

BAB III

LANDASAN TEORI

Dalam bab ini akan dijelaskan tentang beberapa konsep tentang supra desa, *business intelligence*, data warehouse, staging area, ETL, OLAP, ROLAP, Pentaho Data Integration, dan PHP.

3.1 Supra Desa

Supra Desa merupakan subjek dari penelitian ini. Menurut Undang-undang No. 6 Tahun 2014, desa adalah kesatuan masyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur dan mengurus urusan pemerintahan, kepentingan masyarakat setempat berdasarkan prakarsa masyarakat, hak asal usul, dan/atau hak tradisional yang diakui dan dihormati dalam sistem pemerintahan Negara Kesatuan Republik Indonesia (Anon., 2014). Supra Desa adalah bagian dari pemerintahan yang ada di atas desa. Dalam penelitian ini, supra desa yang dimaksud adalah pemerintah kabupaten/kota. Menurut Undang-Undang No. 6 Tahun 2014, pemerintah daerah kabupaten dapat melakukan penataan desa berdasarkan hasil evaluasi tingkat perkembangan pemerintah desa (Anon., 2014).

Penugasan pemerintah daerah kepada desa meliputi penyelenggaraan pemerintahan desa, pelaksanaan pembangunan desa, pembinaan kemasyarakatan desa, dan pemberdayaan masyarakat desa. Pemerintah desa menyusun perencanaan pembangunan desa sesuai dengan kewenangannya dengan mengacu pada perencanaan pembangunan Kabupaten/Kota. Pemerintah daerah memiliki kewajiban

dalam mengembangkan sistem informasi desa dan pembangunan kawasan perdesaan. Sistem informasi desa meliputi data desa, data pembangunan desa, kawasan perdesaan, serta informasi lain yang berkaitan dengan pembangunan desa dan pembangunan kawasan perdesaan. Hal ini membuat pemerintah daerah kabupaten harus menyediakan informasi perencanaan pembangunan kabupaten/kota untuk desa (Anon., 2014).

3.2 Business Intelligence

Business Intelligence merupakan solusi yang akan dikembangkan dalam penelitian ini. *Business intelligence* adalah penyampaian informasi yang akurat dan berguna bagi pengambil keputusan pada waktu tertentu untuk mendukung pengambilan keputusan yang efektif (Larson, 2009). *Business intelligence* adalah kumpulan metodologi, proses, arsitektur, dan teknologi yang berfungsi untuk mentransformasi-kan data menjadi informasi yang mengandung arti yang berguna. Informasi yang dihasilkan *business intelligence* digunakan sebagai landasan dalam pengambilan keputusan dalam sebuah perusahaan (JRP, 2014). *Business intelligence* memiliki arti menggunakan aset data yang dimiliki untuk membuat keputusan bisnis yang lebih baik. *Business intelligence* memiliki banyak konsep, dan akan terus berevolusi dan semakin disempurnakan berdasarkan pengalaman yang diperoleh. Untuk meningkatkan persaingan dalam bisnis, sangat penting untuk organisasi mengefektifkan penggunaan biaya dan memiliki akses yang cepat ke informasi bisnis bagi semua pengguna bisnis. Solusi untuk masalah ini adalah sistem *business intelligence* yang menyediakan

seperangkat teknologi dan produk untuk memberikan kepada pengguna, informasi yang mereka butuhkan untuk menjawab pertanyaan bisnis dan membuat keputusan bisnis secara taktis dan strategis (Almeida, et al., 1999).

Keuntungan dari *business intelligence* adalah:

1. Sistem *business intelligence* tidak hanya mendukung teknologi informasi yang terbaru, tetapi juga menyediakan solusi dengan paket aplikasi.
2. Sistem *business intelligence* berfokus pada akses dan penyampaian informasi bisnis untuk pengguna akhir, dan mendukung baik itu penyedia informasi maupun pengguna informasi.
3. Sistem *business intelligence* mendukung akses untuk segala bentuk informasi bisnis, dan tidak hanya dari informasi yang disimpan di *data warehouse* (Almeida, et al., 1999).

