

## Bab 2

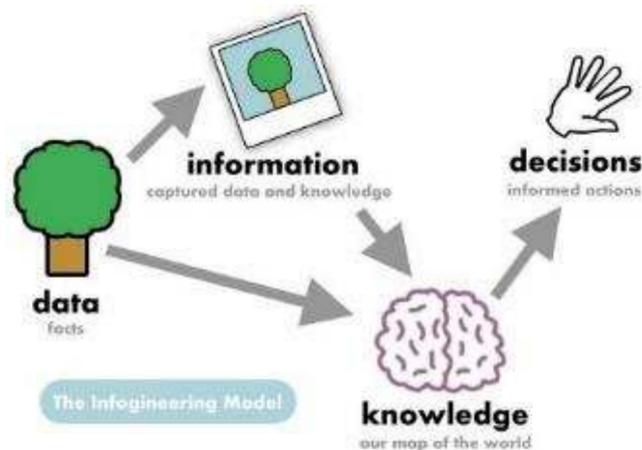
# KOMPONEN DAN KARAKTERISTIK BIG DATA

Dani Anggoro, S.Kom., M.Kom

### 2.1 Komponen Big Data

Dalam perspektif ilmu pengetahuan, kita melihat berbagai perangkat keras seperti HDD, SSD, dan perangkat serupa sebagai sarana penyimpanan yang digunakan. Di sisi lain, dalam konteks jaringan biologi, kita diberikan Otak oleh Sang Pencipta atau Sang Pencipta. Otak ini adalah perangkat penyimpanan biologis yang luar biasa yang memungkinkan kita untuk menyimpan dan memproses informasi dengan canggih.

Pentingnya mengolah data-data kecil hingga membentuk Big Data tak bisa diabaikan. Ini seakan kita mengumpulkan serpihan puzzle kecil yang, ketika digabungkan, membentuk gambaran yang jauh lebih besar dan kaya. Kita bisa membayangkan diri kita sebagai kurator data, mengumpulkan setiap potongan informasi. Sekarang, mari kita fokus pada Gambar 1 yang mengilustrasikan gambaran umum tentang Big Data.



Gambar 1 Gambaran Umum Big Data

Dari ilustrasi Gambar 1 di atas, terlihat beberapa komponen penting dalam konteks Big Data, antara lain:

- Data (Fakta, deskripsi Dunia):
- Informasi mentah yang menggambarkan realitas. Informasi (Data Tertangkap dan Pengetahuan): Proses mencatat atau mengambil Data dan Pengetahuan pada suatu titik waktu tertentu. Penting untuk disadari bahwa Data dan Pengetahuan bersifat dinamis dan dapat berkembang seiring berjalannya waktu.
- Pengetahuan (Peta/Model Pribadi Dunia Kita): Ini adalah representasi pribadi kita tentang dunia, bukan dunia sebenarnya itu sendiri. Saat ini, satu-satunya wadah yang mampu menyimpan Pengetahuan adalah otak kita, dan untuk membangun Pengetahuan, kita memerlukan Informasi dan Data sebagai bahan baku.

Dengan demikian, pemahaman kita tentang Big Data dan proses transformasi dari Data menjadi Pengetahuan sangat bergantung pada bagaimana elemen-elemen ini saling berhubungan dan berinteraksi.

