

LAMPIRAN

Bab 3: Konsep Dasar Analisis Dan Visualisasi Data

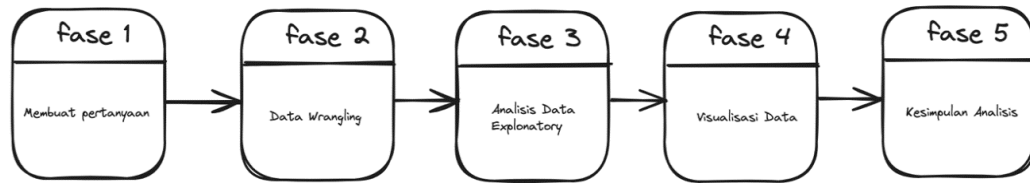
3.5 Teknologi Analisis Data

Teknologi Analisis Data merupakan sebuah mekanisme yang dapat diterapkan untuk mengubah *data raw* (mentah) menjadi sebuah wawasan agar dapat digunakan oleh suatu organisasi untuk pengambilan keputusan ataupun untuk membentuk proses bisnis agar dapat mendorong pertumbuhan pangsa pasar perusahaan, untuk mengubah data mentah menjadi sebuah wawasan diperlukan alat, teknologi dan proses untuk menemukan pola tren yang sedang terjadi ataupun untuk memecahkan permasalahan (AWS, 2023b).

3.5.1 Dasar Analisis Data

Analisis data adalah suatu prosedur yang dapat digunakan untuk menyelidiki, membersihkan, mengubah dan memodelkan data sehingga dari yang sebelum nya data tersebut berupa data mentahan menjadi suatu data yang memiliki wawasan yang berguna, ketika data sudah dapat memberikan wawasan. Data tersebut dapat digunakan sebagai landasan untuk mengambil suatu keputusan, menginformasikan kesimpulan ataupun yang lain nya. Di zaman sekarang data telah menjadi sebuah asset sehingga banyak dari kalangan perusahaan-perusahaan atau pelaku bisnis yang lain nya sedang berlomba untuk memanfaatkan data tersebut menjadi sebuah wawasan yang berharga demi keberlangsungan bisnis yang sedang di jalani nya, karena data sendiri dianggap sebagai asset berharga yang dapat memberikan keuntungan secara kompetitif maka dari itu proses analisis data menjadi langkah penting karena dapat membantu pelaku bisnis untuk mengoptimalkan kinerja nya. Pada dasar nya analisis data memiliki siklus bagaimana data tersebut yang mula-mula nya data mentah menjadi suatu data yang memiliki nilai wawasan yang berharga, berikut

merupakan gambaran atau tahapan dari siklus analisis data.



Gambar 3.5 Siklus Analisis Data.

Siklus Analisis Data merupakan sebuah proses atau langkah-langkah yang dapat digunakan untuk menyelidiki, membersihkan, mengubah dan memodelkan data dari yang sebelumnya mentah menjadi data yang memiliki wawasan yang berguna, dengan proses ini maka data dapat diekstraksi agar data tersebut menghasilkan sebuah informasi yang dapat dimanfaatkan untuk kepentingan organisasi (Dicoding, 2023), Siklus Analisis Data meliputi beberapa tahapan diantaranya :

1) Membuat Pertanyaan

Pada siklus pertama sebelum memulai melakukan analisis data adalah perlu membuat pertanyaan yang relevan terkait tentang permasalahan yang akan diselesaikan, dengan pertanyaan tersebut akan membantu memahami masalah yang sedang dihadapi serta untuk merumuskan tujuan yang akan dicapai.

2) Data Wrangling

Setelah tahap 1 selesai dibuat maka tahap selanjutnya adalah fase data wrangling, fase ini bisa dimulai ketika fase 1 telah selesai membuat pertanyaan terkait tentang permasalahan yang akan diselesaikan. Selanjutnya fase data wrangling dimulai dengan mengumpulkan data-data yang dibutuhkan untuk ditelaah lebih lanjut sehingga harapan menjawab setiap pertanyaan-pertanyaan yang sebelumnya telah dibuat.

3) Analisis Data Explanatory

Pada siklus fase ke 3 merupakan tahap eksplorasi terhadap data-data yang sebelumnya telah dikumpulkan dan dibersihkan pada saat fase ke 2 untuk menjawab pertanyaan-pertanyaan yang telah diajukan pada fase sebelumnya, proses analisis data explanatory merupakan cara untuk memperoleh wawasan, pengetahuan, atau pola yang bermakna dari data yang sebelumnya sudah dikumpulkan. Pada fase ini memungkinkan sebuah data telah bersih dan mudah untuk dipahami isi data mulai dari distribusi, frekuensi, korelasi ataupun yang lainnya. Pada umumnya proses analisis data explanatory dapat dilakukan dengan beberapa cara diantaranya :

- a. Analisis Univariate: Proses analisis secara deskriptif yang hanya dilakukan dengan 1 variabel.
 - b. Analisis Bivariate : Proses analisis secara relasi yang biasa nya dilakukan menggunakan 2 variabel.
 - c. Analisis Multivariate : Proses Analisis secara relasi yang menggunakan lebih dari 2 variabel.
- 4) Visualisasi Data
- Hasil data yang telah diperoleh dari fase analisis data explanatory, selanjut nya alangkah baik nya data tersebut divisualisasikan melalui fase ke 4 ini dengan tujuan untuk mempermudah dalam hal menyampaikan hasil temuan dari hasil data yang telah diolah pada sebelum nya di fase ke 3, dengan demikian harapan nya dengan visualisasi data. Partisipan akan lebih mudah untuk memahami dari hasil pemaparan tentang temuan yang telah dibuat.
- 5) Kesimpulan Analisis
- Pada fase akhir setelah menemukan temuan dari hasil pengolahan data langkah selanjutnya adalah dengan membuat kesimpulan dari hasil proses analisis data tersebut, tentu nya kesimpulan yang akan dibuat harus menjawab pertanyaan-pertanyaan yang telah didefinisikan pada tahap fase awal.

3.5.2 Konsep Statistika Dalam Analisis Data

Statistika merupakan cabang ilmu yang dapat digunakan untuk mengumpulkan, menata, menyajikan, menganalisis dan menginterpretasikan data menjadi sebuah informasi sehingga dapat membantu untuk pengambilan keputusan secara efektif. konsep statistika dalam analisis data merupakan cabang ilmu yang sangat penting dalam proses analisis data dengan tujuan untuk memahami lebih jauh tentang proses analisis data yang sedang dikerjakan nya, sehingga dari proses terjemahan data tersebut harapan nya bisa mendapatkan wawasan untuk dapat dipergunakan sebagai landasan dalam mengambil keputusan (Arifin, 2014)..

