

BAB II

LANDASAN TEORI

2.1 Konsep Dasar Data, Informasi dan Pengetahuan

2.1.1 Data

“Data secara umum menggambarkan kode-kode struktur dari satuan *entity* utama, termasuk juga transaksi yang melibatkan dua atau lebih satuan *entity*” (Vercellis, 2009).

“Data adalah rekaman atau catatan dari fakta, konsep atau instruksi di dalam penyimpanan yang akan digunakan untuk komunikasi, pengambilan data, serta proses penampilan dari informasi yang dapat dimengerti oleh manusia” (Inmon , 2005).

“Data adalah komponen yang paling utama dalam lingkungan *Database Management Systems* dari segi pandang *end-user* dimana data bertindak sebagai suatu jembatan antara komponen mesin dan komponen manusia” (Connolly, et al. 2010).

Berdasarkan pengertian di atas, dapat ditarik kesimpulan bahwa data adalah sekumpulan fakta ataupun angka dan dapat diolah menjadi informasi yang berguna.

2.1.2 Informasi

Informasi merupakan hal yang sangat penting bagi perusahaan dalam mengambil setiap pengambilan keputusan. Secara Etimologi, Informasi berasal dari bahasa Perancis kuno yaitu *informaction* (tahun 1387) yang diambil dari bahasa latin *informationem* yang berarti “garis besar, konsep, ide”. Menurut Krismaji (2015), Informasi adalah “data yang telah diorganisasi dan telah memiliki kegunaan dan manfaat”.

Hal serupa disampaikan oleh Romney dan Steinbart (2015): Informasi (*information*) adalah data yang telah dikelola dan diproses untuk memberikan arti dan memperbaiki proses pengambilan keputusan. Sebagaimana perannya, pengguna membuat keputusan yang lebih baik.

“Informasi merupakan hasil dari kegiatan ekstraksi dan proses yang dibawa oleh data dan berarti untuk siapapun yang menerimanya” (Vercellis, 2009). Sedangkan menurut Inmon (2005) informasi adalah “Data yang dievaluasi oleh umat manusia untuk memecahkan masalah dan membuat keputusan”. Berdasarkan pengertian diatas dapat disimpulkan bahwa pengertian informasi adalah data yang diolah agar bermanfaat dalam pengambilan keputusan bagi penggunanya.

2.1.1 Pengetahuan

“Informasi dirubah menjadi pengetahuan bila digunakan untuk membuat keputusan dan mengembangkan tindakan yang sesuai” (Vercellis, 2009).

Dari pengertian diatas mengenai data, informasi dan pengetahuan, maka dapat disimpulkan bahwa data yang merupakan adalah sebuah fakta apabila dikelola lebih lanjut akan menjadi sebuah informasi yang berguna dalam dasar pengambilan keputusan, dan akan menjadi sebuah pengetahuan dalam menentukan tindakan yang harus diambil secara tepat dan efisien.

2.2 Konsep Dasar *Database*

2.2.1. Pengertian *Database*

Menurut Connolly et al. (2010) menyatakan bahwa “Database adalah suatu kumpulan dari data logical yang terhubung satu sama lain dan deskripsi dari data yang dirancang untuk menemukan informasi yang dibutuhkan sebuah organisasi”.

2.2.2. *Database Management Systems (DBMS)*

Database Management Systems atau DBMS menurut Connolly et al. (2010) merupakan “Suatu sistem perangkat lunak yang memungkinkan user dapat mengidentifikasi, membentuk, memelihara, dan mengatur akses ke database”. Mengacu pada pendapat Conolly et al. (2010) terdapat 5 (lima) komponen utama DBMs, yaitu :

1. *Hardware*. Dalam mengoperasikan DBMS diperlukan peran serta dari *hardware* yang dapat berupa *single personal computer*, *single mainframe*, atau jaringan komputer.

2. *Software.* Komponen *software* terdiri dari DBMS, aplikasi program, sistem operasi, dan perangkat lunak jaringan jika memang diperlukan untuk membuat suatu jaringan.
3. Data. Data yang merupakan data operasional dan *metadata* yang digunakan oleh perusahaan.
4. Prosedur. Prosedur merupakan istruksi atau aturan dan fungsi yang harus ada pada desain database dan DBMS.
5. *People.* People yang terlibat dalam menjalankan DBMS adalah Data *Administration* yang bertugas mengatur sumber daya data meliputi, perencanaan *database*, pengembangan dan pemeliharaan standar, kebijakan dan prosedur, dan *desain database logical* maupun konseptual. *Database Administrator* yang bertugas mengatur realisasi fisik dan aplikasi *database* yang meliputi desain fisik *database* dan implementasi, pengaturan, keamanan, dan kontrol integritas, pengawasan performa sistem dan pengaturan ulang *database*.
Database designer yang bertugas melakukan identifikasi data, hubungan antar data, dan batasan data yang disimpan dalam *database*. *Application developers* yang bertugas mengimplementasikan program aplikasi yang menyediakan kebutuhan bagi *end-user*. *End user* yang dapat digolongkan menjadi dua bagian, yaitu *native user* dan *sophisticated user*. *Native user* adalah *user* yang tidak perlu tahu mengenai DBMS, sedangkan *sophisticated user* adalah user yang mengetahui struktur basis data dan fasilitas DBMS. *Database management system* ini fokus dalam mengolah data dalam jumlah data besar yang terkumpul di dalam database dengan menggunakan bahasa *Database Definition Language* (DDL) dan *Data Manipulation Language* (DML).

2.3 Data Warehouse

2.3.1. Pengertian Data Warehouse

Menurut Han, J. & Kamber, M. (2012) *Data warehouse* adalah “media penyimpanan yang digunakan dalam jangka waktu yang cukup lama dan terdiri dari berbagai jenis format sumber data yang dikelola untuk fasilitas managemen dalam pengambilan keputusan”.

Menurut Inmon (2005) *data warehouse* adalah “Koleksi *database* yang terintegrasi yang dirancang untuk mendukung fungsi sistem pengambilan keputusan, dimana setiap unit data relevan pada beberapa waktu”.

Menurut Connolly et al. (2010) *data warehouse* adalah “suatu kumpulan data yang bersifat *subject oriented, integrated, time variant* dan *non-volatile* dalam mendukung proses pengambilan keputusan”.

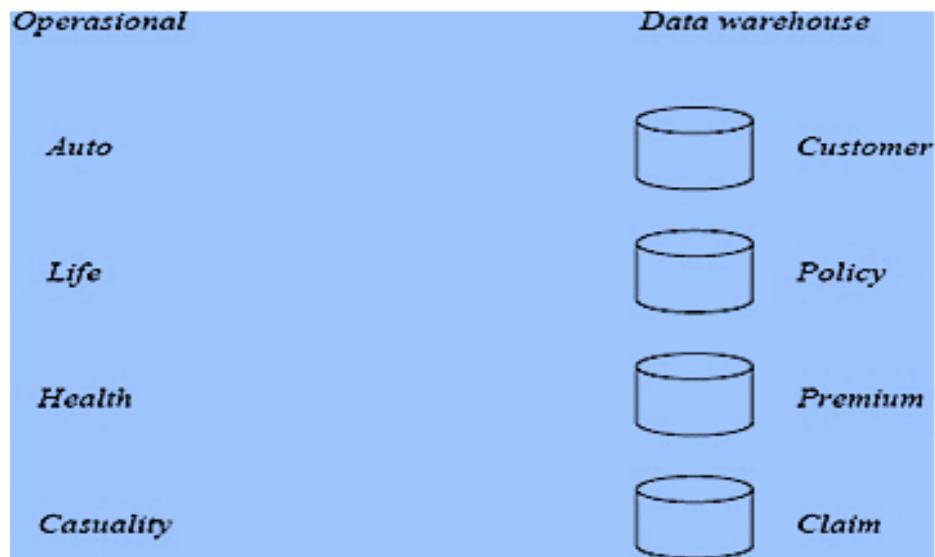
Dari ketiga pengertian diatas, maka dapat disimpulkan bahwa pengertian dari *data warehouse* merupakan media penyimpanan yang terdiri dari kumpulan database yang memiliki sifat *subject oriented, integrated, time variant* dan *non-volatile* yang membantu fasilitas managemen dalam sistem pengambilan keputusan.

