

DATA ENGINEER

UNIT 2 – PART 2 – Fundamental Data Engineer

Presentation by
Nurul Kholifah

BAGIAN 1

Kapan kita harus menggunakan relational database atau nosql database ?

Relational Database digunakan pada data terstruktur dan memiliki schema yang tetap atau paten seperti pada tabel dengan kolom dan baris. Relational Database tepat digunakan untuk aplikasi dengan jumlah data yang tidak terlalu besar.

Sedangkan NoSQL digunakan pada data fleksibel schema dan support semi-struktur data. NoSQL memiliki tipikal distribusi sistem yang akan memecah data menjadi node-node tertentu. Data pada NoSQL berbasis document, graph, key-value, dan wide-column.



BAGIAN 3

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

- Database adalah kumpulan data terstruktur yang disusun sedemikian rupa untuk menyimpan, mengelola, dan menyediakan akses efisien terhadap informasi.
- Data lake digunakan untuk menyimpan data mentah, baik terstruktur maupun tidak, dan memungkinkan analisis mendalam tanpa membatasi format atau jenis data tertentu. Tools yang digunakan biasanya Amazon S3, Google Cloud Storage
- Data Warehouse adalah penyimpanan data terstruktur yang telah diproses dan dioptimalkan untuk hoc analysis sesuai dengan kebutuhan data analysis. Tools yang digunakan Amazon Redshift, Google BigQuery
- Data Mart merupakan bagian atau subset dari data warehouse yang berfokus dan terbagi menjadi beberapa bagian berisikan data khusus untuk mendukung kebutuhan spesifik kelompok atau departemen.

BAGIAN 3

Jelaskan apa itu normalisasi database, dan normalisasikan tabel dibawah!

Normalisasi database adalah proses pengorganisasian struktur data yang bertujuan mengurangi redundansi data dan meningkatkan efisiensi pengelolaan data. Normalisasi database bermanfaat untuk menghindari anomali data, meningkatkan konsistensi data, dan mengoptimalkan desain struktur relasional

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
9	Daniel White	103	Data Engineer	209	Denver	309	Colorado
10	Olivia Martin	101	Software Engineer	210	Atlanta	310	Georgia

normalization

employee_id	employee_name	job_code
1	John Smith	101
2	Alice Johnson	102
3	Bob Davis	103
4	Emily Wilson	101
5	Michael Lee	102
6	Sarah Brown	103
7	James Clark	101
8	Laura Taylor	102
9	Daniel White	103
10	Olivia Martin	101

job_code	job
101	Software Engineer
102	Data Analyst
103	Data Engineer

employee_id	city_code
1	201
2	202
3	203
4	204
5	205
6	206
7	207
8	208
9	209
10	210

city_code	city_name
201	New York
202	Los Angeles
203	Chicago
204	Houston
205	Miami
206	Boston
207	San Francisco
208	Seattle
209	Denver
210	Atlanta

employee_id	province_code
1	301
2	302
3	303
4	304
5	305
6	306
7	307
8	308
9	309
10	310

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

The background is a dark blue field decorated with various geometric elements. It includes numerous small squares in white, light blue, and light orange, some of which are solid while others are outlines. Thin, light blue vertical lines of varying lengths are scattered across the composition, creating a modern, minimalist aesthetic.

TERIMA KASIH