

Big Data Analytics



Big Data
STF1724

Rini Nuraini Sukmana, M.T
0020087901 - 08882024236
rini.nuraini@usbypkp.ac.id



Definisi

- *Big data analytics* mengacu pada proses mengumpulkan, mengorganisasikan dan menganalisa sekumpulan besar data (*big data*) untuk mendapatkan pola-pola dan informasi yang berguna.
- *Big data analytics* tidak hanya membantu untuk memahami informasi yang terkandung di dalam data tapi juga membantu untuk mengidentifikasi data yang paling penting untuk keputusan bisnis saat ini dan masa datang.



- *Big data analytics* pada dasarnya ingin untuk menghasilkan pengetahuan (*knowledge*) dari hasil analisis data.



- Perusahaan tertarik memanfaatkan nilai yang ada di dalam informasi, seperti halnya memanfaatkan sistem *big data* yang dapat menangani data mulai dari *event data* yang berjumlah besar sampai ke *data text* yang berasal dari media sosial.
- Informasi ini sebagian besar belum dimanfaatkan sistem *business intelligence* yang ada saat ini, tapi organisasi mulai mengenali terdapat nilai lebih dari perluasan dan pemanfaatan secara maksimal lingkungan *business intelligence* dan *data warehousing* untuk mengintegrasikan, mengelola dan menganalisa informasi.



Manfaat Big Data Analytics

- Perusahaan terus berusaha mencari dan menemukan *actionable insight* dari data yang mereka miliki.
- Pada beberapa proyek pembangunan big data berasal dari kebutuhan untuk menjawab pertanyaan-pertanyaan bisnis yang bersifat khusus, seperti “Apa yang sebenarnya pelanggan pikirkan mengenai produk dan brand yang perusahaan hasilkan?, Apakah mereka berpendapat positif atau negatif mengenai produk dan brand tersebut?”, “Bagaimana perusahaan dapat meningkatkan sales intelligence?”.
- Big data analytics platform yang tepat mendukung perusahaan dalam meningkatkan penjualan, meningkatkan efisiensi dan memperbaiki operasional perusahaan, pelayanan kepada pelanggan dan meningkatkan kemampuan menangani manajemen risiko.



Tantangan dalam Mengembangkan Big Data Analytics



- Pada hampir semua organisasi, mengembangkan *big data analytics* merupakan suatu tantangan besar.
- Tantangan ini dikarenakan oleh jumlah data yang besar, format data yang beragam (baik pada *structured* atau *unstructured data*) dan berasal dari seluruh bagian yang ada di organisasi serta beragam cara yang digunakan untuk mengkombinasikan, membangun dan menganalisa data untuk mendapatkan pola (*pattern*) tersembunyi dari data dan mendapatkan informasi yang berguna bagi pengguna.
- Tantangan pertama dalam mengembangkan *big data analytics* adalah Bagaimana memecah data untuk dapat memungkinkan mengakses semua data organisasi yang disimpan di tempat penyimpanan yang berbeda dan bahkan juga disimpan pada sistem yang berbeda.
- Tantangan besar kedua adalah Membuat *platform* yang dapat menarik *unstructured data* semudah menarik *structured data*. Volume data ini begitu besar sehingga sulit untuk memprosesnya menggunakan *database* dan metode perancangan *software* yang tradisional.



Enterprise data warehouse

- *Enterprise data warehouse* dapat digunakan untuk menangani *big data* dan beban kerja yang sangat besar, tapi kadang lebih efisien untuk memproses data sebelum ditarik dan disimpan (*loading*) ke dalam warehouse.
- Data yang berasal dari perangkat/alat sensor, contohnya memiliki nilai guna untuk bisnis jika dapat disaring dan diagregasi sebelum digunakan untuk keperluan analisis. Data mentah (*raw data*) biasanya tidak diperlukan untuk proses yang bersifat historis.
- Biaya penyimpanan dan pemrosesan data dapat dikurangi jika data mentah tersebut tidak disimpan di dalam *warehouse*.



