**MAKALAH BIG DATA**

**PENGGAJIAN FRONT-END WEBSITE COMPANY PROFILE UNTUK PT. JASA RAHARJA PERWAKILAN SURABAYA**



**Disusun Oleh:**

|  |  |
| --- | --- |
| Aditya Pratama | 1462000179 |
| Andre Yanuar W | 1462000269 |
| Fadhlyansyah Khair | 1462000170 |
| Mohammad Farhan | 1462000272 |
| Ardha Muktholif | 1462000283 |
| Vegi Burhanudin | 1462000258 |

**PROGRAM SARJANA**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

**2023**

KATA PENGANTAR

Dengan menyebut nama Allah SWT yang Maha Pengasih lagi Maha Panyayang, Kami panjatkan puja dan puji syukur atas kehadirat-Nya, yang telah melimpahkan rahmat, hidayah, dan inayah-Nya kepada kami, sehingga kami dapat menyelesaikan makalah dengan judul BIG DATA.

Pemanfaatan Big Data telah menjadi topik yang semakin penting dalam berbagai bidang, termasuk bisnis, kesehatan, pendidikan, teknologi, dan banyak lagi. Data-data yang terus berkembang dengan cepat telah memberikan kesempatan baru untuk menggali wawasan yang berharga, membuat keputusan yang lebih tepat, dan meningkatkan produktivitas. Namun, seiring dengan potensi manfaatnya, muncul pula berbagai tantangan, termasuk privasi data, keamanan, dan etika penggunaan data.

Dalam makalah ini, kami akan membahas konsep dasar Big Data, termasuk definisinya, karakteristik, dan teknologi yang mendukung pengumpulan dan analisis data. Kami juga akan menyelidiki berbagai sektor yang telah memanfaatkan Big Data, seperti bisnis untuk analisis pasar dan peramalan, kesehatan untuk penelitian medis dan manajemen penyakit, serta pemerintahan untuk pengambilan keputusan berbasis data..

Akhir kata kami berharap semoga makalah ilmiah tentang limbah dan manfaatnya untuk masyarakan ini dapat memberikan manfaat maupun inpirasi terhadap pembaca.

Surabaya, Oktober 2023

Penyusun

**BAB 1**

**PENDAHULUAN**

**Latar Belakang**

Perkembangan jaman hingga saat ini sangat berdampak besar terhadap kehidupan sekarang. Salah satu hal yang berubah adalah cara menggunakan data. Hal tersebut sangat dipengaruhi oleh perkembangan teknologi, karena dapat dilihat sekarang penggunaan tiap individu terhadap data sudah sangat tinggi, hampir semua orang memiliki data dalam setiap perangkatnya (komputer / laptop, smartphone, flashdisk, harddisk eksternal, dll) yang jika dijumlahkan akan menjadi besar sekali. Hal ini dipengaruhi juga dengan mudahnya tiap individu untuk mendapatkan data yang diinginkannya (film, musik, games, dll) melalui internet. Internet menghubungkan tiap individu di seluruh dunia dengan mudah tanpa memperdulikan jarak / lokasi dan waktu. Sekarang dengan terjadinya perkembangan teknologi, data menjadi hal yang penting dalam menjalankan berbagai hal, beberapa diantaranya; mengetahui tren pasar, mengetahui keinginan konsumen saat ini, meningkatkan hasil penjualan, dll. Hasil perubahan ini sangatlah besar, data pun diolah dengan lebih terkomputerisasi sehingga penyimpanan beberapa data dapat menghemat tempat dalam kantor perusahaan dengan cara penyimpanan *softcopy*. Data yang tersimpan ini lama kelamaan menjadi sangat banyak dan besar sehingga semakin susah untuk digunakan, hal tersebut disebut big data. Dengan perkembangan sekarang, big data ini sudah dapat diolah dan digunakan lagi, bahkan memberikan hasil yang lebih baik karena mencakup pengolahan data yang ada di dalam social media.

Dengan perkembangan data inilah big data muncul dan saat ini mulai berkembang. Penggunaannya pun semakin luas, hingga mencakup social media, sehingga dapat menganalisa tren pasar dengan melihat sentimen analisis pelanggan melalui social media. Dengan perkembangan saat ini, ada baiknya untuk memahami lebih dalam mengenai big data, sehingga dapat dimanfaatkan dengan lebih maksimal.

**BAB 2**

**LANDASAN TEORI**

**Pengertian Data**

(R. Kelly Rainer, 2011) Data, menunjuk pada deskripsi dasar akan benda, event, aktivitas, dan transaksi yang terdokumentasi, terklasifikasi,dan tersimpan tetapi tidak terorganisasi untuk dapat memberikan suatu arti yang spesifik.

