

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

1.1 Tinjauan Pustaka

Lukas (2016) *“Implementasi Data Warehouse untuk mendukung Monitoring Kinerja Karyawan berdasarkan Nilai Penilaian Unjuk Kerja”* penelitian ini untuk membangun data warehouse dan Online Analytical Processing (OLAP) yang mampu memberi informasi mengenai kinerja karyawan untuk ditampilkan dalam bentuk grafik untuk membantu bagian SDM dalam melakukan monitoring kinerja karyawan.

Andik (2016) *“Implementasi Data Warehouse pada Sistem Pemetaan Profil Mahasiswa Studi Kasus Universitas Nusantara PGRI Kediri”* penelitian ini untuk mengumpulkan informasi, menganalisis kebutuhan dan permasalahan yang dihadapi oleh universitas, terkait dengan pengambilan keputusan serta merancang aplikasi data warehouse yang dijadikan sebagai alat bantu dalam pengambilan keputusan untuk menentukan strategi pemasaran.

Bakhtiar (2015) *“Pengembangan Sistem Analisis Akademis menggunakan OLAP dan Data Clustering Studi Kasus : Akademik Universitas Sebelas Maret Surakarta”* penelitian ini untuk memenuhi kebutuhan user, sistem dapat melakukan drill up dan drill down untuk membantu merekap data akademik dari berbagai dimensi dan mengelompokkan data lulusan.

Supriyatna.A dan Wahyudi.M (2012) “*Perancangan Data Warehouse Pada Perpustakaan Bina Sarana Informatika*” penelitian ini menggunakan metode OLAP membuat rancangan data warehouse beserta aplikasinya yang dijadikan sebagai pangkalan data yang berisi mengenai informasi perpustakaan yang akan dijadikan sebagai pengambilan keputusan dan menyajikan informasi secara multidimensi yang tepat, akurat, dan cepat.

Difa Isnaini (2023) Penelitian ini menganalisis Penilaian siswa menggunakan metode Star Schema untuk menampilkan data nilai pada periode tertentu secara cepat dan akurat, sehingga dapat mempermudah proses pelaporan nilai. Dengan judul “*Implementasi Data Warehouse untuk Analisa Penilaian Siswa di SD Muhammadiyah Wedi dengan Metode Star Schema*”.

Tabel 2.1 Perbandingan Penelitian

| Penulis | Judul | Metode | Keterangan |
|--------------|---|--------|---|
| Lukas (2016) | Implementasi Data Warehouse untuk mendukung Monitoring Kinerja Karyawan berdasarkan Nilai Penilaian Unjuk Kerja | OLAP | Memberi informasi dalam bentuk grafik untuk membantu bagian SDM dalam melakukan monitoring. |
| Andik (2016) | Implementasi Data Warehouse pada Sistem Pemetaan Profil Mahasiswa Studi Kasus Universitas Nusantara PGRI Kediri | OLAP | Bertujuan menghasilkan informasi sebagai pengambilan keputusan serta merancang aplikasi. |

| | | | |
|-----------------------------------|---|----------------------------|---|
| Bakhtiar (2015) | Pengembangan Sistem Analisis Akademis menggunakan OLAP dan Data Clustering Studi Kasus : Akademik Universitas Sebelas Maret Surakarta | OLAP | Membantu merekap data akademik dari berbagai dimensi dan mengelompokkan data lulusan. |
| Supriyatna.A dan Wahyudi.M (2012) | Perancangan Data Warehouse pada Perpustakaan Bina Sarana Informatika | OLAP Nine-Step Methodology | Membuat rancangan data yang dijadikan sebagai pengambilan keputusan dan menyajikan informasi secara multidimensi yang tepat, akurat, dan cepat. |
| Difa Isnaini (2023) | Implementasi Data Warehouse untuk Analisa Penilaian Siswa dengan Metode Star Schema | OLAP | Menyajikan informasi data nilai siswa pada periode tertentu secara cepat dan akurat. |

2.2. Dasar Teori

2.2.1. OLAP (*Online Analytical Processing*)

OLAP adalah suatu metode khusus untuk melakukan analisis terhadap data yang terdapat dalam media penyimpanan data (basis data) dan kemudian membuat laporannya sesuai dengan permintaan pengguna (Hermawan, 2005).

