



# Teknik Pengelolaan Metadata Big Data

Mata Kuliah Infrastruktur dan Teknologi Big Data

Oleh: Galih Hermawan, S.Kom., M.T.

Program Studi Teknik Informatika. FTIK.

Universitas Komputer Indonesia

# Pengertian Metadata

- Metadata adalah data tentang data, yaitu informasi yang mendeskripsikan sifat, karakteristik, kualitas, dan konteks dari data.
- Contoh metadata adalah nama fail, ukuran fail, tanggal pembuatan, jenis fail, penulis, judul, abstrak, kata kunci, sumber, referensi, dan lainnya.

## Metadata deskriptif

- ✓ mendeskripsikan isi, tujuan, dan sasaran dari data, seperti judul, abstrak, kata kunci, dan sumber.

## Metadata struktural

- ✓ mendeskripsikan struktur, format, dan hubungan antara data, seperti tipe fail, ukuran fail, kolom, baris, dan relasi.

## Metadata administratif

- ✓ mendeskripsikan aspek-aspek administratif dari data, seperti hak cipta, hak akses, keamanan, privasi, dan *audit trail*.



# Pengertian dan Peran Metadata Big Data

- Metadata Big Data adalah metadata yang berhubungan dengan data yang memiliki volume, kecepatan, dan variasi yang sangat besar.
- Data Big Data berasal dari berbagai sumber dan jenis, seperti sensor, media sosial, web, video, audio, teks, gambar, dan lainnya.
- Peran:
  - Membantu mengidentifikasi, mengklasifikasikan, dan mengkatalogkan data yang berasal dari berbagai sumber dan jenis.
  - Membantu mengevaluasi kualitas, keakuratan, dan keandalan data.
  - Membantu menentukan hak akses, keamanan, dan privasi data.
  - Membantu menemukan, mengintegrasikan, dan menganalisis data yang relevan dengan kebutuhan pengguna.
  - Membantu memahami makna, konteks, dan hubungan antara data.



# Teknik Pengelolaan Metadata Big Data

- Menentukan sumber, jenis, dan format data Big Data yang akan dikelola.
- Menentukan standar, aturan, dan prosedur untuk mengumpulkan, menyimpan, dan mengakses metadata Big Data.
- Menggunakan teknologi metadata Big Data yang sesuai dengan kebutuhan dan tujuan pengelolaan data Big Data.
- Melakukan proses ekstraksi, transformasi, dan pemuatan (*Extraction-Transformation-Load* atau ETL) metadata Big Data dari sumber-sumber data ke dalam sistem metadata Big Data.
- Melakukan proses pencarian, penggabungan, dan analisis metadata Big Data untuk mendapatkan informasi yang berguna dan bermakna.
- Melakukan proses pembaruan, penghapusan, dan penyimpanan metadata Big Data secara berkala dan sesuai dengan perubahan data Big Data.



# Teknologi Metadata Big Data: Apache Atlas

- Apache Atlas adalah sebuah platform terbuka yang menyediakan kemampuan untuk mengelola metadata Big Data secara terintegrasi, konsisten, dan komprehensif.
- Apache Atlas dapat mengumpulkan, menyimpan, mengindeks, mencari, dan mengelola metadata Big Data dari berbagai sumber, seperti Hadoop, Hive, HBase, Kafka, Storm, dan lainnya.
- Apache Atlas juga menyediakan fitur-fitur seperti *lineage tracking*, *audit trail*, *security*, dan *governance*.
  - *Lineage tracking* adalah fitur yang menunjukkan asal-usul, perubahan, dan aliran data Big Data dari sumber ke tujuan.
  - *Audit trail* adalah fitur yang mencatat aktivitas-aktivitas yang terjadi pada data Big Data, seperti siapa, kapan, apa, dan bagaimana.
  - *Security* adalah fitur yang mengatur hak akses, keamanan, dan privasi data Big Data, seperti siapa yang boleh melihat, mengubah, dan menghapus data Big Data.
  - *Governance* adalah fitur yang mengatur standar, aturan, dan prosedur untuk pengelolaan data Big Data, seperti bagaimana data Big Data dikumpulkan, disimpan, diakses, dan dianalisis.