Berdasarkan pengertian di atas, data merupakan hal paling mendasar yang dibutuhkan perusahaan yang dapat diperoleh dari proses-proses operasional sehari-hari maupun sumber-sumber luar yang akan diolah menurut keinginan perusahaan.

**Pengertian Information**

(R. Kelly Rainer, 2011) Information, merupakan data yang telah terorganisir agar dapat memberikan arti dan nilai kepada penerima.

Berdasarkan pengertian di atas, hasil penyusunan dan transformasi data yang dapat memberikan makna baru kepada data tersebut.

**Pengertian Knowledge**

(R. Kelly Rainer, 2011) Terdiri dari data atau informasi yang telah terorganisasi dan proses untuk memberikan pemahaman, pengalaman, dan pembelajaran, serta keahlian terhadap problema bisnis yang sedang dihadapi.

Berdasarkan pengertian di atas, *knowledge* menjadi sarana bagi para manajer untuk membuat keputusan - keputusan yang *crucial*dan berdampak besar bagi perusahaan, dimana kesalahan atau kecacatan dalam *knowledge* dapat memberikan dampak buruk bagi perusahaan.

**Pengertian Extract, Tranform, Load (ETL)** (immon liad DWH)

(H.Inmon, 2005) ETL, proses memasukkan data, mengintergrasikan dan menempatkannya dalam data warehouse.

Berdasarkan pengertian tersebut, ETL adalah proses mengubah raw data sebelum masuk kedalam data warehouse.

**Pengertian Data Warehouse**

(R. Kelly Rainer, 2011) Data Warehouse adalah *repository* dari data-data yang bersifat *historical* yang terorganisir berdasarkan subjek yang digunakan untuk mendukung pengambilan keputusan. *Data warehouse* memiliki beberapa karakter dasar seperti. Diatur oleh business dimension or subject dimana data disusun berdasarkan subjeknya dan memiliki informasi yang relevan dengan pemgambilan keputusan dan analisis data.consistent yaitu data memliki bentuk yang sama disetiap atau disemua database. *Historical*, data yang ada merupakan data yang tersimpan dan terkumupl dalam waktu yang lama yang digunakan untuk *forecasting* dan perbandingan untuk meliat tingakt laju suatu perusahaan. *Use only analytical processing* menggunakan *analytical processing* yang berguna unutk mengakumulasi data-data yang ada. Multidimensional, data warehouse menyimpan data dalam lebih dari dua dimensi seperti *data cube*.

Data warehouse merupakan bagian penting dalam struktur / arsitektur suatu BI karena posisinya sebagai tempat penyimpanan data- data yang telah terorganisasi dan yang telah memiliki makna, maka harus memilki struktur data desain yang baik yang dapat mensupport pengambilan data-data dan informasi secara akurat dan cepat dari dalam data warehouse itu sendiri.

**Pengertian Big Data**

Menurut (Eaton, Dirk, Tom, George, & Paul) *Big Data* merupakan istilah yang berlaku untuk informasi yang tidak dapat diproses atau dianalisis menggunakan alat tradisional.

Menurut (Dumbill, 2012) , Big Data adalah data yang melebihi proses kapasitas dari kovensi sistem database yang ada. Data terlalu besar dan terlalu cepat atau tidak sesuai dengan struktur arsitektur database yang ada. Untuk mendapatkan nilai dari data, maka harus memilih jalan altenatif untuk memprosesnya.

Berdasarkan pengertian para ahli di atas, dapat disimpulkan bahwa Big Data adalah data yang memiliki volume besar sehingga tidak dapat diproses menggunakan alat tradisional biasa dan harus menggunakan cara dan alat baru untuk mendapatkan nilai dari data ini.

**BAB 3**

**PEMBAHASAN**

**3.1 Big Data**

Setiap hari, kita menciptakan 2,5 triliun byte data - begitu banyak bahwa 90% dari data di dunia saat ini telah dibuat dalam dua tahun terakhir saja. Data ini berasal dari mana-mana, sensor digunakan untuk mengumpulkan informasi iklim, posting ke situs media sosial, gambar digital dan video, catatan transaksi pembelian, dan sinyal ponsel GPS untuk beberapa nama. Data ini adalah *big data*.

*Big Data* mengacu pada dataset yang ukurannya diluar kemampuan dari database software tools untuk meng-capture, menyimpan,me-manage dan menganalisis. Definisi ini sengaja dibuat subjective agar mampu digabungkan oleh definisi Big Data yang masi belum ada baku. Ukuran big data sekitar beberapa lusin TeraByte sampai ke beberapa PetaByte tergantung jenis Industri

Isi dari Big Data adalah Transaksi+interaksi dan observasi atau bisa di bilang segalanya yang berhubungan dengan jaringan internet, jaringan komunikasi, dan jaringan satelit

*Big Data* dapat juga didefinisikan sebagai sebuah masalahdomain dimana teknologi tradisional seperti relasional database tidak mampu lagi untuk melayani.

