Rakamin Academy

kimia farma

Project Based Internship

Big Data

Penerapan Konsep Big Data dalam Dunia Bisnis

Daftar Isi

A. SV Big Data

5V Big Data memiliki 5 ciri-ciri yang mencerminkan data, yaitu:

1. Volume

   Volume dalam *big data* diartikan sebagai kuantitas atau jumlah data yang dihasilkan dari banyak transaksi serta volume data yang disimpan. Contohnya bisa berbentuk *log history* pengguna seperti *history browser,* pencatatan transaksi pada *ecommerce*, data KTP, data pelanggan pada perbankan dan masih banyak lagi. Ukuran *big data* biasanya menggunakan skala *Terabytes* (per 1.000 *Gigabytes*) dan ukuran *Petabytes* (per 1.000.000 *Gigabytes*), pada contohnya berdasarkan publikasi yg dilakukan oleh facebook di https://research.fb.com/blog/2014/10/facebook-s-top-open-data-problems, facebook menghasilkan sejumlah 400 *petabytes* per hari atau 4.00.000.000 *Gigabytes* per hari, tentu data sebesar ini sudah dikategorikan sebagai *big data*.

2. Variety

   Variety ini artinya variasi tipe dan variasi sifat dari data, apakah data tersebut bersifat terstruktur/*structured*, semi terstruktur ataupun tidak terstruktur. Penjelasannya:

   a. Data terstruktur adalah data yang mempunyai elemen-elemen yang dapat di akses seperti *keys (primary key, relational keys, foreign key)* untuk dapat dianalisis ataupun data yang disimpan pada format tertentu contohnya data yang berada pada *relational database* ataupun *database SQL.*

b. Semi Terstruktur/*Semi-structured data* adalah informasi yang tidak disimpan dalam *relational database* tetapi mempunyai *pattern* atau terorganisir dengan rapi sehingga lebih mudah untuk dianalisa, dengan sedikit pengolahan kita dapat menyimpan data ini ke dalam *relational database* contohnya data pada file XML dan CSV yang sering dipergunakan untuk *export data* pada *database*.

c. Data Tidak Terstruktur/*Unstructured data* adalah informasi atau data yang tidak terorganisir dengan baik karena sifat alaminya, atau tidak memiliki *predefined data model* atau model yang sudah terdefinisi contohnya file gambar, suara, vidio, pdf, log files dan lainnya.

3. Velocity

Velocity dalam *big data* artinya adalah kecepatan dalam men- generate data, mengakses data serta memproses data. *Big data platform* dan *big data analytics software* tentu harus dapat memproses banyak data secepat mungkin ketika ada *request*, contoh *velocity* salah satunya adalah pada *search engine google*, berdasarkan data pada https://www.internetlivestats.com/google-search-statistics/ google harus memproses rata-rata 40.000 pencarian setiap detiknya.

4. Veracity

Veracity adalah data yang ada dapat dipercaya kebenarannya, dapat diandalkan, berkualitas serta dapat dapat diakses dengan baik. Contoh dari *veracity* dalam bisnis adalah ketika data pelanggan atau data keluarga perlu dijaga agar tetap akurat, mengingat perubahan yang terjadi seiring waktu, seperti pergantian nomor handphone, alamat, atau status keluarga. Untuk memastikan data tetap akurat,

perusahaan seringkali melakukan kegiatan *update data*, yang dengan menghubungi pelanggan untuk memperbarui informasi seperti alamat, nomor handphone, dan email. Hal ini penting untuk memastikan analisis data yang akurat dan pengiriman pesan *marketing* yang efektif kepada pelanggan.

5. Value

   Value adalah karakteristik terpenting dalam analisis bisnis dalam kerangka *5V big data*. Ini mengacu pada nilai yang dapat ditemukan dalam data dan sangat tergantung pada isi data serta kemampuan tim *data analyst* yang menganalisisnya. Dengan data yang tepat dan analisis yang cermat, *big data* memiliki potensi untuk menghasilkan informasi yang berharga untuk mendukung pengambilan keputusan. Sebagai contoh, dalam konteks Indonesia, pemerintah dapat menggunakan *big data* untuk mengumpulkan informasi dari berbagai kementerian dan instansi, seperti program ketahanan pangan. Dengan data dari berbagai sumber ini, pemerintah dapat menganalisis kapasitas produksi pangan, stok pangan, dan kebutuhan pangan di Indonesia. Dengan analisis data ini, pemerintah dapat memprediksi potensi kekurangan pangan dan mengambil langkah-langkah, seperti meningkatkan kapasitas produksi atau impor pangan, untuk menjaga ketersediaan pangan di dalam negeri.

