

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Untuk melakukan penelitian, berikut penelitian terdahulu sebagai referensi dalam penelitian ini :

Tabel 2 1 Penelitian terdahulu

No	Sumber	Hasil penelitian	Adopsi
1.	<i>Improve HR Decision-Making Based On Data Mart and OLAP</i> <i>Journal of Physics: Conference Series</i> <i>IOP Publishing, 2020 [7]</i>	Paper ini mempersembahkan kerangka kerja data mart HR dengan implementasi OLAP berdasarkan dataset informasi karyawan pensiunan dari <i>Basra Oil Company</i> , karena data yang digunakan untuk membangun <i>data mart</i> sedikit dibandingkan dengan data yang dimaksudkan untuk mengimplementasikan DW, maka <i>data mart</i> secara <i>bottom up</i> merupakan pilihan terbaik untuk membangun <i>data mart</i> HR departemen. <i>Data mart</i> HR adalah data mart independen yang dapat digunakan nantinya untuk membangun DW perusahaan yang dependen penuh.	Membangun <i>data mart</i> dengan jenis <i>independent data mart</i> dan metode <i>bottom up</i>
2.	<i>Star Schema Implementation For Monitoring in Data Quality Management Tool (A Case Study at A Government Agency)</i> [8] <i>Date of Conference :</i> 16-17 October 2019 <i>Date Added to IEEE Xplore :</i> 10 February 2020 <i>2019 Fourth International Conference on Informatics and Computing (ICIC)</i>	Dalam paper ini, kami mengusulkan sebuah skema bintang yang akan digunakan sebagai referensi untuk mengembangkan pemantauan pada alat manajemen kualitas data sumber terbuka dan dapat digunakan dalam berbagai organisasi. Skema bintang memainkan peran penting dalam pengembangan pemantauan manajemen kualitas data. Menggunakan skema bintang dapat menerjemahkan kebutuhan bisnis yang telah ditetapkan sebelumnya. Visualisasi pemantauan juga memainkan peran sehingga kualitas data yang dimonitor terlihat baik. Penelitian yang sedang kami kerjakan saat ini adalah proses menyelesaikan pengembangan manajemen kualitas data pada tahap pembersihan data, integrasi data, dan pemantauan kualitas data dalam parameter dimensi kualitas data lainnya seperti	Implementasi <i>Star Schema</i> dalam membangun dimensial model <i>data mart</i>

No	Sumber	Hasil penelitian	Adopsi
		kardinalitas data, pola data, distribusi nilai, kesamaan nilai, kelengkapan data, penghapusan data duplikat, dan pengelompokan.	
3.	<p>Rancang Bangun Data Warehouse Berbasis Star Scheme Studi Kasus : SMK Negeri 4 Garut [9] Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) p27-29 February 2020</p>	<p><i>Data warehouse</i> SMK Negeri 4 Garut telah berhasil dibuat menggunakan model star schema. Model star schema dapat diterapkan dalam merancang <i>Data Warehouse</i> pada data akademik sebuah sekolah dan dapat mendukung penggalian informasi yang dibutuhkan bagi <i>top-level management</i> secara cepat. Penelitian ini hanya membahas analisis data-data yang berkaitan dengan data SMK 4 Garut, laporan akademik dan pendataan nilai semester Siswa/I SMK 4 Garut. Penelitian selanjutnya akan mengintegrasikan data Aset dengan <i>data warehouse</i> yang telah dibuat</p>	Implementasi <i>star schema</i> dalam membangun <i>data mart</i> Identitas nasabah.
4.	<p>Implementing data-driven decision support system based on independent educational data mart [10] <i>International Journal of Electrical and Computer Engineering (IJECE)</i> Vol. 11, No. 6, December 2021, pp. 5301~5314</p>	<p>Empat pendekatan yang dapat digunakan dalam merancang dan mengimplementasikan model <i>data mart</i> adalah <i>top-down</i>, <i>bottom-up</i>, <i>inside out</i>, dan campuran. Dalam penelitian ini, pendekatan terbaik adalah <i>bottom-up</i> karena hasilnya diperlukan dengan cepat untuk mendukung keputusan dalam membangun gudang data perusahaan. <i>Data mart</i> pendidikan memberikan banyak manfaat, termasuk menyediakan platform cepat untuk mendukung keputusan pendidikan, serta memberikan solusi kepada para pengambil keputusan akademik, guru, dan manajer sekolah. Skema yang diusulkan untuk <i>data mart</i> pendidikan adalah skema bintang, dipilih karena respons cepat terhadap kueri OLAP, kemudahan penerapan hirarki pada dimensi, dan fleksibilitas dalam menyesuaikan perubahan dalam skema.</p>	Pembangunan <i>data mart</i> dengan metode <i>bottom-up</i> dan <i>star schema</i>
5.	<p>Data Warehouse Design for Multi Finance Companies [11] <i>Turkish Journal of Computer and Mathematics Education</i> Vol.12 No.8 (2021)726-741</p>	<p>Menghasilkan pemahaman tentang perbedaan struktur data antara <i>data warehouse</i> dan sistem operasional (OLTP), di mana <i>data warehouse</i> menggunakan model <i>multidimensional</i> (skema bintang)</p>	Membangun <i>data mart</i> untuk perusahaan <i>mutlifinance</i> dengan model