***Big data* lebih dari hanya masalah ukuran, itu adalah kesempatan untuk menemukan wawasan dalam jenis baru dan muncul data dan konten, untuk membuat bisnis Anda lebih gesit, dan menjawab pertanyaan yang sebelumnya dianggap di luar jangkauan Anda.**

*Big data* dapat di artikan kedalam 9 karakter (IBM) menurut responden sehingga disimpulkan oleh IBM, Big data adalah data yang memiliki scope informasi yang sangat besar, model informasi yang real-time, memiliki volume yang besar, dan berasalkan social media data jadi dapat disimpulkan bahwa Big data adalah dataset yang memiliki volume besar dan salah satu isinya berdasarkan social media data, dan informasi dari Big data selalu yang terbaru (latestdata) sehingga model informasi nya real-time, dan scope informasi nya tidak terfocus pada industri-indrustri kecil saja atau industri-indrustri besar saja melainkan semuanya baik industry kecil maupun besar.

**3.2 Dimensi -Dimensi Big Data**

Ada 3 dimensi awal dalam Big Data yaitu 3V: Volume, Variety dan Velocity

* Volume

perusahaan tertimbun dengan data yang terus tumbuh dari semua jenis sektor, dengan mudah mengumpulkan terabyte bahkan petabyte-informasi.

* Mengubah 12 terabyte Tweet dibuat setiap hari ke dalam peningkatan sentimen analisis produk.
* Mengkonvert 350 milliar pembacaan tahunan untuk lebih baik dalam memprediksi kemampuan beli pasar.

Mungkin karakteristik ini yang paling mudah dimengerti karena besarnya data. Volume juga mengacu pada jumlah massa data, bahwa organisasi berusaha untuk memanfaatkan data untuk meningkatkan pengambilan keputusan yang banyak perusahaan di banyak negara. Volume data juga terus meningkat dan belum pernah terjadi sampai sethinggi ini sehingga tidak dapat diprediksi jumlah pasti dan juga ukuran dari data sekitar lebih kecil dari petabyte sampai zetabyte. Dataset big data sekitar 1 terabyte sampai 1 petabyte perperusahaan jadi jika big data digabungkan dalam sebuah organisasi / group perusahaan ukurannya mungkin bisa sampai zetabyte dan jika hari ini jumlah data sampai 1000 zetabyte, besok pasti akan lebih tinggi dari 1000 zetabyte.

* Variety

Volume data yang banyak tersebut bertambah dengan kecepatan yang begitu cepat sehingga sulit bagi kita untuk mengelola hal tersebut. Kadang-kadang 2 menit sudah menjadi terlambat. Untuk proses dalam waktu sensitif seperti penangkapan penipuan, data yang besar harus digunakan sebagai aliran ke dalam perusahaan Anda untuk memaksimalkan nilainya.

* Meneliti 5 juta transaksi yang dibuat setiap hari untuk mengidentifikasi potensi penipuan
* Menganalisis 500 juta detail catatan panggilan setiap hari secara real-time untuk memprediksi gejolak pelanggan lebih cepat.

Berbagai jenis data dan sumber data. Variasi adalah tentang mengelolah kompleksitas beberapa jenis data, termasuk structured data, unstructured data dan semi-structured data. Organisasi perlu mengintegrasikan dan menganalisis data dari array yang kompleks dari kedua sumber informasi Traditional dan non traditional informasi, dari dalam dan luar perusahaan. Dengan begitu banyaknya sensor, perangkat pintar (smart device) dan teknologi kolaborasi sosial, data yang dihasilkan dalam bentuk yang tak terhitung jumlahnya, termasuk text, web data, tweet, sensor data, audio, video, click stream, log file dan banyak lagi.

* Velocity :

*Big Data* adalah setiap jenis data - data baik yang terstruktur maupun tidak terstruktur seperti teks, data sensor, audio, video, klik stream, file log dan banyak lagi. Wawasan baru ditemukan ketika menganalisis kedua jenis data ini bersama-sama.

* Memantau 100 video masukan langsung dari kamera pengintai untuk menargetkan tempat tujuan.
* Mengeksploitasi 80% perkembangan data dalam gambar, video, dan dokumen untuk meningkatkan kepuasan pelanggan.

