

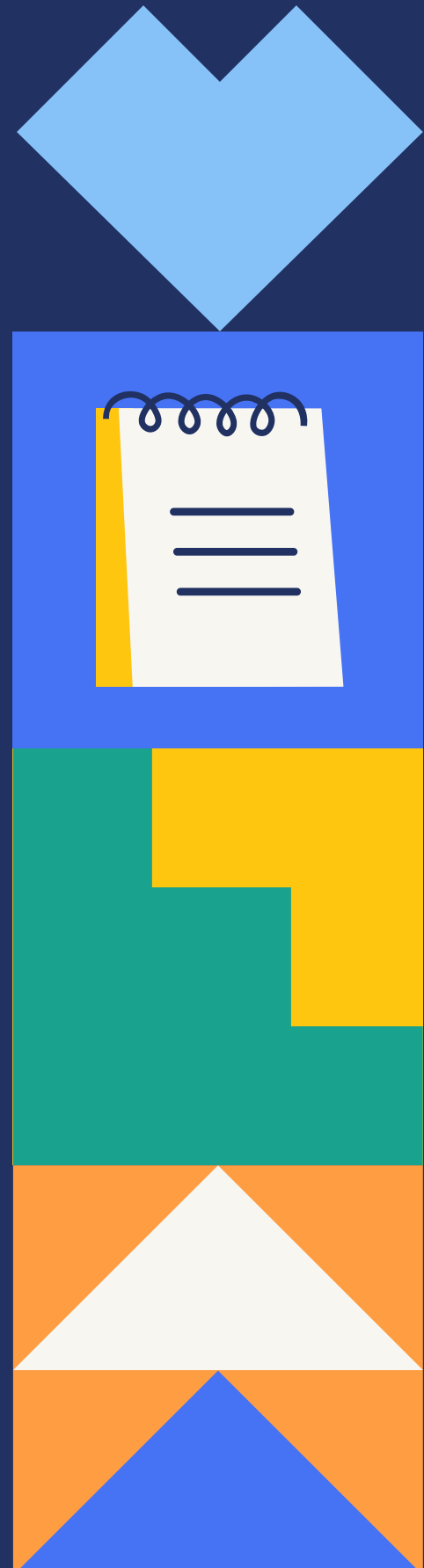
Learning Progress Review

WEEK - 01

KELOMPOK DISCOVRIO

- Scudetto Ciano Syam
- Nugroho Wahyu Saputra
- Radityo Fajar Pamungkas
- Puteri Sakinah Mantikasari A. G.
- Yayas Kasnanda Bintang
- Rima Fitrianti Azahra
- Farhad Salim Sungkar





Outline

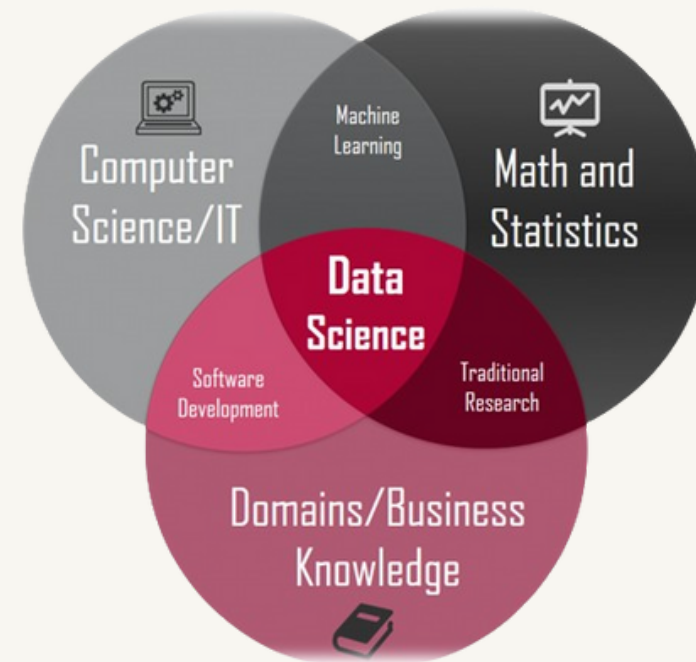
- 01 Introduction to Data Science
- 02 Data Science Methodology
- 03 Introduction to Data & Database

Data Science adalah keterampilan dalam mengolah data untuk menghasilkan informasi berupa:

Insight dari temuan pola dan tren dari data

Algoritma prediksi didasarkan pada pola hubungan keterkaitan informasi data

Data science adalah ilmu yang menggabungkan berbagai bidang seperti statistik, matematika, ilmu komputer, dan domain knowledge untuk mengelola, menganalisis, dan menginterpretasikan data dalam jumlah besar.



