

Learning Progress Review - W1

27 Sep - 2 Okt '25

Group 5 | Seru!



Group 5: Seru!'s Crews

Andi Yuswandi

Bayu Widodo

Samuel Nathanael Sitompul

Dhinar Ardhy Sutrisno

Mesayu Fitri Novianti Azra

Maruli Tyson Hamonangan Sianipar

Muhimah



Get to Know Digital Skola

Class Introduction



Edutech platform di Indonesia yang sudah berdiri dari tahun 2020, memberikan **one-stop solution for learning** untuk mempersiapkan techies dalam memenuhi tuntutan industri saat ini dan masa depan✨

Tutor Data Science kita 🔥



Faiq Miftakhul Falakh
Data Scientist at GovTech
Procurement

More than 5 years
experiences in Data Science



Farah Yulianti
IT Program Development
More than 3 years experience in
Data Science



Hendy Fergus Atheri H
Senior Data Scientist

More than 5 years experience
in Data Science



Havez Fazirani Al Kautsar
AI Product Manager at BUMN
Perbankan

More than 6 years experience in Data
Science



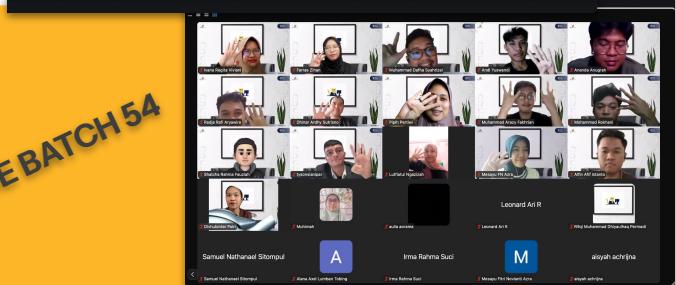
Kak Fikri sebagai **class representative** yang mengakomodir para techies👏

DATA SCIENCE BATCH 54

Dan Pak Dhinar,
sebagai **Ketua Kelas**
dan **Ketua Grup Seru**😎



Dhinar Ardhy Sutrisno



Fundamentals of Data Science & Methodology

Data Science Methodology

Pengolahan Data

- **Pondasi awal** dalam data science.
- Menjadi penentu kualitas model
- Tahap ini meliputi:
 1. Menentukan tujuan pengolahan data
 2. Pengumpulan data
 3. Penggabungan data
 4. Melakukan analisis data

Menemukan Insight

- Tahap **analisis data** untuk menemukan pola, tren, dan hubungan antar-variabel.
- Tahap ini meliputi:
 1. Melakukan analisis deskriptif
 2. Melakukan analisis hubungan diagnostik
 3. Melakukan teknik statistika deskriptif dan uji statistik

Membangun Algoritma Prediksi

- Tahap implementasi data yang sudah diolah, dianalisis, dan diuji.
- Proses ini dikaitkan dengan konsep *Classical Statistics, Machine Learning, dan Human-centered modelling*.
- Selanjutnya, model/ data yang sudah diolah diterapkan dalam dunia nyata, yaitu dengan **diintegrasikan ke sistem, aplikasi, atau dashboard** sehingga bisa digunakan oleh pengguna lain atau sistem bisnis.

Tentang Data Science

Apa itu Data Science ?

Data Science adalah keterampilan mengolah data dengan jumlah yang sangat banyak dan terus bertambah untuk menghasilkan informasi berupa insight dan algoritma prediksi yang akan digunakan untuk membantu bisnis agar berjalan lebih baik.

Profesional yang memiliki kemampuan dalam bidang Data Science disebut **Data Scientist**

3 Bidang Ilmu yang menjadi pilar Data Science:

- Ilmu Matematika atau Statistika
- Computer Science
- Knowledge Domain/keahlian spesifik di bidang tertentu

Hasil Kerja Data Scientist

Berdasarkan tujuan pengolahan data, hasil kerja Data Scientist terdiri dari beberapa kategori inti, yaitu :

1. Decision Support
2. Scoring
3. Recommendation
4. Classification
5. Forecasting
6. Detection



Keterampilan Data Scientist

- Keterampilan Hard Skill :
 - ◆ SQL dan Database
 - ◆ Bahasa Pemrograman seperti Phyton dan R
 - ◆ Visualisasi data dan Business Intelligence
 - ◆ Statistik terapan
 - ◆ Machine Learning & Deep Learning

- Keterampilan Soft Skill :
 - ◆ Data Storytelling
 - ◆ Komunikasi
 - ◆ Berfikir Kritis
 - ◆ Bekerja secara tim