2.3.2. Karakteristik Data Warehouse

Menurut Inmon (2005), *data warehouse* memiliki 4 (empat) karakteristik utama, yaitu :

1. *Subject Oriented*

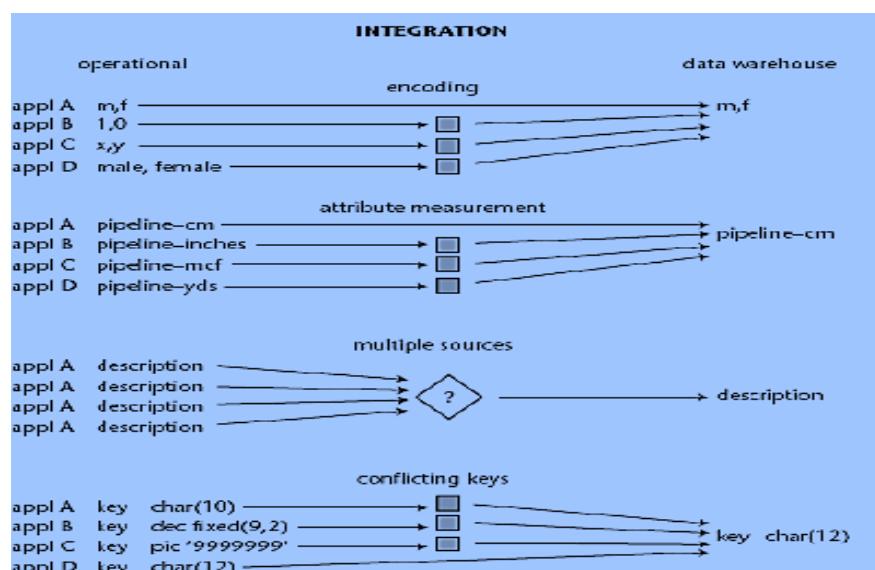
Data warehouse didesain untuk menganalisa data berdasarkan subjek-subjek tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu. *Data warehouse* diorganisasikan disekitar subjek-subjek utama dari organisasi (*customer, products, dan sales*) dan tidak diorganisasikan pada area-area aplikasi utama, hal ini dikarenakan kebutuhan dari *data warehouse* untuk menyimpan data-data yang bersifat sebagai penunjang suatu keputusan, dari pada aplikasi yang berorientasi terhadap data.



Gambar 2.1 Data Warehouse Subject Oriented (Inmon, 2005)

2. Integrated

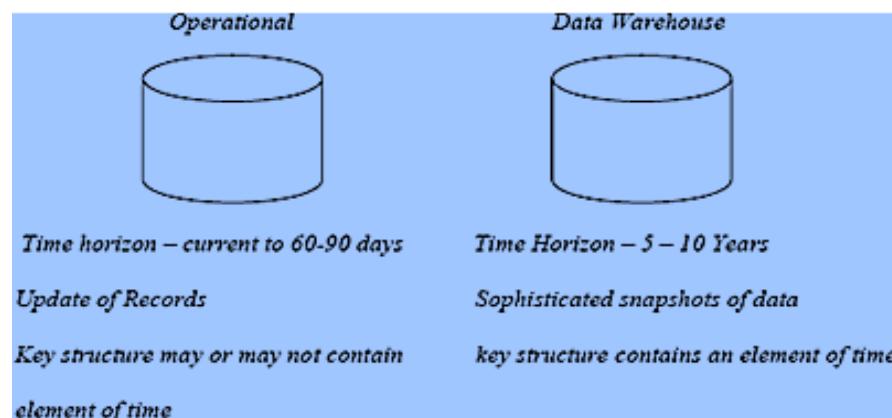
Data warehouse dapat menyimpan data-data yang berasal dari sumber-sumber yang terpisah kedalam suatu format yang konsisten dan saling terintegrasi satu dengan yang lainnya. Syarat integrasi sumber data dapat dipenuhi dengan berbagai cara seperti konsisten dalam penamaan variabel, konsisten dalam ukuran variabel, konsisten dalam struktur pengkodean dan konsisten dalam atribut fisik dari data.



Gambar 2.2 Data Warehouse Integrated (Inmon, 2005)

3. Time Variant

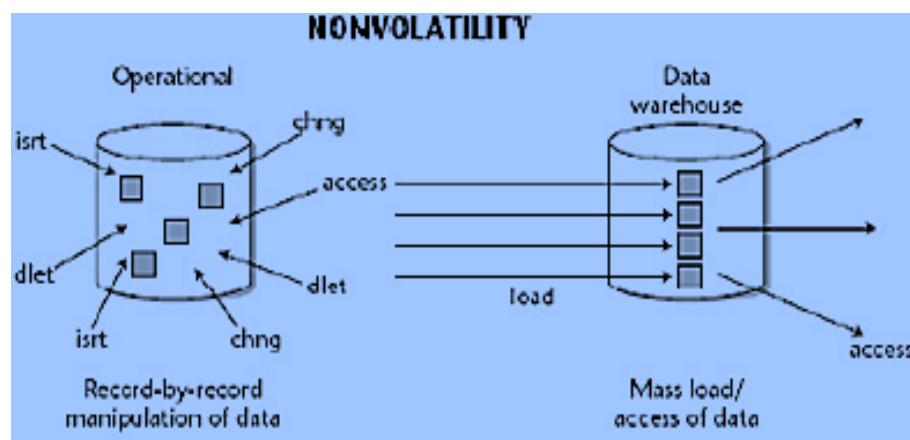
Model analisis yang diterapkan pada sebuah *data warehouse* berfokus pada perubahan data berdasarkan waktu. Data warehouse mampu menyimpan data untuk suatu objek tertentu dalam kurun waktu yang berbeda-beda.



Gambar 2.3 Data Warehouse Time Variant (Inmon, 2005)

4. Non-volatile

Data warehouse tidak diupdate secara *real time* tetapi di *refresh* dari sistem operasi secara regular. Data yang baru selalu ditambahkan sebagai suplemen bagi *database* itu sendiri bukan sebagai perubahan. *Data warehouse* hanya memiliki dua kegiatan memanipulasi data yaitu *loading* data (pengambilan data) dan akses data (mengakses *data warehouse* seperti melakukan *query* atau menampilkan laporan yang dibutuhkan, tidak ada kegiatan *updating* data)



Gambar 2.4 Data Non-Volatile (Inmon, 2005)

Berdasarkan penjelasan diatas dapat disimpulkan bahwa *data warehouse* memiliki empat sifat yaitu *subject oriented* dimana data disusun berdasarkan subjek utama bukan berdasarkan proses dan berorientasi untuk pendukung keputusan, bersifat *integrated* dimana data dari berbagai macam sumber disatukan kedalam satu media penyimpanan dan dapat disimpan dalam waktu yang ditentukan (*time variant*) *data warehouse* juga bersifat *non-volatile* dimana data tidak bisa dilakukan perubahan.

2.4 Pengertian *Business Inteligence*

Menurut Husni dan I. Mukhlash (2014). *Business intelligence* adalah suatu teori, metodologi, proses, arsitektur, dan teknologi yang mengubah data mentah menjadi suatu informasi yang bermakna dan berguna untuk tujuan bisnis. *business intelligence* dapat menangani sejumlah besar informasi untuk membantu mengidentifikasi dan mengembangkan peluang baru. Memanfaatkan peluang baru dan menerapkan strategi yang efektif dapat memberikan keuntungan pasar kompetitif dan stabilitas jangka panjang .

