Teknologi penyimpanan basis data

Berkenalan dengan basis data

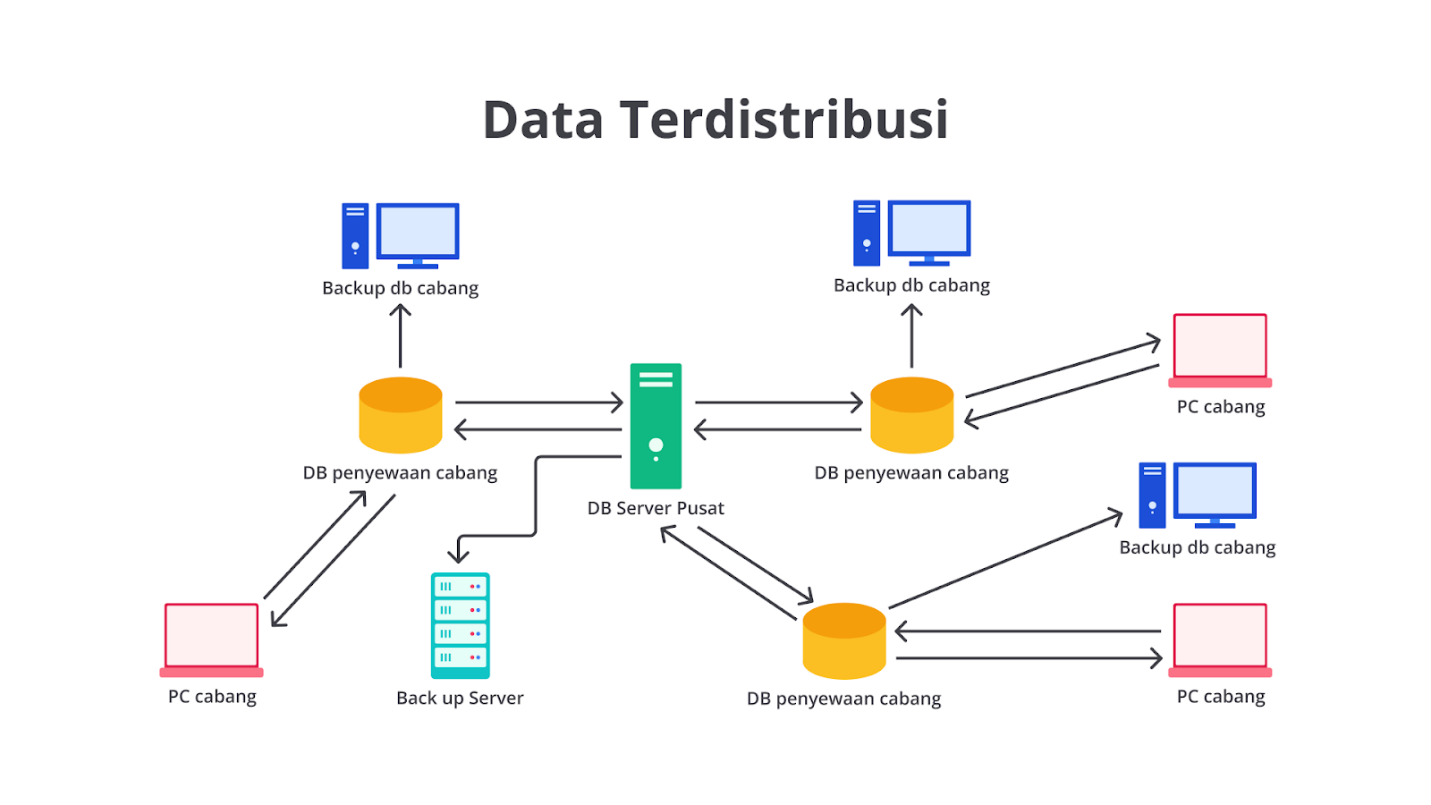
**Pengertian**

Menurut amzon Web service (AWS) basis data adalah koleksi data sistematis di simpan secara electronic.

Menurut Oracale menyatakan bahwa basis data adalah kumpulan terorganisasi dari informasu terstruktur, atau data, biasanya disimpan secara electronic dalam system computer.

