Esai

1. Apa yang dimaksud dengan:
   1. Descriptive Analytics

Suatu analisis untuk menjelaskan peristiwa peristiwa yang telah terjadi. Menanyakan detail dari peristiwa yang sebelumnya. Maka dari situ kita bisa mendapatkan informasi dari data yang telah dikumpulkan. Misalnya: Berapa banyak orang yang datang ke mall A pada hari weekends? Berapa banyak orang yang datang ke mall A pada hari weekdays? Kira kira dari keseluruhan hasil analitik, 80% berupa hasil analitik deskriptif. Secara umum, laporan dari hasil analisis deskriptif ini disjikan dalam bentuk data data historis grid data maupun grafik dan juga statis.

* 1. Diagnostic Analytics

Diagnostic Analytics adalah analisis yang 1 level lebih tinggi dari Descriptive Analytics. Hasil analisis ini menghasilkan value/nilai yang lebih tinggi daripada Descriptive Analytics, namun keahliannya pun dibutuhkan lebih tinggi juga. Analisis ini menganalisa penyebab dari terjadinya suatu peristiwa yang telah terjadi. Sehingga biasanya pertanyaannya mengacu kepada “mengapa”, agar bisa menemukan informasi informasi yang berhubungan dengan kenapa peristiwa tersebut terjadi. Misalnya, Mengapa konsumen Frozen Yoghurt meningkat tiba tiba di bulan Juni? Hasil analitik diagnostic yang diambil dari berbagai sumber kemudian disimpan dalam struktur yang sesuai seperti Analisis drill down yang menyimpan data ringkasan informasi rinci yang meng-highlight hal tertentu, dan juga analisis roll up yang sebaliknya, mendapaatkan ringkasan.

* 1. Predictive Analytics

Predictive Analytics adalah upaya untuk memprediksi hasil peristiwa yang dapat terjadi nantinya, melalui diagnostic analytics, hasil data penyebab dari suatu outcome dan peristiwa yang telah terjadi ditemukan, pola, tren, dan lonjakan data yang tidak normal, jadi dengan penyebab yang mungkin terjadi kedepannya, kita bisa prediksi apa yang akan terjadi di masa depan. Namun ketika kondisi penyebabnya berubah, maka model yang memprediksi peristiwa kedepannya kemudian harus diperbaharui. Maka pertanyaan yang dapat ditanyakan itu seperti “Jika suhu meningkat ke 20 derajat pada pertengahan Juni, berapa peluang hujan terjadi?”. Analisis ini memberikan nilai hasil analisis yang lebih besar daripada descriptive dan diagnostic, namun juga membutuhkan keahlian yang lebih tinggi pula.

1. Karakteristik Big Data

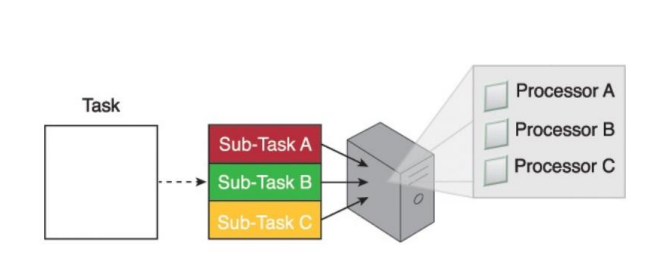
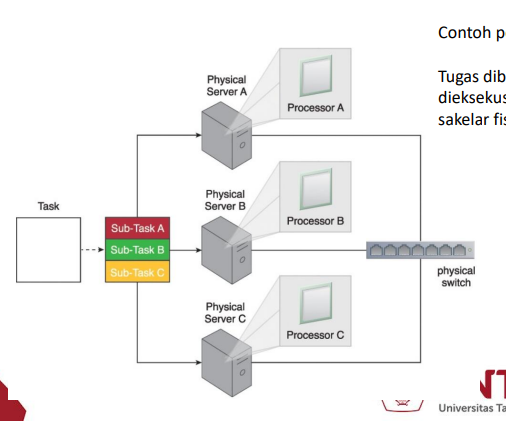
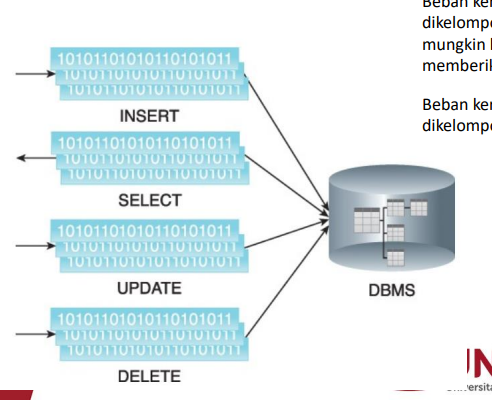
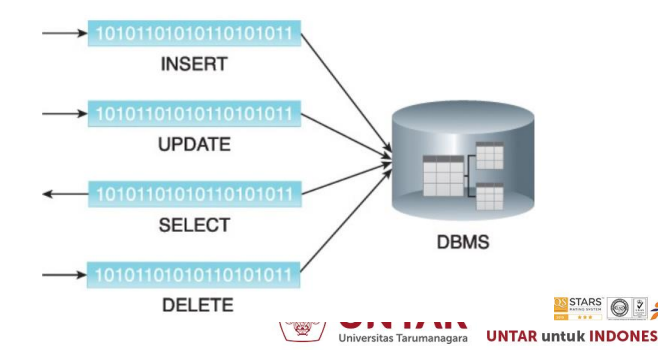
Volume: Besar kecilnya volume data mempengaruhi cara memenuhi permintaan penyimpanan dan cara pemrosesan data, proses persiapan, kurasi, dan juga cara pengelolaan data tambahan yang berbeda beda. Banyak sekali hal hal yang memproduksi banyak sekali data sehingga tinggi sekali volume data, seperti e-commerce, sensor otomatis, dan juga social media.