# Teknologi Metadata Big Data: Apache Hive


- Apache Hive adalah sebuah sistem penyimpanan dan analisis data yang berbasis pada Hadoop.
- Apache Hive menyediakan kemampuan untuk membuat, mengakses, dan memanipulasi tabel data yang disimpan dalam *Hadoop Distributed File System* (HDFS) atau sistem penyimpanan lain yang kompatibel dengan Hadoop.
- Apache Hive juga menyediakan metadata tentang tabel data, seperti nama, tipe, kolom, partisi, statistik, dan lainnya.
- Metadata ini disimpan dalam sebuah komponen yang disebut Hive Metastore.
- Hive Metastore adalah sebuah basis data relasional yang menyimpan informasi tentang tabel data, skema, kolom, partisi, lokasi, statistik, dan lainnya.
- Hive Metastore dapat diakses oleh berbagai aplikasi yang menggunakan data Big Data, seperti Spark, Pig, Presto, Impala, dan lainnya.



# Teknik Pengelolaan Metadata Big Data:

## Menentukan Sumber, Jenis, dan Format Data Big Data

- Langkah pertama dalam pengelolaan metadata Big Data adalah menentukan sumber, jenis, dan format data Big Data yang akan dikelola.
- Sumber data Big Data adalah asal atau tempat di mana data Big Data berasal, seperti sensor, media sosial, web, video, audio, teks, gambar, dan lainnya.
- Jenis data Big Data adalah kategori atau klasifikasi data Big Data berdasarkan karakteristiknya, seperti data terstruktur, data semi-terstruktur, data tidak terstruktur, data streaming, data batch, dan lainnya.
- Format data Big Data adalah bentuk atau cara penyajian data Big Data, seperti CSV, JSON, XML, AVRO, PARQUET, ORC, dan lainnya.
- Menentukan sumber, jenis, dan format data Big Data dapat membantu mengidentifikasi dan mengkatalogkan data Big Data, serta menentukan teknologi metadata Big Data yang sesuai.



# Teknik Pengelolaan Metadata Big Data:

## Menentukan Standar, Aturan, dan Prosedur Metadata Big Data

- Langkah kedua dalam pengelolaan metadata Big Data adalah menentukan standar, aturan, dan prosedur untuk mengumpulkan, menyimpan, dan mengakses metadata Big Data.
- Standar adalah kriteria atau pedoman yang digunakan untuk menjamin kualitas, konsistensi, dan interoperabilitas metadata Big Data, seperti ISO 11179, Dublin Core, *Data Catalog Vocabulary* (DCAT), dan lainnya.
- Aturan adalah ketentuan atau peraturan yang harus dipatuhi dalam pengelolaan metadata Big Data, seperti hak cipta, hak akses, keamanan, privasi, dan *governance*.
- Prosedur adalah langkah-langkah atau cara yang harus dilakukan dalam pengelolaan metadata Big Data, seperti proses ETL, pencarian, penggabungan, analisis, pembaruan, penghapusan, dan penyimpanan metadata Big Data.
- Menentukan standar, aturan, dan prosedur metadata Big Data dapat membantu mengevaluasi, mengatur, dan memanfaatkan metadata Big Data secara efektif dan efisien.






# Teknik Pengelolaan Metadata Big Data: Menggunakan Teknologi Metadata Big Data

- Langkah ketiga dalam pengelolaan metadata Big Data adalah menggunakan teknologi metadata Big Data yang sesuai dengan kebutuhan dan tujuan pengelolaan data Big Data.
- Teknologi metadata Big Data adalah alat atau sistem yang dapat mengelola metadata Big Data dari berbagai sumber dan jenis data, serta menyediakan fitur-fitur tambahan seperti *lineage tracking*, *audit trail*, *security*, dan *governance*.
- Beberapa teknologi metadata Big Data yang populer adalah Apache Atlas dan Apache Hive, yang telah dijelaskan sebelumnya .
- Menggunakan teknologi metadata Big Data dapat membantu mengumpulkan, menyimpan, mengindeks, mencari, dan mengelola metadata Big Data secara terintegrasi, konsisten, dan komprehensif.



