

BIG DATA

(INFORMASI DAN KASUS)



2022

Muhammad Syarif Hartawan M.Kom MM
Susy Rosyida S.Kom M.Kom
Abdul Hamid S.Kom M.Kom
Wulan Dari S.Kom M.Kom
Dr. Arman Syah Putra S.Kom MM M.Kom

BIG

DATA

(INFORMASI DAN KASUS)

2022

Big Data

Informasi dan Kasus

Penulis

Muhammad Syarif Hartawan M.Kom MM || Susy Rosyida S.Kom M.Kom || Abdul Hamid S.Kom M.Kom || Wulan Dari S.Kom M.Kom || Dr. Arman Syah Putra S.Kom MM M.Kom

ISBN

978-623-481-049-3

Editor

Tim Kun Fayakun

Layout

Tim Kun Fayakun

Penyunting

Tim Kun Fayakun

Desain Sampul dan Tata Letak

Tim Kun Fayakun

Penerbit

Tim Kun Fayakun

Redaksi

Tim Kun Fayakun

Jawa Timur

Hp. 0856 07 8802

Email : penulis.kunfayakun@gmail.com

Cetakan pertama : Mei 2022

Hak cipta dilindungi undang-undang

Dilarang memperbanyak karya tulis ini dan dalam bentuk dan dengan cara apapun tanpa izin tertulis dari penerbit
Isi di luar tanggung jawab penerbit dan percetakan

KATA PENGANTAR

Terima kasih kepada Allah SWT dan kedua orang tua kami atas terbit nya buku di tahun 2022 ini, dengan terbit nya buku ini di harapkan berguna untuk semua orang dalam hal riset dan penelitian di bidang data terutama di bidang Big Data, buku ini berisikan tentang pengertian data dan kasus, diharapkan dengan ada buku-buku selanjutnya dan bisa terus berkarya agar bisa membantu dalam hal riset dan penelitian di bidang kecerdasan buatan.

Penulis

Muhammad Syarif Hartawan M.Kom MM

Susy Rosyida S.Kom M.Kom

Abdul Hamid S.Kom M.Kom

Wulan Dari S.Kom M.Kom

Dr. Arman Syah Putra S.Kom MM M.Kom

DAFTAR ISI

COVER	1
KATA PENGANTAR	3
DAFTAR ISI	4
SEJARAH	5
BIGDATA DI INDONESIA	60
BLOCKCHAIN	75
DAFTAR PUSTAKA	94

Sejarah Big Data dan perkembangan trend Big Data

Pada tahun 2005 “teknologi big data” telah dipergunakan namun belum ada penamaan untuk teknologi tersebut akhirnya teknologi tersebut diberikan nama oleh Roger Mougals bernama “big data”, big data yang dimaksudkan adalah kumpulan dari big data yang berukuran besar yang pada saat itu hampir tidak mungkin dikelola dan diproses oleh traditional business intelligence tools pada saat itu, pada tahun yang sama big data platform Hadoop yang dapat menangani Big Data diluncurkan, hadoop dibuat berdasarkan open source software framework yang dinamakan Nutch dan di gabungkan dengan Google’s MapReduce.

Namun agar dapat lebih mengerti asal muasal big data secara lebih mendalam Timeline Sejarah Big Data secara lengkap

Sebenarnya catatan awal tentang penggunaan data untuk melacak dan mengendalikan bisnis berasal dari 7000 tahun yang lalu ketika akuntansi diperkenalkan di Mesopotamia untuk mencatat pertumbuhan tanaman dan ternak. Ilmu akuntansi terus meningkat,

Awal mula sejarah big data dimulai pada tahun 1663

sejarah big data di mulai pada tahun 1663 John Graunt mencatat dan memeriksa semua informasi tentang penyebab kematian di London. John ingin mendapatkan pemahaman dan membangun sistem peringatan untuk wabah pes yang sedang berlangsung saat itu.

Dalam catatan analisis data statistik pertama yang tercatat, ia mengumpulkan temuannya dalam buku *Natural and Political Observations Made on the Bills of Mortality*, yang memberikan wawasan besar tentang penyebab kematian pada abad ketujuh belas. Karena karyanya tersebut, John Graunt dapat dianggap sebagai bapak statistik.

Era Tahun 1887

Pada tahun 1887 Herman Hollerith menemukan mesin komputasi untuk dapat membaca lubang yang dibuat pada kartu kertas untuk mengatur data sensus.

Sejarah Big Data pada Tahun 1937

Proyek mengenai data yang cukup besar pertama dibuat pada tahun 1937 milik Administrasi Franklin D. Roosevelt di Amerika Serikat. Proyek ini muncul setelah Undang-Undang Jaminan Sosial menjadi hukum yang berlaku pada tahun 1937, akibat proyek ini pemerintah harus melacak kontribusi 26 juta orang Amerika dan lebih dari 3 juta pemberi kerja. Maka dari itu IBM dipercaya untuk menyelesaikan proyek besar ini dengan membuat mesin hole punch card.

Era Tahun 1943

Mesin pengolah data pertama kali muncul pada tahun 1943 yang dikembangkan oleh Inggris untuk memecahkan kode tentara Nazi selama Perang Dunia ke 2. Perangkat ini diberi nama Colossus yang bertugas

untuk mencari pola dalam pesan yang disadap oleh inggris. Perangkat ini dapat membaca 5000 karakter per detik yang mampu mengurangi waktu pengerjaan yang tadinya membutuhkan berminggu-minggu hanya menjadi dalam hitungan jam.

Era Tahun 1952

Badan Keamanan Nasional (NSA) Amerika dibentuk pada tahun 1952 dan dalam waktu lebih dari 10 tahun, mereka telah mengontrak 12.000 orang ahli kriptografi karena NSA dihadapkan dengan data yang sangat banyak selama perang dingin.

Era Tahun 1965

Pada tahun 1965 Pemerintah Amerika Serikat memutuskan untuk membangun pusat data pertama untuk menyimpan lebih dari 742 juta pengembalian pajak dan 175 juta set sidik jari dengan mentransfer semua catatan tersebut ke pita komputer magnetik yang harus disimpan di satu lokasi. Tidak lama proyek ini kemudian dihentikan, tetapi secara umum diterima bahwa itu adalah awal dari era penyimpanan data elektronik.

Sejarah Big data yang cukup pesat pada Tahun 1989

Sejarah big data cukup berkembang pada tahun 1989 karena di tahun ini ilmuwan komputer Inggris Tim Berners-Lee akhirnya menciptakan World Wide Web. mereka bertujuan untuk memfasilitasi proses berbagi informasi menggunakan sistem 'hypertext'.

Era Tahun 1995

Pada tahun 1995 data yang ada di dunia sangat banyak karena semakin banyak perangkat yang terhubung ke internet baik itu IoT devices ataupun berbentuk pc. Pada tahun itu juga superkomputer pertama dibangun yang mampu melakukan banyak pekerjaan dalam satu detik daripada kalkulator yang dioperasikan oleh satu orang dalam 30.000 tahun.

Era Tahun 2005

Pada tahun 2005 Roger Mougals dari O'Reilly Media menciptakan istilah Big Data untuk pertama kalinya, Big Data ini mengacu pada kumpulan data besar yang hampir tidak mungkin untuk dikelola dan diproses menggunakan alat intelijen bisnis tradisional.

2005 juga merupakan tahun dimana Hadoop diciptakan oleh Yahoo! dibangun di atas Google MapReduce. Tujuannya adalah untuk mengindeks seluruh World Wide Web dan saat ini Hadoop open-source digunakan oleh banyak organisasi untuk mengolah data dalam jumlah besar.

Munculnya database biometrik terbesar di dunia pada sejarah big data di tahun 2009

Pada tahun 2009 pemerintah India memutuskan untuk melakukan pemindaian iris mata, sidik jari, dan foto dari

semua 1,2 miliar penduduk tersebut. Semua data ini disimpan dalam database biometrik terbesar di dunia.

Perkembangan Era Tahun 2010

Pada tahun 2010 Eric Schmidt berbicara pada konferensi Techonomy di Lake Tahoe di California dan dia menyatakan bahwa “ada 5 exabyte informasi yang diciptakan oleh seluruh dunia antara awal peradaban dan 2003.”

Perkembangan Tahun 2011

Pada tahun 2011 laporan McKinsey tentang Big Data: “The next frontier for innovation, competition, and productivity”, menyatakan bahwa pada tahun 2018 Amerika Serikat akan menghadapi kekurangan 140.000 – 190.000 ilmuwan data scientist dan juga 1,5 juta data manager.

Pada tahun yang sama Facebook meluncurkan Open Compute Project untuk membagikan spesifikasi untuk Data Center yang menggunakan energi secara efisien.

Perkembangan platform big data pada Tahun 2013

Docker diluncurkan sebagai Open source OS Container Software

Perkembangan data center untuk Big Data pada Tahun 2015

Google dan Microsoft memimpin pembangunan data center secara masif

Perkembangan cloud storage di china pada Tahun 2017

Huawei dan tencent bergabung dengan Alibaba membangun Data Center di China

Perkembangan 2018

Pemimpin pasar di dunia data center menggunakan jaringan 400G yang dimana jaringan ini dapat melakukan transfer secepat 400 Gigabyte per detik.

Edge computing lahir di Tahun 2020

Edge Computing mulai muncul dan akan mengubah role “cloud” di sektor utama ekonomi

Transfer rate super cepat pada Tahun 2021

Data center menggunakan jaringan 1000G yang dimana jaringan ini dapat melakukan transfer secepat 1000 Gigabyte per detik

Edge computing Tahun 2025

Data center akan bertambah dan letaknya lebih mendekati ke device untuk mengakomodir kebutuhan edge computing.

Big Data : Panduan lengkap, Pengertian, manfaat, platform

Pengertian Big Data

Big data adalah sekumpulan data digital yang terdiri dari data terstruktur, semi terstruktur dan tidak terstruktur dalam jumlah yang sangat besar yang dihasilkan dan dikumpulkan oleh organisasi yang bertujuan untuk diolah dan dianalisis sehingga menghasilkan informasi yang berharga yang dipergunakan dalam pengambilan keputusan dan masih ada beberapa pengertian big data menurut para ahli lainnya

Pengertian Big data analytics

Big data analytics merupakan 2 gabungan kata yaitu big data dan analytics, arti kata big data sudah dibahas pada sebelumnya dan analytics adalah proses dari menemukan, mengkomunikasikan, menginterpretasi dan mencari pola pada data dengan menggunakan metode komputasi / perhitungan yang sistematis. maka dari itu big data analitik adalah kegiatan dari menemukan pola yang tersembunyi, korelasi pada kumpulan data yang memiliki tipe data dan sumber data yang bervariasi serta memiliki

ukuran yang sangat besar dari ukuran terabytes sampai zettabytes.

Karakteristik Big Data 5v

Big data pada awalnya hanya mempunyai 3 karakteristik pada big data 3v, namun kemudian berkembang menjadi big data 4v dan yang terbaru menjadi big data 5v, berikut karakteristik dan perkembangannya secara urut

Big Data 3V

1. *Volume*
2. *Variety*
3. *Velocity*

Big Data 4V

4. *Veracity*

Big Data 5v

5. *Value*

Untuk lebih detailnya bisa dapat dibaca di karakteristik big data Big Data 3V, Big Data 4v dan Big Data 5v

Sejarah Big Data

Pada tahun 2005 “teknologi big data” telah dipergunakan namun belum ada penamaan untuk teknologi tersebut akhirnya teknologi tersebut diberikan nama oleh Roger Mougals bernama “big data”, big data yang dimaksudkan adalah kumpulan dari big data yang

berukuran besar yang pada saat itu hampir tidak mungkin dikelola dan diproses oleh traditional business intelligence tools pada saat itu, pada tahun yang sama big data platform Hadoop yang dapat menangani Big Data diluncurkan, hadoop dibuat berdasarkan open source software framework yang dinamakan Nutch dan di gabungkan dengan Google's MapReduce

Sebenarnya sejarah big data cukup panjang, untuk dapat membaca versi lengkap dari sejarah big data anda dapat membacanya disini : Sejarah big data

Pemanfaatan Big Data dan Use Case Big Data

Mengapa big data sangat penting, hal itu karena Big data banyak memiliki manfaat, beberapa manfaat yang akan perusahaan dapatkan ketika memanfaatkan big data diantaranya

Manfaat big data :

1. *Optimalisasi biaya perusahaan*
2. *Meningkatkan efektifitas dan efisiensi bisnis*
3. *Product development*
4. *Mendapatkan insight mengenai pelanggan dan pasar*

Sektor sektor yang dapat menerapkan big data :

1. *Healthcare / Kesehatan*
2. *Perbankan*
3. *Ritel / Retail*

4. *Media dan Hiburan*
5. *Agriculture / agrikultur*
6. *Manufaktur*
7. *Minyak dan Gas*
8. *Telekomunikasi*
9. *Pendidikan*
10. *Sains*
11. *Pemerintah*
12. *Teknologi*
13. *Asuransi*
14. *Cybersecurity*

Dan masih banyak lagi, Untuk lebih detail dalam membaca solusi big data kamu bisa merujuk pada artikel ini : Solusi big data : pemanfaatan, use case big data bigbox

Perusahaan big data, big data platform dan big data platform analytics

Untuk mengatasi permasalahan pada big data, banyak perusahaan yang mulai membangun big data platforms dan juga big data platforms analytics diantaranya

Perusahaan big data di Indonesia dan di dunia :

1. *Hadoop*
2. *MapR.*
3. *Bigbox*
4. *Microsoft Azure.*
5. *Cloudera.*
6. *Sisense.*

7. *Collibra.*
8. *Tableau.*
9. *Sparks*
10. *Qualtrics.*
11. *Oracle.*

Untuk lebih detailnya anda dapat membaca

Big Data Platforms : kupas tuntas perusahaan serta big data software nya

Bagaimana big data bekerja

Sebenarnya proses bagaimana big data bekerja cukup panjang, namun ada 3 proses utama yang paling penting

1. Integrasi data

Big data dikumpulkan dari berbagai sumber, bisa dari data transaksi, data pelanggan, foto-foto produk di ecommerce, video-video di social media data lainnya. Data-data tersebut sangatlah banyak baik dari jumlah, jenis serta sumber datanya maka dari itu perlu diintegrasikan kepada sistem big data sehingga dapat diolah dengan baik.

2. Management data

Setelah data tersebut dikumpulkan, tentu kita perlu mengelola data-data tersebut, seperti dimana kita akan mengorganisir dan menyimpan data tersebut, untuk menyimpan data kita dapat menyimpan data tersebut baik

di server on premise, cloud ataupun hybrid / metode gabungan antara on premise dan juga cloud storage.

