



Hak cipta dan penggunaan kembali:

Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:

This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.

2 BAB II

LANDASAN TEORI

2.1 Data dan Informasi

2.1.1 Data

Data adalah fakta yang belum diproses untuk menunjukkan artinya . (Rob, Coronel, & Crockett, 2008)

Data dapat berupa angka, simbol, gambar, suara atau tanda-tanda lain yang dapat diproses untuk menjadi informasi. Dalam kehidupan sehari-hari data dapat ditemukan dalam bentuk angka atau teks pada kertas, bit dalam komputer, dan memori atau fakta yang tersimpan dalam pikiran seseorang.

2.1.2 Informasi

Informasi adalah sebuah hasil proses dari data mentah untuk menunjukkan artinya. (Rob, Coronel, & Crockett, 2008)

Informasi merupakan sebuah hasil dari pengolahan sebuah data yang bermanfaat untuk menambah

Pengetahuan bagi yang menerimanya. Informasi disampaikan sebagai isi pesan atau melalui pengamatan langsung maupun tidak langsung.

2.2 Database

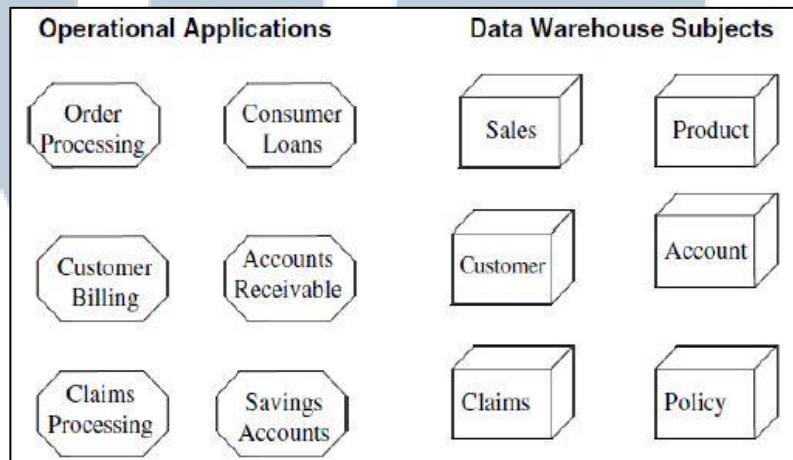
Database merupakan kumpulan data yang saling berelasi secara logis, penjelasan tentang data yang saling terhubung, dan perancangan data sampai sedemikian rupa, sehingga dapat dianalisa untuk menghasilkan informasi yang dibutuhkan. (Thomas & Carolyn, 2010).

Database merupakan komponen penting dalam suatu sistem karena *database* merupakan bagian yang berfungsi sebagai penyimpanan dan pengelolaan data untuk menjadi suatu informasi yang berguna. *Database* dibutuhkan oleh sistem dengan cara mengelompokkan data kedalam tabel yang telah dibuat.

2.3 Data warehouse

Data warehouse yang secara harafiah berarti gudang data sebetulnya memiliki makna lebih dari sekedar gudang. Bill Inmon yang dianggap sebagai pencetus *data warehouse* mendefinisikan bahwa *data warehouse* adalah *database* yang memiliki karakter *subject oriented*, *integrated*, *non-volatile*, dan *time variant*.

