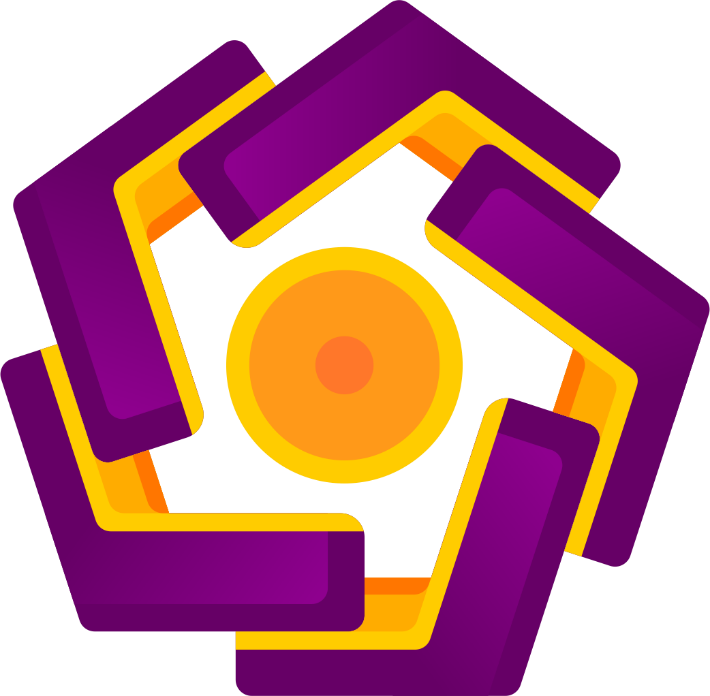
**TUGAS MATA KULIAH *BIG DATA INFRASTRUCTURE***

***CHARACTERISTICS OF BIG DATA***



|  |  |  |
| --- | --- | --- |
|  |  |  |
| Fakultas | : | Ilmu Komputer |
| Program Studi | : | Magister Teknik Informatika (MTI) PJJ |
| Mata Kuliah | : | *Big Data Infrastructure* |
| Dosen Pengampu | : | Dr. I Gede Pramudya Ananta, M.Sc. Ed.D. |
| Tutor | : | M. Syukri Mustafa, M.Kom |
| Anggota Kelompok | : | |  |  | | --- | --- | | Arnila Sandi | 21.55.2154 | | Yoga Adi Prasetyo | 21.55.2155 | | Muhamad Reza Chaedar F. | 21.55.2156 | | Luthfi Nurul Huda | 21.55.2157 | | Indra Irawanto | 21.55.2160 | | Sari Nuralita Nahrin | 21.55.2166 | |

**PROGRAM MAGISTER TEKNIK INFORMATIKA (MTI) PJJ**

**UNIVERSITAS AMIKOM YOGYAKARTA**

**TAHUN 2022**

**TUGAS II  
TASKS: Characteristics of Big Data**

**INSTRUCTIONS:**

1. **Read the tasks carefully.**
2. **Click the links that contain videos to address the tasks.**
3. **Discuss the tasks and your group responses to the tasks. Write your group initial responses to the tasks in English or Bahasa Indonesia using effective language.**
4. **Discuss your group responses with other 3 group, at least. Improve your responses based on the intergroup discussions.**
5. **Submit the responses in a assigned folder at LMS.**

***TASKS: Characteristics of Big Data***

* ***Describe****,* ***explain****,* ***elaborate****, and* ***discuss*** *on six characteristics that represent the vast dimension on big data.*