No	Sumber	Hasil penelitian	Adopsi
		<p>untuk menyimpan data secara efisien dan memfasilitasi analisis yang kuat, sementara sistem operasional menggunakan model ERD untuk menggambarkan hubungan antara entitas dalam basis data operasionalnya.</p> <p>Hal ini memberikan landasan yang kuat bagi organisasi atau perusahaan untuk mengimplementasikan atau meningkatkan penggunaan data warehouse dalam proses pengambilan keputusan strategis dan analisis bisnis mereka.</p>	<p><i>multidimensi</i> untuk mendukung kegiatan analisis bisnis perusahaan keuangan</p>
6.	<p><i>Migrating a research data warehouse to a public cloud: challenges and opportunities</i> [12] <i>Journal of the American Medical Informatics Association</i> 2022, Vol 29 No 4</p>	<p>HDC membuat keputusan awal untuk beralih ke RDW sepenuhnya berbasis <i>cloud</i>. Pada saat itu, ini adalah langkah besar pertama dalam manajemen data pasien di awan publik bagi institusi yang berpartisipasi di HDC. Ini juga merupakan gudang data kesehatan berskala perusahaan pertama di GCP. HDC tidak pernah menyesali keputusan ini.</p>	<p>Membangun <i>data mart</i> menggunakan <i>platform cloud computing</i></p>
7.	<p>Perbandingan Data Warehouse Cloud Computing Menggunakan Konvensional Berbasis Kriptografi [13] Seminar Nasional Teknologi Komputer & Sains (SAINTEKS) Februari 2020 Hal 69-73</p>	<p>Berdasarkan hasil perbandingan <i>data warehouse cloud computing</i> dan konvensional, penggunaan keduanya dapat disesuaikan dengan kebutuhan perusahaan. Dari sisi infrastruktur konvensional akan lebih menyita biaya, sisi keamanan konvensional memiliki kontrol sendiri, namun pada <i>cloud</i> dapat menggunakan <i>private cloud</i>. Untuk perusahaan besar jenis <i>data warehouse</i> yang disarankan adalah <i>hybrid cloud</i> yaitu menggabungkan <i>data warehouse</i> secara <i>cloud</i> dan konvensional.</p>	<p>Dikarenakan pada penelitian ini hanya spesifik membangun <i>data mart</i> identitas nasabah, maka pendekatan yang digunakan yaitu menggunakan <i>cloud computing</i>.</p>
8.	<p>DATA MART DESIGN IN BKPP BANDUNG USING FROM ENTERPRISE MODELS TO DIMENSIONAL MODELS METHOD [14] Jurnal Ilmu Pengetahuan dan Teknologi Komputer Vol 5 No 2 2020</p>	<p>Penelitian ini mengembangkan sebuah data mart kepegawaian dengan menggunakan metode From Enterprise Models to Dimensional Models sebagai metode perancangannya dan pendekatan Bottom-Up Approach dalam pengembangannya. Data mart tersebut dibangun menggunakan basis data</p>	<p>Membangun <i>data mart dimensional model</i> dan pendekatan <i>bottom up</i> menggunakan basis data <i>PostgreSQL</i></p>

No	Sumber	Hasil penelitian	Adopsi
		PostgreSQL dan bahasa pemrograman PHP untuk membuat antarmuka pengguna. Hasilnya adalah sebuah data mart kepegawaian yang menyimpan data historis transaksi mutasi pegawai dan memberikan informasi untuk mendukung proses pengambilan keputusan.	untuk mempermudah pengambilan keputusan.
9.	Analysis and Design of Data Mart Decision Support Systems at PT Marlindo Tirta Nusantara [15] Journal Publications & Informatics Engineering Research Volume 3, Number 2, April 2019	Berdasarkan hasil analisis dan desain data mart di PT Marlindo Tirta Nusantara, dapat diambil beberapa kesimpulan: 1. Dengan adanya data mart, data operasional dan data historis dari PT Marlindo Tirta Nusantara yang awalnya tidak optimal dapat dimanfaatkan dan diproses lebih lanjut sebagai bahan analisis oleh manajemen eksekutif. 2. Data mart mampu memberikan informasi yang dibutuhkan untuk komisi pengiriman. 3. Data mart menyimpan semua informasi tentang komisi transportasi PT Marlindo Tirta Nusantara dengan tujuan membantu atau memudahkan eksekutif dalam mencari atau menganalisis data transportasi yang telah ada selama beberapa tahun terakhir.	Membangun <i>data mart</i> untuk mengoptimalkan data sehingga dapat menjadi bahan analisis manajemen
10.	Financial Technology: From Temporary Disruption to New Financial Industry Model [16]. Journal of Business And Management Review Vol 3 No 6 2022 page 399 - 418	Proses transformasi digital dalam bisnis telah berdampak pada setiap bisnis di seluruh dunia. Dengan menggunakan studi tentang kinerja industri keuangan Indonesia, peneliti ingin mencari tahu variabel-variabel yang mampu menjaga fungsi bisnis digital yang sementara menjadi transformasi permanen dari fungsi bisnis. Hal ini membuktikan bahwa faktor kemampuan dinamis yang tinggi dapat meningkatkan kinerja perusahaan. Orientasi pelanggan juga ditemukan memiliki pengaruh positif namun tidak signifikan terhadap kinerja perusahaan.	Melakukan inovasi teknologi baru yaitu <i>data mart</i> untuk meningkatkan kinerja perusahaan.

