

Tugas 1 Big Data Semester 113

Kelompok 8

Muhammad Aziz Nurhikmah 1313617014

Dwiki Sulthon Saputra Marbi 1313617015

Muhammad Lutfi Albiansyah 1313617033

SOAL

1. Jelaskan pengertian dan keberadaan (ada atau tidak adanya) dari Big Data, dari sudut pandang spesifikasi hardware:
 - a. Jika belum memenuhi kebutuhan Big Data tersebut.
 - b. Jika sudah memenuhi kebutuhan Big Data tersebut.
2. Sebutkan permasalahan apa saja yang sering muncul pada Big Data!
3. Jelaskan apa yang dimaksud dengan Volume, Velocity, Variety, dan Veracity dalam Big Data!
4. Jelaskan perbedaan Analisis dan Analitik (analytics)!
5. Apa pendapat anda antara Big Data vs Big Information vs Big Knowledge, manakah diantara ke-3 hal tersebut yang lebih utama?
6. Berikan deskripsi sekaligus contoh, terkait data terstruktur, semi terstruktur dan tidak terstruktur!

JAWABAN

1. Pengertian big data adalah istilah umum yang banyak digunakan untuk menjelaskan segala kumpulan / kelompok data, baik data terstruktur maupun data yang tidak terstruktur, dalam jumlah sangat besar dan kompleks sehingga untuk penanganannya lebih sulit bila hanya menggunakan manajemen data biasa atau aplikasi pemroses data yang kurang mutakhir. Big data digunakan dalam banyak bisnis, yang digunakan untuk analisis wawasan yang mengarah untuk keperluan pengambilan keputusan dan penyempurnaan strategi bisnis.

Suatu hardware dikatakan belum memenuhi kebutuhan Big Data apabila belum memenuhi kebutuhan big data seperti penyimpanan yang besar, kompleks maupun lain sebagainya.

Hardware yang sudah memenuhi kebutuhan Big Data seperti memiliki penyimpanan yang besar, kecepatan yang super cepat, dan bias mengolah variasi

data yang beragam. Singkatnya Ia sudah memenuhi 4V (Volume, Velocity, Variety, Veracity)

2. Sejumlah kecil data dapat dengan mudah dikelola dan analisis langsung dapat diperoleh darinya. Lagi pula, hanya ada begitu banyak data untuk dipertimbangkan. Tetapi ketika data menjadi besar, masalah besar bisa muncul.

- a. Dimana menaruhnya

Data dalam jumlah kecil dan menengah pun bisa sulit dikelola, baik secara teknis dalam hal cara menyimpannya maupun dalam hal menganalisisnya. Jadi, semakin banyak perusahaan data memiliki masalah yang lebih kompleks dalam pengelolaannya. Apakah Anda membeli perangkat keras? Apakah Anda menyimpannya di cloud? Seberapa sering Anda perlu mengaksesnya? Bisakah Anda mengatasi latensi? Ini adalah jenis pertanyaan yang akan membantu Anda memutuskan cara mengelola data Anda.

- b. Bias besar

Satu masalah dengan banyak data adalah hal itu dapat menimbulkan bias. Misalkan Anda memiliki dua jajak pendapat, akan sangat mudah untuk memahami apa yang dikatakan jajak pendapat tersebut. Sekarang jika Anda menganalisis 100 survei, mungkin ada masalah yang jauh lebih bernuansa dalam data tersebut. Semakin banyak data yang Anda miliki, semakin banyak ruang gerak yang bisa digunakan untuk mempengaruhi statistik.

- c. Kompleksitas besar

Semakin banyak data yang Anda miliki, terkadang semakin sulit untuk menemukan nilai sebenarnya dari data tersebut.

- d. Bukan data yang saya cari

Bayangkan Google Maps memberi Anda petunjuk arah dan menyarankan rute alternatif yang "lebih cepat". Anda mengambilnya hanya untuk mengetahui bahwa itu adalah jalan tanah yang sedang dibangun. Terkadang sistem data besar mengira mereka telah menemukan jalan pintas, tetapi pada kenyataannya, itu tidak persis seperti yang dicari pengguna.

3. **Volume**

Jumlah data yang dihasilkan itu dalam ukuran petabyte (10^{15}) ke zettabyte (10^{21}). Data tersebut sangat besar dan jika volume data itu tidak sebanyak ukuran diatas, data tersebut belum digolongkan kedalam big data.

Velocity

Velocity berkaitan dengan kecepatan aliran data dari sumber seperti proses bisnis, mesin, jaringan, dan interaksi manusia dengan hal-hal seperti situs media sosial, perangkat seluler, dll. Aliran data sangat besar dan berkelanjutan. Data waktu nyata ini dapat membantu peneliti dan bisnis membuat keputusan berharga yang

memberikan keunggulan kompetitif strategis dan ROI jika Anda mampu menanganinya kecepatannya.

