**FUNDAMENTAL OF DE - PART 2**

**TASK 1**

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

Kita harus menggunakan *relational database* jika data yang kita gunakan:

1. Memiliki struktur data yang tetap (*structured data*).
2. Membutuhkan konsistensi yang tinggi, yaitu ACID (Atomicity, Consistency, Isolation, Durability) adalah empat prinsip dalam manajemen basis data (*database*).
3. Memerlukan kueri kompleks yang melibatkan join antara beberapa tabel.

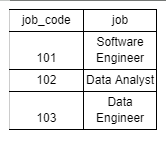
Kita harus menggunakan *NoSQL database* jika data yang kita gunakan:

1. Memiliki data *semi-structured* atau *unstructured*.
2. Memerlukan skalabilitas horizontal untuk menangani volume data yang besar karena NoSQL dapat memberikan fleksibilitas dalam menambahkan node.
3. Membutuhkan kinerja yang sangat tinggi.
4. Memerlukan fleksibilitas *schema* karena NoSQL dapat memberikan fleksibilitas untuk mengubah skema data tanpa perlu menghentikan atau mengubah struktur database.
5. Memerlukan pengelolaan data yang mudah tanpa menggunakan kompleksitas pengelolaan basis data yang relatif tinggi dari *relational database*.
6. Apa perbedaan antara *database, data lake, data warehouse,* dan *data mart*?

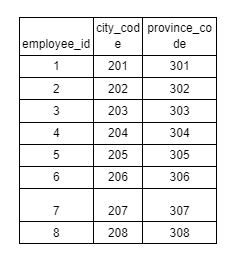
* *Database* adalah kumpulan data yang terorganisir secara terstruktur untuk menyimpan, mengelola, dan mengakses data. Contoh: PostgreSQL, MySQL, Oracle.
* *Data lake* adalah penyimpanan data yang memungkinkan kita menyimpan data mentah dalam berbagai format tanpa mengubahnya dan mengabaikan ukuran data dimana struktur data bisa berupa semi-terstruktur maupun *unstructured*. Contoh: Apache Hadoop, Amazon S3, Google CLoud Storage.
* *Data warehouse* adalah penyimpanan data yang dirancang untuk analisis data terstruktur maupun semi-terstruktur dan pelaporan bisnis dalam suatu tempat yang berasal dari berbagai sumber data. Contoh: Amazon Redshift, Google BigQuery, Snowflake.
* *Data mart* adalah subset dari *data warehouse* yang berfokus pada satu area fungsional atau unit bisnis dimana memiliki struktur data berupa skema bintang atau *snowflake* yang sama dengan data warehouse. Contoh: *Finance data mart, Sales data mart, HR data mart*.

1. Jelaskan apa itu normalisasi *database* dan normalisasikan tabel di bawah ini.

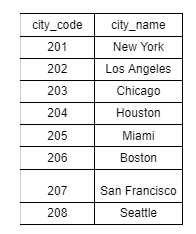
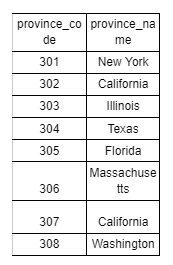
Normalisasi *database* adalah proses desain database yang dilakukan untuk mengorganisir struktur tabel agar sesuai dengan prinsip-prinsip normalisasi yang bertujuan untuk mengurangi redudansi data dan memastikan dependensi data yang konsisten, sehingga menghasilkan desain database yang lebih efisien, fleksibel, dan mudah dipelihara.

Hasil normalisasi data dari soal.









| job\_code | job |
| --- | --- |
| 101 | Software Engineer |
| 102 | Data Analyst |
| 103 | Data Engineer |

| 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 |

| employee\_id | city\_code | province\_code |
| --- | --- | --- |
| 1 | 201 | 301 |
| 2 | 202 | 302 |
| 3 | 203 | 303 |
| 4 | 204 | 304 |
| 5 | 205 | 305 |
| 6 | 206 | 306 |
| 7 | 207 | 307 |
| 8 | 208 | 308 |

| city\_code | city\_name |
| --- | --- |
| 201 | New York |
| 202 | Los Angeles |
| 203 | Chicago |
| 204 | Houston |
| 205 | Miami |
| 206 | Boston |
| 207 | San Francisco |
| 208 | Seattle |

| province\_code | province\_name |
| --- | --- |
| 301 | New York |
| 302 | California |
| 303 | Illinois |
| 304 | Texas |
| 305 | Florida |
| 306 | Massachusetts |
| 307 | California |
| 308 | Washington |