

# DATA WAREHOUSE

NENI PURWATI



# DATA WAREHOUSE

Hak cipta pada penulis  
Hak penerbitan pada penerbit  
Tidak boleh diproduksi sebagian atau seluruhnya dalam bentuk apapun  
Tanpa izin tertulis dari pengarang dan/atau penerbit

**Kutipan Pasal 72 :**  
Sanksi pelanggaran Undang-undang Hak Cipta (UU No. 10 Tahun 2012)

1. Barang siapa dengan sengaja dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam Pasal 2 ayat (1) atau Pasal (49) ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling singkat 1 (satu) bulan dan/atau denda paling sedikit Rp. 1. 000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan atau denda paling banyak Rp. 5. 000.000.000,00 (lima miliar rupiah)
2. Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu Ciptaan atau hasil barang hasil pelanggaran Hak Cipta atau Hak Terkait sebagaimana dimaksud ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp. 500.000.000,00 (lima ratus juta rupiah)

NENI PURWATI

D A T A  
WAREHOUSE



Perpustakaan Nasional RI:  
Katalog Dalam Terbitan (KDT)

**Data Warehouse**

**Penulis**  
Neni Purwati

Rancang Sampul & Penata Isi  
Aura Creative

**ISBN:978-602-51690-2-1**

Cetakan Maret 2018  
xvi + 106 hlm. ; 15.5 x 23 cm

**Penerbit**  
Darmajaya ( DJ ) Press

Alamat :  
Kampus IBI DARMAJAYA  
Jl. Zainal Abidin Pagar Alam No 93,  
Bandar Lampung 35142, INDONESIA

Hak Cipta dilindungi Undang-Undang  
All Rigths Reserved.

Dilarang mengutip atau memperbanyak sebagian  
atau seluruh isi buku ini tanpa izin tertulis dari penerbit

## PRAKATA

**S**egala Puji dan Syukur kepada Allah SWT atas Ridho dan perkenan dari-Nya penyusunan buku ajar ini dapat terlaksana. Buku ajar ini disusun sebagai syarat mengikuti hibah buku ajar Institusi yang dibiayai dari IIB Darmajaya.

Terima kasih kepada IIB Darmajaya, Bapak dan Ibu pihak Manajemen Perguruan Tinggi IIB Darmajaya dan Pengelola LP4M yang telah memberikan dukungan dan masukan yang sangat berharga hingga dapat terselenggara dan terselesaikannya buku ajar ini. Terima kasih pula kepada pihak penyandang dana Hibah yang telah memberikan dukungan dana dalam hal ini.

Semoga budi baik yang telah diberikan, mendapatkan imbalan yang setimpal dari Allah SWT dan semoga buku ajar ini dapat menambah pengetahuan dan manfaat kepada mahasiswa IIB Darmajaya khususnya, dan bagi pengembangan ilmu sistem informasi pada umumnya.

Saya sadari Buku Ajar ini masih jauh dari sempurna, oleh karena itu kritik dan saran yang sangat berguna bagi pengembangan Buku Ajar ini sangat saya harapkan, agar dikemudian hari menjadi lebih baik lagi.

Bandar Lampung, Januari 2017

Penulis

## DAFTAR SINGKATAN

---

1. ACL (Access Control List)
2. Basisdata (Database)
3. BI (Business Intelligence)
4. CRM (Customer Relationship Management)
5. DBMS (Database Management System)
6. DNS (Daftar Nilai Semester)
7. DTS (Data Transformation Service)
8. DSS (Decision Support Systems)
9. DW (Data Warehouse)
10. EIS (Executive Information System)
11. ELT (Extract, Load , Transform)
12. ERP (Enterprise Resource Planning)
13. ETL (Extract, Transform, Load)
14. GDSS (Group Decision Support Systems)
15. HOLAP (Hibrid Online Analytical Processing)
16. HRM (Human Resource Management)
17. IPK (Indeks Prestasi Kumulatif)
18. IPS (Indeks Prestasi Semester)
19. Lulusan (Alumni)
20. MDX (Multidimensional Expressions)

21. MOLAP (Multidimensional Online Analitycal Processing)
22. ODI (Oracle Data Integration)
23. ODS (Operational Data Store)
24. ODSS (Organizational Decision Support Systems)
25. OLTP (Online Transactional Processing)
26. OLAP (Online Analytical Processing)
27. PDI (Pentaho Data Integration)
28. Penghargaan (Reward)
29. PSW (Pentaho Schema Workbench)
30. ROLAP (Relational Online Analitycal Processing)
31. SDM (Sumber Daya Manusia)
32. SQL (Structured Query Language)
33. SSIS (Microsoft SQL Server Integration Service)

# DAFTAR ISI

---

Halaman Judul .....	i
Prakata .....	v
Daftar Singkatan .....	vi
Daftar Isi .....	viii
Daftar Gambar .....	xii
Daftar Tabel .....	xiv
Flowchart Hubungan Antar Pokok Bahasan .....	xv
BAB I TERMINOLOGI DASAR .....	1
1.1 Online Transaction Processing (OLTP) .....	2
1.2 Online Analytical Processing (OLAP) .....	2
1.3 Data Warehouse .....	3
1.4 Data Mart .....	4
1.5 ETL dan ELT .....	5
1.5.1 ETL (Extract, Transform, Load) .....	5
1.5.2 ELT (Extract, Load , Transform) .....	8
Soal latihan.....	9
BAB II BUSINESS INTELLIGENT (BI).....	10
2.1 Definisi Business Intelligence .....	10
2.2 Sejarah BI .....	12

2.3 Manfaat Business Intelligence System Bagi Organisasi .....	13
2.4 Tujuan Business Intelligence System .....	14
2.5 Pendekatan Implementasi Business Intelligence System ...	15
2.6 Arsitektur Business Intelligence.....	17
2.7 DW dan BI .....	23
2.8 Software BI .....	24
2.9 Sistem Kerja DW .....	25
Soal latihan .....	25
<b>BAB III PENGENALAN DATA WAREHOUSE .....</b>	<b>26</b>
3.1 Data Warehouse .....	26
3.1.1 Konsep Dasar Data Warehouse .....	30
3.2 Tujuan Data Warehouse .....	31
3.3 Mekanisme Data pada Data Warehouse .....	32
3.4 Anatomii Data Warehouse .....	33
3.5 Arsitektur Data Warehouse .....	36
3.6 Proses dalam Data Warehouse .....	41
Soal latihan .....	42
<b>BAB IV INFRASTRUKTUR DATA WAREHOUSE .....</b>	<b>43</b>
4.1 Infrastruktur operasional .....	44
4.2 Infrastruktur fisik .....	45
4.2.1 Hardware dan Operating System .....	46
4.2.1.1 Platform .....	48
4.2.1.2 Perangkat Keras Server.....	53
4.2.2 Database Software .....	57
4.2.2.1 Parallel Processing Options.....	57
4.2.3 Collection Of Tools Data Warehouse .....	61
4.2.3.1 Architecture Tools.....	62
4.2.4 Network Software .....	65
Soal latihan .....	66

BAB V DATA MULTIDIMENSIONAL .....	67
5.1 Definisi data .....	67
5.2 Definisi Database .....	68
5.3 Metadata .....	68
5.4 Data Multidimensi .....	70
5.5 Pemodelan Data Multidimensi .....	72
5.5.1 Star Schema .....	72
5.5.2 Snowflakes Schema .....	74
5.6 Cube (Kubus) dan Speadsheet .....	75
5.6.1 Cube (Kubus) .....	75
5.6.1.1 Bentuk Cube (Kubus) pada Data Warehouse	75
5.6.1.2 Model Pemrosesan Data dari Cube .....	77
5.6.1.3 Contoh Data Cube .....	80
5.6.2 Spreadsheet .....	82
Soal latihan .....	83
BAB VI DESAIN DATA WAREHOUSE .....	84
6.1 Desain Data Warehouse .....	84
6.2 Aspek yang perlu dipertimbangkan .....	85
6.2.1 Saat Desain .....	85
6.2.2 Saat Implementasi .....	86
6.2.3 Saat Testing .....	86
Soal latihan .....	87
BAB VII MEMBANGUN DATA WAREHOUSE .....	88
7.1 Empat Langkah Membangun Data Warehouse .....	88
7.2 Enam Tahap Membangun Data Warehouse .....	90
7.3 Metodologi Sembilan Tahap (Nine-step methodology) .....	93
Soal latihan .....	94
BAB VIII IMPLEMENTASI DATA WAREHOUSE .....	95
8.1 Implementasi .....	95
8.2 Tampilan Program .....	98

8.2.1 Tampilan Menu Program .....	98
8.2.2 Tampilan Sub Menu Input Dean List .....	98
8.2.3 Tampilan Sub Menu Laporan .....	99
Daftar Pustaka.....	101

# DAFTAR GAMBAR

---

Gambar 1.1 Data Mart.....	5
Gambar 1.2 ETL vs ELT .....	8
Gambar 2.1 Arsitektur Business Intelligence.....	20
Gambar 2.2 Sistem kerja DW .....	25
Gambar 3.1 Contoh data berorientasi subjek .....	27
Gambar 3.2 Contoh data terintegrasi .....	28
Gambar 3.3 Time Variancy .....	28
Gambar 3.4 Nonvolatility .....	29
Gambar 3.5 Bentuk Data Warehouse Fungsional .....	34
Gambar 3.6 Bentuk Data Warehouse Terpusat .....	35
Gambar 3.7 Bentuk Data Warehouse Terdistribusi .....	36
Gambar 3.8 Arsitektur Sistem Data Warehouse .....	37
Gambar 3.9 Arsitektur Data warehouse .....	38
Gambar 4.1 Infrastruktur Pendukung Arsitektur .....	44
Gambar 4.2 Infrastruktur Fisik .....	46
Gambar 4.3 Pilihan Platforms pada data warehouse .....	50
Gambar 4.4 platform untuk Data Acquisition.....	50
Gambar 4.5 Data Movement Options .....	51
Gambar 4.6 Client / Arsitektur Server untuk Data Warehouse .....	52
Gambar 4.7 Pilihan platform pada Data Warehouse .....	53
Gambar 4.8 Server Hardware Option : SMP .....	54

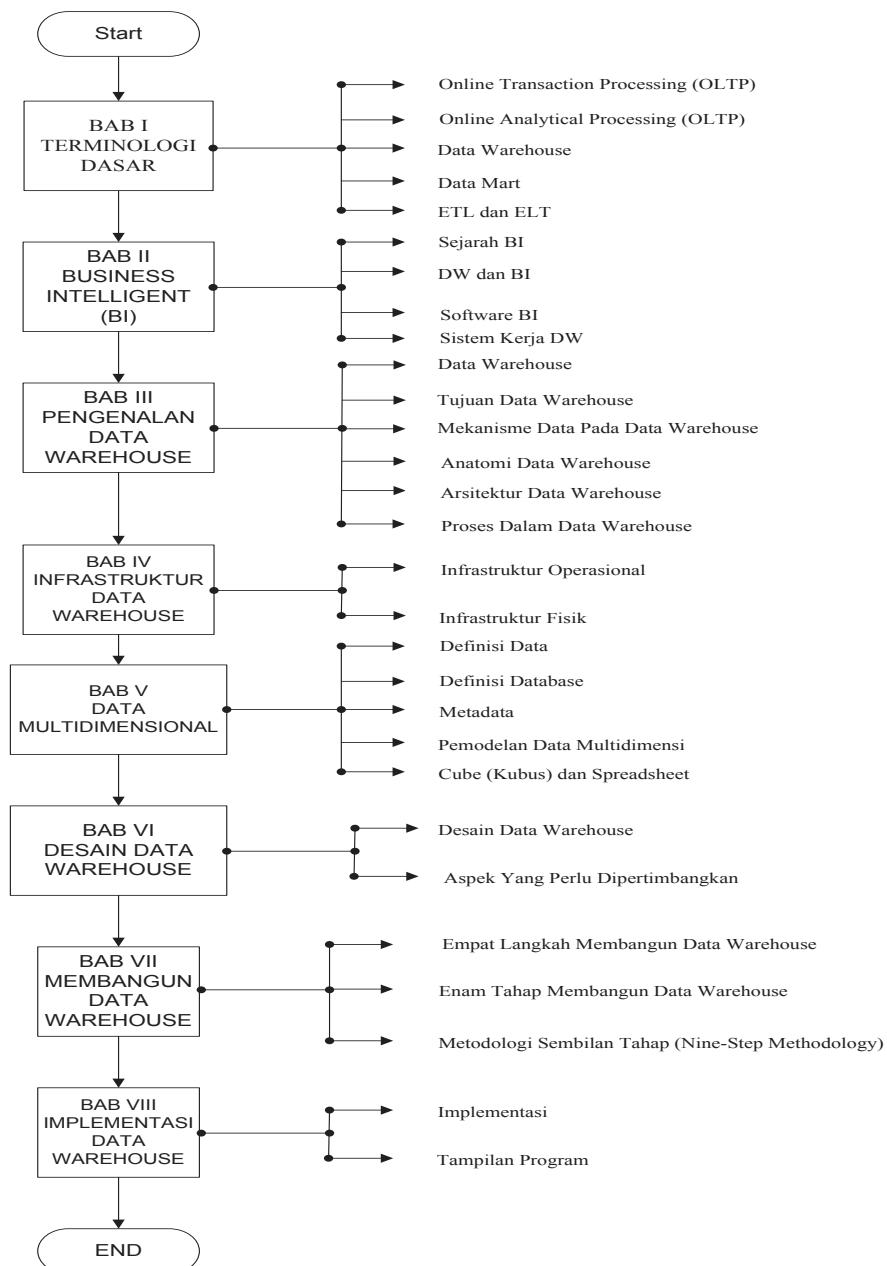
Gambar 4.9 Server Hardware Option : Cluster .....	55
Gambar 4.10 Server Hardware Option : MPP .....	56
Gambar 4.11 Server Hardware Option : NUMA .....	57
Gambar 4.12 Intraquery Paralelization by DBMS .....	60
Gambar 4.13 Tools for Data Warehouse .....	61
Gambar 5.1 Contoh Metadata .....	69
Gambar 5.2 Profesi yang membutuhkan Metadata .....	70
Gambar 5.3 Tabel Relasional .....	71
Gambar 5.4 Data Multidimensi .....	72
Gambar 5.5 Star schema pada data warehouse .....	73
Gambar 5.6 Snowflakes Schema .....	75
Gambar 5.7 Cube .....	76
Gambar 5.8 Model MOLAP .....	78
Gambar 5.9 Model ROLAP .....	79
Gambar 5.10 Model HOLAP .....	80
Gambar 5.11 Contoh Data Cube .....	81
Gambar 5.12 Contoh Spreadsheet .....	83
Gambar 7.1 Alur pembuatan dimensi .....	89
Gambar 7.2 Alur pembuatan Cube .....	89
Gambar 8.1 Star Schema .....	97
Gambar 8.2 Tampilan Program Import IPS .....	98
Gambar 8.3 Tampilan Data Konsistensi IPS .....	99
Gambar 8.4 Tampilan Laporan Dean List .....	100

## **DAFTAR TABEL**

---

Tabel 3.1 Perbedaan OLTP dan data warehouse .....	30
Tabel 3.2 Aliran data dalam mekanisme data warehouse .....	33

# FLOWCHART HUBUNGAN ANTAR POKOK BAHASAN





# BAB I

## TERMINOLOGI DASAR

---

**D**unia korporasi saat ini menghadapi permasalahan yang hampir sama, yaitu membengkaknya data akibat keberhasilan implementasi berbagai sistem komputer. Otomatisasi berjalan dengan baik, namun berbagai laporan yang ingin dihasilkan sebagai output dari sistem tersebut sebagian besar tidak tercapai.

Permasalahannya adalah 85 persen waktu yang diperlukan adalah mempersiapkan data (data preparation). Database yang dirancang untuk sistem entri (OLTP) ternyata sangat buruk digunakan untuk menghasilkan berbagai laporan analisa. Berbagai solusi telah ditawarkan oleh para praktisi IT, dan pendekatan dengan membangun suatu data warehouse adalah yang terbaik. Untuk membangun data warehouse yang baik diperlukan tool yang mampu mengambil, mengolah dan menyimpan data pada berbagai format dengan baik. Tool ini biasa disebut dengan ETL (Extract, Transform and Load).

## 1.1 Online Transaction Processing (OLTP)

Sistem operasional bisnis biasanya berfokus pada mencatat transaksi-transaksi yang terjadi pada berbagai titik operasi bisnis, oleh karena itu mereka dikarakteristikkan sebagai sistem OLTP (*Online Transactional Processing*). Sebuah database OLTP biasanya mengandung data-data yang spesifik terhadap suatu proses bisnis seperti penjualan, produksi, finansial, dan lain-lain.

Beban kerja sistem informasi OLTP difokuskan pada memasukkan data baru, melakukan perubahan dan penghapusan data secara *real time*. Sistem OLTP sangat *mission critical* artinya tidak boleh ada gangguan dalam sistem ini atau operasional tidak bisa berjalan dengan baik.

Beberapa aplikasi yang bisa dikategorikan sebagai OLTP adalah:

- **ERP(Enterprise Resource Planning)**

Beberapa contoh produk ERP : SAP, Compiere/Adempiere, Microsoft Dynamics, dan lain-lain.

- **HRM (Human Resource Management)**

Beberapa contoh produk HRM : OrangeHRM, PeopleSoft, SimpleHRM, dan lain-lain.

- **CRM (Customer Relationship Management)**

Beberapa contoh produk CRM : Microsoft CRM, SugarCRM, Siebel CRM, dan lain-lain.

## 1.2 Online Analytical Processing (OLAP)

*Online Analytical Processing* atau OLAP adalah sistem yang dirancang khusus untuk menghasilkan laporan analisa yang fleksibel, kompleks dan dapat dihasilkan dalam waktu yang cepat.

Database OLAP sangat berbeda dengan OLTP dari sisi beban kerja dimana OLAP dirancang dan difokuskan pada kecepatan pembacaan data sedangkan OLTP pada kecepatan perekaman dan perubahan data.

Umumnya database OLAP dihasilkan melalui suatu proses *batch* dan biasanya dilakukan dalam periode tertentu (tiap tengah malam, tiap minggu, dsbnya).

Database OLAP biasanya juga telah merupakan database yang sudah diperkaya dari berbagai sumber data OLTP, dan biasanya merupakan suatu *data warehouse*.

OLAP mengandung 2 tipe dasar yaitu *measure* dan *dimension*.

1. *Measure* adalah data bilangan yang terukur, misalkan kuantitas (*quantity*), harga (*price*), nilai rata-rata (*average*) dari kelompok nilai tertentu, jumlah (*sum*) dari kelompok nilai tertentu, dan lain sebagainya.
2. *Dimension* mengacu pada kategori yang digunakan untuk mengatur *measure*. Biasanya data dikelompokkan dalam bentuk bertingkat (*level*). Dimensi yang umumnya hampir selalu ada adalah dimensi waktu (*time dimension*). Pada dimensi ini hierarki yang disusun biasanya tahun, kuartal, bulan dan hari. Bisa juga jika diperlukan ada minggu di antara bulan dan hari.

Hampir semua *Database Management System* (DBMS) yang umum seperti Oracle Database, Microsoft SQL Server, IBM DB2, PostgreSQL, MySQL dan yang lainnya dapat digunakan untuk database OLTP maupun OLAP. Yang membedakan antara database OLTP dan OLAP adalah pada skema tabel yang dibentuk. Skema tabel database OLTP umumnya berbentuk *normalization*, sedangkan database OLAP menerapkan skema *star* atau *snowflake*.