Tentang Data Science

Success Story Data Science

1. Air BnB, perusahaan yang menyediakan layanan booking tempat menginap.

Sejak mengaplikasikan Data Science tahun 2008, valuationnya s.d. 2015 meningkat cepat. Data Science yang diimplementasikan adalah A/B Testing, Image Recognition & Analysis, NLP, Predictive Modelling, Regression Analysis, Recommendation System dengan menggunakan metode Collaborative Filtering

2. Zara, mengimplementasi Report/ Dashboarding, Predictive modeling

Sejak menerapkan tersebut, Zara berhasil meningkatkan efisiensi pergudangannya dan proses pembuatan desain-desain baru, meningkatkan revenue bisnis mereka sebesar 200% sejak 2004-2005

Peluang Karir Data Scientist

Peluang Karir Data Scientist di indonesia, contoh diambil dari salah satu startup

- Data Scientist
- Senior Data Scientist
- Data Scientists Lead atau Principal Data Scientist
- Senior Data Scientists Lead atau Senior Principal Data Scientist
- Assistant VP of Data Science
- VP of Data Science
- Director of Data / Director of Business

Output Data Science

Output Data Science terbagi menjadi enam bentuk utama yang membantu organisasi dalam mengambil keputusan, memberikan rekomendasi, melakukan prediksi, hingga mendeteksi pola yang tidak wajar.

Decision Support

- Mengolah data menjadi *insight* untuk mendukung keputusan.
- Membantu menghasilkan keputusan yang lebih akurat dan spesifik.
- Output: Report dan Dashboard.

Scoring

- Menggunakan data scoring historis sebagai dasar model untuk dikembangkan.
- Memperkirakan nilai scoring untuk data berikutnya.
- Output: Nilai skor prediksi.

Recommendation

- Mengeksplorasi hubungan antar-data.
- Memberikan saran untuk pemecahan masalah dan strategi.
- Output: Rekomendasi tindakan atau opsi terbaik.

Output Data Science

Output Data Science terbagi menjadi enam bentuk utama yang membantu organisasi dalam mengambil keputusan, memberikan rekomendasi, melakukan prediksi, hingga mendeteksi pola yang tidak wajar.

Classification

- Berdasarkan data yang sudah memiliki kategori.
- Memprediksi kategori pada data baru.
- Output: Label atau kelas hasil prediksi.

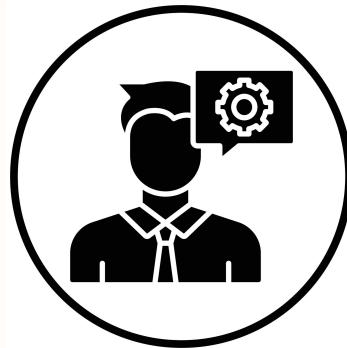
Forecasting

- Menggunakan pola dan tren historis (time series).
- Memprediksi tren dan pola untuk periode mendatang.
- Output: Nilai atau tren masa depan.

Detection

- Analisis aktivitas yang menyimpang dari pola normal.
- Digunakan untuk pencegahan fraud.
- Output: Peringatan (alert) atau deteksi anomali.

Data Science Skillsets



Hard Skill

Hard skill merupakan kemampuan teknis yang bisa dipelajari, diajarkan, diukur, serta diperlukan untuk melaksanakan suatu pekerjaan tertentu.



Soft Skill

Soft skill merupakan keterampilan non-teknis serta karakter personal yang berperan dalam menentukan cara seseorang berinteraksi dan berkomunikasi dengan orang lain secara efektif.

Hard Skill



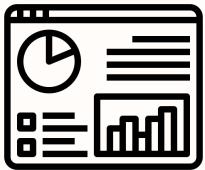
SQL dan Database

Skill ini berguna untuk mengambil data dan penyaringan data sehingga data dapat diolah.



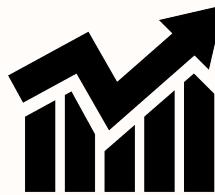
Python dan R

Dua bahasa pemrograman ini akan membantu data scientist dalam mengolah data menjadi informasi baru.



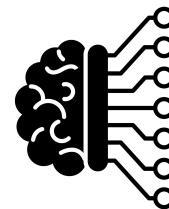
Data Visualization

Kemampuan untuk membuat informasi yang didapatkan menjadi lebih mudah dibaca dan dipahami.



