**Rangkuman Teknologi dan Perangkat Pendukung Data Science**

Menurut**Amazon Web Service (AWS)** basis data adalah koleksi data sistematis yang disimpan secara elektronik. Ini dapat berisi semua jenis data, seperti kata, angka, gambar, video, dan file. Anda dapat menggunakan perangkat lunak (*software*) yang disebut Database Management System (DBMS) atau sistem manajemen basis data.

Ada pula menurut sumber lain, yaitu **Oracle** menyatakan bahwa basis data adalah kumpulan terorganisasi dari informasi terstruktur, atau data, biasanya disimpan secara elektronik dalam sistem komputer. Basis data biasanya dikendalikan oleh Database Management System (DBMS) atau sistem manajemen basis data. Nah, data dan DBMS beserta aplikasi yang terkait dengannya disebut sebagai sistem basis data atau juga bisa disebut basis data saja.

**Tipe-Tipe Basis Data**

1. Basis Data Relasional (*Database Relational*)
2. Basis Data Orientasi Objek (OOD - *Object-oriented Databases*)
3. Basis Data Terdistribusi (*Distributed Databases*)
4. Gudang Data (Data Warehouse)
5. Basis Data NoSQL
6. Basis Data Grafik (Graph Databases)

**Data, Basis Data, Dataset**

Menurut Kamus Besar Bahasa Indonesia, data adalah informasi, fakta, dari keterangan yang benar dan nyata mengenai suatu objek atau kejadian yang bisa dicatat atau disimpan.

| **Basis Data** | **Dataset** |
| --- | --- |
| Basis data adalah kumpulan data yang diatur dan disimpan dengan cara yang memungkinkan akses dan pengambilan yang mudah. Selain itu, ia juga dapat menyimpan berbagai macam tipe data, termasuk teks, nomor, gambar, dan tipe data lainnya. | Dataset adalah kumpulan data diatur dalam format tertentu, seperti spreadsheet, CSV, atau basis data. Umumnya, digunakan untuk keperluan spesifik. |
| Memiliki banyak kumpulan data dan dapat digunakan untuk aplikasi yang berbeda. | Dapat menjadi bagian dari data yang diambil dari basis data yang lebih besar. Bisa juga, dataset diambil dari sumber lain yang mendukung formatnya. |
| Contoh tools: BigQuery, MySQL Clients, DB Browser SQL, dll. | Contoh tools: Google Sheets, Excel, Situs Public Datasets, dll. |

**SQL (Structured Query Language)**

Pada bagian ini Anda akan dikenalkan dengan dua jenis himpunan yang terdapat pada SQL, yaitu DDL dan DML. Simak penjelasannya di bawah ini.

**Data Definition Language**

DDL merupakan subperintah pada SQL yang dimanfaatkan guna membangun kerangka basis data. Tabel di bawah merupakan pernyataan *statements* yang utama dari DDL.

|  |  |  |
| --- | --- | --- |
| CREATE SCHEMA | ALTER DOMAIN | DROP SCHEMA |
| CREATE DOMAIN | DROP DOMAIN |
| CREATE TABLE | ALTER TABLE | DROP TABLE |
| CREATE VIEW | DROP VIEW |

**Data Manipulation Language**

DML merupakan subperintah pada SQL yang dimanfaatkan dalam manipulasi basis data yang sudah dibuat. Pada dasarnya, dalam DML terdapat empat perintah berikut beserta fungsinya.

| **Data Manipulation Language** | **Fungsi** |
| --- | --- |
| INSERT | Digunakan untuk memasukkan data baru dalam tabel. Perintah ini dapat dijalankan saat basis data dan tabel telah selesai dibuat. |
| SELECT | Digunakan untuk mengambil, lalu menampilkan data dari tabel atau sejumlah tabel dengan memanfaatkan relasi. |
| UPDATE | Digunakan untuk memperbarui data dalam tabel. |
| DELETE | Digunakan untuk menghapus data dari tabel. |

**NoSQL (Not Only SQL)**

NoSQL atau disebut juga sebagai “*not only SQL”*, “*non-SQL”*, adalah pendekatan untuk desain basis data yang memungkinkan penyimpanan dan kueri data di luar struktur tradisional yang ditemukan dalam basis data relasional.

Basis data NoSQL umumnya diklasifikasikan ke dalam empat kategori utama, di antaranya

1. *Document Databases*  
   Basis data ini menyimpan data sebagai dokumen semi-terstruktur, seperti JSON atau XML, dan dapat dikueri menggunakan bahasa kueri berorientasi dokumen.
2. *Key-value Stores*  
   Basis data ini menyimpan data sebagai *key-value pairs*dan dioptimalkan untuk operasi baca atau tulis yang sederhana dan cepat.
3. *Column-family stores*  
   Basis data ini menyimpan data sebagai keluarga kolom, yaitu kumpulan kolom yang diperlakukan sebagai satu kesatuan. Mereka dioptimalkan untuk kueri data dalam jumlah besar yang cepat dan efisien.
4. *Graph Databases*  
   Sempat dibahas pada materi sebelumnya, bahwa basis data ini menyimpan data sebagai *node*dan *edge*, serta dirancang untuk menangani hubungan kompleks antar data.

**Database on Premise vs. Cloud**

**Basis Data on Premise**

*On-premise*adalah jenis server berupa *software*yang dijalankan secara internal oleh tim IT perusahaan. Tim IT tersebut bertugas dalam menjalankan aplikasi server, memasang sistem operasi, dan meletakkan server di sebuah gedung.

**Basis Data on Cloud**

Berbeda dengan *on-premise, cloud server*alias *cloud computing*merupakan layanan infrastruktur berupa penyimpanan basis data, server, jaringan, dan *software*berbasis internet. Layanan ini tidak disediakan oleh tim IT perusahaan, tetapi melalui pihak ketiga yang merupakan penyedia layanan *cloud.*

**Tools Pengolahan Data**

Di bawah ini merupakan *tools*pengolahan data yang umumnya sering digunakan.

* Excel
* Google Sheets
* SPSS (Statistical Package for the Social Sciences)

**Tools Visualisasi Data**

Di bawah ini merupakan *tools*visualisasi data yang umumnya sering digunakan.

* Tableau
* Metabase
* Looker Studio

**Bahasa Pemrograman**

**Berkelana dengan Python**

Python adalah bahasa pemrograman yang banyak digunakan dalam aplikasi web, pengembangan perangkat lunak, ilmu data, dan *machine learning* (ML).

**Berkenalan dengan R**

Kemudahan dalam penggunaannya serta mudahnya akses dukungan komunitas, membuat R menjadi salah satu bahasa pemrograman paling populer di dunia. R memiliki aturan atau sintaks yang berbeda dengan bahasa pemrograman lain yang membuatnya memiliki ciri khusus tersendiri.

**Teknologi Machine Learning**

*Machine learning* adalah ilmu pengembangan algoritma yang memanfaatkan konsep matematis dan statistik dalam menjalankan tugas tertentu tanpa instruksi eksplisit. Pada prosesnya, *machine learning* akan berusaha mengenal pola yang terdapat dalam sebuah data serta menggunakannya untuk menghasilkan prediksi.

Secara garis besar, kita dapat mengelompokkan *machine learning*ke dalam tiga kategori yaitu seperti berikut.

* *Supervised learning*.
* *Unsupervised learning.*
* *Reinforcement learning.*