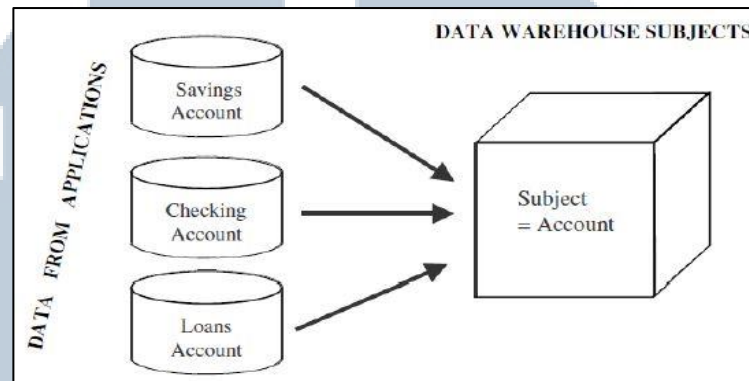
a) *Subject oriented*



Gambar 2.1 karakteristik *subject oriented*

Subject oriented berarti *data warehouse* disusun berdasar subjeknya. Subjek yang dimaksud disini adalah subjek areanya misal *data warehouse* untuk *customer*, produk, atau pelanggan. Dengan kata lain *data warehouse* didesain untuk membantu subjek-subjek tertentu sehingga data yang akan dimasukan ke *data warehouse* merupakan data yang sesuai dengan subjeknya.

b) *Integrated*

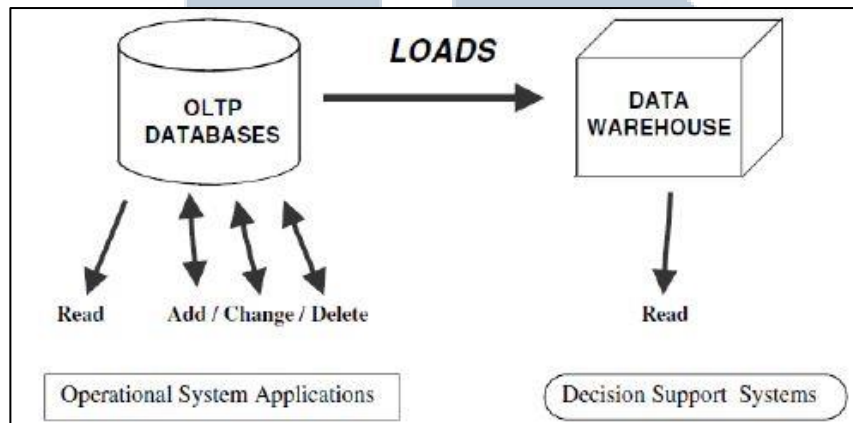


Gambar 2.2 karakteristik data terintegrasi

Integrated berarti data yang bermacam-macam dijadikan satu sehingga mudah untuk diolah. *Data warehouse* dapat menyimpan data-data dari berbagai macam sumber data yang terpisah kedalam suatu *format* yang konsisten dan terintegrasi satu sama lain. Syarat untuk mengintegrasikan satu sumber data dengan sumber data lainnya adalah dengan standarisasi penamaan variabel, ukuran variabel, dan atribut fisik dari *data*. Misal pada sumber data pertama *format* penulisan jenis kelamin menggunakan pria(P) atau wanita(W) sedangkan pada sumber data kedua menggunakan laki-laki(L) atau perempuan(P) maka sebelum dimasukkan kedalam *data warehouse* harus dilakukan standarisasi terlebih dahulu.

UNIVERSITAS
MULTIMEDIA
NUSANTARA

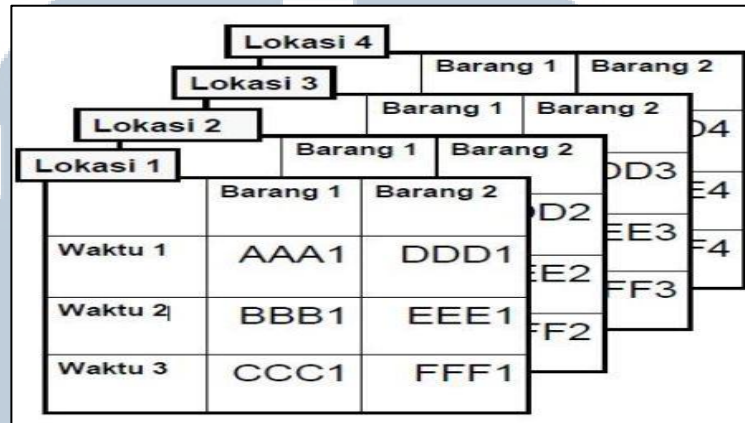
c) *Non-volatile*



Gambar 2.3 karakteristik data *non-volatile*

Non-volatile maksudnya adalah data yang terdapat di *data warehouse* tidak *real time* dan *read only*. Proses *update* pada *data warehouse* dilakukan secara berkala menggunakan *scheduler* sehingga *data* yang disimpan bersifat *historical*. Data yang terdapat pada *data warehouse* bersifat *read only* sehingga proses yang terjadi pada *data warehouse* hanya proses menarik data dan akses data.