3. Analysis

Setelah data telah terkumpul dan terkelola dengan baik, selanjutnya perlu, melakukan eksplorasi data, mencari pola, mengelompokkan data-data sesuai dengan polanya dan melakukan analisa pada data-data tersebut sehingga mendapat informasi yang kita butuhkan.

Untuk lebih detail mengenai bagaimana big data bekerja anda dapat membaca Bagaimana Big data Bekerja

Arsitektur Big data

Arsitektur Big data adalah struktur keseluruhan yang merepresentasikan dari logical dan physical sistem dari big data,

Untuk lebih tahu lebih dalam terkait arsitektur big data, silahkan mengunjungi artikel ini arsitektur big data

Bagaimana cara membangun big data

langkah langkah dalam membangun atau mengimplementasi big data :

1. *Menganalisa masalah*
2. *Memilih Vendor*
3. *Strategi Deployment*
4. *Capacity Planning*

5. *Infrastructure sizing*
6. *Pembuatan Rancangan Disaster Recovery*

Untuk lebih detailnya silahkan klik artikel berikut :
membangun / mengimplementasi big data

Big Data Best Practices and Tips & Trick

1. Mendefinisikan big data business goals dengan matang
2. Melakukan asesmen dan membuat strategi dengan partner ataupun vendor yang dapat dipercaya
3. Tentukan data apa saja yang sudah perusahaan miliki dan juga kemungkinan di masa depan sehingga dapat dengan mudah di scaling
4. Terus menerus berkomunikasi dengan semua stakeholder dan tim IT ataupun bisnis, pastikan scope tidak berubah-ubah dan sudah direncanakan dengan matang.
5. Mulai dengan perlahan-lahan dan hati-hati serta tidak terlalu ambisius, namun harus bereaksi cepat ketika ada perubahan terutama trend.
6. Mengevaluasi big data arsitektur dan hardware secara matang.
7. Membuat multi deployment dan sebaiknya melakukan hybrid untuk mencegah hal yang tidak diinginkan
8. Pelajari hukum terkait data di negara-negara bisnis perusahaan berjalan
9. Kelola tim big data anda, karena tim yang mengelola big data sangat krusial bagi perusahaan.

Permasalahan dan Tantangan serta solusi big data

Dalam mengimplementasi big data, tentu perusahaan pasti akan menghadapi berbagai tantangan diantaranya :

1. *Kurangnya pemahaman dan wawasan big data / lack of understanding and knowledge related to big data*
2. *Permasalahan pertumbuhan data / Data growth issues*
3. *Sulitnya menjaga kualitas data / Difficulty in Managing & maintain data quality*
4. *Kurangnya pengalaman dalam mengalami data yang berjumlah banyak dan kompleks*
5. *mengalami permasalahan dalam mengambil dan mengintegrasikan dari data sources*
6. *Permasalahan upscaling / Upscaling issues*
7. *Bingung dalam memilih teknologi dan platform big data*
8. *Permasalahan budget karena investasi big data memerlukan biaya yang cukup besar*
9. *Masalah keamanan data*

Dari berbagai tantangan tersebut tentu ada solusi dalam menghadapi tantangan dan permasalahan big data tersebut, untuk lebih detail silahkan membaca artikel Permasalahan dan Tantangan serta solusi big data

Pekerjaan terkait big data.

Jika tertarik terhadap teknologi big data dan ingin terjun langsung dalam big data berikut pekerjaan yang berkaitan dengan big data :

1. *Data Analyst*
2. *Data Scientist.*
3. *Data Engineer*
4. *Programmer .*
5. *Big Data Engineer.*
6. *Machine Learning Scientist.*
7. *Business Analytics Specialist.*
8. *Data Visualization Developer.*
9. *Business Intelligence (BI) Engineer.*
10. *BI Solutions Architect.*
11. *Machine Learning Engineer*
12. *AI Engineer*
13. *Statistician*

Pengertian Big Data, Sejarah, Manfaat, Cara Kerja, dan Contoh Penerapannya

Big data atau data besar tentu bukan lagi istilah asing, sebab sudah disebut di banyak media dan digunakan secara luas atau kompleks. Terutama bagi pelaku usaha, yakni bagi para pengembang aplikasi dan perusahaan yang sudah memiliki ekspansi cukup luas. Penggunaan data raksasa tersebut kemudian menjadi hal penting dan menjadi aset.

Hal ini sejalan dengan yang disampaikan oleh Neelie Kroes pada acara bertajuk *Press Conferences on Open*

Data Strategy di tahun 2011 lalu. Kroes dalam acara tersebut, di hadapan banyak reporter menyebutkan bahwa *data is the new gold*” atau data adalah emas baru.

Artinya, data di masa atau era digital revolusi industri 4.0 seperti sekarang sama berharganya dengan emas sehingga oleh Kroes disebut sebagai emas baru atau emas dalam bentuk baru. Pernyataan yang disampaikan oleh Kroes tentu bukan fiktif belaka. Melainkan didasarkan pada sebuah fakta, dimana banyak perusahaan bisa tumbuh dan berkembang lewat data tersebut.

Hal serupa kemudian juga disampaikan oleh Presiden Joko Widodo di tahun 2019 lalu. Melalui Pidato Kenegaraan pada 16 Agustus 2019, Presiden Jokowi menuturkan bahwa “data adalah jenis kekayaan baru bagi bangsa kita, dan kini data lebih berharga dari minyak.”.

Hal ini kemudian menegaskan kepada kita semua bahwa memiliki data raksasa atau dalam jumlah besar adalah aset berharga. Data yang jumlahnya banyak dan kemudian disebut sebagai big data ini mungkin berisi data-data yang penting dan bercampur dengan data tidak penting. Namun poin terpenting bukan jumlah dan tingkat kepentingan data.

Melainkan pada proses pengelolaan data besar tersebut. Jika perusahaan bisa mengelolanya dengan baik, maka akan mendukung perkembangan perusahaan tersebut. Begitu juga sebaliknya, jika kurang perhatian atau salah langkah maka data hanyalah berupa daftar data tanpa arti atau makna.

Lalu, bagaimana memanfaatkannya? Bisa dimulai dengan mengenal data raksasa itu sendiri dengan detail. Simak penjelasan lengkapnya di bawah ini.

Hal pertama yang perlu diketahui mengenai data besar adalah pengertiannya, sehingga bisa mempelajarinya dengan lebih mudah. Sebab sudah paham dasar dari materi mengenai data berukuran raksasa tersebut.

Big data memiliki definisi sebagai suatu istilah yang menggambarkan mengenai volume data yang terbilang besar, baik itu terstruktur maupun tidak terstruktur yang membanjiri bisnis sehari-hari. Pelaku bisnis melalui perusahaannya dijamin akan rutin mendapatkan data setiap harinya, bahkan di setiap jam dan detik.

Jenis data yang didapatkan perusahaan sangat beragam, maka disebut sebagai data yang terstruktur maupun tidak. Data ini kemudian diolah oleh perusahaan tersebut dimana sudah menunjukan orang atau operator khusus. Umumnya perusahaan akan memakai alat bantu atau *tools* sehingga proses mengolah data menjadi lebih mudah, efisien, dan juga efektif.

Sebab sekali lagi, data yang didapatkan perusahaan sebesar atau sebanyak apapun data tersebut bukan poin penting. Poin pentingnya adalah bagaimana si perusahaan ini bisa mengolah data tersebut untuk mendorong pendapatan atau profit perusahaan. Sehingga bisa bertahan di tengah persaingan dan bisa terus berkembang.

Data yang diolah sedemikian rupa dengan teknik dan *tools* tertentu akan membantu perusahaan menentukan strategi atau langkah selanjutnya. Langkah yang diambil ini tentunya akan mempengaruhi masa depan perusahaan tersebut. Apakah akan bertahan, berkembang, atau justru tumbang.

Sejarah Perkembangan Big Data

Melalui penjelasan di atas tentu bisa diketahui bahwa peran dari data besar ini sangat krusial. Namun, bagaimana data semacam ini bisa memiliki peran atau mungkin bagaimana bisa dikenal oleh para pelaku usaha? Perlu diakui bahwa istilah ini semakin populer ketika bisnis berbasis internet atau online semakin berkembang.

Maka bisa mengintip sejarah perkembangannya seperti apa, karena sampai saat ini kebanyakan orang memahami bahwa data raksasa adalah suatu istilah baru. Begitu pula dengan pengelolaan atau pemanfaatannya untuk memberi keuntungan atau manfaat bagi perusahaan.

Sejarah dari data berukuran raksasa ini adalah dimulai pada tahun 2000-an. Yakni ketika seorang analis industri bernama Doug Laney menyampaikan konsep mengenai data berukuran besar tersebut. Laney mengartikulasikan istilah data berukuran besar tersebut menjadi Tiga V. Yaitu:

1. Volume

Istilah V pertama mengarah pada kata volume atau jumlah data, yakni kegiatan dimana perusahaan mulai mengumpulkan data sebanyak mungkin. Data ini didapatkan dari banyak sumber seperti transaksi bisnis, perangkat pintar, media sosial, video, peralatan industri, dan masih banyak lagi sumber data lainnya.

Proses mengumpulkan data ini tentunya terasa sulit di masa lalu, terutama ketika istilah big data masih menjadi istilah asing di telinga. Namun pesatnya perkembangan teknologi membuat banyak layanan pengelolaan data bermunculan. Sehingga prosesnya menjadi lebih mudah dan efisien.

2. Velocity

Istilah V berikutnya yang dipaparkan oleh Laney adalah *velocity*. Istilah ini merujuk pada kecepatan aliran data, sehingga data berukuran besar (volume) kemudian mengalir dengan kecepatan tertentu ke media penyimpanan atau memori. Semakin besar atau semakin banyak jumlah datanya maka aliran ini akan semakin cepat.

3. Varietas

V terakhir yang disampaikan oleh Laney adalah mengarah pada istilah varietas. Yakni mengarah pada aneka macam atau jenis data yang berhasil didapatkan oleh perusahaan

melalui berbagai sumber yang dijelaskan di poin sebelumnya.

Jenis data di dalam big data kemudian semakin kesini semakin beragam. Kebanyakan bentuknya menjadi tidak terstruktur, sedangkan untuk data tradisional sifatnya lebih terstruktur sehingga mudah untuk dikelola dan kemudian dimanfaatkan.

Adapun jenis data yang tidak terstruktur dan kemudian perlu diproses lagi untuk bisa digunakan atau dimanfaatkan. Adalah seperti data dalam bentuk video, audio, dan juga dalam bentuk teks.

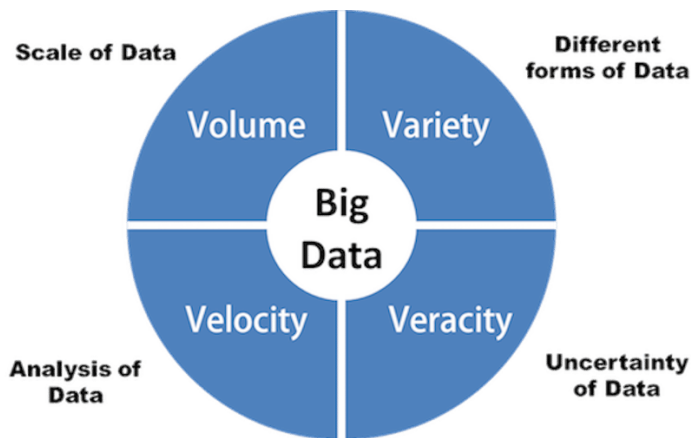
Perlahan kesadaran adanya jumlah data yang sangat besar dan berasal dari banyak sumber, kemudian membuat banyak orang mencoba mengelolanya dengan baik. Adapun sumber data ini umumnya dari media sosial seperti Youtube, Instagram, Facebook, dan sebagainya. Kemudian bisa juga berasal dari website perusahaan dan transaksi offline.

Hal ini kemudian membuat sejumlah *open source* di tahun 2005 mulai mengembangkan kegiatan analisis data, seperti Hadoop dan juga NoSQL. Kegiatan analisis big data tersebut kemudian membantu banyak perusahaan untuk mengelola data berukuran besar yang didapatkan agar menjadi lebih mudah dan cepat.

Perkembangan data besar kemudian terus berjalan, hingga tercetus istilah *Internet of Things* (IoT). Hal ini kemudian semakin mendorong setiap perusahaan untuk

memaksimalkan tata kelola data yang didapatkan secara online. Tujuan utamanya adalah untuk mengukur efektivitas produk maupun jasa yang disediakan terhadap kehidupan konsumen.

Sehingga bisa diketahui, apakah produk yang disediakan memang sudah tepat atau perlu dikoreksi. Pengelolaan data kemudian terus dilakukan agar perusahaan bisa terus inovatif dalam membuat produk dan jasanya bisa lebih berkembang. Sehingga bisa berjalan beriringan dengan kebutuhan dan keinginan konsumennya.



sumber :7wdata.be

Sumber Didapatkan Big Data

Data besar atau big data kemudian seperti yang disampaikan di awal bisa ditemukan di berbagai sumber. Saat ini terdapat beberapa sumber yang terbilang paling sering memberi kontribusi data dalam jumlah besar dan memiliki kualitas baik. Diantaranya adalah:

1. Internet

Sumber pertama adalah internet, dan merupakan penyedia data raksasa paling besar dan paling sering dimanfaatkan perusahaan. Sebab cara kerja internet sendiri adalah menyimpan semua data dari para penggunanya. Jadi, aktivitas pencarian apapun nantinya akan disimpan dan kemudian menjadi kumpulan data.

Inilah alasan kenapa *cache* pada browser maupun perangkat yang dipakai berselancar di internet mudah sekali penuh. Sebab isinya adalah data-data dari hasil aktivitas berselancar di internet. Data ini oleh suatu perusahaan bisa diolah dan menjadi media menentukan strategi perusahaan.

2. Smartphone

Smartphone juga menjadi penyumbang data dalam skala besar, sebab aktivitas apapun di dalamnya akan otomatis disimpan oleh perangkat. Sumber data terbesar di smartphone sendiri adalah aplikasi yang diinstal di dalamnya. Jadi, setiap kali aplikasi ini berjalan maka data akan didapatkan dan dikumpulkan oleh sistem.

3. Media Sosial

Media sosial dimana menjadi media yang banyak dibuka siapa saja di era sekarang bahkan sesaat setelah bangun tidur juga menjadi sumber data dalam skala besar. Setiap foto, video, termasuk cuitan di media sosial sudah disebut

data. Bisa dibayangkan berapa miliar data didapatkan perusahaan saat mengelola media sosial?