## **2.2 Karakteristik Big Data**

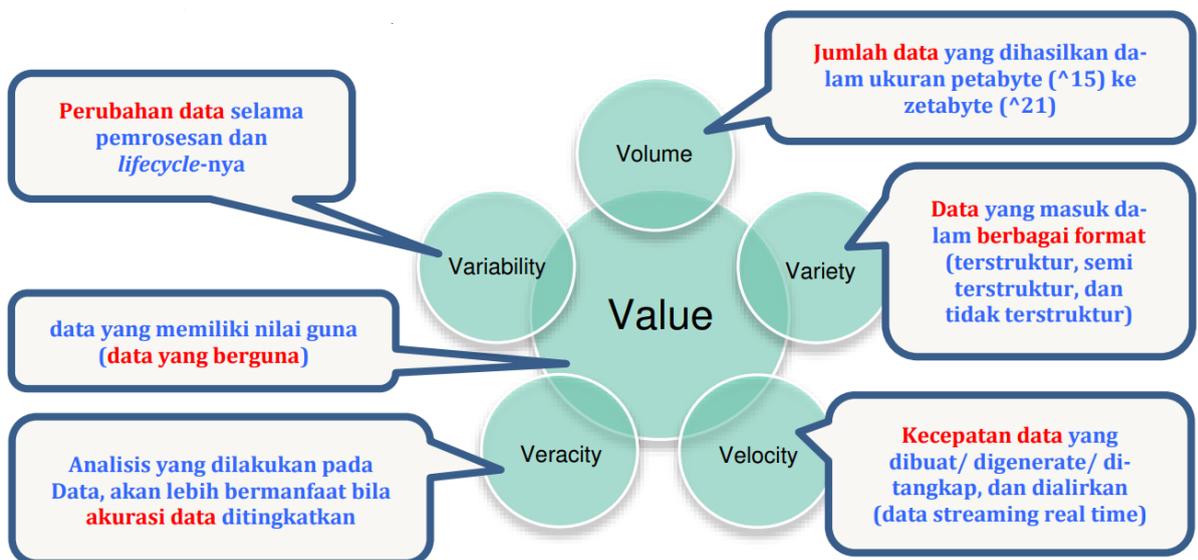
Big Data adalah salah satu konsep yang telah mengubah paradigma dalam pengelolaan data dan analisis di berbagai sektor. Menurut McKinsey Global (2011), Big Data bisa didefinisikan sebagai kumpulan data yang memiliki karakteristik khusus yang meliputi skala (volume), distribusi (velocity), keragaman (variety) yang sangat besar, dan dalam beberapa kasus, juga bersifat abadi. Hal ini berarti bahwa volume data yang dihasilkan dan tersedia saat ini melebihi apa yang bisa diolah dengan metode

tradisional, sehingga membutuhkan pendekatan baru dalam pengumpulan, pengelolaan, dan analisis data untuk mendapatkan wawasan yang bermakna.

Dalam perkembangannya, konsep Big Data telah berkembang lebih jauh dengan munculnya sejumlah dimensi tambahan yang dikenal sebagai "V" dalam dunia Big Data. Seiring dengan Volume, Velocity, dan Variety, terdapat beberapa V tambahan yang juga menjadi bagian integral dari Big Data. Beberapa dari V tambahan tersebut adalah Variability, Veracity, Value, dan Visualization.

- Volume adalah karakteristik yang paling jelas dalam Big Data. Ini mencerminkan jumlah data yang sangat besar yang terus-menerus dihasilkan dan terakumulasi dari berbagai sumber seperti sensor, perangkat mobile, media sosial, dan sistem lainnya. Volume data ini begitu besar sehingga memerlukan infrastruktur yang kuat untuk penyimpanan dan pemrosesan data yang efisien.
- Velocity merujuk pada laju pertumbuhan data yang sangat cepat. Dalam era digital ini, data diproduksi dalam jumlah besar dalam waktu nyata. Contohnya adalah data lalu lintas internet, data transaksi keuangan, atau data sensor dari peralatan industri. Kemampuan untuk mengakses, memproses, dan merespons data ini dengan cepat menjadi kunci dalam memanfaatkan potensi Big Data.
- Variety menggambarkan keragaman tipe data yang ada dalam lingkungan Big Data. Data tidak hanya berupa angka atau teks, tetapi juga dapat berupa gambar, video, audio, data geografis, dan berbagai format data lainnya. Integrasi dan analisis data dari berbagai sumber dan jenis merupakan tantangan penting dalam pengelolaan Big Data.
- Variability menunjukkan fluktuasi dalam data seiring waktu. Data tidak selalu konsisten, dan perubahan dalam struktur atau format data dapat terjadi. Ini memerlukan kemampuan untuk mengidentifikasi perubahan ini dan beradaptasi dengan cepat.

- Veracity adalah tingkat kepastian dan keandalan data. Big Data seringkali mencakup data yang tidak sempurna atau cacat. Veracity mengacu pada kemampuan untuk mengukur dan mengelola kualitas data sehingga keputusan bisnis dapat diandalkan.
- Value adalah inti dari Big Data. Nilai sebenarnya dari Big Data terletak pada kemampuannya untuk memberikan wawasan yang bermanfaat dan memberikan nilai tambah bagi organisasi. Dengan menganalisis Big Data dengan benar, organisasi dapat mengidentifikasi peluang, mengoptimalkan proses, dan mengambil keputusan yang lebih baik.
- Visualization adalah cara untuk menyajikan data yang kompleks dengan cara yang mudah dipahami. Ini melibatkan penggunaan grafik, diagram, dan alat visual lainnya untuk membantu orang memahami pola dan trend dalam data.