d) *Time variant*

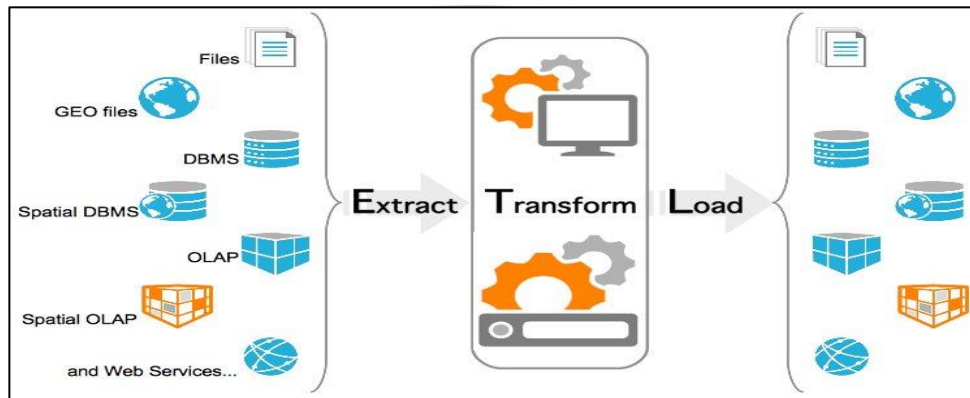


Gambar 2.4 karakteristik data *time variant*

Data warehouse bersifat *time variant* berarti *data warehouse* memiliki dimensi waktu sebagai variabel. Dimensi waktu ini kemudian akan digunakan sebagai pembanding untuk menghasilkan laporan yang diinginkan. Aspek *time variant* pada *data warehouse* memberikan kemampuan untuk menyajikan informasi dalam bentuk *trend*.

2.4 ETL(extract, transform, load)

ETL adalah proses dimana kita melakukan migrasi *data* dari *database* operasional menuju *data warehouse*. *ETL* merupakan proses yang pertama kali dilakukan pada saat pembuatan *data warehouse* atau pada saat *data warehouse* akan di-update. (Golfarelli dan Rizzi 2009, H.15)



Gambar 2.5 ETL

(Sumber: <http://www.dbbest.com>, 2016)

Proses *ETL* meliputi

a) Tahap ekstraksi data(*data extraction*)

Ekstraksi data adalah proses dimana data diambil atau ditarik dari berbagai *source* baik menggunakan *query* atau aplikasi *ETL*. Tujuan utama ekstraksi data adalah menarik data-data yang diperlukan dari berbagai sumber dengan sumber daya seminimal mungkin.

b) Tahap transformasi data(*data transformation*)

Transformasi data adalah tahap dimana data yang ditarik dirubah sesuai dengan standar dan struktur data yang sesuai. Proses ini dilakukan tanpa merubah data pada sumber aslinya.

Hal-hal yang umum dilakukan pada saat transformasi data :

1. *Cleansing*

Menghilangkan atau mengganti unsur tertentu pada data contoh mengganti kata pria menjadi p.

2. *Filtering*

Memilah data yang akan digunakan.

3. *Splitting*

Memisahkan satu kolom menjadi berbagai kolom.

4. *Joining*

Menggabungkan data dari berbagai sumber.

5. Menerapkan *business rules*

Menerapkan standar yang ditentukan atau menghitung sebuah data berdasar rumus yang sudah ditentukan.

c) Pengisian data(*loading*)

Tahap ini adalah tahap pengisian data yang sudah ditransformasi ke dalam wadah(*data mart, database, data warehouse*).