Variety

Variety merupakan kumpulan dari berbagai macam data, baik data yang terstruktur, semi terstruktur maupun tidak terstruktur. Variasi mengacu pada banyak sumber dan jenis data baik yang terstruktur maupun yang tidak terstruktur. Biasanya tersimpan data dari sumber seperti spreadsheet dan database. Sekarang data datang dalam bentuk email, foto, video, perangkat pemantauan, PDF, audio, dll. Berbagai data tidak terstruktur ini menimbulkan masalah untuk penyimpanan, penambangan, dan analisis data.

Veracity

Veracity mengacu pada bias, noise, dan ketidaknormalan dalam data. Apakah data yang disimpan, dan ditambang berarti untuk masalah yang dianalisis. Inderpal merasa kebenaran dalam analisis data adalah tantangan terbesar jika dibandingkan dengan hal-hal seperti volume dan kecepatan. Dalam menentukan strategi big data, Anda perlu meminta tim dan partner Anda bekerja untuk membantu menjaga data tetap bersih dan proses untuk menjaga agar 'data kotor' tidak terakumulasi di sistem Anda. Veracity adalah analisis yang dilakukan pada data, akan lebih bermanfaat bila akurasi data ditingkatkan. Veracity ini menyangkut kevalidan suatu data, dikarenakan data yang tersedia terkadang berantakan dan kualitas akurasi kurang terkontrol. Hal ini terjadi karena banyaknya variasi data yang dihasilkan oleh berbagai jenis dan sumber data yang berbeda.

4. Analisis

Analisis adalah proses untuk mengurai, membedakan, memilah dan evaluasi dari sesuatu (data) untuk digolongkan dan dikelompokkan kembali menurut kriteria tertentu kemudian dicari keterkaitan antar data dan ditafsir makna dalam data tersebut. Analisis adalah proses memecah topik atau substansi yang kompleks menjadi bagian-bagian yang lebih kecil untuk mendapatkan pemahaman yang lebih baik tentangnya. Teknik ini telah diterapkan dalam studi matematika dan logika sejak sebelum Aristoteles (384-322 SM), meskipun analisis sebagai konsep formal adalah perkembangan yang relatif baru.

Analitik

Analitik adalah metode atau ilmu untuk mempelajari data dengan menggunakan analisis statistik. Digunakan untuk penemuan, interpretasi, dan komunikasi pola bermakna dalam data. Ini juga memerlukan penerapan pola data menuju pengambilan keputusan yang efektif. Ini dapat bermanfaat di daerah yang kaya dengan informasi yang terekam; analitik bergantung pada aplikasi statistik, pemrograman komputer, dan penelitian operasi secara bersamaan untuk mengukur kinerja.

5. Dilihat dari pendefinisian antar Data, Informasi dan Knowledge. Data merupakan fakta yang ada pada dunia, sementara Information merupakan proses pengambilan Data dan Knowledge pada satu waktu tertentu dan dapat berubah dan bertambah dari waktu ke waktu. Sedangkan Knowledge merupakan apa yang selama ini kita ketahui dan merupakan hasil dari penggabungan antara Data dan Information.

Tak bisa dipungkiri bahwa Data merupakan hal yang paling utama kedalam 3 elemen tersebut dikarenakan Data merupakan sumber dari 2 elemen lainnya yaitu Information dan Knowledge.

6. **Data terstruktur**

Merupakan kumpulan data yang terorganisir, umumnya sudah didefinisikan tipe dan ukuran datanya. Data terstruktur mendefinisikan tipe data atribut, yang ada pada record maupun tuple. Setiap record memiliki field yang sama. Data ada dalam entitas dan entitas yang sama dikelompokkan menjadi satu menggunakan relasi atau kelas. Entitas dalam kelompok yang sama memiliki atribut yang sama pula. Deskripsi untuk semua entitas dalam skema memiliki kesamaan format.. Contoh: MySQL, MangoDB dll.

Data semi terstruktur

Merupakan kumpulan data yang semi terorganisir, umumnya bentuk data tidak sesuai dengan struktur formal data. Data semiterstruktur diatur dalam bentuk entitas. Entitas yang sama dijadikan satu kelompok, namun dapat berbeda atributnya. Contoh: JSON, XML, dll.

Data tidak terstruktur

Merupakan data yang tidak terorganisir, umumnya bentuk data tidak cocok dengan struktur baris dan kolom dari relasional database biasanya membutuhkan pemrosesan secara analitik. Data dapat berbentuk dalam tipe apapun tanpa perlu mengikuti format, aturan atau alur tertentu. Sebagai contoh data yang ditampilkan pada halaman web termasuk data tidak terstruktur dan tidak beraturan. Data tidak terstruktur biasanya ada pada teks, file, video, email, laporan (report), presentasi power point, pesan suara (voice mail), memo, dan citra.