Gambar 1.
Ruang Lingkup Data Science
(Source: [DigitalSkola](#))

Orang yang bekerja dibidang data science biasa disebut **Data Scientist**. Seorang data scientist mengolah berbagai bentuk data yang dapat berupa data tabular, gambar, text, video, audio, dan lain-lain.

01

Introduction to Data Science



Hasil Kerja Data Scientist

1

Decision Support

Data yang dianalisis oleh Data Scientist bertujuan untuk menghasilkan insight sebagai pendukung bagi pengambilan keputusan.

2

Scoring

Bertujuan untuk melakukan prediksi nilai penilaian pada data berikutnya berdasarkan model atau algoritma yang telah dikembangkan.

3

Recommendation

Algoritma ini dibangun melalui analisis data untuk menemukan hubungan-hubungan antar informasi dengan tujuan seperti pemecahan masalah atau perencanaan langkah-langkah strategis di masa mendatang.

4

Classification

Algoritma ini didasarkan pada data yang sudah diklasifikasikan sebelumnya, yang kemudian dapat diubah menjadi suatu rumus untuk memprediksi kategori berikutnya.

5

Forecasting

Algoritma ini dikembangkan berdasarkan pola dan tren tiap periode sebelumnya untuk memprediksi pola dan tren pada periode berikutnya.

6

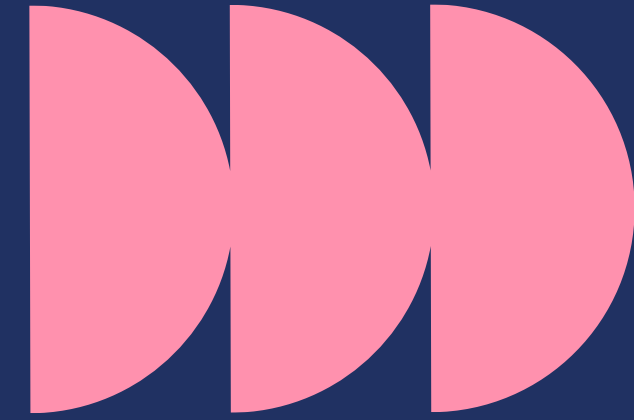
Detection

Algoritma ini dikembangkan untuk mendeteksi aktivitas yang tidak ada pada aktivitas sebelumnya sebagai tindakan antisipasi terhadap potensi fraud.

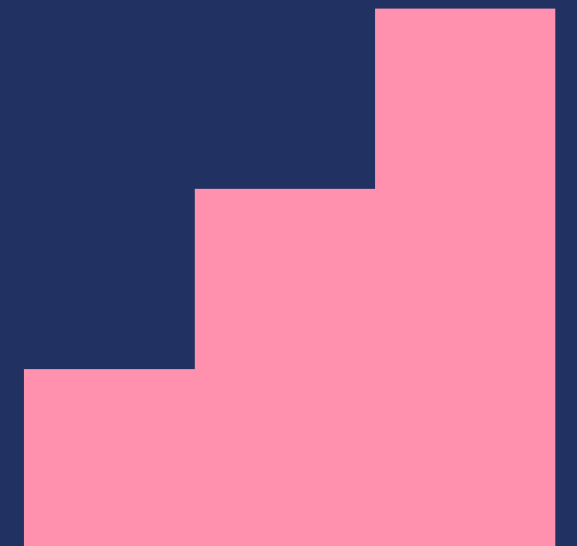


Hard Skills

- **SQL & Database**
Membantu proses pengumpulan, penyaringan dalam tahapan awal proses pengolahan data.
- **Python & R**
Membantu proses saat menerapkan statistika, machine learning, dan visualisasi informasi.
- **Visualisasi Data & Business Intelligence**
Membantu menemukan metode analisis informasi yang kemudian informasi tersebut dipaparkan dalam bentuk visualisasi data.
- **Statistika Terapan**
Membantu menentukan strategi yang sesuai dalam mengolah data tersedia dan sesuai dengan tujuan yang dicapai.
- **Machine Learning dan Deep Learning**
Membantu kerja data dengan pemrograman komputer.



Keterampilan Data Scientist



Soft Skills



Data Storytelling



Komunikasi



Berpikir Kritis

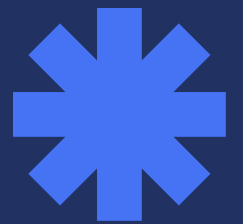


Bekerja Secara Tim



Keterampilan Data Scientist





Contoh Keberhasilan Data Scientist

Contoh keberhasilan beberapa perusahaan yang memanfaatkan keterampilan data science untuk proses bisnisnya.

01 Amazon

Revenue tahunan meningkat \$13.2 miliar. Pada tahun 2021, Amazon Web Services (AWS) menghasilkan revenue sebesar \$62.3 miliar.

Implementasi: Recommendation, Pricing

02 Walmart

Menghemat \$1 miliar per tahun melalui optimasi rantai pasokan.