Kemudian dikemukakan oleh Grossmann (2015) yang menjelaskan bahwa *business intelligence* adalah Yang berarti bahwa *business intelligence* merupakan sebuah sistem otomatis yang dikembangkan untuk menyebarkan informasi secara efisien ke banyak sektor seperti di industry dan sebuah organisasi.

Beberapa keuntungan yang bisa didapatkan bila suatu organisasi mengimplementasikan *Business intelligence* adalah (Turban, Rainer, & Potter, 2011) :

- Meningkatkan nilai data dan informasi organisasi Dalam penerapan Business Intelligence seluruh data dan informasi dapat diintegrasikan sehingga mampu menghasilkan sebuah kesimpulan dari keadaan bisnis saat ini. Hal ini bertujuan untuk membantu pihak manajerial dalam membuat serta pengambilan keputusan yang lebih baik.

- Memudahkan pengukuran kinerja organisasi Pengukuran tingkat kinerja suatu organisasi dapat dilakukan menggunakan *Key Performance Indicator*. *business intelligence* mampu menunjukkan pencapaian *Key Performance Indicator* suatu organisasi dengan mudah dan cepat. Hal ini akan membantu pihak yang terlibat dalam pengambilan keputusan guna mempersiapkan langkah antisipasi apabila terdapat indikator yang menunjukkan adanya masalah atau belum tercapainya suatu target.
- Meningkatkan nilai investasi teknologi informasi yang sudah ada *business intelligence* dalam penerapannya tidak selalu mengubah atau menggantikan sistem informasi yang sudah ada, namun *business intelligence* dapat hanya menambahkan layanan pada sistem tersebut sehingga informasi dapat direpresentasikan dengan lebih baik.
- Meningkatkan efisiensi biaya *business intelligence* mampu meningkatkan efisiensi biaya dikarenakan tidak terlalu banyak usaha yang dilakukan dalam mendapatkan informasi yang dibutuhkan, hal ini secara otomatis dapat meningkatkan efisiensi biaya

2.5 Arsitektur *Business Inteligence*

Arsitektur dari sebuah sistem *business intelligence* terdiri atas enam komponen utama (Vercellis, 2009) yaitu :

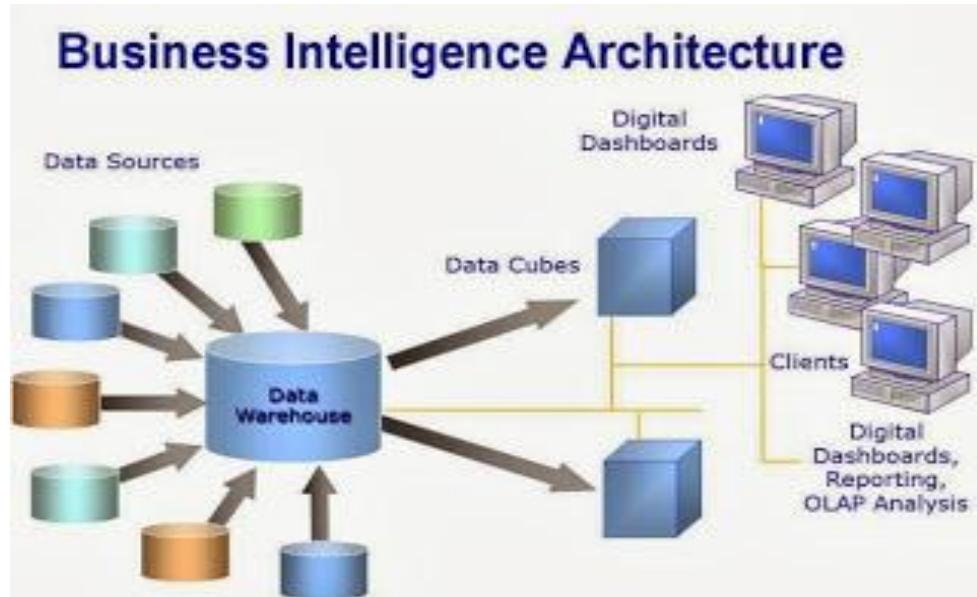
1. *Data Source*

Pada tahap pertama ini diperlukan proses untuk mengumpulkan dan mengintegrasikan data yang disimpan dalam berbagai sumber yang bervariasi yang saling berbeda baik itu asal maupun jenisnya. Sumber data ini berasal dari data yang terdapat pada *operational system*, tetapi juga bisa berasal dari dokumen yang tidak terstruktur seperti e-mail.

2. *Data Warehouse*

Pada tahap ini proses menggunakan extraction dan transformation *tools* yang dikenal sebagai ETL (*Extract, Transform, Load*), data yang berasal dari berbagai sumber yang berbeda disimpan ke dalam basis data yang ditujukan untuk mendukung proses analisis *business intelligence*.

Arsitektur dari sistem *business intelligence* dapat dilihat pada gambar 2.5



Gambar 2.5 Arsitektur *Business Inteligence* (sis.binus.ac.id, 2017)

3. *Data Exploration*

Pada tahap ini, *tools* yang berfungsi untuk keperluan analisis *business intelligence* pasif digunakan. *Tools* ini terdiri dari *query* dan *reporting system*, serta *statistical methods*. Metodologi ini bersifat pasif dikarenakan para pengambil keputusan harus mengambil keputusan berdasarkan hiposkripsi mereka sendiri atau mendefinisikan kriteria dari data *extraction*, kemudian menggunakan *tools* analisis untuk menemukan jawaban dan mencocokannya dengan hipotesa awal mereka.

4. *Data Mining*

Pada tahap ini proses terdiri sejumlah metodologi *business intelligence* bersifat aktif yang tujuannya untuk mengekstrak informasi dan pengetahuan dari data tersebut. Metodologi ini berisi sejumlah model matematika untuk pengenalan pola (*pattern*), pembelajaran mesin (*machine learn*) dan teknik *data mining*.

5. *Optimization*

Pada tahap ini menghasilkan solusi dimana solusi terbaik harus dipilih dari sekian solusi alternatif yang ada, dan biasanya sangat banyak dan beragam atau bervariasi.