Pada OLTP bahasa query yang umum digunakan adalah *Structured Query Language* (SQL), sedangkan pada OLAP bahasa query yang digunakan adalah *Multidimensional Expressions* (MDX).

### 1.3 Data Warehouse

**Data Warehouse (DW)** adalah suatu database khusus yang digunakan sebagai “gudang data” atau data yang telah terkonsolidasi

dari sumber-sumber data berbagai sistem informasi yang ada pada suatu organisasi / perusahaan.

Misalkan jika kita memiliki sistem pembelian (*purchasing*), sumber daya manusia (*human resources*), penjualan (*sales*), inventori, dan lain-lain. Maka data warehouse adalah hasil konsolidasi atau penggabungan data-data dari berbagai sistem OLTP tersebut dalam satu kesatuan sehingga kaitan antara satu dengan yang lainnya menjadi jelas dan kuat.

Dengan adanya konsolidasi tersebut, diharapkan banyak informasi berharga yang bisa didapatkan sehingga menciptakan nilai strategik yang tinggi dalam kaitan pengambilan keputusan.

Struktur data warehouse tentunya perlu dioptimalkan untuk berbagai kepentingan sistem pelaporan terutama yang bersifat analisa menyeluruh. Dan dengan kaitan perkembangan OLAP yang semakin penting, data warehouse saat ini lebih banyak mengadopsi pemodelan yang cocok untuk dikonsumsi sistem OLAP dengan baik.

#### **1.4 Data Mart**

Setelah data warehouse terbentuk, informasi yang bisa diambil tentunya jauh lebih luas dan lengkap. Pada data warehouse kemungkinan besar terjadi proses pembersihan data (*data cleansing*) dan pengayaan data (*content enrichment*).

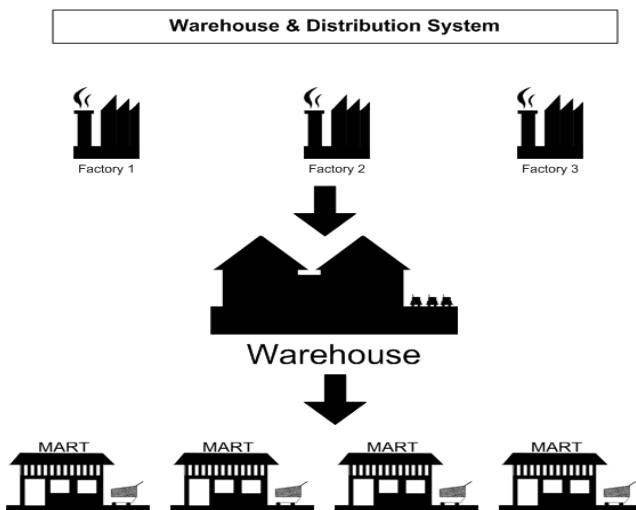
Sebagai contoh, pada sistem penjualan kita hanya mengetahui siapa saja tenaga penjual (*sales person*) dengan kinerja yang baik. Tetapi kita tidak dapat mengetahui latar belakang kenapa beberapa sales performanya baik dan yang lainnya tidak. Dan dengan integrasi data dari sistem SDM yang lebih lengkap profilnya, kita dapat memetakan kinerja dengan latar belakang pendidikan atau lebih luasnya ke data demografis dan hasil uji saringan masuk.

Dari data warehouse ini sering dibutuhkan subset data / informasi yang hanya dibutuhkan oleh divisi:divisi tertentu. Dan kadang jauh lebih baik apabila secara fisik database ini terpisah sendiri untuk divisi tersebut.

Subset dari data warehouse ini disebut dengan istilah **Data Mart**.

Ilustrasi atau analogi akan hubungan sumber data (produsen / pabrik data), data warehouse dan data mart sangat mirip akan sistem distribusi seperti terlihat pada gambar di bawah ini.

Gambar 1.1 Data Mart (Multimatics, 2015)



## 1.5 ETL dan ELT

### 1.5.1 ETL (Extract, Transform, Load)

ETL adalah sekumpulan proses untuk mengambil dan memproses data dari satu atau banyak sumber menjadi sumber baru, misalnya mengolah database OLTP menjadi database OLAP. Sumber data yang diolah ETL bisa dari berbagai sumber data, tidak hanya dari database OLTP saja, tetapi bisa juga dari website, file teks, spreadsheet, database, email, dan lain sebagainya.

ETL terdiri dari 3 bagian utama, setiap bagian memiliki fungsi sesuai namanya.

1. **Extract**, semua proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi

proses-proses selanjutnya.

Contoh :

- Membaca file Microsoft Excel
- Mengambil data dari database
- Mengambil data kurs mata uang dari situs sebuah bank
- Dan lain-lain

2. **Transform**, bagian ini mengacu pada fungsi apa saja yang berfungsi untuk mengubah data yang masuk menjadi data yang dikehendaki.

Fungsi-fungsi dapat berupa :

- Pemindahan data
- Validasi data sesuai aturan yang ditetapkan
- Modifikasi isi, tipe atau struktur data
- Integrasi atau penggabungan data dari sumber-sumber lain
- Perhitungan, dan lain-lain

Contoh :

- Mengubah tipe data dari String menjadi Date
- Memeriksa apakah data nomor kartu kredit sesuai format kartu kredit yang baku atau tidak
- Melakukan lookup untuk mendapatkan nilai tertentu berdasarkan kunci
- dan lain-lain

3. **Load**, semua proses yang diperlukan untuk mengisi data ke target.

Contoh :

- Hasil dari proses sebelumnya disimpan ke dalam file Microsoft Excel

- Hasil dari proses sebelumnya disimpan ke dalam database OLAP
- dan lain-lain.

Aplikasi ETL sangat membantu sekali dalam merancang dan membentuk suatu *data warehouse*.

Beberapa kemampuan yang harus dimiliki oleh ETL adalah sebagai berikut :

- Dapat membaca dan mengirim data dari dan ke berbagai sumber seperti file teks, file Excel, table-table database relasional, web services, dan lain sebagainya.
- Mampu melakukan manipulasi ( transformasi ) data dari yang sederhana sampai ke tahap rumit. Beberapa pengolahan data yang harus bisa dilakukan seperti melakukan normalisasi dan denormalisasi data, memecah kolom, melakukan perhitungan yang terkait tanggal dan jam, dan lain-lain.
- Script dapat dirancang dan diimplementasikan dengan tingkat produktivitas yang tinggi.
- Menghasilkan informasi meta data pada setiap perjalanan transformasi.
- Memiliki audit log yang baik.
- Dapat ditingkatkan performanya dengan scale up dan scale out.

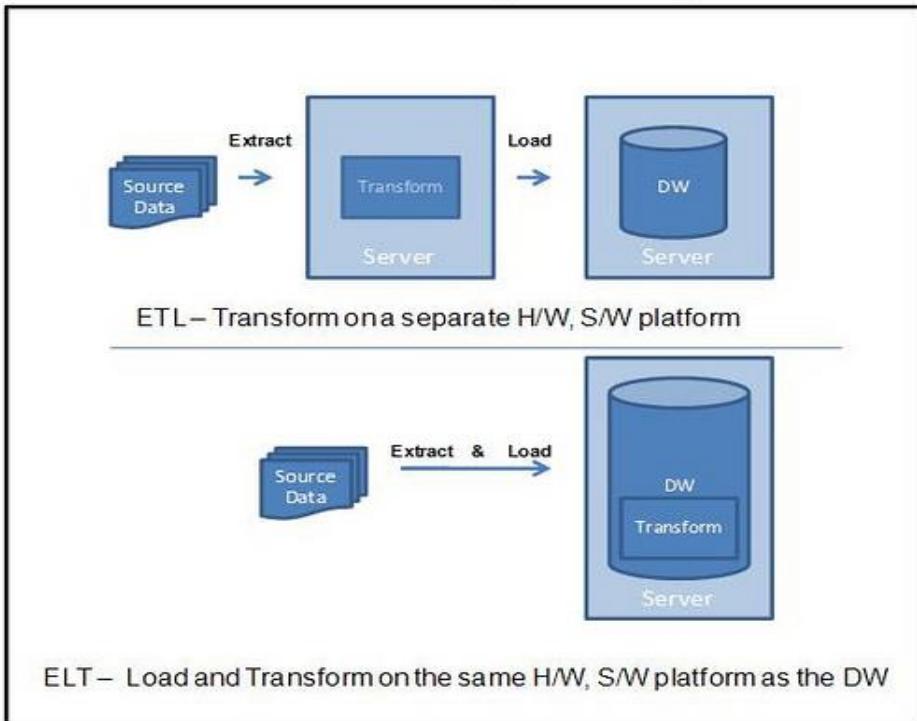
Walaupun membangun solusi ETL dengan programming biasa seperti *coding* dengan bahasa Java atau .NET sangat dimungkinkan, namun tidak disarankan karena produktivitasnya yang cukup rendah dan berakibat ke pemborosan waktu dan biaya.

Berbagai utilitas ETL telah tersedia di pasaran dan menunjukkan kinerja yang baik contohnya adalah Informatica, Data Transformation Service (DTS), Microsoft SQL Server Integration Service (SSIS), Pentaho Data Integration (Kettle), dan lain sebagainya.

### 1.5.2 ELT (Extract, Load , Transform)

Pada proses ELT transformasi terjadi justru pada target database dan bahkan bisa terjadi pada sumber data. Karena proses transformasi terjadi pada target database, biasanya perangkat ELT dikhususkan untuk target database tertentu.

#### **ETL vs. ELT as a Data Integration Approach**



Gambar 1.2 ETL vs ELT (Mulyana JRP, 2014)

Contoh beberapa perangkat ETL/ELT yang umum digunakan adalah sebagai berikut :

1. Pentaho Data Integration (Kettle)
2. Talend Open Studio for Data Integration
3. Microsoft SQL Server Integration Services (SSIS)
4. Oracle Data Integration (ODI)

**Soal latihan :**

1. Jelaskan apa yang dimaksud dengan *mission critical* ?
2. Jelaskan apa yang dimaksud dengan OLTP dan OLAP ?
3. Jelaskan perbedaan OLTP dengan OLAP ?
4. Jelaskan yang dimaksud dengan *Measure* dan *Dimension* ?
5. Jelaskan yang dimaksud dengan *Data Warehouse* dan *Data Mart* ?
6. Jelaskan apa perbedaan ELT dengan ETL ?

# BAB II

## BUSINESS INTELLIGENT (BI)

**P**ada Bab I telah dibahas pengertian dan latar belakang dikembangkannya Data Warehouse, pada Bab II berikut ini akan diuraikan keterkaitan antara Business Intelligent dengan Data Warehouse.

### 2.1 Definisi Business Intelligence

Beberapa pendapat para pakar tentang definisi Business Intelligence (BI) adalah :

1. Business Intelligence (BI) merupakan seperangkat solusi sistem informasi yang dapat menunun kepada percepatan pengambilan keputusan dalam tingkat akurasi yang tinggi (Thia, 2011). BI dapat didefinisikan sebagai kemampuan perusahaan untuk memahami dan menggunakan informasi dalam rangka meningkatkan kinerja (Curko, 2008). Dapat disimpulkan bahwa BI menjelaskan tentang suatu teknologi mengkonversi data berdasarkan sistem yang berbasiskan data menghasilkan informasi yang tepat waktu untuk meningkatkan kualitas pengambilan keputusan bisnis.

2. Business Intelligence adalah suatu cara untuk mengumpulkan, menyimpan, mengorganisasikan, membentuk ulang, meringkas data serta menyediakan informasi, baik berupa data aktivitas bisnis internal perusahaan, maupun data aktivitas bisnis eksternal perusahaan termasuk aktivitas bisnis para pesaing yang mudah diakses serta dianalisis untuk berbagai kegiatan manajemen (David, 2000). Business Intelligence merupakan sebuah proses untuk melakukan ekstraksi data-data operasional perusahaan dan mengumpulkannya dalam sebuah data warehouse. Selama proses ekstraksi juga dapat dilakukan transformasi dengan menerapkan berbagai formula, agregasi, maupun validasi sehingga didapat data yang sesuai dengan kepentingan analisis bisnis. Selanjutnya data di data warehouse diproses menggunakan berbagai analisis statistik dalam proses data mining, sehingga didapat berbagai kecenderungan atau pattern dari data. Hasil penyederhanaan dan peringkasan ini disajikan kepada end user yang biasanya merupakan pengambil keputusan bisnis. Dengan demikian manajemen dapat mengambil keputusan berdasarkan fakta-fakta aktual, dan tidak hanya mengandalkan intuisi dan pengalaman kuantitatif saja (Imelda, 2013). Business Intelligence menjelaskan tentang suatu konsep dan metode bagaimana untuk meningkatkan kualitas pengambilan keputusan bisnis berdasarkan sistem yang berbasiskan data. BI seringkali dipersamakan sebagaimana briefing books, report and query tools, dan sistem informasi eksekutif. BI merupakan sistem pendukung pengambilan keputusan yang berbasiskan data-data (Power, 2002). Business Intelligence adalah alat analisis yang digunakan untuk mengkonsolidasikan data, menganalisis, menyimpan dan mengakses banyak data untuk membantu dalam pembuatan keputusan, seperti perangkat lunak untuk query database dan pelaporan, alat untuk analisis data multidimensi, dan data mining (Laudon, Kenneth C, Jane P., 2008).
3. Business Intelligence System/Sistem Bisnis Intelijen (SIB) : business intelligence merupakan sistem dan aplikasi yang

berfungsi untuk mengubah data-data dalam suatu perusahaan/organisasi (data operasional, data transaksional, atau data lainnya) ke dalam bentuk pengetahuan. Aplikasi ini melakukan analisis data-data di masa lampau, menganalisisnya dan kemudian menggunakan pengetahuan tersebut untuk mendukung keputusan dan perencanaan organisasi.

4. Steve and Nancy Williams "The Profit Impact of Business Intelligence" (2007) : Business Intelligent as business information and business analyses within the context of key business processes that lead to decisions and actions and that result in improved business performance. In particular, BI means leveraging information assets within key business processes to achieve improved business performance.
5. SIB yang sebagaimana oleh Powers (2002) : "Business Intelligence System menjelaskan tentang suatu konsep dan metode bagaimana untuk meningkatkan kualitas pengambilan keputusan bisnis berdasarkan sistem yang berbasiskan data. Business Intelligent seringkali dipersamakan sebagaimana briefing books, report and query tools, dan sistem informasi eksekutif. Business Intelligent merupakan sistem pendukung pengambilan keputusan yang berbasiskan data-data".

## 2.2 Sejarah BI

Sejarah perkembangan BI dapat dijelaskan sebagai berikut :

- Teori tentang pengambilan keputusan organisasi dikembangkan di Carnegie Institute of Technology (akhir tahun 1950an – awal tahun 1960an).
- Implementasi DSS tersebut dalam bentuk sistem komputer interaktif dilakukan di Massachusetts Institute of Technology (tahun 1960an)
- Konsep DSS menjadi area riset (pertengahan 1970 – 1980an).

- Dikembangkan Executive Information Systems (EIS), Group Decision Support Systems (GDSS), dan Organizational Decision Support Systems (ODSS) untuk *single user* berbasis model (pertengahan 1980)
- Dikembangkan *data warehousing* dan *On-Line Analytical Processing* (OLAP) (awal tahun 1990).
- Dikembangkan aplikasi analitik berbasis web (awal tahun 2000) dikenal dengan Business Intelligence (BI).

### **2.3. Manfaat Business Intelligence System Bagi Organisasi**

Menurut Steadman (2003) Beberapa manfaat yang bisa didapatkan bila suatu organisasi mengimplementasikan SIB antara lain :

1. Pertama meningkatkan nilai data dan informasi organisasi. Melalui pembangunan BI, maka seluruh data dan informasi dapat diintegrasikan sedemikian rupa sehingga menghasilkan dasar pengambilan keputusan yang lengkap. Informasi-informasi yang dulunya tidak dicakupkan sebagai salah satu faktor pengambilan keputusan (terisolasi) dapat dengan mudah dilakukan 'connect and combine' dengan menggunakan BI. Data dan informasi yang dihasilkan pun juga menjadi lebih mudah diakses dan lebih mudah untuk dimengerti (friendly-users infos).
2. Kedua memudahkan pemantauan kinerja organisasi. Dalam mengukur kinerja suatu organisasi seringkali dipergunakan ukuran yang disebut Key Performance Indicator (KPI). KPI tidak melulu diukur dengan satuan uang, namun dapat juga berdasarkan kecepatan pelaksanaan suatu layanan.
3. Ketiga meningkatkan nilai investasi Teknologi Informasi (TI) yang sudah ada. SIB tidak perlu/harus mengubah atau menggantikan sistem informasi yang sudah digunakan sebelumnya. Sebaliknya, SIB hanya menambahkan layanan pada sistem-sistem tersebut sehingga data dan informasi yang sudah ada dapat menghasilkan

informasi yang komprehensif dan memiliki kegunaan yang lebih baik.

4. Keempat menciptakan pegawai yang memiliki akses informasi yang baik (well-informed workers). Dalam melaksanakan pekerjaannya sehari-hari, seluruh level dari suatu organisasi (mulai dari pegawai/bawahan sampai dengan pimpinan) selalu berkaitan dan/atau membutuhkan akses data dan informasi. BI mempermudah seluruh level pegawai dalam mengakses data dan informasi yang diperlukan sehingga membantu membuat suatu keputusan. Jika kondisi seperti ini tercapai, maka misi dan strategi organisasi yang sudah ditetapkan dapat dengan lebih mudah terlaksana serta terpantau tingkat pencapaiannya.
5. Kelima meningkatkan efisiensi biaya. BI dapat meningkatkan efisiensi karena mempermudah seseorang dalam melakukan pekerjaan : hemat waktu dan mudah pemanfaatannya. Waktu yang dibutuhkan untuk mencari data dan mendapatkan informasi yang dibutuhkan menjadi semakin singkat dan cara untuk mendapatkannya pun tidak memerlukan pengetahuan (training) yang rumit.

## 2.4 Tujuan Business Intelligence System

Implementasi Business Intelligence System Sistem SIB yang biasanya digunakan di organisasi. Tujuan utamanya adalah untuk meningkatkan layanan, efisiensi biaya, dan efektifitas kerja yang diembankan kepada organisasi dimaksud. Dalam mengimplementasikan SIB di suatu perusahaan pemerintah, hal utama yang harus diperhatikan adalah bahwa SIB harus mendukung pencapaian visi, misi, dan strategi organisasi dalam mencapai tingkat kinerja organisasi (organization performance) yang diinginkannya.

## 2.5 Pendekatan Implementasi **Business Intelligence System**

Pada tahap pendekatan implementasi SIB, ada beberapa yang harus dilewati untuk mulai dari identifikasi awal penentuan bahan yang diperlukan hingga deployment yang akan dijelaskan di tiap-tiap langkanya. Ada enam stage dalam pendekatan implementasi SIB (Moss dan Atre, 2003) :

### 1. Justification

Menilai kebutuhan bisnis yang menimbulkan proyek rekayasa baru.

### 2. Planning

Mengembangkan rencana strategis dan taktis yang merencanakan bagaimana proyek rekayasa akan tercapai dan disebarluaskan.

### 3. Business Analysis

Menunjukkan analisis rinci mengenai masalah bisnis atau peluang usaha untuk mendapatkan pemahaman yang kuat tentang persyaratan bisnis.

### 4. Design

Menyusun sebuah produk yang memecahkan masalah bisnis atau memungkinkan peluang bisnis.

### 5. Construction

Membangun produk yang seharusnya memberikan pengembalian investasi dalam jangka waktu yang ditentukan.