Tabel penelitian 2.1 ini merupakan list penelitian yang menjadi referensi dalam membangun penelitian ini, berikut perbedaan dan persamaan yang diadopsi dari referensi penelitian kepada penelitian ini. Pertama, penelitian ini

akan melakukan inovasi pada teknologi digital dengan membangun *data mart* seperti penelitian [15] namun penelitian ini berfokus pada identitas nasabah. Hal ini dikarenakan menurut penelitian [16] perusahaan yang melakukan inovasi dapat meningkatkan kinerja perusahaan. Kedua, penelitian ini hanya spesifik berfokus pada identitas nasabah, berbeda dengan penelitian sebelumnya [7], [9], [11], memiliki cakupan yang lebih luas atau berfokus pada aspek lain dari manajemen data. Ketiga, penelitian ini mengusulkan pendekatan baru dalam membangun solusi dengan memanfaatkan teknologi *cloud* seperti pada penelitian [12] dan pengolahan data tidak terstruktur menggunakan OCR, sementara penelitian sebelumnya [7], [8], [9], [10], [14] lebih terpusat pada pendekatan dan teknologi yang berbeda, namun pada penelitian ini teknologi *cloud* seperti pada penelitian [13] yang digunakan adalah AWS, Sedangkan penelitian [12] menggunakan GCP. Keempat, penelitian ini membawa kontribusi baru dalam mengatasi tantangan yang spesifik yang dihadapi oleh industri *multifinance* dalam mengelola data identitas nasabah, seperti integrasi data tidak terstruktur dan pemahaman yang lebih dalam tentang preferensi nasabah seperti penelitian [11]. Dengan demikian, penelitian ini memberikan kontribusi berharga untuk literatur dalam bidang manajemen data nasabah dan dapat menjadi landasan untuk penelitian lebih lanjut dalam topik ini.

2.2 Teori tentang Topik Skripsi

2.2.1 Industri *Multifinance*

Perusahaan *multifinance*, sebagai lembaga keuangan non-bank, memainkan peran krusial dalam aktivitas pembiayaan di Indonesia [17]. Keterlibatan mereka telah terbukti signifikan dalam mendistribusikan sumber daya keuangan kepada pelaku usaha dan masyarakat. Perusahaan-perusahaan ini menyediakan pembiayaan untuk barang-barang produktif yang dibutuhkan oleh pelaku usaha, serta barang-barang konsumtif yang menjadi kebutuhan masyarakat umum [18]. Perusahaan *multifinance* terlibat dalam kegiatan pembiayaan atau memberikan jaminan kepada *lessee* (nasabah) terkait dengan aset berwujud yang digunakan oleh *lessee* tersebut. Meskipun demikian, perusahaan *multifinance* ini tetap memiliki

kepemilikan atas aset berwujud yang digunakan oleh *lessee*, sehingga *multifinance* dapat dianggap sebagai perantara antara *lessee* dan pihak penyedia barang atau *supplier*. Dalam hal pembiayaan kendaraan, terutama mobil, *multifinance* bekerja sama dengan *showroom* atau *dealer* untuk memfasilitasi proses transaksi jual beli yang dilakukan dengan skema kredit terhadap produk tersebut [17]. Dalam melakukan kegiatan pembiayaan, perusahaan membutuhkan informasi mengenai identitas nasabah. Informasi-informasi tersebut datang dengan berbagai macam bentuk, baik berupa data terstruktur, maupun data tidak terstruktur.

2.2.2 Data Tidak Terstruktur

Data tidak terstruktur dapat ditemukan dalam berbagai format *file*, seperti dokumen, video, suara, *email*, dan lampiran lainnya [19]. Sebagai pembandingan, data terstruktur memiliki representasi yang lebih terorganisir, memungkinkan pengolahan dan analisis yang lebih mudah. Tantangan utama dengan data tidak terstruktur adalah keragaman bentuk representasinya, sehingga pengintegrasian dengan sumber data lain dan analisisnya menjadi sangat sulit [20]. Data ini tidak dapat diorganisir, dicari, atau dianalisis dengan metode yang sama seperti data terstruktur, sehingga memerlukan pendekatan yang baru. Oleh karena itu, diperlukan metode dan proses inovatif untuk mengekstrak informasi yang memiliki nilai, berbagi pengetahuan, dan menyampaikan keunggulan yang terkandung dalam data tidak terstruktur tersebut [21]. Transformasi data tidak terstruktur menjadi data terstruktur menjadi hal yang penting untuk memanfaatkan potensi informasi yang tersembunyi dalam beragam format data.