Statistika Terapan

Statistika terapan berguna untuk menerapkan hal-hal apa saja yang dibutuhkan untuk mencapai hasil yang diinginkan



Machine Learning dan Deep Learning

Dengan adanya kemampuan ini data scientist mampu membuat algoritma, prediksi dan klasifikasi data.

Soft Skill



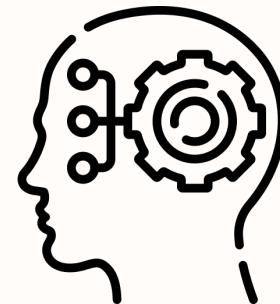
Data Storytelling

Kemampuan untuk menjelaskan data-data yang didapat menjadi informasi yang mudah dipahami oleh audience.



Komunikasi

Data scientist harus memiliki kemampuan komunikasi yang baik sehingga penyampaian hasil analisis data yang didapat tersampaikan dengan jelas.



Berfikir Kritis

Berpikir kritis sangat penting bagi data scientist untuk melakukan proses kerja seperti pengumpulan, analisis data dan mempresentasikan hasil kerja.

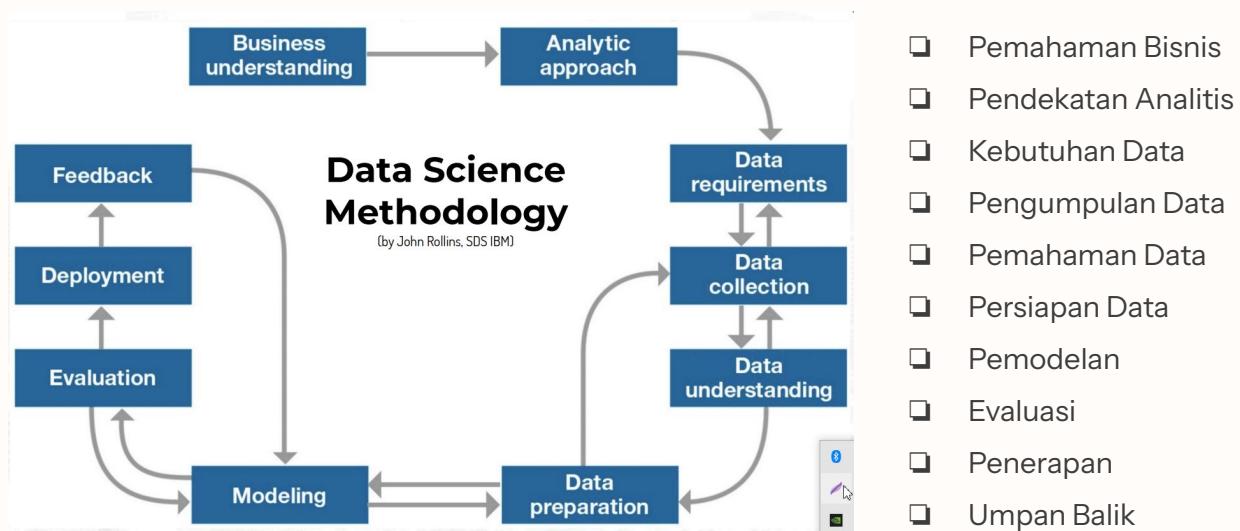


Kerjasama Tim

Data scientist harus memiliki kerjasama tim yang baik agar proses kerja dapat berjalan dengan lancar.

Framework Data Science: CRISP-DM (1/3)

- Data scientist melakukan beberapa **tahapan aktivitas** pengolahan data untuk mendapatkan **insight** dan melatih **model prediksi**.
- Salah satu metode siklus dan kerangka kerja yang dapat digunakan adalah **CRISP-DM** (Cross-Industry Standard Process for Data Mining)



Framework Data Science: CRISP-DM (2/3)



Pemahaman Bisnis

- Memahami ruang lingkup permasalahan bisnis yang ingin diselesaikan dengan pengembangan model algoritma prediksi



Pendekatan Analitis

- Mengoptimalkan kerja pengolahan data sesuai dengan tujuan bisnis yang ditentukan



Kebutuhan Data

- Menentukan dan mendefinisikan data apa saja yang dibutuhkan untuk diolah lebih lanjut



Pengumpulan Data

- Mendapatkan data sesuai dengan kebutuhan bisnis dan tujuan dari pengembangan model algoritma prediksi



Pemahaman Data

- Memeriksa dan mengidentifikasi masalah data yang telah dikumpulkan

Framework Data Science: CRISP-DM (3/3)



Persiapan Data

- Memperbaiki masalah data seperti noise, inkonsistensi, missing values



Pemodelan

- Membangun algoritma prediksi untuk mendapat insight berdasarkan data tersedia



Evaluasi

- Melakukan pengujian model untuk memastikan akurasi model sesuai kebutuhan



Penerapan

- Algoritma prediksi diimplementasi untuk dapat meningkatkan bisnis



Umpatan Balik

- Sebagai masukan apakah model dikembangkan, diubah atau tidak digunakan lagi

Success Story



AirBnB
Homestay Business

A sharp increase in company valuation, driven by a rapidly growing user base: 25 million users across 192 countries (2015 data).