Tipe-Tipe Basis data

1. Basis data relasional(database relational)
2. Basis data orientasi objek(OOD – Object-oriented Databases)
3. Object oriented database adalah sebuah model basis data yang menyimpan informasi dalam bentuk objek. Beberapa produk yang biasa digunakan adalah IBM Db2, InterSystems IRIS, Google Cloud Storage for Firebase, dll.
4. Basis data Terdistribusi



Basis data terdistribusi adalah kumpulan data yang digunakan bersama-sama dan memiliki hubungan logis fisik tersebar di jaringan komputer. Konsep ini mendasarkan pada prinsip bahwa data dibuat, disediakan, dan dikelola secara terpusat.

1. Gudang Data (data Werhouse)

Sebuah sistem atau platform yang dirancang untuk mengumpulkan, menyimpan, menganalisis, dan mengelola data dari berbagai sumber yang berbeda untuk tujuan analisis bisnis dan pengambilan keputusan. Beberapa tools data warehouse di antaranya Amazon Redshift, Snowflake, Google BigQuery, dll.

1. Basis Data NoSQL



NoSQL atau basis data non-relasional, memungkinkan data tidak terstruktur dan semi terstruktur untuk disimpan dan dimanipulasi (berbeda dengan basis data relasional, yang menentukan semua data yang dimasukkan ke dalam basis data harus disusun). Beberapa contoh basis data NoSQL antara lain MongoDB, Cassandra, dan Couchbase.

1. Basis data grafik(graph Databases)

Basis data everywhere

Berbisnis dengan hati nurani sangat bermanfaat” -

Ingvar Kamprad (CEO IKEA)

Quoets di atas bisa kita kontekskan kedalam manfaat dari basis data yg di gunakan industri untuk berbagai kasus penggunaanya.

1. **Deteksi penipuan**  
   Basis data grafik membantu manajemen identitas dan mendeteksi penipuan. Ketika digabungkan dengan algoritma *machine learning,* ia dapat menemukan pola dan mendeteksi aktivitas penipuan secara otomatis.
2. **Manajemen dokumen**  
   Basis data NoSQL juga dapat menyimpan dan mengelola dokumen, seperti artikel dan kontrak. Basis data NoSQL juga memungkinkan organisasi untuk melakukan kueri dan mengindeks dokumen.
3. **Gaming dan hiburan**  
   Banyak perusahaan *gaming* dan hiburan menggunakan basis data secara luas untuk memberikan pengalaman bermain yang lebih baik, seperti akses masuk simultan untuk jutaan pengguna.

Data, Basis data, dataset

Apa sih perbedaan dari 3 hal di atas?

Data menurut kbbi adalah informasi, fakta, dari keterangan yang benar dan nyata mengenai suatu objek atau kejadian yang bisa dicatat atau disimpan.