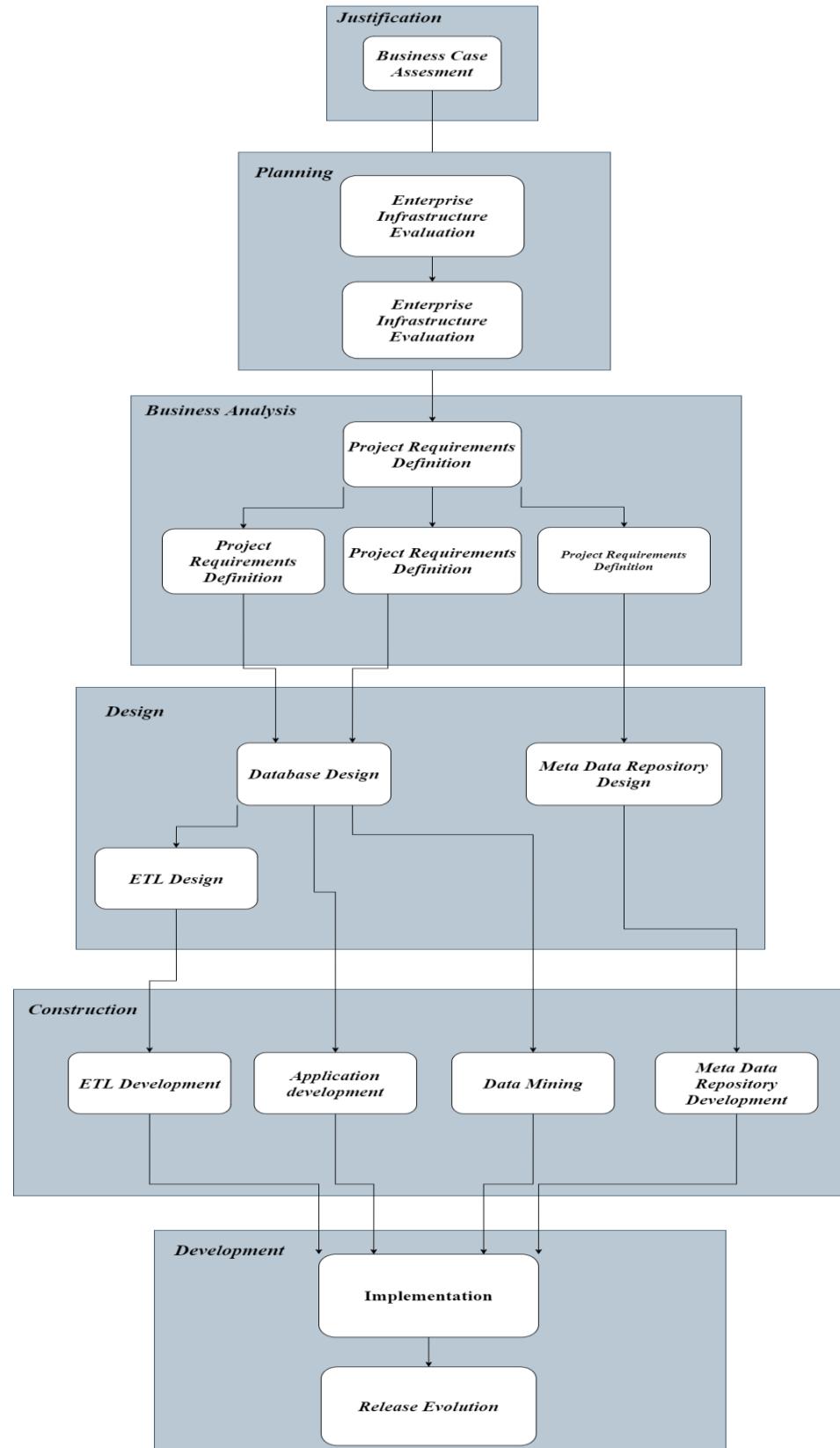
6. *Decisions*

Pada tahap ini yang menjadi persoalan utama merupakan bagaimana menentukan keputusan akhir yang akan diambil yang dikenal sebagai *decision making process*. Walaupun metodologi *business intelligence* berhasil diterapkan, pilihan untuk mengambil sebuah keputusan tetap ada ditangan para pengambil keputusan

2.6 *Business Inteligence Roadmap*

Dalam melakukan rancangan dan pengimplementasian *business intelligence* dapat menggunakan metode *business intelligence roadmap* yang merupakan salah satu metode pengembangan *business intelligence* yang dapat pakai karena sifatnya yang adaptif.

Dalam penelitian ini, metode yang dibahas adalah menggunakan pendekatan seperti yang dapat dilihat pada Gambar 2.6.



Gambar 2.6 *Business Intelligence Project Life Cycle* (Moss dan Atre, 2003)

Moss dan Atre, (2003), *business intelligence roadmap* ada beberapa langkah yang harus dilakukan berikut langkah-langkah yang dilalui dalam penerapan *business intelligence roadmap* ini:

2.6.1. *Justification*

Dalam tahap *justification*, dilakukan *business case assessment* yang merupakan langkah awal yang menjadi pertimbangan dalam mengembangkan *business intelligence*. Kegiatan yang dilakukan pada tahap ini adalah menentukan kebutuhan bisnis, mengevaluasi sistem pengambilan keputusan yang sedang berjalan, mengevaluasi sumber data operasional dan prosedur yang sedang berjalan, , menentukan objektif dari pembangunan *business intelligence*, mengajukan sebuah solusi *business intelligence*, menampilkan *cost-benefit analysis*, menampilkan analisis risiko, dan menulis laporan evaluasi.

2.6.2. *Planning*

Tahap ini merupakan tahap perencanaan strategis dan taktik yang akan digunakan untuk menghasilkan *business intelligence* yang sesuai dengan kebutuhan bisnis perusahaan untuk mencapai tujuan strategis bisnis. Proses yang ada dalam tahap ini adalah sebagai berikut :

1. *Enterprise Infrastructure Evaluation*

Sebelum membangun suatu *business intelligence* perlu dilakukan kesiapan perusahaan sebelum diterapkannya *business intelligence*. *Enterprise infrastructure* terdiri dari dua komponen utama, yaitu teknikal infrastruktur seperti *hardware*, *middleware* dan *database management systems* (DBMSs) dan non teknikal infrastruktur seperti standar, *metadata* aturan bisnis dan kebijakan perusahaan.

1) *Technical Infrastructure Evaluation*

Hardware yang dibutuhkan dalam penerapan *business intelligence* adalah *hardware* yang memiliki cukup *power* untuk mengatasi kebutuhan akan akses dan analisis data dalam jumlah besar.

2) *Nontechnical Infrastructure Evaluation*

Kegiatan yang dilakukan pada tahap ini adalah menilai komponen *non-teknikal infrastruktur* seperti standar *metadata*, standar penamaan data, model data-data *logical* perusahaan, metodologi, panduan prosedur pengujian, proses pengawasan perubahan, prosedur dalam masalah-masalah manajemen, dan penyelesaian masalah.

2. Project Planning

Proyek pengambilan keputusan pada pembangunan *business intelligence* sangat dinamis. Perubahan ruang lingkup, *staff*, anggaran, teknologi, gambaran bisnis, dan sponsor-sponsor bisa sangat berdampak pada keberhasilan sebuah proyek. Karena itu diperlukan perencanaan proyek yang matang dan terperinci sesuai dengan kemajuan yang ingin dicapai. Aktivitas yang terlibat didalam project planning yaitu :

1) *Project Goal and Objective*

Dalam membuat proyek *business intelligence* faktor penting yang harus ditentuan adalah apakah tujuan dan sasaran dibuatnya aplikasi *business intelligence* pada organisasi. Masalah bisnis apa yang dimiliki oleh organisasi dan aplikasi *business intelligence* apa yang cocok dalam menyelesaikan masalah yang ada. Pada akhirnya apakah aplikasi *business intelligence* ini berhasil atau tidak bergantung pada tingkat efektivitas yang diperoleh dari diterapkannya aplikasi *business intelligence*.

2) *Project Scope*

Pembangunan *business intelligence* pada suatu organisasi membutuhkan banyak hal yang harus diperhitungkan. Untuk itu, dibutuhkan batasan proyek agar dapat diketahui dengan pasti berapa banyak sumber daya, biaya , data dan hal lain yang dibutuhkan dalam melaksanakan proyek *business intelligence*.

3) *Project Risk*

Setiap proyek yang dilaksanakan memiliki resiko-resiko yang harus diperhitungkan dan diperkirakan agar tidak mengganggu jalannya jadwal dan hasil dari pelaksanaan proyek. Oleh karena itu harus dibuat perkiraan resiko dan upaya pengedalian resiko yang Mungkin terjadi. Hal yang menjadi fokus perhatian dalam membuat *project risk* adalah *triggers* dan *mitigation plan*. *Triggers* adalah pemicu adanya potensi resiko yang menyebabkan terganggunya jadwal dan hasil dari proyek yang dapat menyebabkan kerugian. Sedangkan *mitigation plan* adalah upaya yang dilakukan dalam mengatasi triggers sehingga potensi terganggunya jadwal dan hasil proyek yang menimbulkan kerugian dapat dicegah.

4) *Project Constraint*

Batasan projek harus difikirkan yaitu lamanya waktu yang dibutuhkan untuk melakukan penggeraan proyek dan pembiayaan yang dibutuhkan terkait dengan seluruh komponen yang digunakan dalam pelaksanaan proyek.