OLAP merupakan sebuah *tools* (alat) digunakan untuk melakukan analisis data secara dinamis pada skala enterprise yang mana didalamnya mencakup tahapan menciptakan (*create*), manipulasi, menampilkan, serta memformulasikan model analisis data. OLAP mendukung analisis data multidimensi, yang memungkinkan

pengguna untuk melihat data yang sama dengan cara yang berbeda menggunakan berbagai dimensi. OLAP juga dapat melihat data secara detail (*drill-down*) dan menuju kearah global (*roll-up*), serta mengkaitkan data atau informasi dari beberapa sumber (*drill-throught*).

Data warehouse adalah sekumpulan data yang bersifat integrated, subject-oriented, time variant dan nonvolatile dalam mendukung pengambilan keputusan manajemen (Inmon, 2005). Karakteristik data warehouse menurut Inmon :

1. *Subject-oriented* (Berorientasi Subjek)

Data warehouse didesain untuk menganalisa data berdasarkan subject-oriented tertentu dalam organisasi, bukan pada proses atau fungsi aplikasi tertentu.

2. *Integrated* (Terintegrasi)

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

3. *Time Variant* (Rentang Waktu)

Seluruh data pada data warehouse dapat dikatakan akurat atau valid pada rentang waktu tertentu. Cara-cara yang digunakan untuk melihat interval waktu yang digunakan dalam mengukur keakuratan suatu data warehouse :

- a. Cara yang paling sederhana adalah menyajikan data warehouse pada rentang waktu tertentu.
- b. Menggunakan variasi/perbedaan waktu yang disajikan dalam data warehouse baik implicit maupun explicit, secara explicit dengan unsure waktu dalam hari, minggu, bulan, dan waktu tertentu.
- c. Variasi yang disajikan data warehouse melalui serangkaian snapshot yang panjang.

4. *Nonvolatile*

Data pada data warehouse tidak di-update secara real time, tetapi di-refresh dari sistem operasional secara reguler. Data yang baru selalu ditambahkan bagi basis data itu sebagai sebuah perubahan. Basis data tersebut secara terus menerus menyerap data baru ini, kemudian disatukan dengan data sebelumnya. Hal ini berbeda dengan basis data operasional yang dapat melakukan update, insert dan delete terhadap data, sedangkan pada data warehouse hanya ada dua kegiatan manipulasi data yaitu loading data (mengambil data) dan akses data.

Menurut Williams dalam (Gustiarahman, 2006) data warehouse dikembangkan untuk menjalankan tugas – tugas sebagai berikut :

1. Data Mining

Data mining merupakan proses untuk menggali pengetahuan dan informasi baru dari data yang berjumlah banyak pada data warehouse, dengan menggunakan kecerdasan buatan, statistic dan matematika.

2. Pembuatan Laporan

Pembuatan laporan merupakan salah satu manfaat dari data warehouse yang paling umum dilakukan. Laporan yang dihasilkan dapat disesuaikan dengan kebutuhan pimpinan dengan memilih sendiri subjek yang akan dianalisis.

3. *Online Analytical Processing* (OLAP)

OLAP mendayagunakan konsep data multi dimensi dan memungkinkan pemakai menganalisa secara detail tanpa harus menuliskan perintah *Structure Query Language* (SQL). Hal ini dimungkinkan karena pada konsep multidimensi data yang berupa fakta yang sama dapat dilihat dengan menggunakan fungsi yang berbeda. Dengan OLAP, informasi dapat dilihat dengan detail yang disebut dengan drilldown, atau secara lebih ringkas disebut dengan roll up.