4. Digitalisasi

Digitalisasi media tertentu kemudian menciptakan sumber data dan tentu bisa dikelola oleh suatu perusahaan. Misalnya saja perusahaan menyediakan data dalam bentuk musik, yang kemudian dirilis ke masyarakat dalam bentuk audio dan bisa dinikmati di sejumlah *platform* atau aplikasi.

Misalnya di Joox, maka setiap kali orang mendengarkan musik tersebut di Joox perusahaan akan mendapatkan data. Data inilah yang kemudian dikumpulkan dan diolah untuk menentukan strategi perusahaan di masa mendatang.

Namun data ini tentu bukan data pribadi dan spesifik seperti nama, usia, tempat tinggal, dan sebagainya. Namun data kasar, misalnya untuk musik tadi perusahaan bisa mengetahui musik mana saja yang ketika dirilis banyak didengarkan.

Maka perusahaan bisa mengolah data tersebut untuk menentukan genre musik apa yang akan dikembangkan lagi di masa mendatang. Jadi, saat menggunakan aplikasi dan smartphone tidak perlu cemas data pribadi bocor kemana-mana.

Manfaat Big Data

Manfaat dari big data sendiri kemudian sangat beragam dan membuatnya semakin dibutuhkan oleh perusahaan. Perusahaan kemudian akan berusaha sebaik mungkin mengolah data-data yang didapatkan agar bisa berguna di kemudian hari. Adapun manfaat tersebut adalah:

– Bidang Analytic

Manfaat yang pertama datang dari kegiatan analisis atau *analytic* yang dilakukan perusahaan, khususnya yang meluncurkan website atau *platform* dan aplikasi. Adanya kumpulan data besar membantu perusahaan untuk menemukan masalah penyebab kegagalan dari website dan aplikasi yang diluncurkan.

Data ini juga bisa digunakan perusahaan untuk menemukan suatu anomali atau perilaku menyimpang di dalam struktur bisnis. Sehingga data ini bisa menemukan suatu kesalahan dan masalah, dan membantu perusahaan menemukan solusi terbaik untuk mengatasinya dan kemudian menyempurnakan aplikasi atau website.

– Bidang Bisnis

Dalam bidang bisnis, data berskala besar ini juga memberi manfaat yang sangat besar. Seperti membantu perusahaan meningkatkan sistem operasional bisnis, menyusun CRM (*Customer Relationship Management*) yang tepat, dan juga meningkatkan pengalaman pengguna aplikasi yang diluncurkan dengan segala perbaikannya.

– Bidang Informasi

Big data kemudian juga bermanfaat dalam mengelola dan menyediakan media sosial terbaik, dan membantu lebih banyak perusahaan menemukan data yang sesuai kebutuhan. Selain itu, data besar akan membantu setiap perusahaan termasuk suatu negara untuk menciptakan perangkat cerdas yang memudahkan aktivitas manusia.

Cara Kerja Big Data

Melalui uraian di atas tentu sering mendengar istilah bahwa data berskala besar ini bermanfaat bagi perusahaan. Kenapa bisa demikian? Memang kumpulan data bisa mendorong perusahaan untuk menentukan keputusan atau strategis bisnis yang tepat.

Supaya kamu para pelaku usaha bisa memanfaatkannya dengan baik, maka wajib paham juga cara kerja data raksasa ini seperti apa. Berikut detailnya:

1. Integrasi Data

Data yang kebanyakan didapatkan di era digital merupakan data tidak terstruktur yang tentu perlu diproses secara bertahap untuk kemudian bisa dimanfaatkan dengan baik. Jadi, cara kerja data ini dimulai dengan mengatur integrasi data sebaik mungkin.

Misalnya mengumpulkan semua data berdasarkan kategori bisa dari kumpulan komentar pelanggan di media sosial, unggahan video pelanggan dengan produk, dan

sebagainya. Setelah data sudah dibuat terstruktur maka baru bisa diproses oleh bagian analisis data di perusahaan.

2. Manage Data

Berikutnya adalah mengatur dan menyimpan data yang sudah terstruktur tadi ke dalam media penyimpanan terbaik. Saat ini media yang paling banyak dianjurkan adalah media penyimpanan online atau *cloud*. Sehingga bisa memiliki kapasitas besar dan kemudian bisa diakses dari mana saja dan kapan saja.

3. Analisis Data

Tahap berikutnya adalah melakukan analisis data, dan jika dilakukan manual maka akan memakan waktu dan tenaga sekaligus pikiran. Maka perusahaan masa kini sudah menggunakan *tools* atau alat bantu untuk analisis data. Supaya big data yang jumlahnya susah dihitung bisa dianalisis dengan baik dan efektif.

Contoh Big Data dalam Kehidupan Sehari-Hari

Kemudian untuk aplikasi atau contoh penerapan dari data besar dalam kehidupan sehari-hari sudah tentu cukup banyak. Pastikan pula diterapkan oleh berbagai perusahaan di berbagai bidang. Adapun sedikit contoh aplikasinya adalah sebagai berikut:

– Perusahaan Perbankan

Perusahaan perbankan kini sudah jamak memanfaatkan olah data dalam skala besar untuk menyediakan produk dan layanan yang semakin beragam dan berkualitas. Perbankan akan mendapatkan data dari aplikasi perbankan maupun transaksi perbankan offline di berbagai kantor cabang.

Hal ini memudahkan bank mengetahui apa saja yang dilakukan nasabah terhadap produk dan layanan yang disediakan. Supaya kedepannya perbankan tersebut bisa meningkatkan kualitas layanan. Sehingga semakin banyak nasabah yang merasa terbantu dan menjadi pengguna jasa yang loyal.

– Perusahaan Retail

Banyak perusahaan retail kini menawarkan layanan membership, menyediakan aplikasi belanja online, membangun website, dan lain sebagainya. Semau ini membantu perusahaan mendapatkan data berskala besar tidak hanya dari transaksi langsung di lokasi retail.

Namun juga mendapatkan data dari sumber lain, sehingga membantu perusahaan retail ini untuk berkembang. Misalnya menentukan promosi seperti apa yang tepat, menarik, dan sesuai kebutuhan konsumen. Kemudian produk apa saja yang termasuk jenis produk baru dan perlu segera di disediakan.

Perusahaan retail kemudian bisa terus berkembang dan bertahan di tengah kencangnya arus persaingan. Sebab mampu mengumpulkan dan mengolah data berukuran besar dengan baik.

Masih banyak lagi penerapan dari big data untuk berbagai kegiatan perusahaan di berbagai bidang. Mayoritas data-data ini digunakan untuk menentukan arah kebijakan dan strategis marketing perusahaan. Sehingga bisa menyediakan produk yang terus variatif, inovatif, sekaligus memberi layanan yang terus berkualitas.

Saat ini, big data adalah salah satu bagian yang sangat penting dan dibutuhkan dalam berbagai macam bidang. Big data merupakan istilah yang digunakan untuk menggambarkan sebuah volume data cukup besar, baik terstruktur maupun tidak terstruktur. Di mana kita bisa menemukan big data?

Jadi big data ini digunakan di berbagai macam bidang, salah satunya adalah penggunaan media sosial. Jika kamu mengupload foto dan video di Instagram, tentu menggunakan big data untuk penyimpanannya. Tidak hanya itu, tetapi big data ini juga digunakan dalam toko online, dimana memudahkan para pembeli untuk mencari berbagai macam kebutuhan barang yang diinginkan.

Sejarah Big Data Pertama Kali

Big data adalah salah satu hal yang sangat penting di zaman modern seperti ini karena bisa mendukung berbagai macam bidang. Bagaimana awal mulanya big

data ini ditemukan? Jadi istilah big data ini semakin populer ketika adanya penggunaan bisnis berbasis internet atau online semakin berkembang.

Namun sebenarnya, sejarah perkembangan big data ini dimulai pada tahun 2000-an dimana seorang analis industri bernama Doug Laney menyampaikan konsep terkait data dengan ukuran besar. Dari tahun 2000 tersebut, nama big data sudah mulai digunakan dan dikenal dengan berbagai macam kebutuhan.

Doug Laney menyampaikan istilah data berukuran besar tersebut menjadi 3V, yaitu;

Volume

Istilah pertama yang digunakan adalah volume atau jumlah data, yaitu kegiatan dimana perusahaan mulai mengumpulkan data sebanyak mungkin. Data ini didapatkan dari berbagai macam sumber seperti transaksi bisnis, media sosial, perangkat pintar, video, peralatan industri, dan masih banyak beberapa sumber lainnya.

Proses mengumpulkan data ini pada awalnya terasa sangat sulit, apalagi ketika istilah dari big data ini menjadi sesuatu yang asing di telinga. Namun dengan berkembangnya teknologi saat ini, tentu membuat layanan pengelolaan data bermunculan sehingga prosesnya menjadi lebih mudah dan efisien.

Velocity

Selanjutnya ada bagian velocity atau kecepatan aliran data. Kecepatan aliran data menjadi salah satu hal yang penting sebab data berukuran besar akan mengalir dengan kecepatan tertentu dari media penyimpanan atau memori. Semakin besar atau semakin banyak jumlah data, maka alirannya akan semakin cepat.

Varietas

Varietas mengarah pada aneka macam atau jenis data yang berhasil didapatkan dari berbagai macam sumber. Jenis big data ini kemudian semakin beragam, bahkan bentuknya menjadi tidak terstruktur. Sedangkan untuk data tradisional biasanya lebih terstruktur, sehingga mudah untuk dikelola dan dimanfaatkan.

Perkembangan big data kemudian terus berjalan hingga tercetus Internet of Things atau IoT. Hal tersebut digunakan untuk mendorong setiap perusahaan memaksimalkan tata kelola data sehingga mengukur efektivitas produk menjadi lebih mudah.

Dengan menggunakan hal tersebut, produk yang disediakan oleh perusahaan bisa diketahui secara tepat tanpa perlu adanya koreksi. Pengelolaan data kemudian dilakukan agar perusahaan bisa terus berjalan secara inovatif dalam membuat produk maupun jasa agar lebih berkembang.

Penggunaan Big Data dalam Kehidupan Sehari-Hari

Penggunaan big data saat ini sudah digunakan dalam kehidupan sehari-hari, namun tidak banyak orang yang menyadari hal tersebut. Big data ini sudah diterapkan dari berbagai macam perusahaan di bidang tertentu, seperti berikut.

- Perusahaan perbankan saat ini sudah memanfaatkan olah data dalam skala besar untuk menyediakan produk maupun layanan semakin beragam dan berkualitas. Dengan penggunaan big data tersebut, tentu pihak perbankan akan mendapatkan data dari aplikasi maupun transaksi perbankan secara offline di berbagai macam kantor cabang. Hal tersebut akan memudahkan untuk mengetahui apa saja yang sudah dilakukan nasabah terhadap produk dan layanan.
- Perusahaan ritel menawarkan layanan, menyediakan aplikasi belanja online, membangun sebuah website dan lain sebagainya. Hal tersebut untuk membantu perusahaan untuk mendapatkan data berskala besar, tidak hanya melalui transaksi secara langsung di lokasi ritel.

Masih banyak penggunaan big data yang terjadi dalam aktivitas sehari-hari. Big data juga digunakan dalam media sosial yang sering kita gunakan dalam kehidupan sehari-hari loh. Penggunaan big data di zaman modern seperti sekarang ini sangat penting.

Sejarah Big Data dan cara kerja Big Data.

29 Apr 2021

- Popular Articles

Big data pertama kali diperkenalkan oleh seorang analis industri yang bernama Doug Laney, yang menyampaikan terkait 3 masalah utama yang dikenal sebagai *the 3V of data* yang terdiri dari *Volume*, *Velocity* dan *Variety*.

- *Volume*

Merupakan masalah data yang disebabkan karena volume data yang dimiliki oleh sebuah perusahaan atau organisasi atau bisnis sangat besar. Sebagai contoh perusahaan telekomunikasi yang melakukan *record* dari aktivitas *browsing* pengguna dalam satu hari yang dimana di dalam 1 hari bisa tercatat lebih dari 1 juta hingga 1 milyar record data pengguna. Apabila data tersebut di dalam 1 bulan, berarti ada sekitar 30 juta hingga 30 milyar *record* data.

Data tersebut saja bisa disimpan dalam sebuah *warehouse*, namun tentunya biaya yang perlu dikeluarkan tentunya juga besar dan akan menjadi lebih mahal. Hadoop merupakan sebuah platform big data dapat menjadi solusi penyimpanan data – data tersebut dan dapat digunakan juga untuk mengolah dan menganalisa data, sehingga data yang awalnya tidak memiliki ekonomis dapat menjadi nilai tambah dan menghasilkan keuntungan bagi bisnis.

- *Velocity*

Merupakan masalah data yang disebabkan karena kecepatan data yang akan dihasilkan. Masalah ini muncul biasanya terjadi karena besarnya data yang dimiliki oleh perusahaan, membuat kecepatan data yang akan di proses memiliki waktu yang lama. Masalah ini tidak hanya terjadi pada pengolahan data dengan volume yang besar saja, namun terjadi juga pada data yang masuk secara *realtime* dengan jumlah yang besar sehingga membuat perusahaan membutuhkan sebuah perangkat lunak untuk pemrosesan data tersebut secara *realtime* juga. Sebagai contoh perusahaan di bidang telekomunikasi yang melakukan pemrosesan data pelanggan secara real-time yang ingin mengukur dan mengetahui kualitas koneksi internet. Dapat dibayangkan berapa banyaknya data yang dapat masuk setiap menitnya. Dengan pemanfaatan *big data*, data tersebut dapat disimpan dan di proses ke *software big data* dan dianalisa pada saat itu juga untuk mengetahui kualitas jaringan internet, lalu hasil dari analisa tersebut akan dapat dikirimkan kepada tim teknis jaringan sehingga perusahaan pun mengetahui lokasi mana saja yang memperoleh gangguan koneksi internet.

- *Variety*

Merupakan masalah data yang disebabkan oleh variasi jenis data. Masalah ini merupakan salah satu masalah yang sulit dihadapi apabila perusahaan masih menggunakan sebuah platform data tradisional baik menggunakan database maupun *warehouse*, karena data yang disimpan terdapat berbagai jenis file dan tipe data

yang berbeda, sehingga dalam proses pengolahan data membutuhkan waktu yang lebih lama.