Gambar 2 Big Data dengan 6V+ 1V(Visualization)=7V

Big data merupakan istilah untuk sekumpulan data yang begitu besar atau kompleks dimana tidak bisa ditangani lagi dengan sistem teknologi komputer konvensional (Hurwitz, et al., 2013). Kapan suatu data dapat dikatakan sebagai Big Data?



Gambar 3 Big Data dengan 10V

Dengan begitu banyak V yang berkontribusi pada Big Data, jelas bahwa Big Data bukanlah sekadar volume data yang besar, tetapi juga melibatkan berbagai aspek yang mencakup karakteristik, kualitas, kecepatan, dan nilai data. Dalam dunia bisnis dan ilmu pengetahuan, pemahaman yang mendalam tentang berbagai aspek ini menjadi kunci untuk mengambil keputusan yang informasional dan berdaya saing dalam era Big Data ini.

### Karakteristik Big Data 3 V

- Volume

Beberapa contoh volume data dalam kehidupan sehari-hari

- 1) Facebook menciptakan sebanyak 10 terabyte data segar dalam sehari, sementara Twitter menghasilkan sekitar 7 terabyte.
- 2) Sebuah pesawat Boeing 737, ketika melakukan penerbangan dari satu wilayah di Amerika Serikat ke wilayah lainnya, menghasilkan sebanyak 240 terabyte data penerbangan selama perjalanan.
- 3) Menurut pernyataan Steve Ballmer pada tahun 2013, Microsoft saat ini mengoperasikan sekitar satu juta server, jumlah ini lebih besar dibandingkan dengan Amazon, namun masih kalah dari jumlah server yang dimiliki oleh Google.

Catatan: Kita dapat menggunakan semua data ini untuk memberitahu kita sesuatu, jika kita mengetahui pertanyaan yang tepat untuk bertanya.

Gambar 3. menjelaskan perbedaan volume pada penyimpanan tradisional seperti database dengan big data

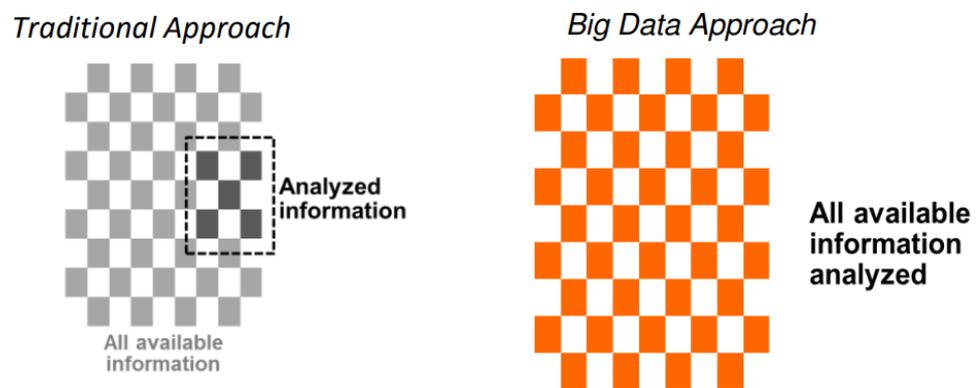


Gambar 3, Perbandingan Volume Penyimpanan Data tradisional dengan Big Data

Kata Big Data mengarah kepada manajemen informasi skala besar (large-scale) dan teknologi analisis yang melebihi kapabilitas teknologi data secara tradisional. Terdapat perbedaan yang paling utama dalam tiga hal antara

Tradisional dengan Big Data, yaitu amount of data (volume), the rate of data generation and transmission (velocity), dan the types of structured and unstructured data (variety).

Pada gambar 4 digambarkan gambaran traditional approach dan big data approach



Gambar 4 Ilustrasi bagaimana proses analisa pada big data

Big data menganalisa semua jenis file, sementara penyimpanan tradisional hanya menganalisa bagian tertentu yang memiliki jenis file yang sama.