3.3 Data warehouse

Data warehouse merupakan sumber data yang akan dianalisis dalam *business intelligence* yang dikembangkan dalam penelitian ini. *Data warehouse* adalah pusat data yang dibentuk dari hasil penggabungan dan pengolahan data dari beragam sumber data (JRP, 2014). *Data warehouse* adalah kumpulan data yang berorientasi subjek, terintegrasi, *non volatile*, dan bersifat *time variant* untuk mendukung keputusan manajemen (Inmon, 2002). Berorientasi subjek maksudnya adalah *data warehouse* didesain untuk menganalisa data berdasarkan subjek-subjek tertentu dalam organisasi, bukan pada proses atau fungsi dari aplikasi tertentu. Terintegrasi maksudnya sumber data yang ada dalam *data warehouse* tidak hanya

berasal dari data operasional (*internal source*), tetapi juga berasal dari data di luar sistem (*external source*). *Non-volatile* maksudnya pada *data warehouse* hanya ada dua kegiatan memanipulasi data, yaitu *loading data* (mengambil data) dan akses data. *Time variant* maksudnya sistem operasional mengandung data yang bernilai sekarang, sedangkan data dalam *data warehouse* mengandung data tidak hanya data terkini, tetapi juga data masa lampau.

Keuntungan dari *data warehouse* adalah:

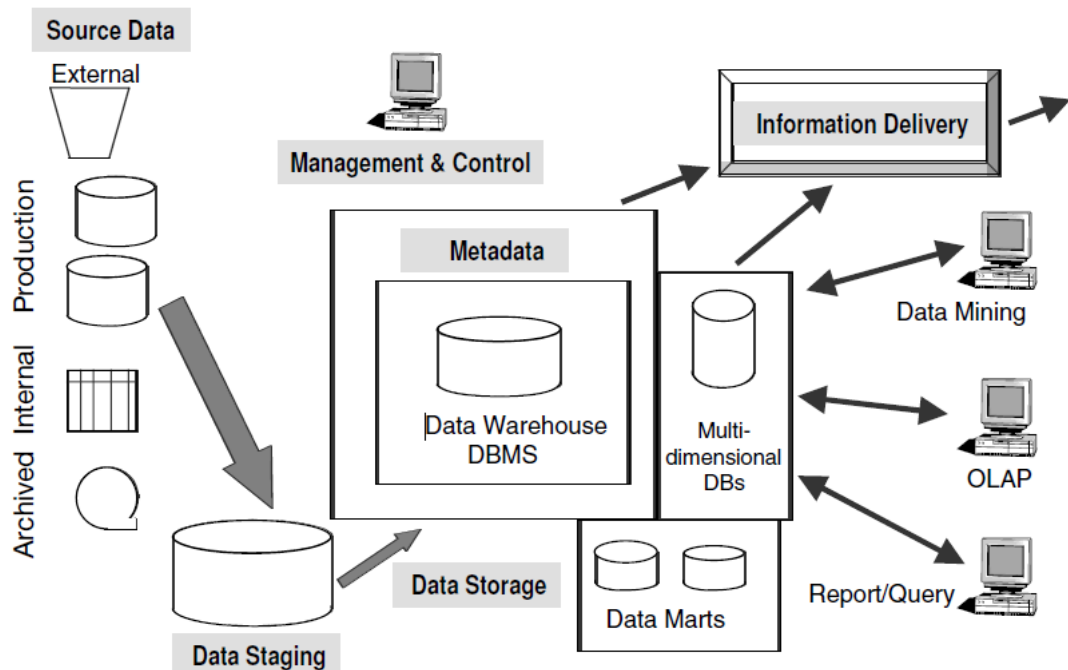
- a. *Data warehouse* didesain untuk memenuhi kebutuhan pengguna bisnis dan bukan aplikasi operasional sehari-hari.
- b. Informasi untuk *data warehouse* telah dibersihkan dan konsisten, dan disimpan dalam bentuk yang dapat dimengerti oleh pengguna bisnis.
- c. Tidak seperti sistem operasional, yang hanya mengandung data saat ini, *data warehouse* dapat menyediakan informasi yang *historical* maupun informasi yang telah diringkas.
- d. Penggunaan komputasi *client server* menyediakan kepada pengguna *data warehouse* antarmuka pengguna yang telah ditingkatkan dan alat yang sangat berguna untuk mendukung keputusan (Almeida, et al., 1999).

