

LAPORAN
RENCANA TUGAS MANDIRI (RTM) Ke-1
MATA KULIAH BIG DATA KELAS B
**“MERINGKAS MATERI TENTANG KONSEP BIG DATA &
ANALITIK DATA”**



DISUSUN OLEH:

Fadlila Agustina (21083010050)

DOSEN PENGAMPU:

Tresna Maulana Fahrudin S.ST., M.T. (NIP. 199305012022031007)

PROGRAM STUDI SAINS DATA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAWA TIMUR
2023

ISI RINGKASAN

Data merupakan *resource* atau sumber dari suatu informasi. Penggunaan data dalam suatu informasi harus bersifat akurat agar informasi yang didapatkan menjadi acuan dari suatu output informasi yang diinginkan. Data-data yang sangat besar, kompleks, dan selalu bertambah tiap waktunya merupakan definisi dari Big Data. Big data juga dapat diartikan sebagai kumpulan data, seperti identitas, proses, sifat dan lain-lain yang terdiri dari beberapa aspek dalam jumlah yang besar yang terstruktur maupun tidak terstruktur.

Big data mempunyai 9 karakteristik. Yang pertama **Volume**, dapat dikatakan seperti karakter utama pada big data karena karakteristik big data dapat dilihat dari ukuran dan jumlah data secara keseluruhan maupun berdasar pada platform yang mengelolanya. Selain **volume**, karakteristik big data juga dapat dilihat dari **Velocity** atau kecepatan data dalam artian seberapa cepat data yang dihasilkan, diproses, dan dianalisis untuk memenuhi kebutuhan. **Variety** merupakan karakter big data yang mempunyai sekumpulan big data yang bervariasi di setiap platform yang dibagi menjadi *structured data*, *semi-structured data*, dan *unstructured data*. **Value** adalah karakteristik big data yang dapat dilihat dari nilai atau manfaat *user*. **Visualization** pada big data dapat dilihat dari bagaimana data tersebut divisualisasikan (grafik). **Volatility** atau perubahan data yang terkumpul dan berdampak pada homogenitas data. **Validity** pada big data dapat dilihat dari validnya data yang terkumpul sebelum digunakan. **Vulnerability** atau keamanan dari big data yang harus terlindungi. **Variability** pada big data dapat dilihat dari perubahan data yang berdampak pada karakteristik yang lain.

Siklus aliran dari proses *source* sampai ke aplikasi merupakan pengertian dari arsitektur big data yang dapat memudahkan *user* melakukan transaksi dan AHP (Analytical Hierarchy Process). Arsitektur big data memiliki beberapa komponen, yaitu dari **Data Source** atau sumber data yang umumnya dipompa masuk big data menggunakan operasional file system seperti transfer file. Selanjutnya **Data Aggregator** merupakan *software* yang mengumpulkan dan menyalurkan data dari sumber ke beberapa jenis pengolahan. **Real Time Streaming Processor** merupakan metode untuk menganalisis data yang bersifat *real time* dan *streaming*. **Hadoop** merupakan *software* yang digunakan untuk menyimpan dan menangani big data dengan efisien tanpa mengeluarkan dana yang banyak. **Non Real Time Processor** atau pemrosesan untuk data besar yang terdapat pada Hadoop Distributed File System. **Data Store** merupakan tools untuk menyimpan data hasil proses *real time* dan *on real time*. Pada komponen terakhir, yaitu **Application** atau aplikasi yang berhubungan langsung dengan *user* yang dapat mengakses data yang berada pada data store.

Data analitik mempunyai 4 tipe, yaitu yang pertama ada **Descriptive Analytics** atau proses untuk mendapatkan gambaran umum dari beberapa data yang terkumpul. Adapun tipe **Diagnostic Analytics** yang membantu analisis deskriptif untuk menggali lebih dalam agar menemukan jawabannya. Selanjutnya ada **Predictive Analytics** yang memberikan hasil prediksi sesuatu yang akan datang dengan penambahan daya komputasi untuk menjalankan banyak model dengan cepat dan adopsi teknik prediktif. **Prescriptive Analytics** adalah proses penghasilan jawaban atas masalah kenapa sesuatu akan terjadi serta memberi saran terhadap sesuatu yang mendatang.

Pada mode pemrosesan analitik, terdapat beberapa langkah, yaitu **Data Selection** atau proses pemilihan data yang akan digunakan sesuai kebutuhan. **Preprocessing Data**, yaitu tahapan untuk menghilangkan beberapa masalah yang mengganggu saat pemrosesan data. **Modelling Data** adalah proses menciptakan representasi visual untuk menentukan

pengumpulan informasi. **Visualization Data** merupakan rangkaian proses yang akan dilakukan untuk menampilkan data dalam bentuk grafik, dan lain sebagainya. **Interpretation Data** merupakan serangkaian proses dari meninjau data melalui beberapa proses yang sudah ditentukan. **Action** digunakan untuk mengirim data saat sudah menekan tombol tertentu

Analitik big data dapat diimplementasikan dalam perkembangan bisnis, contohnya seperti membantu manajemen dan staff untuk membuat keputusan lebih baik dan meningkatkan kemampuan analitik. Selain membantu mengambil keputusan, analitik big data juga dapat memetakan profil pelamar kerja dengan cara menggabungkan beberapa data dari social media, profil perusahaan, dan database pencarian sehingga memudahkan proses filtering CV dan lebih efisien. Selain itu, analitik big data dibutuhkan oleh perbankan dan sekuritas untuk memantau pasar keuangan melalui monitor aktivitas jaringan pengolah *natural language* untuk mengurangi penipuan.

Terdapat tools yang digunakan terkait dengan penggunaan basis data skala besar :

- **Hadoop**, digunakan untuk meningkatkan dari server tunggal ke ribuan mesin.
- **Apache Hbase**, yaitu database NoSQL sumber terbuka yang dibuat di Apache Hadoop dan dimodelkan setelah Google BigTable.
- **MapReduce**, yaitu model pemrograman yang digunakan untuk memproses data berukuran besar secara terdistribusi dan paralel.
- **Pig**, yaitu tools yang digunakan untuk menghasilkan kode MapReduce untuk menganalisis dataset yang besar.
- **Hive**, yaitu tools yang digunakan untuk menyiapkan data sebelum memuatnya ke tujuan data.
- **Apache Avro**, yaitu system serialisasi data yang umum digunakan di dunia streaming.
- **Zookeeper**, yaitu proyek *software* dari Apache yang menyediakan konfigurasi terpusat untuk system terdistribusi besar.
- **Ambari**, yaitu tools untuk memudahkan melakukan provisioning, manajemen, dan monitoring terhadap Apache Hadoop clusters.
- **Apache Spark**, yaitu tools yang digunakan untuk membangun jalur pemrosesan data dengan mudah.
- **Apache Sqoop**, yaitu tools yang digunakan untuk mentransfer data antara cluster Hadoop dan database hubungan.
- **Oozie**, yaitu system alur kerja dan koordinasi yang mengelola pekerjaan Apache Hadoop.
- **Hcatalog**, yaitu sistem manajemen metadata dan table untuk mendukung Hadoop yang lebih luas.