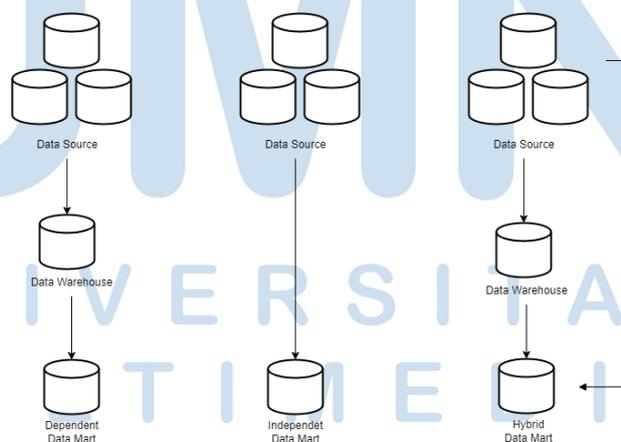
2.2.3 Data Mart

Data mart merupakan platform penyimpanan data yang menampung informasi khusus untuk unit bisnis dalam suatu organisasi. *Data mart* hanya memuat sebagian kecil dan terpilih dari seluruh data yang disimpan oleh perusahaan dalam berbagai sistem penyimpanan yang lebih besar. Perusahaan memanfaatkan *data mart* untuk melakukan analisis informasi yang terkait dengan departemen-departemen secara lebih efisien [22].

Fungsi utama *data mart* adalah untuk mengambil dan menganalisis informasi spesifik. Perbedaan antara *data mart* dengan basis data relational adalah basis data relational merupakan sistem penyimpanan terstruktur yang digunakan untuk menyimpan, mencari, mengambil, dan menganalisis informasi secara umum. Data dalam basis data relational dihubungkan oleh kunci unik yang membedakan setiap entitas. Sementara itu, *data mart* hanya berfokus pada informasi yang relevan dengan subjek tertentu, seperti departemen sumber daya manusia dan dukungan pelanggan.

Dalam perbandingan dengan *data warehouse*, *data warehouse* adalah platform penyimpanan data yang menyimpan informasi seluruh bisnis dalam organisasi. *Data warehouse* mencakup data dari berbagai departemen dan topik di seluruh perusahaan. Di sisi lain, *data mart* hanya memuat data yang berkaitan dengan satu departemen saja, namun *Data mart* dapat diintegrasikan dengan *data warehouse* yang lebih besar. Ini memungkinkan data yang relevan dengan departemen tertentu diakses dari satu tempat [23]. Keuntungan utama membangun *data mart* adalah bahwa departemen yang bertanggung jawab atas *data mart* memiliki kendali penuh atas proses pemuatan dan pengelolaan data.

Terdapat 3 jenis *data mart* yang sering digunakan, yaitu *Dependent*, *Independen*, dan *Hybrid*.



Gambar 2. 1 Data Mart

Sumber [9]

Dapat dilihat pada Gambar 2.1, dependen berarti *data mart* yang sumber datanya berasal dari *Data Warehouse*. *Independent* berarti *data mart* yang sumber datanya langsung berasal dari sumber data sistem yang kemudian nantinya *data mart independent* memungkinkan menyumbang data untuk *data warehouse*. *Hybrid* berarti gabungan dari dependen dan independen *data mart*. Sumber data bisa berasal dari sumber data langsung, ataupun dari *data warehouse*. Tantangan terbesar dari *data mart* adalah bagaimana mengintegrasikan data dari berbagai sumber atau berbagai tipe data. Oleh karena itu, solusi dari tantangan tersebut adalah dengan menggunakan ETL [24].

2.2.4 Extract, Transform, Loading (ETL)

ETL merupakan proses yang digunakan dalam analisis data dan manajemen informasi untuk memindahkan data dari satu lokasi atau sumber ke tempat lain, biasanya dari sumber operasional ke *data warehouse* atau data lake untuk tujuan analisis. ETL digunakan untuk mengekstraksi data dari sistem, membersihkan data untuk meningkatkan kualitas data dan menyesuaikannya dengan format yang diinginkan, dan memuat data ke dalam database target. ETL memberikan kontribusi penting bagi *data warehouse*, mulai dari kontrol kualitas, integrasi data, dan pengelolaan berbagai sumber data. Gambar 2.2, menunjukkan 3 langkah yang terlibat di dalam proses ETL [25].



Gambar 2. 2 ETL

Sumber [25]

Extraction (Ekstraksi), Proses pertama dari ETL melibatkan ekstraksi data dari sumber yang berbeda, seperti basis data, file teks, atau aplikasi bisnis. Data ini dapat berupa data transaksi, data nasabah, atau data lainnya yang relevan untuk analisis. *Transform* (Transformasi), Setelah data diekstraksi, proses transformasi melibatkan pembersihan, penggabungan,

dan perubahan format data sesuai dengan kebutuhan bisnis atau analisis. *Loading* (Pemuatan), Setelah data diubah, langkah terakhir adalah memuatnya ke dalam penyimpanan data yang dituju, seperti *data warehouse*, *data lake*, atau basis data lainnya. Data ini kemudian siap untuk digunakan dalam analisis, pelaporan, atau aplikasi bisnis lainnya. ETL merupakan proses kunci dalam pengelolaan data yang efektif, karena memungkinkan organisasi untuk mengintegrasikan data dari berbagai sumber, membersihkan dan mempersiapkan data untuk analisis, dan menyimpannya dalam bentuk yang dapat diakses dan dipelajari didalam sebuah database dengan menggunakan relasi yang disebut *Entity Relationship Diagram (ERD)*.

