

BAB XII *BIG DATA*

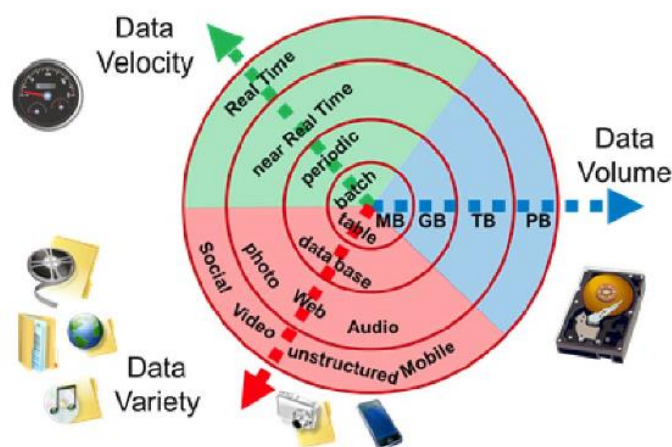
Tujuan Pembelajaran (Chapter VIII *Learning Objectives*)

- Mahasiswa dapat menjelaskan pengertian dan karakteristik *Big Data*;
- Mahasiswa dapat menyebutkan tipe-tipe *Big Data*;
- Mahasiswa dapat menjelaskan teknologi *Big Data*;
- Mahasiswa dapat mengidentifikasi manfaat *Big Data*;
- Mahasiswa dapat menjelaskan contoh pemanfaatan *Big Data* di berbagai bidang.

A. Pengertian dan Karakteristik *Big Data*

Big Data bukanlah sebuah teknologi, teknik, maupun inisiatif yang berdiri sendiri. *Big Data* adalah suatu trend yang mencakup area yang luas dalam dunia bisnis dan teknologi. *Big Data* merupakan jumlah data yang sangat besar yang diproduksi di era digital. Data dalam jumlah besar ini, baik terstruktur maupun tidak terstruktur, mencakup semua data web yang dihasilkan oleh email, situs web, dan situs jejaring sosial (Holmes, 2017).

Big Data menunjuk pada teknologi dan inisiatif yang melibatkan data yang begitu beragam, cepat berubah, atau berukuran super besar sehingga terlalu sulit bagi teknologi, keahlian, maupun infrastruktur konvensional untuk dapat menanganinya secara efektif. Dengan kata lain, *Big Data* memiliki ukuran (volume), kecepatan (velocity), atau ragam (variety) yang terlalu ekstrim untuk dikelola dengan teknik konvensional. Tiga hal tersebut yang kemudian menjadi karakteristik dari *Big Data*.



Gambar 1. Karakteristik *Big Data*
(sumber: https://commons.wikimedia.org/wiki/File:Big_Data.png)

Big Data melibatkan proses pembuatan data, penyimpanan, penggalian informasi, dan analisis yang menonjol dalam hal *volume*, *velocity*, dan *variety* (Kamal & Saxena, 2019).

1. Volume (Ukuran)

Sesuai dengan namanya, kita perlu memproses *volume* data yang cukup besar untuk *Big Data*. Pada tahun 2000, PC biasa pada umumnya memiliki kapasitas penyimpanan sekitar 10 gigabytes. Saat ini, Facebook menyedot sekitar 500 terabytes data baru setiap harinya; sebuah pesawat Boeing 737 menghasilkan sekitar 240 terabytes data penerbangan dalam satu penerbangan melintasi Amerika; makin menjamurnya penggunaan ponsel pintar (smartphone), bertambahnya sensor-sensor yang disertakan pada perangkat harian, akan terus mengalirkan jutaan data-data baru, yang terus ter-update, yang mencakup data-data yang berhubungan dengan lingkungan, lokasi, cuaca, video bahkan data tentang suasana hati si pengguna ponsel pintar.

2. Velocity (kecepatan)

Velocity disini adalah kecepatan yang sangat cepat di mana data diterima dan (mungkin) langsung digunakan. Clickstreams maupun ad impressions mencatat perilaku pengguna Internet dalam jutaan event per detik; algoritma jual-beli saham dalam frekwensi tinggi dapat mencerminkan perubahan pasar dalam hitungan microseconds; proses-proses yang melibatkan hubungan antara suatu mesin dengan mesin lainnya telah melibatkan pertukaran data antar jutaan perangkat; peralatan sensor dan perangkat-perangkat pada infrastruktur menghasilkan log data secara real time; sistem game online dapat melayani jutaan pengguna secara bersamaan, yang masing-masing memberikan sejumlah input per detiknya.