Implementasi: Supply Chain Optimization, Inventory Management

03 JPMorgan Chase

Mengurangi risiko penipuan sebesar 30%.

Implementasi: Fraud Detection, Risk Management

Peluang Karir Data Scientist



Berdasarkan riset dari World Economic Forum 2020, keterampilan pada bidang data berada di urutan teratas sebagai profesi di masa mendatang karena adanya transformasi bisnis ke arah digital yang memberikan banyak keuntungan bagi perusahaan.

<https://intelligence.weforum.org/topics/a1G6800000004C9hEAE>



Berdasarkan prediksi dari McKinsey Global Institute menyatakan bahwa Indonesia membutuhkan 1,5 juta data scientist pada tahun 2030

<https://www.mckinsey.com/capabilities/quantumblack/careers-and-community/data-scientist>



Berdasarkan penelusuran pada Mei 2024, terdapat lebih dari 800 lowongan pekerjaan data scientist di Indonesia di LinkedIn

<https://www.linkedin.com/>



Pekerjaan Data Scientist

Dalam pekerjaannya, seorang data scientist akan melakukan beberapa **proses yang saling berkaitan**, yaitu:

Data Processing

- Menentukan tujuan analisis
- Mengumpulkan data sesuai kebutuhan
- Menggabungkan data
- Menganalisis Data

Insights Finding

Memanfaatkan:

- Informasi deskriptif (sorting, chart viewing)
- Informasi diagnostik (hubungan dan sebab akibat)
- Uji statistik

Generating Prediction Algorithm

Menggunakan:

- Classical Statistics
- Machine Learning
- Human Centered Modelling

02

Data Science Methodology



Methodology Framework

Cr oss

I ndustry

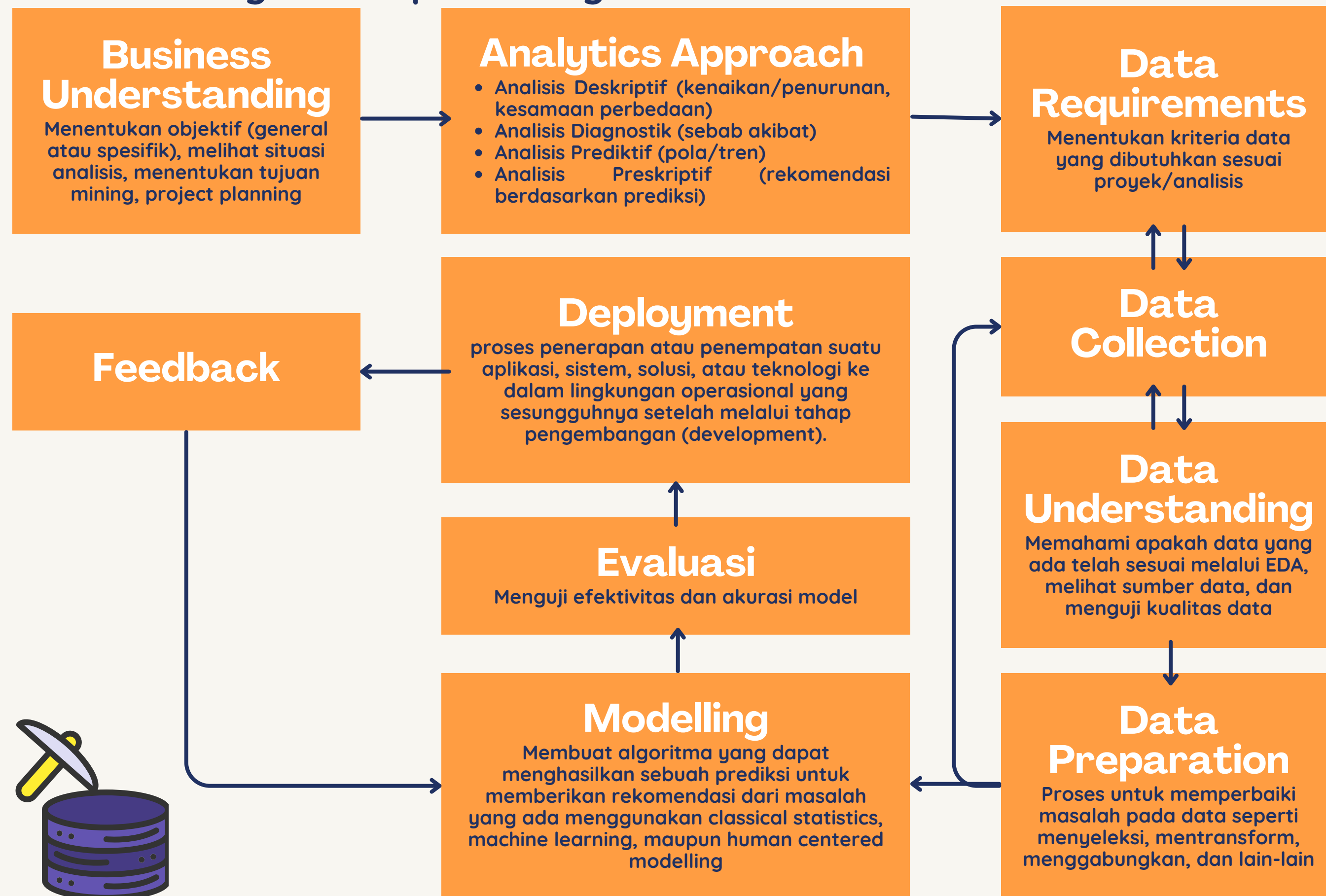
S tandard

P rocess

D ata

M ining

CRISP-DM adalah framework yang paling umum digunakan oleh data scientist dengan tahapan sebagai berikut:



Data Science Products



Rekomendasi Produk

BERRYBENKA

Ad • Berrybenka



65%



Berrybenka - Sepatu...

Rp106.000 ~~Rp299.000~~

★ 4.8 • 100+ terjual

64%



Berrybenka - Sepatu...

Rp133.000 ~~Rp369.000~~

★ 4.9 • 500+ terjual

71%



Berrybenka - S...

Rp86.000 ~~Rp249.000~~

★ 4.7 • 250+ terjual

72%



Rekomendasi Video



What Happened Before History? Human Origins

Kurzgesagt – In a Nutshell · 22M views · 8 years ago



Pengertian Data

Secara umum, pengertian data dapat didefinisikan sebagai **kumpulan fakta-fakta** yang diperoleh dan dapat diolah untuk mendapat pemahaman lebih lanjut