Menurut Poniah (2001), *data warehouse* memiliki beberapa komponen seperti yang ada dalam gambar 3.1 dibawah, yaitu sumber data (*data source*), *data staging*, penyimpanan data (*data storage*), manajemen dan kontrol (*management & control*), *metadata*, dan penyampaian

informasi (*information delivery*). Sumber data terdiri dari empat kategori, yaitu:

1. *Production data*. Data pada kategori ini diperoleh dari sistem operasional yang ada di perusahaan.
2. *Internal data*. Data pada kategori ini merupakan data yang sifatnya pribadi/internal, seperti *spreadsheet*, dokumen, profil pelanggan, dan data departemen yang bersifat *privat*.
3. *Archived data*. Data pada kategori ini diperoleh dari data sistem operasional yang telah dijadikan arsip.
4. *External data*. Data pada kategori ini berasal dari luar, seperti data statistik hasil penelitian, dan lainnya.

Pada komponen *data staging* menyediakan tempat dan area dengan kumpulan fungsi untuk membersihkan, mengubah, menggabungkan, menghilangkan duplikasi, dan mempersiapkan sumber data dari penyimpanan dan digunakan pada *data warehouse*. Komponen *data storage* menyimpan dan mengelola *data warehouse*. Selain itu, *data storage* juga menyimpan *metadata*, yang menyimpan informasi mengenai struktur data, *file*, alamat, *index*, dan sebagainya. Komponen manajemen dan kontrol mengkoordinasikan layanan dan aktivitas yang berkaitan dengan *data warehouse*. Komponen *information delivery* adalah cara penyampaian informasi kepada pengguna terhadap data yang tersimpan di *data warehouse* untuk dilakukan analisis, seperti data mining, OLAP, dan Report/Query (Ponniah, 2001).



Gambar 3.1 Komponen data warehouse (Ponniah, 2001)

3.4 Staging Area

Pada penelitian ini, dalam pembentukan *data warehouse* digunakan *staging area* untuk mempersiapkan data yang diperoleh dari database transaksional untuk disimpan di *data warehouse*. *Staging area* adalah tempat dikumpulkannya semua data yang telah diekstraksi, dan kemudian dipersiapkan untuk dimuat kedalam *data warehouse* (Ponniah, 2001). Pada area ini, dapat dilakukan pengujian terhadap setiap file yang diekstraksi, melihat aturan bisnis, melakukan transformasi data yang beragam, pengurutan dan penggabungan data, menghilangkan inkonsistensi data, dan membersihkan data. Menurut Inmon (2005), *staging area* adalah tempat dimana data transit ditempatkan, dan biasanya data tersebut datang dari sumber data sebelum memasuki lapisan ETL (Inmon, 2005). *Staging area* adalah

lokasi sementara dimana data dari sistem sumber di salin. *Staging area* dibutuhkan dalam arsitektur data warehouse karena alasan waktu. Jadi, *staging area* menyimpan semua data yang diperlukan sebelum data diintegrasikan ke dalam *data warehouse*. Hal ini dikarenakan bervariasinya siklus bisnis, siklus pemrosesan data, perangkat keras dan keterbatasan sumber daya jaringan dan faktor geografis, sehingga tidak memungkinkan semua data diambil dari *database* operasional dalam waktu yang bersamaan (<http://data-warehouses.net>, n.d.).