### 6. Deployment

Melaksanakan atau menjual produk jadi, kemudian mengukur efektivitas untuk menentukan apakah solusi memenuhi atau gagal untuk memenuhi hasil yang diharapkan dari investasi.

Dalam membangun dan mengimplementasikan SIB di suatu organisasi, terdapat 3 (tiga) pendekatan yang bisa digunakan. Masing-masing dari pendekatan tersebut memiliki kelebihan dan kelemahan, dimana pilihan dari strategi tersebut berdasarkan kondisi dan kebutuhan organisasi

yang akan membangun SIB (Von Halle, 2001). Pendekatan tersebut adalah antara lain :

1. Top-down Approach

Kelebihan dari pendekatan ini yaitu : Pembangunan SIB langsung mencakup data seluruh organisasi, Kerangka SIB akan lebih terstruktur, bukan gabungan dari berbagai data mart (data parsial), Penyimpanan data menjadi terpusat dan Kontrol informasi dapat dilakukan secara tersentralisasi.

Adapun kelemahan pendekatan ini yang harus diantisipasi diantaranya: Waktu implementasi lebih lama, Risiko kegagalan relatif tinggi karena kerumitannya dan Membutuhkan biaya yang relatif besar.

2. Bottom-up Approach

Kebalikan dengan pendekatan sebelumnya, dalam pendekatan bottom-up SIB yang akan disusun justru dari tingkat departemental (departement data warehouse) baru kemudian diintegrasikan menjadi data warehouse organisasi secara keseluruhan. Pendekatan ini sangat tepat bagi kebutuhan suatu organisasi yang memprioritaskan pembangunan BI di suatu departemen terlebih dahulu. Kemudian setelah sukses di departemen tersebut akan dilanjutkan de departemen lain.

Kelebihan dari pendekatan ini antara lain: Implementasi lebih mudah untuk dikelola, Risiko kegagalan relatif lebih kecil, Bersifat incremental, dimana data mart yang penting dapat dijadwalkan lebih awal dan Memungkinkan anggota tim proyek untuk belajar dengan baik.

Kelemahan pendekatan ini yang harus diantisipasi yaitu: Tiap data mart merupakan departmental-view, Memungkinkan terjadinya duplikasi data di setiap data mart di masing-masing departemen, Data tidak konsisten dan data sulit direkonsiliasi dan Terdapat banyak interface yang sulit.

### 3. Practical Approach.

Pendekatan ini mengkombinasikan ke-dua pendekatan sebelumnya untuk mendapatkan kelebihannya. Dalam pendekatan ini, pengembangan SIB di suatu organisasi akan dimulai dengan perencanaan dan pendefinisian arsitektur kebutuhan data warehouse organisasi secara keseluruhan (standardisasi). Baru kemudian akan dilakukan serangkaian pembuatan SIB pada tiap departemen yang membutuhkan.

## 2.6 Arsitektur Business Intelligence

Sering kali kita mendengar istilah intelijen bisnis atau lebih terkenal dengan Business Intelligence (BI). Intelijen bisnis sebenarnya merupakan pengolahan data yang khusus untuk informasi bisnis. Intelijen bisnis mempunyai komponen berupa seperangkat teori, metodologi, proses, arsitektur, dan teknologi yang mampu mengubah data mentah menjadi informasi yang bermakna dan berguna untuk tujuan bisnis. Intelijen bisnis dapat menangani sejumlah besar informasi untuk membantu mengidentifikasi dan mengembangkan peluang baru dalam dunia bisnis. Dengan memanfaatkan intelijen bisnis, kita akan bisa mendapatkan peluang baru dan menerapkan strategi yang efektif sehingga mampu menghasilkan keuntungan pasar kompetitif dan stabilitas jangka panjang.

Dengan bantuan teknologi informasi, intelijen bisnis mampu memberikan gambaran tentang sejarah operasi bisnis, kondisi bisnis saat ini dan prediksi operasi bisnis dimasa datang. Fungsi umum dari teknologi intelijen bisnis adalah melaporkan, pengolahan analisis online, data mining, data mining, pengolahan informasi kompleks, bisnis manajemen kinerja, analisis prediktif dan analisis preskriptif.

Istilah intelijen bisnis pertama kali digunakan oleh Hans Peter, seorang peneliti di IBM, dalam sebuah artikel pada tahun 1958. Bisnis intelijen merupakan evolusi dari sistem pendukung keputusan Decision Support Systems (DSS) yang dimulai pada tahun 1960 dan dikembangkan pada

tahun 1980-an . DSS berasal dari model dibantu komputer dibuat untuk membantu pengambilan keputusan atau *Executive Information Systems* (EIS) dan perencanaan. Dari DSS, Data Warehouse, *Executive Information System* (EIS), OLAP dan akhirnya menjadi intelijen bisnis. Suatu metamorfosa yang hebat.

Pada tahun 1989, Howard Dresner, seorang analis Gartner Group, mengusulkan agar istilah intelijen bisnis dipakai untuk menggambarkan konsep dan metode untuk meningkatkan pengambilan keputusan bisnis dengan menggunakan sistem pendukung berbasis fakta. Pada akhir tahun 1990, *Bussiness Intelligence* menjadi berkembang demikian pesat.

Seringkali aplikasi *Bussiness Intelligence* menggunakan data yang dikumpulkan dari data warehouse atau data mart . Sebuah gudang data adalah salinan dari data transaksional yang memfasilitasi dukungan keputusan. Namun, tidak semua gudang data yang digunakan untuk intelijen bisnis, juga tidak semua aplikasi bisnis intelijen memerlukan data warehouse .

Sekarang istilah *Bussiness Intelligence* bisa didefinisikan menjadi seperangkat metodologi, proses, arsitektur, dan teknologi yang mengubah data mentah menjadi informasi yang bermakna dan berguna digunakan untuk memungkinkan wawasan strategis yang lebih efektif, taktis, dan operasional dan pengambilan keputusan. Apabila menggunakan definisi ini, intelijen bisnis juga mencakup teknologi seperti integrasi data, kualitas data, data pergudangan, magister manajemen data, teks dan analisis konten, dan banyak orang lain yang pasar kadang-kadang benjolan ke segmen manajemen Informasi .

Dalam rangka mendorong nilai bisnis, Bisnis intelijen dapat diterapkan untuk tujuan bisnis berikut :

## 1. Pengukuran

Aplikasi/program yang mampu menciptakan hirarki metrik kinerja dan benchmarking yang menginformasikan pemimpin bisnis tentang kemajuan menuju tujuan bisnis.

## 2. Analytics

Aplikasi/program yang mampu membangun proses kuantitatif untuk bisnis untuk sampai pada keputusan yang optimal dan untuk melakukan bisnis penemuan pengetahuan. Pada proses analisis ini, sering melibatkan : data mining, data process, analisis statistik, analisis prediktif, pemodelan prediktif, pemodelan proses bisnis, pengolahan informasi kompleks dan analisis preskriptif .

## 3. Pelaporan

Aplikasi/program yang mampu membangun infrastruktur untuk pelaporan strategis untuk melayani manajemen strategis bisnis, bukan pelaporan operasional. Sering melibatkan visualisasi data, sistem informasi eksekutif dan OLAP .

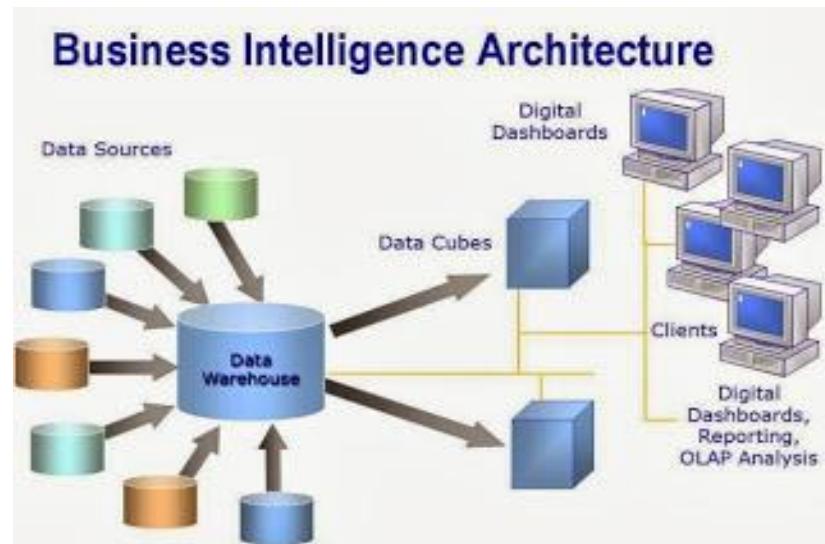
## 4. Kolaborasi/platform kolaborasi

Aplikasi/program yang mampu mendapat area yang berbeda (baik di dalam dan di luar bisnis) untuk bekerja sama melalui berbagi data dan pertukaran data elektronik .

## 5. Pengetahuan manajemen

Aplikasi/program yang mampu membuat data perusahaan didorong melalui strategi dan praktik untuk mengidentifikasi, menciptakan, merepresentasikan, mendistribusikan, dan memungkinkan adopsi wawasan dan pengalaman yang pengetahuan bisnis sejati.

Selain di atas, intelijen bisnis juga dapat memberikan pendekatan proaktif, seperti fungsi alarm untuk mengingatkan pengguna aplikasi. Ada banyak jenis peringatan, misalnya jika beberapa nilai bisnis melebihi nilai ambang jumlah dalam laporan, maka intelijen bisnis akan memberikan peringatan dan analis bisnis bisa menjadi lebih waspada. Kadang-kadang mail alert akan dikirim ke pengguna juga. Berikut gambar Arsitektur Bisnis Intellijen :



Gambar 2.1 Arsitektur Business Intelligence (sis.binus.ac.id, 2017)

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 email.

2. Data Warehouse

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

### 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 hipotesis 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 solusis 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 tersebut.

Menurut (Turban et al., 2011), business intelligence terbagi ke dalam lima jenis atau kategori yaitu :

1. Enterprise Reporting digunakan untuk menghasilkan laporan-laporan statis yang didistribusikan ke banyak orang. Jenis laporan ini sangat sesuai untuk laporan operasional dan dashboard.

2. Cube Analysis digunakan untuk menyediakan analisis OLTP multidimensional yang ditujukan untuk manajer bisnis dalam lingkungan terbatas.
3. Ad Hoc Query and Analysis digunakan untuk memberikan akses kepada user agar dapat melakukan query pada basis data, dan menggali informasi sampai pada tingkat paling dasar dari informasi transaksional. Query ini berfungsi untuk mengeksplor informasi yang dilakukan oleh user.
4. Statistical Analysis and Data Mining digunakan untuk melakukan analisis prediksi atau menentukan korelasi sebab akibat diantara dua matrik.
5. Delivery Report and Alert digunakan secara proaktif untuk mengirimkan laporan secara lengkap atau memberikan peringatan kepada populasi user yang besar atau banyak.

Menurut Ronald (2008) ada beberapa bagian dalam solusi business intelligence yaitu, keseluruhan proses dalam business intelligence dapat diterjemahkan menjadi langkah-langkah dibawah ini :

1. Identifikasi masalah bisnis yang perlu diselesaikan dengan gudang data dan menentukan data yang diperlukan untuk menyelesaikan masalah tersebut.
2. Identifikasi lokasi dari data-data yang diperlukan dan mengambilnya dari sumber penyimpanannya.
3. Merubah data yang diperoleh dari beragam sumber tersebut ke dalam sebuah data yang konsisten.
4. Mengambil data yang telah dirubah tersebut ke dalam lokasi yang tersentralisasi.
5. Membuat sebuah gudang data dengan data yang ada dalam lokasi yang tersentralisasi tersebut
6. Memasang sebuah produk atau aplikasi yang dapat memberikan akses ke data yang ada dalam cube tadi. Ada berbagai macam jalan

dan cara untuk berbagai macam tipe pekerjaan ketika berurusan dengan cube.

Pembentukan sebuah data *warehouse* dimulai dengan tahapan praproses data (Kamber, 2006) yang akan mengubah data yang bersifat tidak jelas/rusak, data yang kekurangan nilai atributnya, dan data yang tidak konsisten.

Tahapan praproses data ini terdiri dari :

- a. Integrasi data

Integrasi data adalah penggabungan data dari berbagai sumber penyimpanan data untuk menjadi satu kesatuan data yang koheren.

- b. Reduksi data

Teknik reduksi data diterapkan untuk memperoleh representasi tereduksi dari sejumlah data yang berimplikasi pada volume yang jauh lebih kecil.

- c. Pembersihan data

Proses ini merupakan tahapan pembersihan data, yaitu mengisi data yang hilang, mengatasi data yang kotor dan rusak, mengidentifikasi atau membuang data pencilan, memperbaiki data yang tidak konsisten.

- d. Transformasi data

Transformasi data yaitu proses pengubahan data menjadi bentuk yang tepat. Proses ini dilakukan agar kondisi data tetap konsisten dan dapat digunakan untuk proses analisis.

## 2.7 DW dan BI

Pada dasarnya Data Warehouse (DW) adalah database, dan merupakan pusat data yang dibentuk dari hasil penggabungan dan pengolahan data dari beragam sumber data. Biasanya DW digunakan untuk keperluan pelaporan dan analisis data.

Business Intelligent (BI) adalah Kumpulan metodologi, proses, arsitektur, dan teknologi yang berfungsi untuk mentransformasi data menjadi informasi yang mengandung arti dan berguna. Biasanya informasi ini ditampilkan dalam bentuk laporan, analitik dan dashboard. Pada umumnya BI menggunakan DW sebagai sumber datanya. Informasi yang dihasilkan oleh BI digunakan sebagai landasan dalam pengambilan keputusan dalam sebuah perusahaan.

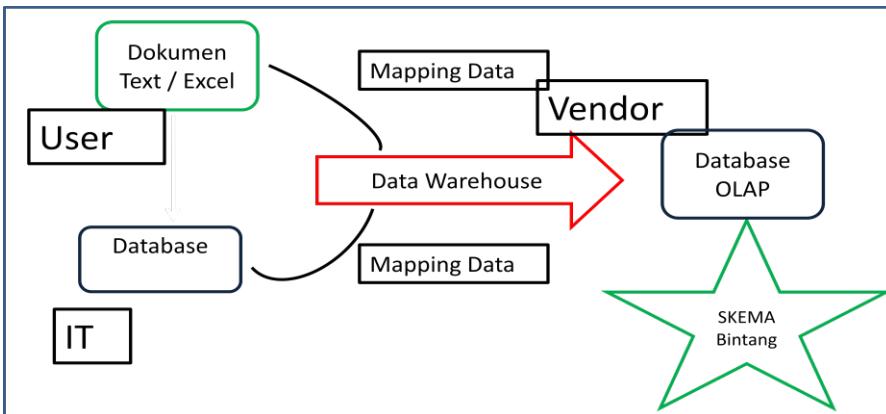
## 2.8 Software BI

Perangkat lunak atau software BI yang banyak digunakan di antaranya :

- Pentaho (2004 : <http://www.pentaho.com>) Pentaho Foundation
- Eclipse BI (2005) Eclipse Foundation
- SpagoBI (OWk Consortium:<http://www.spagobi.org>)
- Birst (<http://www.birst.com/>)
- IBM Cognos  
(<http://www-01.ibm.com/software/analytics/cognos/express/>)
- Microsoft SSAS (Server 2005 Analysis Services), dan Microsoft SSRS (Server Report Services)  
<http://www.microsoft.com/en-us/server:cloud/solutions/business:intelligence/analysis.aspx>
- Oracle BI  
(<http://www.oracle.com/us/solutions/business:analytics/business:intelligence/enterprise:edition/overview/index.html>)
- SAP BI

## 2.9 Sistem Kerja DW

Sistem kerja DW dijelaskan pada gambar berikut :



Gambar 2.2 Sistem kerja DW (Multimatics, 2015)

### Soal latihan :

1. Jelaskan sejarah perkembangan Business Intelligence pada pertengahan tahun 1980 ?
2. Jelaskan yang dimaksud dengan Business Intelligence ?
3. Jelaskan hubungan antara Data Warehouse dengan Business Intelligence ?
4. Jelaskan aplikasi Business Intelligence yang dikembangkan oleh Microsoft ?
5. Jelaskan secara singkat sistem kerja Data Warehouse ?

## BAB III

# PENGENALAN DATA WAREHOUSE

**P**ada Bab sebelumnya telah dipelajari tentang keterkaitan antara Business Intelligent dengan *Data Warehouse*, pada Bab ini akan diuraikan tentang konsep dasar dan hal-hal yang berkaitan dengan *Data Warehouse*.

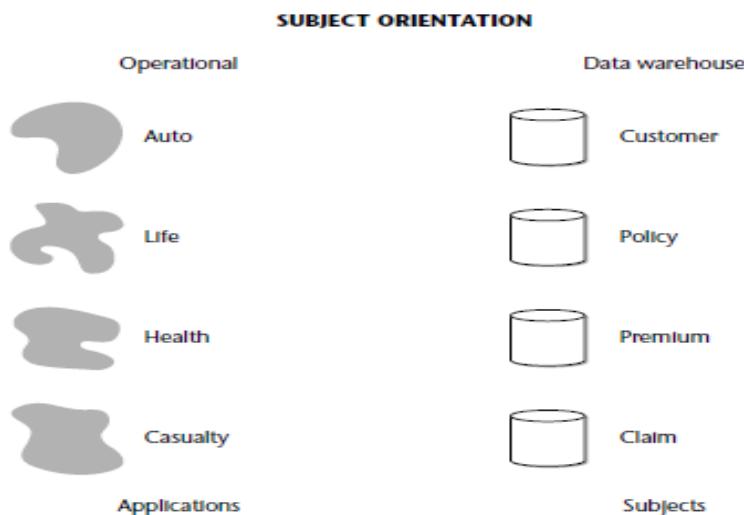
### 3.1 Data Warehouse

Menurut Inmon ***data warehouse*** adalah merupakan kumpulan data yang berorientasi subyek, terintegrasi, tidak dapat di *update*, memiliki dimensi waktu, yang digunakan untuk mendukung proses manajemen pengambilan keputusan dan kecerdasan bisnis. Berdasarkan definisi tersebut, maka *data warehouse* memiliki karakteristik sebagai berikut :

#### 1. Berorientasi subjek

Data diorganisasi oleh subjek detail (misal berdasarkan pelanggan), berisi informasi yang relevan untuk mendukung keputusan. *Data warehouse* berbeda dengan database operasional. Pada umumnya, *database* operasional mempunyai

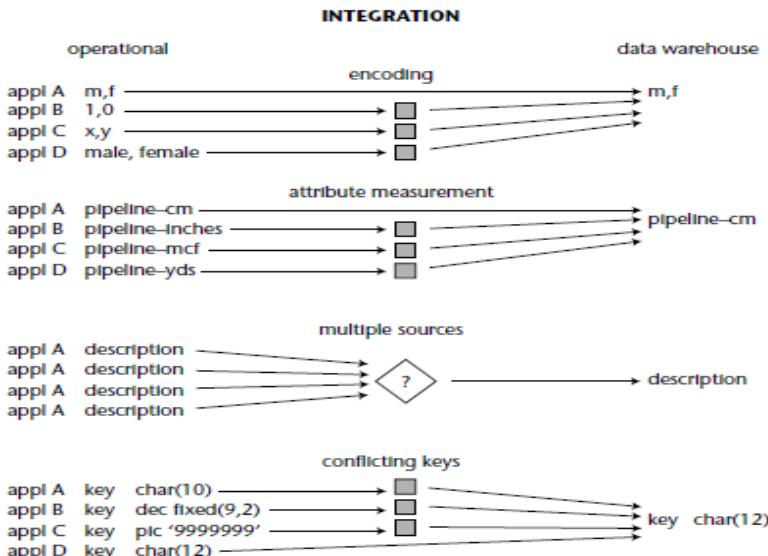
sebuah orientasi produk untuk menangani transaksi yang memperbarui *database*.