3. Variety (ragam)

Variety yang dimaksud disini adalah berbagai jenis data yang tersedia. Jenis data tradisional biasanya lebih terstruktur. Dengan semakin berkembangnya *Big Data*, ada juga data yang belum terstruktur. *Big Data* tidak hanya menyangkut data yang berupa angka-angka, data tanggal, dan rangkaian teks. *Big Data* juga meliputi data-data ruang / geospasial, data 3D, audio dan video, dan data-data teks tak berstruktur termasuk file-file log dan media sosial. Sistem database tradisional didesain untuk menangani data-data berstruktur, yang tak terlalu sering mengalami update atau updatenya dapat diprediksi, serta memiliki struktur data

yang konsisten yang volumenya tak pernah sebesar *Big Data*. Selain itu, sistem database tradisional juga didesain untuk digunakan dalam satu server yang berdiri sendiri, yang berakibat pada keterbatasan dan mahalnya biaya untuk peningkatan kapasitas, sedangkan aplikasi sudah dituntut untuk mampu melayani pengguna dalam jumlah yang jauh lebih besar dari yang pernah ada sebelumnya. Dalam hal ini, database *Big Data* seperti halnya MongoDB maupun HBase, dapat memberikan solusi yang feasible yang memungkinkan peningkatan profit perusahaan secara signifikan.

B. Tipe *Big Data*

Setelah mengetahui pengertian dari *Big Data*, selanjutnya berikut beberapa prinsip dasar mengenai *Big Data* yang akan membuatnya lebih sederhana untuk memahami apa itu *Big Data*:

1. *Big Data* mengacu pada sejumlah besar data yang terus tumbuh secara eksponensial seiring waktu.
2. *Big Data* sangat banyak sehingga tidak dapat diproses atau dianalisis menggunakan teknik pemrosesan data konvensional.
3. *Big Data* data mining, data storage, data analysis, data sharing, dan data visualization.
4. *Big Data* memiliki beberapa tipe data sehingga terdapat alat dan teknik yang digunakan untuk memproses dan menganalisis data tersebut.

Lebih lanjut, mari kita bahas mengenai tipe-tipe *Big Data* sebagai berikut:

1. *Structured* (terstruktur)

Terstruktur adalah salah satu jenis *Big Data*, yang dimaksud dengan data terstruktur adalah data yang dapat diproses, disimpan, dan diambil dalam format tetap. Ini mengacu pada informasi yang sangat terorganisir yang dapat dengan mudah disimpan dan diakses dari database dengan algoritma mesin pencari sederhana. Misalnya, tabel karyawan dalam database perusahaan akan terstruktur karena detail karyawan, posisi pekerjaan, gaji, dll sehingga dapat ditampilkan secara terorganisir.

2. *Unstructured* (tidak terstruktur)

Data tidak terstruktur mengacu pada data yang tidak memiliki bentuk atau struktur tertentu apa pun. Ini membuatnya sangat sulit dan memakan waktu untuk

memproses dan menganalisis data yang tidak terstruktur. Email adalah contoh data yang tidak terstruktur. Terstruktur dan tidak terstruktur adalah dua jenis *Big Data* yang penting.

3. *Semi-structured (semi-terstruktur)*

Semi terstruktur adalah jenis *Big Data* ketiga. Data semi terstruktur berkaitan dengan data yang memuat kedua format tersebut di atas, yaitu data terstruktur dan tidak terstruktur. Tepatnya, ini mengacu pada data yang meskipun belum diklasifikasikan di bawah repositori (database) tertentu, namun berisi informasi atau tag penting yang memisahkan elemen individu dalam data.

