[NPM]_[NamaLengkap].ipynb

Contoh:

Assignment3_1906438834_FransiscoErwando.ipynb

Apabila Anda mencari referensi dari internet, Anda WAJIB menyertakan sumbernya. Dilarang keras menyontek. Plagiarisme tidak ditoleransi dan akan dikenai penalti atau nilai akhir E.

Pengurangan nilai akibat keterlambatan pengumpulan tugas akan ditentukan berdasarkan jumlah menit keterlambatan Anda dalam mengumpulkan. Misalnya, apabila terlambat 1 menit, nilai akhir akan dikurangi 1 poin, apabila terlambat 10 menit, nilai akhir akan dikurangi 10 poin, dan seterusnya.