Data dalam gerak. Kecepatan di mana data dibuat, diolah dan dianalisis terus menerus. Berkontribusi untuk kecepatan yang lebih tinggi adalah sifat penciptaan data secara real-time, serta kebutuhan untuk memasukkan streaming data ke dalam proses bisnis dan dalam pengambilan keputusan. Dampak Velocity latency, jeda waktu antara saat data dibuat atau data yang ditangkap, dan ketika itu juga dapat diakses. Hari ini, data terus-menerus dihasilkan pada kecepatan yang mustahil untuk sistem tradisional untuk menangkap, menyimpan dan menganalisis. Jenis tertentu dari data harus dianalisis secara real time untuk menjadi nilai bagi bisnis.

**3.3 Dimensi Ketidakpastian data**

Dalam industri untuk meningkatan sebuah data lebih berkualitas dibutuhkan dimensi ke empat yaitu Veracity, pencantuman Veracity dapat menekankan pengelolahan dan penanganan untuk suatu ketidakpastian yang melekat dalam beberapa jenis data.

Veracity

1 dari 3 pemimpin bisnis tidak mempercayai informasi yang mereka gunakan untuk membuat keputusan. Bagaimana Anda dapat bertindak atas informasi yang anda tidak percaya? Membangun kepercayaan atas *big data* menghadirkan tantangan besar sebagai variasi dan sumber untuk pertumbuhan perusahaan.

ketidakpastian data. Veracity mengacu pada tingkat keandalan yang terkait dengan jenis tertentu dari data. Berjuang untuk kualitas data yang tinggi merupakan syarat big data penting dan tantangan, tapi bahkan metode pembersihan data yang terbaik tidak dapat menghapus ketidakpastian yang melekat pada beberapa data, seperti cuaca, ekonomi, atau aktual keputusan membeli pelanggan masa depan. Kebutuhan untuk mengakui dan merencanakan ketidakpastian adalah dimensi data besar yang telah diperkenalkan sebagai eksekutif berusaha untuk lebih memahami dunia di sekitar mereka.  
 Beberapa data tidak pasti, misalnya: sentimen dan kebenaran pada manusia, sensor GPS memantul antara pencakar langit Manhattan, cuaca kondisi-kondisi, faktor ekonomi, dan masa depan. Ketika berhadapan dengan jenis data, tidak ada metode pembersihan data dapat memperbaiki untuk semua itu. Namun, meski ketidakpastian, data masih mengandung informasi yang berharga. Kebutuhan untuk mengakui dan menerima ketidakpastian ini merupakan ciri dari data.Uncertainty besar memanifestasikan dirinya dalam data besar dalam banyak cara. Sekarang dalam skeptisisme yang mengelilingi data yang dibuat dalam lingkungan manusia seperti jaringan sosial, dalam ketidaktahuan bagaimana masa depan akan terungkap dan bagaimana orang-orang, alam atau kekuatan pasar yang tak terlihat akan bereaksi terhadap variabilitas dari dunia di sekitar mereka.

Untuk mengelola ketidakpastian, analis perlu menciptakan konteks sekitar data. Salah satu cara untuk mencapai ini adalah melalui data fusion, di mana menggabungkan beberapa sumber yang kurang dapat diandalkan menciptakan lebih akurat dan berguna point data, seperti komentar sosial ditambahkan ke geospasial informasi lokasi. Cara lain untuk mengelola ketidakpastian adalah melalui matematika canggih yang mencakup hal itu, seperti teknik optimasi yang kuat dan pendekatan fuzzy logic. Manusia secara alami, tidak menyukai ketidakpastian, tetapi hanya mengabaikannya dapat menciptakan lebih banyak masalah daripada ketidakpastian itu sendiri. Dalam era big data, eksekutif akan perlu pendekatan dimensi ketidakpastian berbeda. Mereka akan perlu untuk mengakuinya, menerimanya dan menentukan.

**3.4 Arsitektur Big Data**

Traditional Information Architecture Capabilities

Untuk memahami level aspek arsitektur yang tinggi dari Big Data, sebelumnya harus memahami arsitektur informasi logis untuk data yang terstruktur. Pada gambar di bawah ini menunjukkan dua sumber data yang menggunakan teknik integrasi (ETL / Change Data Capture) untuk mentransfer data ke dalam DBMS data warehouse atau operational data store, lalu menyediakan bermacam-macam variasi dari kemampuan analisis untuk menampilkan data. Beberapa kemampuan analisis ini termasuk,; dashboards, laporan, EPM/BI Applications, ringkasan dan query statistic, interpretasi semantic untuk data tekstual, dan alat visualisasi untuk data yang padat. Informasi utama dalam prinsip arsitektur ini termasuk cara memperlakukan data sebagai asset melalui nilai, biaya, resiko, waktu, kualitas dan akurasi data.