Terdapat beberapa konsep statistika yang biasa nya sering digunakan untuk mendeskripsikan data yang sedang di proses, proses mendeskripsikan data tersebut dikenal sebagai *descriptive statistics* (Setiawan, 2023), secara umum konsep tersebut adalah :

1) *Measuring Central Tendency*

Measuring Central Tendency adalah sebuah konsep ilmu statistika yang digunakan untuk mengidentifikasi suatu nilai khas ataupun nilai tengah dalam suatu kumpulan data. Secara umum *Measuring Central Tendency* terdiri dari :

a. Mean

Mean adalah salah satu parameter dalam ilmu statistika yang sering digunakan untuk memperhitungkan suatu nilai khas pada *central* dari suatu data yang telah terkumpul, untuk mencari nilai *mean* dapat diperoleh dengan cara menghitung terlebih dahulu jumlah dari keseluruhan data dan kemudian dibagi dengan banyak nya data yang telah terkumpul, Rumus yang dapat digunakan untuk mencari nilai *mean* dalam ilmu statistika.

$$Mean = \frac{\sum x}{n}$$

Keterangan :

$$\sum x = \text{jumlah keseluruhan data}$$

$$n = \text{jumlah banyak data}$$

Sebagai contoh untuk menggunakan rumus di atas hal pertama yang harus dilakukan adalah mengumpulkan jumlah keseluruhan data terlebih dahulu, semisal untuk contoh pada kasus ini akan menggunakan 10 data secara bebas, dengan demikian maka jumlah banyak nya data akan sama persis dengan jumlah keseluruhan data, sehingga data tersebut dapat dihitung dengan rumus dibawah.

$$\frac{2+3+4+6+9+8+5+1+7+2}{10} = \frac{47}{10} = 4,7$$

Sehingga hasil yang diperoleh dari nilai khas tersebut ataupun nilai *mean* dari sekumpulan data yang telah terkumpul diatas adalah 4,7.

b. Median

Median adalah salah satu parameter dalam ilmu statistika yang dapat digunakan untuk mencari nilai *sentral* dari keseluruhan data yang telah terkumpul, hampir mirip seperti *mean* dan *median* memiliki kelebihan, yaitu toleran terhadap suatu nilai *outlier*. Untuk mencari nilai *sentral*/menggunakan median terbagi menjadi 2 cara, jika terdapat total keseluruhan data berjumlah ganjil maka nilai *median* dapat diperoleh dengan cara mengambil langsung nilai tengah dari data yang telah dikumpulkan, namun dengan syarat data-data tersebut diurutkan terlebih dahulu mulai dari yang terkecil hingga ke yang terbesar. Namun akan sedikit berbeda jika total keseluruhan nilai berjumlah genap, untuk memperoleh nilai tengah dari kumpulan data yang berjumlah genap bisa diperoleh dengan mencari rata-rata 2 nilai yang berada di tengah. Diberikan cara yang dapat digunakan untuk mencari nilai *median* dalam ilmu statistika, berikut merupakan nilai acak yang dapat dijadikan sebagai contoh akan untuk menghitung median.

Bab 4: Infrastruktur Cloud Untuk Analisis Data

4.9 Model Penyebaran Infrastruktur Cloud

Model penyebaran infrastruktur cloud merupakan mekanisme fundamental yang menentukan bagaimana sumber daya komputasi, baik dalam bentuk perangkat lunak maupun perangkat keras, dapat disediakan kepada pengguna melalui jaringan internet. Model ini mengilustrasikan bagaimana sumber daya seperti server virtual, penyimpanan berbasis cloud, serta infrastruktur jaringan dapat diakses secara fleksibel tanpa adanya ketergantungan pada perangkat fisik tertentu. Dengan demikian, pengguna dapat memanfaatkan layanan cloud untuk mengelola, menyimpan, dan memproses data secara efisien tanpa harus membangun serta memelihara infrastruktur sendiri. Secara garis besar, terdapat tiga model utama dalam penyebaran infrastruktur cloud, yaitu *Infrastructure as a Service* (IaaS), *Platform as a Service* (PaaS), dan *Software as a Service* (SaaS), yang masing-masing menawarkan tingkat abstraksi dan kontrol yang berbeda bagi penggunaanya.

1. *Infrastructure as a Service* (IaaS)

Infrastructure as a Service (IaaS) merupakan model layanan yang memungkinkan pengguna untuk memperoleh akses terhadap sumber daya komputasi berbasis *cloud*, termasuk server virtual, jaringan, dan penyimpanan data yang disediakan oleh penyedia layanan *cloud*. Dalam model ini, pengguna memiliki kontrol penuh terhadap sistem operasi, *middleware*, serta aplikasi yang dijalankan di atas infrastruktur *cloud*, sedangkan penyedia layanan bertanggung jawab terhadap pemeliharaan perangkat keras, jaringan, dan lingkungan virtualisasi (AWS, 2023b). Keunggulan utama dari model IaaS adalah fleksibilitas serta skalabilitas yang tinggi, di mana pengguna dapat menyesuaikan kapasitas sumber daya komputasi sesuai dengan kebutuhan mereka tanpa harus melakukan investasi dalam infrastruktur fisik. Selain itu, dengan adopsi model *pay-as-you-go*, pengguna hanya dikenakan biaya

berdasarkan sumber daya yang mereka gunakan, sehingga dapat mengoptimalkan efisiensi biaya dalam pengelolaan infrastruktur TI.

2. *Platform as a Service (PaaS)*

Platform as a Service (PaaS) adalah model layanan cloud yang menyediakan lingkungan pengembangan dan eksekusi aplikasi berbasis *cloud*, termasuk *framework*, *middleware*, serta alat pengembangan perangkat lunak yang dapat digunakan oleh pengembang tanpa perlu mengelola infrastruktur dasar. Dalam model ini, penyedia layanan *cloud* bertanggung jawab atas pengelolaan sistem operasi, *runtime*, *database*, serta lingkungan pengembangan, sehingga pengguna dapat lebih fokus pada proses pengembangan aplikasi tanpa harus mempertimbangkan aspek infrastruktur (Konermann, 2016). Salah satu keunggulan utama dari PaaS adalah kemampuannya untuk mengurangi kompleksitas dalam proses pengembangan aplikasi, terutama dalam aspek kompatibilitas platform dan integrasi layanan. Dengan adanya dukungan otomatis terhadap berbagai sistem operasi dan lingkungan pemrograman, pengembang dapat dengan mudah mengembangkan, menguji, serta menerapkan aplikasi mereka tanpa harus khawatir mengenai kompatibilitas dengan perangkat keras atau konfigurasi sistem yang rumit. Selain itu, banyak platform PaaS yang telah mencakup pengelolaan lisensi perangkat lunak, sehingga pengguna tidak perlu menanggung biaya tambahan terkait hak penggunaan perangkat lunak yang mereka gunakan dalam lingkungan cloud.