Gambar 3.1 contoh data berorientasi subjek (Inmon WH, 2005)

## 2. Terintegrasi

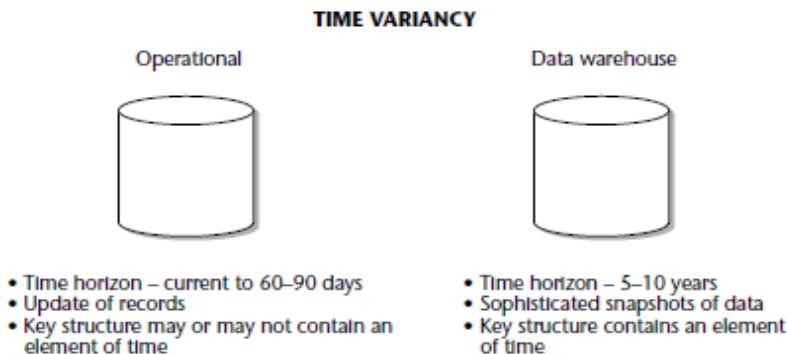
Data pada sumber berbeda dapat di *encode* dengan cara yang berbeda. Sebagai contoh data jenis kelamin dapat di *encode* sebagai 0 dan 1 di satu tempat dan “m” dan “f” di tempat lain. Di dalam *data warehouse*, *encode* tersebut dibersihkan atau dibuat ke dalam satu format sehingga datanya dapat distandarisasi dan konsisten. Banyak organisasi menggunakan terminologi yang sama untuk data dari jenis yang berbeda. Sebagai contoh “penjualan bersih” bisa berarti “komisi bersih” untuk departemen pemasaran, tetapi “retur penjualan kotor” bagi departemen akuntansi. Data yang terintegrasi mengatasi inkonsistensi dan menyediakan istilah yang seragam di organisasi keseluruhan, juga format waktu dan data yang bervariasi.



Gambar 3.2 contoh data terintegrasi (Inmon WH, 2005)

### 3. Time-variability

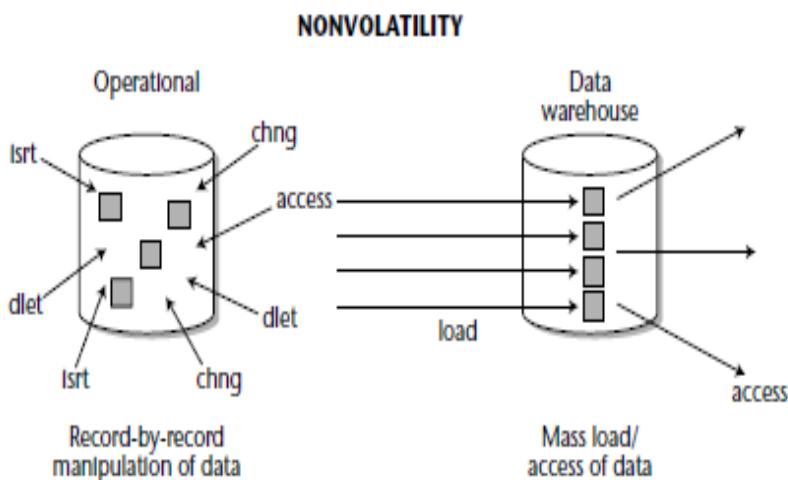
Data tidak menyediakan status saat ini, namun disimpan untuk 5 atau 10 tahun bahkan lebih dan digunakan untuk tren, peramalan, dan perbandingan. Ada kualitas sementara pada sebuah *data warehouse*. Waktu adalah dimensi penting yang harus didukung oleh semua *data warehouse*. Data untuk analisis dari berbagai sumber berisi berbagai poin waktu (misal, harian, mingguan, bulanan).



Gambar 3.3 Time Variancy (Inmon WH, 2005)

#### 4. Non Volatile

Sekali dimasukkan ke dalam *data warehouse*, data adalah *read only*, mereka tidak bisa diubah atau diperbarui. Data yang dibuang, dan perubahan direkam sebagai data baru. Ini memungkinkan *data warehouse* untuk disesuaikan hamper secara ekslusif untuk akses data. Sebagai contoh, sejumlah besar ruang kosong (untuk pertumbuhan data) umumnya tidak diperlukan dan reorganisasi database dapat dijadwalkan bersama dengan operasi pengisian sebuah *data warehouse*.



Gambar 3.4 Nonvolatility (Inmon WH, 2005)

Data warehouse dapat dianggap sebagai suatu salinan data transaksional/OLTP (*Online Transaction Processing*) yang terstruktur untuk kebutuhan analisis, *reporting*, dan *data mining*.

Berikut perbedaan sistem OLTP (*Online Transaction Processing*) dan sistem *data warehouse* :

Tabel 3.1 Perbedaan OLTP dan *data warehouse*

OLTP	Data Warehouse
◆ Menangani data saat ini	◆ Lebih cenderung menangani data masa lalu
◆ Data disimpan pada beberapa platform	◆ Data disimpan dalam satu platform
◆ Data diorganisasi berdasarkan fungsi atau operasi, seperti : penjualan, produksi, dan pemrosesan pesanan	◆ Data diorganisasi menurut subjek seperti : pelanggan atau produk
◆ Pemrosesan bersifat berulang	◆ Pemrosesan sewaktu-waktu, tak terstruktur, dan bersifat heuristik
◆ Untuk mendukung keputusan harian (operasional)	◆ Untuk mendukung keputusan yang strategis
◆ Melayani banyak pemakai operasional	◆ Untuk mendukung pemakai manajerial yang berjumlah relatif sedikit
◆ Berorientasi pada transaksi	◆ Berorientasi pada analisis

### 3.1.1 Konsep Dasar Data Warehouse

Konsep dasar *data warehouse* antara lain sebagai berikut :

1. Mengambil data dari berbagai operational systems
2. Jika dibutuhkan dapat menyertakan data yang relevant dari sumber lain diluar sistem seperti indikator keberhasilan
3. Mengintegrasikan seluruh data dari berbagai sumber
4. Menghilangkan data yang tidak konsisten dan merubah bentuk sesuai kebutuhan

5. Menyimpan data dalam format yang mudah digunakan untuk pengambilan keputusan.

### 3.2 Tujuan Data Warehouse

Tujuan dari *data warehouse* adalah sebagai berikut :

1. Menurut Connolly dan Begg (2005), tujuan data warehouse adalah mengintegrasikan data organisasi dengan jumlah yang besar menjadi satu wadah dimana pengguna dapat dengan mudah menjalankan query, membuat laporan serta melakukan analisis. Singkatnya, data warehouse adalah teknologi manajemen dan analisis data.
2. Menurut Mark W. Humphries, Michael W. Hawkins, Michelle C. Dy tujuan data warehouse adalah :
  - a. Menyediakan kalangan bisnis untuk mengakses data.

Data warehouse menyediakan layanan sehingga kalangan bisnis dapat mengakses data, yang sebenarnya rumit dan sulit dipahami, dengan cukup mudah.
  - b. Menyediakan data yang valid.

Sebagai contoh kasus, customer bukan merupakan sebuah istilah yang diberikan kepada semua klien. Ada sebuah patokan dimana klien pantas diberi title customer atau tidak. Dengan adanya data warehouse, sebuah title customer dapat ditentukan mungkin dari jumlah total pembelian atau faktor lainnya.
  - c. Untuk menyimpan data yang lama dengan akurat.

Hal ini sangat diperlukan ketika diperlukan sebuah perbandingan antara laporan sekarang dengan yang lalu.

- d. Slice and Dice data.

Sebuah ketersediaan data detail dapat meningkatkan bisnis analisis dengan mengurangi waktu dan usaha yang diperlukan untuk mengumpulkan data kembali.

- e. Memisahkan antara proses analisis dan operasional.

Biasanya data warehouse diperlukan untuk mendukung proses analisis karena sifatnya yang cepat dalam mengolah data yang sangat banyak walaupun dengan kemampuan terbatas.

Beberapa sifat antara lain :

Proses analisis -> READ

Proses operasional -> READ, WRITE, UPDATE

- f. Mendukung re-engineering pada decisional process. Dengan sifatnya yang memfokuskan pada pengambilan keputusan bisnis, data warehouse adalah sebuah sistem yang ideal apabila diperlukan re-engineering pada proses pengambilan keputusan bisnis.

### 3.3 Mekanisme Data pada Data Warehouse

*Data warehouse* memiliki 4 (empat) level data (Kimball, 2004) yaitu :

- 1. Level operasional

Level ini khusus menyimpan data yang berorientasi pada aplikasi secara detil dan diutamakan untuk memenuhi kebutuhan operasional. Akses pada level ini umumnya memiliki frekuensi yang sangat tinggi, sering juga disebut dengan OLTP (*Online Transactional Processing*).

## 2. Level integrasi

Level ini menyimpan data historis yang terintegrasi serta berorientasi subjek. Data pada level ini lebih sederhana daripada di level operasional dan tidak dapat di-update.

## 3. Level *data mart*

Level ini dirancang sesuai dengan kebutuhan dari pengguna dan merupakan data agregasi

## 4. Level individual

Level ini merupakan tempat terjadinya analisis dan *reporting*. Data pada level ini bersifat temporal, *ad hoc*, heuristic dan non retitif.

Tabel 3.2 Aliran data dalam mekanisme *data warehouse*

Operasional	Integrasi	Data Mart	Individual
<ul style="list-style-type: none"><li>♦ Detil</li><li>♦ Harian</li><li>♦ Nilai aktual</li><li>♦ Orientasi aplikasi</li></ul>	<ul style="list-style-type: none"><li>♦ Granular</li><li>♦ Time variant</li><li>♦ Terintegrasi</li><li>♦ Orientasi subjek</li><li>♦ Ringkas</li></ul>	<ul style="list-style-type: none"><li>♦ Spesialisasi sesuai pengguna</li><li>♦ Data agregasi</li></ul>	<ul style="list-style-type: none"><li>♦ Temporal</li><li>♦ <i>Ad hoc</i></li><li>♦ Heuristik</li><li>♦ Non-repetitif</li></ul>

### 3.4 Anatomi Data Warehouse

Penerapan awal dari arsitektur data warehouse dibuat berdasarkan konsep bahwa data warehouse mengambil data dari berbagai sumber dan memindahkannya ke dalam pusat pengumpulan data yang besar. Konsep ini sebenarnya lebih cenderung kepada sebuah lingkungan mainframe yang terpusat.

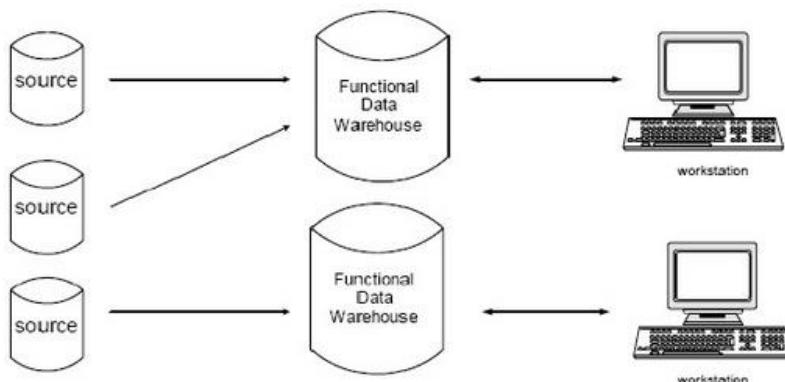
Keunggulan teknologi Client Server memungkinkan data warehouse diterapkan dalam berbagai macam cara untuk menampung kebutuhan pemakai sistem secara lebih proporsional. Dalam suatu kasus, misalkan saja pemakai tertentu perlu menggabungkan data dari sebuah sistem pengumpulan data yang statis dengan data dari sistem operasional yang dinamis hanya dengan sebuah query saja.

Berikut ini adalah tiga jenis dasar sistem Data Warehouse :

## 1. Functional Data Warehouse (Data Warehouse Fungsional)

Data operasional disini merupakan database yang diperoleh dari kegiatan sehari-hari. Data warehouse dibuat lebih dari satu dan dikelompokkan berdasar fungsi-fungsi yang ada di dalam perusahaan seperti fungsi keuangan/financial, marketing, personalia dan lain-lain.

Keuntungan dari bentuk data warehouse seperti ini adalah, sistem mudah dibangun dengan biaya relatif murah sedangkan kerugiannya adalah resiko kehilangan konsistensi data dan terbatasnya kemampuan dalam pengumpulan data bagi pengguna.

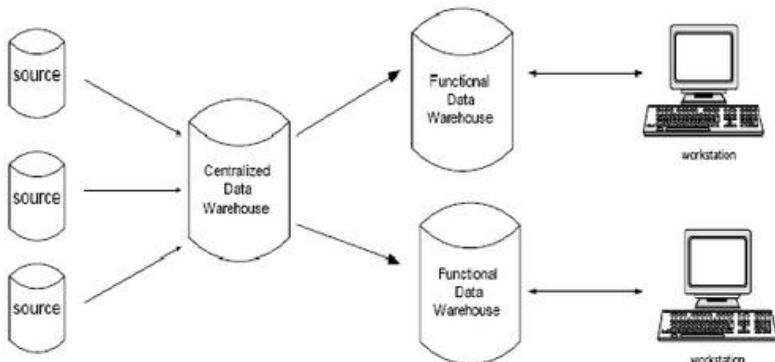


Gambar 3.5 Bentuk Data Warehouse Fungsional (Deliana et al, 2009)

## 2. Centralized Data warehouse (Data Warehouse Terpusat)

Bentuk ini terlihat seperti bentuk data warehouse fungsional, namun terlebih dahulu sumber data dikumpulkan dalam satu tempat terpusat, kemudian data disebar ke dalam fungsinya masing-masing, sesuai kebutuhan persusahaan. Data warehouse terpusat ini, biasa digunakan oleh perusahaan yang belum memiliki jaringan eksternal.

Keuntungan bentuk dari bentuk ini adalah data benar-benar terpadu karena konsistensinya yang tinggi sedang kerugiannya adalah biaya yang mahal serta memerlukan waktu yang cukup lama untuk membangunnya.

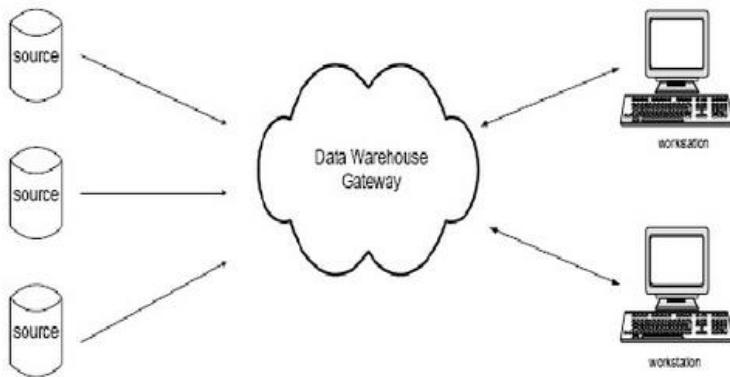


Gambar 3. 6 Bentuk Data warehouse Terpusat (Deliana et al, 2009)

### 3. Distributed Data Warehouse (Data Warehouse terdistribusi)

Pada data warehouse terdistribusi ini, digunakan gateway yang berfungsi sebagai jembatan penghubung antara data warehouse dengan workstation yang menggunakan sistem beraneka ragam. Dengan sistem terdistribusi seperti ini memungkinkan perusahaan dapat mengakses sumber data yang berada diluar lokasi perusahaan (eksternal).

Keuntungannya adalah data tetap konsisten karena sebelum data digunakan data terlebih dahulu di sesuaikan atau mengalami proses sinkronisasi. Sedangkan kerugiannya adalah lebih kompleks untuk diterapkan karena sistem operasi dikelola secara terpisah juga biayanya yang paling mahal dibandingkan dengan dua bentuk data warehouse lainnya.



Gambar 3. 7 Bentuk Data warehouse Terdistribusi (Deliana et al, 2009)

### 3.5 Arsitektur Data Warehouse

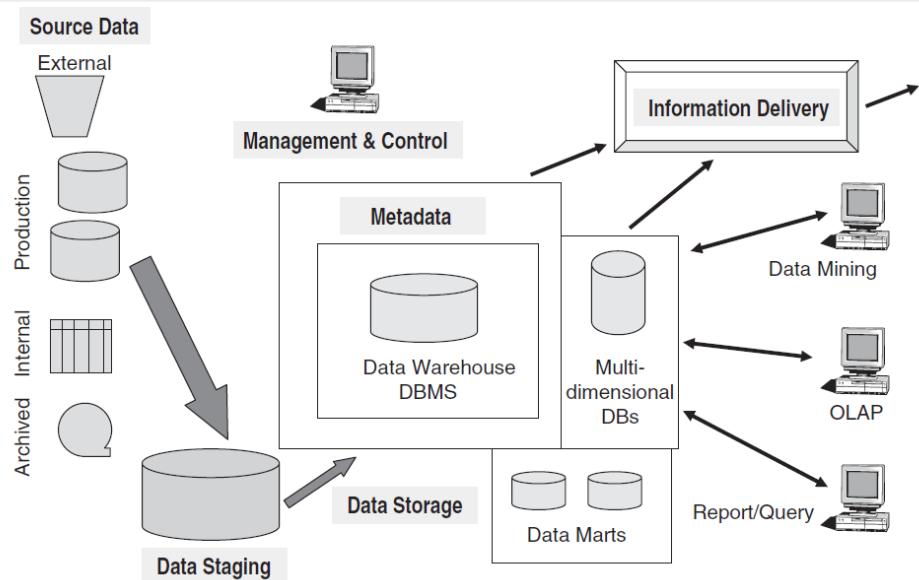
Menurut Poe arsitektur adalah sekumpulan atau struktur yang memberikan kerangka untuk keseluruhan rancangan suatu sistem atau produk. Ada arsitektur client-server, arsitektur networking dan masih banyak arsitektur lainnya. Arsitektur data menyediakan kerangka dengan mengidentifikasi dan memahami bagaimana data akan pindah melalui sistem dan digunakan dalam perusahaan. Arsitektur data untuk data warehouse mempunyai komponen utama yaitu read-only database.

Karakteristik arsitektur data warehouse (Poe) :

1. Data diambil dari sistem asal (sistem informasi yang ada), database dan file.
2. Data dari sistem asal diintegrasikan dan ditransformasi sebelum disimpan ke dalam Database Management System (DBMS) seperti Oracle, Ms SQL Server, Sybase dan masih banyak yang lainnya.
3. Data warehouse merupakan sebuah database terpisah bersifat hanya dapat dibaca yang dibuat khusus untuk mendukung pengambilan keputusan.

- Pemakai mengakses data warehouse melalui aplikasi front end tool.

Berikut adalah gambar arsitektur sistem *data warehouse* :



Gambar 3. 8 arsitektur sistem *data warehouse* (Ponniah Paulraj, 2010)

Sumber data yang masuk ke dalam *data warehouse* dapat dikelompokkan ke dalam empat kategori besar, sebagai berikut :

1. Production Data

Kategori data ini berasal dari berbagai sistem operasional perusahaan. Ini biasanya mencakup sistem keuangan, sistem manufaktur, sistem supply chain, dan sistem manajemen hubungan pelanggan.

2. Internal Data

Setiap organisasi, pengguna menyimpan spreadsheet "pribadi", dokumen, profil pelanggan, dan terkadang bahkan database

departemen. Ini adalah data internal, bagian yang bisa berguna di data warehouse.