Supply Chain



- Cara terbaik untuk menangani *big data* dan proses-proses bisnis yang melibatkan *big data* adalah mengoptimalkan penggunaan solusi *hardware* dan *software* untuk memproses beban kerja *big data* dan kemudian mengkombinasikan solusi tersebut dengan *data warehouse* perusahaan yang telah ada untuk membuat dan mengintegrasikan *information supply chain*.
- Tujuan dari *information supply chain* adalah menggunakan dan mengintegrasikan beberapa variasi dari data mentah yang ada di dalam prganisasi, kemudian menganalisis data tersebut dan kemudian menghasilkan dan menyampaikan hasil analisis ke pengguna (pengguna dalam kegiatan bisnis).



Information Supply Chain



- Dalam beberapa tahun terakhir telah terjadi kecenderungan yang sama terhadap proses analisis dalam mencapai kinerja dari sistem-sistem yang bersifat umum.
- Saat ini, *multi-terabyte data warehouse* tidak lagi memenuhi beban kerja dari proses analisis yang makin kompleks.
- Hasilnya adalah sekali lagi organisasi membutuhkan optimasi sistem untuk memenuhi tantangan beban kerja yang makin berat. Respon dari pihak industri telah menawarkan solusi *hardware* dan *software* yang keduanya terkait dengan optimasi proses analisis.



MENGELOLA DAN MENGANALISIS BIG DATA

- Tantangan utama dalam menangani *big data* dan beban kerjanya yang besar adalah, keragaman dan volume data, kompleksitas beban kerja analitis dan beban kerja.
- Hal ini dapat dilihat pada *information supply chain*, input pada *information supply chain* terdiri atas data mentah yang dibutuhkan untuk analisis.
- Selama dua dekade terakhir hampir semua *business analytics* telah dibuat menggunakan *structured data* yang diekstraksi dari sistem operasional dan dikonsolidasikan ke dalam suatu *data warehouse*.
- *Big data* meningkat secara drastis dari segi jumlah data mentah dan keragaman serta volume data yang diperlukan untuk analisis.
- Prosentase tertinggi berupa *multi-structured data* (untuk membedakannya dengan *structured data* pada data operasional yang digunakan untuk dimasukkan pada *data warehouse*).
- Pada sebagian besar organisasi, *multi-structured data* terus bertambah jumlahnya dibandingkan dengan *structured data*.



Teknik Utama Menganalisis Big Data



Terdapat dua teknik utama untuk menganalisis big data :

1. *the store and analyze approach*, dan
2. *the analyze and store approach*.



The Store and Analyze Approach



- *The store and analyze approach*, mengintegrasikan sumber data ke dalam *consolidated data store* sebelum data dianalisa.
- Pendekatan ini digunakan oleh sistem *data warehouse* tradisional untuk menghasilkan *data analytics*.
- Pada *data warehousing system*, *consolidated data store* biasanya berupa *enterprise data warehouse* atau *data mart* yang dikelola oleh *relational DBMS* atau *multidimensional DBMS*.



Dua trend big data yang penting untuk mendukung *Store and analyze approach* adalah :

- relational DBMS yang dioptimalkan untuk menganalisis data (sering disebut juga dengan Analytic RDBMS atau ARDBMS)
- non-relational system (kadang disebut dengan NoSQL systems) yang digunakan untuk memproses *multi-structured data*.

Non-relational system dapat digunakan untuk menghasilkan suatu analisis dari *big data*, atau untuk pra-proses *big data* sebelum dikonsolidasikan ke dalam *data warehouse*.



Keuntungan The Store and Analyze Approach

Meningkatkan integrasi data dan meningkatkan kemampuan mengelola kualitas data serta kemampuan untuk mengelola historical information.

Kerugian The Store and Analyze Approach

Dibutuhkan data storage tambahan dan meningkatkan latency data (waktu yang dibutuhkan untuk menyimpan dan mengirim paket data).



The Analyze and Store Approach

- *Analyze and Store Approach* menganalisis data yang mengalir pada proses bisnis, yang ada di antara *network* dan *systems*.
- Hasil analisis kemudian dapat dipublikasikan pada *interactive dashboards* dan/atau dipublikasikan ke dalam *data store* (seperti *data warehouse*) untuk diakses pengguna, menghasilkan *historical reporting* dan analisis tambahan lainnya.
- Pendekatan ini dapat juga digunakan untuk menyaring dan mengagregasi *big data* sebelum dimasukkan ke dalam *data warehouse*.