2. *Software as a Service (SaaS)*

Software as a Service (SaaS) merupakan model layanan *cloud* yang menyediakan akses langsung terhadap aplikasi perangkat lunak yang dijalankan di infrastruktur penyedia layanan cloud. Dalam model ini, pengguna tidak perlu melakukan instalasi, konfigurasi, atau pemeliharaan aplikasi secara langsung, karena semua proses tersebut ditangani oleh penyedia layanan. Pengguna cukup mengakses aplikasi melalui antarmuka berbasis web atau API, yang memungkinkan aksesibilitas yang lebih fleksibel dan efisien. Keunggulan utama dari SaaS adalah kemudahan penggunaan, skalabilitas, serta efisiensi biaya, di mana pengguna dapat mengakses berbagai aplikasi tanpa perlu mengkhawatirkan pengelolaan perangkat keras maupun lisensi perangkat lunak. Model ini juga memungkinkan pengguna untuk secara otomatis mendapatkan pembaruan perangkat lunak, perbaikan *bug*, serta fitur keamanan terbaru, yang dapat meningkatkan stabilitas dan keamanan sistem. SaaS banyak digunakan dalam berbagai sektor

industri, termasuk perkantoran, pendidikan, kesehatan, serta bisnis *e-commerce*, dengan contoh aplikasi seperti Google Workspace, Microsoft 365, serta Salesforce yang menyediakan berbagai layanan berbasis cloud yang dapat diakses kapan saja dan di mana saja.

Model penyebaran infrastruktur cloud menawarkan solusi yang fleksibel dan efisien bagi organisasi maupun individu dalam memanfaatkan sumber daya komputasi tanpa harus berinvestasi dalam infrastruktur fisik yang kompleks. IaaS memberikan fleksibilitas dalam pengelolaan sumber daya infrastruktur, PaaS menyederhanakan pengembangan aplikasi dengan menyediakan platform yang telah dikonfigurasi sebelumnya, sedangkan SaaS memungkinkan pengguna untuk mengakses perangkat lunak tanpa perlu mengkhawatirkan aspek teknis dalam pengelolaannya. Dengan memahami perbedaan serta keunggulan dari masing-masing model ini, pengguna dapat memilih solusi yang paling sesuai dengan kebutuhan mereka dalam mengadopsi teknologi *cloud computing* secara optimal.

4.10 Layanan Big Data

Apache Hadoop dan *Apache Spark* adalah dua teknologi utama dalam pemrosesan data skala besar. Keduanya digunakan dalam *big data analytics*, tetapi memiliki perbedaan mendasar dalam arsitektur dan kinerja. Berikut perbedaan utama antara *Hadoop* dan *Spark* serta keunggulan masing-masing.

1. *Apache Hadoop* adalah kerangka kerja open-source yang dirancang untuk penyimpanan dan pemrosesan data dalam skala besar dengan model pemrograman *MapReduce*. *Hadoop* terdiri dari beberapa komponen utama:
 - a. HDFS (*Hadoop Distributed File System*) memiliki arti sebagai sistem penyimpanan terdistribusi yang memungkinkan data disimpan di berbagai node.
 - b. *MapReduce* merupakan Model pemrograman yang memungkinkan pemrosesan data dalam bentuk peta dan pengurangan.
 - c. YARN (*Yet Another Resource Negotiator*) berarti manajer sumber daya untuk menangani penjadwalan tugas dan pengelolaan kluster.

Hadoop memiliki keunggulan yang dapat menangani *dataset* yang sangat besar, dapat pada skalabilitas tinggi, mendukung berbagai format data dan *Open-source* dan didukung oleh komunitas besar. Adapun kelemahan yang dihadapi pada hadoop yaitu kecepatan pemrosesan lambat karena berbasis *disk* dan kurang efisien untuk pemrosesan data *real-time*.

2. *Apache Spark* adalah kerangka kerja pemrosesan data yang dikembangkan untuk meningkatkan kecepatan pemrosesan dibandingkan dengan *Hadoop MapReduce*. *Spark* menggunakan konsep *in-memory computing* yang memungkinkan data disimpan dalam memori untuk meningkatkan kinerja. *Spark* memiliki komponen utama yang diberikan diantaranya:
 - a. *Spark Core* adalah Mesin pemrosesan utama.
 - b. *Spark SQL* ini merupakan modul untuk pemrosesan data terstruktur
 - c. *Spark Streaming* ini juga mendukung pemrosesan data *real-time*.
 - d. *MLlib (Machine Learning Library)* adalah sebuah pustaka di *machine learning*
 - e. *GraphX* yaitu Modul untuk analisis graf.

Apache Spark memiliki beberapa keunggulan, seperti kecepatan pemrosesan yang lebih tinggi dibandingkan *Hadoop MapReduce* serta dukungan untuk pemrosesan *real-time*. Dengan *in-memory processing*, *Spark* mampu mengurangi latensi secara signifikan. Selain itu, fleksibilitasnya terlihat dari dukungan terhadap berbagai bahasa pemrograman, termasuk Java, Scala, Python, dan R. Namun, *Spark* juga memiliki beberapa kelemahan, seperti konsumsi memori yang lebih besar dibandingkan *Hadoop* serta efisiensi yang lebih rendah dalam pemrosesan *batch* pada *dataset* yang sangat besar, berikut adalah perbandingan utama antara *Apache Hadoop* dan *Apache Spark*.