B. Distributed File System

Distributed File System (DFS) adalah sistem file yang didistribusikan di beberapa *server file* atau beberapa lokasi. Hal ini memungkinkan program untuk mengakses atau menyimpan file terisolasi seperti yang mereka lakukan

dengan yang lokal, memungkinkan *programmer* untuk mengakses file dari jaringan atau komputer manapun.

Tujuan utama dari *Distributed File System (DFS)* adalah untuk memungkinkan pengguna sistem terdistribusi secara fisik untuk berbagi data dan sumber daya mereka dengan menggunakan *Common File System.* Kumpulan *workstation* dan *mainframe* yang dihubungkan oleh *Local Area Network (LAN)* merupakan konfigurasi pada *Distributed File System.* DFS dijalankan sebagai bagian dari sistem operasi. Di DFS, *namespace* dibuat dan proses ini transparan untuk klien.

DFS memiliki dua komponen:

1. Transparansi Lokasi – Transparansi Lokasi dicapai melalui komponen *namespace*.
2. Redundansi – Redundansi dilakukan melalui komponen replikasi file.

Fitur DFS:

1. Transparansi

   Klien tidak perlu mengetahui jumlah atau lokasi server file dan perangkat penyimpanan. Beberapa file server harus disediakan untuk kinerja, kemampuan beradaptasi, dan ketergantungan.
2. Transparansi akses

   File lokal dan jarak jauh harus dapat diakses dengan cara yang sama. Sistem file harus secara otomatis ditempatkan pada file yang diakses dan mengirimkannya ke sisi klien.
3. Transparansi nama file

   Seharusnya tidak ada petunjuk dalam nama file ke lokasi file. Setelah

nama diberikan ke file, itu tidak boleh diubah selama mentransfer dari satu node ke node lain.

4. Transparansi replikasi

   Jika file disalin pada beberapa node, salinan file dan lokasinya harus disembunyikan dari satu node ke node lainnya.

5. Performance

   Kinerja didasarkan pada jumlah rata-rata waktu yang dibutuhkan untuk meyakinkan permintaan klien. Waktu ini mencakup waktu CPU + waktu yang dibutuhkan untuk mengakses penyimpanan sekunder + waktu akses jaringan. Disarankan agar kinerja Sistem File Terdistribusi serupa dengan sistem file terpusat.

6. Ketersediaan

   Sistem File Terdistribusi harus dapat melanjutkan jika terjadi kegagalan sebagian seperti kegagalan tautan, kegagalan *node*, atau *crash drive* penyimpanan. Sistem file terdistribusi yang otentik dan mudah beradaptasi harus memiliki *server file* yang berbeda dan independen untuk mengontrol perangkat penyimpanan yang berbeda dan independen.

7. Skalabilitas

   Karena menumbuhkan jaringan dengan menambahkan mesin baru atau menggabungkan dua jaringan adalah rutinitas, sistem terdistribusi pasti akan tumbuh seiring waktu. Akibatnya, sistem file terdistribusi yang baik harus dibangun untuk menskalakan dengan cepat seiring dengan bertambahnya jumlah node dan pengguna dalam sistem. Layanan tidak boleh terganggu secara substansial karena jumlah node dan pengguna bertambah.

8. Integritas data

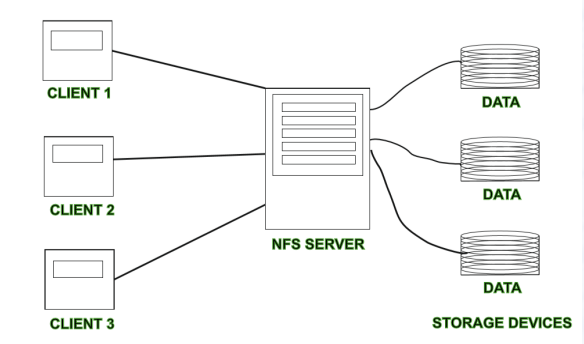
   Beberapa pengguna sering berbagi sistem file. Integritas data yang disimpan dalam file bersama harus dijamin oleh sistem file. Artinya, permintaan akses bersamaan dari banyak pengguna yang bersaing untuk akses ke file yang sama harus disinkronkan dengan benar menggunakan metode kontrol konkurensi. Transaksi atom adalah mekanisme manajemen konkurensi tingkat tinggi untuk integritas data yang sering ditawarkan kepada pengguna oleh sistem file.