2.5 Pentaho Business Analytics

Pentaho business analytics adalah sekumpulan *open source business intelligence product* yang menyediakan fungsi *ETL, data integration, OLAP services, reporting, dashboarding, dan data mining*.

Kelebihan *pentaho business analytics* :

- a) *Pentaho* bersifat *open source*.
- b) Mempunyai komunitas yang memberikan *support* ke pengguna.
- c) Dapat beroperasi dengan baik di berbagai macam sistem operasi.

2.5.1 Pentaho Data Integration

Pentaho Data Intergration adalah salah satu alat intergrasi data yang digunakan secara luas untuk semua jenis manipulasi data seperti migrasi data antar aplikasi atau *database*, mengekspor data dari *database* ke *flat file, data cleansing*, dan banyak lagi (Pulvirenti & Carina, 2010).

2.5.2 Pentaho Schema Workbench

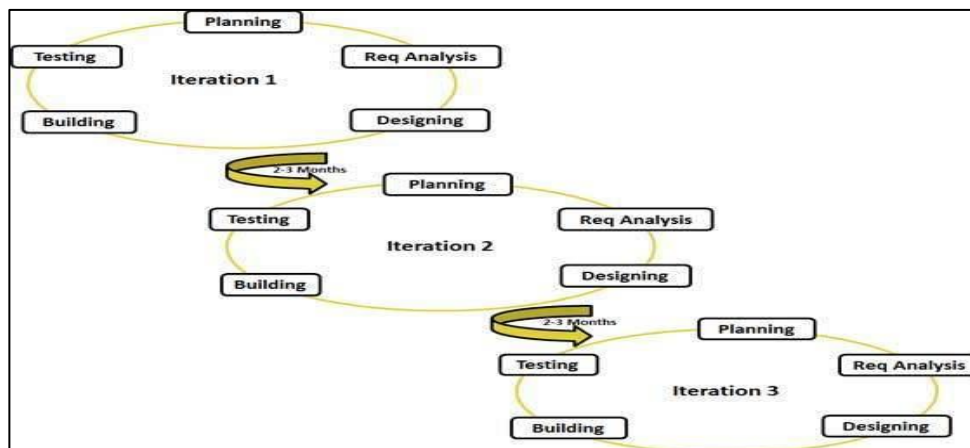
Pentaho schema workbench adalah sebuah aplikasi yang digunakan untuk membuat *OLAP cube* dengan format *xml*. *Pentaho schema workbench* menyediakan fitur *text editor* dan *GUI* yang dapat digunakan oleh *user* untuk membuat *cube*. *Cube* yang dibuat menggunakan *pentaho schema workbench* akan ditampilkan pada *pentaho bi-server*.

2.5.3 Pentaho BI-Server

Pentaho Bi-server adalah sebuah aplikasi dari *pentaho* yang berjalan pada *J2EE webserver*s. *Pentaho bi-server* berfungsi sebagai tempat untuk menampilkan *cube*, *dashboard*, dan *report*.

2.6 SDLC Agile

SDLC agile berdasar pada sebuah gagasan pembuatan program yang *iterative* dan *incremental*. Dalam *SDLC agile* pembangunan program proses pembuatan dibagi menjadi beberapa bagian yang disebut *increment* atau *iteration*.



gambar 2.6 *sdlc agile*

(sumber : https://www.tutorialspoint.com/sdlc/sdlc_agile_model.htm)

2.7 Critical Success Factor for Data Warehouse Engineering

Berikut ini adalah *critical success factor* dalam membangun *data warehouse*.

a) *Sponsorship and Involvement*

Dalam pembuatan *data warehouse* semua pihak yang akan menggunakan *data warehouse* harus ambil bagian dalam proses pembuatannya selain itu eksekutif perusahaan dan manager harus mensponsori pembuatan *data warehouse*. Pihak yang dimaksud disini adalah pihak manajemen, jajaran eksekutif, dan semua *potensial user*. Pada umumnya tanpa sponsor dari kedua belah pihak dan keterlibatan berbagai pihak pembuatan *data warehouse* gagal.