Basis data

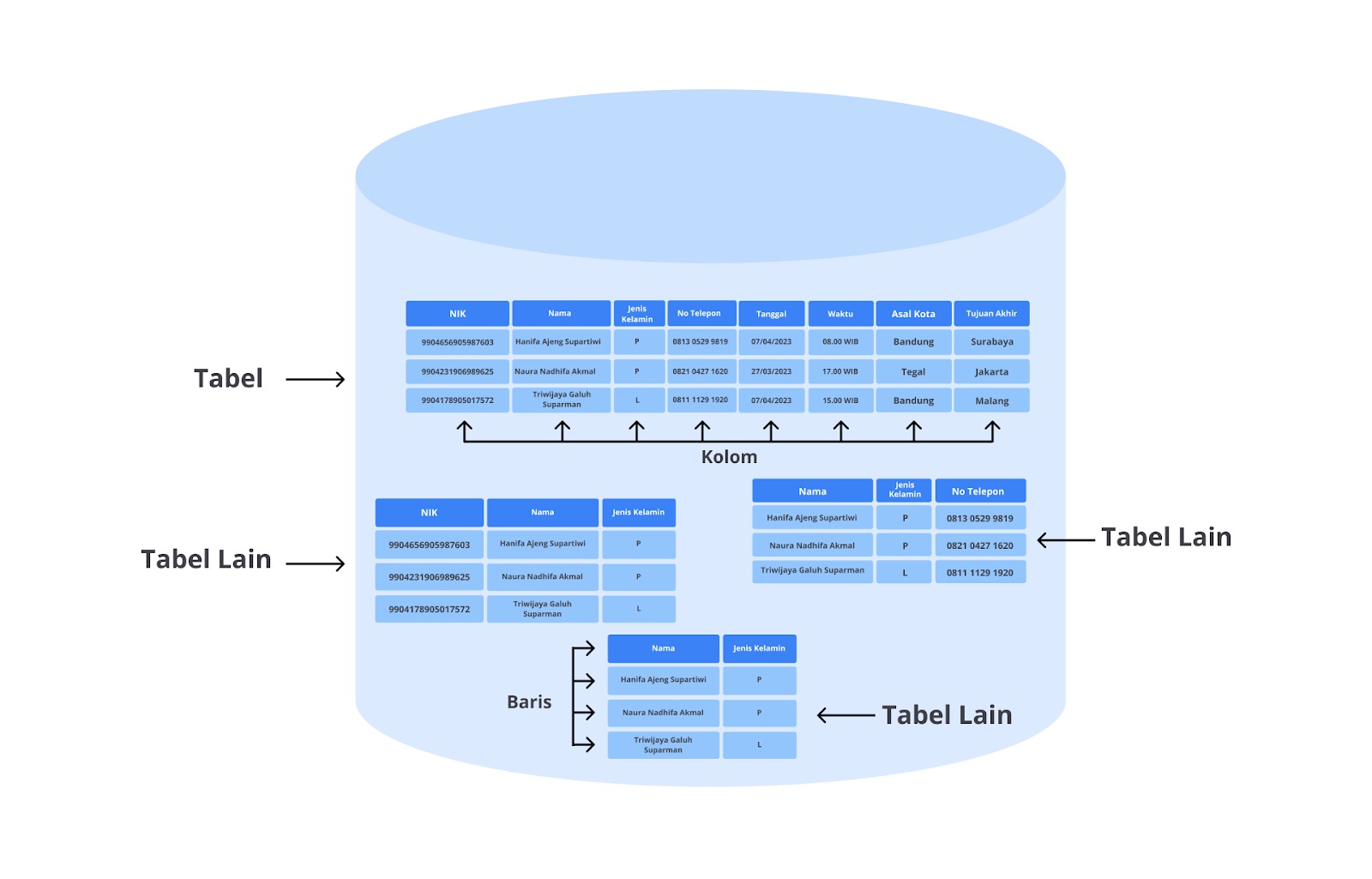
Basis data adalah kumpulan data yang diatur dan disimpan dengan cara yang memungkinkan akses dan pengambilan yang mudah. Selain itu, ia juga dapat menyimpan berbagai macam tipe data, termasuk teks, nomor, gambar, dan tipe data lainnya

Contoh tools: BigQuery, MySQL Clients, DB Browser SQL, dll.