Sebagai contoh tipe data yang sering kita jumpai seperti teks, numerik, video, suara, gambar, dan masih jenis tipe data lainnya, dengan pemanfaatan Hadoop perusahaan dapat melakukan sortir data sesuai dengan tipe file dan jenisnya, kemudian dapat dilakukan Analisa sehingga dapat menghasilkan data yang bernilai ekonomis.

Tentunya dalam pemanfaatan *big data* dalam perusahaan ada beberapa langkah yang perlu dilakukan terlebih dahulu agar implementasi *big data* yang akan dilakukan tidak sia – sia, antara lain :

- Perusahaan atau pelaku bisnis, harus dapat menetapkan strategi yang tepat untuk menggunakan *big data*.
- Perusahaan atau pelaku bisnis harus dapat melakukan identifikasi secara menyeluruh sumber dari *big data*.
- Perusahaan atau pelaku bisnis, harus dapat menentukan terlebih dahulu dalam melakukan pengelolaan dan penyimpanan data apakah menggunakan *warehouse database*, *cloud*, ataupun Hadoop.
- Setelah melakukan hal tersebut, perusahaan atau pelaku bisnis tentu harus melakukan analisa data.
- Perusahaan atau pelaku bisnis harus dapat mengambil keputusan yang tepat dari hasil analisis data yang telah didapatkan sebelumnya.

Big Data adalah sebuah istilah umum yang mengacu pada teknologi dan teknik untuk memproses dan menganalisa sekumpulan data yang memiliki jumlah yang sangat besar, baik yang terstruktur, semi terstruktur dan tidak terstruktur. Ada banyak tantangan yang akan dihadapi ketika berhubungan dengan big data, mulai dari bagaimana data diambil, disimpan, hingga masalah keamanan data.

Walaupun Istilah big data sudah sering didengar dan diucapkan, masih banyak diantara kita yang bertanya-tanya: Apa yang dimaksud dengan big data? Apa kegunaan big data? Apa saja teknologi big data? Mengapa big data diperlukan dalam berbagai bidang?

Apa Yang Dimaksud Dengan Big Data?

Tidak ada definisi yang baku mengenai big data. Secara garis besar big data adalah sekumpulan data yang memiliki jumlah yang sangat besar atau struktur yang kompleks sehingga teknologi pemrosesan data tradisional tidak lagi dapat menanganinya dengan baik. Saat ini istilah big data juga sering digunakan untuk menyebut bidang ilmu atau teknologi yang berkaitan dengan pengolahan dan pemanfaatan data tersebut.

Aspek yang paling penting dari big data sebenarnya bukan sekedar pada berapa besarnya data yang bisa disimpan dan diolah, akan tetapi kegunaan atau nilai tambah apa yang dapat diperoleh dari data tersebut. Jika kita tidak dapat mengekstrak nilai tambah tersebut, maka data hanya

akan menjadi sampah yang tidak berguna. Nilai tambah ini dapat digunakan untuk berbagai hal, seperti meningkatkan kelancaran operasional, ketepatan penjualan, peningkatan kualitas layanan, prediksi atau proyeksi pasar, dan lain sebagainya.

Dalam bidang informatika dikenal sebuah istilah “Garbage in Garbage out” atau masukan sampah akan menghasilkan keluaran sampah juga. Maksudnya adalah jika masukan yang kita berikan ke sistem adalah input berkualitas rendah, maka kualitas outputnya tentu akan rendah juga. Input yang dimaksud di sini adalah data.

Untuk itu, memastikan kualitas input maupun output dalam setiap tahap pengolahan data untuk mendapatkan keluaran akhir yang berkualitas adalah sebuah keharusan dalam implementasi big data.

Karakteristik Big Data

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

- 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.

- 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.
- 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.
- 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.

Apa Saja Teknologi Big Data?

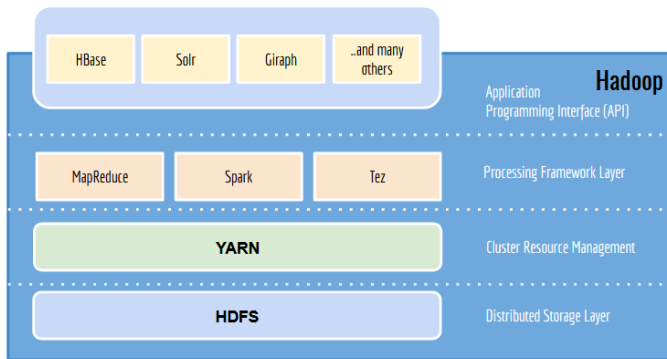
Perkembangan teknologi big data tidak bisa dilepaskan dari teknologi atau konsep open source. Istilah Big Data terus bergaung seiring dengan pesatnya perkembangan teknologi open source yang mendukungnya. Banyak perusahaan besar berkontribusi teknologi big data yang mereka buat dan mereka gunakan ke komunitas open source. Hal inilah yang kemudian menjadi salah satu pendorong utama berkembangnya big data.

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

1. 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:
 - o 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 klaster Hadoop. HDFS terinspirasi dari Google File System.

- Framework pemrosesan data secara paralel dan terdistribusi MapReduce adalah model pemrograman untuk melakukan pemrosesan data besar secara terdistribusi dalam klaster Hadoop. MapReduce bekerja dan mengolah data-data yang berada dalam HDFS.
- Resource management terdistribusi YARN merupakan tools yang menangani resource manajemen dan penjadwalan proses dalam klaster 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 klaster Hadoop. YARN memungkinkan banyak framework komputasi paralel lain, seperti Spark, Tez, Storm, dsb, untuk bekerja diatas klaster Hadoop dan mengakses data-data dalam HDFS.



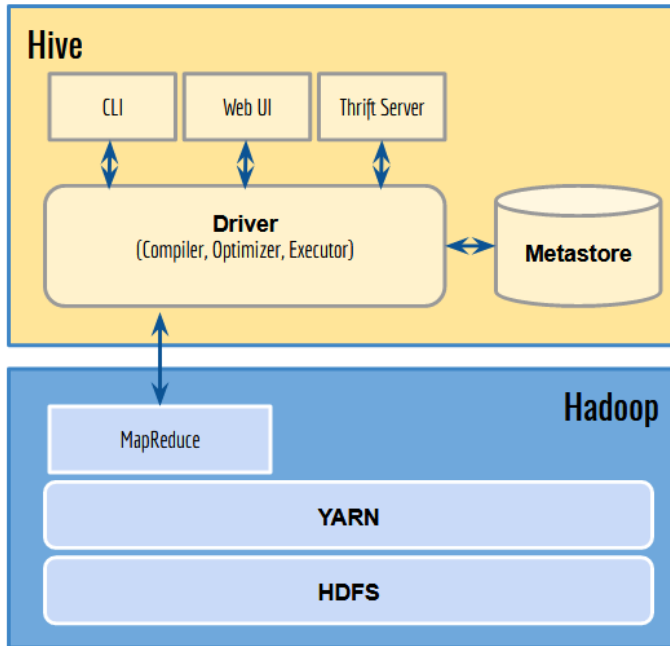
Komponen-komponen Apache Hadoop

2. 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.



Komponen Utama Apache Hive

Pada dasarnya Hive hanya sebuah layer untuk menerjemahkan perintah-perintah SQL ke dalam framework komputasi terdistribusi. Hive dapat bekerja menggunakan berbagai framework yang berjalan diatas Hadoop, seperti MapReduce, Tez ataupun Spark.

3. 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 di atas 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.

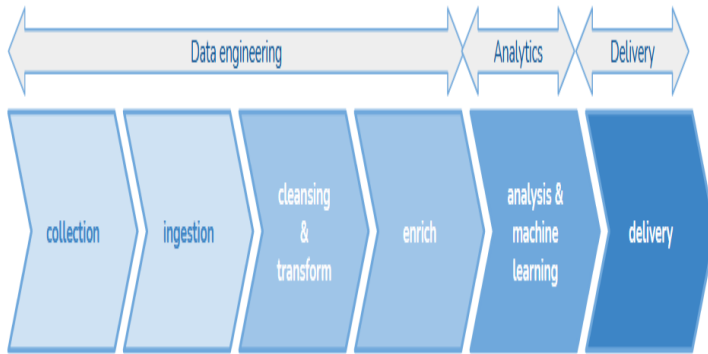
Selain 3 teknologi tersebut, sebenarnya masih sangat banyak teknologi dan framework big data lainnya yang bersifat open source seperti HBase, Cassandra, Presto, Storm, Flink, NiFi, Sqoop, Flume, Kafka dan lain sebagainya.

Big Data Pipeline

Untuk dapat memberikan nilai yang bermanfaat, data harus melalui berbagai tahapan pemrosesan terlebih dahulu. Mulai dari pencatatan/pembuatan, pengumpulan, penyimpanan, pengayaan, analisis dan pemrosesan lebih lanjut, hingga penyajian. Rangkaian proses data ini biasa disebut dengan Data Pipeline.

Secara garis besar Big Data Pipeline dapat dibagi menjadi 3, yaitu :

- - Data Engineering: tercakup di dalamnya data collection, ingestion, cleansing, transformation dan enrichment.
 - Data Analytics / Machine Learning: mencakup feature engineering dan komputasi.
 - Data Delivery: penyajian data, termasuk penerapan model dalam aplikasi atau sistem, visualisasi, dan lain sebagainya.



Data Processing Pipeline

Big Data Analytics

Saat ini jika kita berbicara mengenai big data, maka biasanya yang dimaksud adalah big data analytics. Hal ini cukup wajar, karena ketika sebuah proyek big data dimulai, tentu saja hasil akhir yang diharapkan adalah mendapatkan insight yang bermanfaat, yang dapat membantu pengambilan keputusan.

Data Analytics sendiri adalah serangkaian proses untuk menggali informasi atau insight dari kumpulan data. Informasi tersebut dapat berupa pola, korelasi, trend, dan lain sebagainya. Data analytics seringkali melibatkan teknik dan algoritma pengolahan data yang cukup kompleks seperti data mining maupun perhitungan statistik.

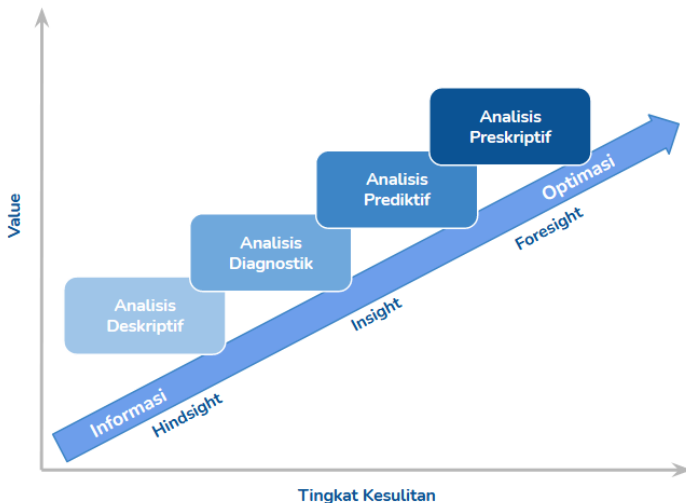
Dalam Big Data Analytics, tingkat kesulitannya semakin besar karena data yang diproses diperoleh dari berbagai sumber dengan bentuk dan jenis yang berbeda-beda, dan ukuran serta kecepatan yang besar pula. Oleh karena itu Big Data Analytics banyak menggunakan teknik dan algoritma yang lebih advance seperti predictive model dan machine learning untuk melihat trend, pola, korelasi dan insight lainnya.

Secara umum big data analytics terbagi 4 kategori yaitu:

1.

1. Descriptive Analytics
Analisis ini digunakan untuk menjawab pertanyaan mengenai apa yang sedang terjadi. Hampir semua organisasi telah mengimplementasikan analisis jenis ini.
2. Diagnostic Analytics
Setelah mengetahui apa yang terjadi, biasanya pertanyaan berikutnya adalah mengapa bisa terjadi. Analisa jenis ini menggunakan drill-down data untuk mencari alasan lebih mendalam mengenai apa yang sedang terjadi.
3. Predictive Analytics
Analisis prediktif memberikan prediksi mengenai apa yang akan terjadi berdasarkan data-data yang ada. Analisa jenis ini menggunakan teknik dan algoritma machine learning dan artificial intelligence untuk menghasilkan model prediksi berdasarkan data-data historis.

4. **Prescriptive Analytics**
Memanfaatkan analisis deskriptif dan prediktif, analisis jenis ini memberikan insight untuk dapat memperoleh hasil yang sesuai dengan apa yang telah diprediksikan.



Sumber : Gartner, Maret 2012

Jenis dan Tingkatan Data Analytics

Implementasi Big Data dalam Bisnis

Kebiasaan manusia dan persaingan bisnis di era yang semakin terbuka saat ini menjadikan pengambilan keputusan yang tepat adalah kunci untuk bertahan dalam bisnis. Data adalah salah satu penentu keberhasilan dalam pengambilan keputusan.

Customer

Profiling

Pola dan profil pelanggan dapat dipelajari melalui data-data yang dibuat oleh pelanggan ketika sedang berinteraksi dengan produk, baik secara langsung, melalui website ataupun menggunakan aplikasi. Saat ini data profil pelanggan dapat diperluas lagi dengan menyertakan informasi geolokasi, bahkan data-data sosial media yang mereka buat.

Semakin banyak data yang dikumpulkan, serta makin canggihnya proses pengolahan data tersebut, maka informasi yang akurat dan detail mengenai profil pelanggan dapat diperoleh. Produsen atau penyedia layanan dapat memberikan rekomendasi yang tepat kepada pelanggan sehingga dapat meningkatkan penjualan maupun loyalitas pelanggan.

Product

Development

Membangun produk dari sebuah ide yang pada akhirnya dapat diterima dengan baik oleh pasar merupakan sebuah tantangan. Big data dapat memberikan insight yang mendalam

untuk mengidentifikasikan kebutuhan pasar, melihat respon pelanggan melalui komentar pada forum atau sosial media, mengevaluasi kinerja penjualan produk di pasar dengan cepat, mengoptimalkan rantai distribusi, hingga mengoptimalkan strategi pemasaran.

Semakin baik pengelolaan data dan semakin cepat ketersediaan data akan dapat terus untuk membuat produk yang berkesinambungan sehingga memberikan nilai yang baik di bagi pelanggan dan pengguna.

Price

Optimization

Harga bisa menjadi kunci bagi pelanggan untuk menentukan produk yang akan dibeli. Akan tetapi perang harga pun dapat memberikan pengaruh buruk bagi produk itu sendiri. Big data dapat memberikan peta dan pola harga yang ada di pasar, sehingga produsen dapat menentukan harga yang optimal dan promosi harga yang sesuai dengan kebutuhan pasar.