Zara
Clothing Company

Annual revenue grew 220% between 2004 and 2015. Production rose sharply, and warehousing became more efficient.



Amazon
Online Store

Customer response time reduced by 40%. 29% increase of order value. 68% more click through rate. Significant conversion rate.

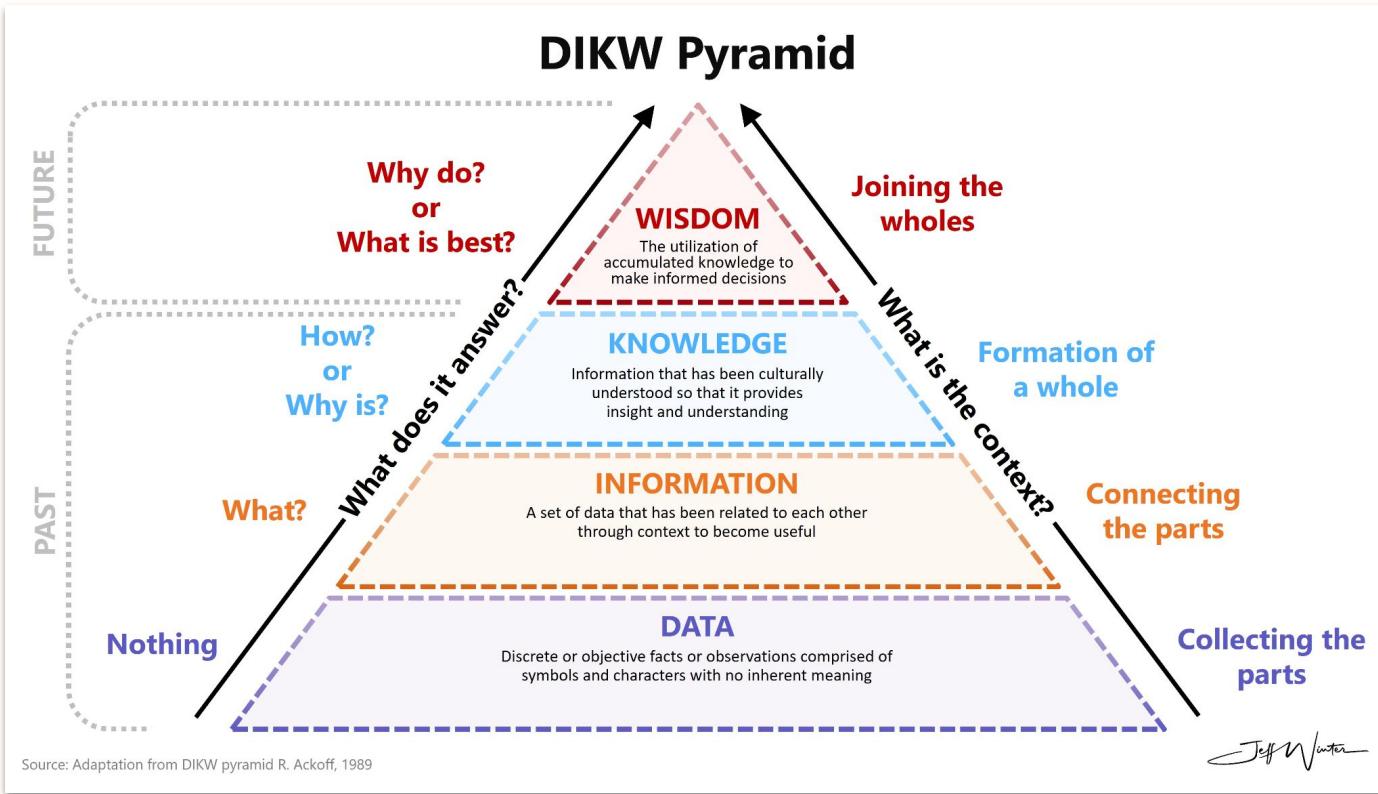


Uber
Transportation Tech

Matching algorithms led to around 20% reduction in travel time. More efficient fuel. Reduced passengers waiting time. Massive increase of customers