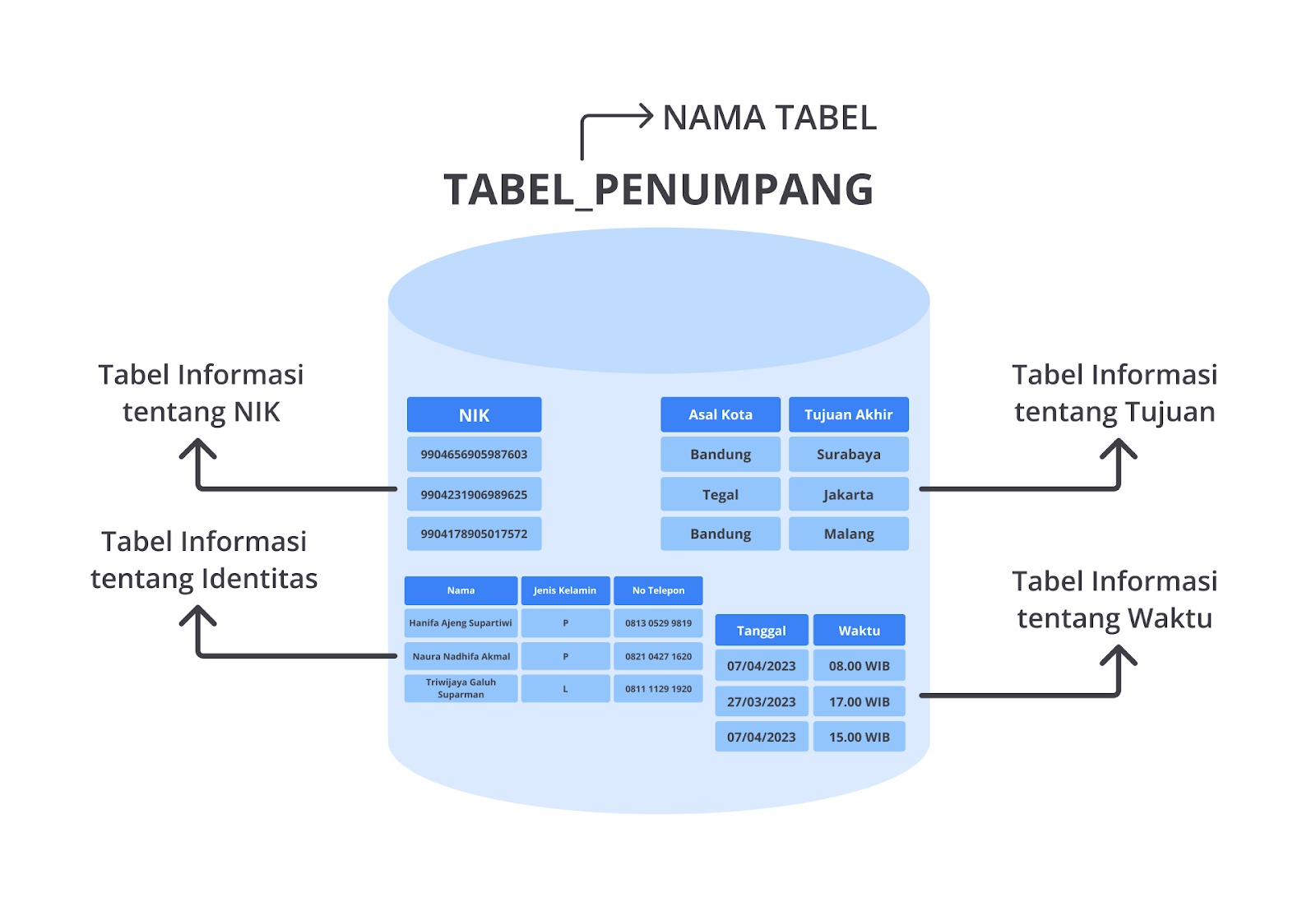
Dataset

Dataset adalah kumpulan data diatur dalam format tertentu, seperti spreadsheet, CSV, atau basis data. Umumnya, digunakan untuk keperluan spesifik.

Contoh tools: Google Spreadsheet, Excel, Situs Public Datasets, dll.