Big Data untuk Telekomunikasi

Telekomunikasi merupakan salah satu sektor yang mau tidak mau harus berurusan dengan big data. Terlebih lagi saat ini layanan telekomunikasi bisa dibilang adalah jantung dari dunia digital kita. Jika data sering disebut sebagai ‘the new oil’, maka penyedia layanan telekomunikasi seperti memiliki sebuah tambang minyak yang sangat produktif. Ada banyak sekali sumber data yang ada dalam sebuah perusahaan telekomunikasi. Sebut saja data operasional jaringan, data transaksi percakapan, data koneksi internet, data pelanggan, dan data produk. Jika semua data-data tersebut dapat diintegrasikan dengan baik, maka akan dapat memberikan insight yang dapat digunakan untuk optimalisasi jaringan, meningkatkan pelayanan, pembuatan produk dan program promosi, serta meningkatkan loyalitas pelanggan.

Big Data untuk Kesehatan

Data dalam bidang kesehatan adalah salah satu contoh big data karena volume, kompleksitas, keragaman serta tuntutan ketepatan waktunya. Disamping itu layanan kesehatan juga melibatkan banyak sekali pihak, diantaranya yaitu berbagai rumah sakit, lab, klinik, dan asuransi kesehatan. Oleh karena itu bidang kesehatan termasuk sektor yang memiliki tantangan besar di bidang big data.

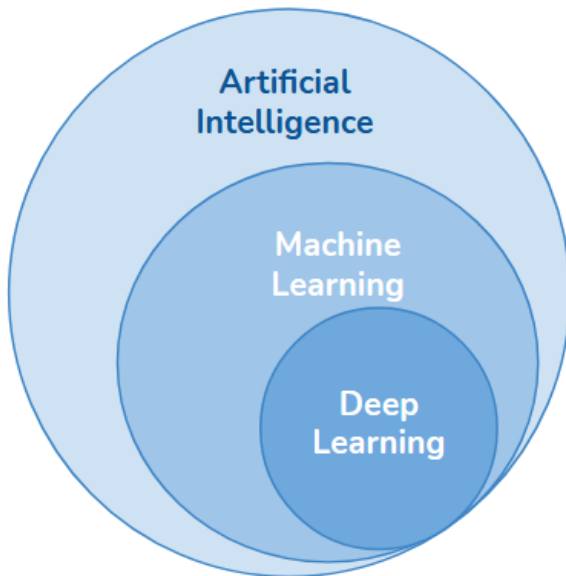
Integrasi data, akurasi data dan kecepatan perolehan data merupakan hal yang sangat penting dalam bidang kesehatan, karena hal ini menyangkut keselamatan pasien. Tidak hanya itu, jumlah tenaga medis dan rumah sakit pun masih sangat kurang dibanding dengan potensi pasien, terlebih di masa pandemi seperti saat ini. Insight yang diperoleh melalui big data dapat digunakan untuk membantu mengatasi permasalahan tersebut, diantaranya yaitu untuk penegakan diagnosa yang lebih akurat, personalisasi obat-obatan, peningkatan pelayanan rumah sakit hingga optimalisasi operasional rumah sakit.

Kecerdasan Buatan (Artificial Intelligence) dan Big Data

Setelah implementasi big data dalam arti pengelolaan dan analisa data dapat dilakukan dengan baik, tantangan berikutnya adalah bagaimana dengan data itu kita bisa melatih mesin untuk dapat belajar sehingga dapat bekerja dan memberikan insight secara otomatis, cepat, dan akurat. Maka Artificial Intelligence, Machine Learning

dan Deep Learning muncul kembali dan menjadi trend baru di masa kini.

Lalu apa perbedaan AI, machine learning dan deep learning? Secara ruang lingkup, deep learning merupakan bagian dari machine learning, dan machine learning merupakan bagian dari artificial intelligence. Inti ketiganya adalah bagaimana membuat mesin atau komputer menjadi cerdas. Tujuan utamanya yaitu untuk mengurangi campur tangan manusia dalam memberikan insight ataupun dalam melakukan berbagai pekerjaan manusia.



Hubungan AI, Machine Learning, dan Deep Learning

Artificial intelligence sendiri bukan merupakan hal baru, akan tetapi bidang ini mulai berkembang dengan sangat pesat dan menjadi sebuah trend setelah munculnya big data. Hal ini dikarenakan ketersediaan data yang melimpah, yang telah dapat ‘ditaklukkan’ dengan big data, yang merupakan materi utama bagi mesin untuk belajar dan menjadi cerdas.

Tidak hanya data, teknologi juga memegang peranan penting bagi perkembangan artificial intelligence. Berbagai perangkat dan teknologi dengan performa yang sangat tinggi saat ini sudah tersedia secara relatif murah dan terjangkau. Jika semula artificial intelligence dianggap sebagai sesuatu yang canggih dan hanya bisa diterapkan menggunakan teknologi yang tinggi dan mahal, saat ini artificial intelligence sudah dapat diimplementasikan pada berbagai perangkat dan sistem yang digunakan sehari-hari.

Komunitas Big Data Indonesia

idBigData adalah komunitas big data Indonesia yang dideklarasikan pada tanggal 2 Desember 2014. Pada saat itu big data masih menjadi sebuah hal yang relatif baru di Indonesia. Belum banyak orang yang paham mengenai apa itu big data, apa kegunaannya, dan bagaimana memanfaatkannya. Maka dibentuknya idBigData sebagai komunitas big data Indonesia bertujuan untuk menjadi wadah berkumpulnya komponen masyarakat dari berbagai bidang untuk saling bertukar pengetahuan dan pengalaman, serta menjalin berbagai kerja sama dalam

bidang big data serta pemanfaatannya, termasuk di dalamnya data science dan artificial intelligence.

Klasifikasi Teknologi Big Data – Dari namanya kita sudah tahu bahwa Big Data merupakan data dengan volume yang besar. Sejauh ini belum ada definisi dan juga arti Big Data namun para ahli mencoba untuk memberikan definisi terhadap Big Data ini agar dipahami oleh masyarakat luas khususnya pebisnis.

Big Data merupakan data dengan volume yang besar sehingga tidak bisa diproses dengan menggunakan alat tradisional biasa dan harus diproses dengan alat atau teknologi baru untuk bisa mendapatkan nilai dan juga informasi dari data ini. Tanpa disadari setiap hari manusia menghasilkan banyak data dan bisa berasal dari mana saja seperti postingan sosial media, gambar digital, video, dan lain sebagainya. Data inilah yang kemudian disebut dengan Big Data.

Pengelompokan atau Klasifikasi Big Data

Jika ditanya tentang klasifikasi Big Data maka jawabannya ada dua yaitu Big Data Operasional dan Big Data Analytic. Beban kerja operasional dan juga analitis yang berbeda inilah yang menyebabkan kebutuhan sistem berlawanan satu sama lain. Sistem Big Data inilah yang kemudian berevolusi menangani kedua jenis Big Data secara khusus, terpisah, dengan cara yang berbeda. Untuk memahami klasifikasi teknologi Big Data lebih lanjut, Anda bisa menyimak penjelasan di bawah ini:

- Big Data Operasional

Klasifikasi teknologi Big Data yang pertama adalah operasional. Untuk menangani pekerjaan-pekerjaan yang berhubungan dengan Big Data operasional telah dibangun sistem dengan database NoSQL seperti dengan database berbasis dokumen yang mana bisa ditunjukkan dengan beberapa tipe aplikasi, database key-value stores, column family store, dan juga database graph yang mana bisa dioptimalkan menjadi lebih spesifik.

NoSQL sendiri dikembangkan karena kekurangan database relasional yang berada di lingkungan komputasi lebih modern. Teknologi NoSQL dikenal lebih cepat, mudah, dan lebih murah. Hasilnya adalah NoSQL ini didesain sedemikian rupa untuk bisa memanfaatkan keunggulan arsitektur berbasis cloud yang sudah muncul dalam dua dekade ini. Dengan hal ini memungkinkan dijalankannya komputasi dengan skala besar secara efisien dengan biaya yang relatif murah. Hasilnya adalah NoSQL dengan komputasi awan menjadi perangkat kerja operasional Big Data yang mudah dikelola dan bisa di implementasikan lebih cepat.

Baca Juga: Mengapa Perlu Mengintegrasikan Big Data untuk Bisnis?

- Big Data Analytic

Klasifikasi Teknologi Big Data selanjutnya adalah jenis analytic dimana pekerjaan yang berhubungan dengan Big Data ini justru di proses dengan implementasi sistem

database MPP dan juga MapReduce. Munculnya teknologi ini juga menjadi reaksi terhadap keterbatasan dan juga kekurangan kemampuan relational database tradisional dan dalam skala lebih dari satu server. Selain itu MapReduce ini juga menawarkan metode baru yang mampu menganalisa data yang mana fungsinya sebagai pelengkap terhadap kapabilitas SQL ini.

Populernya penggunaan berbagai jenis aplikasi dari pengguna ini terus menerus memproduksi data dari pemakaian aplikasi tersebut dan dilakukan sejumlah upaya analisa yang benar-benar memberikan nilai berarti terhadap kemajuan bisnis yang mengimplementasikan teknologi Big Data tersebut. Sistem NoSQL ini juga mampu menyediakan beberapa fungsi MapReduce bawaan yang mana memungkinkan proses analisis dalam data operasional.

Sayangnya tidak semua orang menguasai teknologi dan analisis Big Data sehingga jika perusahaan ingin mengimplementasikan Big Data harus menggunakan jasa vendor yang benar-benar ahli dan berpengalaman. Anda bisa menggunakan jasa SOLTIUS yang mana menjadi perusahaan konsultan dan penyedia IT terbaik di Indonesia. Kami memiliki produk Big Data Indonesia yang mana mampu melakukan analisis data bisnis Anda secara maksimal. Dengan solusi teknologi Big Data yang kami berikan Anda bisa mengambil keputusan lebih cepat, akurat, dan tepat.

Berbicara teknologi memang tidak pernah ada matinya. Bagi Anda yang setia mengikuti perkembangan teknologi

saat ini, mungkin cukup familiar dengan istilah Big Data Indonesia. Apakah Anda sudah familiar dengan istilah tersebut? kira-kira apa manfaat yang diberikan sehingga big data begitu *booming*? dan siapa yang sudah menggunakan dan merasakan manfaatnya?

Pengertian Big Data Indonesia

Dalam dunia teknologi, big data berarti terobosan baru yang berkaitan dengan mengolah, menyimpan, dan menganalisis data dalam berbagai format, dimana jumlah data yang diolah, disimpan dan di analisis sangatlah besar dan data yang bertambah dengan cepat.

Dari istilahnya, Anda bisa memprediksi apa yang ada di dalamnya, yaitu data yang serba besar, dan melalui proses simpan, olah, dan analisis yang cepat. Dibandingkan dengan database sebelumnya, sebut saja MySQL, Big data Indonesia lebih unggul secara kapasitasnya.

Ada tiga kriteria data yang masuk dalam Big Data Indonesia, yaitu:

- Dari segi jumlah atau volumenya, tentu sangat besar. Adapun total ukuran data yang masuk biasanya dalam ukuran terabytes dan selebihnya.
- Bicara *velocity* atau pertumbuhan data dalam Big Data Indonesia, sangat cepat. Data bertambah dengan sangat cepat dan dalam kurun waktu yang relatif singkat.
- Dari segi format data yang masuk cukup beraneka ragam. Data tersebut adalah mulai dari data tabel, text, excel, dan beragam jenis format yang lain.

Tiga kriteria data tersebut selalu ada dalam Big Data Indonesia. Dari segi kuantitasnya sangat banyak, dari segi *velocity* data sangat cepat, dan dari segi varietas datanya sangat bervariasi. Untuk jenis data dengan kriteria tersebut, cocok sekali jika menggunakan Big Data Indonesia.

Manfaat Big Data Indonesia

Dari deskripsi Big Data Indonesia di atas? Beberapa manfaat yang bisa Anda dapatkan dengan menggunakan aplikasi ini adalah:

- Mampu menyuguhkan gambaran data yang lebih lengkap dibandingkan dengan aplikasi sebelumnya dengan jenis data yang biasa diolah adalah data terstruktur. Anda tidak perlu membagi data yang masuk ke dalam beberapa aplikasi karena kuantitasnya yang cukup banyak, atau karena ragamnya yang bervariasi. Kesatuan data yang utuh bisa Anda temukan disini.
- Beberapa data yang masuk dapat digunakan untuk merancang strategi pemasaran yang bagus sehingga meningkatkan omset perusahaan. Misalnya, data yang Anda dapatkan dari facebook atau social media yang lain akan diolah dengan *rigid* oleh Big Data Indonesia. Bagaimana tingkah laku dan respon konsumen, apa saja produk yang mendapatkan respon positif dan sebagainya. Dari situ, pihak perusahaan menghubungkan tingkah laku calon konsumen dengan database relasional yang sudah ada. Dari Sini, akan ditemukan strategi, dan solusi untuk perkembangan perusahaan selanjutnya.
- Big data Indonesia siap dimanfaatkan untuk berbagai bidang. Mulai dari bisnis, pemerintahan, perbankan, dan sebagainya. Dalam bidang pemerintahan, big data dimanfaatkan untuk mempercepat pengambilan keputusan, monitoring, dan evaluasi.

Hingga kini, pemanfaatan big data Indonesia dalam bidang pembangunan masih kurang maksimal. Ini

dikarenakan kurangnya para ilmuwan yang sanggup mengadopsi kecanggihan teknologi ini sehingga pemanfaatannya di bidang ini terhambat. Walau begitu, ada cukup banyak instansi pemerintahan yang sudah memanfaatkan kecanggihan teknologi yang satu ini. Dengan Big data Indonesia pekerjaan akan semakin cepat selesai, data yang masuk beragam dan dalam jumlah yang besar, serta perubahan data yang cepat pun bisa diproses dengan cepat pula.

Belum selesai Indonesia heboh membahas sistem cloud, sekarang sudah heboh big data. Apalagi itu? Begitu cepat sekali teknologi berkembang. Sepertinya lebih cepat teknologi untuk berkembang daripada telur menetas, ya? Memang, dengan hadirnya sistem cloud ditambah dengan teknologi big data, data dengan jumlah & volume besar yang dulu tidak ekonomis, kini menjadi “aset” bagi perusahaan.

Apakah kamu familiar dengan istilah era informasi? Era dimana informasi kini memiliki “harga” & bersifat ekonomis.

