

Part 2 – Fundamental DE

1. Kapan kita harus menggunakan Relational Database atau NoSQL Database?
Penggunaan Relational Database dan NoSQL Database tergantung pada kebutuhan. Berikut adalah penjelasan lebih detailnya :
 - 1) Relational Database
Relational Database Management System (RDBMS) seperti MySQL, PostgreSQL, Oracle dan lain sebagainya memanfaatkan SQL (Structured Query Language) untuk manajemen data dan menggunakan tabel untuk menyimpan data. RDBMS sendiri biasa digunakan pada sektor perbankan, manajemen inventaris, atau aplikasi berbasis ERP (Enterprise Resource Planning), lalu untuk penggunaannya yaitu :
 - I. Ketika data yang akan di proses dalam bentuk terstruktur atau bisa disebut structured data dan dapat dibuat dalam skema tabel.
 - II. Ketika ingin menjalankan sebuah query yang kompleks, seperti join, union, agregasi, function, ataupun sub-query.
 - III. Jika kompatibilitas dengan alat dan aplikasi lain yang mendukung SQL.
 - 2) NoSQL Database
NoSQL Database seperti MongoDB, Cassandra, Redis dan lain sebagainya dapat menyimpan data dalam berbagai format, seperti dokumen, graf, kolom, dan key-value. NoSQL Database sendiri biasa digunakan pada sektor media sosial, analisis, maupun big data. Kemudian untuk penggunaannya yaitu :
 - I. Ketika data yang akan di proses dalam bentuk semi terstruktur (semi-structured data) atau tidak terstruktur (unstructured data) seperti JSON, data graf, atau key-value.
 - II. Ketika mengelola data dalam jumlah besar.
 - III. Ketika melakukan sebuah pengembangan yang cepat dan perubahan skema yang sering serta lebih fleksibel saat mengubah data.
2. Apa perbedaan antara Database, Data Lake, Data Warehouse, dan Data Mart?
 - 1) Database digunakan untuk menyimpan data operasional yang mendukung operasi bisnis yang memerlukan data terstruktur atau memproses transaksi.
 - 2) Data Lake digunakan untuk menyimpan raw data dalam jumlah besar dari berbagai sumber yang diperlukan untuk big data atau integrasi data dan terdiri dari macam macam format data seperti teks, gambar, video, JSON, CSV, dan lain sebagainya tanpa batas ukuran file.
 - 3) Data Warehouse digunakan untuk menyimpan historis data terstruktur, pembuatan pelaporan, atau business intelligence yang diambil dari berbagai sistem operasional dengan ETL (Extract, Transform, Load) sebelum disimpan. Biasanya menggunakan skema snowflake atau star untuk mengatur data.
 - 4) Data Mart digunakan untuk menyimpan subset data dari Data Warehouse yang di klasifikasikan sesuai fungsi dan kebutuhan tertentu sehingga ruang lingkupnya lebih sempit dan terfokus. Data Mart memiliki akses yang lebih cepat ke data tertentu yang dibutuhkan untuk kepentingan analisis dan pelaporan 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 Brown	103	Data Engineer	206	Boston	306	Massachusetts
7	James Clark	101	Software Engineer	207	San Francisco	307	California
8	Laura Taylor	102	Data Analyst	208	Seattle	308	Washington

- 1) Normalisasi database merupakan sebuah proses mengatur tabel pada sebuah database untuk menghilangkan duplikasi data dan memastikan bahwa data terkait disimpan dengan cara yang logis dan efisien.
- 2) Berikut ini merupakan hasil dari normalisasi tabel diatas :

I. Tabel Employee

employee_id	employee_name	job_code	city_code
1	John Smith	101	201
2	Alice Johnson	102	202
3	Bob Davis	103	203
4	Emily Wilson	101	204
5	Michael Lee	102	205
6	Sarah Brown	103	206
7	James Clark	101	207
8	Laura Taylor	102	208

II. Tabel Job

job_code	job
101	Software Engineer
102	Data Analyst
103	Data Engineer

III. Tabel City

city code	city name	province code
201	New York	301
202	Los Angeles	302
203	Chicago	303
204	Houston	304
205	Miami	305
206	Boston	306
207	San Francisco	307
208	Seattle	308

IV. Tabel Province

province code	province name
301	New York
302	California
303	Illinois
304	Texas
305	Florida
306	Massachusetts
307	California
308	Washington