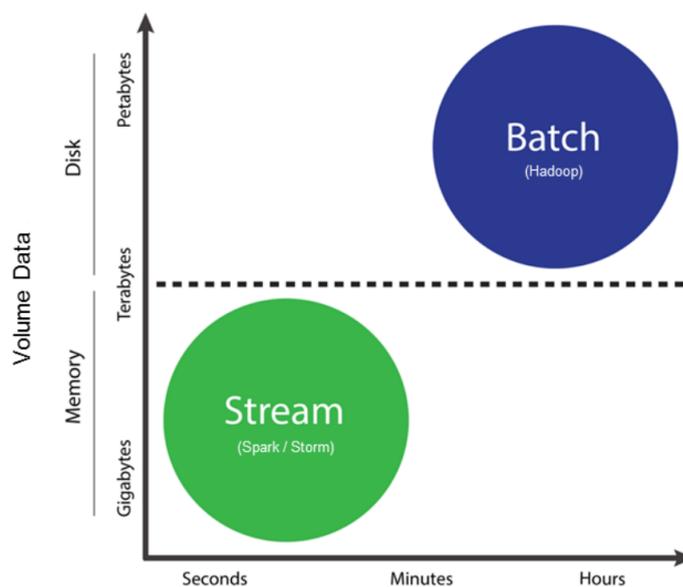
Teknologi Big Data memiliki dua kelompok utama dalam pemrosesannya: batch processing yang digunakan untuk menganalisis data yang sudah terkumpul pada satu waktu tertentu (data at rest), dan streaming processing yang digunakan untuk menganalisis data yang terus-menerus mengalir dan terupdate setiap saat (data in motion).

Contoh batch processing adalah ketika sebuah perusahaan ingin menganalisis data penjualan tahunan mereka untuk mengidentifikasi tren dan pola. Mereka mengumpulkan semua data penjualan selama satu tahun dan kemudian

menerapkan analisis batch untuk menghasilkan wawasan tentang kinerja bisnis mereka selama periode tersebut.

Contoh streaming processing adalah ketika platform media sosial seperti Twitter atau Facebook mengamati dan menganalisis aliran data yang terus berjalan dari pengguna mereka. Mereka menggunakan streaming processing untuk mendeteksi tren topik yang sedang populer, mengidentifikasi konten yang viral, atau memonitor aktivitas yang mencurigakan secara real-time.

Kedua pendekatan ini memiliki peran penting dalam ekosistem Big Data, di mana batch processing cocok untuk analisis data historis yang besar dan statis, sementara streaming processing memungkinkan organisasi untuk mengambil tindakan segera berdasarkan data yang terus berubah.



Gambar 5 Batch dan Stream Batch

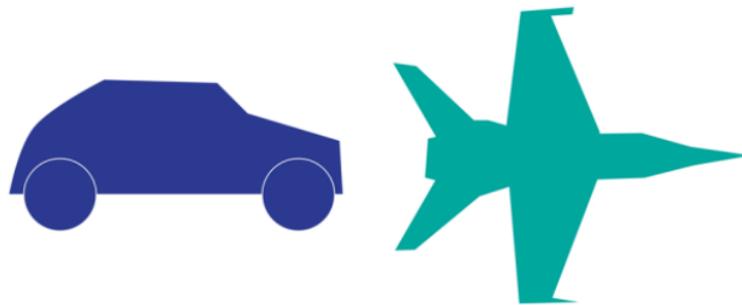
- Velocity

Velocity adalah istilah yang mengacu pada kecepatan data yang masuk atau diterima dalam suatu sistem atau platform, dan kecepatan ini dapat diukur dalam berbagai satuan waktu seperti per jam atau per detik. Dalam konteks Big Data, ini sering terkait dengan aliran data yang berkelanjutan, seperti clickstreams (log web) dan transfer data asynchronous yang mampu mencatat semua aktivitas yang dilakukan oleh jutaan pengguna secara bersamaan.

Clickstreams atau log web adalah salah satu sumber data yang sangat menarik dalam analisis data. Mereka mencatat setiap tindakan yang dilakukan oleh pengunjung pada sebuah situs web. Dengan kata lain, setiap kali seseorang meminta halaman web, clickstream akan mencatatnya. Ini berarti bahwa clickstream dapat merekam setiap gerakan atau tindakan yang dilakukan oleh pengunjung, yang pada gilirannya dapat memberikan informasi berharga tentang kebiasaan dan preferensi pengguna tersebut.

Dengan menggunakan data clickstream, kita dapat mencoba mengidentifikasi berbagai aspek, seperti sesi yang sukses dan yang tidak, tingkat kepuasan pengunjung, serta bagian dari situs web yang paling efektif dalam menarik perhatian pengunjung. Dengan demikian, clickstream memungkinkan kita untuk memahami lebih baik bagaimana pengguna berinteraksi dengan suatu situs web atau platform online, dan bagaimana kita dapat meningkatkan pengalaman mereka.