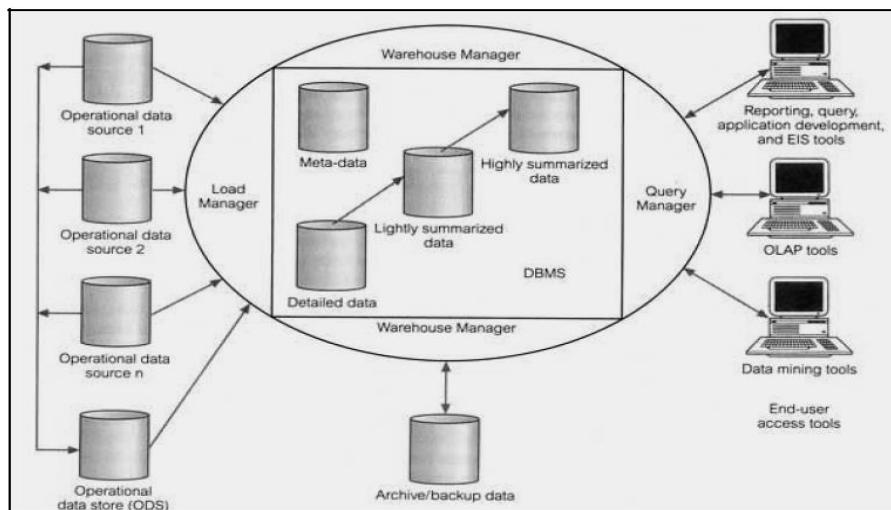
### 3. Archived Data

Disetiap sistem operasional, secara berkala dilakukan pengambilan data lama dan menyimpannya di file arsip.

### 4. External Data

Kebanyakan eksekutif atau top level management bergantung pada data dari sumber eksternal dengan persentase tinggi untuk informasi yang digunakan.

Menurut Connolly (2010), komponen-komponen utama dalam sebuah data warehouse antara lain dapat dilihat pada gambar berikut :



Gambar 3.9 Arsitektur Data warehouse (Connolly, 2010)

Keterangan gambar :

#### 1. Operational Data

Data untuk data warehouse berasal dari:

- Mainframe data operasional yang terdapat pada generasi pertama, yaitu hierarki dan basis data jaringan.

- Data departemen yang berada pada sistem *file*, seperti VSAM, RMS, dan relasional DBMS (seperti Informix dan Oracle).
- Data pribadi yang berada pada *workstation* dan server pribadi.
- Sistem-sistem eksternal seperti internet, *database* yang tersedia secara komersil, atau *database* yang berhubungan dengan pemasok atau pelanggan perusahaan.

## 2. *Operational Data Store*

*Operational Data Store* (ODS) merupakan tempat penyimpanan data operasional terkini dan terintegrasi, yang digunakan untuk memenuhi kebutuhan analisis. ODS menyimpan data yang telah diekstrak dan telah dibersihkan dari sumber data. Dengan demikian, proses pengintegrasian dan restrukturisasi data untuk *data warehouse* menjadi lebih sederhana.

## 3. *ETL Manager*

*ETL manager* melakukan semua operasi yang berhubungan dengan fungsi ETL (Extract, Transform, Loading) data ke dalam *data warehouse*. Data dapat diekstrak dari sumber-sumber data atau pada umumnya diambil dari *Operational Data Store*.

## 4. *Warehouse Manager*

*Warehouse manager* melakukan semua operasi yang berhubungan dengan manajemen data dalam *data warehouse*, seperti : analisis data untuk memastikan konsistensi, transformasi dan penyatuan sumber data dari media penyimpanan sementara ke tabel *data warehouse*, membentuk indeks dan *view* pada tabel, *generate* proses denormalisasi, *generate* aggregasi, dan melakukan *back up* dan *archiving* data

## 5. *Query Manager*

*Query manager* melakukan semua operasi yang berhubungan dengan pengaturan *query* yang dimasukkan oleh *user*. Operasi

yang dilakukan komponen ini berupa pengarahan *query* pada tabel-tabel yang tepat dan penjadwalan eksekusi *query*.

#### 6. Detailed Data

Komponen ini menyimpan semua data detil dalam skema basis data. Pada umumnya beberapa data tidak disimpan secara *online*, tetapi dapat dilakukan secara aggregasi. Secara periodik data detil ditambahkan ke *data warehouse* untuk mendukung aggregasi data.

#### 7. Lightly and Highly Summarized Data

Komponen ini menyimpan semua data yang sudah diringkas (diaggregasi), yang digenerate oleh *warehouse manager*. Data perlu diringkas dengan tujuan untuk mempercepat performa *query*. Ringkasan data terus diperbaharui seiring dengan adanya data yang baru yang masuk ke dalam *data warehouse*.

#### 8. Archive / Backup Data

Komponen ini menyimpan data detil dan ringkasan data dengan tujuan untuk menyimpan dan *backup* data. Walaupun ringkasan data diperoleh dari data detil, ringkasan perlu dibackup juga apabila data tersebut disimpan melampaui periode tertentu dalam penyimpanan data detil.

#### 9. Metadata

Komponen ini menyimpan semua definisi metadata (informasi mengenai data) yang digunakan dalam proses *data warehouse*. Metadata digunakan untuk berbagai tujuan, diantaranya: proses *extracting* dan *loading*, untuk memetakan sumber data dalam *warehouse*; dalam proses manajemen *warehouse*, metadata digunakan untuk mengotomatisasi pembentukan tabel ringkasan; sebagai bagian dari proses manajemen *query*, metadata digunakan untuk mengarahkan sebuah *query* pada sumber data yang tepat.

## 10. End-User Access Tools

Tujuan utama dari *data warehouse* adalah mendukung dalam proses pembuatan keputusan yang strategis dalam berbisnis. Para pengguna berinteraksi dengan *data warehouse* menggunakan *end-user access tools*. Berdasarkan kegunaannya, terdapat empat kategori *end-user access tools*, yaitu : *Reporting and Query Tools*, *Application Development Tools*, *Online Analytical Processing (OLAP) Tools*, dan *Data Mining Tools*.

### 3.6 Proses dalam *Data Warehouse*

Beberapa proses yang terjadi dari basis data operasional menuju *data warehouse* adalah sebagai berikut :

1. Melakukan ekstraksi data dari berbagai sumber

*Data warehouse* biasanya melibatkan lebih dari satu sumber data sehingga proses ekstraksi harus dilakukan terhadap bermacam-macam sumber data.

2. Melakukan integrasi data ke tempat yang umum

Setelah ekstraksi selesai, data tersebut disaring dan diperiksa keterhubungan dengan data lainnya kemudian data tersebut diintegrasikan ke basis data yang sama.

3. Menyimpan data dalam format yang dapat digunakan oleh pengguna

Umumnya data yang berasal dari sumber yang berbeda memiliki format data yang berbeda pula, agar dapat memanfaatkan informasi yang terdapat di dalamnya, data tersebut harus dikonversi/diubah menjadi satu format sehingga format datanya sama.

4. Menyediakan mekanisme bagi pengguna agar dapat mengakses *data warehouse*

Pengguna *data warehouse* harus memanfaatkan sudut pandang baru yang lebih kompleks untuk mendukung perubahan analisis perusahaan. Untuk memenuhi perubahan bagi kebutuhan pengguna terhadap basis data, maka *data warehouse* harus memiliki mekanisme yang fleksibel.

**Soal latihan :**

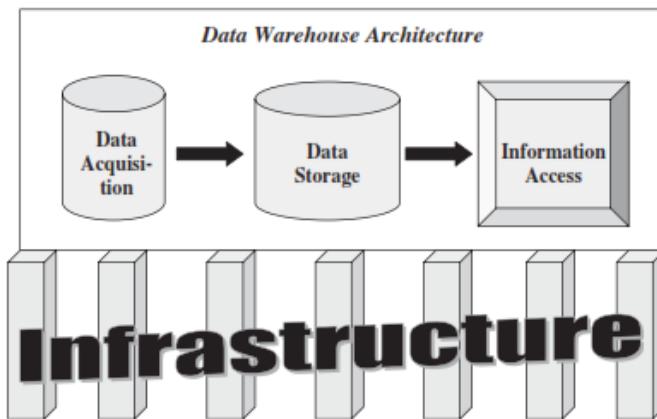
1. Jelaskan karakteristik *Data Warehouse* ?
2. Jelaskan perbedaan antara *OLTP* dengan *Data Warehouse* ?
3. Jelaskan konsep dasar *Data Warehouse* ?
4. Jelaskan Tujuan *Data Warehouse* menurut Conolly dan Begg ?
5. Menurut Kimball *Data warehouse* memiliki 4 (empat) level data, jelaskan !
6. Jelaskan tiga jenis dasar sistem *Data Warehouse* ?
7. Jelaskan secara singkat arsitektur *Data Warehouse* ?
8. Jelaskan beberapa proses yang terjadi dari basis data operasional menuju *data warehouse* ?

## BAB IV

### INFRASTRUKTUR DATA WAREHOUSE

**P**ada Bab sebelumnya telah dijelaskan tentang konsep dasar dan hal-hal yang berkaitan dengan *data warehouse*, sedangkan pada Bab ini akan dijelaskan tentang infrastuktur, karena membangun *data warehouse* harus didukung oleh infrastuktur yang baik agar hasilnya maksimal.

Infrastruktur data warehouse mencakup semua elemen dasar yang memungkinkan arsitektur untuk diimplementasikan. Singkatnya, infrastruktur mencakup beberapa elemen seperti perangkat keras server, sistem operasi, perangkat lunak jaringan, perangkat lunak database, LAN dan WAN, alat vendor untuk setiap komponen arsitektur, orang, prosedur, dan pelatihan. Untuk lebih jelasnya dapat dilihat pada gambar berikut :



Gambar 4.1 Infrastruktur pendukung arsitektur (Poniah, 2010)

Unsur-unsur infrastruktur data warehouse dapat diklasifikasikan menjadi dua kategori : infrastruktur operasional dan infrastruktur fisik. Perbedaan ini penting karena elemen di setiap kategori berbeda dalam sifat dan fiturnya dibandingkan dengan kategori lainnya. Pertama, akan membahas elemen-elemen yang dikelompokkan sebagai infrastruktur operasional. Infrastruktur fisik jauh lebih luas dan lebih mendasar. Setelah memperoleh pemahaman dasar tentang elemen-elemen arsitektur fisik, kita akan memeriksa elemen-elemen spesifik secara lebih rinci.

#### 4.1 Infrastruktur operasional

Untuk memahami infrastruktur operasional, diambil contoh pementasan data (data staging). Salah satu bagian dari infrastruktur dasar mengacu pada perangkat keras komputasi dan perangkat lunak. Diperlukan perangkat keras dan perangkat lunak untuk menjalankan fungsi data staging dan memberikan layanan yang sesuai. Dibutuhkan perangkat lunak untuk melakukan transformasi data. Diperlukan perangkat lunak untuk membuat file output. Diperlukan perangkat keras disk untuk menempatkan data dalam file staging area. Infrastruktur operasional terdiri dari aspek berikut :

- ◆ Orang-orang
- ◆ Prosedur
- ◆ Pelatihan
- ◆ Manajemen perangkat lunak

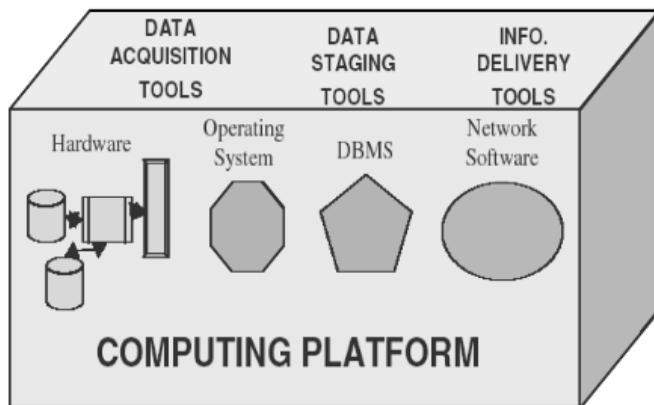
Bukan hanya orang-orang dan prosedur yang diperlukan untuk mengembangkan data warehouse. Diperlukan perlindungan untuk menjaga data warehouse. Elemen-elemen ini sama pentingnya dengan perangkat keras dan perangkat lunak yang menjaga data warehouse tetap berjalan. Mendukung pengelolaan data warehouse dan mempertahankan efisiensinya.

Pengembang data warehouse memberikan banyak perhatian pada perangkat keras dan elemen infrastruktur sistem perangkat lunak. Tetapi infrastruktur operasional sering diabaikan. Meskipun mungkin memiliki perangkat keras dan perangkat lunak yang tepat, data warehouse memerlukan infrastruktur operasional agar berfungsi dengan baik. Tanpa infrastruktur operasional yang tepat, data warehouse cenderung hanya berjalan dengan lambat dan tidak efektif.

#### **4.2 Infrastruktur fisik**

Infrastruktur fisik terdiri dari aspek berikut :

- ◆ Hardware (Perangkat Keras) dan Operating System (Sistem Operasi)
- ◆ DBMS
- ◆ Network Software



Gambar 4.2 Infrastruktur Fisik (Firdaus Solihin, 2009)

#### 4.2.1 Hardware Dan Operating System

Perangkat keras dan Sistem operasi menyusun lingkungan untuk data warehouse. Semua penyaringan data, transformasi, integrasi, dan penjadwalan kerja dijalankan pada hardware yang terpilih dan sistem operasi pilihan.

Beberapa hal yang harus diperhatikan untuk memilih perangkat keras dan sistem operasi antara lain :

1. Skalabilitas

Pastikan bahwa perangkat keras yang dipilih dapat ditingkatkan saat data warehouse jumlah pengguna semakin banyak, jumlah pertanyaan, dan kompleksitasnya dari queri. Data Warehouse tumbuh sangat cepat. Bersama perangkat keras dan perangkat lunak database, sistem operasi harus dapat mendukung peningkatan jumlah pengguna dan aplikasi.

2. Dukungan/Support

Dukungan vendor sangat penting untuk pemeliharaan perangkat keras. Pastikan bahwa dukungan dari vendor perangkat keras berada pada level setinggi mungkin.

### 3. Referensi Vendor

Penting untuk memeriksa referensi vendor dengan situs lain untuk menggunakan perangkat keras dari vendor ini.

### 4. Stabilitas Vendor

Periksa stabilitas dan daya tahan vendor.

### 5. Keamanan/Security

Ketika beberapa client mengakses server, sistem operasi harus dapat melindungi setiap client dan sumber daya terkait. Sistem operasi harus menyediakan lingkungan yang aman pada setiap client.

### 6. Keandalan/Reliability

Sistem operasi harus dapat melindungi aplikasi dari malfungsi (tidak berfungsi).

### 7. Ketersediaan/ Availability

Ketersediaan ini adalah untuk keandalan. Lingkungan komputasi harus terus tersedia dari keadaan aplikasi yang tidak normal.

### 8. Preemptive Multitasking

Perangkat keras server harus mampu menyeimbangkan alokasi waktu dan sumber daya di antara banyak tugas/pekerjaan. Sistem operasi harus dapat memprioritaskan tugas yang lebih tinggi atau menunggu tugas lain ketika diperlukan.

### 9. Gunakan Pendekatan Multithreaded

Sistem operasi harus dapat melayani beberapa permintaan secara bersamaan dengan mendistribusikan ke beberapa prosesor dalam konfigurasi perangkat keras multiprosesor. Fitur ini sangat penting karena konfigurasi multiprosesor adalah arsitektur pilihan dalam lingkungan data warehouse.

## 10. Perlindungan Memori

Dalam lingkungan data warehouse beberapa permintaan akan dijalankan secara bersamaan. Fitur perlindungan memori dalam sistem operasi mencegah pelanggaran terhadap memori yang lain.

Pilihan Umum Hardware Data Warehouse adalah sebagai berikut :

### a. Mainframe

Mainframe yang dipilih dengan mempertimbangkan beberapa hal berikut :

1. Hardware yang teruji kemampuannya
2. dirancang untuk OLTP dan bukan untuk aplikasi pendukung keputusan
3. tidak hemat biaya untuk data warehouse.

### b. Server Open System

Server Open System dipilih dengan mempertimbangkan hal-hal berikut :

1. Server UNIX, pilihan medium untuk kebanyakan data warehouse
2. biasanya Sempurna
3. Sesuai untuk pengolahan paralel

### c. NT Server

1. Mendukung data warehouse ukuran menengah
2. Kemampuan proses paralel yang terbatas
3. Hemat biaya untuk data warehouse kecil dan ukuran menengah

#### 4.2.1.1 Platform

Platform komputasi yang diperlukan untuk melakukan beberapa fungsi dari berbagai komponen arsitektur data warehouse. Platform komputasi

adalah seperangkat komponen perangkat keras, sistem operasi, jaringan, dan perangkat lunak jaringan. Fungsi sistem OLTP atau sistem pendukung keputusan seperti data warehouse, fungsi tersebut harus dilakukan pada platform komputasi.

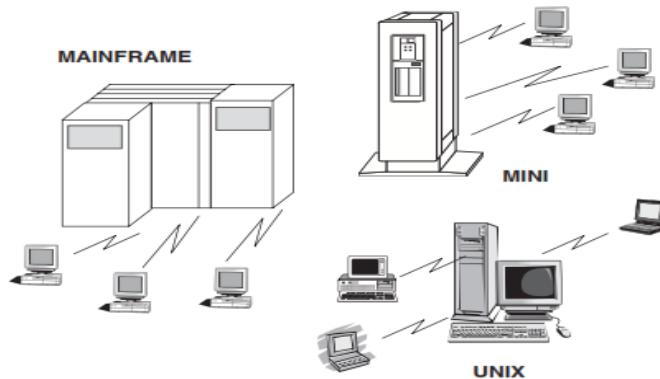
Berikut fungsi dan layanan komponen arsitektur di tiga bidang utama platform sebagai berikut :

1. Data Acquisition : ekstraksi data, transformasi data, pembersihan data, integrasi data, dan pementasan data.
2. Data Storage : pemuatan data, pengarsipan, dan manajemen data.
3. Information Delivery : pembuatan laporan, pemrosesan queri, dan analisis yang kompleks.

Berikut jenis platform pada komputasi antara lain :

1. Single Platform
  - a. Merupakan pilihan yang paling sederhana untuk implementasi arsitektur *data warehouse*
  - b. Semua fungsi dari *backend* penyaringan data pada *front-end* pengolahan *query* dilakukan pada komputasi *platform* tunggal.
2. Hybrid Platform

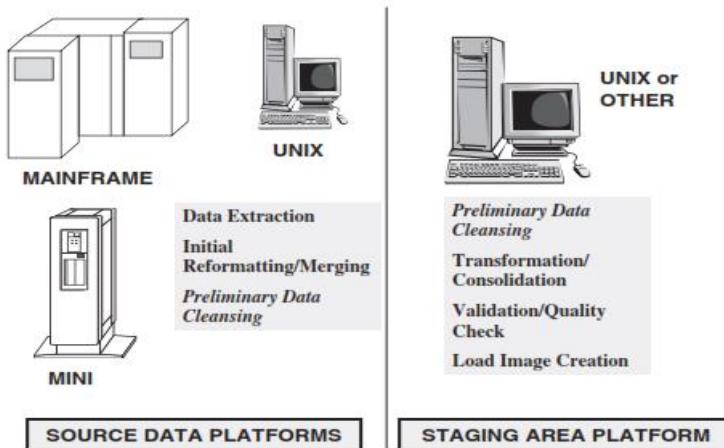
Berdasarkan sistem warisan dan aplikasi yang lebih modern pada perusahaan, kemungkinan besar akan memutuskan bahwa pendekatan Single-Platform tidak dapat dilakukan untuk data warehouse. Inilah kesimpulan yang paling sering dihadapi perusahaan. Di sisi lain, jika perusahaan termasuk dalam kategori dimana platform lama akan mengakomodasi data warehouse, maka diambil pendekatan solusi Single Platform. Sekali lagi solusi Single-Platform adalah solusi yang lebih mudah.