2.6.3. Business Analysis

Tahap ini menunjukkan rincian analisis dari masalah bisnis atau peluan bisnis untuk memperoleh sebuah pemahaman penuh dari kebutuhan bisnis untuk menjadi solusi yang potensial.

1. Project Requirement Definition

Pada bagian ini akan dibahas beberapa hal mengenai cara untuk menentukan persyaratan, perbedaan antara bisnis pada umumnya dengan syarat yang berhubungan dengan proyek. Pembatasan penggeraan proyek juga perlu ditentukan agar pekerjaan menjadi lebih terfokus dan tidak meluas. Aktivitas yang ada pada tahap ini adalah mendefinisikan kebutuhan tambahan infrastruktur teknikal. Mendefinisikan kebutuhan tambahan non-teknikal, mendefinisikan kebutuhan laporan, mendefinisikan kebutuhan sumber daya, meninjau ruang lingkup proyek, memperluas logical data model, mendefinisikan

persetujuan *preliminary service level*, menulis dokumen kebutuhan aplikasi.

1) *Technical infrastructure Requirement*

Pada tahap ini adalah menentukan infrastruktur teknik yang dibutuhkan dalam peningkatan perbaikan infrastruktur yang dibutuhkan dalam proyek *business intelligence*. Pada tahap ini harus sudah dapat dipastikan apakah komponen yang ada pada organisasi sudah memenuhi dalam mendukung aplikasi *business intelligence* ataukah ada perubahan yang dibutuhkan.

2) *Nontechnical Infrastructure Requirement*

Selain melakukan pengecekan kebutuhan teknikal, kebutuhan non teknikal juga harus dilakukan pengecekan. Pengecekan dilakukan dalam hal sebagai berikut :

a. *Roles and Responsibility*

Apakah terjadi perubahan atau tanggung jawab pengguna data apabila aplikasi *business intelligence* diterapkan dalam organisasi. Siapakah yang dapat bertanggung jawab dan mempunyai wewenang dalam melakukan akses data *business intelligence*.

b. *The Requirement of Source Data*

Dalam tahap ini ditentukan data apa sajakah yang dibutuhkan dalam pelaksanaan proyek *business intelligence*. Darimanakan sumber files dan database yang dibutuhkan dalam aplikasi *business intelligence*.

c. *Review The Project Scope*

Melakukan pengecekan kembali batasan-batasan yang digunakan dalam pembuatan aplikasi BI apakah terdapat perubahan yang harus dinegosiasi atau tidak terkait dengan pelaksanaan proyek.

2. *Data Analysis*

Dalam pembuatan proyek *business intelligence* dan proyek pengambilan keputusan sejenis adalah kualitas data yang dimiliki oleh perusahaan. dengan data yang berkualitas, maka akan dapat mendukung keputusan yang lebih tepat sasaran. Kegiatan yang dilakukan pada tahap ini adalah analisis sumber data *eksternal*, memperbaiki *logical data model*, menganalisis kualitas sumber data, memperluas *logical data model* perusahaan, menyelesaikan perbedaan data, dan menulis spesifikasi data.

3. *Application Prototyping*

Implementasi merupakan penyampaian fungsional yang memungkinkan *stakeholder* bisnis melihat potensi dan batasan bisnis yang memberikan kesempatan bagi mereka untuk menyesuaikan kebutuhan proyek dengan harapan mereka. Kegiatan yang ada pada tahap ini adalah menganalisis kebutuhan akses, menentukan cakupan Implementasi pemilihan alat untuk untuk Implementasi, mempersiapkan Implementasi *chapter*, desain laporan dan *query*, membangun Implementasi, dan menunjukkan hasil *prototyping*.

4. *Metadata Repository Analysis*

Terdapat dua kategori *metadata*, yaitu *business metadata* dan *technical metadata*. *Business metadata* menyediakan pebisnis dengan *roadmap* untuk mengakses data bisnis dalam lingkungan pendukung keputusan *business intelligence*. Sedangkan *technical metadata* mendukung teknisi dan *power user* dengan menyediakan para teknisi dengan informasi tentang aplikasi dan *database* mereka untuk menjaga aplikasi *business intelligence*. Kegiatan pada *metadata repository analysis* adalah menganalisis kebutuhan *repository metadata*, menganalisis kebutuhan antarmuka untuk *repository metadata*, *penganalisis repository metadata* dan kebutuhan laporan, membuat *logical data model* dan membuat *metadata*.

2.6.4. Design

Pada tahap ini adalah proses pemahaman lebih lanjut mengenai sumberdaya yang menyelesaikan masalah bisnis atau meningkatkan peluang bisnis. Proses yang ada pada tahap ini adalah sebagai berikut.

1. Database Design

Satu atau lebih *database business intelligence* akan menyimpan data bisnis dalam bentuk rincian atau kumpulan, tergantung pada kebutuhan laporan dari komunitas bisnis. Tidak semua laporan kebutuhan strategis, dan tidak semua *multidimensional*. Skema perencanaan *database* harus mencocokan kebutuhan-kebutuhan pengaksesan informasi dari komunitas bisnis. Kegiatan yang ada pada *database design* adalah meninjau kebutuhan data akses, menentukan agregasi dan ringkasan kebutuhan, merancang target *database business intelligence*, merancang *struktur database physical*, mengembangkan perawatan *prosedur database*, menyiapkan *monitor* dan *tune* rancangan *database*.

2. ETL Design

Merupakan proses esensial dari kegiatan perancangan *business intelligence*. Semakin buruk kualitas sumber data, maka semakin banyak waktu yang diperlukan untuk menjalankan transformasi dan program pembersihan. Sumber data untuk aplikasi *business intelligence* berasal dari berbagai sumber *platform*, dimana sumber tersebut diatur oleh beberapa aplikasi dan sistem operasi. Tujuan dari penggunaan *platform* ini adalah untuk menggabungkan data dari *platform-platform* yang berbeda kedalam sebuah format standar untuk *database* tujuan *business intelligence* didalam lingkungan pengambilang keputusan. Kegiatan yang ada di dalam tahap ini adalah membuat dokumentasi pemetaan sumber ke target, mencoba fungsi tool ETL, merancang aliran proses ETL, merancang program ETL, dan *set up ETL staging area*.

3. Metadata Repository Design

Jika tempat penyimpanan *metadata* berlisensi, maka kemungkinan besar harus ditingkatkan dengan fitur yang didokumentasikan pada *meta model* yang logis tetapi tidak disediakan oleh produk. Jika tempat penyimpanan

meta data sedang dibangun, maka keputusan yang harus dibuat adalah desain tempat penyimpanan *database*, apakah akan berbasis entitas-hubungan atau berorientasi objek. Kegiatan pada tahap ini adalah merancang *meta data repository database*, *install* dan tes produk *meta data repository*, merancang proses *migrasi meta data*, dan merancang aplikasi *meta data*.

2.6.5. Construction

Pada tahap ini dilakukan pembangunan business intelligence yang dengan tenggang waktu yang telah ditetapkan. Proses yang ada pada tahap ini adalah sebagai berikut :

1. ETL Development

Banyak *tools* yang tersedia untuk proses ETL. Sebuah tool ETL bisa atau tidak bisa menjadi solusi terbaik. Dalam kasus lainnya, pemrosesan awal data dan penulisan tambahan untuk mendukung kemampuan dari tool ETL sering dibutuhkan. Kegiatan yang dilakukan pada tahap ini adalah membangun dan uji coba unit proses ETL, *integrasi* dan *regresi* proses ETL, menampilkan uji coba proses ETL, jaminan kualitas dan uji coba proses ETL, dan persetujuan uji coba proses ETL.