Ilustrasi bagaimana kecepatan big data tergambar pada gambar 6. Jika penambahan data pada data tradisional secepat mobil maka pada big data proses penambahan file secepat pesawat jet.



Gambar 6 Perbandingan Kecepatan Data Tradisional dengan Big Data

- Variety

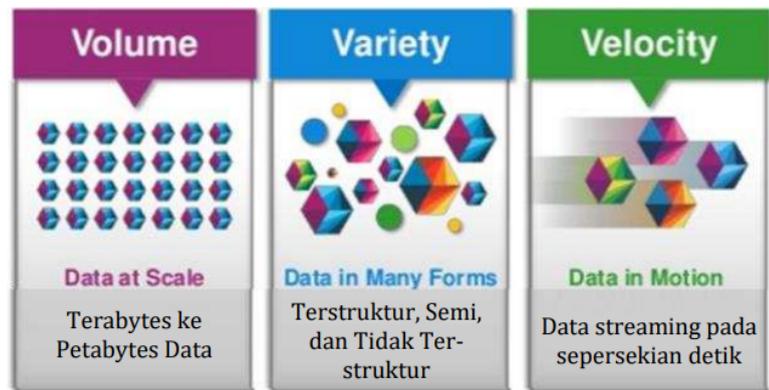
Variety adalah konsep yang mencakup beragam jenis data, termasuk data yang memiliki struktur yang jelas, data yang memiliki struktur sebagian, dan data yang sama sekali tidak memiliki struktur (yang biasanya lebih mendominasi). Dengan Variety, kita dapat melihat data dalam berbagai bentuk dan format yang berbeda, sehingga memberikan pandangan yang lebih lengkap dan komprehensif tentang informasi yang tersedia.



Gambar 7. Variasi data pada big data

Data terstruktur, seperti yang terdapat dalam database, data semi terstruktur, yang bisa ditemui dalam file, dan data tidak terstruktur, seperti yang sering muncul dalam media sosial atau percakapan digital, adalah tiga jenis data yang berbeda berdasarkan tingkat keteraturan dan pengorganisasiannya.

- 1) Data Terstruktur: Data terstruktur adalah jenis data yang memiliki format dan tata letak yang terorganisir dengan baik. Data ini sering disimpan dalam tabel atau basis data yang telah ditentukan sebelumnya. Contoh data terstruktur adalah catatan penjualan dalam basis data yang memiliki kolom-kolom seperti nama pelanggan, tanggal transaksi, jumlah belanja, dan produk yang dibeli.
- 2) Data Semi Terstruktur: Data semi terstruktur adalah data yang memiliki sebagian struktur, tetapi tidak selengkap data terstruktur. Data ini mungkin memiliki beberapa aturan atau format, tetapi juga dapat mengandung elemen yang lebih bebas atau tidak terorganisir. Sebagai contoh, file XML atau JSON adalah jenis data semi terstruktur. Di dalamnya, ada format dan hierarki, tetapi juga dapat memiliki bidang tambahan yang tidak selalu terdapat dalam setiap entitas.
- 3) Data Tidak Terstruktur: Data tidak terstruktur adalah jenis data yang tidak memiliki format atau struktur yang jelas. Data ini seringkali berupa teks bebas atau informasi yang dicatat tanpa pengaturan yang khusus. Contoh data tidak terstruktur adalah untaian percakapan di media sosial seperti Twitter, di mana setiap pengguna dapat mengirimkan pesan dengan panjang dan format yang berbeda-beda.



Gambar 8. Rangkuman 3V

### 2.3 Kesimpulan

Dalam Big Data, terdapat tiga komponen utama: Data sebagai fakta tentang realitas, Informasi sebagai hasil dari proses mencatat dan mengambil Data serta Pengetahuan, yang merupakan representasi pribadi tentang dunia. Data dan Pengetahuan bersifat dinamis dan berkembang seiring waktu, dan untuk membangun Pengetahuan, kita memerlukan Informasi dan Data sebagai bahan dasar.

Big data umumnya memiliki 3 karakteristik data yaitu Volume: Big Data memiliki volume yang sangat besar, dengan jumlah data yang terus bertambah secara signifikan dari berbagai sumber. Velocity: Data dalam Big Data bergerak dengan kecepatan tinggi, termasuk aliran data real-time yang terus-menerus. Variety: Big Data mencakup beragam jenis data, termasuk data terstruktur, semi terstruktur, dan tidak terstruktur, yang memerlukan pendekatan yang berbeda dalam pengelolaan dan analisisnya.