Data yang semula tidak memiliki nilai ekonomi, kini justru banyak dijual oleh orang-orang yang tidak bertanggung jawab. Parahnya, ada oknum yang membeli data ilegal tersebut. Ya, memang ada istilah “hal baru, masalah baru”.

Tetapi, ayo lihat dari sisi positifnya. Setelah itu, kamu baru boleh melakukan justifikasi terhadap teknologi big data, ya?

Untuk benar-benar bisa mengenal big data. Kamu harus paham dulu tentang pengetahuan dasar big data. Tenang saja, tidak akan ada hitung-hitungan di dalam artikel ini. Cukup duduk santai, sambil minum kopi panas diselingi dengan cemilan.

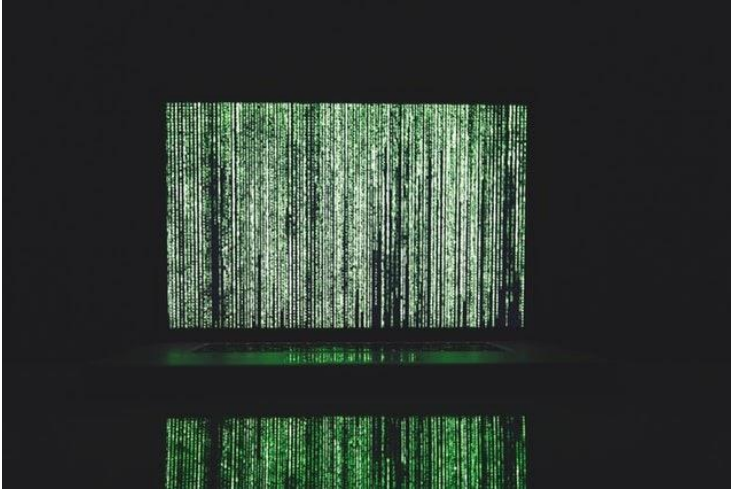
Apa itu Big Data?

Berikut adalah pengertian big data dari beberapa sumber:

Menurut TechTarget, Big Data adalah istilah yang dipakai untuk mendeskripsikan kumpulan data yang terstruktur, semi-terstruktur, hingga tidak terstruktur / abstrak dalam volume besar yang memiliki potensi untuk diolah sebagai informasi yang akan berguna untuk proyek *machine learning* dan aplikasi analisis canggih lainnya.

Menurut Investopedia, Big Data adalah data yang bervariasi dengan format beragam yang terkumpul dalam jumlah besar dalam waktu singkat.

Menurut perusahaan teknologi raksasa IBM, Big Data adalah istilah yang diberikan pada sekumpulan data yang tidak dapat dianalisis menggunakan metode analisa data tradisional karena tidak dapat menyimpan, mengolah, dan memproses data dalam waktu yang singkat.



Dari ketiga sumber tersebut, Big Data bisa disimpulkan sebagai:

1. Data dengan volume besar (terabyte - exabyte) - *Volume*
2. Struktur data memiliki variabel yang beragam (terstruktur - abstrak) - *Variety*
3. Berkumpul dalam waktu yang singkat (low latency) - *Velocity*

Meskipun begitu, ada persyaratan yang harus terpenuhi untuk bisa menyebut sekumpulan data dengan istilah Big Data.

Jika ada salah satu dari ketiga poin tadi tidak terpenuhi maka kumpulan data tersebut tidak bisa dikategorikan sebagai Big Data. Misalnya, ada sekumpulan data yang memenuhi poin *volume* & *variety* tetapi butuh waktu 10

Tahun untuk mengumpulkan data sebanyak 1 terabyte. Kenapa tidak dianalisis sebelum datanya menumpuk?

Atau misalnya, ada data yang terkumpul dalam waktu singkat dengan volume yang besar namun tidak memiliki variable. Bisa jadi itu big-size data saja. Masih jaman?

Atau contoh yang terakhir, sekumpulan data yang memiliki variabel yang beragam dan terkumpul dalam waktu singkat, namun dengan volume yang tidak terlalu besar. Itu Big Data, atau tugas kuliah?

Intinya, tidak semua kumpulan data bisa dikategorikan sebagai Big Data. Jangan kamu anggap tugas kuliah sebagai Big Data kalau data dengan volume besar terkumpul karena kamu males mencicil tugas, ya.

Sampai sini jelas? Kalau masih bingung, coba ulangi baca 3 poin di atas agar kamu bisa terus mengikuti alur artikel ini

Ada Apa Dengan Big Data?

Seberapa sakti sih Big Data itu? Kenapa sekarang banyak yang mengelu-elukan big data? Jadi gini lo, big data itu merupakan hal yang cukup baru. Baru dalam artian baru bisa dioptimalisasikan penggunaannya. Itu karena, sebelum hadirnya teknologi pendukung pengolahan big data, nilai dari big data sangat tidak ekonomis. Butuh hardware yang cukup mahal. Selain itu, jika data disimpan dalam jangka waktu tertentu, akan butuh biaya tambahan untuk memastikan data tersebut tidak rusak.

Tetapi, dengan hadirnya teknologi yang sudah mumpuni, kini bahkan perusahaan IT raksasa seperti IBM, Google, Amazon dan Microsoft, berlomba-lomba untuk menyediakan *platform data processing* (tempat untuk menganalisa data) untuk big data agar hasil analisis yang didapat bisa muncul secara *near to real time*. Dengan begitu, perusahaan / pelaku bisnis tidak perlu menyimpan data yang begitu besar, banyak dan variatif yang memakan akan banyak biaya maintenance. Sehingga, big data menjadi aset yang memiliki nilai ekonomi menjanjikan.



Hasil dari *data processing* Big Data dapat menghasilkan informasi yang akan berguna untuk membuat keputusan bisnis. Apalagi, dengan hadirnya teknologi Internet of Things, Artificial Intelligence, dan Smart City di

Indonesia, menjadikan Big Data sebagai acuan untuk mendapatkan hasil informasi yang presisi dalam waktu singkat (*near to real time*).

Apa hubungan teknologi IoT, AI, & smart city dengan big data?

Untuk dapat memanfaatkan teknologi IoT, AI, dan smart city, big data diperlukan untuk membaca kebiasaan pengguna / *user*, kebutuhan pengguna / *user*, hingga pola “aneh” yang terjadi. Dan tentunya, untuk bisa membaca kebiasaan & kebutuhan pengguna / *user*, hingga pola “aneh” dari pengguna / *user* diperlukan data yang sangat banyak. Mulai dari bangun tidur hingga tidur kembali. Tujuannya yaitu untuk dapat menampilkan hasil analisa yang presisi.

Coba kamu ingat-ingat aktivitas apa yang kemarin kamu lakukan? Berapa banyak kamu minum air putih? Berapa lama kamu menggunakan ponsel? Berapa banyak uang yang kamu habiskan selama 1 hari untuk belanja online? Atau mungkin, berapa tensi darah kamu kemarin?

Mungkin kamu hanya akan ingat kejadian yang secara sadar kamu lakukan. Tetapi, tentang berapa banyak kamu minum air putih? Atau mungkin tensi darah kamu kemarin? Well, itulah fungsi dari big data.

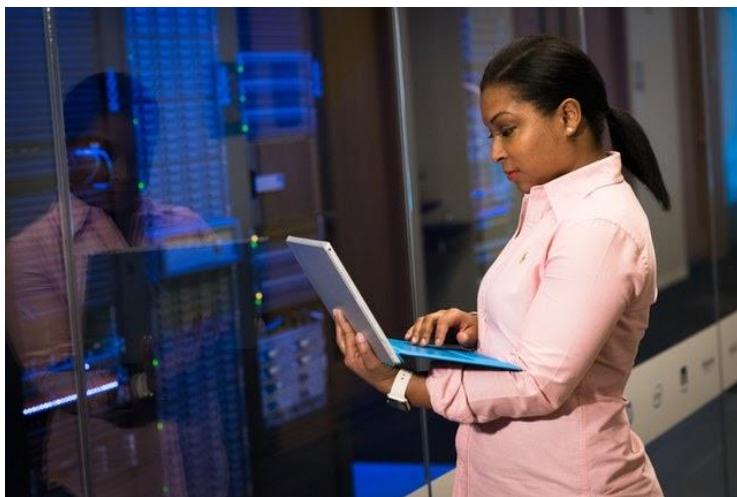
Perangkat IoT, AI atau apapun itu yang memiliki fungsi untuk merekam informasi, akan mengumpulkan data dengan jumlah yang sangat banyak dan diolah untuk memastikan tidak ada pola “aneh” pada aktivitas kamu

sehari-hari. Itu baru secara individual. Bagaimana jika data 1 RT, 1 RW, Kelurahan, Kecamatan, atau bahkan 1 kota? Berapa banyak data per detik yang direkam?

Bayangkan jika kamu mengolah data tersebut masih menggunakan Microsoft Excel. Jangankan untuk dibayangkan, untuk didengar saja sudah membuat pusing tujuh keliling. Dengan hadirnya teknologi pendukung Big Data, *automated data processing* bukan suatu hal yang mustahil.

Teknologi Pendukung Big Data

Big data yang sedang naik daun, tidak terlepas dari beberapa teknologi pendukung yang turut serta menstimulasi peningkatan minat pasar terhadap teknologi big data. Yuk simak penjelasan berikut.



The Hadoop Ecosystem

Ketika kamu mencari tentang Big Data di Google, kamu akan selalu melihat kata Hadoop. Termasuk di artikel ini. Hehehe.

Sebenarnya, apa itu hadoop?

Hadoop ecosystem itu bukanlah sebuah aplikasi, tetapi merupakan sebuah solusi untuk masalah Big Data. Hadoop bersifat *open source*. Itu berarti, sistem hadoop bisa kamu modifikasi untuk bisa memenuhi kebutuhan dari Big Data bisnismu. Atau bahasa gaul anak sekarang, hadoop itu bisa kamu oprek. Menurut Shubham Sinha dari edureka, Hadoop itu sebuah frameworks yang terdiri dari beberapa bagian dengan tugas yang berbeda-beda. Ada yang bertugas untuk mengumpulkan data, hingga bagian yang bertugas untuk menganalisis data.

Yuk simak bagian serta tugas dari anggota ekosistem Hadoop.

HDFS (Storage)

HDFS merupakan singkatan dari Hadoop Distributed File System. HDFS ini memiliki 2 bagian yang memiliki peran yang berbeda. Nah loh, beranak pinak kali ya ini sistem. Ya, HDFS ini terdiri dari Name node dan Data node.

Name node memiliki peran untuk merekam metadata dari sebuah file. Misalnya, file size, permission, lokasi penyimpanan file dan lainnya. Sedangkan Data node

memiliki peran untuk rekam data sesungguhnya. Misalnya, angka binary dan lainnya.

Pusing ya? Sama. Eh loh. Enggak kok bercanda.

Simple nya begini, HDFS itu layaknya telur asin. Ada cangkang nya, ada bagian telurnya. Nah, cangkangnya ini ibarat Name node. Kamu bisa tahu itu telur asin dari warna cangkangnya, cap produksi telur asinnya (kalau ada), cap expired nya juga. Tetapi, cangkangnya itu bukan sesuatu yang bisa kamu makan.

Ketika kamu beli telur asin, pasti kamu ingin memakan telurnya, bukan cangkangnya. Nah, telur di dalam cangkang itu lah yang diibaratkan Data node. Kamu bisa memakan telur asin nya, merasakan tekstur nya, rasanya, dll. Sama seperti Data node yang merekam file sesungguhnya.

Akan tetapi, Name node dan Data node itu sama-sama penting. Seperti telur asin. Memang kamu pernah beli telur asin tetapi tidak ada cangkangnya? Atau beli telur asin tidak ada telurnya? Seperti mata uang yang tidak bisa dipisahkan pokoknya.

Yarn

Apa pula ini yarn? Simple nya Yarn itu yang mengatur alur data processing di HDFS. Tidak mungkin dong file yang keluar masuk itu gak diatur. Lalu lintas saja ada traffic lights, masa big data enggak? Hehehe.

Sama seperti HDFS, Yarn juga dibagi menjadi 2 bagian. Bagian yang pertama dinamakan Resources Manager. Nah, ini fungsinya mirip traffic light. Kalau ada lampu hijau (request diterima), data / file baru bisa di record (write & read). Begitu pula jika ada lampu merah (request ditolak), maka data / file tidak bisa di record (write & read).

Jika sudah ada sinyal hijau atau merah perihal request dari system, Node manager lah yang akan bertindak seperti polisi lalu lintas. Memastikan lalu lintas write & read data berjalan sesuai dengan request. Node manager ini sudah pasti terpasang di setiap Data node (lihat bagian HDFS).

Nah, HDFS & Yarn ini dua sistem yang harus ada di dalam Hadoop ecosystem. Karena Hadoop ecosystem ini bersifat *open source*, kamu bisa menyesuaikan banyak sedikitnya sistem yang dipakai dalam Hadoop ecosystem ini. Ada beberapa list tools pendukung untuk Hadoop Ecosystem.

- Sparks (In-memory, data flow engine)
- Pig (Scripting)
- Hive & Drill (Analytical SQL-on-Hadoop)
- HBase (No SQL Database)
- Dan lain-lain

List diatas sifatnya sangat teknis, sehingga sulit dijelaskan dan akan memperlebar pembahasan tentang Big Data. So, Lanjut ke teknologi pendukung selanjutnya, ya?

Artificial Intelligence

Ya, disadari atau tidak bahwa hebohnya teknologi Artificial Intelligence beberapa tahun terakhir ini turut memperbanyak permintaan pasar akan Big Data.

Apa korelasi nya?

Ya, artificial intelligence atau bahasa Indonesianya adalah kecerdasan buatan, merupakan teknologi yang memungkinkan untuk melakukan prediksi akan suatu kondisi dan memberikan saran atau *insight*. Nah, untuk dapat menghasilkan prediksi akan suatu kondisi, teknologi Artificial Intelligence memerlukan Big Data untuk diproses sehingga mendapatkan hasil analisis yang lebih akurat.

Sudah paham kan korelasi Artificial Intelligence dengan Big Data?

Bahkan, tren artificial intelligence sudah mulai disuntikkan bukan hanya kedalam smartphone saja. Tetapi juga ke dalam website e-commerce sehingga bisa memberikan pengalaman pengguna “seperti memiliki asisten belanja”.

Tentunya, hal tersebut tidak bisa terjadi atas bantuan Big Data. Rekam jejak transaksi pengguna, hal apa yang sering pengguna cari di internet, hingga barang apa yang kali terakhir pengguna lihat.

Semua jejak terekam ke dalam Big Data yang akan diproses dan dianalisis. Nantinya, website e-commerce dapat memberikan rekomendasi produk yang “kemungkinan” kamu suka dan akan kamu beli.