Mendefinisikan kemampuan memproses untuk *big data architecture*, diperlukan beberapa hal yang perlu dilengkapi; volume, percepatan, variasi, dan nilai yang menjadi tuntutan. Ada strategi teknologi yang berbeda untuk *real-time* dan keperluan *batch* *processing*. Untuk *real*-*time*, menyimpan data nilai kunci, seperti NoSQL, memungkinkan untuk performa tinggi, dan pengambilan data berdasarkan indeks. Untuk batch processing, digunakan teknik yang dikenal sebagai *Map Reduce*, memfilter data berdasarkan pada data yang spesifik pada strategi penemuan. Setelah data yang difilter ditemukan, maka akan dianalisis secara langsung, dimasukkan ke dalam unstructured database yang lain, dikirimkan ke dalam perangkat mobile atau digabungkan ke dalam lingkungan data warehouse tradisional dan berkolerasi pada data terstruktur.

Sebagai tambahan untuk *unstructured* *data* yang baru, ada dua kunci perbedaan untuk *big* *data*. Pertama, karena ukuran dari data set, *raw* *data* tidak dapat secara langsung dipindahkan ke dalam suatu data warehouse. Namun, setelah proses *Map* *Reduce* ada kemungkinan akan terjadi reduksi hasil dalam lingkungan data warehouse sehingga dapat memanfaatkan pelaporan *business* *intelligence*, statistik, semantik, dan kemampuan korelasi yang biasa. Akan sangat ideal untuk memiliki kemampuan analitik yang mengkombinasikan perangkat BI bersamaan dengan visualisasi big data dan kemampuan query. Kedua, untuk memfasilitasi analisis dalam laingkungan Hadoop, lingkungan sandbox dapat dibuat.

Untuk beberapa kasus, big data perlu mendapatkan data yang terus berubah dan tidak dapat diperkirakan, untuk menganilisis data tersebut, dibutuhkan arsitektur yang baru. Dalam perusahaan retail, contoh yang bagus adalah dengan menangkap jalur lalu lintas secara real-time dengan maksud untuk memasang iklan atau promosi toko di tempat strategis yang dilewati banyak orang, mengecek peletakan barang dan promosi, mengamati secara langsung pergerakan dan tingkah laku pelanggan.

Dalam kasus lain, suatu analisis tidak dapat diselesaikan sampai dihubungkan dengan data perusahaan dan data terstruktur lainnya. Sebagai contohnya, analisis perasaan pelanggan, mendapatkan respon positif atau negatif dari *social* *media* akan memiliki suatu nilai, tetapi dengan mengasosiasikannya dengan segala macam pelanggan (paling menguntungkan atau bahkan yang paling tidak menguntungkan) akan memberikan nilai yang lebih berharga. Jadi, untuk memenuhi kebutuhan yang diperlukan oleh big data BI adalah konteks dan pemahaman. Menggunakan kekuatan peralatan statistikal dan semantik akan sangat memungkinkan untuk dapat memprediksikan kemungkinan – kemungkinan di masa depan.

An Integrated Information Architecture

Salah satu tantangan yang diteliti dalam pemakaian Hadoop dalam perusahaan adalah kurangnya integrasi dengan ekosistem BI yang ada. Saat ini BI tradisional dan ekosistem big data terpisah dan menyebabkan analis data terintegrasi mengalami kebingungan. Sebagai hasilnya, hal ini tidaklah siap untuk digunakan oleh pengguna bisnis dan eksekutif biasa.

Pengguna big data yang pertama kali mencoba menggunakan, seringkali menulis kode khusus untuk memindahkan hasil big data yang telah diproses kembali ke dalam database untuk dibuat laporan dan dianalisa. Pilihan – pilihan ini mungkin tidak layak dan ekonomis untuk perusahaan IT. Pertama, karena menyebabkan penyebaran salah satu data dan standar yang berbeda, sehingga arsitekturnya mempengaruhi ekonomi IT. Big data dilakukan secara independen untuk menjalankan resiko investasi yang redundan, sebagai tambahannya, banyak bisnis yang sama sekali tidak memiliki staff dan ketrampilan yang dibutuhkan untuk pengembangan pekerjaan yang khusus.

Pilihan yang paling tepat adalah menggabungkan hasil big data ke dalam data warehouse. Kekuatan informasi ada dalam kemampuan untuk asosiasi dan korelasi. Maka yang dibutuhkan adalah kemampuan untuk membawa sumber data yang berbeda-beda, memproses kebutuhan bersama – sama secara tepat waktu dan analisis yang berharga.

