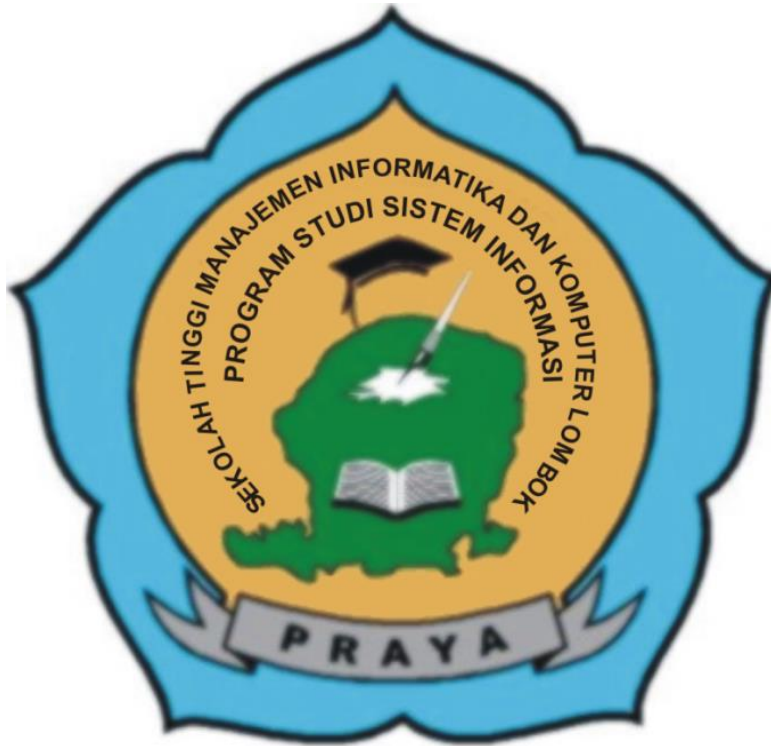


MAKALAH BIG DATA



OLEH:

RIZAL MAHENDRA (TI17200045)

JURUSAN TEKNIK INFORMATIKA (A)

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER

2022

KATA PENGANTAR

Puji syukur saya ucapkan atas kehadiran tuhan yang maha esa, dengan rahmat taufik dan hidayah-Nya saya dapat menyelesaikan makalah ini, meskipun masih banyak kekurangan yang ada didalamnya. Dan pada kesempatan ini, saya ingin menyampaikan ucapan terimakasih kepada semua pihak yang telah membantu saya menyelesaikan penyusunan makalah ini.

Terutama kepada orang tua saya, bapak/ibu dosen dan teman-teman saya. Saya berharap semoga dengan tersusunnya makalah ini yang berjudul *BIG DATA* dapat membantu pembaca terutama saya sendiri dalam rangka menambah wawasan kita tentang peranan ilmu logika dalam bidang matematika, elektronika, informatika dan dalam bidang ilmu kesehatan. Dan kami juga menyadari sepenuhnya bahwa makalah ini jauh dari kesempurnaan. Oleh karna itu, saya sangat mengharapkan adanya kritik dan saran yang bersifat membangun, demi perbaikan makalah yang akan saya buat dikesempatan selanjutnya.

Praya, 17 september 2022

Hormat penulis

DAFTAR ISI

KATA PENGANTAR	2
BAB I.....	4
PENDAHULUAN.....	4
A. Latar belakang.....	4
B. Rumusan masalah.....	4
C. Tujuan	4
BAB II	5
PEMBAHASAN	5
A. Pengertian Big Data.....	5
B. Karakteristik Big Data	6
C. Teknologi Big Data	6
D. Hasil Pengolahan Big Data.....	8
BAB III.....	10
PENUTUP	10
A. KESIMPULAN.....	10
B. SARAN.....	10

BAB I

PENDAHULUAN

A. Latar belakang

Penting nya Big Data di era modern ini. Dengan mengikuti perkembangan Big Data maka kita tidak ketinggalan informasi atau ketinggalan jaman. Dengan menggunakan Big Data ini juga apat memudahkan kita mencari informasi misalnya melalui google. Di google kita dapat mencari informasi dari yang terbaru hingga yang sudah lampau.

Misalnya dari facebook, maka kita dapat chatingan, bahkan kita dapat melihat aktivitas seseorang. Dengan menggunakan salah satu Banyak orang membutuhkan pengolahan Big Data, antara lain untuk mengetahui topik yang sedang hangat saat ini di Twitter, mencari teman lama secara cepat melalui Facebook, dan lain-lain. Perusahaan perlu mengolah Big Data untuk pengambilan keputusan bisnis yang harus cepat. Misal, untuk mengetahui kebiasaan dan kesukaan pelanggan tanpa harus bertanya, mengetahui selera pembaca portal berita di web untuk disesuaikan dengan iklan yang ditampilkan, mengatur perjalanan pesawat agar tidak delay, mengendalikan wabah penyakit, dan sebagainya.