4. Proses Informasi Eksekutif

Data warehouse dapat membuat ringkasan informasi yang penting, namun tetap dapat pula diketahui rinciannya. Hal ini memudahkan bagi pimpinan tingkat atas dalam pengambilan keputusan.

2.2.2. Teknik Pemodelan Data Warehouse

Pada data warehouse digunakan teknik pemodelan data yang disebut *dimensional modelling technique*. Menurut (Pusadan, 2013) “Pemodelan dimensional adalah suatu model berbasis pemanggilan yang mendukung akses query volume tinggi. Star Schema adalah alat dimana pemodelan dimensional diterapkan dan berisi sebuah tabel fakta pusat.”

Menurut Silver dalam (Supriyatna, 2011) terdapat dua teknik pemodelan data warehouse, yaitu :

1. Skema Bintang (*Star Schema*)

Skema yang mengikuti bentuk bintang, dimana terdapat satu tabel fakta (*fact table*) di pusat bintang dengan beberapa tabel dimensi (*dimensional tables*) yang mengelilinginya. Semua tabel dimensi berhubungan dengan tabel fakta. Tabel fakta memiliki beberapa key yang merupakan kunci indeks individual dalam tabel dimensi.

2. Skema Salju (*Snowflake Schema*)

Skema bola salju merupakan perluasan dari skema bintang dengan tambahan beberapa tabel dimensi yang tidak berhubungan secara langsung dengan tabel fakta. Tabel dimensi tersebut berhubungan dengan tabel dimensi yang lain.

Berdasarkan codd Rules OLAP mempunyai 5 aturan didalam codd rules yaitu *FASMI (Fast Analysis of Shared Multidimensional Information)*.

a. Fast

Sistem memiliki target untuk memberikan respon secepat mungkin pada pengguna menurut analisis yang dilakukan. Kecepatan dalam pemrosesan data, pengumpulan data dari berbagai sumber data, analisis data dan data multidimensi.

b. Analysis

Kemampuan OLAP untuk dapat melakukan analisis data, dan diperlukan adanya proses pelaporan secara digital (Reporting) dari hasil analisis data tersebut pada OLAP.

c. Shared

Sistem melakukan keseluruhan keperluan pengamanan data, yang mana apabila diperlukan banyak akses penulisan tentang data, maka perlu disesuaikan dengan level pengguna. Karena tidak seluruh aplikasi memerlukan pengguna menulis data kembali, maka sistem harus bisa menyelesaikan multiple update dalam satu waktu dengan aman.

d. Multidimensional

Multidimensional adalah kemampuan OLAP untuk dapat menyediakan, mendukung sudut pandang dan konsep data multidimensi beserta dengan hirarki data. Hal ini akan mempermudah dalam analisis data sesuai dengan tujuan dari organisasi. Namun jumlah dimensi dimodelkan harus sesuai dengan kebutuhan pengguna dan target.

e. Information

Informasi merupakan output dari OLAP dan data warehouse kepada pengguna. Informasi harus dapat disampaikan kepada pengguna dengan baik, serta dapat disimpan dalam media penyimpanan. Selain itu informasi didalam Codd Rules ini menggambarkan kemampuan OLAP didalam menangani dan mengolah data yang diperoleh dari informasi yang dihasilkan. Didalam sistem OLAP terdapat lima kemampuan dalam

menangani data, yaitu : Duplikasi data, Kebutuhan memori (RAM), Utilisasi penyimpanan (Hardisk), Integrasi terhadap data warehouse serta performa sistem.