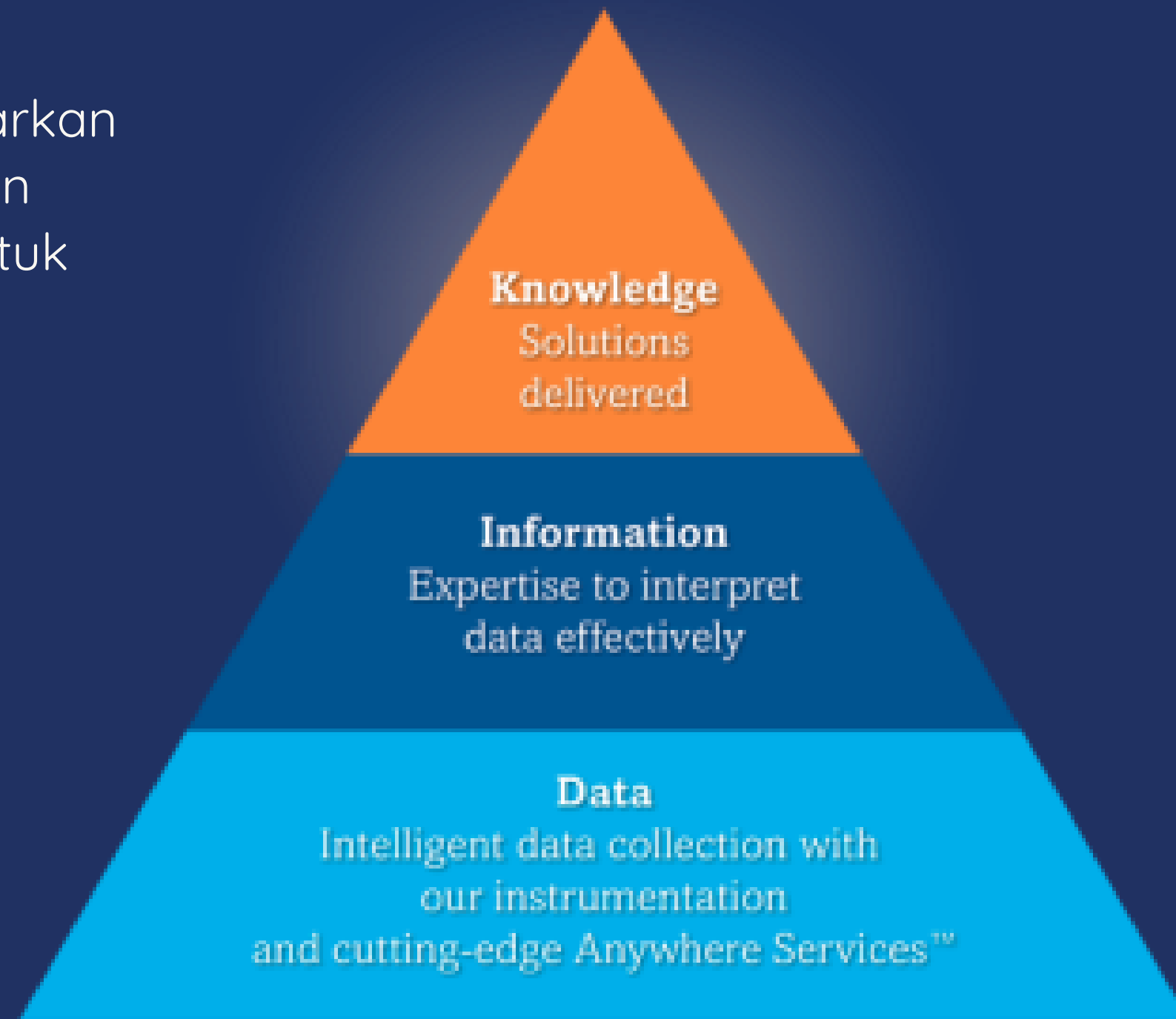
03

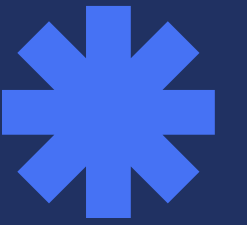
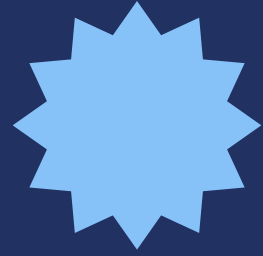
Introduction to Data & Database



Data, Information, and Knowledge

Data, information, dan knowledge merupakan term yang menggambarkan **tingkatan hierarkis** dari data, urutan tersebut dapat disajikan dalam bentuk **piramida data** berikut





KATEGORI DATA

Berdasarkan Jenis Data

1 DATA KUALITATIF

- Nominal : tidak ada peringkat atau tingkat kedudukan, hanya sebagai kategori
- Ordinal : memiliki kedudukan atau derajat

2 DATA KUANTITATIF

Data berupa angka, umumnya didapat dari hasil pengukuran/ pengamatan atas data-data diskrit atau kontinu

Berdasarkan Sumber

1 DATA PRIMER

Data yang didapatkan langsung dari wawancara/ eksperimen/ survei dan kuesioner

2 DATA SEKUNDER

Data yang diperoleh dari sumber-sumber tersedia atau sudah dikumpulkan sebelumnya

DATABASE

Database/ basis data adalah **kumpulan data yang terhubung dan terorganisir** yang disimpan secara elektronik. Database umumnya berada pada perangkat komputer dan dikelola secara otomatis oleh **Database Management System (DBMS)**.

Tujuan dan Fungsi

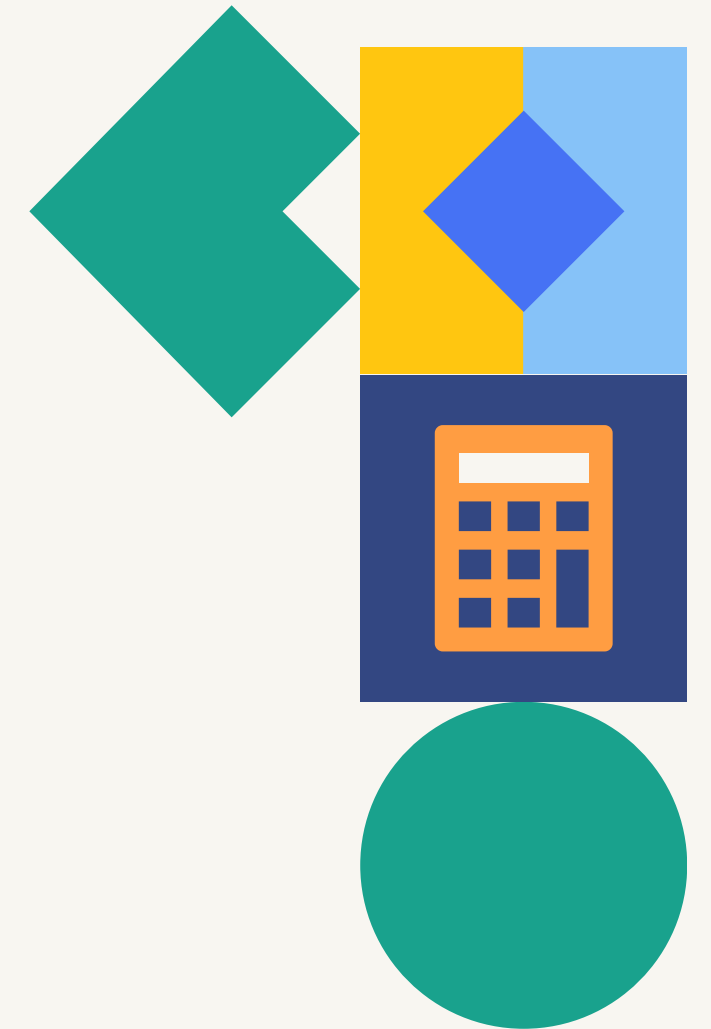
- **Kecepatatan dan kemudahan (Speed)**, update atau manipulasi cepat dan mudah
- **Efisien ruang penyimpanan (Space)**, hemat ruang karena redudansi/pengulangan data
- **Keamanan (Security)**, ada aspek keamanan misal hak akses atau jenis operasi yang boleh dilakukan
- **Kemudahan membagi (Shareability)**, dapat digunakan secara bersamaan
- **Standarisasi (Standardization)**, keterpusatan kontrol data, sehingga terdapat di satu ruang penyimpanan tunggal
- **Keakuratan (Accuracy)**, data yang dikumpulkan akurat dan unik
- **Konsistensi Data (Consistent)**, bila suatu data berubah di satu perangkat maka data akan berubah semua