B. Rumusan masalah

Berdasarkan latar belakang diatas, saya dapat merumuskan beberapa masalah yang akan dibahas dalam makalah ini antara lain:

1. Pengertian Big Data
2. Karakteristik Big Data
3. Teknologi Big Data
4. Hasil pengolahan dari Big data

C. Tujuan

Untuk mengetahui tentang Big Data dan metodologi-metodologi, teknologi dan hasil dari pengolahan big data tersebut. dan sebagai media pembelajaran untuk pembaca dan penulis.

BAB II

PEMBAHASAN

A. Pengertian Big Data

Big data adalah sebuah data elektronik yang sangat besar, yang mempunyai banyak macam jenis dan data elektronik ini sangat cepat berubah. Big data saat ini banyak di gunakan orang-orang untuk membantu kebutuhan. Banyak sekali data yang tersebar dan tersimpan di komputer-komputer di internet misalnya seperti teks, gambar, video, suara, animasi, blog, buku, cuaca, GPS, temperatur, dan masih banyak lagi jenis nya.

Banyak orang yang menggunakan big data ini untuk mencari informasi yang sedang populer saat ini, misal nya berita terbaru. Misalnya facebook, di facebook kita bisa menemui banyak orang, bahkan teman lama yang sama-sama menggunakan facebook.

Untuk mengolah Big Data menjadi informasi yang lebih berguna, perlu program “big” yang artinya bukan program “biasa”. Jika data konvensional selama ini hanya berisi teks dan angka biasa seperti data keuangan, maka cukup diolah dengan database biasa pula, misal MS Access, MS SQL Server, dan lain-lain yang selama ini hanya untuk mengolah data terstruktur. Big Data tidak dapat diolah hanya dengan program database konvensional yang disebut SQL (Structured Query Language) atau RDBMS (Relational Database Management System). Big Data membutuhkan program database yang mendukung NoSQL (Not only SQL), yang mampu mengolah data tidak terstruktur.

Dalam bahasa Inggris, Big Data terkait dengan 3V, yakni Volume (ukuran data sangat besar), Velocity (kecepatan transfer/perubahan data sangat tinggi), dan Variety (variasi atau jenis data sangat banyak). Ada juga yang menjadikan 4V, ditambah Value, karena sangat besarnya nilai bisnis yang dihasilkan, sehingga menjadi besar pula peluang kerja bagi profesional di bidang pemrograman komputer, pengolahan data statistik, dan Cloud Computing. Kebutuhan SDM di bidang Big Data pada 2015 ini diproyeksi sekitar 4,4 juta orang. Di Amerika saja butuh 190.000 orang pada 2011 dan akan tambah butuh lagi 490.000 pada 2018 (Sumber: McKinsey Global Institute, 2011).

B. Karakteristik Big Data

Karakteristik Big Data biasa disebut dengan singkatan 4V, yaitu:

a) Volume

Mengacu pada ukuran data yang perlu diproses. Saat ini satuan volume data di dunia telah melampaui zettabyte (1021 byte), bahkan telah banyak perusahaan atau organisasi yang perlu mengolah data sampai ukuran petabytes perharinya. Volume data yang besar ini akan membutuhkan teknologi pemrosesan yang berbeda dari penyimpanan tradisional.

b) Velocity

Adalah kecepatan data yang dihasilkan. Data yang dihasilkan dengan kecepatan tinggi membutuhkan teknik pemrosesan yang berbeda dari data transaksi biasa. Contoh data yang dihasilkan dengan kecepatan tinggi adalah pesan Twitter dan data dari mesin ataupun sensor.

c) Variety

Big Data berasal dari berbagai sumber, dan jenisnya termasuk salah satu dari tiga kategori berikut: data terstruktur, semi terstruktur, dan tidak terstruktur. Tipe data yang bervariasi ini membutuhkan kemampuan pemrosesan dan algoritma khusus. Contoh data dengan variasi tinggi adalah pemrosesan data sosial media yang terdiri dari teks, gambar, suara, maupun video.

d) Veracity

Mengacu pada akurasi atau konsistensi data. Data dengan akurasi tinggi akan memberikan hasil analisis yang berkualitas. Sebaliknya, data dengan akurasi rendah mengandung banyak bias, noise dan abnormalitas. Data ini jika tidak diolah dengan benar akan menghasilkan keluaran yang kurang bermanfaat, bahkan dapat memberikan gambaran atau kesimpulan yang keliru. Veracity merupakan tantangan yang cukup berat dalam pengolahan Big Data.