b) *Business Requirement*

Membangun *data warehouse* tanpa mengetahui *business requirement* sebuah perusahaan merupakan sebuah langkah yang salah. *Business requirement* dapat didapat dengan menganalisa *enterprise strategic planning* dan *performance measure*. karena hal ini merupakan dasar dari sebuah arsitektur perusahaan yang kemudian akan menjadi arsitektur dan desain dari *data warehouse*.

c) *Enterprise Architecture*

Dalam pembuatan *data warehouse* *enterprise architecture* merupakan hal yang sangat penting untuk dianalisa. *Enterprise architecture* menunjukan komponen-komponen, relasi antar bagian, *SOP*, dan hal-hal lainnya dalam perusahaan. Dengan menganalisa *enterprise architecture* dapat diketahui dibagian mana saja yang akan diimplementasi *data warehouse*. *Enterprise architecture* dapat diketahui dengan cara menganalisa dan menghubungkan *enterprise business architecture*(EBA) dengan *enterprise information architecture*(EIA), *enterprise service component architecture*(ETA), dan *enterprise technical architecture*(ETA).

d) *Data warehouse Architecture and Design*

Salah satu faktor penentu berhasil atau tidaknya sebuah *data warehouse* adalah *data warehouse architecture and design*. Desain dan arsitektur sebuah *data warehouse* harus dapat mencerminkan *performance measurement* dan *business requirement* dari sebuah perusahaan. Dalam membuat desain dan arsitektur *data warehouse* hal-hal yang harus diperhatikan adalah *data warehouse data model, component, structure*, dan *metadata* harus berdasarkan kebutuhan informasi internal bukan kebutuhan teknologi tertentu.

e) *Data warehouse Technology*

Setelah menentukan arsitektur *data warehouse* langkah selanjutnya adalah menentukan dan melakukan implementasi teknologi yang mendukung *data warehouse*. Jika tidak maka besar kemungkinan *data warehouse* tidak memenuhi kebutuhan perusahaan. Sebuah *data warehouse* yang dirancang hanya untuk teknologi tertentu secara spesifik dapat berakibat pada penggantian teknologi pendukung. Seiring dengan berkembangnya teknologi makin banyak teknologi pendukung *data warehouse*. Secara umum teknologi tersebut dapat dikelompokkan menjadi *user interface, warehouse engine, hardware platform, system software*, dan *security*.

f) *Information Quality*

Salah satu faktor yang menentukan sukses atau tidaknya sebuah *data warehouse* adalah kualitas informasi yang terdapat dalam *data warehouse*. Informasi yang terdapat di dalam *data warehouse* haruslah *accurate, relevant, complete, concise*, disajikan dengan jelas dan mudah dipahami. Jika informasi didalam *data warehouse* memiliki kualitas yang buruk maka seiring dengan waktu *data warehouse* akan tidak

digunakan dan *data warehouse* tersebut gagal. Lebih parahnya jika kualitas informasi didalam *data warehouse* buruk dan perusahaan tidak menyadarinya serta *data warehouse* tersebut tetap digunakan maka ada kemungkinan perusahaan tersebut dapat jatuh.

g) *Development Environment*

Dalam pembuatan *data warehouse* lingkungan kerja juga mempengaruhi sukses atau tidaknya pembangunan sebuah *data warehouse*. Agar sebuah perusahaan secara konsisten melakukan *develop, design, dan implementasi data warehouse* perusahaan tersebut harus memiliki *development environtment* yang menerapkan *best practice, dan technique*. Faktor-faktor yang mempengaruhi *development environtment* adalah *project teams, methodology, dan tools*.