2. Application Development

Setelah upaya *prototyping* telah memenuhi persyaratan fungsional, pengembangan aplikasi akses dan analisis yang sebenarnya dapat dimulai. Mengembangkan aplikasi dapat menjadi masalah sederhana dari penyelesaian *Implementasi operasional*, dan atau menjadi upaya pengembangan lebih dengan menggunakan hal yang berbeda. Kegiatan yang ada pada tahap ini adalah menentukan kebutuhan final proyek, merancang program aplikasi, membangun dan uji coba unit program aplikasi, uji coba program aplikasi, dan menyediakan akses data dan *analysis training*.

3. Data Mining

Tahap ini adalah menemukan informasi tersembunyi yang didapatkan dari data organisasi yang ada. Aktivitas yang ada pada tahap *data mining* ini adalah mengelompokan masalah bisnis, mengumpulkan data,

konsolidasi dan membersihkan data, menyiapkan data, membangun model data *analytical*, menginterpretasikan hasil *data mining*, menampilkan hasil *validasi eksternal*, dan mengawasi waktu analisis data model.

4. *Metadata Repository Development*

Kegiatan yang ada pada tahap ini adalah membangun *database meta data repository*, membangun dan uji coba unit proses migrasi *meta data*, membangun dan uji coba unit *meta data* proses migrasi, membangun dan uji coba unit aplikasi *meta data*, uji coba program *meta data repository* atau fungsi produk, menyiapkan mea data repository untuk produksi, dan menyediakan pelatihan *meta data repository*.

Kelebihan *business intelligence roadmap* menurut Moss & Atre (2003) adalah business intelligence *roadmap* merupakan panduan dalam melakukan pengembangan *business intelligence* dimana metode ini dirancang untuk beradaptasi dengan kebutuhan spesifik dari stakeholder dan organisasi yang menjelaskan kompleksitas dari proyek *business intelligence*.

Pembahasan di fokuskan pada masing- masing tanggung jawab dari representatif bisnis, sponsor bisnis, manajer proyek, dan pihak teknisi. *Business intelligence roadmap* ini jelas dan komprehensif untuk digunakan pada pengembangan *business intelligence* dengan adanya pembahasan kebutuhan bisnis, pelaksanaan proyek, hingga desain dari proses ETL hingga desain *physical database*. Metode ini menyediakan kerangka kerja untuk pelaksanaan proyek *business intelligence* dengan permulaan yang lebih mudah. Organisasi dapat meneruskan tahap *roadmap* berikutnya apabila telah memenuhi kriteria pada setiap tahapnya.

2.6.6. *Deployment*

Setelah melewati tahap *development*, selanjutnya adalah melakukan proses deployment, di mana tahap ini terdapat 2 kegiatan utama, yaitu:

1. *Implementation*

Setelah tim tester sudah melakukan proses *testing* terhadap semua komponen di dalam aplikasi BI, tim kemudian melakukan proses *deployment* terhadap database dan aplikasi BI. Training dijadwalkan untuk para eksekutif yang akan menggunakan aplikasi BI dan *meta data repository*. Bagian *support* mulai bekerja seperti melakukan proses *help desk*, pemeliharaan terhadap *database* BI, penjadwalan dan menjalankan proses ETL *batch job*, monitoring terhadap kinerja aplikasi serta tuning *database*. Kegiatan-kegiatan yang dilakukan pada tahap ini adalah merencanakan implementasi, membangun lingkungan produksi, instalasi semua komponen aplikasi BI, setup jadwal produksi, *load database* produksi, dan mempersiapkan *support*.

2. *Release Evaluation*

Dalam tahapan ini, semua proses, teknik, petunjuk, dan lain-lain yang sudah di-*release* sebelumnya dipelajari dan dievaluasi untuk menghasilkan aplikasi yang lebih baik dan efisien. Kegiatan-kegiatan yang dilakukan pada tahap ini adalah mempersiapkan *review* pasca implementasi, mengorganisasikan *review meeting* pasca implementasi, melakukan *meeting* untuk *re-review* pasca implementasi, dan melakukan *follow-up* hasil *meeting* pasca implementasi.

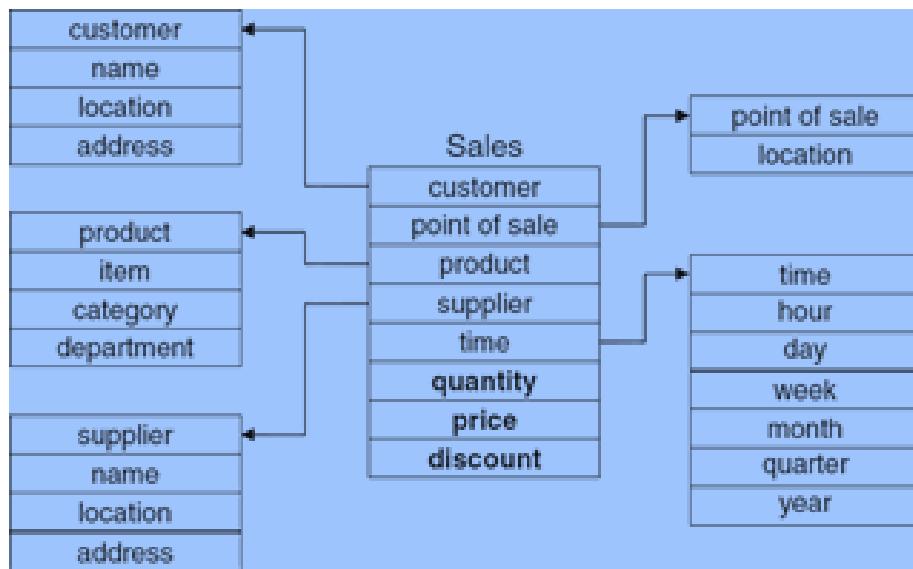
2.7 Model Dimensional

Model dimensional merupakan sebuah perancangan logika yang bertujuan untuk menampilkan data dalam bentuk standar dan intuitif yang mengizinkan pengaksesan dengan performa tinggi (Connolly et al., 2010). Model dimensional menggunakan konsep model hubungan antar entitas dengan beberapa batasan yang penting. Setiap model dimensi terdiri dari sebuah tabel dengan sebuah *composit key*, disebut dengan tabel fakta. Sedangkan untuk satu set tabel yang lebih kecil disebut tabel dimensi. Setiap tabel dimensi memiliki sebuah *simple primary key* yang merespon tepat pada satu komponen *composit key* di tabel fakta. Dengan kata lain, *primary key* dari sebuah tabel fakta terdiri atas dua atau lebih *foreign key*. Struktur karakteristik ini disebut dengan skema bintang atau join bintang.

Fitur penting lainnya dari sebuah model dimensional adalah semua *natural keys* diganti dengan kunci pengganti (*surrogate keys*). Setiap kali *join* antar tabel fakta dengan tabel dimensi selalu didasari oleh kunci pengganti, bukan *natural keys*. Kegunaan dari kunci pengganti adalah memperbolehkan data pada data warehouse untuk memiliki beberapa kebebasan dalam penggunaan data, tetapi tidak seperti halnya yang diproduksi oleh sistem OLTP. Model dimensional yang sering digunakan adalah *star schema* atau *snowflake* yang mudah dimengerti dan sesuai dengan kebutuhan bisnis, mendukung query sederhana dari bisnis, dan menyediakan performa *query* yang superior dengan meminimalisasi tabel-tabel *join*.

Menurut Vercellis (2009) *star schema* merupakan representasi multidimensi didasarkan pada *star schema* yang berisi dua jenis tabel yaitu tabel dimensi dan tabel fakta. Sedangkan menurut Connolly et al. (2010) *star schema* adalah struktur logikal yang mempunyai sebuah tabel fakta yang berisi data fakta di tengah dan dikelilingi oleh tabel-tabel dimensi yang berisikan data referensi atau keterangan yang biasanya dapat didenormalisasikan.