9. Reliability/Keandalan

   Penting untuk meminimalkan risiko kehilangan data dalam sistem file terdistribusi yang tepat. Sebagai alternatif membuat salinan cadangan oleh pengguna, sistem file seharusnya secara otomatis membuat salinan cadangan dari file kunci untuk mengatasi kemungkinan kehilangan data. Banyak sistem file mengandalkan penyimpanan yang stabil dengan keandalan tinggi untuk tujuan ini.

10. Keamanan

    Sistem file terdistribusi harus aman sehingga penggunanya dapat percaya bahwa data mereka akan dijaga kerahasiaannya. Untuk menjaga informasi yang terkandung dalam sistem file dari akses yang tidak diinginkan & tidak sah, mekanisme keamanan harus diterapkan.



Keuntungan	Kekurangan
1. DFS memungkinkan banyak pengguna untuk mengakses atau menyimpan data. 2. Memungkinkan data untuk dibagikan dari jarak jauh. 3. Meningkatkan ketersediaan file, waktu akses, dan efisiensi jaringan. 4. Peningkatan kapasitas untuk mengubah ukuran data dan juga meningkatkan kemampuan untuk bertukar data. 5. Sistem File Terdistribusi memberikan transparansi data bahkan jika server atau disk gagal.	1. Dalam Sistem File Terdistribusi, node dan koneksi perlu diamankan oleh karena itu kita dapat mengatakan bahwa keamanan dipertaruhkan. 2. Ada kemungkinan kehilangan pesan dan data dalam jaringan saat berpindah dari satu node ke node lainnya. 3. Koneksi database dalam kasus Sistem File Terdistribusi rumit. 4. Juga penanganan database tidak mudah dalam Sistem File Terdistribusi dibandingkan dengan sistem pengguna tunggal. 5. Ada kemungkinan overloading akan terjadi jika semua node mencoba mengirim data sekaligus.

C. Cloud Computing

Cloud Computing adalah ketersediaan sumber daya komputasi berdasarkan permintaan sebagai layanan melalui internet. *Cloud Computing* menghilangkan kebutuhan perusahaan untuk mendapatkan, mengkonfigurasi, atau mengelola

sumber daya sendiri, dan mereka hanya membayar untuk apa yang mereka gunakan.

Memahami jenis sumber daya komputasi awan dapat memakan waktu dan mahal. Perusahaan perlu membeli *server* fisik dan infrastruktur lainnya melalui proses pengadaan yang dapat memakan waktu berbulan-bulan, dan mendukung arsitektur komputasi awan. Sistem yang diperoleh memerlukan ruang fisik, biasanya ruang khusus dengan daya dan pendinginan yang cukup. Setelah mengkonfigurasi dan menerapkan sistem, perusahaan membutuhkan personel ahli untuk mengelolanya. Proses panjang ini sulit untuk diukur ketika permintaan melonjak atau bisnis berkembang. Perusahaan dapat memperoleh lebih banyak sumber daya komputasi daripada yang dibutuhkan, berakhir dengan angka pemanfaatan yang rendah. Contoh komputasi awan ini salah satunya adalah *Google Cloud.*

Keuntungan *Cloud Computing:*

1. Fleksibel

   Karena arsitektur komputasi awan, perusahaan dan penggunanya dapat mengakses layanan *cloud* dari mana saja dengan koneksi internet, meningkatkan atau menurunkan skala layanan sesuai kebutuhan.

2. Efisien

   Perusahaan dapat mengembangkan aplikasi baru dan dengan cepat memasukkannya ke dalam produksi tanpa mengkhawatirkan infrastruktur yang mendasarinya.

3. Menawarkan nilai strategis

   Karena penyedia *cloud* selalu mengikuti inovasi terbaru dan menawarkannya sebagai layanan kepada pelanggan, perusahaan dapat

memperoleh keunggulan yang lebih kompetitif dan laba atas investasi yang lebih tinggi dibandingkan jika mereka berinvestasi pada teknologi yang akan segera usang.

4. Aman

   Keamanan komputasi awan umumnya diakui lebih kuat daripada di pusat data perusahaan, karena kedalaman dan luasnya mekanisme keamanan yang diterapkan oleh penyedia awan. Selain itu, tim keamanan penyedia cloud dikenal sebagai pakar top di bidangnya.