C. Jenis Teknologi *Big Data*

Berdasarkan pengertiannya, *Big Data* menggambarkan kumpulan data yang begitu besar dan kompleks yang tak memungkinkan lagi untuk dikelola dengan tools software tradisional. Dalam hal Teknologi, bentangan *Big Data* didominasi oleh dua jenis teknologi *Big Data* yaitu: (1) *Big Data* operasional dan (2) *Big Data* analitis. Beban kerja operasional dan analitis terhadap *Big Data* telah menyebabkan kebutuhan sistem yang berlawanan satu sama lain, dan sistem *Big Data* saat ini telah berevolusi untuk menangani kedua jenis kerja tersebut secara khusus, terpisah, dan dengan cara yang sangat berbeda. Namun dalam keberadaannya, kedua jenis teknologi *Big Data* ini bersifat saling melengkapi dan sering digunakan secara bersamaan.

Akses data terhadap sistem operasional dapat dilakukan dengan berbagai pilihan kriteria. Dilain pihak, sistem analitis cenderung berfokus pada penanganan arus data yang lebih besar, query-query yang ditujukan pada data tersebut bisa sangat kompleks, dan setiap kali dieksekusi dapat melibatkan sebagian besar atau keseluruhan data yang ada dalam sistem. Baik sistem *Big Data* operasional maupun sistem *Big Data* analitis, kedua-duanya dioperasikan dengan melibatkan sejumlah servers yang tergabung dalam suatu cluster komputer, dan digunakan untuk mengelola puluhan atau ratusan terabytes data yang memuat miliaran record.

1. Teknologi *Big Data* Operasional

Big Data operasional merupakan sistem yang memiliki kapabilitas operasional untuk pekerjaan-pekerjaan bersifat interaktif dan real time dimana data pada umumnya diserap dan disimpan. Untuk menangani pekerjaan-pekerjaan *Big Data* Operasional, telah dibangun sistem *Big Data* dengan database NoSQL seperti halnya

database berbasis dokumen (document based database) yang dapat ditujukan untuk berbagai tipe aplikasi, database key-value stores, column family stores, dan database graph yang dioptimalkan untuk aplikasi yang lebih spesifik. Teknologi NoSQL, yang telah dikembangkan untuk mengatasi kekurangan dari database relasional (relational database) pada lingkungan komputasi modern, dikenal lebih cepat serta lebih mudah dan murah dalam hal peningkatan skala (more scalable) dibanding relational databases.

Terlebih lagi, sistem *Big Data* dengan database NoSQL telah didesain untuk memanfaatkan keunggulan dari arsitektur cloud computing (komputasi awan) yang telah muncul dalam dekade terakhir ini. Hal ini memungkinkan dijalankannya komputasi berskala besar secara efisien dan dengan biaya yang relatif lebih murah. Sebagai hasilnya, sistem NoSQL dengan komputasi awan ini telah menjadikan perangkat kerja *Big Data* operasional lebih mudah dikelola, serta dapat diimplementasikan dengan lebih murah dan cepat.

2. Teknologi *Big Data* Analitis

Big Data analitis merupakan sistem yang menyediakan kapabilitas analitis untuk mengerjakan analisis yang kompleks dan retrospektif yang dapat melibatkan sebagian besar atau bahkan keseluruhan data. Pekerjaan-pekerjaan *Big Data* analitis cenderung diproses dengan mengimplementasikan sistem database MPP dan MapReduce. Munculnya teknologi ini juga merupakan reaksi terhadap keterbatasan dan kurangnya kemampuan relational database tradisional untuk mengelola database dalam skala lebih dari satu server (terdistribusi). Disamping itu, MapReduce juga menawarkan metode baru dalam menganalisa data yang dapat berfungsi sebagai pelengkap terhadap kapabilitas SQL.

Dengan semakin populernya penggunaan berbagai jenis aplikasi dan para penggunanya terus menerus memproduksi data dari pemakaian aplikasi tersebut, terdapat sejumlah upaya analisa retrospektif yang benar-benar dapat memberikan nilai berarti terhadap kemajuan bisnis. Ketika upaya-upaya tersebut mesti melibatkan algoritma yang lebih rumit, MapReduce telah menjadi pilihan pertama untuk melakukan analisa retrospektif tersebut. Beberapa sistem NoSQL juga menyediakan fungsi MapReduce bawaan yang memungkinkan proses analisa diterapkan pada data operasional. Sebagai alternatif lain, data juga dapat dikopi

dari sistem NoSQL ke dalam sistem analitis seperti halnya Hadoop dengan MapReduce-nya.