# Teknik Pengelolaan Metadata Big Data: Melakukan Proses ETL Metadata Big Data

- Langkah keempat dalam pengelolaan metadata Big Data adalah melakukan proses ekstraksi, transformasi, dan pemuatan (ETL) metadata Big Data dari sumber-sumber data ke dalam sistem metadata Big Data.
- Proses ETL adalah proses yang melibatkan tiga tahap, yaitu:
  - Ekstraksi: tahap ini melibatkan pengambilan atau pengunduhan metadata Big Data dari sumber-sumber data, seperti Hadoop, Hive, HBase, Kafka, Storm, dan lainnya.
  - Transformasi: tahap ini melibatkan perubahan atau penyesuaian metadata Big Data sesuai dengan standar, aturan, dan prosedur yang telah ditentukan, seperti membersihkan, memvalidasi, menggabungkan, dan meng enrich metadata Big Data.
  - Pemuatan: tahap ini melibatkan penyimpanan atau pengunggahan metadata Big Data ke dalam sistem metadata Big Data, seperti Apache Atlas atau Hive Metastore.
  - Proses ETL dapat dilakukan secara manual, otomatis, atau semi-otomatis, tergantung pada teknologi metadata Big Data yang digunakan dan kebutuhan pengelolaan data Big Data.



# Teknik Pengelolaan Metadata Big Data: Melakukan Proses Pencarian, Penggabungan, dan Analisis Metadata Big Data

- Langkah kelima dalam pengelolaan metadata Big Data adalah melakukan proses pencarian, penggabungan, dan analisis metadata Big Data untuk mendapatkan informasi yang berguna dan bermakna.
- Proses pencarian adalah proses yang melibatkan pencarian atau query metadata Big Data berdasarkan kriteria atau kata kunci tertentu, seperti nama, tipe, sumber, kualitas, dan lainnya.
- Proses penggabungan adalah proses yang melibatkan penggabungan atau integrasi metadata Big Data dari berbagai sumber dan jenis data, untuk mendapatkan gambaran yang lebih lengkap dan holistik tentang data Big Data.
- Proses analisis adalah proses yang melibatkan analisis atau eksplorasi metadata Big Data untuk mendapatkan wawasan atau pengetahuan baru tentang data Big Data, seperti makna, konteks, hubungan, pola, tren, dan lainnya.
- Proses pencarian, penggabungan, dan analisis dapat dilakukan dengan menggunakan teknologi metadata Big Data yang telah disediakan, seperti Apache Atlas atau Hive Metastore, atau dengan menggunakan alat atau aplikasi analisis data lainnya, seperti Spark, Pig, Presto, Impala, dan lainnya.



# Teknik Pengelolaan Metadata Big Data: Melakukan Proses Pembaruan, Penghapusan, dan Penyimpanan Metadata Big Data

- Langkah keenam dan terakhir dalam pengelolaan metadata Big Data adalah melakukan proses pembaruan, penghapusan, dan penyimpanan metadata Big Data secara berkala dan sesuai dengan perubahan data Big Data.
- Proses pembaruan adalah proses yang melibatkan pembaruan atau perubahan metadata Big Data sesuai dengan perubahan data Big Data, seperti penambahan, pengurangan, atau modifikasi data Big Data.
- Proses penghapusan adalah proses yang melibatkan penghapusan atau penghapusan metadata Big Data yang tidak lagi relevan, valid, atau diperlukan, seperti data Big Data yang sudah kedaluwarsa, rusak, atau salah.
- Proses penyimpanan adalah proses yang melibatkan penyimpanan atau *backup* metadata Big Data untuk mencegah kehilangan, kerusakan, atau pencurian metadata Big Data, seperti menggunakan sistem penyimpanan awan, lokal, atau hibrida.
- Proses pembaruan, penghapusan, dan penyimpanan dapat dilakukan dengan menggunakan teknologi metadata Big Data yang telah disediakan, seperti Apache Atlas atau Hive Metastore, atau dengan menggunakan alat atau aplikasi manajemen data lainnya, seperti Hadoop, HDFS, atau sistem penyimpanan lainnya.





# Tanya Jawab Diskusi