Relational Database Management System (RDBMS)

Relational Database Management System (RDBMS) merupakan sistem manajemen basis data berdasarkan data relasional atau berbentuk tabel

Contoh RDBMS

- 1 SQLite
- 2 MySQL
- 3 PostgreSQL
- 4 OracleDB
- 5 SQL Server



Jenis-jenis RDBMS

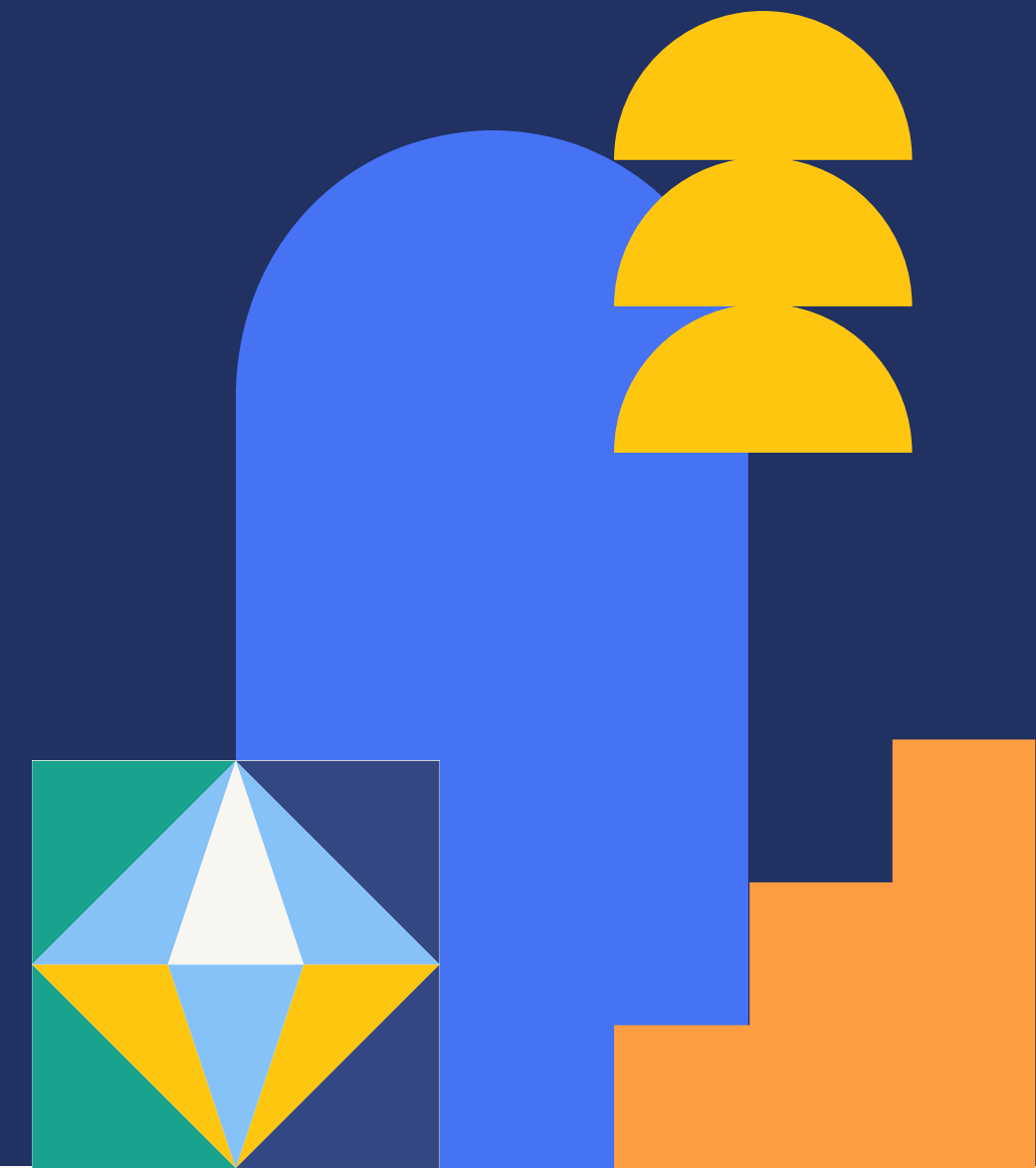
1.Operational Database merupakan tipe basis data yang berfungsi sebagai tempat untuk mengelola data dinamis secara langsung dan real time.