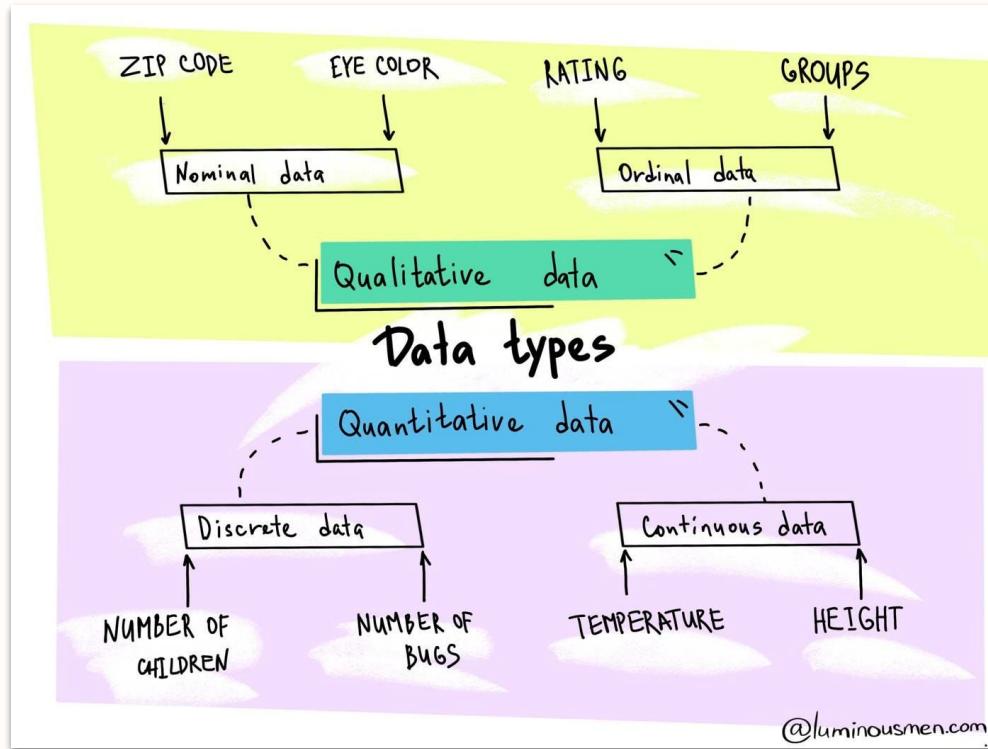
Essentials of Data and Databases

Piramida Data



Source: <https://www.jeffwinterinsights.com/insights/dikw-pyramid>

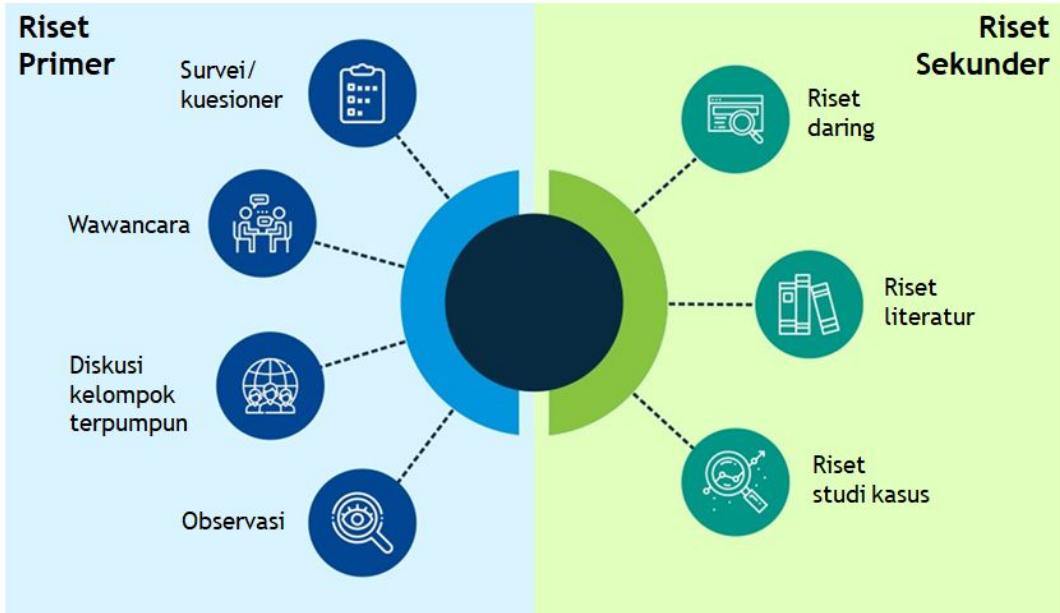
Kategori Data Berdasarkan Pengukurannya



Berdasarkan pengukurannya data dapat dibagi dalam 2 kelompok, yaitu:

- Data Kualitatif:**
Data yang dibuat/disajikan dalam bentuk deskriptif (bukan angka).
- Data Kuantitatif:**
Data yang dapat disajikan dalam bentuk angka dan dihitung.