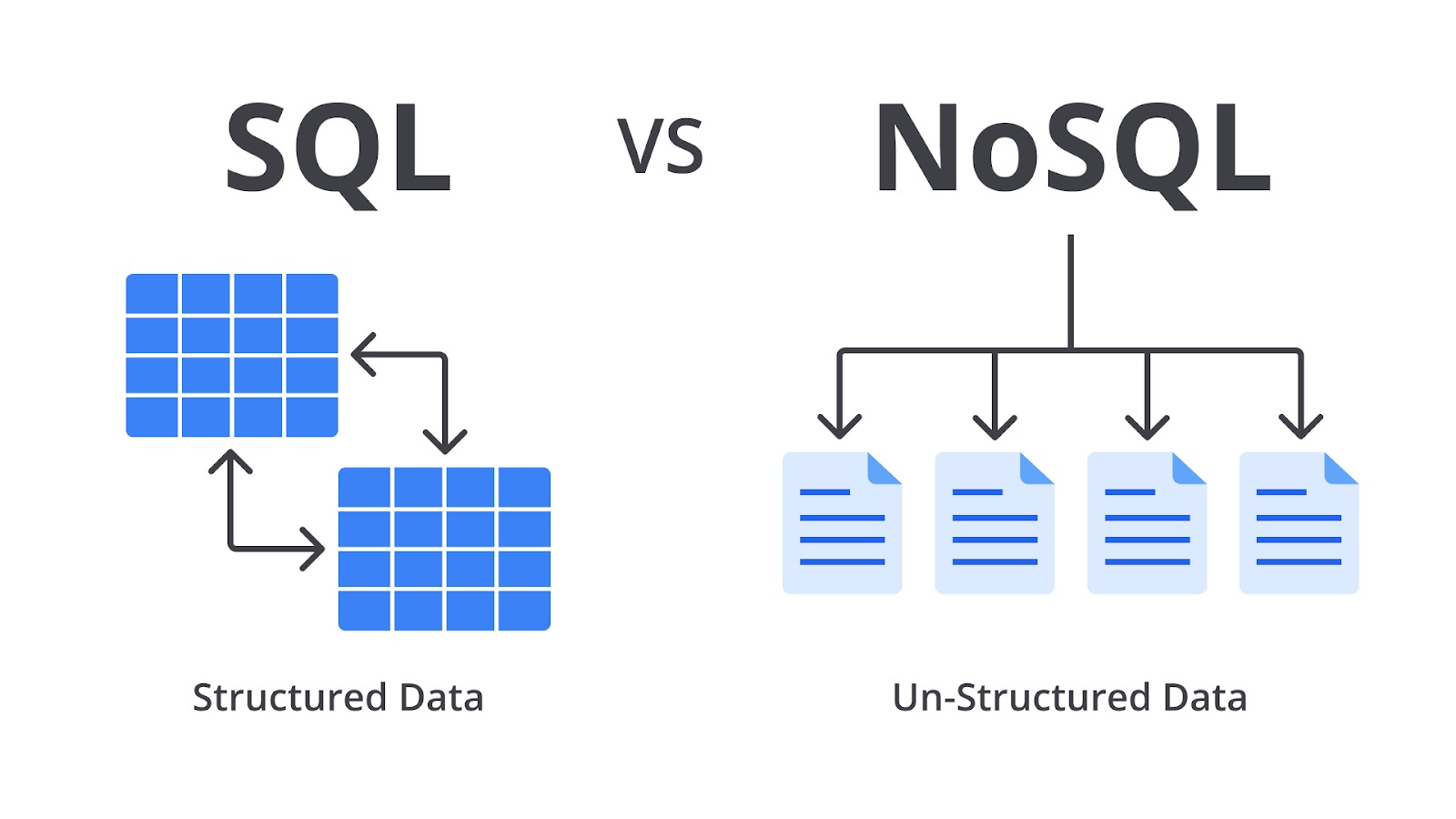


Sederhananya, gambar tabung di atas merupakan representasi dari sebuah basis data. Tabel yang terdapat di dalamnya adalah struktur dalam basis data yang diatur dengan bentuk kolom dan baris, serta basis data juga terdiri dari data-data yang saling terhubung seperti berikut.



Namun, apabila tabel-tabel tersebut dibungkus dengan menggunakan format tertentu, ia disebut sebuah dataset. Sederhananya adalah basis data terdiri dari dataset yang saling berhubungan satu sama lainnya.

Rupa-Rupa Sistem Basis data



Perlu Anda ketahui bahwa SQL (Structured Query Language) dikembangkan oleh IBM pada tahun 1970 dengan tujuan untuk memanipulasi dan menarik data yang tersimpan pada IBM *database management system*(DBMS) yang disebut System R.

Seiring dengan perkembangan teknologi, di awal abad 21, muncul teknologi baru yang dikenal dengan NoSQL (No SQL atau Not Only SQL). Teknologi ini dipicu oleh kebutuhan-kebutuhan perusahaan Web 2.0 yang mengacu pada website dengan titik berat pada user, penggunaan yang mudah, dan sesuai dengan sistem yang digunakan oleh end user.

1. **SQL (Structured Query Language)**

#### NoSQL (Not Only SQL)

#### 

### 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.

Gambar di atas merupakan penyimpanan basis data on-premise yang biasa disebut RAM atau ROM. Dapat dibilang on-premise adalah sebuah vendor, yakni pemilik layanan on-premise memberikan lisensi penuh kepada tim tersebut untuk mengelola basis data di dalamnya. Jadi, on-premise bukanlah jaringan server secara jarak jauh atau remote atau secara sederhananya basis data on-premise disimpan di dalam memori internal komputer dari sebuah vendor.

#### 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.

Berikut merupakan 5 alasan mengapa Anda perlu belajar dan menggunakan SQL dalam data science.