2.Data Warehouse merupakan tipe basis data yang merupakan elemen utama dalam Business Intelligence (BI).

3.Relational Database merupakan tipe basis data relasional yang merupakan basis data untuk mengorganisir sesuai model hubungan data.

4.Distributed Database merupakan tipe basis data yang terdiri dari gabungan situs dan tersebar di banyak lokasi berbeda.

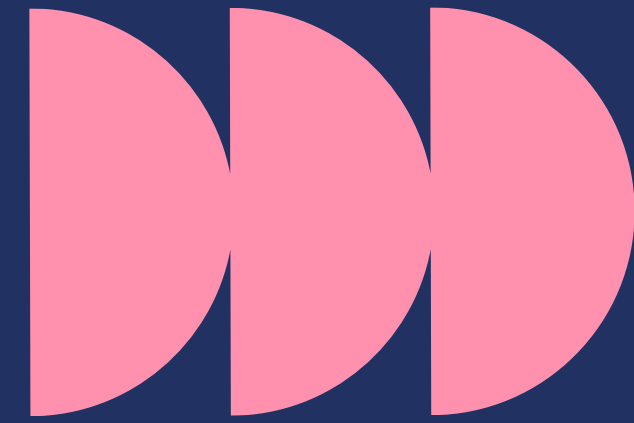
5.End-User Database merupakan jenis database yang dibuat dan diatur oleh pengguna akhir menggunakan perangkat atau workstation sendiri.



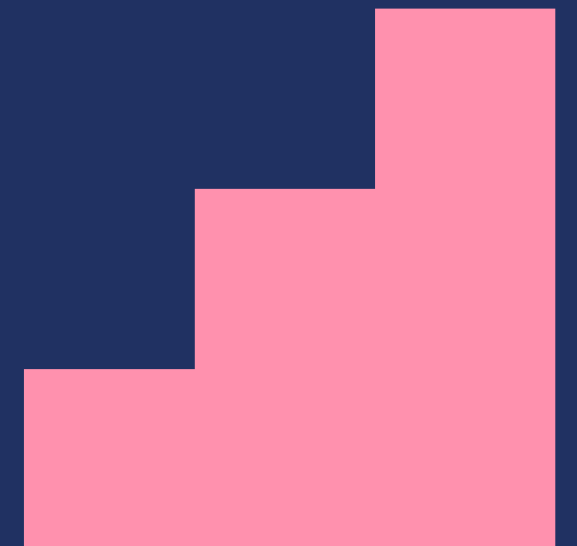
Data Definition Language (DDL)

DDL adalah suatu bentuk dari SQL yang dapat dipergunakan untuk menciptakan ataupun membuat suatu database, struktur tabel, tabel, dan lainnya.

- **CREATE** : untuk membuat table, database, atau view table
- **RENAME** : untuk mengubah nama table
- **ALTER** : untuk mengubah struktur suatu table
- **DROP** : untuk menghapus suatu table atau bahkan basis data



Jenis Perintah SQL



Data manipulation language(DML)

DML merupakan sekumpulan perintah yang dipergunakan dalam pengelolaan database seperti menampilkan, menghapus, menambah, dan membuat baru.

- **SELECT** : untuk memilih data yang ingin didapatkan
- **INSERT** : untuk menambahkan data baru
- **UPDATE** : untuk mengubah isi data yang sudah ada
- **DELETE** : untuk menghapus data



Jenis Perintah SQL



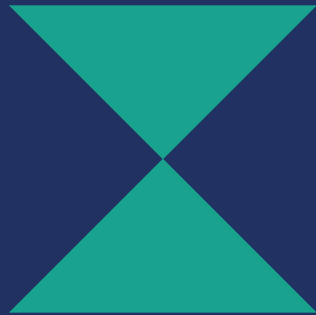
Tipe-tipe Data



Numerik

- **TINYINT** (-128 s/d 127),
- **SMALLINT** (-32.768 s/d 32.767),
- **MEDIUMINT** (-8.388.608 s/d 8.388.607),
- **INT** (-2.147.483.648 s/d 2.147.483.647),
- **BIGINT** (+ 9,22 x 10¹⁸)
- **FLOAT** (3.402823466E+38 s/d -1.175494351E-38, 0, dan 1.175494351E-38 s/d 3.402823466E+38)
- **DOUBLE**

