

# Data Lake Attributes

# Data Ingestion

---

Data Ingestion : Proses dimana data berpindah dari satu sumber ke tujuan berupa data store yang kemudian akan di analisis.

Dengan manajemen data ingestion, data baik yang terstruktur maupun tidak akan terhubung dengan metadatanya.

---

**Metadata** adalah informasi terstruktur yang mendeskripsikan, menjelaskan, menemukan, atau setidaknya menjadikan suatu informasi mudah untuk ditemukan kembali, digunakan, atau dikelola. **Metadata** sering disebut sebagai data tentang data atau informasi tentang informasi. Data Ingestion : Proses dimana data berpindah dari satu sumber ke tujuan berupa data store yang kemudian akan di analisis.

Dengan manajemen data ingestion, data baik yang terstruktur maupun tidak akan terhubung dengan metadatanya.

**Metadata** adalah informasi terstruktur yang mendeskripsikan, menjelaskan, menemukan, atau setidaknya menjadikan suatu informasi mudah untuk ditemukan kembali, digunakan, atau dikelola. **Metadata** sering disebut sebagai data tentang data atau informasi tentang informasi.

## Data Democratization :

---

Setiap personal memiliki akses ke data dan tidak ada sistem yang menghalangi proses tersebut

Positif :

- Setiap orang bisa memberikan keputusan berdasar data yang diambil
- Semakin banyak orang yang bisa mengakses data meningkatkan kecepatan organisasi mengidentifikasi dan mengambil tindakan

Negatif :

- Kekhawatiran kesalahan interpretasi data oleh karyawan non teknis
- Masalah keamanan sistem

## Data Discovery

---

Proses mengumpulkan data dari berbagai basis data dan simpanan lokal, dan menggabungkannya menjadi satu sumber yang dapat dengan mudah dan langsung dievaluasi

Business Intelligence akan mendapatkan pola data berdasar wilayah, produk, penduduk dan lainnya.

Dengan Data Discovery pengguna data mencari pola dari data set

## Data Lineage

---

- Menyatakan dari mana data berasal, ke mana ia pergi, dan transformasi apa yang diterapkan padanya saat ia mengalir melalui berbagai proses.
- Siklus hidup data
- Berkaitan dengan aturan proses, aturan perhitungan yang benar, input output yang kompatibel dan lainnya

## Self Services Capabilities

---

- Pendekatan untuk analisis data yang memungkinkan pengguna bisnis mengakses dan bekerja dengan data perusahaan bahkan jika mereka tidak memiliki latar belakang dalam analisis statistic penambangan data.
- Pengguna dapat memfilter, mengurutkan, menganalisis, dan memvisualisasikan data tanpa melibatkan expert organisasi

# Metadata Management

---

- Proses, peran, dan aturan yang terkait yang melibatkan pemodelan data, tata kelola data, dan administrasi metadata
- Melibatkan pengelolaan metadata tentang data lain → data konten.
- Bentuk metadata yang lebih lama adalah katalog, kamus, dan taksonomi