Aspek	<i>Apache Hadoop</i>	<i>Apache Spark</i>
Kecepatan	Lambat (<i>disk-based</i>) dalam satu <i>batch</i> besar.	Cepat (<i>in-memory</i>) kontinu.
Pemrosesan	<i>Batch Processing</i>	<i>Batch & Streaming</i>
Skalabilitas	Sangat Skalabel	Skalabel, tetapi membutuhkan lebih banyak RAM
Kompleksitas	Relatif Kompleks	Lebih Sederhana
Keperluan Memori	Rendah (<i>disk</i>)	Tinggi (RAM)

Tabel 4.2 Perbandingan *Apache Hadoop* dan *Apache Spark*

Bab 5: Keamanan Data

5.1 Pengantar

Di zaman yang serba digital seperti sekarang ini, tidak bisa dipungkiri bahwa data merupakan suatu hal yang lumrah ditemui di dalam aktivitas kehidupan sehari-hari, namun seiring masif nya dalam penggunaan data tersebut maka menimbulkan berbagai masalah yang serius yaitu tentang ancaman keamanan data, fakta nya bahwa banyak kasus yang terjadi pada saat ini seperti penyalahgunaan atau pencurian data oleh sekelompok pihak yang tidak berwenang, dan tentu nya itu apabila data tersebut berhasil dicuri ataupun di salah gunakan oleh sekelompok orang yang tidak memiliki wewenang maka akan merugikan seseorang ataupun organisasi (Suryantoro *et al.*, no date), sebagai contoh semisal terdapat sebuah bank A dan kemudian bank tersebut memiliki tugas untuk mengelola keuangan nasabah-nasabah nya. Namun seiring nya berjalan nya waktu tidak dapat sangka bahwa bank A tersebut mengalami insiden dengan berhasilnya diretas oleh sekelompok *cracker*, sehingga *cracker* tersebut dapat mengakses data nasabah-nasabah bank A, tentu nya dengan keberhasilan *cracker* tersebut dalam mengakses data nasabah maka akan banyak sekali pihak yang dirugikan oleh insiden tersebut terutama nasabah dan bank A, selanjut nya kemungkinan yang sangat besar terjadi setelah insiden tersebut adalah berkurang nya kepercayaan publik ataupun seorang nasabah terhadap bank A tersebut, karena pada dasar nya nasabah akan menganggap bahwa bank A telah gagal dalam mengamankan suatu data sensitif dan privasi nasabah, tentu nya itu akan berdampak besar dan merugikan suatu bank, karena bank tersebut telah kehilangan kepercayaan dan reputasi baik di depan publik ataupun nasabah, karena atas kegagalan dalam melindungi keamanan data-data sensitif dan privasi, Maka dari itu keamanan data akan menjadi suatu aspek yang sangat penting di zaman yang serba digital seperti sekarang ini dengan tujuan untuk menghindari kerugian-kerugian yang sewaktu-waktu bisa saja terjadi seperti pada skenario diatas. Maka dari itu pembahasan tentang keamanan data akan menjadi suatu aspek yang harus diketahui oleh masyarakat umum guna menghindari resiko-resiko yang kemungkinan dapat merugikan seseorang ketika menggunakan layanan *platform* digital. Untuk memahami tentang resiko dan ancaman terhadap keamanan data, pada bab ini telah disajikan berupa istilah yang umum dijumpai di kehidupan sehari-hari tentang resiko ancaman terhadap data.

5.1.1 Definisi dan Tujuan Keamanan Data

Keamanan data merupakan suatu proses ataupun praktik yang dapat dilakukan untuk memproteksi data dari akses yang tidak sah, tujuan dari praktik keamanan data sendiri yaitu mengacu pada upaya untuk menjaga kerahasiaan (*Confidentiality*), integritas (*Integrity*), ketersediaan (*Availability*) data yang telah disimpan di dalam suatu sistem ataupun insfastruktur media yang lain (The Open University, 2023). Namun perlu di ketahui bahwa dengan menerapkan praktik keamanan data tidak ada jaminan sepenuhnya bahwa data yang disimpan akan aman dari ancaman *cyber*, akan tetapi setidaknya setelah menerapkan praktik keamanan data, suatu data dalam sistem dapat diminimalisir dari tindakan yang tidak diinginkan seperti perubahan akses, penghapusan, ataupun modifikasi data yang dilakukan secara tidak sah. Mengamankan data merupakan suatu kewajiban yang harus disadari oleh seseorang ataupun organisasi karena mengamankan informasi dari tindakan yang tidak diinginkan seperti akses, perubahan, penghapusan, atau pengungkapan yang tidak sah merupakan inti dari tujuan keamanan data.



Gambar 5.1 Elemen Keamanan Data.

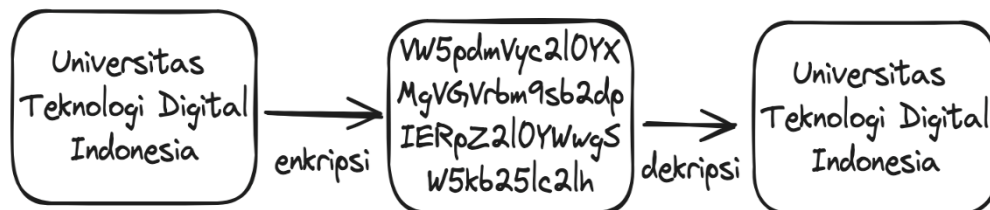
Secara garis besar terdapat beberapa aspek penting yang dapat dilakukan untuk menerapkan praktik terbaik untuk keamanan data tersebut, diantara nya adalah :

a. Kerahasiaan (*Confidentiality*)

Aspek kerahasiaan pada keamanan data mengacu kepada kondisi dimana setiap akses terhadap suatu data hanya akan diberikan kepada pihak yang memiliki hak yang sah ataupun otoritas, pada aspek *confidentiality* harus dapat menjamin bahwa data hanya dapat diakses dan digunakan oleh pihak yang memiliki wewenang. Metode umum yang biasa nya digunakan untuk

menjaga kerahasiaan pada data adalah dengan metode enkripsi.

Enkripsi merupakan salah satu metode kriptografi yang banyak digunakan ketika pengguna mempertukarkan informasi data pada sistem, pada dasarnya enkripsi digunakan untuk menjaga aspek kerahasiaan pada data agar proses pertukaran data yang dikirim melewati sistem tidak dapat dilihat oleh orang lain, cara kerja dari enkripsi sendiri adalah untuk mengubah informasi data yang dapat dibaca (*plaintext*) menjadi bentuk yang sama sekali tidak dapat dibaca (*ciphertext*) dan kemudian mengembalikan lagi ke semula agar data dapat dibaca, berikut merupakan suatu contoh ilustrasi sederhana bagaimana cara kerja enkripsi terhadap suatu data.



Gambar 5.2 Contoh Ilustrasi Enkripsi.