2.2.5 *Entity Relationship Diagram (ERD)*

ERD adalah singkatan dari *Entity-Relationship Diagram*. Ini adalah alat visual yang digunakan untuk merancang basis data untuk menggambarkan hubungan antara entitas atau objek yang terlihat beserta atributnya dalam sebuah sistem [26]. ERD terdiri dari 3 elemen dasar yaitu entitas, atribut, dan relasi antara entitas tersebut.

1. Entitas

Entitas adalah objek yang menjadi perhatian dalam suatu basis data [27]. Contoh objek dalam sistem dapat berupa orang, tempat, benda, dan sebagainya. Setiap entitas biasanya memiliki atribut yang mendefinisikan karakteristik atau properti unik dari entitas tersebut.

2. Atribut

Atribut adalah karakteristik atau informasi yang menggambarkan entitas tersebut. Setiap entitas harus memiliki *primary key* sebagai ciri dari entitas dan atribut deskriptif [27]. Contoh jika entitasnya adalah orang maka atributnya bisa berupa nama, alamat, dan sebagainya.

3. Relasi

Relasi menggambarkan cara entitas terhubung satu sama lain dalam basis data baik satu entitas atau lebih, seperti hubungan *one-to-one* (satu entitas hanya berhubungan dengan satu entitas), *one-to-many* (satu entitas bisa berhubungan dengan banyak entitas lain), atau *many-to-many* (banyak entitas bisa berhubungan dengan banyak entitas). Contohnya, dalam konteks penelitian ini, relasi antara entitas *data_id* dan *invoice* adalah satu nasabah dapat melakukan banyak pesanan (*one-to-many*). Relasi ini membantu dalam memahami bagaimana data terkait antar entitas dan bagaimana entitas tersebut berinteraksi dalam konteks sistem yang direpresentasikan. Tabel 2.2 merupakan tabel komponen ERD dan Tabel 2.3 merupakan relasi ERD.

Tabel 2.2 tabel komponen ERD

Sumber [28]

Notasi	Komponen
	Entitas
	Atribut
	Relasi

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A

Tabel 2 3 tabel relasi ERD

Sumber [28]

Relasi	Keterangan
	One to One
	One to Many
	Many to Many

Dengan menggunakan ERD, perancang basis data dapat memvisualisasikan struktur data dengan jelas dan memahami bagaimana entitas berinteraksi satu sama lain dalam sistem yang direncanakan. *Entity-Relationship* juga digunakan sebagai salah satu metode pemodelan basis data yang digunakan untuk menghasilkan skema konseptual untuk jenis/model data semantik sistem seperti model konseptual *Data Warehouse* [26].

2.2.6 Model Konseptual *Data Warehouse*

Model konseptual *data warehouse* merupakan gambaran bagaimana informasi data akan diorganisir dan diakses dalam *data warehouse*. Ini seperti membuat peta besar yang menunjukkan semua tempat di mana data disimpan dan bagaimana mereka terhubung satu sama lain [28]. Model konseptual *data warehouse* menyoroti entitas utama yang relevan untuk bisnis, seperti nasabah, transaksi, dan sebagainya, serta hubungan dan aturan bisnis yang mengatur interaksi antara entitas tersebut. Tujuan utama dari model konseptual *data warehouse* adalah untuk memberikan pemahaman yang jelas tentang data yang disimpan dalam *data warehouse*, sehingga pemangku kepentingan bisnis dapat mengakses, menganalisis, dan mengambil keputusan yang tepat berdasarkan informasi yang tersedia.

Model konseptual *data warehouse* sering kali diwakili dengan menggunakan diagram atau notasi yang intuitif, seperti *Entity-Relationship Diagram* (ERD), untuk memvisualisasikan entitas, atribut, dan hubungan di

antara mereka. Ini membantu dalam merancang dan mengembangkan arsitektur *data warehouse* yang memenuhi kebutuhan analisis bisnis dan kebutuhan pengambilan keputusan perusahaan. Terdapat istilah yang berkaitan dengan model ini, yaitu tabel fakta dan tabel dimensi.

- Tabel Fakta

Tabel fakta merupakan tabel utama dalam model konseptual *data warehouse*. Ini adalah tempat di mana kita menyimpan data fakta atau informasi yang ingin kita analisis atau jelajahi. Tabel fakta biasanya berisi kumpulan data dari *primary key* table lainnya dan juga nilai kuantitatif yang terukur, seperti jumlah angsuran. Tabel fakta juga berhubungan dengan tabel-tabel dimensi yang membantu menjelaskan atau menguraikan data fakta tersebut.