Gambar 4.3 pilihan platform pada data warehouse (Ponniah, 2010)

### 3. Options for the Staging Area

Dalam pembahasan langkah-langkah akuisisi data, disoroti platform komputasi yang optimal untuk setiap langkah. Langkah-langkah kunci terjadi di staging area. Platform yang paling sesuai untuk staging area tergantung pada sumber status platform.



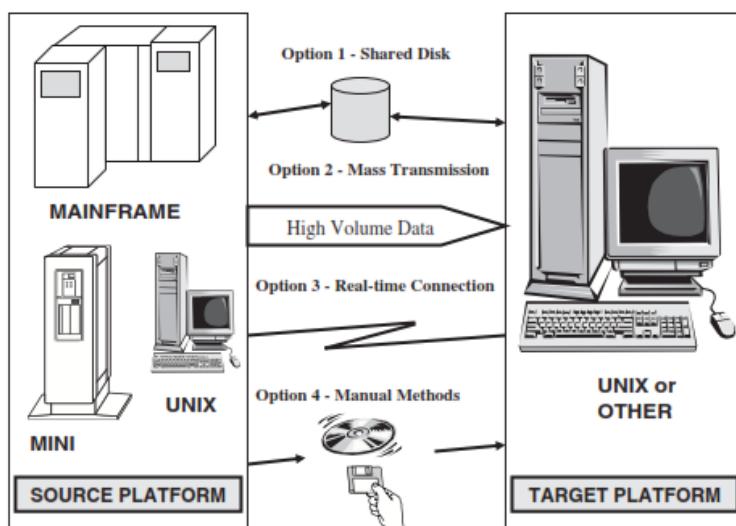
Gambar 4.4 Platform untuk Data Acquisition

### 4. Pertimbangan Pergerakan Data

Pada platform komputasi mana saja langkah-langkah individu dari akuisisi data dan penyimpanan data yang terjadi, data harus bergerak melintasi platform. Tergantung pada platform sumber di

perusahaan dan pilihan platform untuk pementasan data dan penyimpanan data, harus disediakan transportasi data di berbagai platform.

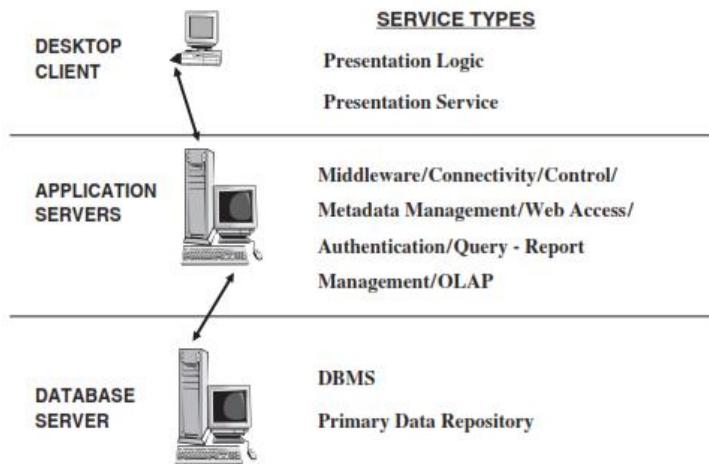
Ditemukan bahwa satu pendekatan saja tidak cukup. Untuk memiliki kombinasi yang seimbang dari pendekatan yang berbeda. Dalam setiap gerakan data di dua platform komputasi, pilih opsi yang paling sesuai. Penjelasan singkat tentang opsi standar ikuti.



Gambar 4.5 Data Movement Options (Poniah, 2010)

##### 5. Client / Arsitektur Server untuk Data Warehouse

Meskipun platform mainframe dan komputer mini digunakan dalam penerapan awal data warehouse, pada umumnya, warehouse saat ini dibangun menggunakan arsitektur client/server. Sebagian besar adalah arsitektur client / server multitier, generasi kedua. DBMS Data Warehouse mengeksekusi pada komponen server data. Repositori data dari gudang data ada di mesin ini. Komponen server ini adalah komponen utama dan dibahas secara rinci di bagian selanjutnya.



Gambar 4.6 Client / Arsitektur Server untuk Data Warehouse  
(Poniah, 2010)

## 6. Pertimbangan untuk Workstation Client

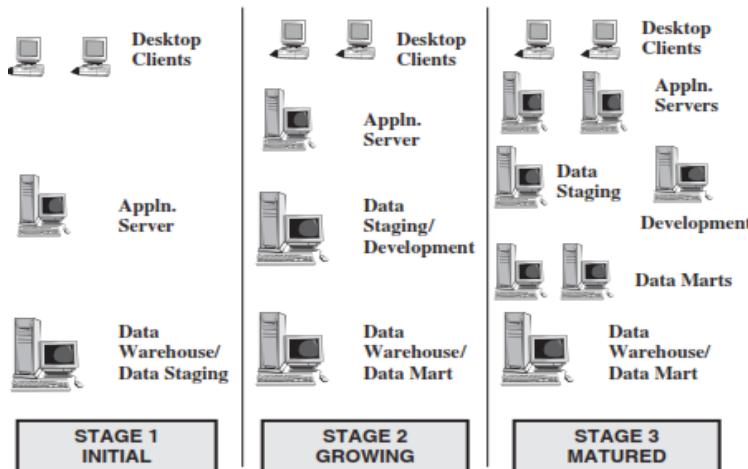
Hal yang harus dipertimbangkan saat konfigurasi untuk mesin workstation, harus disadari bahwa perlu melayani berbagai jenis pengguna. Jangan hanya mempertimbangkan kebutuhan di workstation yang berkaitan dengan pengiriman informasi dari data warehouse. Pengguna mungkin terpuaskan dengan mesin yang dapat menjalankan browser Web untuk mengakses laporan HTML. Seorang analis membutuhkan mesin workstation yang lebih besar dan lebih kuat.

## 7. Options as the Data Warehouse Matures

Setelah semua pembahasan tentang platform komputasi untuk data warehouse, bahwa pilihan platform telah diperbaiki segera setelah pilihan awal dibuat. Sangat menarik untuk dicatat bahwa sebagai data warehouse di setiap perusahaan, pengaturan platform juga berkembang.

Pementasan data dan penyimpanan data dapat dimulai pada platform komputasi yang sama. Seiring berjalananya waktu dan

semakin banyak pengguna mulai bergantung pada data warehouse untuk pengambilan keputusan strategis, bahwa pilihan platform mungkin harus disusun kembali.



Gambar 4.7 Pilihan platform pada Data Warehouse (Poniah, 2010)

#### 4.2.1.2 Perangkat Keras Server

Memilih perangkat keras server adalah salah satu keputusan terpenting yang dihadapi. Mungkin, untuk sebagian besar data warehouse, pemilihan perangkat keras server dapat menjadi keputusan. Skalabilitas dan kinerja query yang optimal adalah kunci utamanya.

Tujuan utama data warehouse untuk memberikan informasi kepada pengguna. Query yang kompleks dari data warehouse adalah metode yang paling umum untuk pengiriman informasi. Jika perangkat keras server tidak mendukung pemrosesan queri lebih cepat, seluruh proyek dalam bahaya. Ketika data warehouse matang, akan melihat peningkatan tajam dalam jumlah pengguna dan dalam jumlah queri. Biasanya, jumlah pengguna aktif menjadi dua kali lipat dalam enam bulan. Sekali lagi, ketika data warehouse matang, akan meningkatkan konten dengan memasukkan lebih banyak bidang subjek bisnis dan menambahkan lebih banyak lagi data mart. Data warehouse perusahaan

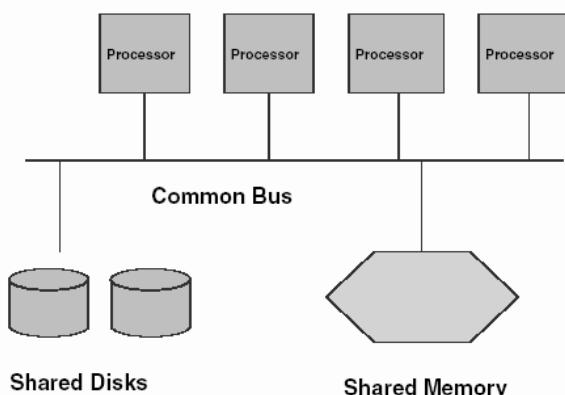
mulai dari sekitar 200 hingga 300 GB hingga terabyte dalam waktu 18 hingga 24 bulan.

Pilihan perangkat keras untuk skalabilitas dan pemrosesan query yang kompleks terdiri dari empat jenis arsitektur paralel. Awalnya, arsitektur paralel paling masuk akal. Queri diselesaikan lebih cepat jika menambah jumlah prosesor, setiap prosesor mengerjakan bagian-bagian queri secara bersamaan. Bisakah untuk tidak membagi lagi permintaan besar menjadi tugas-tugas terpisah dan menyebarluaskan tugas di antara banyak prosesor. Pemrosesan paralel dengan beberapa mesin komputasi memang memberikan berbagai manfaat, tetapi tidak ada opsi pemrosesan tunggal yang melakukan segalanya dengan benar.

## 1. **Symmetric Multiprocessing (SMP)**

Ciri – ciri :

- suatu arsitektur yang shared-everything,
- Mesin olah paralel yang paling sederhana.
- Masing-Masing processor mempunyai akses penuh kepada memori secara bersama melalui suatu umum bus.
- Komunikasi antara processor terjadi melalui memori umum.
- Disk dapat dikontrol oleh semua prosesor

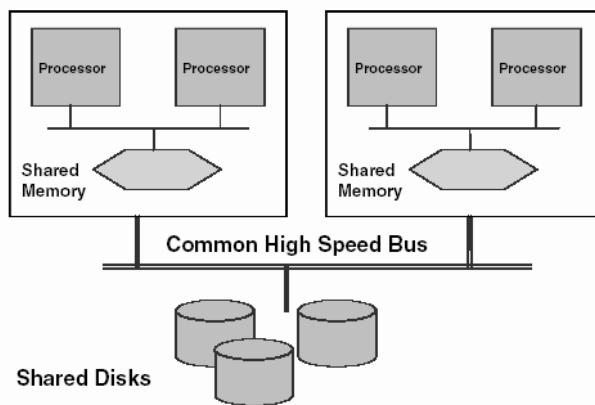


Gambar 4.8 Server Hardware Option : SMP (Poniah, 2010)

## 2. **Clusters**

Ciri – ciri :

- Masing-masing node terdiri dari satu atau lebih processor dan berhubungan dengan memori.
- Memori tidak shared di antara node tapi shared hanya di dalam masing-masing node.
- Komunikasi terjadi dalam kecepatan tinggi.
- Masing-masing node mempunyai akses umum ke setiap disk.
- Arsitektur ini adalah suatu cluster dari node.



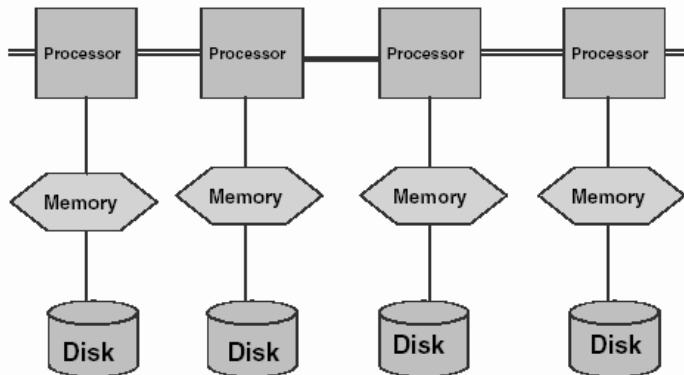
Gambar 4.9 Server Hardware Option : Cluster (Poniah, 2010)

## 3. **Massively Parallel Processing (MPP)**

Ciri – ciri:

- Ini adalah suatu arsitektur yang shared-nothing.
- Arsitektur ini lebih terkait dengan mengakses disk dibanding akses memori.
- Bekerja secara baik dengan suatu sistem operasi yang mendukung akses disk transparan.

- Jika suatu table database ditempatkan pada disk tertentu , akses ke disk itu tergantung seluruhnya pada processor yang memiliki disk
- Internode komunikasi adalah dengan koneksi antar processor ke processor lainnya.



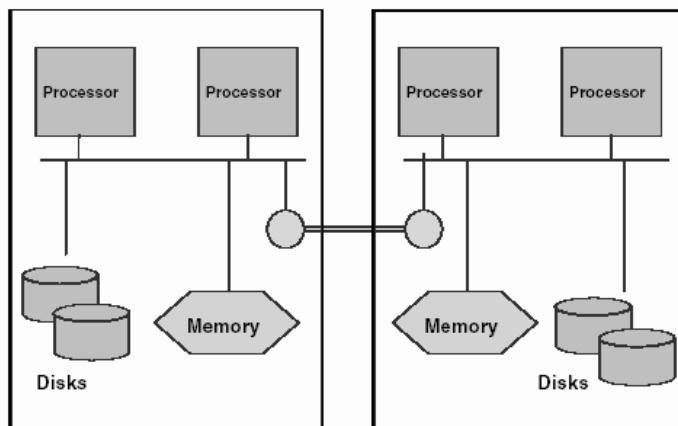
Gambar 4.10 Server Hardware Option : MPP (Poniah, 2010)

#### 4. **Cache-coherent Nonuniform Memory Architecture (NUMA)**

Ciri-ciri :

- Ini adalah arsitektur yang paling baru; dikembangkan pada awal 1990.
- NUMA arsitektur seperti suatu SMP besar mematahkan SMPs yang lebih kecil. Dimana yang lebih mudah untuk dibangun.
- Perangkat keras pertimbangkan semua unit memori sebagai suatu raksasa memori. Sistem mempunyai memori nyata tunggal menunjuk pada keseluruhan alamat memori mesin dimulai dengan 1 pada tangkai / node yang pertama dan berlanjut ke tangkai / node berikutnya. Masing-Masing tangkai / node berisi suatu direktori alamat memori di dalam tangkai / node tersebut.

- Di dalam arsitektur ini, jumlah waktu yang diperlukan untuk mendapatkan kembali suatu nilai memori bervariasi sebab tangki/node yang pertama mungkin memerlukan nilai yang berada di memori dari tangki / node ketiga. Hal inilah yang merupakan alasan mengapa arsitektur ini disebut arsitektur akses memori non uniform.



Gambar 4.11 Server Hardware Option : NUMA (Poniah, 2010)

#### 4.2.2 Database Software

##### 4.2.2.1 Parallel Processing Options

Pilihan pemrosesan paralel dalam perangkat lunak basis data hanya ditujukan untuk mesin dengan banyak prosesor. Sebagian besar perangkat lunak database saat ini dapat memparalelkan sebagian besar operasi. Operasi ini meliputi : pemuatan massal data, pemindai tabel, queri dengan kondisi pengecualian, queri dengan pengelompokan, pemilihan dengan nilai berbeda, agregasi, pengurutan, pembuatan tabel menggunakan subqueri, membuat dan membangun kembali indeks, menyisipkan baris ke dalam tabel dari tabel lain, memungkinkan kendala, transformasi bintang (teknik pengoptimalan saat memproses queri terhadap skema STAR), dan seterusnya.

Sekarang memeriksa yang terjadi ketika pengguna memulai queri di workstation. Setiap sesi mengakses database melalui proses server. queri dikirim ke DBMS dan pengambilan data berlangsung dari database. Data diambil dan hasilnya dikirim kembali, semua di bawah kendali proses server khusus. Perangkat lunak bertanggung jawab untuk membagi pekerjaan, mendistribusikan unit yang akan di proses dari kumpulan proses server queri yang tersedia, dan menyeimbangkan beban. Akhirnya, hasil proses queri dirakit dan dikembalikan sebagai satu set hasil gabungan.

a. Interquery Parallelization

Dalam metode ini, beberapa proses server menangani beberapa permintaan secara bersamaan. Beberapa queri dapat dilayani berdasarkan konfigurasi server dan jumlah prosesor yang tersedia. Fitur ini berhasil memanfaatkan DBMS pada sistem SMP, sehingga meningkatkan dan mendukung pengguna yang bersamaan.

Namun, paralelisme interquery terbatas. Beberapa queri diproses secara bersamaan, tetapi setiap queri masih diproses secara serial oleh satu proses server. Misalkan sebuah query terdiri dari membaca indeks, membaca data, gabungan, dan mengurutkan operasi, operasi ini dilakukan dalam urutan ini. Setiap operasi harus selesai sebelum yang berikutnya bisa dimulai. Bagian dari permintaan yang sama tidak dijalankan secara paralel. Untuk mengatasi keterbatasan ini, banyak vendor DBMS datang dengan versi produk mereka untuk menyediakan paralelisasi intraquery.

b. Intraquery Parallelization

Pemahaman ini akan membantu untuk mencocokkan pilihan perangkat keras server dengan RDBMS pilihan. Permintaan dari salah satu pengguna terdiri dari indeks yang dibaca, data yang dibaca, data gabungan, dan jenis data dari database data warehouse. Proses serial DBMS akan memproses queri ini dalam urutan operasi dasar dan menghasilkan hasil yang ditetapkan. Namun, saat queri ini dijalankan pada satu prosesor di sistem SMP, queri lain dapat dijalankan secara paralel. Metode ini adalah

paralelisasi interqueri. Menggunakan teknik paralelisasi intraquery, DBMS membagi queri ke dalam operasi tingkat yang lebih rendah dari membaca indeks, membaca data, menggabungkan data, dan mengurutkan data. Kemudian masing-masing operasi dasar ini dijalankan secara paralel pada satu prosesor. Hasil akhir yang ditetapkan adalah konsolidasi hasil perantara. Berikut tiga cara DBMS dapat menyediakan paralelisasi intraquery, yaitu paralelisasi bagian-bagian operasi dalam queri yang sama, antara lain :

### 1. Horizontal Parallelism

Data dipartisi di beberapa disk. Pemrosesan paralel terjadi dalam setiap tugas tunggal dalam query, misalnya : pembacaan data yang dilakukan pada beberapa prosesor secara bersamaan pada set data yang berbeda untuk dibaca dari beberapa disk.

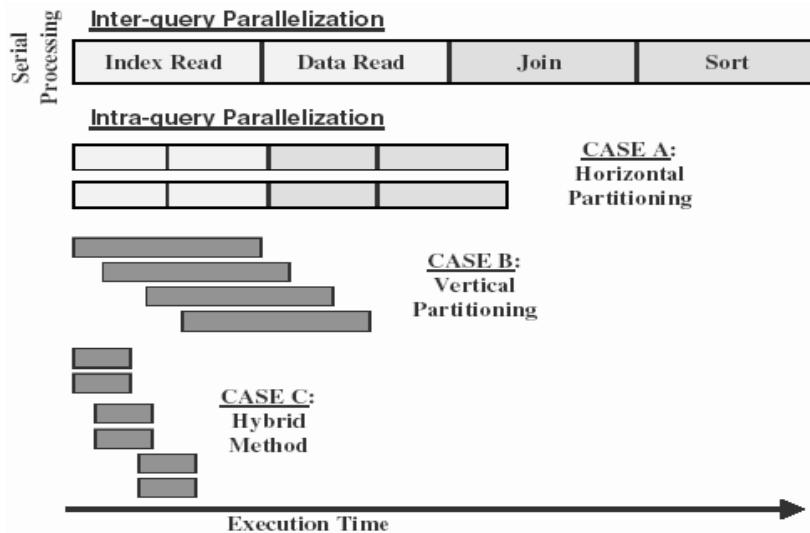
Setelah tugas pertama diselesaikan dari semua bagian yang relevan dari data yang dipartisi, tugas selanjutnya dari query itu dilakukan, dan kemudian yang berikutnya setelah tugas itu, dan seterusnya. Masalah dengan pendekatan ini adalah menunggu hasil sampai semua data yang diperlukan dibaca.