3.5 ETL

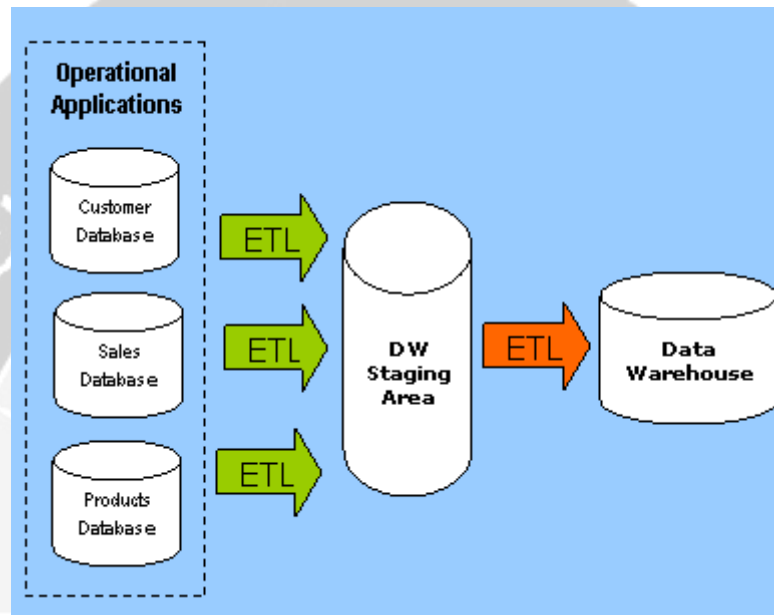
Pada penelitian ini, dilakukan proses ETL dari sumber data ke *staging area*, maupun dari *staging area* ke *data warehouse* seperti gambar 3.2 dibawah. ETL (*Extract Transform dan Load*) adalah sekumpulan proses untuk mengambil dan memproses data dari satu atau banyak sumber menjadi sumber baru. Sumber data yang diolah ETL adalah *database OLTP*, *website*, *file teks*, *spreadsheet*, *database*, *email*, dan sebagainya.

ETL terdiri dari 3 bagian utama, yaitu:

1. *Extract* merupakan proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi proses-proses selanjutnya.
2. *Transform* merupakan proses yang berfungsi untuk mengubah data yang masuk menjadi data yang dikehendaki. Fungsi-fungsi tersebut dapat berupa pemindahan data, validasi data sesuai aturan yang ditetapkan, modifikasi isi, tipe atau struktur data,

integrasi atau penggabungan data, perhitungan dan lain-lain.

3. *Load* merupakan proses yang diperlukan untuk mengisi data ke target (JRP, 2014).



Gambar 3.2 Arsitektur data warehouse (<http://data-warehouses.net>, n.d.)

3.6 OLAP

Pada penelitian ini, sistem yang dibangun adalah OLAP yang merupakan salah satu teknologi untuk analisis. *Online Analytical Processing* (OLAP) adalah teknologi untuk kebutuhan analitik. Terdapat pula *Online Transactional Processing* (OLTP) yang merupakan teknologi untuk mengelola aplikasi yang berorientasi pada transaksi. Database yang digunakan pada OLAP disebut database OLAP dan Database yang digunakan pada OLTP disebut database OLTP. Database OLAP merupakan database yang dimaksimalkan untuk kecepatan dalam pembacaan (*select query*), sedangkan database OLTP merupakan

database yang digunakan untuk aplikasi yang berorientasi pada transaksi, yaitu aplikasi yang cenderung lebih banyak melakukan proses insert, update, dan delete secara *real-time*. Database OLTP merupakan salah satu sumber data untuk membentuk OLAP dengan melalui proses ETL.

OLAP mengandung dua tipe data dasar, yaitu:

1. *Measures* adalah data bilangan yang terukur, misalkan kuantitas (*quantity*), harga (*price*), nilai rata-rata (*averages*) dari kelompok nilai tertentu, jumlah (*sum*) dari kelompok nilai tertentu, dan lain sebagainya.
2. *Dimension* mengacu pada kategori yang digunakan untuk mengatur *measures*. Biasanya dikelompokkan dalam bentuk bertingkat (level/hirarki). Dimensi yang umumnya hampir selalu ada adalah dimensi waktu (*time dimension*). Pada dimensi ini, hirarki yang disusun biasanya tahun, kuartal, bulan, dan hari (JRP, 2014).

Perbedaan antara database OLTP dan OLAP terdapat pada skema tabel yang dibentuk. Skema tabel database OLTP umumnya berbentuk *normalization*, sedangkan database OLAP menerapkan skema *star* atau *snowflake*. (JRP, 2014).