- Tabel Dimensi

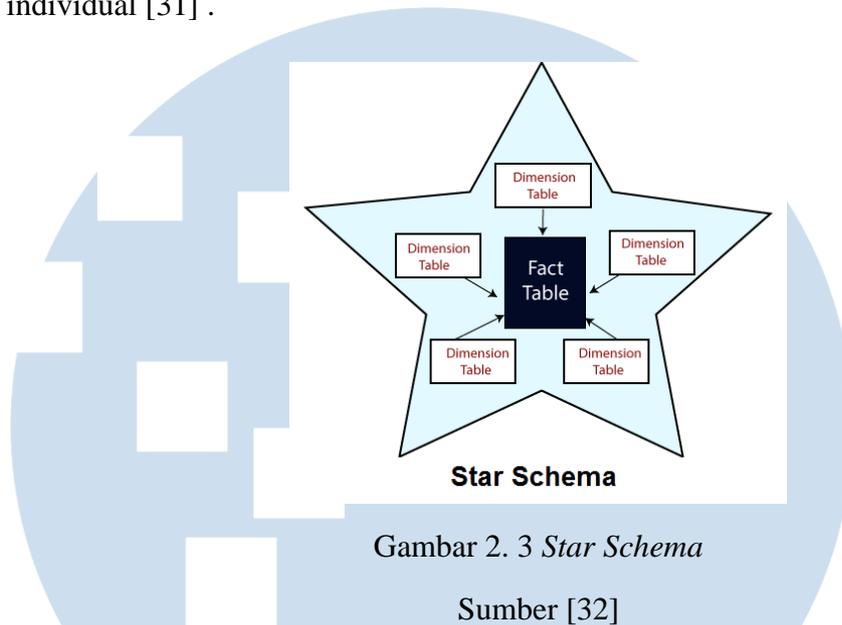
Tabel dimensi adalah tabel yang digunakan untuk menggambarkan atau menjelaskan data dalam tabel fakta. Mereka menyediakan konteks atau detail tambahan untuk data fakta yang disimpan dalam tabel fakta. Misalnya, jika tabel fakta berisi data NIK KTP, tabel dimensi berisi informasi tentang KTP tersebut (misalnya, nama alat, dan sebagainya). Tabel dimensi biasanya memiliki kunci primer yang terhubung dengan kunci asing dalam tabel fakta.

Salah satu model konseptual *data warehouse* yang paling populer adalah *star schema* [29].

2.2.7 *Star Schema*

Star Schema atau skema bintang merupakan desain *database* yang dapat digunakan untuk memodelkan *data warehouse* yang memungkinkan pengambilan data analitis. Dibandingkan dengan skema lain seperti skema *snowflakes*, skema ini lebih populer, hal ini dikarenakan strukturnya yang lebih sederhana dan memfasilitasi operasi OLAP untuk menganalisa data dengan efisien [30]. Skema bintang dapat berhubungan dengan dua jenis data: data dimensi, yang menjelaskan bagaimana data umumnya

diagregasi, dan data fakta atau peristiwa, yang menjelaskan transaksi individual [31] .



Gambar 2.3, menunjukkan struktur dari skema bintang yang terdiri dari satu tabel fakta yang berada di tengah-tengah, dikelilingi oleh tabel-tabel dimensi yang terhubung padanya melalui kunci. Hal ini menciptakan bentuk visual yang menyerupai bintang, dengan tabel fakta di pusatnya dan tabel dimensi yang merupakan "lengan" bintangnya.

2.2.8 Optical character recognition (OCR)

Optical character recognition (OCR) adalah bagian dari ilmu komputer yang berkaitan dengan proses membaca teks dari dokumen fisik dan mengubahnya menjadi format yang bisa dipahami dan dimanipulasi oleh komputer. Sistem OCR memungkinkan Anda untuk mengambil buku atau artikel majalah, memasukkannya langsung ke dalam *file* komputer elektronik, dan kemudian mengedit *file* tersebut menggunakan pemroses kata. Semua sistem OCR mencakup pemindai optik untuk membaca teks dan perangkat lunak canggih untuk menganalisis gambar [33]. Proses Optical Character Recognition (OCR) melibatkan beberapa langkah yang penting untuk mengubah gambar teks menjadi teks yang dapat dikenali secara digital. Langkah-langkah ini dimulai dengan akuisisi data, di mana gambar teks dipindai atau diambil. Selanjutnya, pra-pemrosesan dilakukan

untuk memperbaiki kekurangan gambar, seperti kemiringan, dan menghasilkan gambar biner yang hanya berisi teks. Setelah itu, gambar dibagi menjadi karakter-karakter individu melalui proses segmentasi. Fitur-fitur dari karakter tersebut diekstraksi, dan akhirnya, klasifikasi dilakukan untuk mencocokkan karakter dengan karakter prototipe yang telah ditentukan. Dengan melakukan langkah-langkah ini, proses OCR dapat menghasilkan teks yang akurat dari gambar teks yang diambil [34]. Salah satu teknologi OCR yang tersedia adalah Textract.

2.3 Teori tentang Framework / Algoritma yang digunakan

2.3.1 Textract

Amazon Textract adalah layanan yang disediakan oleh Amazon Web Services (AWS) yang bertujuan untuk mengekstrak teks dan data struktural dari berbagai jenis dokumen dengan menggunakan teknologi kecerdasan buatan (AI). Dengan Textract, pengguna dapat dengan mudah mengekstrak informasi dari gambar dan dokumen tercetak dalam berbagai format, seperti PDF, JPEG, PNG, dan TIFF. Layanan ini memiliki kemampuan untuk mengenali struktur dokumen, termasuk tabel, formulir, dan teks tulisan tangan, serta mengekstrak data yang terkandung di dalamnya. Dengan demikian, Textract memberikan kemudahan bagi pengguna dalam mengotomatisasi proses pengekstrakan informasi dari dokumen, menghemat waktu dan tenaga yang diperlukan untuk melakukan tugas tersebut secara manual. Dengan fitur-fitur yang canggih dan akurat, Amazon Textract menjadi solusi yang efektif untuk berbagai kasus penggunaan, seperti pengolahan dokumen, analisis data, dan integrasi dengan sistem bisnis yang ada [35]. Textract ini merupakan teknologi yang berada di layanan *cloud computing*.