| ***Characteristics*** | ***Describe***  ***(what)*** | ***Explain***  ***(why)*** | ***Elaborate***  ***(how)*** | ***Discuss*** |
| --- | --- | --- | --- | --- |
| Volume | Volume = ukuran data dalam jumlah besar.  Volume di Big Data adalah dimensi big data yang terkait dengan ukuran dan pertumbuhan potensialnya.  Pertumbuhan dari tahun ke tahun semakin besar. | Semakin meningkatnya volume data, semakin meningkat kinerja dan biayanya.. Maka dari itu industri bisnis memerlukan strategi untuk menangani pemrosesan data skala jumlah besar agar dapat menghemat biaya dan mendapatkan keuntungan. | *More data = better safety*  Semakin banyak data yang diperoleh dan dipelajari artinya semakin baik sebuah produk yang diproduksi. Contohnya di bidang safety (keamanan) transportasi pesawat terbang, kapal layar, kereta api dll, fasilitas healthcare, tata Kelola pemerintahan dan reliability (uji ketahanan produk).  Salah satu tujuan industri bisnis adalah mengubah data dalam jumlah besar menjadi bermanfaat dan menguntungkan. | Tantangan untuk ukuran data dalam jumlah besar adalah storage (penyimpanan). Semakin besar jumlah data, semakin besar pula kapasitas yang diperlukan.  *In house Storage VS Cloud Storage.*  Tantangan :   * Biaya * Skalabilitas * Kinerja storage, akses dan pemrosesannya. |
| Velocity | Secara bahasa velocity berarti kecepatan. Dimana dalam kasus big data velocity bermakna data yang dihasilkan atau yang digunakan bisa diproses secara real-time. Yang mana data yang didapatkan dari sumber data baik seperti sensor, polling, atau yang lain dapat dikirim secara langsung ke dalam server untuk disimpan dan kemudia diproses dalam analisis dalam big data dan dikembalikan lagi kepada user sebagai informasi matang yang siap dikonsumsi. | Penerapan velocity bisa dikatakan urgent, karna pemrosesan data secara real-time ditujukan agar data yang dihasilkan sesuai dengan tingkat produksinya sejalan dengan tujuan khusus dari analitik big data. Tidak dapat dipungkiri real-time sendiri masih memiliki delay-time dimana dalam beberapa kasus data harus diproses lebih mendalam melewati koridor-koridor atau tahap-tahap pemrosesan untuk menghasilkan data yang akurat nantinya. | Dari penjelasan tentang real-time dan delay-time atau batch, dari keduanya bisa kita gabungkan. Bagaimana kedua type velocity ini bisa bekerja bersamaan dengan tujuan yang sama. Tingkat generasi, pengambilan, atau pengolahan data sebagai jawaban dari kebutuhan kita. Kebutuhan akan tindakan real-time dalam kasus bisnis misalnya adalah apa yang pada akhirnya menentukan kecepatan analitik atas big data. Terkadang ketepatan satu menit dibutuhkan. Kadang setengah hari.  Ketika ketepatan waktu dari informasi yang diproses tidak berperan dalam pengambilan keputusan, kecepatan data yang dihasilkan menjadi tidak relevan. Dengan kata lain, kita bisa menunggu selama yang dibutuhkan untuk memproses data. Hari, bulan, minggu. Dan setelah pemrosesan selesai, kita akan melihat hasilnya dan mungkin membagikannya dengan orang lain. Ketika ketepatan waktu tidak menjadi masalah, kita dapat memilih salah satu dari beberapa jalur. |  |
| Variety | Apa itu Variety?  Variety merupakan sekumpulan big data yang bervariasi . Ragam data dikelompokkan dalam structured data, semi-structured data, dan unstructured data. Contohnya ada pada media sosial yang menampilkan banyak macam data, mulai dari foto, video, formulir, hingga filter-filter unggulan. Dalam urusan bisnis, data ini juga beragam bentuknya, bisa berupa dokumen, tabel, dan sebagainya. | mengapa kita perlu membahas atau mempelajari tentang tipe dan format data terlebih dahulu ?  karna sebagai seorang data scientist harus selalu mengetahui model atau algoritma apakah sudah sesuai atau tidak dengan data yg dimiliki dan harus memiliki pemahaman terkait tipe atau format data.  jadi tujuan dari Variety ialah untuk mengenalkan berbagai tipe data atau format data yg cocok untuk model di algoritma nantinya.  Sebagai contohnya kita sedang meneliti bagaimana mengoleksi data covid dari pasien covid – 19. dimana pasien nanti diminta untuk memotret satu anggota tubuhnya yang diharapkan bisa memberikan indikator tentang penyakit tersebut dan kemudian digabungkan dengan data sensor yang dihasilkan dari smartwatch, dengan cara menggabungkan dua data tersebut diharapkan bisa mengidentifikasi lebih akurat. Dari kasus ini kita bisa melihat bahwa data yang digunakan untuk sebuah analisis tidak hanya mengandalkan data PCR, melainkan menggunakan data JPEG atau IMAGE dari kamera dan data PCR atau teks dari sebuah sensor. Dalam konteks ini kita bisa menarik sebuah kesimpulan bahwa Big Data tidak terlalu memperdulikan tentang variasinya tapi yang terpenting adalah kontennya. | Semakin banyaknya data yang bervariasi,  bagaimana semisal dalam analisis big data kita menggunakan data data yg berbeda format ataupun tipe?  Dalam konteks big data maka kita tidak lagi melihat bahwa data itu harus single format atau single type. Melainkan bisa multy type atau variety nya sangat besar dan bisa saja jenisnya bermacam-macam. |  |