Ketika bermacam – macam data telah didapatkan, data tersebut dapat disimpan dan diproses ke dalam DBMS tradisional, simple files, atau sistem cluster terdistribusi seperti NoSQL dan Hadoop Distributed File System (HDFS).

Secara arsitektur, komponen kritikal yang memecah bagian tersebut adalah layer integrasi yang ada di tengah. Layer integrasi ini perlu untuk diperluas ke seluruh tipe data dan domain, dan menjadi jembatan antara data penerimaan yang baru dan tradisional, dan pengolahan kerangka. Kapabilitas integrasi data perlu untuk menutupi keseluruhan spektrum dari kecepatan dan frekuensi. Hal tersebut diperlukan untuk menangani kebutuhan ekstrim dan volume yang terus bertambah banyak. Oleh karena itu diperlukan teknologi yang memungkinkan untuk mengintegrasikan Hadoop / Map Reduce dengan data warehouse dan data transaksi.

Layer berikutnya digunakan untuk *Load* hasil reduksi dari big data ke dalam data warehouse untuk analisis lebih lanjut. Diperlukan juga kemampuan untuk mengakses data terstruktur seperti informasi profil pelanggan ketika memproses dalam big data untuk mendapatkan pola seperti mendeteksi aktivitas yang mencurigakan.

Hasil pemrosesan data akan dimasukkan ke dalam ODS tradisional, data warehouse, dan data marts untuk analisis lebih lanjut seperti data transaksi. Komponen tambahan dalam layer ini adalah Complex Event Processing untuk menganalisa arus data secara real-time. Layer business intelligence akan dilengkapi dengan analisis lanjutan, dalam analisis database statistik, dan visualisasi lanjutan, diterapkan dalam komponen tradisional seperti laporan, dashboards, dan query. Pemerintahan, keamanan, dan pengelolaan operasional juga mencakup seluruh spektrum data dan lanskap informasi pada tingkat enterprise.

Dengan arsitektur ini, pengguna bisnis tidak melihat suatu pemisah, bahkan tidak sadar akan perbedaan antara data transaksi tradisional dan big data. Data dan arus analisis akan terasa mulus tanpa halangan ketika dihadapkan pada bermacam – macam data dan set informasi, hipotesis, pola analisis, dan membuat keputusan.

**3.5 Big Data for development**

Big Data untuk keperluan development berkaitan dengan, tetapi berbeda dari, 'tradisional Data pembangunan '(misalnya data survei, statistik resmi), dan sektor swasta dan media mainstream menyebutnya 'Big Data’.

Big Data untuk sumber Pengembangan umumnya memiliki beberapa / semua fitur ini:

1. *Digitally generated*

data yang dihasilkan secara digital (sebagai lawan yang didigitalkan manual), dan dapat disimpan dengan menggunakan rangkaian satu dan nol, dan dengan demikian dapat dimanipulasi oleh komputer

2. *Passively produced*

Data ini merupakan data yang dihasilkan atau produk dari kehidupan kita sehari-hari atau interaksi dengan jasa digital.

3. *Automatically collected*

Data-data yang terbentuk dari data-data operasional dan transaksi yang dikumpulkan dan telah diproses (ETL) dan si simpan kedalam data mart

4. Geographically or temporally trackable

Data –data yang menunjukan lokasi atau posisi, misalnya data lokasi ponsel atau durasi waktu panggilan

5. Continuously analysed

informasi yang relevan dengan kesejahteraan manusia dan pembangunan dan dapat dianalisis secara real-time

**3.6 Pengunaan Big Data dalam perusahaan**

* IT logs Analytics

Penyimpanan Log jangka panjang, digunakan untuk analisa proses sistem yang sedang berjalan untuk mencegah dan menaggulangi kegagalan dalam sistem, mengunakan hasil analisa log untuk menemukan dan mentukan secara pasti kegagalan apa yang terjadi didalam sistem, menyiapkan langkah-langkah pasti yang dapat digunakan sebagai solusi masalah sistem.

* Fraud Detection Pattern

Banyak digunakan dalam Bidang keuangan atau dimana saja transaksi finasial terlibat, Memaksimalkan pengunaan data-data yang ada untuk memberikan kemampuan unutk mendeteksi fraud ketika transaksi sedang berlangsung

* The Social Media Pattern

Pengunaan Big data untuk analisa media social dan sentiment pelangan, memberikan kemampuan bagi perusahan untuk mengetahui keinginan customer secara luas, mendapatkan feedback secara langsung, dan mengenali langsung dampak sentimen terhadap penjualan, serta efektivitas dan penerimaan pelangan terhadap pemasaran yang dilakukan.