Karena hadirnya Big Data, data & informasi menjadi sangat sensitif. Perusahaan-perusahaan yang memang merekam jejak pengguna harus benar-benar memiliki integritas tinggi.

Kenapa? Jangan sampai data pengguna dijual kepada pihak ketiga yang ingin membaca perilaku pengguna di Internet.

Apakah kamu pernah mendengar istilah “data is new oil”? Kini banyak perusahaan yang menginginkan data kebiasaan pengguna. Gunanya, tentu saja untuk mengembangkan bisnis. Data tersebut untuk keperluan product development, hingga customer experience.

Nah, integritas perusahaan pemilik data pengguna akan banyak diuji dengan tawaran-tawaran menarik untuk menjual data pengguna kepada pihak ketiga.

Kasus yang pernah menghebohkan dunia yaitu Facebook. Memang, facebook tidak menjual data pengguna ke pihak ketiga. Hanya saja facebook menyalahgunakan wewenang dengan mengizinkan pihak ketiga untuk menggunakan data pengguna Facebook. Lalu, tentu saja Facebook diputuskan bersalah.

Jadi, hati-hati ya dalam menggunakan Big Data atau data apapun. Sebagai perusahaan, tentu data pengguna tentang apapun itu tidak boleh disalahgunakan.

Selain itu, perusahaan juga harus memastikan bahwa pengguna produk tahu jika data mereka akan ditampung dalam Big Data. Amannya, berikan pilihan kepada pengguna untuk menyetujui atau tidaknya rekam jejak digital pengguna untuk disimpan dan digunakan untuk tujuan pengembangan perusahaan.

Sebagai pengguna, kamu harus tahu bahwa aktivitas digital kamu akan direkam. Sehingga, gunakanlah produk digital dengan lebih bijak.

Blockchain

Bitcoin. Ketika mendengar istilah blockchain, banyak orang yang mengaitkan dengan bitcoin. Padahal, blockchain memiliki arti yang lebih luas dari sekedar cryptocurrency.

Lalu, apa hubungan blockchain dengan Big Data?

Yuk, simak penjelasan berikut ini.

Blockchain merupakan metode distribusi data yang terbilang baru. Metode distribusi blockchain ini unik. Sekali data terekam (write) maka tidak akan bisa dimodifikasi ataupun dihapus. Sehingga, sangat efektif untuk menekan angka “fraud” pada transaksi.

Hal tersebut bisa terjadi karena sifat desentralisasi pada metode distribusi data blockchain. Sehingga, setiap transaksi akan terekam pada beberapa channel terkait (bank, perusahaan A, B, C, supplier A, B, C, government). Dengan demikian, seorang senior manager sekalipun memerlukan ijin akses (permission) untuk bisa membuka data dari setiap channel terkait dalam suatu transaksi.

Misalnya begini. Ketika kamu membeli sesuatu di toko online, dan ternyata kamu ditipu, pelacakan terhadap pelaku penipuan akan sangat mudah. Itu karena bank, selaku channel terkait pada saat transaksi, dapat melacak siapa saja yang terkait dalam proses transaksi tersebut. Menariknya, selain menemukan pelaku, bank pun bisa

mendata siapa saja orang yang pernah ditipu dengan modus yang sama.

Nah, hubungannya dengan big data adalah ketika terjadi transaksi, big data mampu menampung dan merekam aktivitas tersebut dalam skala yang besar. Data dengan volume dan variasi yang banyak distimulasi oleh banyaknya channel terkait sebuah transaksi. Dengan hadirnya blockchain, permintaan akan big data juga akan semakin besar.

Cuma sebagai tambahan, blockchain digadang-gadang dapat memberikan keamanan data yang jauh lebih baik. Sehingga, perusahaan besar IT di dunia mulai mengembangkan dan membuat simulasi kasus blockchain. Misalnya IBM, AWS, Microsoft dan masih banyak perusahaan perintis (startups), mereka sudah mulai mengembangkan metode ini untuk banyak kasus.

Sebenarnya masih banyak teknologi yang turut mendukung perkembangan Big Data. Meskipun demikian, 3 teknologi di atas (Hadoop ecosystem, Artificial Intelligence, dan Blockchain) merupakan highlight karena dinilai akan / sudah dirasakan secara langsung oleh kamu, yang sedang membaca artikel ini.

Manfaat Big Data

Setelah kamu paham apa itu big data dan bagaimana masa depan big data, sekarang saatnya membahas manfaat dari big data.



Ada 3 Manfaat Big Data bagi bisnis:

Mengoptimalisasi Cost

Big Data mampu menampung berbagai macam bentuk data dari berbagai macam sumber. Selain itu, kemampuan untuk melakukan data processing serta analisis secara mendalam melalui ecosystem seperti Hadoop, dapat membentuk pola baru yang akan sulit ditemui oleh tools analisis tradisional.

Nah, biasanya, pola baru ini memberikan efisiensi lebih terhadap bisnis proses. Entah melakukan cutting value chain, atau menemukan kelompok target market baru

dengan cost yang lebih sedikit. Sehingga, perusahaan dapat menggunakan cost secara optimal.

Misalnya, ketika kamu memiliki perusahaan kosmetik. Seringkali perusahaan kosmetik disangkut pautkan dengan wanita. Namun, setelah bisnismu memiliki website e-commerce.

Data menunjukkan bahwa memang benar, wanita lebih sering berbelanja di website e-commerce kamu. Tetapi, meskipun memiliki persentase yang lebih kecil, kaum adam mampu menghabiskan budget lebih terhadap produk yang mereka beli. Hal ini tentu sinyal hijau bagi pelaku bisnis. Bagaimana jika mereka fokus untuk mengembangkan produk khusus pria? Mereka dapat meraup margin keuntungan yang lebih besar. Contoh sederhananya seperti itu.

Meningkatkan efektifitas & efisiensi bisnis

Efektifitas dapat diartikan dalam 2 hal. Arti yang pertama yaitu berkurangnya waktu produksi akan suatu pekerjaan dengan kualitas hasil yang optimal. Misalnya, ketika teknologi internet belum sanggup untuk melakukan streaming film. Kamu harus mengunduh film dari internet agar bisa menonton film.

Waktu yang diperlukan untuk menonton film adalah waktu unduh + waktu menonton. Tetapi, ketika teknologi internet sudah mampu untuk melakukan streaming film. Kamu tidak perlu lagi mengunduh. Cukup tonton saja film yang kamu suka langsung dari internet browser di gadget

mu. Waktu yang diperlukan untuk menonton akan jauh lebih efektif.

Arti kedua dari efektifitas adalah bertambahnya hasil produksi dengan kualitas yang optimal dalam waktu satu kali kerja. Misalnya, ketika kamu masih sekolah dulu. Ketika teknologi kamera ponsel belum bisa memproduksi gambar yang tajam.

Kamu tidak bisa mengambil foto papan tulis dimana gurumu menerangkan perjalanan. Alih-alih mengambil foto, kamu harus menulis semua tulisan yang ada di papan tulis. Sehingga, kamu tidak fokus terhadap pelajaran. Tetapi, setelah teknologi kamera ponsel berkembang dan diperbolehkan penggunaannya di sekolah-sekolah, kamu kini tidak perlu lagi memecah konsentrasi antara menyalin & mendengarkan penjelasan oleh guru.

Kini kamu bisa mendapatkan dua hal dalam satu kali kerja. Yang pertama adalah pemahaman materi karena konsentrasi yang tidak terpecah. Yang kedua adalah dokumentasi.

Dari kedua analogi di atas, menunjukan pada satu benang merah yang sama, yaitu efisiensi.

Lalu, apa hubungannya dengan big data?

Ya, big data memiliki kemampuan untuk mencari jalan tercepat untuk menghasilkan sesuatu. Entah itu flow bisnis yang lebih efektif & efisien, hingga konsumen mana yang memiliki potensi “willingness to purchase”

tinggi namun terlewat. Itu semua bisa dideteksi lebih awal dengan big data.

Big data memungkinkan kamu untuk melakukan analisa secara menyeluruh. Selain itu, informasi yang dihasilkan oleh big data juga bisa dibilang *near-to-real-time*, sehingga menimbulkan kesan instan. Jika terdapat pola janggal, bisa segera ditangani tanpa meninggalkan snowball efek.

Product Development

Dengan kemampuan membaca kebiasaan & pola pengguna di internet, informasi yang dihasilkan oleh big data dapat digunakan oleh perusahaan untuk menyesuaikan produk dengan minat pasar. Kemampuan untuk menampilkan trend atau peningkatan demand terhadap suatu barang secara *near-to-real-time*, akan menguntungkan bagi bisnis.

Ketika terjadi lonjakan awal terhadap suatu hal, pelaku bisnis bisa memanfaatkan momen tersebut untuk mengembangkan produk yang disesuaikan dengan minat pasar. Sehingga, pelaku bisnis bisa merasakan manfaat dari “hype” konsumen akan sesuatu.

Selain itu, big data juga bisa memberikan informasi perihal jenis produk yang paling banyak digemari konsumen, alasan mengapa produk 1 disukai dan produk 2 kurang disukai, hingga kekosongan pasar. Selain akan menciptakan peluang baru bagi produk perusahaan, kemampuan untuk menyesuaikan produk dengan

ekspektasi konsumen maupun calon konsumen juga bisa terpenuhi.

Cara packing, warna package, aroma produk, tekstur produk, semua bisa disesuaikan dengan minat pasar berdasarkan hasil analisis dari big data.

Kesimpulan

Sejak memasuki era informasi. Kini banyak orang yang berlomba-lomba untuk memiliki informasi penting yang berguna dalam pengembangan bisnis atau bahkan bisa dikomersialisasikan. Namun, untuk mendapatkan informasi-informasi penting / pola penting terhadap suatu industri memerlukan data yang jumlahnya tidak sedikit. Bisa memang menggunakan metode sampling, namun untuk menghasilkan informasi yang lebih presisi dibutuhkan data yang memiliki volume tidak sedikit.

Oleh karena itu, munculah istilah big data. Tidak semua data yang memiliki size besar itu bisa disebut big data. Ada 3 kategori yang harus dipenuhi. Yang pertama adalah volume. Kedua yaitu variasi. Ketiga adalah velocity.

Dengan hadirnya big data, tentu membutuhkan teknologi pendukung untuk memunculkan sifat keekonomian data. Dengan naik daunnya teknologi big data, banyak teknologi pendukung yang bermunculan. Dari banyaknya teknologi pendukung big data, ada beberapa teknologi pendukung big data yang perlu kamu tahu. Pertama yaitu hadoop ecosystem. Selain itu adalah teknologi Artificial Intelligence. Yang terakhir adalah blockchain.

Big Data membawa beberapa manfaat yang akan berguna untuk bisnis. Manfaat yang pertama adalah meningkatkan efektifitas dan efisiensi bisnis. Dengan peningkatan efektifitas dan efisiensi bisnis, big data memberikan manfaat lain yaitu berupa optimalisasi cost. Selain itu, big data juga bisa menghadirkan informasi yang berguna untuk pengembangan produk. Produk dapat semakin disesuaikan dengan minat pasar dan lainnya.

Dari sini juga bisa diambil kesimpulan, bahwa Big Data berperan penting untuk mewujudkan *goal* dari perusahaan. Big Data juga memudahkan kamu dalam menentukan langkah selanjutnya dari bisnis yang kamu jalankan.

E-Commerce merupakan salah satu “wadah” yang bisa kamu gunakan untuk memperoleh data untuk kebutuhan bisnismu. Oleh karena itu sebelum kamu membangun website E-Commerce untuk bisnismu, kamu perlu tahu bagaimana nantinya sistem kamu dapat mengolah dan mengintegrasikan data dengan baik.

Softwareseni dapat membantu kamu untuk mewujudkan sistem E-Commerce impian kamu. Softwareseni akan membantu mulai dari tahap estimasi, *scoping* sampai proses *adjustment* sistem sesuai dengan *goal* dari bisnis kamu.

Dengan lebih dari 170 Staf profesional, Softwareseni yakin dapat menyediakan segala bentuk kebutuhan sistem E-Commerce kamu, sehingga kamu dapat memperoleh

informasi-informasi penting mengenai perilaku pengguna di website E-Commerce kamu.

Data adalah Emas



Pada pembukaan Press Conference on Open Data Strategy tahun 2011, Neelie Kroes, yang saat itu menjabat sebagai Vice-President of the European Commission responsible for the Digital Agenda, menyampaikan pidatonya yang berjudul "Data is the New Gold". Pesannya adalah bahwa pada era digital ini, data telah menjadi bagian yang sangat penting bagi peradaban manusia seperti halnya minyak bumi, yang telah mendapat julukan *black gold*. Pernyataan tersebut disampaikan berdasar fakta bahwa data telah menjadi

sumber laba bagi para pelaku bisnis di dunia maya Internet. Mereka hidup dari data yang mereka berdayakan.

Di Indonesia, Presiden Jokowi juga memiliki pandangan yang serupa akan pentingnya Data. Dalam pidato kenegaraan Presiden Jokowi pada 16 Agustus 2019, beliau menyatakan bahwa: "data adalah jenis kekayaan baru bangsa kita, kini data lebih berharga dari minyak".

Sejalan dengan pernyataan Presiden Jokowi dan Neelie Kroes tersebut, kini, manajemen data bukan lagi hanya menjadi kompetensi yang penting bagi suatu organisasi, melainkan telah menjadi bagian kritis yang berperan sebagai penentu kemenangan dalam penguasaan pasar maupun dalam pencapaian misi. Saat ini, perusahaan-perusahaan Fortune 1000 dan institusi-institusi pemerintah telah mulai memetik keuntungan dari inovasi-inovasi yang telah dikembangkan oleh para pionir dalam bisnis web services. Para decision maker pada organisasi-organisasi tersebut sedang berupaya untuk mengembangkan inisiatif baru dan mengevaluasi strategi-strategi yang mereka miliki demi menemukan cara bagaimana mereka dapat memanfaatkan Big Data untuk mengembangkan bisnisnya. Dalam proses tersebut, mereka pun belajar untuk memahami apa itu Big Data; mulai dari definisi Big Data, jenis-jenis teknologi Big Data, manfaat yang mungkin diperoleh dari implementasi teknologi Big Data, hingga bagaimana memilih teknologi Big Data yang tepat bagi kebutuhan mereka.