1. **Sangat kuat**  
   SQL adalah bahasa yang kuat karena dapat digunakan untuk memanipulasi data, membuat tabel baru, memasukkan data ke dalam tabel, dan mengambil hasil kueri.
2. **Familier**  
   SQL adalah bahasa standar untuk berinteraksi dengan basis data. Ini berarti bahwa jika Anda tahu cara menulis kueri dalam SQL, Anda dapat menggunakan keterampilan yang sama dalam aplikasi atau alat basis data apa pun.
3. **Mudah dibagikan**  
   Para data scientist perlu memahami bahasa ini karena memungkinkan mereka untuk berkolaborasi dengan orang lain di organisasi atau perusahaan yang memiliki keterampilan berbeda tetapi membutuhkan akses ke informasi yang sama. Nah, SQL merupakan sebuah jawaban untuk hal tersebut karena mudah diakses dan fleksibel.
4. **Sangat umum**  
   SQL merupakan bahasa yang digunakan para data scientist, data analyst, dan pengguna bisnis untuk membuat kueri basis data. Ini adalah bahasa yang paling umum untuk membuat kueri *data warehouse* dan *data lakes*.
5. **Relevan**  
   SQL relevan karena digunakan untuk banyak tugas data science. Anda menggunakan SQL untuk menjelajahi data dan memahaminya dengan lebih baik, membersihkan data, menyiapkan data untuk analisis, membangun model di atas kumpulan data yang telah dibersihkan, dan disiapkan memvisualisasikan hasil dan melaporkannya.

PERANGKAT PENGELOLA DATA

### Excel dan Spreadsheet

### SPSS (Statistical Package for the Social Sciences)

SPSS adalah program komputer yang dipakai untuk analisis statistika. SPSS digunakan untuk pengolahan dan menganalisis data yang memiliki kemampuan analisis statistik serta sistem manajemen data dengan lingkungan grafis.

PERANGKAT VISUALISASI DATA

Visualisasi data adalah mempermudah komunikasi, mempercepat pengambilan keputusan, memperkaya perspektif tim, dan mengidentifikasikan masalah.

1. Tableau <https://www.tableau.com/products/public>
2. Metabase <https://www.metabase.com/>
3. Looker Studio <https://lookerstudio.google.com/>

Bahasa pemrograman Data Science

“The limits of my language mean the limits of my world”

* + Ludwig Wittgenstein

Peribahasa di atas cukup relate denga napa yg sekarang di hadapu dalam konteks ilmu data. Bahasa pemrograman adalah cara berkomunikasi untuk menyampaikan maksud

1. Pyhton

Pada data scientist dalam mengerjakan proyek data science

1. Memperbaiki dan menghapus data yang salah atau biasa disebut dengan pembersihan data.
2. Mengekstraksi dan memilih berbagai fitur data.
3. Pelabelan data (*data labeling*), yaitu menambahkan nama yang bermakna untuk data.
4. Melakukan proses analisis data.
5. Memvisualisasikan data dengan menggunakan diagram dan grafik seperti diagram garis, grafik batang, histogram, dan diagram lingkaran.
6. Mengembangkan model machine learning.
7. Bahasa R

Pengenalan machine learning

Machine Learning adalah ilmu pengembangan algoritma yg memanfaatkan konsep matematis dan statistika dalam menjalankan tugas tertentu tanpa instruksi eksplisit.

3 kategori dalam pengelompokan machine learning

1. Supervised learning
2. Unsupervised learning
3. Reinforcemenet Learning.

Teknologi pendukung Github

Github adalah software yg memungkinkan control proyek melalui alat yg di kenal sebagai git. Ini memungkinkan pengguna untuk mendistribusikan kode, berkolaborasi dalam proyek, membantu penerapan proyek melalui CI/CD (Integrasi berkelanjutan/penerapan berkelanjutan), kembali ke versi proyek sebelumnya, dll.

Adapun repositories github yg digunakan data science

* [TensorFlow](https://github.com/tensorflow/tensorflow)
* [The Algorithms](https://github.com/TheAlgorithms)
* [Awesome Machine Learning](https://github.com/josephmisiti/awesome-machine-learning)
* [Data Science I-Python Notebooks](https://github.com/donnemartin/data-science-ipython-notebooks)
* [Homemade Machine Learning](https://github.com/trekhleb/homemade-machine-learning)
* [Awesome Data Science](https://github.com/academic/awesome-datascience)
* [Deep Learning Drizzle](https://github.com/kmario23/deep-learning-drizzle)