### 2. Vertical Parallelism

Paralisme semacam ini terjadi di antara tugas-tugas yang berbeda, bukan hanya satu tugas dalam sebuah pertanyaan seperti pada kasus paralelisme horizontal. Semua operasi query komponen dijalankan secara paralel, tetapi secara pipeline. Ini mengasumsikan bahwa RDBMS memiliki kemampuan untuk menguraikan queri menjadi subtugas, setiap subtugas memiliki semua operasi indeks yang dibaca, membaca data, menggabungkan dan mengurutkan. Kemudian masing-masing subtugas mengeksekusi data secara serial. Dalam pendekatan ini, catatan basis data idealnya diproses dengan satu langkah dan segera diberikan ke langkah berikutnya untuk diproses, sehingga menghindari waktu tunggu. Tentu saja, dalam metode ini, DBMS harus memiliki tingkat kecanggihan yang sangat tinggi dalam penguraian tugas.

### 3. Hybrid Method

Dalam metode ini, queri memartisi queri secara horizontal dan vertikal. Pendekatan ini menghasilkan hasil terbaik. Pemanfaatan sumber daya terbesar, kinerja optimal, dan skalabilitas tinggi.



Gambar 4.12 Intraquery Paralelization by DBMS (Poniah, 2010)

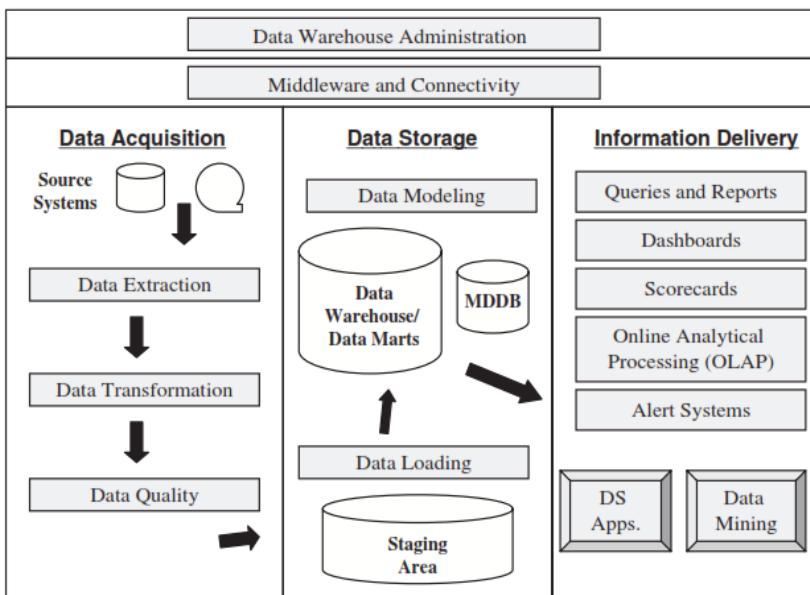
Pemilihan DBMS dilakukan berdasarkan beberapa hal berikut :

- Query governor (untuk mengantisipasi dan membatalkan queri yang sedang berjalan)
- Query optimizer (untuk mengurai dan mengoptimalkan queri pengguna)
- Query management (untuk menyeimbangkan eksekusi berbagai jenis queri)
- Load utility (untuk pemuatan, pemulihan, dan pemulihan data berkinerja tinggi)
- Metadata management (dengan katalog data atau kamus aktif)
- Scalability (baik dari segi jumlah pengguna dan volume data)

- Extensibility (memiliki ekstensi hibrida ke database OLAP)
- Portability (di seluruh platform)
- Query tool Application Program Interfaces (APIs) ( untuk alat dari vendor terkemuka)
- Administration (menyediakan dukungan untuk semua fungsi DBA)

#### 4.2.3 Collection Of Tools Data Warehouse

Memilih alat untuk proyek data warehouse, daftar berikut berfungsi sebagai referensi praktis, antara lain :



Gambar 4.13 Tools for Data Warehouse (Poniah, 2010)

Membahas komponen arsitektur dan mempelajari fungsi layanan individu. Sekarang, lanjutkan ke bagian berikutnya tentang asas penting layanan tersebut, antara lain :

#### 4.2.3.1 Architecture Tools

Subbagian ini akan dibahas : memilih alat; desain arsitektur, pilih alat untuk mencocokkan fungsi dan layanan yang ditetapkan untuk komponen arsitektur.

Tenaga vendor penjual alat menjanjikan kesuksesan. Merancang komponen arsitektur pengiriman informasi. Persyaratan bisnis adalah kekuatan pendorong. Kelompok pengguna terbesar adalah kelompok pengguna daya. Melakukan analisis yang kompleks terdiri dari drilling, slicing dan cutting data, dan visualisasi ekstensif dari rangkaian hasil. Membutuhkan komponen pengiriman informasi yang paling canggih. Fungsi dan layanan penyampaian informasi harus terlibat dan sangat kuat. Alat pengiriman informasi dipilih sebelum komponen arsitektur didirikan. Alat ini tidak memenuhi persyaratan karena hal ini telah direfleksikan dalam arsitektur.

Jenis perangkat lunak untuk data warehouse, antara lain sebagai berikut :

##### 1. Data Modeling

- Memungkinkan pengembang untuk membuat dan memelihara model data untuk sistem sumber dan basis data tujuan pada penyimpanan data. Jika perlu, model data dapat dibuat untuk area staging
- Menyediakan kemampuan teknik untuk menghasilkan skema database.
- Menyediakan dan menghasilkan model data dari entri kamus data dan dari basis data sumber yang ada
- Menyediakan kemampuan pemodelan perancang data untuk membuat skema STAR.

##### 2. Data Extraction

- Ekstraksi massal untuk penyegaran dan replikasi berbasis perubahan untuk beban tambahan.

- Pilihan alat bergantung pada faktor-faktor berikut : platform sistem sumber dan basis data, dan tersedia fasilitas ekstraksi dan duplikasi di dalam sistem sumber.

### 3. Data Transformation

- Mentransformasi data yang diekstrak ke dalam format dan struktur data yang tepat.
- Fitur utama termasuk pemisahan bidang, konsolidasi, standarisasi, dan deduplikasi.

### 4. Data Loading

- Memuat data yang diubah dan dikonsolidasikan ke dalam data warehouse.
- Beberapa pemuat data menghasilkan kunci primer untuk tabel yang dimuat.
- Untuk memuat data pada mesin RDBMS yang sama dengan data warehouse, prosedur yang disimpan sebelumnya pada database itu sendiri dapat digunakan untuk memuat.

### 5. Data Quality

- Bantu dalam menemukan dan memperbaiki kesalahan data
- Dapat digunakan pada data di area pementasan atau pada sistem sumber secara langsung
- Membantu menyelesaikan inkonsistensi data dalam memuat data.

### 6. Queries and Reports

- Memungkinkan pengguna untuk menghasilkan laporan yang canggih, grafis yang intensif dan canggih.
- Bantu pengguna untuk merumuskan dan menjalankan queri.
- Klasifikasi utama adalah penulis laporan dan server laporan.

7. Dashboards

- Memberikan informasi real-time atau mendekati waktu nyata kepada pengguna secara interaktif. Kebanyakan dashbor waktu nyata ditautkan langsung ke sistem operasional.
- Memungkinkan beberapa kemampuan untuk pengguna seperti menelusuri, perubahan parameter pada tampilan, berbagai jenis tampilan, dan sebagainya.

8. Scorecards

- Memungkinkan pengguna untuk memilih indikator kinerja utama dengan mudah dalam pelaporan.
- Memberikan perbandingan kinerja saat ini terhadap target dan kinerja masa lalu.
- Fokus pada kejelasan dan kemudahan penggunaan.

9. Online Analytical Processing (OLAP)

- Memungkinkan pengguna menjalankan queri dimensi yang kompleks.
- Memungkinkan pengguna untuk membuat queri
- Dua kategori pengolahan analitik online adalah pengolahan *Multidimensional online analytical processing* (MOLAP) dan *Relational online analytical processing* (ROLAP). MOLAP bekerja dengan database multidimensional proprietary yang menerima umpan data dari data warehouse. ROLAP menyediakan kemampuan pemrosesan analitik online dari database relasional dari data warehouse itu sendiri.

10. Alert Systems

- Fokus pada pengguna berdasarkan pengecualian yang didefinisikan.
- Memberikan peringatan dari database data warehouse untuk mendukung keputusan strategis.

- Tiga jenis peringatan dasar adalah : dari sumber sistem individu, dari data warehouse perusahaan yang terintegrasi, dan dari data mart individu.
11. Middleware and Connectivity
- Akses transparan ke sistem sumber di lingkungan yang heterogen.
  - Akses transparan ke database dari berbagai jenis platform.
  - Alat cukup mahal tetapi terbukti tidak ternilai untuk menyediakan interoperabilitas di antara berbagai komponen data warehouse.
12. Data Warehouse Administration
- Membantu administrator data warehouse dalam manajemen sehari-hari.
  - Beberapa alat fokus pada proses pemuatan dan melacak beban riwayat.
  - Alat-alat lain melacak jenis dan jumlah permintaan pengguna.

#### **4.2.4 Network Software**

Network Software (Perangkat Lunak Jaringan), jika dilakukan upgrade baik perangkat keras dan perangkat lunak dengan cara sedikit demi sedikit membuka masalah kompatibilitas. Ini karena perangkat keras dan perangkat lunak tidak dirancang khusus dan disesuaikan untuk data warehouse dengan volume data yang

besar dan mekanisme pengiriman intelijen bisnis yang kompleks. Akibatnya, organisasi mulai menggunakan apa yang dikenal sebagai peralatan data warehouse. Alat data warehouse adalah perangkat dengan komponen perangkat keras dan perangkat lunak yang dirancang khusus dan diarsipkan untuk data warehouse. Melalui arsitekturnya, alat data warehouse mengintegrasikan perangkat keras prosesor,

penyimpanan data, sistem operasi, dan perangkat lunak database menjadi satu kesatuan yang utuh.

**Soal latihan :**

1. Jelaskan infrastruktur fisik data warehouse ?
2. Jelaskan pilihan umum *Hardware Data Warehouse* ?
3. Jelaskan yang dimaksud dengan *platform* dan pilihan *platform* pada *data warehouse* ?
4. Jelaskan tugas *Data Warehouse Management* ?
5. Jelaskan tugas dari *Middleware* dan *Connectivity* ?

# BAB V

## DATA MULTIDIMENSIONAL

**P**ada Bab sebelumnya telah dijelaskan tentang infrastruktur *data warehouse*, dan pada Bab ini akan dijelaskan tentang *data*, *database*, *metadata*, *data multidimensi*, *pemodelan data multidimensi*, *Cube* (Kubus) dan *Spreadsheet*.

### 5.1 Definisi data

Definisi data menurut Rainer dan Cegielski (2011) merupakan deskripsi dasar mengenai sesuatu, peristiwa, aktifitas, dan transaksi yang dicatat, diklasifikasikan, dan disimpan tapi tidak terorganisasi dalam menghasilkan suatu makna yang spesifik. Keterlibatan data dalam pelaksanaan operasional perusahaan saat ini telah berperan penting mendukung setiap kegiatan yang dilakukan. Oleh karena itu, diperlukan penampung dalam mengintegrasikan data-data yang ada. Sistem yang dapat mendukung kegiatan operasional perusahaan ini dikenal dengan sistem *database*.

## 5.2 Definisi Database

Menurut Connolly dan Begg (2010), *database* adalah sekumpulan data dan deskripsi dari data yang berhubungan secara logikal didesain untuk memenuhi kebutuhan perusahaan. Melalui pengaplikasian sistem *database* tentunya dapat mempermudah *user* dalam memproses data, menampilkan data, serta mengurangi kemungkinan kesalahan yang akan timbul dalam pelaksanaan operasional perusahaan.

Seiring dengan berjalananya waktu, jumlah transaksi mengalami peningkatan yang cukup signifikan, hal ini ditandai dengan bertambahnya jumlah *record* pada tiap tabel yang terdapat dalam *database*. Peningkatan data tersebut dapat menimbulkan penurunan performa pada sistem operasional perusahaan, sehingga upaya mempertahankan data historis dinilai tidak sebanding dengan manfaat yang ada karena data tersebut akan menghambat kegiatan operasional yang berlangsung. Namun jika dilihat dari sisi lain, keberadaan data historis dapat dijadikan sebagai acuan dalam mengetahui pola atau kecenderungan yang akan terjadi di masa yang akan datang berdasarkan data yang sudah terjadi. Melalui penerapan sistem *data warehouse* dapat memberikan dampak positif bagi perusahaan, diantaranya proses analisis ataupun pengelolaan informasi berdasarkan data historis yang terdapat di perusahaan menjadi lebih mudah, sehingga proses pengambilan keputusan strategis dapat dilakukan dengan cepat berdasarkan proses penganalisisan yang akurat karena didasarkan pada data historis yang telah terjadi selama ini.

## 5.3 Metadata

Metadata dalam *data warehouse* digunakan untuk menyimpan catatan mengenai data warehouse itu sendiri.

Adapun definisi lain tentang metadata adalah :

- Catatan mengenai data Data about the data
- Daftar isi dari data

- Catalog dari data
- Roadmap Data warehouse roadmap
- Directory Data warehouse
- Glue that holds the data warehouse
- Contents together
- Tongs to handle the data
- The nerve center

Entity Name: Customer
Alias Names: Account, Client

**Definition:** A person or an organization that purchases goods or services from the company.

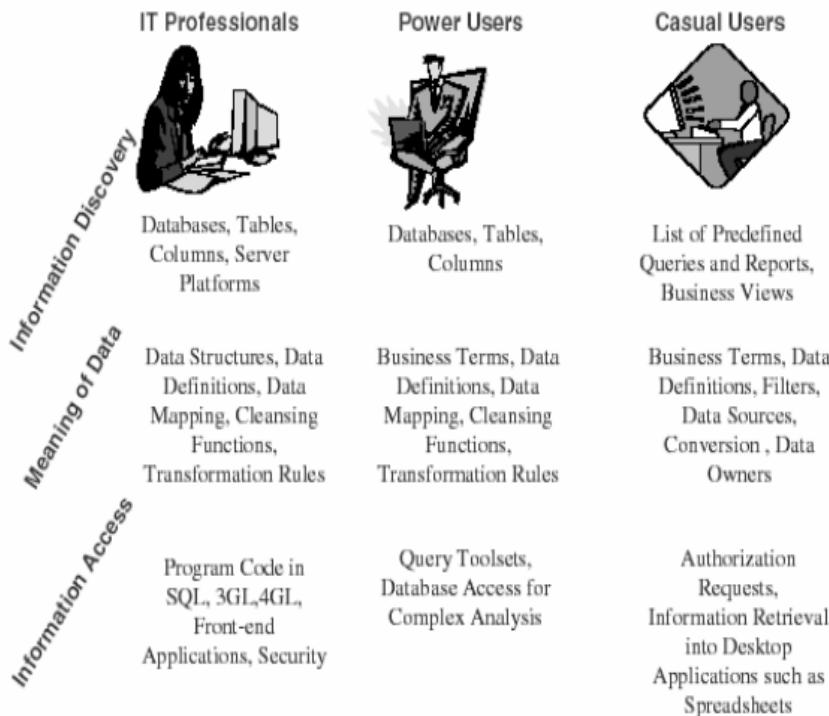
**Remarks:** Customer entity includes regular, current, and past customers.

**Source Systems:** Finished Goods Orders, Maintenance Contracts, Online Sales.

<b>Create Date:</b>	January 15, 1999
<b>Last Update Date:</b>	January 21, 2001
<b>Update Cycle:</b>	Weekly
<b>Last Full Refresh Date:</b>	December 29, 2000
<b>Full Refresh Cycle:</b>	Every six months
<b>Data Quality Reviewed:</b>	January 25, 2001
<b>Last Deduplication:</b>	January 10, 2001
<b>Planned Archival:</b>	Every six months
<b>Responsible User:</b>	Jane Brown

Gambar 5.1 Contoh Metadata (Firdaus Solihin, 2009)

Profesi yang membutuhkan metadata antara lain seperti pada gambar berikut :



Gambar 5.2 Profesi yang membutuhkan Metadata (Firdaus Solihin, 2009)

#### 5.4 Data Multidimensi

Pada sistem OLTP ( *Online Transactional Processing* ) digunakan suatu teknik pemodelan data yang disebut sebagai E-R (Entity-Relationship).

Tabel relasional terdiri atas baris dan kolom, seperti pada gambar berikut :

StudentID	Student	BirthDate
61521	Jones	12/05/77
61300	Devi	04/28/78
61425	Tony	11/01/78
61230	Paula	06/14/77

SC1 table.

CourseID	Course	Credit
MAT231	Math	3
ECO220	Economic	3
HST211	History	2
ENG202	English	2

SC2 table.

StudentID	CourseID	Grade	Weight
61521	MAT231	B	3
61521	ECO220	A	4
61521	HST211	B	3
61300	HST211	A	4
61425	ENG202	C	2
61425	MAT231	B	3
61230	MAT231	B	3
61230	ENG202	C	2

SC3 table.

Gambar 5.3 Tabel Relasional (Firdaus Solihin, 2009)

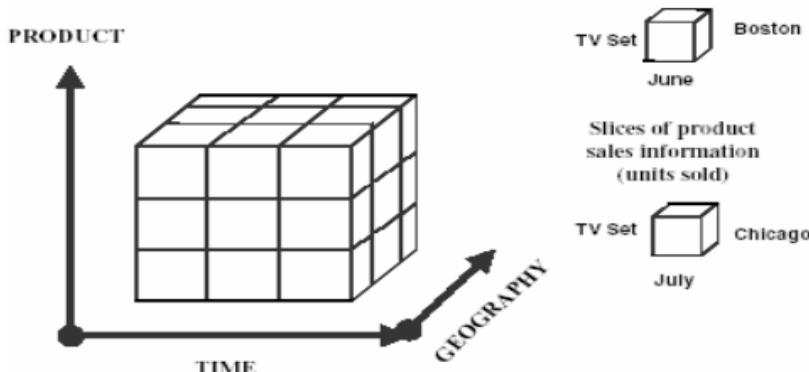
Adapun penjelasan gambar di atas adalah sebagai berikut :

- Setiap record atau baris merepresentasikan data student yang berbeda-beda.
- Satu baris dengan baris lainnya diidentifikasi dengan sebuah key yaitu *primary key*.
- Kolom seperti student, birthdate, menyimpan fakta yang sama atau sejenis, dimana setiap fakta tersebut merujuk pada *primary key* yaitu Student ID.
- Hal inilah yang menunjukkan bahwa tabel relasional hanya mempunyai satu dimensi.

Sedangkan pada data *warehouse* digunakan teknik pemodelan data yang disebut *dimensional modelling technique*. Pemodelan dimensional adalah suatu model berbasis pemanggilan yang mendukung akses *query volume* tinggi.