3.7 ROLAP (Relational On Line Analytical Processing)

Arsitektur OLAP yang dikembangkan dalam penelitian ini adalah ROLAP. ROLAP memungkinkan data dianalisis dari basis data relasional yang menyimpan *data warehouse*. Data dasar yang diperoleh dari data transaksional disimpan dan kemudian diagregasikan di basis data relasional. Data *multi dimensi* disimpan di *data warehouse* dengan skema bintang (*star schema*). Hal

ini bertujuan untuk memastikan akses yang cepat dan fleksibel untuk menangani permintaan.

Arsitektur ROLAP yang memiliki ketergantungan akan tabel relasional memiliki kelebihan, yaitu dapat menyimpan data dalam jumlah besar dibandingkan arsitektur OLAP yang lain. Kekurangannya dari arsitektur ROLAP adalah dalam melakukan agregasi, lebih lambat dari arsitektur OLAP yang lain (Larson, 2009).

3.8 Pentaho Data Integration (Kettle)

Pada penelitian ini, aplikasi untuk melakukan proses ETL adalah Pentaho Data Integration atau nama lainnya Kettle. Pentaho Data Integration (Kettle) adalah aplikasi ETL yang bersifat *open source* dan sangat populer di dunia. Kettle merupakan bagian dari aplikasi Pentaho. Proyek ini berdiri sendiri dan merupakan inisiatif dari Matt Casters dan diakuisisi oleh Pentaho pada tahun 2006. Sejak diakuisisi, Kettle berubah nama menjadi Pentaho Data Integration (<http://pentaho.phi-integration.com/>, n.d.). Kettle terdiri dari beberapa bagian, yaitu:

- a. Step adalah sebuah blok bagian atau node dari transformation yang memiliki fungsi dan tugas tertentu, yang diwakili dengan ikon-ikon unik.
- b. Transformation, yaitu komponen Kettle yang menangani proses manipulasi aliran data. Semua proses ETL dilakukan dalam transformation. Dalam sebuah transformation, terdapat satu atau lebih step.
- c. Job, yaitu komponen Kettle yang terdiri dari satu atau lebih job entry yang dijalankan dalam urutan tertentu.

- d. Job entry, yaitu blok bagian atau node dari job yang diwakili dengan ikon-ikon unik.
- e. Spoon, yaitu aplikasi grafis berbasis swing yang digunakan untuk merancang, menyunting, dan menjalankan file skema job dan transformation (<http://pentaho.phi-integration.com/>, n.d.).

3.9 PHP

Pada penelitian ini, bahasa pemrograman yang digunakan untuk mengembangkan sistem OLAP berplatform web adalah bahasa pemrograman PHP. PHP yang merupakan singkatan dari *PHP Hypertext Preprocessor* adalah bahasa yang dirangkai oleh relawan programmer dan telah digunakan pada lebih dari dua puluh juta situs web dan lebih dari sepertiga web server didunia. PHP bersifat fleksibel, dapat dikembangkan, dapat diperluas, stabil, dan terbuka, sehingga membuat PHP menjadi bahasa pemrograman yang paling populer didunia dalam pengembangan web (Vaswani, 2009).

PHP adalah bahasa pemrograman untuk membangun situs web yang dinamis dan interaktif. Sebagai aturan umum, program PHP berjalan di web server, dan menyediakan halaman web sesuai permintaan pengunjung. Salah satu fitur kunci dari PHP adalah kode PHP dapat ditempatkan di dalam halaman web HTML, sehingga sangat mudah untuk membuat konten yang dinamis dengan cepat. Halaman web yang dinamis maksudnya adalah halaman yang isinya dapat berubah secara otomatis setiap kali halaman ditampilkan. Situs web yang interaktif maksudnya adalah situs yang memberikan tanggapan sesuai masukan dari pengunjung (Doyle, 2010). Pada penelitian ini, juga akan digunakan

PHP dengan framework MVC (Model View Controller), yaitu framework Code Igniter (CI) karena framework CI merupakan framework PHP yang paling populer di Indonesia (Skvorc, 2015).