Sepintas jika dilihat pada gambar diatas, mula-mula nya data masih dapat dibaca secara jelas, namun ketika pengguna ingin berbagi informasi data tersebut dengan teman nya maka data yang akan dikirimkan harus melewati proses enkripsi, proses enkripsi sendiri akan mengubah bentuk pesan asli menjadi bentuk yang sama sekali tidak dapat dibaca sehingga yang mula-mula nya data dapat dibaca atau dipahami akan menjadi data yang sama sekali tidak dapat dibaca, kemudian ketika data telah melewati proses enkripsi maka data akan dikirimkan kepada teman nya melalui suatu sistem komunikasi. Ketika pesan telah sampai kepada penerima maka yang semula nya data telah dienkripsi akan di dekripsi terlebih dahulu sebelum diterima oleh penerima, proses deskripsi dilakukan agar data informasi yang telah dienkripsi kembali menjadi pesan asli yang dapat dibaca ataupun dipahami oleh penerima.

b. Integritas (Integrity)

Aspek integritas pada keamanan data mengacu kepada upaya untuk menjaga data dari percobaan modifikasi ataupun manipulasi yang dilakukan secara tidak sah oleh pihak yang tidak memiliki otoritas atau wewenang. Pada aspek *integrity* harus dapat menjamin bahwa data sifat nya harus murni atau tidak dimanipulasi oleh pihak yang tidak memiliki otoritas, metode umum untuk menjaga integritas pada data adalah dengan menggunakan metode hashing atau tanda tangan digital. Hashing sendiri merupakan salah satu metode kriptografi juga yang banyak digunakan untuk menjaga keaslian pada data, keaslian yang dimaksud adalah bahwa

ketika data sedang diakses/digunakan maka data tersebut ketika dalam proses pengiriman pada sistem harus bisa menjamin bahwa sumber data yang digunakan harus selalu dalam keadaan *original* seperti sedia kala pada sumber nya, cara kerja hash mirip seperti enkripsi yaitu untuk mengacak sebuah *plaintext* yang ditulis menjadi *chipertext*, namun yang membedakan dari enkripsi bahwa hash fungsi utama nya adalah untuk menjaga suatu integritas data, semisal jika seseorang mencoba untuk mengunduh suatu file, maka file tersebut harus menjamin bahwa file yang telah di unduh nya merupakan file asli dan tidak ada proses perubahan data selama proses pengunduhan nya, untuk mengetahui apakah file tersebut tidak dimodifikasi oleh pihak yang tidak berwenang maka cara sederhana nya bisa membandingkan nilai hash dari sumber file yang diunduh nya.

c. Ketersediaan (Availability)

Aspek ketersediaan pada keamanan data mengacu kepada kondisi dimana pengguna berhak mengakses data kapanpun selagi membutuhkan, pada aspek *availability* harus dapat menjamin bahwa ketersediaan data yang telah disimpan pada infrastruktur harus dikelola dengan baik untuk menjaga ketersediaan data secara *real-time*. Ideal nya data yang telah disimpan dalam infrastruktur dilakukan pencadangan (*backup*) data secara teratur dan memiliki rencana pemulihan (*recovery*) jika suatu waktu terjadi kegagalan sistem.

d. Autentikasi (Authentication)

Aspek autentikasi pada keamanan data mengacu kepada proses untuk melakukan verifikasi identitas seorang pengguna ataupun sistem sebelum akses data diberikan ke pengguna., pada aspek *authentication* harus dapat menjamin bahwa pengguna bisa membuktikan kepada sistem bahwa diri nya merupakan *user* yang memiliki otoritas yang sah atas penggunaan data tersebut.

e. Non-Repudiation

Aspek non-repudiation pada keamanan data mengacu kepada proses untuk memastikan bahwa pengguna yang terlibat dalam pengaksesan sebuah data dicatat pada log suatu sistem, aspek ini akan menjadi penting karena untuk memastikan bahwa pengguna tidak dapat menyalah gunakan data berdasarkan peraturan yang telah disepakati. Non-repudiation sendiri akan menjadi bukti digital dan dapat diaudit apabila pengguna yang mencoba untuk menyalah gunakan data menyangkal dalam keterlibatan penyalahan gunaan data tersebut, sehingga dengan prinsip non-repudiation maka suatu sistem dapat di lacak aktivitas atas penggunaan data nya.

Bab 6: Studi Kasus

6.1 Pengantar

Pada era digital seperti saat ini, kebutuhan untuk melakukan analisis dan visualisasi data tidak hanya dapat dilakukan di komputer pribadi ataupun perusahaan saja, saat ini kebutuhan analisis dan visualisasi data dapat dilakukan dengan menggunakan layanan *cloud*, tentu nya dengan menggunakan layanan *cloud* tersebut akan jauh lebih cepat dan efisien ketika memproses suatu data dalam jumlah kapasitas yang besar, saat ini *cloud computing* telah menjadi suatu solusi yang dapat menawarkan masalah tantangan seperti ruang penyimpanan data yang terbatas, kelambatan dalam mengelola data dalam skala besar, dan masih banyak lagi manfaat yang dapat dirasakan ketika menggunakan layanan *cloud* berkat keunggulan nya dalam menawarkan fitur skalabilitas, fleksibilitas dan efisiensi biaya. Pada studi kasus kali ini pembahasan yang akan dipaparkan adalah bagaimana suatu data dapat diolah di lingkungan *cloud* dan kemudian memvisualisasi kan nya kedalam bentuk grafik, tujuan dari divisualisasikan ke dalam grafik tentu nya agar lebih mudah untuk dipahami oleh partisipan ataupun oleh para pemangku kepentingan lain nya. Selanjut nya pada pembahasan kali ini akan di paparkan secara lebih rinci bagaimana suatu data dapat dianalisis dilingkungan *cloud* dan kemudian hasil nya akan diformat ke dalam bentuk grafik agar seseorang yang membaca data tersebut mudah untuk memahami nya.