Setelah MapReduce (yang telah menjadi komponen utama dari Hadoop), kemudian muncul Apache Spark sebagai jawaban terhadap kebutuhan akan platform analitik yang lebih cepat dan terpadu. Dikatakan bahwa kemampuan processing Apache Spark adalah 100 kali lebih cepat dibanding Hadoop. Lebih detail tentang Apache Spark ada disini: *Apache Spark, Platform Analitik Terpadu dan Super Cepat untuk Big Data*. Apache Spark dapat dicoba di perangkat komputer sendiri (local environment) maupun dengan menggunakan layanan berbasis cloud computing.

D. Manfaat *Big Data*

Salah satu manfaat terbesar dari *Big Data* adalah analisis prediktif. Analitik *Big Data* dapat memprediksi hasil secara akurat, sehingga memungkinkan bisnis dan organisasi untuk membuat keputusan yang lebih baik, sekaligus mengoptimalkan efisiensi operasional dan mengurangi risiko. Manfaat lain dari *Big Data* khususnya di bidang bisnis, sebagai berikut:

1. Dengan memanfaatkan data dari platform media sosial menggunakan alat analitik *Big Data*, bisnis di seluruh dunia dapat merampingkan strategi pemasaran digital untuk meningkatkan pengalaman konsumen secara keseluruhan. *Big Data* memberikan wawasan tentang poin kesulitan pelanggan dan memungkinkan perusahaan untuk meningkatkan produk dan layanan mereka.
2. Meningkatkan akurasi, *Big Data* menggabungkan data yang relevan dari berbagai sumber untuk menghasilkan wawasan yang sangat dapat ditindaklanjuti. Hampir 43% perusahaan tidak memiliki alat yang diperlukan untuk menyaring data yang tidak relevan, yang pada akhirnya menghabiskan jutaan dolar untuk memilah-milah data yang berguna dari massal. Analitik *Big Data* dapat membantu mengurangi ini, menghemat waktu dan uang.
3. Analitik *Big Data* dapat membantu perusahaan menghasilkan lebih banyak arahan penjualan yang secara alami akan berarti peningkatan pendapatan. Bisnis menggunakan alat analitik *Big Data* untuk memahami seberapa baik kinerja produk / layanan mereka di pasar dan bagaimana pelanggan meresponsnya.
4. Dengan wawasan *Big Data*, perusahaan bisa selangkah lebih maju dari pesaingnya. Perusahaan dapat menyaring pasar untuk mengetahui jenis promosi dan penawaran apa yang disediakan pesaingnya, dan kemudian dapat memberikan

penawaran yang lebih baik untuk pelanggan. Selain itu, wawasan *Big Data* memungkinkan perusahaan mempelajari perilaku pelanggan untuk memahami tren pelanggan dan memberikan pengalaman yang sangat 'dipersonalisasi' kepada mereka.

E. Contoh Implementasi *Big Data*

Banyak bidang saat ini yang telah menggunakan *Big Data*, berikut contoh beberapa bidang diantaranya:

1. Kesehatan

Big Data sudah mulai membuat perbedaan besar di sektor perawatan kesehatan. Dengan bantuan analitik prediktif, profesional medis sekarang dapat memberikan layanan perawatan kesehatan yang dipersonalisasi untuk setiap pasien. Selain itu, perangkat kebugaran, telemedicine, pemantauan jarak jauh yang didukung oleh *Big Data* dan Artificial Intelligence dapat membantu mengubah hidup menjadi lebih baik.

2. Akademisi

Big Data juga membantu meningkatkan pendidikan saat ini. Pendidikan tidak terbatas pada batasan fisik ruang kelas - ada banyak kursus pendidikan online untuk dipelajari. Institusi akademis berinvestasi dalam kursus digital yang didukung oleh teknologi *Big Data* untuk membantu perkembangan pelajar pemula secara menyeluruh.

3. Perbankan

Sektor perbankan mengandalkan *Big Data* untuk deteksi penipuan. Alat *Big Data* dapat secara efisien mendeteksi tindakan penipuan secara real-time seperti penyalahgunaan kartu kredit / debit, pengarsipan jalur inspeksi, perubahan yang salah dalam statistik pelanggan, dll.