2.3.2 Cloud Computing

Cloud computing merupakan model komputasi yang menggunakan layanan jaringan dengan akses *remote*. Ini berarti pengguna dapat mengakses dan menggunakan sumber daya komputasi, seperti penyimpanan data, aplikasi, dan layanan lainnya, melalui internet atau jaringan lainnya

tanpa harus memiliki infrastruktur komputasi fisik sendiri. Teknologi ini mirip dengan layanan publik karena dapat disesuaikan dengan kebutuhan pengguna dan mudah dikontrol. Salah satu keunggulan utama *cloud computing* adalah skalabilitasnya yang hampir tanpa batas, artinya pengguna dapat menambah atau mengurangi sumber daya komputasi sesuai kebutuhan mereka tanpa harus memperhatikan batasan infrastruktur fisik [36].

Secara lebih rinci, layanan *cloud computing* mencakup berbagai macam layanan jaringan yang memberikan skalabilitas, jaminan kualitas layanan (*Quality of Service/QoS*), personalisasi, dan merupakan platform komputasi yang ekonomis dan dapat diakses secara luas. Ini memungkinkan pengguna untuk memanfaatkan sumber daya komputasi tanpa harus berinvestasi dalam infrastruktur fisik yang mahal. Salah satu *cloud computing* yang terkenal adalah Amazon Web Services.

2.4 Teori tentang Tools / Software yang digunakan

2.4.1 Amazon Web Services (AWS)

Amazon Web Services (AWS) adalah satu penyedia layanan *cloud computing* terdepan yang terpercaya dan aman. Dengan berbagai layanan termasuk komputasi, penyimpanan, basis data, *content delivery network* (CDN), dan banyak lagi, AWS membantu berbagai bisnis untuk tumbuh dan mengoperasikan aplikasi secara efisien. Dengan berbagai produknya, AWS memudahkan pembangunan dan pengelolaan bisnis. Istilah yang tepat untuk menggambarkan AWS adalah *Infrastructure as a Service* (IaaS), di mana pengguna dapat menyewa infrastruktur komputasi, seperti server *virtual* dan penyimpanan, sebagai layanan. AWS juga menawarkan layanan lainnya seperti *Software as a Service* (SaaS), di mana pengguna dapat menggunakan aplikasi yang di-*hosting* secara online, dan *Platform as a Service* (PaaS), di mana pengguna dapat mengembangkan, menjalankan, dan mengelola aplikasi mereka tanpa harus memikirkan infrastruktur di belakang layar. Melalui model ini, AWS memungkinkan bisnis untuk mengakses dan menggunakan sumber daya IT secara fleksibel dan efisien sesuai kebutuhan

mereka [37], [38]. Gambar 2.4 merupakan logo AWS. Amazon Web Services (AWS) menyediakan berbagai layanan *cloud* yang mencakup beragam kebutuhan IT. Beberapa layanan AWS yang paling populer termasuk Amazon S3, Amazon RDS, Amazon EC2, AWS Lambda, dan Textract.

2.4.2 Amazon S3

Amazon Simple Storage Service (Amazon S3) merupakan layanan penyimpanan objek yang disediakan oleh Amazon Web Services (AWS). Dengan menawarkan skalabilitas, ketersediaan data yang tinggi, keamanan yang andal, dan kinerja terdepan dalam industri, Amazon S3 memungkinkan pelanggan dari berbagai ukuran dan sektor untuk menyimpan dan melindungi volume data besar dengan berbagai macam kasus penggunaan. Ini termasuk penyimpanan data lake, hosting situs web, pengembangan aplikasi seluler, pencadangan dan pemulihan data, arsip, aplikasi perusahaan, integrasi dengan perangkat IoT, dan analisis data besar. Amazon S3 dilengkapi dengan beragam fitur manajemen yang memungkinkan pengguna untuk mengoptimalkan, mengatur, dan mengonfigurasi data mereka sesuai dengan kebutuhan bisnis mereka. Dengan demikian, Amazon S3 menjadi solusi yang sangat fleksibel dan andal untuk kebutuhan penyimpanan data yang berkembang dan beragam. [39]. Gambar 2.5 merupakan *bucket* S3

2.4.3 AWS Lambda

AWS Lambda adalah layanan komputasi tanpa server (*serverless*) yang disediakan oleh Amazon Web Services (AWS). Layanan ini memungkinkan Anda menjalankan kode tanpa perlu mengelola infrastruktur server secara langsung. Lambda adalah layanan komputasi yang ideal untuk skenario aplikasi yang perlu ditingkatkan dengan cepat, dan skala turun ke nol saat tidak diminati [40]. AWS Lambda mendukung berbagai bahasa pemrograman, termasuk Python yang memberikan fleksibilitas bagi pengembang untuk menulis fungsi dalam bahasa sesuai keinginan. Lambda juga dapat diintegrasikan dengan layanan AWS lainnya,

seperti Amazon S3, Amazon DynamoDB, Amazon API Gateway, dan lain-lain, sehingga Anda dapat membuat arsitektur *serverless* yang lengkap. Gambar 2.6 merupakan logo dari AWS Lambda.