6.1.1 Latar Belakang Kasus

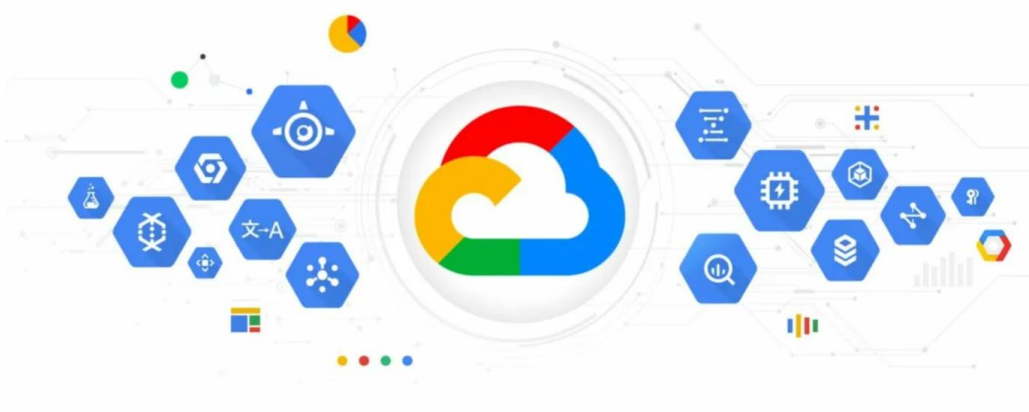
Saat ini cloud computing menjadi suatu solusi yang ideal untuk menghadapi tantangan penyimpanan dan pengolahan data dalam skala besar, berkat kemampuannya dalam menawarkan skalabilitas, fleksibilitas, dan efisiensi biaya. Pada kasus yang akan dipaparkan di pembahasan bab ini rencananya nya akan membuat suatu analisis data yang dihasilkan oleh *platform e-commerce*, dan kemudian harapan nya setelah data di analisis dapat menemukan wawasan ataupun karakteristik pola yang berharga untuk keberlangsungan suatu bisnis agar organisasi ataupun perusahaan dapat mengambil suatu keputusan ataupun tindakan yang tepat sesuai dengan trend pada karakteristik data, karena pada *era* saat ini perkembangan platform e-commerce telah menghasilkan sumber data yang cenderung sangat berlimpah dan bervariasi, tentu nya dari data yang dihasilkan oleh platform *e-commerce* tersebut mencakup tentang informasi-informasi yang berharga seperti data transaksi, perilaku

konsumen dalam membeli produk, inventaris, dan kinerja selama rentang waktu penjualan. Namun apabila data pada *platform e-commerce* tersebut dibiarkan begitu saja, maka dari data-data yang dihasilkan tidak dapat memberikan wawasan yang berharga untuk keberlangsungan bisnis yang sedang di jalani nya, sehingga untuk mendapat wawasan dari data yang telah dihasilkan tersebut perlu ada nya kemampuan khusus untuk melakukan analisis data dengan cepat dan efektif, karena dengan kemampuan tersebut akan menjadi salah satu kunci keberhasilan suatu bisnis *e-commerce* yang sedang dijalani nya.

6.1.2 Tujuan dan Lingkup Kasus

Tujuan dari kasus ini bagaimana melakukan analisis data yang dihasilkan oleh *platform e-commerce* menggunakan layanan *cloud*, karena tentu nya hal ini akan menjadi suatu tantangan tersendiri karena akan banyak sekali tools yang tentu nya akan digunakan untuk melakukan analisis dan visualisasi data di lingkungan *cloud*. Selain menjadi tantangan tersendiri dalam situasi dan kondisi ini bagaimana suatu data dapat dikelola secara efisien dan efektif, bagaimana menganalisis pola pembelian terhadap konsumen dan yang paling penting adalah bagaimana cara menyajikan hasil analisis data tersebut kedalam bentuk format visualisasi agar mudah dimengerti oleh partisipan ataupun para pemangku kepentingan lain nya. Namun lingkup pembahasan yang dipaparkan pada bab ini adalah mengidentifikasi pola pembelian terhadap konsumen, tren penjualan dan optimalisasi strategi pemasaran sesuai dengan hasil analisis.

6.2 Implementasi Analisis Data Pada Platform Cloud



Gambar 6.1 Layanan Google Cloud.

Untuk kasus analisis dan visualisasi data pada pembahasan ini akan dilakukan pada lingkungan *cloud*, tentu nya sebelum memulai tahapan implementasi maka yang harus dilakukan terlebih dahulu adalah memilih

platform cloud yang akan digunakan untuk melakukan proses analisis dan visualisasi data, akan tetapi pada kasus ini *platform cloud* yang akan digunakan adalah Google Cloud. Alasan menggunakan platform tersebut adalah mudah digunakan dan interaktif, selain alasan tersebut Google Cloud memiliki fitur modern yang tentu nya dapat digunakan untuk melakukan proses analisis dan visualisasi data dengan cepat dan mudah, dan terakhir yang paling menonjol dari penawaran platform ini adalah dapat di uji coba secara gratis dengan kredit senilai 300 USD ketika ingin membangun solusi untuk optimasi pengelolaan data dan reporting (Dzikriyana and Sulianta Feri, 2022).

6.2.1 Data Yang Digunakan

Data yang akan digunakan untuk kasus analisis dan visualisasi data pada pembahasan ini adalah theLook eCommerce yang dikembangkan oleh Looker tim, dataset tersebut dipilih karena ini merupakan data publik yang bersifat bebas untuk digunakan ataupun di eksplorasi, sehingga dengan memanfaatkan dataset yang disediakan oleh Google Cloud maka untuk implementasi akan lebih menitik beratkan kepada analisis beserta visualisasi di lingkungan *cloud*, dari data tersebut telah di deskripsikan secara lebih rinci tentang informasi-informasi data. Data tersebut memberikan informasi data tentang pelanggan, produk, pesanan, logistik, dan kampanye pemasaran produk secara digital (Hana Haura, 2023), sehingga dari informasi yang didapat bahwa data yang dipilih selaras dengan tujuan studi kasus yang akan dibuat pada bahasan bab ini. Berikut merupakan informasi tabel dasar yang dimuat pada dataset tersebut.

No	Nama Tabel
1	distribution_centers.csv
2	<i>events.csv</i>
3	inventory_items.csv
4	order_items.csv
5	orders.csv
6	products.csv
7	users.csv

Tabel 6.1 Informasi Dasar Dataset.

Namun pada studi kasus pembahasan ini, dataset yang akan digunakan justru hanya empat tabel, diantara nya adalah orders_items.csv, orders.csv, product.csv, dan users.csv, alasan memilih empat tabel tersebut karena pada dasar nya tabel yang dipilih telah memenuhi syarat dan relevansi untuk di eksplorasi lebih lanjut pada pembahasan bab ini.