Bagaimana dengan di Indonesia? Tentu, Indonesia juga

tak mau ketinggalan. Sejumlah institusi pemerintah seperti Bank Indonesia dan Ditjen Pajak (Dengan 1,5 Triliun Rupiah, Ditjen Pajak Berdayakan Big Data untuk Amankan Pajak), BUMN seperti Bank Mandiri (Bank Mandiri Alokasikan 136 Miliar untuk Berdayakan Big Data), juga perusahaan swasta seperti GoJek, telah mengadopsi teknologi big data demi tingkatan kinerja dan produktivitas.

Mengingat kembali Definisi Big Data

Untuk membahas apa itu Big Data, baiknya dimulai dari kesepakatan tentang definisi Big Data itu sendiri. Big Data bukanlah sebuah teknologi, teknik, maupun inisiatif yang berdiri sendiri. Big Data adalah suatu trend yang mencakup area yang luas dalam dunia bisnis dan teknologi. Big Data menunjuk pada teknologi dan inisiatif yang melibatkan data yang begitu beragam, cepat berubah, atau berukuran super besar sehingga terlalu sulit bagi teknologi, keahlian, maupun infrastruktur konvensional untuk dapat menanganinya secara efektif. Dengan kata lain, Big Data memiliki ukuran (volume), kecepatan (velocity), atau ragam (variety) yang terlalu ekstrim untuk dikelola dengan teknik konvensional.

Big Data melibatkan proses pembuatan data, penyimpanan, penggalian informasi, dan analisis yang menonjol dalam hal volume, velocity, dan variety.

1. Volume (Ukuran). Pada tahun 2000 lalu, PC biasa pada umumnya memiliki kapasitas penyimpanan sekitar 10 gigabytes. Saat ini, Facebook menyedot sekitar 500 terabytes data baru setiap harinya; sebuah pesawat Boeing 737 menghasilkan sekitar 240 terabytes data penerbangan

dalam satu penerbangan melintasi Amerika; makin menjamurnya penggunaan ponsel pintar (smartphone), bertambahnya sensor-sensor yang disertakan pada perangkat harian, akan terus mengalirkan jutaan data-data baru, yang terus ter-update, yang mencakup data-data yang berhubungan dengan lingkungan, lokasi, cuaca, video bahkan data tentang suasana hati si pengguna ponsel pintar.

2. Velocity (kecepatan). Clickstreams maupun ad impressions mencatat perilaku pengguna Internet dalam jutaan event per detik; algoritma jual-beli saham dalam frekwensi tinggi dapat mencerminkan perubahan pasar dalam hitungan microseconds; proses-proses yang melibatkan hubungan antara suatu mesin dengan mesin lainnya telah melibatkan pertukaran data antar jutaan perangkat; peralatan sensor dan perangkat-perangkat pada infrastruktur menghasilkan log data secara real time; sistem game online dapat melayani jutaan pengguna secara bersamaan, yang masing-masing memberikan sejumlah input per detiknya.

3. Variety (ragam). Big Data tidak hanya menyangkut data yang berupa angka-angka, data tanggal, dan rangkaian teks. Big Data juga meliputi data-data ruang / geospasial, data 3D, audio dan video, dan data-data teks tak berstruktur termasuk file-file log dan media sosial. Sistem database tradisional didesain untuk menangani data-data berstruktur, yang tak terlalu sering mengalami update atau updatenya dapat diprediksi, serta memiliki struktur data yang konsisten yang volumenya tak pernah sebesar Big Data. Selain itu, sistem database tradisional juga didesain untuk digunakan dalam satu server yang berdiri sendiri, yang berakibat pada keterbatasan dan

mahalnya biaya untuk peningkatan kapasitas, sedangkan aplikasi sudah dituntut untuk mampu melayani pengguna dalam jumlah yang jauh lebih besar dari yang pernah ada sebelumnya. Dalam hal ini, database Big Data seperti halnya MongoDB maupun HBase, dapat memberikan solusi yang feasible yang memungkinkan peningkatan profit perusahaan secara signifikan.

Singkatnya, Big Data menggambarkan kumpulan data yang begitu besar dan kompleks yang tak memungkinkan lagi untuk dikelola dengan tools software tradisional.

Jenis Teknologi Big Data : Big Data Operasional dan Big Data Analitis

Dalam hal Teknologi, bentangan Big Data didominasi oleh dua jenis teknologi Big Data yaitu: (1) Big Data operasional: sistem yang memiliki kapabilitas operasional untuk pekerjaan-pekerjaan bersifat interaktif dan real time dimana data pada umumnya diserap dan disimpan; (2) Big Data analitis: sistem yang menyediakan kapabilitas analitis untuk mengerjakan analisis yang kompleks dan retrospektif yang dapat melibatkan sebagian besar atau bahkan keseluruhan data. Dalam keberadaannya, kedua jenis teknologi Big Data ini bersifat saling melengkapi dan sering digunakan secara bersamaan.

Beban kerja operasional dan analitis terhadap Big Data telah menyebabkan kebutuhan sistem yang berlawanan satu sama lain, dan sistem Big Data saat ini telah berevolusi untuk menangani kedua jenis kerja tersebut secara khusus, terpisah, dan dengan cara yang sangat berbeda. Baik kebutuhan kerja operasional maupun

analitis untuk Big Data, masing-masing telah mendorong penciptaan arsitektur-arsitektur teknologi baru. Sistem operasional, seperti halnya NoSQL database, berfokus pada pelayanan terhadap permintaan akses yang tinggi yang terjadi dalam waktu bersamaan, dengan tetap memberikan respon yang seketika (low latency) terhadap permintaan akses tersebut. Akses data terhadap sistem operasional ini dapat dilakukan dengan berbagai pilihan kriteria. Di lain pihak, sistem analitis cenderung berfokus pada penanganan arus data yang lebih besar, query-query yang ditujukan pada data tersebut bisa sangat kompleks, dan setiap kali dieksekusi dapat melibatkan sebagian besar atau keseluruhan data yang ada dalam sistem. Baik sistem Big Data operasional maupun sistem Big Data analitis, kedua-duanya dioperasikan dengan melibatkan sejumlah servers yang tergabung dalam suatu cluster komputer, dan digunakan untuk mengelola puluhan atau ratusan terabytes data yang memuat miliaran record.

Teknologi Big Data Operasional

Untuk menangani pekerjaan-pekerjaan Big Data Operasional, telah dibangun sistem Big Data dengan database NoSQL seperti halnya database berbasis dokumen (document based database) yang dapat ditujukan untuk berbagai tipe aplikasi, database key-value stores, column family stores, dan database graph yang dioptimalkan untuk aplikasi yang lebih spesifik. Teknologi NoSQL, yang telah dikembangkan untuk mengatasi kekurangan dari database relasional (relational database) pada lingkungan komputasi modern, dikenal lebih cepat serta lebih mudah dan murah dalam hal

peningkatan skala (more scalable) dibanding relational databases.

Terlebih lagi, sistem Big Data dengan database NoSQL telah didesain untuk memanfaatkan keunggulan dari arsitektur cloud computing (komputasi awan) yang telah muncul dalam dekade terakhir ini. Hal ini memungkinkan dijalankannya komputasi berskala besar secara efisien dan dengan biaya yang relatif lebih murah. Sebagai hasilnya, sistem NoSQL dengan komputasi awan ini telah menjadikan perangkat kerja Big Data operasional lebih mudah dikelola, serta dapat diimplementasikan dengan lebih murah dan cepat.

Berbicara NoSQL, saat ini MongoDB adalah database NoSQL yang sedang Naik Daun di Era Big Data.

Teknologi Big Data Analitis

Dilain pihak, pekerjaan-pekerjaan Big Data analitis cenderung diproses dengan mengimplementasikan sistem database MPP dan MapReduce. Munculnya teknologi ini juga merupakan reaksi terhadap keterbatasan dan kurangnya kemampuan relational database tradisional untuk mengelola database dalam skala lebih dari satu server (terdistribusi). Disamping itu, MapReduce juga menawarkan metode baru dalam menganalisa data yang dapat berfungsi sebagai pelengkap terhadap kapabilitas SQL.

Dengan semakin populernya penggunaan berbagai jenis aplikasi dan para penggunanya terus menerus

memproduksi data dari pemakaian aplikasi tersebut, terdapat sejumlah upaya analisa retrospektif yang benar-benar dapat memberikan nilai berarti terhadap kemajuan bisnis. Ketika upaya-upaya tersebut mesti melibatkan algoritma yang lebih rumit, MapReduce telah menjadi pilihan pertama untuk melakukan analisa retrospektif tersebut. Beberapa sistem NoSQL juga menyediakan fungsi MapReduce bawaan yang memungkinkan proses analisa diterapkan pada data operasional. Sebagai alternatif lain, data juga dapat dikopi dari sistem NoSQL ke dalam sistem analitis seperti halnya Hadoop dengan MapReduce-nya.

Setelah MapReduce (yang telah menjadi komponen utama dari Hadoop), kemudian muncul Apache Spark sebagai jawaban terhadap kebutuhan akan platform analitik yang lebih cepat dan terpadu. Dikatakan bahwa kemampuan *processing* Apache Spark adalah 100 kali lebih cepat dibanding Hadoop. Lebih detail tentang Apache Spark ada disini: Apache Spark, Platform Analitik Terpadu dan Super Cepat untuk Big Data. Apache Spark dapat dicoba di perangkat komputer sendiri (*local environment*) maupun dengan menggunakan layanan berbasis *cloud computing*.

Manfaat Pemberdayaan Big Data

Serangkaian teknologi baru yang ditujukan untuk memberdayakan Big Data telah memungkinkan direalisasikannya suatu nilai dari Big Data. Sebagai contoh, pebisnis retail online dapat mempelajari perilaku para pengunjungnya berdasarkan data hasil web click

tracking. Dengan mengetahui perilaku konsumen maupun calon konsumennya, maka dimungkinkan untuk menerapkan strategi baru guna meningkatkan penjualan, mengatur harga dan stok barang secara efisien. Institusi pemerintah maupun Google dapat mendeteksi timbulnya suatu wabah penyakit dengan memanfaatkan informasi yang mengalir di media sosial. Perusahaan minyak dan gas dapat menggunakan output dari sensor-sensor pada peralatan pengeboran untuk menemukan teknik pengeboran yang lebih aman dan efisien.

Jadi, dengan mendayagunakan database Big Data, operasional perusahaan dapat melakukan penghematan pengeluaran, meningkatkan keuntungan, dan mencapai sasaran-sasaran bisnis lainnya. Dalam hal ini paling tidak, ada 3 hal yang dapat diraih oleh perusahaan yang menerapkan teknologi Big Data, yakni:

1. Membuat aplikasi baru. Big Data memungkinkan suatu perusahaan untuk mengumpulkan data-data real time dari produk-produk yang mereka pasarkan, dari sumber daya yang digunakan, dan data-data yang berkaitan dengan pelanggannya. Data-data ini dapat dimanfaatkan untuk mengoptimalkan kepuasan pelanggan ataupun untuk efisiensi penggunaan sumber daya. Sebagai contoh, sebuah kota besar di Amerika Serikat telah menggunakan MongoDB, sebuah document based NoSQL database, untuk menurunkan angka kejahatan dan meningkatkan pelayanan umum dengan mengumpulkan dan menganalisa data geospatial secara real-time dari 30 departemen yang berbeda.

2. Meningkatkan efektifitas dan menurunkan biaya dari aplikasi yang telah ada. Teknologi Big Data dapat menggantikan sistem berspesifikasi tinggi yang mahal dengan sistem yang dapat dijalankan dengan spesifikasi standar. Disamping itu, karena banyak teknologi Big Data yang sifatnya open source, tentu mereka dapat diimplementasikan dengan biaya yang lebih murah dibandingkan teknologi yang hanya dimiliki dan dijual oleh suatu perusahaan.

3. Meningkatkan loyalitas pelanggan. Dengan semakin banyaknya data yang bisa diakses oleh berbagai bagian dalam suatu organisasi, juga dengan semakin cepatnya update yang dilakukan pada data-data tersebut, akan memungkinkan respon yang makin cepat dan akurat pula terhadap berbagai permintaan pelanggan.

Data adalah emas. Organisasi manapun yang menguasai emas, dapat dipastikan kekayaan dan kekuasaan ada ditangannya. Begitu juga halnya dengan data. Namun demikian, seperti halnya emas, data mesti digali, diproses dan dianalisa dengan serentetan teknologi tertentu demi mendapatkan nilai yang berharga dari lautan data pada era digital sekarang ini. Lautan data tersebut kemudian dikenal dengan istilah Big Data, kumpulan data yang begitu besar dan kompleks yang tak memungkinkan lagi untuk dikelola dengan tools software tradisional. Terdapat dua type teknologi untuk memberdayakan Big data, yaitu: (1) teknologi untuk memproses Big Data demi kebutuhan operasional, yakni: database NoSQL (MongoDB, HBase,

...), dan (2) teknologi untuk memproses Big Data guna kebutuhan analitis seperti halnya Hadoop. Dengan mengimplementasikan kedua type teknologi Big Data ini, akan memungkinkan didapatkannya nilai-nilai baru yang dapat memberikan manfaat pada operasional perusahaan berupa penghematan pengeluaran, peningkatan keuntungan, dan pencapaian sasaran-sasaran bisnis lainnya.

DAFTAR PUSTAKA

<https://bigbox.co.id/blog/big-data-panduan-lengkap-pengertian-manfaat-platform-dan-perusahaan-big-data/>

<https://bigbox.co.id/blog/sejarah-big-data-dan-trend/>

<https://www.universitas123.com/news/ketahui-sejarah-big-data-pertama-kali>

<https://penerbitbukudeepublish.com/big-data/>

<https://sis.binus.ac.id/2021/04/29/27751/>

<https://idbigdata.com/official/big-data-definisi-teknologi-dan-implementasinya/>

<https://www.soltius.co.id/blog/read/apa-itu-big-data-indonesia>

<https://www.soltius.co.id/id/blog/read/dua-klasifikasi-teknologi-big-data>

<https://www.softwareseni.co.id/blog/4-hal-tentang-big-data>

<https://www.teknologi-bigdata.com/2015/12/apa-itu-big-data-definisi-big-data-jenis-big-data-manfaat-big-data.html>

PENULIS BUKU BIG DATA



Muhammad Syarif Hartawan M.Kom MM
Universitas Krisnadipayana



Susy Rosyida, S.Kom, M.Kom
Universitas Nusa Mandiri



Abdul Hamid, S.Kom, M.Kom
Universitas Bina Sarana Informatika



Wulan Dari, S.Kom, M.Kom
Universitas Nusa Mandiri



Dr. Arman Syah Putra S.Kom MM M.Kom
Universitas Bina Nusantara