* The Call centere Mantra

Penyimpanan hasil perbincangan atau laporan customer dalam bentuk text yang kemudian digunakan sebagai data untuk analisa masalah yang dihadapai customer, memberikan kemampuan bagi perusahaan untuk memberikan tanggapan yang cepat maupun secara langsung terhadap masalah yang dihadapi customer, serta kemampuan unutk mendeteksi penurunan loyalitas customer dikarenakan masalah dan ketidakpuasaan.

* Risk: Patterns for Modeling and Management

Memberikan kempuaan pengunaan data secara penuh dan analisis dalam pemodelan resiko dan menejemen resiko untuk memberikan pengetahuan akan resiko dan penanggulangannya secara tepat dan langsung

* Big data and The Energy Sector

Memberikan kemampuan penyimpanan dan pemrosesan data secara langsung dari berbagai sumber(sensor), analisa dan kemudahan dalam pengenalan *noise* untuk memisahkannya dari signal*.*

**3.7 Membangun *Big Data* Platform**

Seperti data pergudangan, toko web atau platform TI, infrastruktur untuk data yang besar memiliki kebutuhan yang unik. Dalam mempertimbangkan semua komponen platform data yang besar, penting untuk diingat bahwa tujuan akhir adalah untuk dengan mudah mengintegrasikan data yang besar dengan data perusahaan Anda untuk memungkinkan Anda untuk melakukan analisis mendalam pada set data gabungan.

Infrastructure Requirements

Requirement dalam *big data* infrastruktur :

* *data acquisition*,
* *data organization*
* *data analysis*

***Data acquisition***

Tahap akuisisi adalah salah satu perubahan besar dalam infrastruktur pada hari-hari sebelum *big data*. Karena *big data* mengacu pada aliran data dengan kecepatan yang lebih tinggi dan ragam yang bervariasi, infrastruktur yang diperlukan untuk mendukung akuisisi data yang besar harus disampaikan secara perlahan, dapat diprediksi baik di dalam menangkap data dan dalam memprosesnya secara cepat dan sederhana, dapat menangani volume transaksi yang sangat tinggi , sering dalam lingkungan terdistribusi, dan dukungan yang fleksibel, struktur data dinamis.

Database NoSQL sering digunakan untuk mengambil dan menyimpan *big data*. Mereka cocok untuk struktur data dinamis dan sangat terukur. Data yang disimpan dalam database NoSQL biasanya dari berbagai variasi/ragam karena sistem dimaksudkan untuk hanya menangkap semua data tanpa mengelompokkan dan *parsing* data.

Sebagai contoh, database NoSQL sering digunakan untuk mengumpulkan dan menyimpan data media sosial. Ketika aplikasi yang digunakan pelanggan sering berubah, struktur penyimpanan dibuat tetap sederhana. Alih-alih merancang skema dengan hubungan antar entitas, struktur sederhana sering hanya berisi kunci utama untuk mengidentifikasi titik data, dan kemudian wadah konten memegang data yang relevan. Struktur sederhana dan dinamis ini memungkinkan perubahan berlangsung tanpa reorganisasi pada lapisan penyimpanan.

***Data Organization***

Dalam istilah Data pergudangan klasik, pengorganisasian data disebut integrasi data. Karena ada volume/jumlah data yang sangat besar, ada kecenderungan untuk mengatur data pada lokasi penyimpanan aslinya, sehingga menghemat waktu dan uang dengan tidak memindah-midahkan data dengen volume yang besar. Infrastruktur yang diperlukan untuk mengatur data yang besar harus mampu mengolah dan memanipulasi data di lokasi penyimpanan asli. Biasanya diproses didalam *batch* untuk memproses data yang besar, beragam format, dari tidak terstruktur menjadi terstruktur.

Apache Hadoop adalah sebuah teknologi baru yang memungkinkan volume data yang besar untuk diatur dan diproses sambil menjaga data pada cluster penyimpanan data asli. Hadoop Distributed File System (HDFS) adalah sistem penyimpanan jangka panjang untuk log web misalnya. Log web ini berubah menjadi perilaku browsing dengan menjalankan program MapReduce di cluster dan menghasilkan hasil yang dikumpulkan di dalam cluster yang sama. Hasil ini dikumpulkan kemudian dimuat ke dalam sistem DBMS relasional.