6.2.2 Pemilihan Tools Google Cloud Untuk Analisis Data

Dengan menggunakan platform Google Cloud, berbagai tools ataupun fitur yang ditawarkan oleh penyedia layanan *cloud* tersebut dapat digunakan untuk melakukan proses analisis data dalam kapasitas volume data yang besar, namun dari platform yang ditawarkan oleh penyedia layanan cloud tersebut terdapat suatu fitur unggulan yang dapat digunakan untuk melakukan layanan analisis data yang memungkinkan dapat menjalankan analisis data dengan query kompleks pada dataset dan berkinerja tinggi, nama fitur tersebut adalah Google BigQuery (Google Cloud, 2024). BigQuery merupakan suatu fitur yang ditawarkan oleh Google untuk mengelola penyimpanan ataupun menganalisis data yang besar berbasis SQL, sehingga fitur ini dapat membantu proses pencarian wawasan dari data yang diolah untuk kebutuhan bisnis yang sedang dijalani oleh organisasi ataupun perusahaan. BigQuery sendiri memiliki beberapa kelebihan untuk melakukan proses analisis data, diantara nya adalah :

a. Skalabilitas dan Performa

BigQuery di desain untuk menangani jumlah data yang besar, sehingga seseorang memungkinkan dapat menjalankan query untuk melakukan proses analisis data tanpa harus mengelola infrastruktur dari perangkat tersebut, tentu nya kondisi ini sangat berbeda jika seseorang mengelola analisis data di server on-premise yang selalu harus memperhatikan infrastruktur yang digunakan untuk proses analisis data, selanjut nya BigQuery sendiri menggunakan arsitektur yang terdistribusi sehingga pemrosesan query dapat dibagi-bagi ke beberapa node untuk mempercepat proses eksekusi query walaupun dataset yang diolah memiliki jumlah data dalam kapasitas yang besar.

b. Serverless

Karena BigQuery dilakukan di atas layanan cloud, maka seseorang yang menggunakan layanan ini tidak perlu repot untuk mengelola server seperti instalasi, penyediaan infrastruktur, konfigurasi, penskalaan terhadap infrastruktur dan yang lain nya. Sehingga pengguna akan lebih cenderung fokus dan menitik beratkan terhadap proses analisis untuk mencari wawasan terhadap data yang sedang dianalisis. Selain

serverless BigQuery akan menyiapkan skalasi otomatis untuk menyediakan sumber daya yang sesuai dengan jumlah ukuran query yang diproses, sehingga pengguna tinggal menggunakannya saja tanpa harus khawatir tentang *over-provisioning* atau *under-provisioning*.

c. Efisiensi Biaya

BigQuery menerapkan prinsip *pay-as-you-go* yang artinya bahwa beban pembayaran untuk menggunakan layanan ini akan dihitung berdasarkan jumlah data yang diproses pada query, sehingga untuk menggunakan layanan ini seseorang hanya perlu membayar untuk data yang di analisis saja, akan tetapi yang menjadi masalah ketika menggunakan prinsip *pay-as-you-go* adalah terkadang untuk mengontrol penggunaan query tidak luput lupa dari batas yang telah ditentukan, sehingga itu akan menimbulkan pembengkakan biaya untuk menggunakan layanan ini, namun untuk mengatasi masalah ini layanan google cloud memiliki fitur kontrol biaya yang bisa dimanfaatkan untuk mengatur batas biaya query untuk menghindari biaya tagihan yang tidak terduga. Akan tetapi pada kondisi saat ini BigQuery dapat di uji coba secara gratis tanpa ada biaya yang harus dikeluarkan sehingga ini cocok digunakan seseorang untuk mengeksplorasi layanan ini sebelum akhirnya memutuskan menggunakan BigQuery.

d. Terintegrasi dengan tools Google lainnya

Dengan menggunakan layanan BigQuery maka tools pendukung lain seperti Google Data Studio/Looker (Layanan Visualisasi Data) ataupun tools lainnya dapat terintegrasi dengan mudah, sehingga ini akan memudahkan

Bab 7: Tantangan Dan Solusi

7.1 Pengantar

Dalam era transformasi digital yang kian pesat, analisis dan visualisasi data telah menjadi komponen strategis yang mendasari proses pengambilan keputusan berbasis data (data-driven decision making) di berbagai sektor industri. Kemampuan untuk mengolah, menganalisis, dan memvisualisasikan data secara efektif bukan lagi sekadar keunggulan kompetitif, melainkan sebuah kebutuhan mendesak bagi organisasi dan perusahaan yang ingin tetap relevan dan adaptif terhadap perubahan pasar dan perilaku konsumen yang dinamis. Seiring dengan meningkatnya volume, kecepatan, dan variasi data yang dihasilkan baik dari sistem internal, sensor IoT, media sosial, maupun transaksi digital dengan paradigma pengelolaan data pun mengalami pergeseran yang signifikan. Jika sebelumnya infrastruktur data banyak bergantung pada perangkat keras dan perangkat lunak on-premise, saat ini organisasi cenderung mengadopsi pendekatan berbasis cloud computing untuk mendukung proses analisis dan visualisasi data secara lebih elastis, skalabel, dan hemat biaya.

Cloud computing memberikan fleksibilitas dan kapabilitas komputasi tinggi untuk menangani big data dalam skala besar dengan kecepatan dan efisiensi yang lebih optimal. Layanan cloud modern, seperti Amazon Web Services (AWS), Microsoft Azure, dan Google Cloud Platform (GCP), menyediakan ekosistem lengkap untuk integrasi, pemrosesan, analisis, serta visualisasi data dalam satu lingkungan yang terdistribusi dan mudah diakses. Namun, meskipun menawarkan berbagai keunggulan, implementasi analisis dan visualisasi data di lingkungan cloud tidak lepas dari sejumlah tantangan teknis maupun non-teknis. Tantangan-tantangan tersebut mencakup manajemen volume data yang terus bertambah secara eksponensial, kebutuhan akan arsitektur penyimpanan dan komputasi yang andal, keterbatasan dalam optimisasi performa analisis real-time, serta isu-isu kritis terkait keamanan, privasi, dan kepatuhan terhadap regulasi data. Di samping itu, kurangnya kompetensi sumber daya manusia dalam mengelola tools cloud dan pipeline data modern seringkali menjadi kendala tambahan dalam proses implementasi.

Sehingga pada era sekarang, bahwa analisis dan visualisasi data telah menjadi bagian yang sangat penting demi keberlangsungan suatu bisnis perusahaan ataupun organisasi agar terus beroperasi, sehingga pada bab

ini akan dibahas dan mengelaborasi secara mendalam berbagai tantangan utama yang umum ditemui dalam proses analisis dan visualisasi data di cloud, serta menyajikan pendekatan solusi yang relevan baik dari sisi teknologi maupun strategi organisasi, untuk memastikan proses analitik yang berkelanjutan dan bernilai tambah di era komputasi awan.