Di samping 4V tersebut, ada juga yang menambahkan satu lagi sehingga menjadi 5V, yaitu value. Value ini sering didefinisikan sebagai potensi nilai sosial atau ekonomi yang mungkin dihasilkan oleh data. Keempat karakteristik di atas (volume, velocity, variety dan veracity) perlu diolah dan dianalisis untuk dapat memberikan value atau manfaat bagi bisnis maupun kehidupan. Oleh karena itu, karakteristik yang kelima ini berkaitan erat dengan kemampuan kita mengolah data untuk menghasilkan output yang berkualitas.

C. Teknologi Big Data

Ada banyak sekali teknologi open source yang populer dalam ekosistem big data, berikut ini beberapa di antaranya:

a) Apache Hadoop

Apache Hadoop adalah sebuah framework yang memungkinkan untuk melakukan penyimpanan dan pemrosesan data yang besar secara terdistribusi dalam kluster komputer menggunakan model pemrograman sederhana. Hadoop terinspirasi dari teknologi yang dimiliki oleh Google seperti Google File System dan Google Map Reduce.

Hadoop menawarkan 3 hal utama yaitu:

1. Sistem penyimpanan terdistribusi

Hadoop memiliki sebuah file sistem yang dinamakan Hadoop Distributed File System atau lebih dikenal dengan HDFS. HDFS merupakan sistem penyimpanan file

atau data terdistribusi dalam kluster Hadoop. HDFS terinspirasi dari Google File System.

2. Framework pemrosesan data secara paralel dan terdistribusi

MapReduce adalah model pemrograman untuk melakukan pemrosesan data besar secara terdistribusi dalam kluster Hadoop. MapReduce bekerja dan mengolah data-data yang berada dalam HDFS.

3. Resource management terdistribusi

YARN merupakan tools yang menangani resource manajemen dan penjadwalan proses dalam kluster Hadoop. YARN mulai diperkenalkan pada Hadoop 2.0. YARN memisahkan antara layer penyimpanan (HDFS) dan layer pemrosesan (MapReduce). Pada awalnya Hadoop hanya mensupport MapReduce sebagai satu-satunya framework komputasi paralel yang dapat bekerja diatas kluster Hadoop. YARN memungkinkan banyak framework komputasi paralel lain, seperti Spark, Tez, Storm, dsb, untuk bekerja diatas kluster Hadoop dan mengakses data-data dalam HDFS.

b) Apache Hive

Apache Hive adalah sebuah framework SQL yang berjalan di atas Hadoop. Hive mendukung bahasa pemrograman SQL yang memudahkan untuk melakukan query dan analisis data berukuran besar di atas Hadoop. Selain Hadoop, Hive juga dapat digunakan di atas sistem file terdistribusi lain seperti Amazon AWS3 dan Alluxio.

Dukungan Hive terhadap SQL ini sangat membantu portabilitas aplikasi berbasis SQL ke Hadoop, terutama sebagian besar aplikasi data warehouse yang membutuhkan sistem penyimpanan maupun komputasi yang besar. Pada awalnya Hive dikembangkan oleh Facebook untuk digunakan sebagai sistem data warehouse mereka. Setelah disumbangkan ke komunitas open source, Hive berkembang dengan pesat dan banyak diadopsi serta dikembangkan oleh perusahaan besar lainnya seperti Netflix dan Amazon.

c) Apache Spark

Apache Spark merupakan framework komputasi terdistribusi yang dibangun untuk pemrosesan big data dengan kecepatan tinggi. Apache spark memiliki algoritma yang berbeda dengan MapReduce, tetapi dapat berjalan diatas Hadoop melalui YARN. Spark menyediakan API dalam Scala, Java, Python, dan SQL, serta dapat digunakan untuk menjalankan berbagai jenis proses secara efisien, termasuk proses ETL, data streaming, machine learning, komputasi graph, dan SQL. Selain HDFS, Spark juga dapat digunakan di atas file system lain seperti Cassandra, Amazon AWS3, dan penyimpanan awan yang lain. Fitur utama Spark adalah komputasi cluster dalam memori. Penggunaan memori ini dapat meningkatkan kecepatan pemrosesan aplikasi secara drastis. Untuk kasus tertentu, kecepatan pemrosesan Spark bahkan dapat mencapai 100 kali dibanding pemrosesan menggunakan disk seperti MapReduce. Jika MapReduce lebih sesuai digunakan untuk pemrosesan batch dengan dataset yang sangat besar, maka Spark sangat sesuai untuk pemrosesan iteratif dan live-streaming, sehingga Spark banyak dimanfaatkan untuk machine learning. Spark adalah salah satu sub project Hadoop yang dikembangkan pada tahun 2009 di AMPLab UC Berkeley. Sejak tahun 2009, lebih dari 1200 developer telah berkontribusi pada project Apache Spark.