Velocity: Data dapat terproses dengan waktu yang sangat singkat, berarti dengan kecepatan yang sangat tinggi. Dari sisi korporat dan perusahaan, velocity berarti jumlah waktu yang dipakai dalam meproses

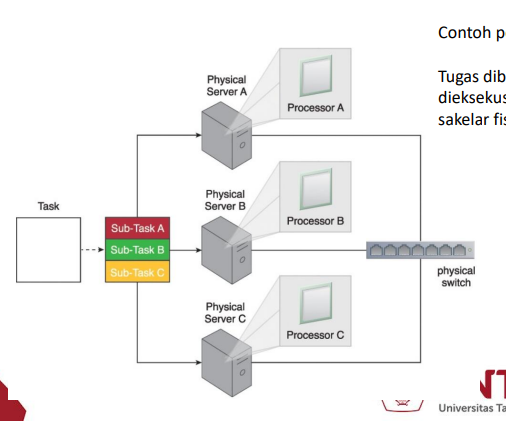
Variety

Veracity

Value

* Pemrosesan Data Paralel •
  + Pemrosesan data paralel melibatkan eksekusi simultan dari beberapa sub-tugas yang secara kolektif terdiri dari tugas yang lebih besar. •
  + Tujuannya adalah untuk mengurangi waktu eksekusi dengan membagi satu tugas yang lebih besar menjadi beberapa tugas yang lebih kecil yang berjalan secara bersamaan. •
  + Meskipun pemrosesan data paralel dapat dicapai melalui beberapa mesin jaringan, hal ini lebih sering dicapai dalam batas-batas satu mesin dengan beberapa prosesor atau cores.
  + 
  + Sebuah tugas dapat dibagi menjadi tiga sub-tugas yang dieksekusi secara paralel pada tiga prosesor berbeda dalam mesin yang sama
  + Pemrosesan data terdistribusi terkait erat dengan pemrosesan data paralel di mana prinsip yang sama "divide-and-conquer" diterapkan. •
  + Pemrosesan data terdistribusi selalu dicapai melalui mesin yang terpisah secara fisik yang terhubung ke jaringan bersama sebagai sebuah cluster.
  + 
  + Contoh pemrosesan data terdistribusi
  + Tugas dibagi menjadi tiga sub-tugas yang kemudian dieksekusi pada tiga mesin berbeda yang berbagi satu sakelar fisik.
* Memproses Beban Kerja (Workloads) •
  + Pemrosesan beban kerja (workloads) dalam Big Data didefinisikan sebagai jumlah dan sifat data yang diproses dalam jangka waktu tertentu. •
  + Beban kerja (workloads) dibagi menjadi dua jenis: •
    - Batch •
    - transactional
  + Batch •
    - Pemrosesan batch dikenal sebagai pemrosesan offline, melibatkan pemrosesan data dalam batch dan biasanya menimbulkan penundaan, yang menghasilkan respons latensi tinggi. •
    - Beban kerja batch biasanya melibatkan data dalam jumlah besar dengan baca/tulis berurutan dan terdiri dari grup kueri baca atau tulis. •
    - Kueri bisa rumit dan melibatkan banyak operasi join. •
    - Sistem OLAP memproses beban kerja dalam batch. •
    - BI dan analitik strategis berorientasi pada batch karena merupakan tugas yang sangat intensif membaca yang melibatkan volume data yang besar.
    - 
    - Beban kerja batch terdiri dari baca /tulis yang dikelompokkan yang memiliki jejak data besar dan mungkin berisi operasi join yang kompleks dan memberikan respons latensi tinggi .
    - Beban kerja batch dapat mencakup baca /tulis yang dikelompokkan sebagai INSERT, SELECT, UPDATE, dan DELETE.
  + Transaksional •
    - Pemrosesan transaksional juga dikenal sebagai pemrosesan online. •
    - Pemrosesan beban kerja transaksional mengikuti pendekatan di mana data diproses secara interaktif tanpa penundaan, menghasilkan respons latensi rendah. •
    - Beban kerja transaksi melibatkan sejumlah kecil data dengan pembacaan dan penulisan acak. •
    - OLTP dan sistem operasional, yang umumnya intensif menulis, termasuk dalam kategori ini. •
    - Meskipun beban kerja ini berisi campuran kueri baca/tulis, umumnya lebih intensif menulis daripada membaca. •
    - Beban kerja transaksional terdiri dari pembacaan/penulisan acak yang melibatkan lebih sedikit operasi join daripada intelijen bisnis dan beban kerja pelaporan.
    - 

Beban kerja transaksional memiliki sedikit operasi join dan respons latensi lebih rendah daripada beban kerja batch

* + Pemrosesan data terdistribusi terkait erat dengan pemrosesan data paralel di mana prinsip yang sama "divide-and-conquer" diterapkan. •
  + Pemrosesan data terdistribusi selalu dicapai melalui mesin yang terpisah secara fisik yang terhubung ke jaringan bersama sebagai sebuah cluster.
  + 
  + Contoh pemrosesan data terdistribusi
  + Tugas dibagi menjadi tiga sub-tugas yang kemudian dieksekusi pada tiga mesin berbeda yang berbagi satu sakelar fisik.