5. Hemat biaya

   Apa pun model layanan komputasi awan yang digunakan, perusahaan hanya membayar sumber daya komputasi yang mereka gunakan. Mereka tidak perlu meningkatkan kapasitas pusat data untuk menangani lonjakan permintaan atau pertumbuhan bisnis yang tidak terduga, dan mereka dapat mengerahkan staf TI untuk mengerjakan inisiatif yang lebih strategis

D. Data Security

Data security adalah praktik melindungi informasi digital dari akses tidak sah, korupsi, atau pencurian di seluruh siklus hidupnya. Ini adalah konsep yang mencakup setiap aspek keamanan informasi mulai dari keamanan fisik perangkat keras dan perangkat penyimpanan hingga kontrol administratif dan akses, serta keamanan logis aplikasi perangkat lunak. Ini juga mencakup kebijakan dan prosedur organisasi.

Ketika diterapkan dengan benar, strategi keamanan data yang kuat akan melindungi aset informasi organisasi dari aktivitas kejahatan dunia maya, tetapi juga melindungi dari ancaman orang dalam dan kesalahan manusia,

yang tetap menjadi salah satu penyebab utama pelanggaran data saat ini. Keamanan data melibatkan penerapan alat dan teknologi yang meningkatkan visibilitas organisasi ke tempat data penting berada dan bagaimana data tersebut digunakan. Idealnya, alat ini harus dapat menerapkan perlindungan seperti enkripsi, penyembunyian data, dan redaksi file sensitif, dan harus mengotomatiskan pelaporan untuk merampingkan audit dan mematuhi persyaratan peraturan.

1. Strategi Data Security

Strategi keamanan data yang komprehensif menggabungkan orang, proses, dan teknologi. Menetapkan kontrol dan kebijakan yang tepat adalah pertanyaan budaya organisasi yang sama pentingnya dengan penerapan perangkat yang tepat. Ini berarti menjadikan keamanan informasi sebagai prioritas di semua area perusahaan.

- Keamanan fisik server dan perangkat pengguna

  Terlepas dari apakah data disimpan di tempat, di pusat data perusahaan, atau di *cloud* publik, perlu memastikan bahwa fasilitas diamankan dari penyusup dan memiliki tindakan pencegahan kebakaran dan kontrol iklim yang memadai. Penyedia *cloud* akan bertanggung jawab atas tindakan perlindungan ini atas nama organisasi.

- Manajemen dan kontrol akses

  Memberikan akses database, jaringan, dan akun administratif kepada sesedikit mungkin orang, dan hanya mereka yang benar-benar membutuhkannya untuk menyelesaikan pekerjaan mereka.

- Keamanan dan penambalan aplikasi

  Semua perangkat lunak harus diperbarui ke versi terbaru sesegera mungkin setelah *patch* atau versi baru dirilis.

- Cadangan

  Mempertahankan salinan cadangan yang dapat digunakan dan diuji secara menyeluruh dari semua data penting adalah komponen inti dari setiap strategi keamanan data yang kuat. Selain itu, semua cadangan harus tunduk pada kontrol keamanan fisik dan logis yang sama yang mengatur akses ke database utama dan sistem inti.

- Pengetahuan karyawan

  Melatih karyawan tentang pentingnya praktik keamanan yang baik dan kebersihan kata sandi dan mengajari mereka untuk mengenali serangan rekayasa sosial untuk menjaga data.

- Pemantauan dan kontrol keamanan jaringan dan titik akhir

  Menerapkan rangkaian lengkap alat dan platform manajemen ancaman, deteksi, dan respons di seluruh lingkungan lokal dan platform *cloud* dapat mengurangi risiko dan mengurangi kemungkinan pelanggaran.

2. Jenis- Jenis Data Security

- Enkripsi

  Menggunakan algoritma untuk mengubah karakter teks normal menjadi format yang tidak dapat dibaca, kunci enkripsi mengacak data sehingga hanya pengguna yang berwenang yang dapat membacanya. Solusi enkripsi file dan *database* berfungsi sebagai

garis pertahanan terakhir untuk volume sensitif dengan mengaburkan isinya melalui enkripsi atau tokenisasi. Sebagian besar solusi juga menyertakan kemampuan manajemen kunci keamanan.

- Penghapusan Data

  Lebih aman daripada penghapusan data standar, penghapusan data menggunakan perangkat lunak untuk sepenuhnya menimpa data pada perangkat penyimpanan apa pun. Ini memverifikasi bahwa data tidak dapat dipulihkan.