Kategori Data Berdasarkan Sumbernya



Source: <https://dqlab.id/peneliti-wajib-tahu-ini-perbedaan-data-sekunder-dan-data-primer>

Berdasarkan sumbernya data dapat dibagi dalam 2 kelompok:

1. Data Primer

Data yang sumbernya diperoleh secara langsung seperti hasil eksperimen, survei dan wawancara.

2. Data Sekunder

Data yang sumbernya diperoleh dari sumber lain yang tersedia seperti literatur dan hasil penelitian orang lain.

Basis Data

Basis data adalah kumpulan data terhubung yang tersimpan secara elektronik, memudahkan pengaturan, penyimpanan, dan pengorganisasian data untuk mendukung pengolahan informasi dan efisiensi kerja.

Tujuan & Fungsi Basis Data

Kecepatan

Mempermudah penyimpanan, update, dan akses data secara cepat.

Efisiensi Ruang

Menghemat ruang penyimpanan dengan mengurangi redundansi data.

Keamanan

Menyediakan batasan akses data melalui password dan kontrol lainnya.

Kemudahan Membagi

Data dapat digunakan bersama oleh beberapa aplikasi.

Standardisasi

Data terpusat mempermudah kontrol dan pertukaran data.

Keakuratan

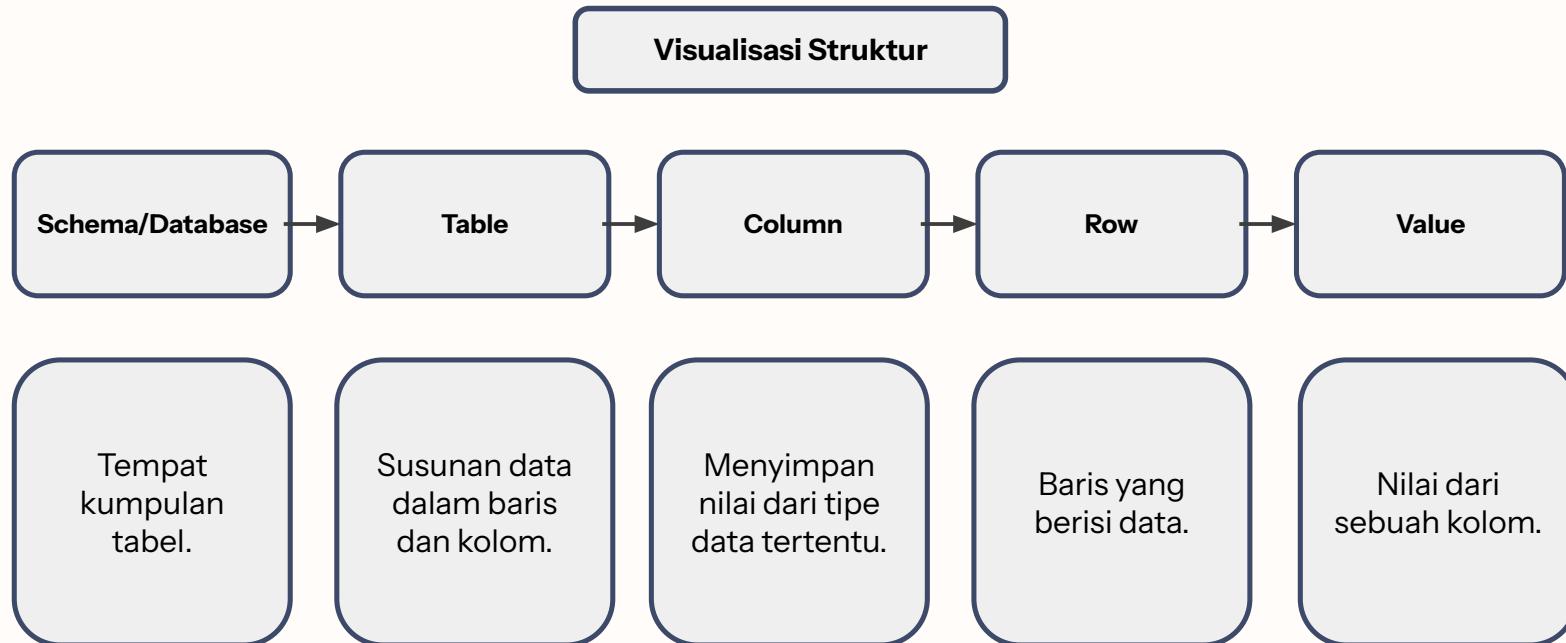
Menjamin data unik, tidak ganda, dan sesuai relasi.