Tipe-tipe Data



Date and Time

- **TINYINT** (-128 s/d 127),
- **DATE** (1000-01-01 s/d 9999-12-31 (YYYY-MM-DD))
- **TIME** (-838:59:59 s/d +838:59:59 (HH:MM:SS))
- **DATETIME** ('1000-01-01 00:00:00' s/d '9999-12-31 23:59:59')
- **YEAR** (1900 s/d 2155)

Tipe-tipe Data



String

- **CHAR** (untuk menyimpan text ukuran tetap 0-255 character)
- **VARCHAR** (untuk menyimpan text ukuran dinamis (0 s/d 65.535 karakter)
- **TEXT** (untuk menyimpan text panjang)

Data Lifecycle Management (DLM)

Definisi

Data Lifecycle Management (DLM) adalah proses yang memastikan data dikelola dengan baik dari pembuatan hingga penghancuran. DLM bertujuan agar data tetap dapat diakses, digunakan, dan sesuai dengan regulasi sepanjang siklus hidupnya.

Pentingnya DLM

- Meningkatkan keamanan dan kepatuhan data
- Memperbaiki aksesibilitas dan kegunaan data
- Mengoptimalkan biaya dan efisiensi penyimpanan



Data Lifecycle Management (DLM)

Lima Tahapan DLM

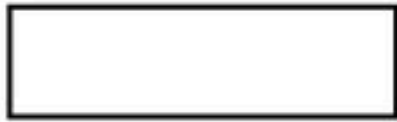
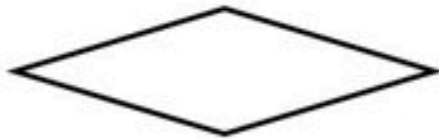


- 1 **CREATION** : Data dihasilkan dan ditangkap dengan akurat
- 2 **STORAGE** : Data disimpan dengan aman dan efisien
- 3 **USAGE** : Data diakses dan digunakan oleh aplikasi dan pengguna yang berwenang.
- 4 **ARCHIVING** : Penyimpanan jangka panjang untuk data yang jarang diakses.
- 5 **DESTRUCTION** : Pembuangan data usang dengan aman dan sesuai regulasi.



Entity-Relationship Diagram (ERD)

ERD (Entity Relationship Diagram) atau diagram hubungan entitas adalah **diagram yang digunakan untuk perancangan suatu database** dan **menunjukkan relasi** antar objek atau entitas beserta atribut-atributnya secara detail.

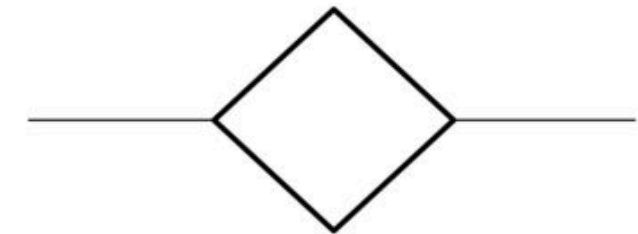
Komponen ERD

| Simbol | Keterangan |
|---|---|
|  | Entitas, yaitu kumpulan dari objek yang dapat diidentifikasi secara unik |
|  | Relasi, yaitu hubungan yang terjadi antara salah satu lebih entitas. Jenis hubungan antara lain. one to one, One to many, dan many to many. |
|  | Atribut, yaitu karakteristik dari entitas atau relasi yang merupakan penjelasan detail tentang entitas. |
|  | Hubungan antara entitas dengan atributnya dan himpunan entitas dengan himpunan relasinya. |

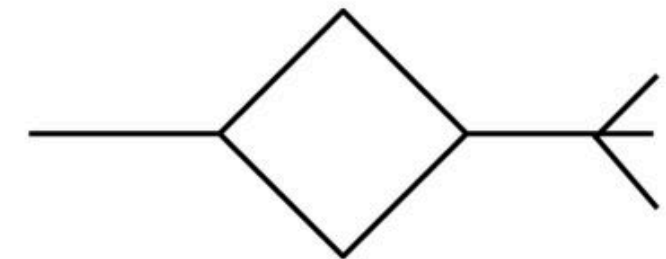
Entity-Relationship Diagram (ERD)

Simbol Kardinalitas

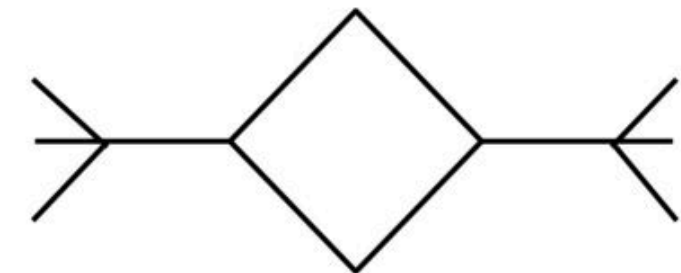
- 1 to 1 (one to one)



- 1 to n (one to many)



- N to n (many to many)





Thank you!