Menurut Vercellis (2009) *star schema* dibagi menjadi dua macam yaitu tabel dimensi dan tabel fakta. Tabel fakta adalah tabel yang umumnya mengandung sesuatu yang dapat diukur (*measure*) dan bersifat historis, dan merupakan kumpulan *foreign key* dari *primary key* yang terdapat pada masing-masing dimensional tabel. Tabel fakta cenderung mengacu pada transaksi dan berisikan dua jenis data yaitu pertama *link* ke tabel dimensi yang diperlukan sebagai acuan dari referensi informasi yang terkandung dalam setiap tabel fakta, dan kedua adalah nilai numeric dari atribut yang menjadi ciri transaksi yang termasuk dan yang mewakili target sebenarnya dari OLAP analisis. Tabel dimensi adalah tabel yang berisikan kategori-kategori dengan ringkasan detail yang dapat berupa laporan. Secara umum dapat disebut dengan dimensi yang terkait dengan entitas sekitarnya dalam proses organisasi.



Gambar 2.7 Star Schema (Vercellis, 2009).

2.8 Pengertian OLTP (*Online Transactional Processing*)

OLTP adalah sistem yang memiliki performa tinggi, dirancang untuk menangani proses transaksi yang terjadi pada perusahaan dan data perusahaan tersebut ke dalam *database*" (Connolly et al., 2010). Sistem OLTP memungkinkan banyak user untuk mengakses data pada waktu yang bersamaan serta melakukan proses yang mereka perlukan. Seluruh penjumlahan yang disimpan dari transaksi harian perusahaan adalah *history* dari perusahaan tersebut.

2.9 Pengertian OLAP (*Online Analytical Processing*)

Menurut Turban, Sharda, Delen dan King (2011) OLAP adalah kemampuan dari memanipulasi data secara efisien dari beberapa pandangan (perspektif). Struktur operasional utama pada OLAP berdasarkan pada konsep yang disebut *cube* (kubus). *Cube* dalam OLAP merupakan struktur data multidimensional (aktual/virtual) yang memungkinkan analisa data secara cepat. Susunan data pada cube berfungsi untuk mengatasi keterbatasan *database relational*. *Database relational* tidak sesuai untuk analisa secara cepat dari jumlah data yang besar, yang lebih sesuai adalah dengan memanipulasi *record* (dengan cara menghapus, menambahkan serta memperbarui data) yang mewakili berbagai transaksi.

Menurut Adnyana, (2013) dalam jurnalnya yang berjudul Analisis Perancangan Business Intelligence Berbasis SaaS *Cloud Computing* menjelaskan bahwa *Online Analytical Processing* (OLAP) merupakan kunci dari *business intelligence*, digunakan untuk meningkatkan analisis bisnis, merupakan perhitungan *Decision Support System* (DSS) dan *Expert Infotmation System* (EIS) yang dilakukan oleh *end-user* pada sistem online. OLAP digunakan pada banyak aplikasi, mulai dari pelaporan (*reporting*) perusahaan sampai DSS. Beberapa aktivitas yang dilakukan OLAP antara lain *generate query*, meminta laporan *ad hoc*, mendukung analisis statistik, analisis interaktif, dan membangun aplikasi multimedia. Untuk memfasilitasi OLAP ini diperlukan data warehouse

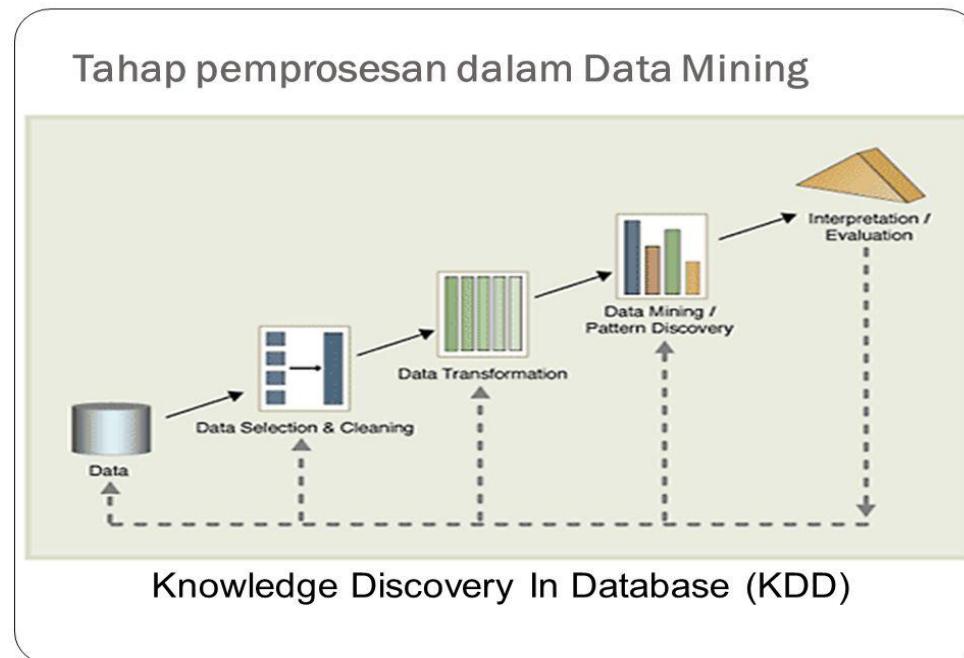
2.10 Data Mining

Menurut Hermawati (2013), Data mining adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan secara otomatis. Data mining merupakan proses iteratif dan interaktif untuk menemukan pola atau model baru yang sah (sempurna), bermanfaat dan dapat dimengerti dalam suatu database yang sangat besar (*massive database*).

Data mining berisi pencarian *trend* atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti. Selain itu juga data mining disebut sebagai suatu proses untuk menemukan hubungan yang berarti pola dan kecenderungan dengan memeriksa dalam sekumpulan data besar yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Larose, 2005).

Data mining mampu membantu perusahaan untuk mendapatkan pola dari data – data yang tersimpan di dalam basis data perusahaan. Pengetahuan yang diperoleh tersebut akan menjadi pedoman dalam mengambil tindakan

- tindakan bisnis sebagai upaya pemeliharaan dan peningkatan fungsi – fungsi berikut: deskripsi, estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi (Nurjoko dan Kurniawan, 2016).



Gambar 2.8 Tahapan Data Mining (Han & Kamber, 2006)

Tahap-tahap data mining ada 6 yaitu :

1. Pembersihan data (*data cleaning*) Pembersihan data merupakan proses menghilangkan *noise* dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh, baik dari database suatu perusahaan maupun hasil eksperimen, memiliki isian-isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga hanya sekedar salah ketik. Selain itu, ada juga atribut-atribut data yang tidak relevan dengan hipotesa data mining yang dimiliki. Data-data yang tidak relevan itu juga lebih baik dibuang. Pembersihan data juga akan mempengaruhi performasi dari teknik data mining karena data yang ditangani akan berkurang jumlah dan kompleksitasnya.
2. Integrasi data (*data integration*) Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam satu *database* baru. Tidak jarang data yang diperlukan untuk data mining tidak hanya berasal dari satu *database* tetapi juga berasal dari beberapa *database* atau *file teks*.

Integrasi data dilakukan pada atribut-atribut yang mengidentifikasi entitas-entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan dan lainnya. Integrasi data perlu dilakukan secara cermat karena kesalahan pada integrasi data bisa menghasilkan hasil yang menyimpang dan bahkan menyesatkan pengambilan aksi nantinya. Sebagai contoh bila integrasi data berdasarkan jenis produk ternyata menggabungkan produk dari kategori yang berbeda maka akan didapatkan korelasi antar produk yang sebenarnya tidak ada.