Selain 3 teknologi tersebut, sebenarnya masih sangat banyak teknologi dan framework big data lainnya yang bersifat open source seperti HBase, Cassandra, Presto, Storm, Flink, NiFi, Sqoop, Flume, Kafka dan lain sebagainya.

D. Hasil Pengolahan Big Data

Berikut ini terdapat beberapa hasil penerapan dari Big Data antara lain:

a) Netflix

Netflix, sebuah layanan streaming yang berbasis langganan ini mendulang sukses dengan memberikan layanan konsumen yang sifatnya personal. Hal ini tidak lepas dari peran big data analytics software yang mengumpulkan lebih dari 150 juta pelanggannya dan mengamati interaksi dan respon mereka terhadap sebuah tayangan. Mulai dari jam berapa mereka biasa menonton, device apa yang mereka gunakan, dan apakah seseorang menonton sebuah tayangan sampai habis atau tidak.

b) Starbucks

Gerai kopi Starbucks menggunakan solusi big data yang mereka dapatkan dari survei dan transaksi pembelian pelanggan untuk menyajikan menu-menu yang mereka sukai. Data-data ini diolah dengan big data analytics software untuk mendapatkan menu-menu pilihan berdasarkan kegemaran pelanggan.

c) Amazon

Amazon mengumpulkan data individu setiap pelanggannya saat mereka menggunakan situs web. Selain apa yang dibeli pelanggan, Amazon mengamati produk-produk lain dilihat oleh pelanggan tersebut dengan mengintegrasikan big data analytics software ke dalam website ecommerce-nya. Kemudian Amazon juga mengamati alamat pengiriman, dan apakah pelanggan meninggalkan ulasan.

Big Data telah membantu mendorong Amazon ke puncak e-commerce. Perusahaan terhubung dengan tiap penjual dan agar dapat memastikan inventaris mereka. Amazon juga dapat melacak gudang terdekat yang dapat dipilih sehingga biaya pengirimannya dapat berkurang hingga 40 %. Dengan selalu update mengenai data-data ini, Amazon selalu dapat memprediksi kemungkinan tren di masa mendatang agar tidak tertinggal oleh pesaing dan selalu terdepan dalam melayani konsumen.

d) Gojek

Gojek menghimpun data untuk meningkatkan pengalaman pelanggan sekaligus mitranya. Selama masa pandemi ini misalnya, big data analytics software digunakan oleh Gojek untuk mengembangkan fitur contactless dalam pengiriman makanan agar dapat menerapkan social distancing. Selain itu, Gojek juga memberikan update bagi para driver, area mana saja yang tidak bisa dilewati karena PSBB.

e) Telkomsel

Sebagai pionir dalam mengadakan jaringan 2G, 3G dan 4G LTE di Indonesia, Telkomsel menerapkan big data analytics untuk memprediksi kemungkinan konsumen untuk melakukan pembelian lewat rekomendasi-rekomendasi yang diberikan secara personal oleh Telkomsel.

f) Traveloka

Traveloka melakukan personalisasi dengan menerapkan solusi big data untuk membaca kebutuhan konsumennya. Dengan melacak kebiasaan konsumen serta lokasi mereka, Traveloka dapat merekomendasikan promosi yang berada di area pelanggannya. Traveloka juga menawarkan hotel maupun tiket perjalanan sesuai dengan kebiasaan-kebiasaan konsumen dalam melakukan pencarian.

g) Ovo

OVO menggunakan solusi big data untuk lebih memahami keinginan konsumen. Dengan riset konsumen dan mengolahnya lewat big data analytics software, OVO dapat lebih mengenal tipe-tipe konsumen dan memberikan berbagai jenis layanan yang memang disukai oleh para penggunanya.

BAB III

PENUTUP

A. KESIMPULAN

Dari pemaparan diatas dapat kita simpulkan bahwa teknologi big data sudah mulai dimanfaatkan dan akan sangat berguna untuk difahami lebih dalam untuk mengimbangi perkembangan jaman ke arah teknologi dan analisis yang lebih praktis dan efektif.

B. SARAN

Semoga dengan adanya makalah ini kita bisa mebuatgetahui bagaimana pentingnya untuk Memahami teknologi big data, penerapan dan hasil dari pengolahannya.

Saya sebagai penulis juga menyadari masih banyak kesalahan kesalahan, baik dalam hal penulisan maupun dalam hal yang terkait pada makalah ini, saya sangat membutuhkan kritikan yang bersifat membangun guna untuk menjadi lebih baik lagi kedepannya.