

PART 2 - Fundamental DE

TASK 1

1. Kapan kita harus menggunakan relational database atau nosql database ?

Jawab:

Relational database

- Jika aplikasi memerlukan transaksi yang konsisten dan integritas data yang tinggi (misalnya, aplikasi keuangan), RDBMS yang mendukung ACID (Atomicity, Consistency, Isolation, Durability) lebih cocok.
- Jika data memiliki struktur yang jelas dan tetap (misalnya, tabel dengan kolom tetap), RDBMS sangat cocok karena mendukung skema yang ketat.
- Jika sering bekerja dengan data yang memiliki banyak relasi dan perlu melakukan join antar tabel, RDBMS lebih efisien karena mendukung operasi join dengan baik
- RDBMS biasanya memiliki fitur keamanan dan manajemen transaksi yang kuat, yang penting untuk aplikasi yang membutuhkan kontrol akses yang ketat dan logging transaksi yang detail.

NoSQL

- Saat kita perlu mengelola dan menyimpan data dalam jumlah yang sangat besar, khususnya dalam skala Big Data yang dimana, NoSQL Database sanggup menangani data dalam jumlah besar tersebut, serta dapat memberikan performa pemrosesan yang cepat.
- Saat perlu mengelola data yang sifatnya semi-terstruktur ataupun tidak terstruktur seperti dokumen, grafik, ataupun data key-value pairs.
- NoSQL dapat menjadi pilihan yang baik ketika kita memerlukan skalabilitas yang mudah ditingkatkan secara horizontal (kemampuan untuk meningkatkan kapasitas sistem dengan menambahkan lebih banyak server atau node ke dalam cluster).
- Ketika ACID (Atomicity, Consistency, Isolation, Durability) sudah tidak dibutuhkan.
- Adanya hambatan antara objek untuk berelasi.
- RDBMS sudah tidak bisa dikembangkan lagi karena tidak bisa dibuat menjadi sistem terdistribusi.
- Ketika data yang bersifat temporal (sementara). Contoh ada wish list belanja pada sebuah marketplace, ketika wish list tersebut sudah tercapai maka data tersebut akan hilang dari wish list.

2. Apa perbedaan antara database, data lake, data warehouse, dan data mart ?

Jawab:

- Database adalah sekumpulan data yang terorganisir yang disimpan dalam sebuah sistem manajemen basis data (DBMS). Data disimpan dalam tabel yang terhubung satu sama lain melalui kunci dan relasi.
- Data Lake adalah repositori terpusat yang dirancang untuk menyimpan, memproses, dan mengamankan jumlah besar data semi-terstruktur dan tidak terstruktur. Ini dapat menyimpan data dalam format aslinya dan memproses berbagai jenis data, tanpa memperhatikan batasan ukuran.
- Data warehouse adalah sistem perusahaan yang digunakan untuk analisis dan pelaporan data terstruktur dan semi-terstruktur dari berbagai sumber, seperti transaksi titik penjualan, otomatisasi pemasaran, manajemen hubungan pelanggan, dan lain-lain. Data warehouse cocok untuk analisis ad hoc serta pelaporan kustom. Data warehouse dapat menyimpan baik data saat ini maupun historis dalam satu tempat dan dirancang untuk memberikan pandangan jangka panjang atas data dari waktu ke waktu.
- Data Mart adalah subset dari data warehouse yang difokuskan pada garis bisnis, departemen, atau area subjek tertentu. Data mart membuat data spesifik tersedia untuk kelompok pengguna yang ditentukan, yang memungkinkan pengguna tersebut untuk dengan cepat mengakses wawasan penting tanpa membuang waktu mencari melalui seluruh data warehouse. Contohnya banyak perusahaan mungkin memiliki data mart yang selaras dengan departemen tertentu dalam bisnis, seperti keuangan, penjualan, atau pemasaran.

3. Jelaskan apa itu normalisasi database, dan normalisasikan tabel dibawah !

employee_id	employee_name	job_code	job	city_code	city_name	province_code	province_name
1	John Smith	101	Software Engineer	201	New York	301	New York
2	Alice Johnson	102	Data Analyst	202	Los Angeles	302	California
3	Bob Davis	103	Data Engineer	203	Chicago	303	Illinois
4	Emily Wilson	101	Software Engineer	204	Houston	304	Texas
5	Michael Lee	102	Data Analyst	205	Miami	305	Florida
6	Sarah	103	Data Engineer	206	Boston	306	Massachusetts

	Brown						ts
7	James Clark	101	Software Engineer	207	San Fransisco	307	California
8	Laura Taylor	102	Data Analyst	208	Seattle	308	Washington
9	Daniel White	103	Data Engineer	209	Denver	309	Colorado
10	Olivia Martin	101	Software Engineer	210	Atlanta	310	Georgia

Jawab:

- Database normalization adalah prinsip desain basis data untuk mengorganisir data secara terstruktur dan konsisten.
- Tujuan utama dari normalisasi basis data adalah menghindari kompleksitas, menghilangkan duplikasi, menjaga integritas basis data, dan mengorganisir data secara konsisten. Dalam proses normalisasi, data dibagi menjadi beberapa tabel yang saling terhubung melalui hubungan.
- Normalisasi juga membantu menghilangkan karakteristik yang tidak diinginkan terkait dengan penyisipan, penghapusan, dan pembaruan data.

Normalisasi Tabel

Tabel Employee

Employee_id	Employee_name
1	John Smith
2	Alice Johnson
3	Bob Davis
4	Emily Wilson
5	Michael Lee
6	Sarah Brown
7	James Clark
8	Laura Taylor
9	Daniel White
10	Olivia Martin

Tabel Job

Job_code	Job
101	Software Engineer
102	Data Analyst
103	Data Engineer

Tabel City

City_code	City_name
201	New York
202	Los Angeles
203	Chicago
204	Houston
205	Miami
206	Boston
207	San Fransisco
208	Seattle
209	Denver
210	Atlanta

Tabel Province

Province_code	Province_name
301	New York
302	California
303	Illinois
304	Texas
305	Florida
306	Massachusetts
307	California
308	Washington
309	Colorado
310	Georgia