2.4.4 Amazon RDS

Amazon Relational Database Service (RDS) adalah layanan manajemen *database* yang disediakan oleh Amazon Web Services (AWS). Layanan ini memudahkan pengguna untuk membuat, mengelola, dan mengelola *database* relasional dalam lingkungan *cloud* AWS tanpa harus mengurus detail konfigurasi dan pemeliharaan infrastruktur *database*. RDS mendukung beberapa jenis *database* relasional populer, salah satunya adalah PostgreSQL. Ini memberikan fleksibilitas dalam memilih *database* yang sesuai dengan kebutuhan aplikasi. RDS menyediakan alat pemantauan dan logging terintegrasi, yang memungkinkan pengguna untuk memantau kinerja *database* dan mendeteksi potensi masalah [41]. Gambar 2.7 merupakan logo Amazon RDS.

2.4.5 PostgreSQL

PostgreSQL adalah sistem manajemen basis data relasional yang mendapatkan popularitas tinggi dan bersifat *open source* yang berarti bahwa *source code* dari PostgreSQL dapat digunakan secara bebas [42]. Ini mengadopsi model relasional (RDBMS), di mana data disimpan dalam bentuk tabel dan kolom yang terstruktur dan terhubung melalui kunci asing. Dukungan terhadap perintah SQL memudahkan *administrator database* dalam manipulasi data, seperti *insert*, *update*, dan *delete*. Perintah *select* pada PostgreSQL juga lebih fleksibel dengan berbagai klausa yang ada, memungkinkan *query* data dari tabel-tabel dengan lebih mudah. PostgreSQL juga mendukung berbagai bahasa pemrograman seperti Python yang membuatnya menjadi pilihan yang fleksibel dan dapat diandalkan untuk berbagai jenis aplikasi [43]. PostgreSQL juga memiliki kemampuan untuk membuat *VIEW* yang dapat mengoptimalkan *query* yang sering digunakan untuk menghindari eksekusi kueri yang sama dan mengurangi beban. Berikut fungsi dari *VIEW* :

- Dapat meningkatkan keamanan data karena hanya menampilkan kolom yang dibutuhkan.
- Dapat menyimpan *query* yang kompleks sehingga mengoptimalkan kinerja *query*.
- Kolom yang ditampilkan dapat menyesuaikan keinginan dan kebutuhan user.

Dengan menggunakan pendekatan ini, PostgreSQL dapat menyimpan data dengan cara yang terorganisir dan mudah diakses. Pada penelitian ini skema dan tabel yang dibuat pada RDS Postgre membutuhkan bantuan dari Amazon EC2.

2.4.6 Amazon EC2

Amazon Elastic Compute Cloud (Amazon EC2) adalah layanan komputasi yang tersedia di Amazon Web Services (AWS) Cloud. EC2 memungkinkan pengguna untuk menyewa server virtual sesuai dengan kebutuhan mereka, sehingga mengurangi biaya perangkat keras dan mempercepat pengembangan dan *deployment* aplikasi. Dengan EC2, pengguna dapat meluncurkan server virtual sebanyak atau sesedikit yang diperlukan, mengonfigurasi keamanan dan jaringan, serta mengelola penyimpanan. Layanan ini memungkinkan pengguna untuk menyesuaikan kapasitas komputasi mereka sesuai dengan beban kerja yang berubah-ubah, seperti proses bulanan atau tahunan, atau lonjakan lalu lintas situs *web*. Beberapa fitur tingkat tinggi dari Amazon EC2 meliputi:

- Instans: Server virtual yang dapat disesuaikan dengan berbagai konfigurasi CPU, memori, penyimpanan, dan kapasitas jaringan.
- Amazon Machine Images (AMI): Templat yang telah dikonfigurasi untuk instans, termasuk sistem operasi dan perangkat lunak tambahan.
- Pasangan kunci: Memastikan keamanan dengan kunci publik-privat yang digunakan untuk login ke instans.

- Volume penyimpanan: Untuk data sementara atau persisten menggunakan Amazon Elastic Block Store (Amazon EBS).
- Wilayah dan Zona: Lokasi fisik untuk sumber daya, seperti instans dan volume Amazon EBS.
- Grup keamanan: Firewall virtual untuk mengontrol lalu lintas masuk dan keluar instans.
- Alamat IP Elastis: Alamat IPv4 statis untuk komputasi cloud dinamis.
- Tag: Metadata yang dapat ditambahkan dan diatur ke sumber daya Amazon EC2.
- Virtual Private Cloud (VPC): Jaringan virtual yang terisolasi logis dari sisa AWS Cloud, yang dapat dihubungkan ke jaringan internal pengguna.

Dengan fitur-fitur ini, Amazon EC2 memberikan fleksibilitas, skalabilitas, dan keamanan yang diperlukan untuk memenuhi kebutuhan komputasi berbasis *cloud* pengguna [44].