7.2 Tantangan Umum Dalam Analisis dan Visualisasi Data Di Cloud

Ketika sebuah organisasi ataupun perusahaan telah memutuskan untuk mengelola data menggunakan lingkungan *cloud*, maka itu adalah keputusan yang tepat untuk mengelola analisis dan visualisasi data dalam kapasitas data yang memiliki ukuran volume yang besar, pada dasar nya solusi *cloud* telah menawarkan waktu yang cepat dan efisien dalam mengelola data yang memiliki jumlah ukuran volume yang besar, akan tetapi pada kenyataannya ketika memutuskan untuk mengelola data menggunakan lingkungan *cloud* maka terdapat beberapa tantangan yang harus dicermati secara seksama ketika sudah memutuskan menggunakan layanan *cloud*, beberapa cara dalam mengelola tantangan di lingkungan *cloud* pada saat ini dapat dikatakan semakin sulit dan kompleks. Berikut merupakan tantangan-tantangan umum beserta solusi yang sering dijumpai dan dilakukan ketika mencoba melakukan implementasi layanan *cloud*.



Gambar 7.1 Diagram Flow Tantangan Umum.

7.2.1 Kesulitan Dalam Integrasi Data

Kesulitan dalam melakukan integrasi data terhadap layanan *cloud* merupakan hal lumrah yang umum ditemui di lapangan, karena data yang telah di ambil dari berbagai sumber secara umum memiliki bentuk format data yang berbeda-beda sehingga memerlukan waktu dan proses yang cukup lama dalam mengintegrasikan data tersebut kedalam lingkungan *cloud* (Saadia, 2021). Berikut merupakan tantangan dalam melakukan integrasi data kedalam lingkungan cloud.

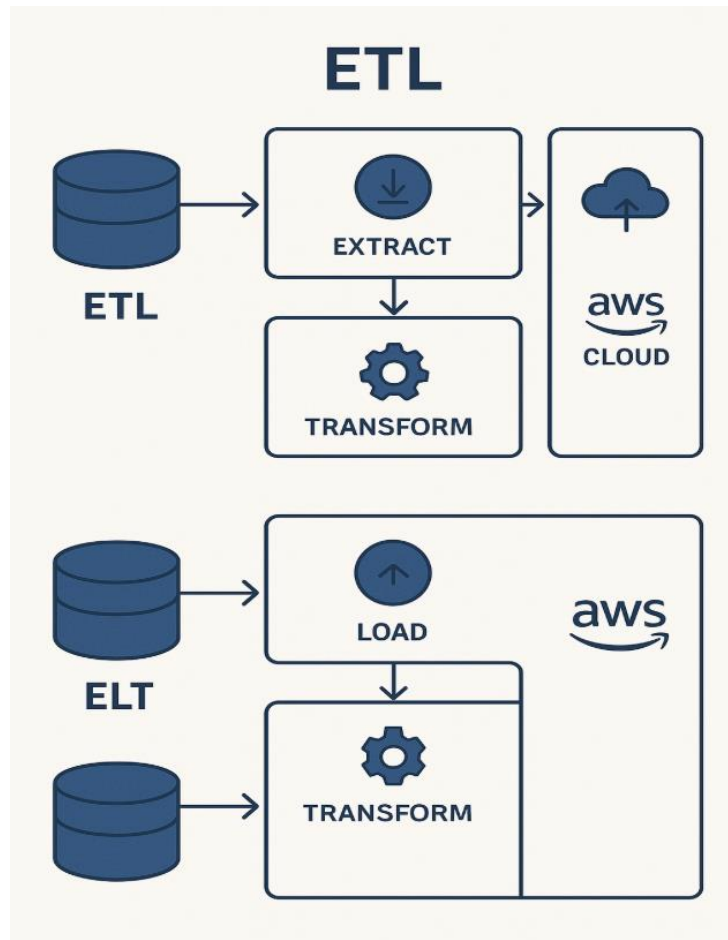
a. Tantangan Integrasi Data Dari Berbagai Sumber

Data yang telah berhasil dikumpulkan dari berbagai sumber seringkali memiliki bentuk ataupun format yang berbeda-beda, sehingga itu akan membuat sulit seseorang ketika berupaya untuk melakukan integrasi data, sebagai contoh semisal terdapat suatu dokumen yang telah berhasil dikumpulkan dari sumber X dan dari salah satu data tersebut memiliki nilai nya berisi tanggal pemrosesan data semisal (2 Juli 2024), namun seseorang telah mendapatkan data kembali yang bersumber dari Y dan dari salah satu data tersebut memiliki nilai yang berisi tanggal pemrosesan data seperti yang terdapat pada sumber X namun data tersebut ditulis (July 02, 2024), dengan perbedaan format tersebut tentu nya akan menyulitkan seseorang ketika akan melakukan integrasi data, namun solusi untuk kasus seperti demikian dapat diatasi dengan cara sebagai berikut :

1. Konsolidasi Data

Konsolidasi data merupakan sebuah teknik yang dapat digunakan untuk melakukan aksi membersihkan atau mengesktraksi data sebelum data yang berhasil dikumpulkan dari berbagai sumber tersebut pada akhir nya disimpan di lokasi penyimpanan pusat data (Amazon Web Services, 2023), dengan teknik konsolidasi maka data akan diseleksi terlebih dahulu dari yang mula-mula nya data tersebut dalam keadaan kotor, dan setelah dibersihkan data tersebut kemudian diseragamkan dalam bentuk format yang telah disepakati, untuk prinsip konsolidasi data

terbagi menjadi dua bagian besar diantara nya adalah:



Gambar 7.2 Konsolidasi Data dengan ETL dan ELT AWS

- **ETL**

ELT (*Extract, Transform, Load*) merupakan sebuah teknik yang dapat dilakukan untuk memformat atau menata ulang data dari berbagai sumber dan kemudian mengubah nya menjadi suatu standar atau format yang telah disepakati bersama oleh organisasi ataupun perusahaan, sehingga harapan nya setelah melewati proses ETL maka data dapat disimpan di lokasi penyimpanan data, data yang telah rapih selanjut nya dapat digunakan untuk berbagai macam kebutuhan seperti *Business Intelligence*, pencarian pola dengan menggunakan *machine learning* ataupun yang lain nya.

- **ELT**

ELT (*Extract, Load, Transform*) mirip seperti

pada proses ETL namun ELT mengalihkan dua proses data akhir secara berurutan, paradigma ELT pada dasarnya ketika data berhasil diekstrak maka data-data tersebut akan disimpan terlebih dahulu di lokasi penyimpanan data, data tersebut akan diubah ataupun diproses ketika memang akan dibutuhkan saja.