2.8 Studi Literatur

Pada penelitian ini penulis banyak belajar dari jurnal-jurnal. Dalam studi literatur ini penulis akan membahas secara singkat salah satu jurnal yang dibaca penulis dengan judul “PEMANFAATAN DATA WAREHOUSE SEBAGAI SARANA PENUNJANG PENYUSUNAN BORANG AKREDITASI STANDAR 3 PADA FAKULTAS TEKNOLOGI INFORMASI UNIVERSITAS BUDI LUHUR”. Jurnal ini disusun oleh Windarto untuk menyelesaikan program pasca sarjana magister komputer pada Universitas Budi Luhur.

Identifikasi masalah pada penelitian ini adalah :

- a) Belum adanya *data warehouse* yang khusus dikembangkan untuk keperluan pengisian borang, sehingga pihak fakultas masih mengalami kesulitan dalam melakukan pengolahan data untuk mendukung penyusunan boring standar 3 mengenai data kemahasiswaan dan lulusan.

- b) Fakultas saat ini masih sulit mendapatkan informasi yang berkaitan dengan trend kualitas mahasiswa per jurusan dari periode ke periode yang dapat digunakan sebagai dasar dalam mengevaluasi kinerja dari setiap jurusan nya.

Tujuan dari penelitian ini secara garis besar adalah mendesain model *data warehouse* yang sesuai untuk diterapkan pada Fakultas Teknologi Informasi Universitas Budi Luhur serta membangun *data warehouse* sesuai desain sehingga mampu menyajikan informasi mengenai data kemahasiswaan dan lulusan yang dapat membantu pihak fakultas dalam pengisian boring akreditasi standar 3.

Dibawah ini adalah *screen shot* prototipe *data warehouse* yang dibangun oleh Windarto.

		Measures		
periode_dim	prodi_dim	jum_mhs	jum_botmhs	jum_lulusan
20072008	All periode_dim.periode_sks	1,744	5,793	1,811
	All prodi_dim.prodi_sks			
	Komputerisasi Akuntansi	0	7	0
	Sistem Informasi	884	3,066	1,016
	Sistem Komputer	33	177	56
	Teknik Informatika	827	2,543	739
	All prodi_dim.prodi_sks	779	2,891	915
	Komputerisasi Akuntansi	0	6	0
20082009	Sistem Informasi	442	1,570	502
	Sistem Komputer	15	100	33
	Teknik Informatika	322	1,215	380
	All prodi_dim.prodi_sks	965	2,902	896
	Komputerisasi Akuntansi	0	1	0
	Sistem Informasi	442	1,496	514
	Sistem Komputer	18	77	23
	Teknik Informatika	505	1,328	359

gambar 2.7 informasi jumlah mahasiswa baru, total mahasiswa, dan lulusan per tahun ajaran dan program studi

		Measures		
prodi_dim	status_mhs_dim	jum_mhs	jum_totmhs	jum_lulusan
All prodi_dim.prodi_sks	All status_mhs_dim.statmhs_sks	1,744	5,793	1,811
	Reguler	1,593	5,641	1,775
	Transfer	151	152	36
Komputerisasi Akuntansi	All status_mhs_dim.statmhs_sks	0	7	0
	Reguler	0	7	0
	Transfer			
Sistem Informasi	All status_mhs_dim.statmhs_sks	884	3,066	1,016
	Reguler	773	2,953	992
	Transfer	111	113	24
Sistem Komputer	All status_mhs_dim.statmhs_sks	33	177	56
	Reguler	27	167	56
	Transfer	6	10	0
Teknik Informatika	All status_mhs_dim.statmhs_sks	827	2,543	739
	Reguler	793	2,514	727
	Transfer	34	29	12

gambar 2.8 informasi jumlah mahasiswa baru, total mahasiswa per program studi, dan status mahasiswa

Pada penelitian ini model data pada gambar 2.8 ini memberikan inspirasi bagi penulis dalam menentukan struktur *hierarchy* pada proses *multidimensional modelling*.

2.9 Count Validation

Testing dilakukan dengan cara menghitung data awal pada *source* lalu dibandingkan dengan data pada *database*.