4. Manufaktur

Menurut TCS Global Trend Study, manfaat paling signifikan dari *Big Data* di bidang manufaktur adalah meningkatkan strategi pasokan dan kualitas produk. Di sektor manufaktur, *Big Data* membantu menciptakan infrastruktur yang transparan, sehingga memprediksi ketidakpastian dan ketidakmampuan yang dapat mempengaruhi bisnis secara negatif.

5. Teknologi Informasi (TI)

Salah satu pengguna *Big Data* terbesar, perusahaan TI di seluruh dunia menggunakan *Big Data* untuk mengoptimalkan fungsinya, meningkatkan produktivitas karyawan,

dan meminimalkan risiko dalam operasi bisnis. Dengan menggabungkan teknologi *Big Data* dengan Machine Learning dan Artificial Intelligence, sektor TI terus mendorong inovasi untuk menemukan solusi bahkan untuk masalah yang paling kompleks sekalipun.

6. Ritel

Big Data telah mengubah cara kerja di toko ritel. Selama bertahun-tahun, perusahaan ritel telah mengumpulkan sejumlah besar data dari survei demografis lokal, pemindai POS, RFID, kartu loyalitas pelanggan, inventaris toko, dan sebagainya. Sekarang, mereka mulai memanfaatkan data ini untuk menciptakan pengalaman pelanggan yang dipersonalisasi, meningkatkan penjualan, meningkatkan pendapatan, dan memberikan layanan pelanggan yang luar biasa. Mereka juga mengumpulkan data media sosial untuk memahami apa yang pelanggan katakan tentang merek mereka, layanan mereka, dan menyesuaikan desain produk dan strategi pemasaran mereka.

7. Transportasi

Big Data Analytics memiliki nilai yang sangat besar untuk industri transportasi. Di negara-negara di seluruh dunia, baik perusahaan transportasi swasta maupun yang dikelola pemerintah menggunakan teknologi *Big Data* untuk mengoptimalkan perencanaan rute, mengontrol lalu lintas, mengelola kemacetan jalan, dan meningkatkan layanan. Selain itu, layanan transportasi bahkan menggunakan *Big Data* untuk manajemen pendapatan, mendorong inovasi teknologi, meningkatkan logistik, dan tentu saja, untuk mendapatkan keunggulan di pasar.

F. Tantangan dalam *Big Data*

Meskipun *Big Data* memberi banyak kemudahan, bukan berarti tidak ada tantangan dalam menggunakannya. Salah satu kesalahan yang sering terjadi adalah perusahaan atau organisasi hanya menggunakan data tanpa melibatkan manusia. Padahal untuk bisa menggunakan data tersebut dengan baik dan benar, perlu campur tangan manusia untuk mempelajarinya. Oleh karenanya kemudian memunculkan beberapa karir yang berkaitan dengan *Big Data*, data science, dan data analytics.

Big Data memiliki jumlah yang besar, meskipun sudah ada tools yang bisa membantu untuk menyimpan data-data ini, volume data terus meningkat setiap tahunnya, sehingga masih banyak organisasi yang kesulitan menampung data-datanya. Tidak hanya untuk ditampung, data-data ini juga perlu diolah. Untuk

mengolah data agar data-data yang dimiliki sudah bersih atau sesuai dengan kebutuhan klien dan organisasi membutuhkan banyak pekerjaan. Data scientist menghabiskan 50-80% waktu mereka untuk menganalisa dan mempersiapkan data agar bisa digunakan.

Akhirnya, teknologi *Big Data* terus berubah dengan sangat cepat. Beberapa tahun yang lalu, Apache Hadoop adalah teknologi populer yang digunakan untuk menangani data besar. Kemudian Apache Spark diperkenalkan pada tahun 2014. Saat ini, kombinasi dari dua software ini tampaknya merupakan pendekatan terbaik. Mengikuti teknologi *Big Data* masih menjadi suatu tantangan berkelanjutan baik untuk individual maupun organisasi.