Konsistensi Data

Perubahan pada data berlaku otomatis di semua aplikasi terkait.

Struktur Basis Data

Struktur basis data terdiri dari schema, tabel, kolom, baris, dan nilai yang saling terhubung untuk menyusun, menyimpan, serta mengatur data agar mudah diakses dan dikelola.



Relational Database Management System (RDBMS)

Sistem basis data yang menyimpan dan menyediakan data dalam bentuk relasional, disajikan dalam bentuk tabel.

Berdasarkan cara menyimpan basis data :

1. RDBMS

Secara terstruktur, berbentuk tabel.

2. Non -RDBMS

Secara tidak terstruktur, berbentuk document, pair-keys, graph, dll.

RDBMS memungkinkan user melakukan tugas berikut:

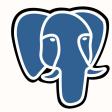
1. Data Definition → Membuat, memodifikasi, dan menghapus organisasi data.

2. Data Updation → Menyisipkan, memodifikasi, dan menghapus data aktual.

3. Data Retrieval → Mengambil data untuk dipergunakan berbagai keperluan.

4. User Administration → Mengelola user dalam hal pendaftaran dan pemantauan.

Contoh database RDBMS



PostgreSQL



Relational Database Management System (RDBMS)

Sistem basis data yang menyimpan dan menyediakan data dalam bentuk relasional, disajikan dalam bentuk tabel.

Jenis-jenis RDBMS :

Operational Database

Memperbarui data secara real-time/OLTP
(On-Line Transaction Processing) → JSON, XML

Relational Database

Menggunakan bahasa query SQL → MySQL,
PostgreSQL, Oracle DB

Data Warehouse

Repositori sentral data, memuat kumpulan
informasi yang diekstrak dari basis data waktu ke
waktu

Distributed Database

Basis data saling terhubung, tersebar pada lokasi
fisik berbeda pada unit/jejaring komputer di seluruh
dunia

End-user Database

Basis data yang ditanam untuk penyimpanan
lokal/klien dalam software → SQLite

Fungsi RDBMS :

1. Data dictionary management
2. Data storage management
3. Data transformation and presentation
4. Security management
5. Multi User access control
6. Backup & recovery management
7. Data integrity management
8. Database communication
9. Interfaces
10. Transaction management



Structured Query Language

Structured Query Language atau sering disingkat SQL adalah bahasa pemrograman yang digunakan untuk mengelola data yang tersimpan di RDBMS. Query adalah perintah atau permintaan untuk mengambil, memanipulasi atau mengelola data dari sebuah basis data.

Komponen SQL

DDL (Definition Data Language)

DDL merupakan perintah yang bertujuan untuk membuat, mendesain dan menghapus struktur database.

Beberapa perintah DDL:

- Create, membuat tabel dalam database.
- Alter, mengubah struktur database.
- Drop, menghapus tabel dari database

DCL (Data Control Language)

DCL merupakan perintah untuk memberikan izin atau akses ke dalam database.

Beberapa perintah DCL:

- Grant, memberikan akses kepada pengguna lain ke database.
- Revoke, mencabut hak akses pengguna dari database.

DML (Data Manipulation Language)

DML merupakan perintah untuk melakukan proses manipulasi data yang ada di database.

Beberapa perintah DML:

- Select, menarik data dari database.
- Insert, menyisipkan data baru ke dalam database.
- Update, memperbarui data yang ada dalam tabel.
- Delete, menghapus data dari database.

Tipe Data SQL

Tipe Numerik

Tipe Numerik adalah tipe data untuk menyimpan nilai-nilai angka. Tipe numerik juga dapat melakukan operasi aritmatika yaitu penambahan, pengurangan, perkalian dan pembagian.