3. Seleksi Data (*Data Selection*) Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus *market basket analysis*, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan saja.
4. Transformasi data (*Data Transformation*) Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining. Beberapa metode data mining membutuhkan format data yang khusus sebelum bisa diaplikasikan. Sebagai contoh beberapa metode standar seperti analisis asosiasi dan clustering hanya bisa menerima input data kategorikal. Karenanya data berupa angka numerik yang berlanjut perlu dibagibagi menjadi beberapa interval. Proses ini sering disebut transformasi data.
5. Proses mining, Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.
6. Evaluasi pola (*pattern evaluation*), Untuk mengidentifikasi pola-pola menarik kedalam knowledge based yang ditemukan. Dalam tahap ini hasil dari teknik data mining berupa pola-pola yang khas maupun model prediksi dievaluasi untuk menilai apakah hipotesa yang ada memang tercapai. Bila ternyata hasil yang diperoleh tidak sesuai hipotesa ada beberapa alternatif yang dapat diambil seperti menjadikannya umpan balik untuk memperbaiki proses data mining, mencoba metode data mining lain yang lebih sesuai, atau menerima hasil ini sebagai suatu hasil yang di luar dugaan yang mungkin bermanfaat.

7. Presentasi pengetahuan (*knowledge presentation*), Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Tahap terakhir dari proses data mining adalah bagaimana memformulasikan keputusan atau aksi dari hasil analisis yang didapat. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bisa dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, visualisasi juga bisa membantu mengkomunikasikan hasil data mining (Han, 2006)

Menurut Gartner Group, *Datamining is the process of discovering meaningful new correlations, patterns and trends by sifting through large amounts of data stored in repositories, using pattern recognition technologies as well as statistical and mathematical techniques.* Selain definisi di atas, ada juga beberapa definisi maupun pengertian mengenai data mining seperti yang tertera sebagai berikut:

- a. Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data.
- b. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.
- c. Data mining ialah campuran dari statistic, kecerdasan buatan dan riset basis data yang masih berkembangkan.

Menurut Moss & Atre (2003), data mining adalah analisis data dengan maksud untuk menemukan informasi tambahan yang tersembunyi dalam jumlah data yang besar yang telah didapatkan pada saat bisnis berjalan. Data mining memiliki teknik-teknik untuk melakukan implementasi spesifik dari algoritma yang digunakan dalam proses data mining, teknik-teknik data mining menurut Moss & Atre (2003) adalah sebagai berikut :

1. *Associations Discovery*

Teknik ini digunakan untuk mengidentifikasi perilaku peristiwa tertentu dalam suatu proses. Teknik ini memungkinkan untuk menemukan keterkaitan atas terjadinya suatu proses dengan proses lainnya.

2. *Sequential Pattern Discovery*

Teknik ini memiliki kemiripan dengan teknik asosiasi kecuali pola sekuensial dari waktu ke waktu menentukan bagaimana suatu item atau objek berhubungan satu sama lain dari waktu ke waktu.

3. *Classification*

Teknik klasifikasi adalah teknik yang banyak digunakan dalam data mining. Klasifikasi melihat pada perilaku dan atribut dari kelompok yang telah ditentukan.

4. *Clustering*

Teknik ini digunakan untuk menemukan kelompok yang berbeda dalam data. teknik ini serupa dengan klasifikasi kecuali bahwa ada dari kelompok-kelompok tersebut yang belum ditetapkan pada awal menjalankan data mining tool.

5. *Forecasting*

Teknik data mining *forecasting* yaitu *time series* memperkirakan hanya tergantung pada waktu nilai data. metode ini membuat peramalan dengan menggunakan asumsi bahwa masa depan adalah fungsi dari masa lalu. Tujuannya adalah untuk menentukan pola dalam deret data historis dan menterjemahkan pola tersebut ke masa depan. Misalkan, menentukan tingkat kecelakaan selama musim liburan berdasarkan jumlah kecelakaan yang terjadi selama musim liburan yang sama pada tahun sebelumnya. Untuk mengukur kesesuaian suatu metode peramalan tertentu untuk suatu kumpulan data yang diberikan maka dibutuhkan ketepatan peramalan. Ketepatan peramalan dipandang sebagai kriteria penolakan untuk memilih suatu metode peramalan.

Dalam *time series*, dari data masa lalu dapat diramalkan situasi yang akan terjadi pada masa yang akan datang. Untuk menguji kebenaran peramalan, digunakanlah ketepatan peramalan. Salah satu kriteria yang

digunakan untuk menguji ketepatan peramalan adalah Nilai Tengah *Galat Absolut (Mean Absolute Error)* yaitu rata-rata nilai *absolute error* dari kesalahan meramal. Formula dalam ketepatan peramalan *Mean Absolute Error* (MAE).

2.11 Analisa SWOT

Menurut Pearce dan Robinson (2013) analisis SWOT perlu dilakukan untuk mencocokan fit antara sumber daya internal dan sumber daya eksternal perusahaan. Analisis yang baik dapat memaksimalkan kekuatan dan peluang perusahaan serta meminimalkan kelemahan dan ancaman. Empat faktor analisis SWOT adalah sebagai berikut :

1. *Strength*

Sebuah kekuatan dalam perusahaan adalah sumber daya dan kemampuan yang dapat digunakan sebagai dasar dalam pengembangan *competitive advantage*.

2. *Weakness*

Kelemahan merupakan sifat-sifat dalam perusahaan yang cenderung mengurangi nilai- nilai kompeten dan faktor internal yang menghambat perusahaan dalam mencapai tujuannya.

3. *Opportunity*

Peluang-peluang yang ada dalam perusahaan yang dapat menambahkan peluang baru bagi sebuah perusahaan untuk meraih keuntungan.

4. *Threat*

Kejadian-kejadian yang dapat terjadi yang mengakibatkan kerugian tertentu dalam perusahaan serta dapat menghambat perusahaan.

2.12 Strategi Marketing

pemasaran merupakan suatu proses sosial dan manajerial yang didalamnya individu dan kelompok mendapatkan apa yang mereka butuhkan dan inginkan dengan menciptakan, menawarkan, dan mempertukarkan produk yang bernilai kepada pihak lain (Hartono et al., 2012). Penjelasan di atas menjelaskan bahwa pemasaran merupakan suatu proses yang dilakukan untuk menjangkau orang dan mengubahnya menjadi pelanggan untuk produk yang kita sediakan.

2.13 Pentaho Data Integraton/Kettle

Pentaho Data Integration atau Kettle adalah software dari Pentaho yang dapat digunakan untuk proses ETL (*Extraction, Transformation dan Loading*). *Pentaho Data Integration* dapat digunakan untuk migrasi data, membersihkan data, *loading* dari *file* ke *database* atau sebaliknya dalam volume besar, (Mulyana, 2014), *Pentaho Data Integration* menyediakan *graphical user interface* dan *drag-drop* komponen yang memudahkan *user*. Elemen utama dari *Pentaho Data Integration* adalah *Transformation* dan *Job*. *Transformation* adalah sekumpulan instruksi untuk merubah *input* menjadi *output* yang diinginkan (*input-proses-output*). Sedangkan *Job* adalah kumpulan instruksi untuk menjalankan *transformasi*. Ada tiga komponen dalam *Pentaho Data Integration* : *Spoon*, *Pan* dan *Kitchen*. *Spoon* adalah *user interface* untuk membuat *Job* dan *Transformation*. *Pan* adalah *tools* yang berfungsi membaca, merubah dan menulis data. Sedangkan *Kitchen* adalah program yang mengeksekusi *jo*.

2.14 Software Weka

Weka (*Waikato Environment for Knowledge Analysis*) merupakan utilitas data mining yang sangat populer dan digunakan untuk *predictive analysis*. Mendukung data sumber/source berformat ARFF (*Attribute-Relation File Format*), CSV (*Comma Separated Value*), maupun *table/view* dari relasional *database*, mendukung data nominal dan numerik (*continuous maupun discrete*). Pentaho Data Mining juga mendukung algoritma yang cukup banyak untuk operasi dasar data mining seperti *Classification*, *Clustering* dan *Association*. Utilitas grafis yang membantu pengolahan dan pemahaman data mulai dari *preprocessing*, data visualization dalam bentuk *plot* maupun *curve, tree visualization*.