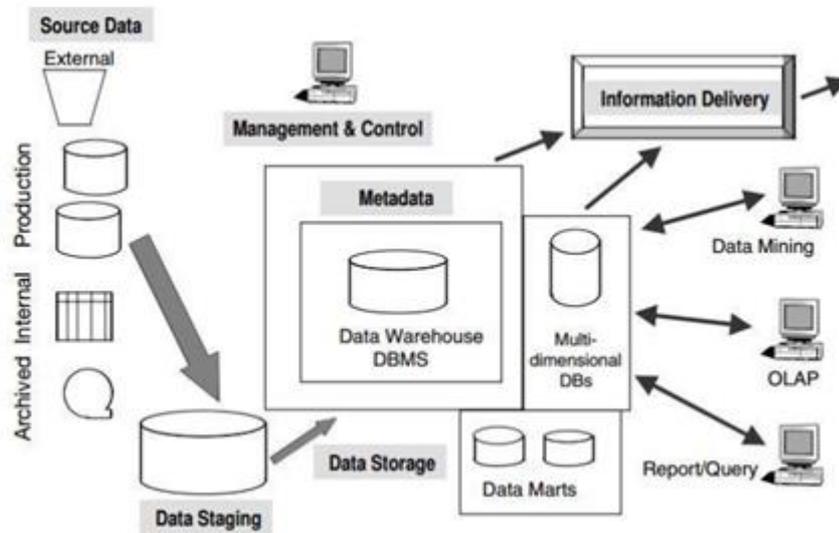
2.2.3. Data Warehouse

Menurut Inmon dalam Imhoff, Galemo dan Geiger (2003). Data warehouse adalah sekumpulan data yang berorientasi subjek, terintegrasi, memiliki rentang waktu dan tidak mudah berubah yang digunakan dalam mendukung pembuatan keputusan strategis. Menurut Silvers (2008) terdapat tujuh karakteristik pada data warehouse, yaitu : *Enterprise data, Subject Orientation, Data Inegration, Non Volatily, Time Variant, One Version of The Truth, Long Term Investent*.

Komponen Data Warehouse

Menurut Ponniah (2001), komponen data warehouse dapat divisualisasikan seperti yang terlihat pada gambar 1. Pada gambar 1, komponen source data terletak di sebelah kiri. Data source bisa berasal dari berbagai sumber baik internal source maupun external source. Komponen data staging sebagai blok pembangunan berikutnya setelah source data. Komponen data staging ini digunakan dalam proses ETL, komponen ini melakukan extract dari data source kemudian melakukan transform terhadap data itu, lalu data hasil transform ditransfer ke data warehouse. Pada bagian tengah dapat dilihat komponen data storage yang mengelola data warehouse, komponen ini tidak hanya menyimpan dan mengelola data, tetapi juga menjaga bagian data yang disebut metadata repository. Komponen information

delivery berada di sebelah kanan. Komponen tersebut terdiri dari semua hal yang berhubungan dengan penyediaan informasi dari data warehouse bagi pengguna.



Gambar 1. Komponen Data Warehouse

a. Model Dimensional Data Warehouse

Menurut Silvers (2008) terdapat dua konsep model data pada data warehouse, yaitu:

1. Fakta (*Fact*)

Fakta dikenal sebagai suatu kejadian atau transaksi. Fakta adalah sesuatu yang terjadi pada sebuah table fakta menggabungkan entitas yang di identifikasikan dalam logical data model.

2. Dimensi (*Dimensions*)

Dimensi adalah data yang menggambarkan kualifikasi entitas perusahaan yang terlibat dalam fakta. Hubungan antara tabel fakta dan dimensi dapat digambarkan dengan skema dimensional, skema dimensional yang paling umum adalah skema bintang (*Star Schema*) yang terdiri dari sebuah tabel fakta yang dikelilingi oleh tabel dimensi.

2.2.4. Pentaho Data Integration

Pentaho Data Integration (PDI) atau Kettle adalah software dari Pentaho yang dapat digunakan untuk proses ETL (*Extraction, Transformation dan Loading*). PDI dapat digunakan untuk migrasi data, membersihkan data, loading dari file ke database atau sebaliknya dalam volume besar. PDI menyediakan graphical user interface dan drag-drop komponen yang memudahkan user.

Elemen utama dari PDI 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 PDI: 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 job.