***Data Analysis***

Karena data tidak selalu bergerak selama fase organisasi, analisis ini juga dapat dilakukan dalam lingkungan terdistribusi, di mana beberapa data akan tinggal di mana data itu awalnya disimpan dan diakses secara transparan dari sebuah data warehouse. Infrastruktur yang diperlukan untuk menganalisis data yang besar harus mampu mendukung analisis yang lebih dalam seperti analisis statistik dan data mining, pada data dengan jenis yang beragam dan disimpan dalam sistem yang terpisah, memberikan waktu respon lebih cepat didorong oleh perubahan perilaku; dan mengotomatisasi keputusan berdasarkan model analitis. Yang paling penting, infrastruktur harus mampu mengintegrasikan analisis pada kombinasi data yang besar dan data perusahaan tradisional. Wawasan baru datang bukan hanya dari analisis data baru, tapi dari menganalisisnya dalam konteks yang lama untuk memberikan perspektif baru tentang masalah lama.

Misalnya, menganalisis data persediaan dari mesin penjual otomatis cerdas dalam kombinasi dengan acara kalender untuk tempat di mana mesin penjual otomatis berada, akan menentukan kombinasi produk yang optimal dan jadwal pengisian untuk mesin penjual otomatis.

**3.8 Tantangan dalam pemanfaatan Big Data**

Dalam usaha pemanfaatan *Big Data* dapat terdapat banyak hambatan dan tantangan, beberapa hal diantaranya berhubungan dengan data dimana melibatkan acquisition, sharing dan privasi data, serta dalam analisis dan pengolahan data

* Privasi

Privasi merupakan isu yang paling sensitif, dengan konseptual, hukum, dan teknologi, Privasi dapat dipahami dalam arti luas sebagai usaha perusahaan untuk melindungi daya saing dan konsumen mereka. Data-data yang digunakan / disimpan sebagai big data

* *Access* dan *sharing*

Akses terhadap data, baik data lama maupun data baru dapat menjadi hambatan dalam mendapatkan data untuk big data, terlebih pada data lama dimana data- data tersimpan dalam bentuk – bentuk yang berbeda-beda dan beragam ataupun dalam bentuk fisik, akses terhadap data baru juga membutuhkan usaha yang lebih kerana diperlukannya izin dan lisensi untuk mengakses data-data non-public secara legal.

* Analisis

Bekerja dengan sumber data baru membawa sejumlah tantangan analitis. relevansi dan tingkat keparahan tantangan akan bervariasi tergantung pada jenis analisis sedang dilakukan, dan pada jenis keputusan yang akhirnya akan bisa diinformasikan oleh data.

Tergantung dari jenis data terdapat 3 kategori dalam analisis data

* + Penentuan gambaran yang benar

Masalah ini biasanya ditemukan dalam penanganan *unstructured user-generated text-based data* dimana data yang didapatkan belum tentu benar karena data atau sumber yang salah.

* + *Interpreting Data*

Kesalahan –kesalahan seperti Sampling selection bias merupakan hal yang sering ditemukan dimana data yang ada tidak dapat digunakan untuk mepresentasikan semua populasi yang ada, dan apophenia, melihat adanya pola walaupun tidak benar-benar ada dikarenakan jumlah data yang besar, dan kesalahan dalam menginterpreasikan hubungan dalam data.

* + Defining and detecting anomalies

Tantangan sensitivitas terhadap spesifisitas pemantauansistem. Sensitivitas mengacu pada kemampuan sistem pemantauan untuk mendeteksi semua kasus sudah diatur untuk mendeteksi sementara spesifisitas mengacu pada kemampuannya untuk mendeteksi hanya kasus-kasus yang relevan. kegagalan untukmencapai hasil yang terakhir "Tipe I kesalahan keputusan", juga dikenal sebagai "positif palsu"; kegagalanuntuk mencapai mantan "Type II error", atau "negatif palsu." Kedua kesalahan yang tidak diinginkan ketika mencoba untuk mendeteksi malfungsi atau anomali, bagaimanapun didefinisikan, untuk berbagai alasan. Positif palsu merusak kredibilitas sistem sementara negatif palsu dilemparkan ragu pada relevansinya. Tapi apakah negatif palsu lebih atau kurang bermasalah daripada positif palsu tergantung pada apa yang sedang dipantau, dan mengapa itu sedang dipantau.

**BAB 4**

**Penutup**

**Simpulan**

Dapat disimpulkan bahwa big data sudah mulai dimanfaatkan dan akan sangat berguna untuk dipahami lebih dalam untuk mengimbangi perkembangan jaman ke arah teknologi dan analisis yang lebih praktis.

**Saran**

* Mempelajari lebih dalam mengenai big data, karena hal ini akan sangat berguna bagi perusahaan, dimana perusahaan jaman sekarang mulai mempelajari tren pasar dan pola hidup konsumen.
* Dalam penerapannya, perusahaan akan membutuhkan orang yang khusus dalam bidang data mining, karena hal ini cukup rumit dan diperlukan orang yang paham cara untuk menggunakan data.

**Daftar Pustaka**