Contoh:

- TINYINT, (-128 s/d 128)
- SMALLINT, (-32.768 s/d 32.767)
- MEDIUMINT, ($\pm 8,39 \times 10^6$)
- INT, ($\pm 2.15 \times 10^9$)
- BIGINT, ($\pm 9,22 \times 10^{18}$)
- FLOAT,
 $(-1.79769313486231570E+308$
s/d
 $1.79769313486231570E+308$
- DOUBLE

Tipe Date and Time

Tipe Date and Time adalah tipe data yang menyimpan tanggal dan waktu.

Contoh:

- DATE, (YYYY-MM-DD)
- TIME, (HH:MI:SS)
- DATETIME, (YYYY-MM-DD HH:MI:SS)
- YEAR, (YYYY)

Tipe String (Text)

Tipe String adalah tipe data yang menyimpan text.

Contoh:

- CHAR (0 - 255 karakter)
- VARCHAR (0 - 65.535 karakter)
- TEXT, (untuk menyimpan text lebih panjang)

DLM (Data Lifecycle Management)

- ❖ **Data Lifecycle Management** merupakan suatu model pengorganisasian siklus dan aliran data yang berada di dalam suatu sistem.

Tahap-tahap Data Lifecycle Management

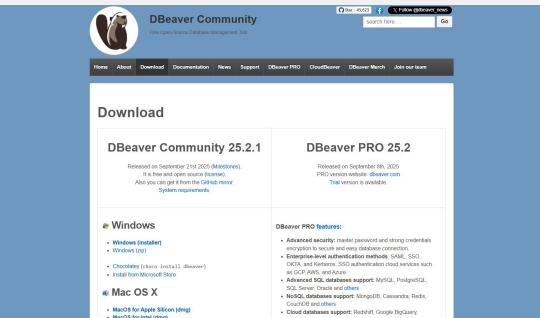


DLM (Data Lifecycle Management)

*Manfaat Penerapan
Data Lyfecycle
Management*

- 01 DLM memungkinkan perusahaan untuk memenuhi kebutuhan data yang diajukan oleh sektor-sektor bisnis yang membutuhkan
- 02 DLM dapat menjamin suatu infrastruktur proteksi data yang baik
- 03 DLM menjamin adanya data-data berkualitas tinggi dengan nilai-nilai yang selalu diperbarui (update value). Melalui ekstraksi dan pemeliharaan siklus-siklus data yang dilakukan.
- 04 Dengan menerapkan DLM, data-data berkualitas tinggi, aman, dan optimal dapat selalu tersedia bagi seluruh pengguna. Artinya, hal itu dapat meningkatkan kinerja bisnis perusahaan.

DBeaver: Configure, Develop, & Analyse Databases



What is DBeaver?

- A universal and open-source **SQL Client** and database administration tool.
- **Graphical User Interface (GUI)** makes it easier to manage various type of database from one place.

Key Advantages

- Supports **multiple databases** (MySQL, PostgreSQL, Oracle, SQL Server)
- Multi-platform, multi-OS, multi-language.
- User friendly.
- Many extension (plugin) choices.

Installation

- Visit dbeaver.io/download.
- Download installer from the official website based on your OS.
- Start the installation → activate (check) "**Associate .SQL files**".



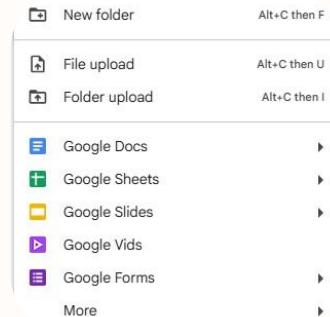
Menginstal Google colaboratory



1. Buka drive.google.com



2. Klik New di pojok kiri atas



3. Muncul kotak dialog, pilih more

- Google Drawings
- Google My Maps
- Google Sites
- Google Apps Script
- Google Colaboratory
- + Connect more apps

4. Muncul kotak dialog lagi, pilih Google Colaboratory



5. Anda akan diarahkan ke tab baru. Muncul lembar kerja Google Colab dan siap digunakan



Mengapa Google Colaboratory?

Google Colaboratory merupakan platform berbasis cloud yang memungkinkan kita untuk menulis, menjalankan, dan berbagi kode Python melalui web browser.

Dengan demikian, kelebihan Google Colaboratory dibandingkan dengan aplikasi Coding Python yang lain adalah:

1

Tidak perlu instalasi.

Google Colab adalah layanan berbasis *cloud* dari Google.

2

Kolaborasi

Google Colab memungkinkan untuk melakukan kolaborasi dengan klien atau tim dengan lebih mudah.

3

Pengolahan Data Cepat

Karena terhubung dengan aplikasi Google Drive, kita bisa menyimpan serta mengolah data dengan lebih praktis

Thank you!