| Veracity | Veracity dalam Big Data ini sebenarnya berhubungan dengan truthfulness, realiability, quality and availability of the data. Atau dalam Bahasa Indonesianya dapat diartikan bahwa data tersebut dapat dipercaya keakuratannya, keasliannya, dapat diandalkan, memiliki kualitas yang baik serta dapat diakses dengan baik. | Semakin besar data biasanya akan semakin sulit untuk di jaga tingkat keakuratannya, lalu bagaimana sebuah big data dapat memiliki keakuratan yang baik? Caranya adalah dengan melakukan kegiatan update atau pembaharuan data secara berkala. | Karena data-data yang didapat akan digunakan untuk menganalisis sehingga kita tidak mungkin mendapatkan data yang tidak akurat. Karena dapat menghasilkan Analisa yang kurang tepat atau dapat dapat dikatakan Analisa yang salah. | 1. Manfaat dari Veracity  * Melindungi data dari noise * Memperoleh data yang akurat dan terpercaya * Mendapatkan hasil Analisa yang baik * Mendapatkan data dengan kualitas yang baik  1. Kelebihan  * Kekauratan data * Kualitas data yang baik * Validitas data yang terjamin  1. Kekurangan  * Untuk memperoleh data yang akurat harus melakukan update data secara berkala. |
| Valence | Valensi, adalah arahan mengacu pada apakah informasi mendukung keyakinan atau menyangkal keyakinan (Littlejohn,2014:111). Seberapa besar informasi tersebut dapat mendukung kepercayaan yang telah dimiliki sebelumnya. Hal ini memberikan pengaruh pada kekuatan informasi untuk membentuk sikap. Jika informasi tersebut mendukung kepercayaan yang suudah ada, maka informasi tersebut dianggap sebagai informasi positif. | Valensi diartikan sebagai lebih menguntungkan atau kurang menguntungkan penilaian yang dibuat mengenai suatu produk (Larceneux, 2013). Artinya pendapat yang dianggap kurang menguntungkan akan menyebabkan suatu produk menjadi dianggap tidak baik, sementara produk yang mendapatkan ulasan yang menguntungkan akan menaikkan ekspektasi yang baik dari khalayak kepada produk tersebut (Larceneux, 2013). | mengacu pada bagaimana big data dapat terikat satu sama lain, membentuk koneksi antara kumpulan data yang berbeda  Ukuran konektivitas  Konektivitas Data: Dua item data terhubung saat mereka saling terkait  Valensi: Fraksi item data yang terhubung dari jumlah total kemungkinan koneksi  Demikian pula semakin banyak data yang terhubung berarti semakin banyak valensi yang sekali lagi akan membantu menghubungkan data yang terhubung. Kumpulan data valensi tinggi lebih padat. Hal ini membuat banyak kritik analitik reguler menjadi sangat tidak efisien.[3] http://ignited.in/I/a/200754 | Tantangan:  Algoritma eksplorasi data yang lebih kompleks  Perubahan valensi pemodelan dan prediksi  Analisis perilaku yang muncul  Banyak masalah muncul karena perilaku dinamis data karena data terus berubah seiring waktu dan volume. Jika ada lebih banyak valensi di antara data maka kompleksitas analisis akan meningkat. |
| Value | *Value* berarti *big data* memiliki nilai yang sangat tinggi apabila diolah dengan cara yang tepat guna atau dapat juga dikatakan seberapa bernilainya atau bermaknanya suatu data. | Data yang semakin besar dan semakin banyak pasti ada data yang bernilai sangat penting untuk pemrosesan data pada sebuah system. | Sebagai contoh, biodata karyawan suatu perusahaan penjualan bahan baku makanan tidak akan bernilai untuk kepentingan analisis prediksi penjualan bahan baku ke customer. Data tersebut mungkin tidak penting dan tidak bernilai untuk satu hal, namun bisa sangat penting dan sangat bernilai untuk hal lain. Data yang tidak memiliki nilai di bagian mana pun tidak akan terfilter di sistem aplikasi analisis Big data. | Data yang sangat penting memiliki beberapa resiko terhadap keamanan data yang mungkin banyak orang yang menginginkannya. |