G. Rangkuman

Data merupakan hal yang penting dalam kehidupan sehari-hari, terutama bagi kalangan pebisnis, peneliti, lembaga pemerintahan, lembaga survei, dan yang pasti bagi data scientist. Saat ini pun data sangat erat kaitannya dengan dunia digital (teknologi informasi). Semua peralatan teknologi yang kita gunakan dapat memproduksi dan membaca data sehingga pada akhirnya dapat membentuk *Big Data*. *Big Data* di era digital ini menjadi sangat penting dan banyak dibicarakan. Definisi *Big Data* sering dikaitkan dengan data yang berjumlah besar dan teknologi komputer. Hal tersebut dikarenakan perangkat konvensional tidak akan mampu mengelola data tersebut secara efektif. Terdapat beberapa karakteristik *Big Data* yang membedakan dengan kumpulan data-data biasa. Karakteristik ini biasa disebut dengan istilah 3V, meliputi: volume, velocity, dan variety. Berdasarkan jenisnya, *Big Data* terbagi atas: structured, unstructured, dan semi structured. Saat ini telah banyak bidang yang menganalisis *Big Data* dan memanfaatkannya untuk meningkatkan performa kinerja dan keuntungan bisnis.

H. Referensi

- Holmes, D. E. (2017). *Big Data: A Very Short Introduction*. Oxford University Press. Retrieved from <https://books.google.co.id/books?id=NXw7DwAAQBAJ>
- Kamal, R., & Saxena, P. (2019). *BIG DATA ANALYTICS: Introduction to Hadoop, Spark, and Machine-Learning*. McGraw-Hill Education. Retrieved from <https://books.google.co.id/books?id=jN-gDwAAQBAJ>

I. Latihan

1. Apa yang dimaksud dengan istilah 'Big Data'?
 - a) Sejumlah data disimpan dalam basis data relasional.
 - b) Kumpulan data dan informasi yang besar dan beragam
 - c) Koneksi internet data kecepatan tak terbatas
 - d) Kumpulan informasi yang kecil dan tidak beragam
 - e) Media sosial yang beragam
2. Big Data penting karena semua hal berikut, kecuali ...
 - a) Dapat memprediksi hasil secara akurat, sehingga memungkinkan untuk membuat keputusan yang lebih baik
 - b) Dapat membantu dalam mengidentifikasi tren data
 - c) Memungkinkan untuk menentukan berapa banyak data yang dapat disimpan database
 - d) Dapat digunakan untuk membuat koneksi dalam data
 - e) Membantu memecahkan pertanyaan yang mungkin dimiliki perusahaan
3. Apa yang dimaksud dengan 3 v's yang merupakan karakteristik Big Data?
 - a) *volume, velocity, variety*
 - b) *variety, various, volume*
 - c) *volume, verification, vigorous*
 - d) *verification, valuable, volatile*
 - e) *velocity, variety, volatile*
4. Ada beberapa jenis atau tipe big data, berikut ini manakah yang termasuk jenis big data ...
 - a) Structured
 - b) Unstructure
 - c) Semi-structured
 - d) Analytics
 - e) A, B, dan C
5. Tinggi, usia, kecepatan, panjang adalah contoh dari jenis data berikut ...
 - a) *Exploratory*
 - b) *Explanatory*
 - c) *Qualitative*
 - d) *Quantitative*
 - e) *Subjective*

6. Manakah dari berikut ini yang biasanya membuat kumpulan data yang besar?
- a) Search Engines
 - b) Situs Social Media
 - c) Rumah Sakit
 - d) Bank
 - e) Semua Benar
7. Apa tujuan Amazon / Lazada / Shopee menggunakan Big Data?
- a) Untuk menyimpan detail pelanggan
 - b) Untuk menganalisis tren pembelian dan merekomendasikan pembelian di masa mendatang
 - c) Untuk menyimpan detail pembayaran
 - d) Untuk memprediksi pendapatan perusahaan di masa depan
 - e) Untuk memperbarui data produk penjualan
8. Proses pemeriksaan untuk data set yang sangat besar dan beragam adalah _____.
- a) *Big Data Analytics*
 - b) *Machine Learning*
 - c) *Data warehousing*
 - d) *Cloud Computing*
 - e) *Internet of Things*
9. Apa salah satu karakteristik utama pemrosesan Big Data
- a) Pemrosesan data terpusat
 - b) Penyimpanan data terpusat di Gudang Data
 - c) Pengumpulan dan komputasi data yang didistribusikan
 - d) Pengumpulan informasi melalui sistem komputer
 - e) Berbagi data di semua perangkat
10. Berikut ini adalah tantangan utama untuk Big Data, kecuali ...
- a) *Complex workloads*
 - b) *Prolonged training time*
 - c) *High Intelligence*
 - d) *Data deluge*
 - e) *Data Volume*