- Data Masking

  Dengan menyembunyikan data (*masking*), organisasi dapat mengizinkan tim untuk mengembangkan aplikasi atau melatih orang menggunakan data nyata. Ini menutupi *personally identifiable information (PII)* jika diperlukan sehingga pengembangan dapat terjadi di lingkungan yang sesuai.

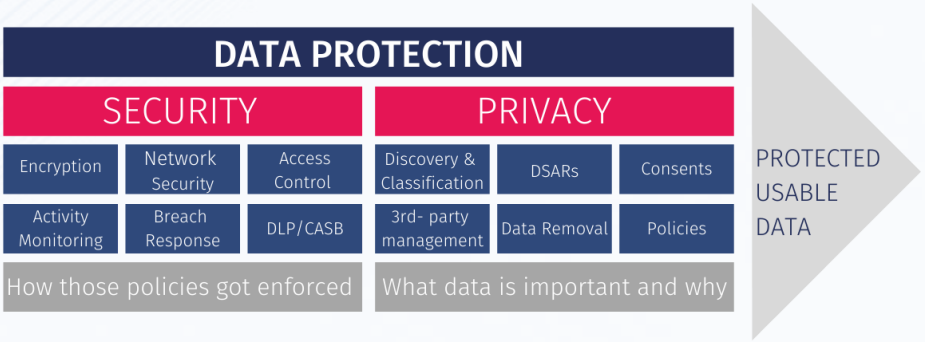
- Data Resiliency

  Ketahanan ditentukan oleh seberapa baik organisasi bertahan atau pulih dari segala jenis kegagalan, mulai dari masalah *hardware* hingga kekurangan daya dan peristiwa lain yang mempengaruhi ketersediaan data. *Data Resiliency* sangat penting untuk meminimalkan dampaknya.

E. Data Privacy

Data privacy adalah bagian dari area perlindungan data yang berhubungan dengan penanganan data yang tepat dengan fokus pada kepatuhan terhadap peraturan perlindungan data. Privasi Data berpusat pada bagaimana data harus

dikumpulkan, disimpan, dikelola, dan dibagikan dengan pihak ketiga mana pun, serta kepatuhan terhadap undang-undang privasi yang berlaku (seperti California Consumer Privacy Act- CCPA atau General Data Protection Regulation GDPR).



1. Elemen Data Privacy

- Hak individu untuk dibiarkan sendiri dan memiliki kendali atas data pribadinya
- Prosedur untuk penanganan, pemrosesan, pengumpulan, dan pembagian data pribadi yang tepat
- Kepatuhan dengan undang-undang perlindungan data

2. Mengapa Data Privacy Penting?

Undang-undang perlindungan data di seluruh dunia bertujuan untuk mengembalikan kendali individu atas data, memberdayakan mereka untuk mengetahui bagaimana data mereka digunakan, oleh siapa dan mengapa, memberi mereka kendali atas bagaimana data pribadi mereka diproses dan digunakan.

Pada tahun 2019, 73% pelanggan mengatakan bahwa kepercayaan pada perusahaan lebih penting daripada tahun lalu, dan kami dapat berasumsi bahwa jumlahnya telah meningkat. Baca 100 Statistik Privasi Data dan Keamanan Data untuk wawasan lebih lanjut.

Itulah sebabnya organisasi perlu mempelajari cara memproses data pribadi sambil melindungi preferensi privasi individu. Inilah yang diharapkan individu dari organisasi. Ini adalah visi privasi mereka. Pentingnya Privasi Data semakin ditingkatkan dengan diperkenalkannya *General Data Protection Regulation.*

References

5 things you need to know about Data Privacy [Definition & Comparison] – Data Privacy Manager. (2023). Data Privacy Manager. Retrieved October 27, 2023, from https://dataprivacymanager.net/5-things-you-need-to-know-about-data-privacy/

Jain, S. (2023). *What is DFS (Distributed File System)?* GeeksforGeeks. Retrieved October 27, 2023, from https://www.geeksforgeeks.org/what-is-dfsdistributed-file-system/

What is Cloud Computing? (n.d.). Google Cloud. Retrieved October 27, 2023, from https://cloud.google.com/learn/what-is-cloud-computing

What is Data Security? Data Security Definition and Overview. (n.d.). IBM. Retrieved October 27, 2023, from https://www.ibm.com/topics/data-security