OLAP (Online Analytical Processing) sering juga disebut analisis data multidimensional. Sedangkan definisi data multidimensional yaitu : data yang dapat dimodelkan sebagai atribut dimensi (dimension) dan atribut ukuran (fact/measure). OLAP dapat digunakan untuk melakukan :

1. Konsolidasi adalah melibatkan pengelompokan data.
2. Drilldown adalah suatu bentuk yang merupakan kebalikan dari konsolidasi, yang memungkinkan data yang ringkas dijabarkan menjadi data yang lebih detail.
3. Slicing dan dicing (pivoting) adalah menjabarkan pada kemampuan untuk melihat data dari berbagai sudut pandang. Pivoting biasa dikenal dengan istilah cross tabulation, yaitu : sebuah tabel dengan nilai sebuah atribut (A) membentuk judul baris dan atribut (B) membentuk judul kolom, dan nilai masing-masing sel yang merupakan persilangan kedua atribut menyatakan nilai perpaduan untuk kedua atribut.



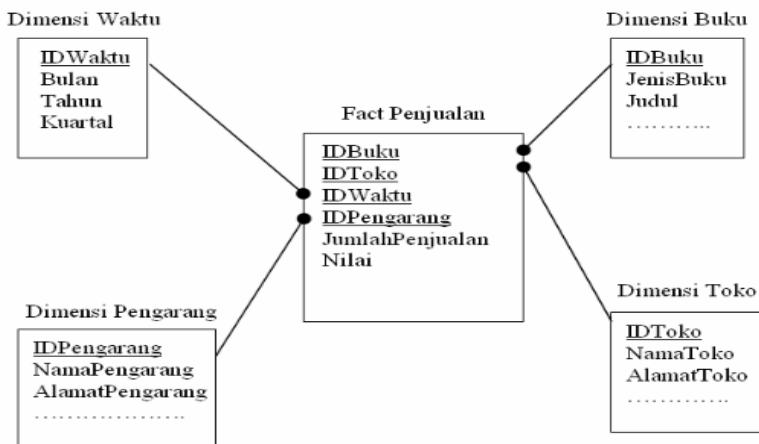
Gambar 5.4 Data Multidimensi (Firdaus Solihin, 2009)

## 5.5 Pemodelan Data Multidimensi

### 5.5.1 Star Schema

Star Schema adalah alat dimana pemodelan dimensional berisi sebuah tabel fakta pusat. Tabel fakta berisi atribut deskriptif yang

digunakan untuk proses *query* dan *foreign key* untuk menghubungkan ke tabel dimensi. Atribut analisis keputusan terdiri dari ukuran performa, metrik operasional, ukuran agregat, dan semua metrik yang lain yang di perlukan untuk menganalisis performa organisasi. Tabel fakta menunjukkan apa yang didukung oleh data *warehouse* untuk analisis keputusan. Tabel dimensi mengelilingi tabel fakta pusat. Tabel dimensi berisi atribut yang menguraikan data yang dimasukkan dalam tabel fakta. Tabel dimensi menunjuk bagaimana data akan dianalisis.



Gambar 5.5 Star schema pada *data warehouse* (Firdaus Solihin, 2009)

Tabel dimensi waktu adalah suatu tabel dimensi yang harus ada dalam sebuah data *warehouse* karena setiap data *warehouse* adalah *time series*. Waktu merupakan dimensi pertama yang harus diperhatikan dalam proses *sort order* dalam sesuatu *database* karena ketika hal itu pertama kali dilakukan, *loading* data secara berturut-turut dalam interval waktu tertentu akan masuk kedalam tempat yang sebenarnya dalam sebuah disk.

Data *designer* menyatakan secara eksplisit mengenai kebutuhan tabel dimensi waktu karena jika kunci waktu didalam tabel fakta adalah tipe data tanggal, beberapa SQL *query* mendapat kendala langsung dari kunci waktu pada tabel fakta dan menggunakan *natural SQL* untuk kendala pada bulan dan tahun.

Karakteristik Star Schema adalah sebagai berikut :

1. Pusat dari star disebut *fact table*
2. Fact table mempunyai sebuah nilai aggregate dari data-data yang berasal dari tabel dimensi
3. Setiap tabel dimensi berelasi langsung dengan *fact table*
4. Tabel dimensi beisikan data tentang informasi atau waktu
5. Relasi antara *fact table* dengan dimensi-dimensinya adalah 1 – N (*one to many*)
6. Primary key pada tabel dimensi akan menjadi key pada *fact table* atau dapat diakatakan bahwa *fact table* memiliki kombinasi key dari tabel dimensi tersebut.

KELEBIHAN model star adalah :

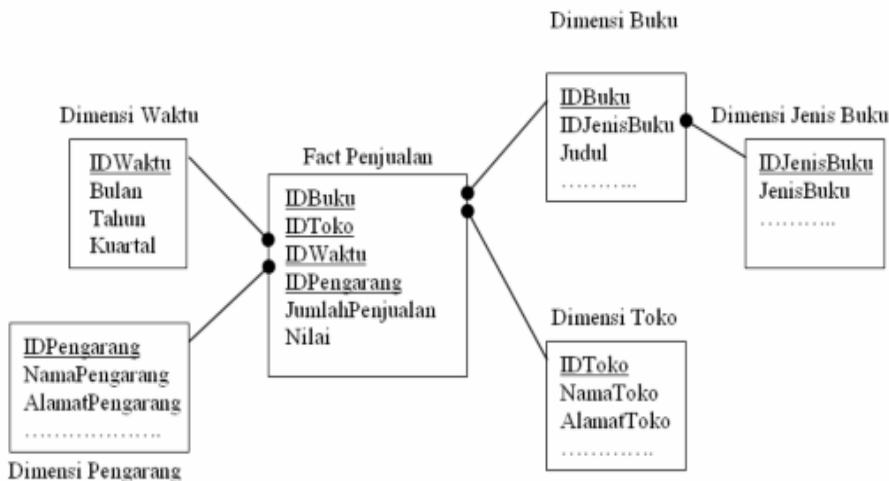
1. Lebih simple
2. Mudah dipahami
3. Hasil dari proses *query* juga relatif lebih cepat.

KEKURANGAN model star adalah : lebih boros dalam space.

### 5.5.2 Snowflakes Schema

Model snowflake merupakan perluasan dari star, yang sama-sama punya satu atau lebih dimensi. Pada snowflake, tabel yang berelasi pada fact table hanya tabel dimensi utama, sedangkan tabel yang lain dihubungkan pada tabel dimensi utama. Model snowflake ini hampir sama seperti teknik normalisasi.

Berikut adalah gambar pemodelan snowflakes schema :



Gambar 5.6 Snowflakes Schema (Firdaus Solihin, 2009)

KELEBIHAN model snowflakes schema adalah :

1. Pemakain space yang lebih sedikit
2. Update dan maintenance yang lebih mudah.

KEKURANGAN model snowflakes schema adalah :

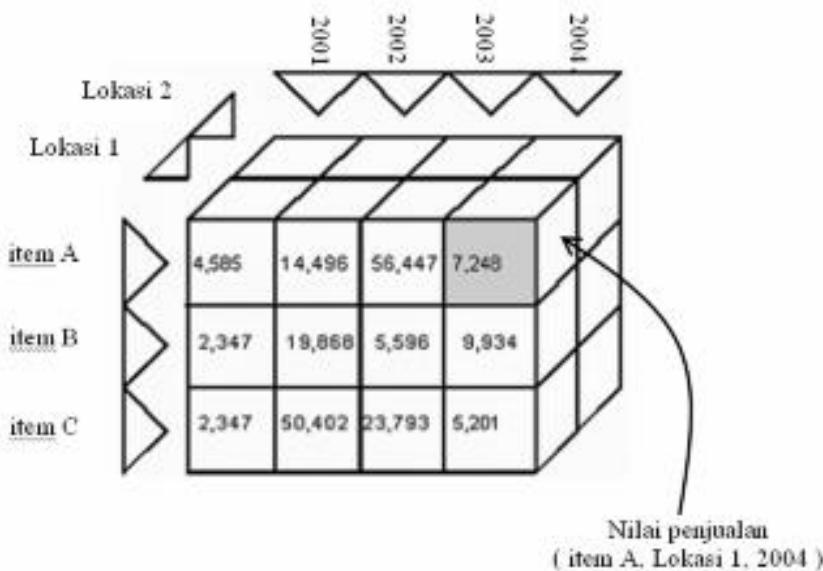
1. Model lebih komplek dan rumit
2. Proses *query* lebih lambat
3. Performance yang kurang bagus.

## 5.6 Cube (Kubus) dan Spreadsheet

### 5.6.1 Cube (Kubus)

#### 5.6.1.1 Bentuk Cube (Kubus) pada Data Warehouse

Cube adalah bagian utama dari OLAP. Cube berisi kumpulan banyak data yang telah disatukan (di agregasi) sehingga mempercepat hasil *query*.



Gambar 5.7 Cube (Firdaus Solihin, 2009)

Berikut ini terdapat bagian dari cube, yaitu :

- ♦ *Storage Location*

*Storage Location* (Lokasi penyimpanan) cube data dapat disimpan di sejumlah server analisis menggunakan distribusi baru dan dipartisi oleh fasilitas kubus (cube). Hal ini juga memungkinkan untuk penyebaran sejumlah cube di server, berbagai analisis dan akses semua hanya dengan menghubungkan ke salah satu server. Hal ini terkait teknologi cube memungkinkan keuntungan yang terbesar adalah untuk mengurangi biaya penyimpanan, karena sekarang masing-masing cube perlu disimpan hanya sekali (pada server tunggal).

- ♦ *Action*

*Action* adalah set operasi yang telah didefinisikan di awal. Sebagai pengguna dapat memanggil cube atau hanya digunakan untuk keperluan analisis. *Action* bahkan dapat melakukan tugas eksternal seperti mengirimkan parameter ke aplikasi.

- ◆ Dimension

Dimension adalah bagian dari *cube* yang menggambarkan suatu kategori data, yang digunakan sebagai dasar untuk merangkum data. Misalnya : dimensi waktu atau area. Dimensi biasanya memiliki hirarki atau jenjang (tingkatan). Dimensi waktu memiliki tingkatan berupa : tahun, bulan, minggu, hari dan sebagainya. Dimensi area memiliki tingkatan berupa : propinsi, kabupaten, kecamatan dan desa. Tabel yang digunakan sebagai dimensi disebut *dimension table*.

- ◆ Measure

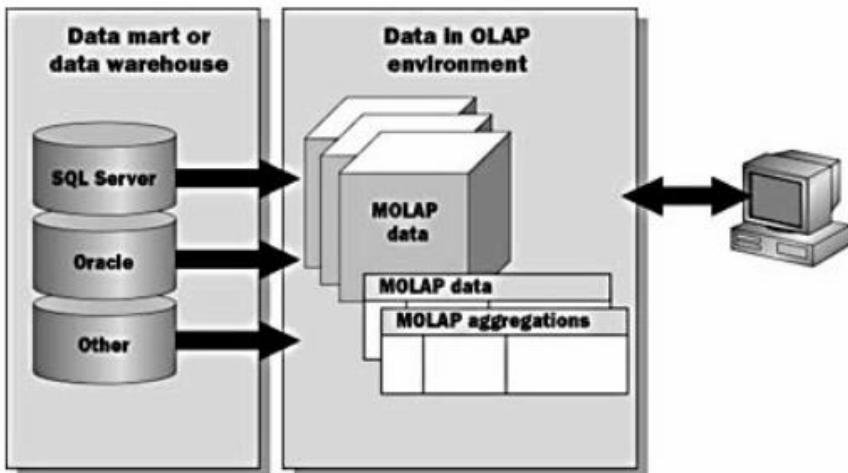
Measure adalah bagian dari *cube* yang berisi kumpulan data yang akan dirangkum, seperti data jumlah penduduk laki-laki, kepadatan penduduk, dan lain-lain. Umumnya data yang tergolong dalam *measure* ini tidak memiliki hirarki atau jenjang, dan dapat dilakukan proses matematis (penjumlahan, pengurangan, perkalian, dan lain-lain). Tabel yang berisi data-data dalam kelompok *measure* ini disebut dengan *Fact Table*.

#### 5.6.1.2 Model Pemrosesan Data dari Cube

Model pemrosesan data dari *cube* terdiri dari 3 macam, antara lain :

a. MOLAP

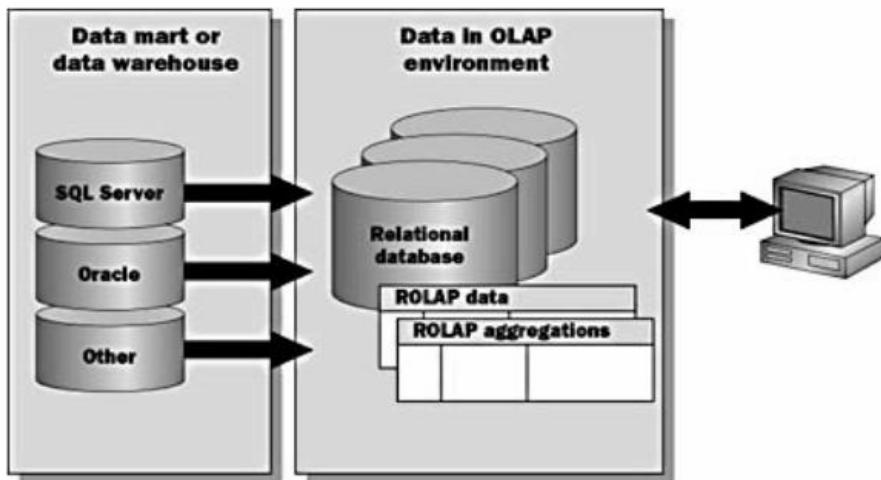
- ◆ Multidimensional online analytical processing (MOLAP) menyimpan data dan aggregasi pada struktur data multidimensi.
- ◆ Struktur MOLAP tidak tersimpan pada data warehouse tapi tersimpan pada OLAP server. Sehingga performa query yang dihasilkan olehnya sangat bagus.
- ◆ Model penyimpanan ini sesuai untuk database dengan ukuran kecil sampai sedang.



Gambar 5.8 Model MOLAP (Firdaus Solihin, 2009)

b. ROLAP

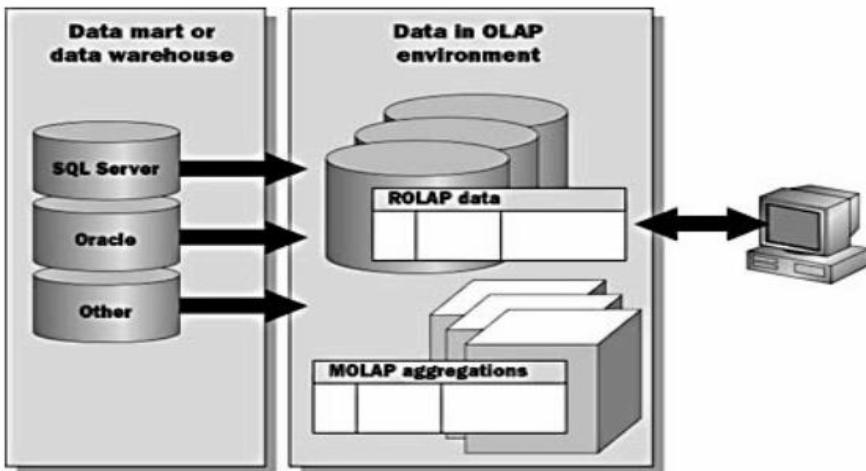
- ROLAP (Relational online analytical processing) menggunakan tabel pada database relasional datawarehouse untuk menyimpan detil data dan aggregasi kubus.
- ROLAP tidak menyimpan salinan database, dan mengakses langsung pada tabel fact ketika membutuhkan jawaban sebuah *query* sehingga *query* pada ROLAP mempunyai *response time* yang lebih lambat dibandingkan ROLAP maupun HOLAP
- Karakteristik model ini digunakan untuk menyimpan data yang besar dan jarang dilakukannya proses *query*. Misalkan, data histori dalam jumlah besar dari beberapa tahun yang sebelumnya.



Gambar 5.9 Model ROLAP (Firdaus Solihin, 2009)

c. HOLAP

- ◆ Gabungan model MOLAP dan ROLAP dapat kita peroleh dari HOLAP (Hibrid online analytical processing) Detil data tersimpan pada tabel relasional tapi aggregasi data disimpan dalam format multidimensi.
- ◆ Kubus HOLAP lebih kecil daripada kubus MOLAP tapi response time query masih lebih cepat jika dibandingkan dengan ROLAP. Penyimpanan HOLAP ini sesuai untuk kubus yang membutuhkan performa query yang bagus dengan jumlah data yang besar.



Gambar 5.10 Model HOLAP (Firdaus Solihin, 2009)

#### 5.6.1.3 Contoh Data Cube

- 0 dimensi:
  - Total penjualan sampai sekarang

Total	XXX
-------	-----

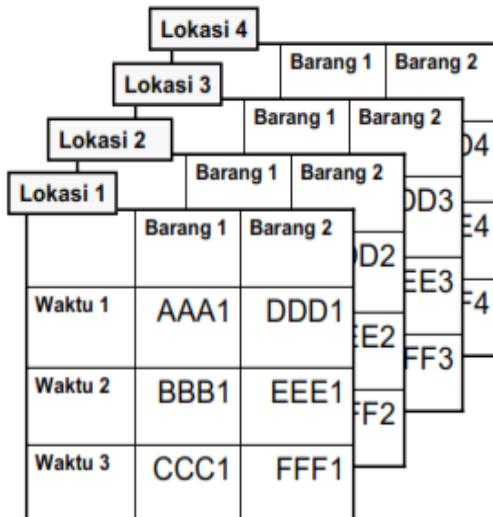
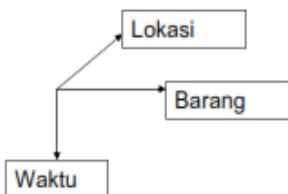
- 1 dimensi:
  - Total penjualan untuk **waktu** tertentu

<b>Waktu 1</b>	AAA
<b>Waktu 2</b>	BBB
<b>Waktu 3</b>	CCC

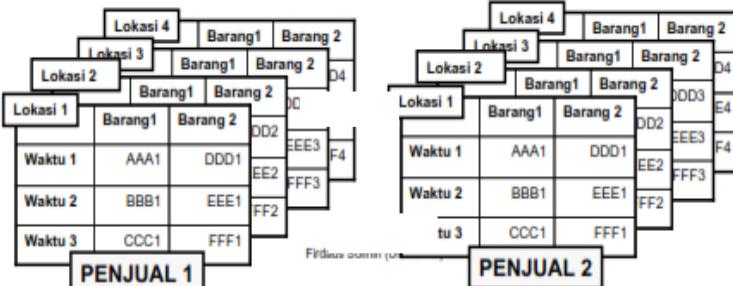
- 2 dimensi:
  - Total penjualan pada **waktu** tertentu untuk **barang**

	<b>Barang 1</b>	<b>Barang 2</b>
<b>Waktu 1</b>	AAA	DDD
<b>Waktu 2</b>	BBB	EEE
<b>Waktu 3</b>	CCC	FFF

- 3 dimensi:
    - Total penjualan pada **waktu** tertentu untuk **barang** pada **lokasi** tertentu



- 4 dimensi:
    - Total penjualan untuk barang **penjual** tertentu



Gambar 5.11 Contoh Data Cube (Firdaus Solihin, 2009)

## 5.6.2 Spreadsheet

Definisi Spreadsheet adalah dokumen yang menyimpan data dalam grid baris (row) horizontal dan kolom (column) vertikal. Baris biasanya diberi label menggunakan nomor (1, 2, 3, dan seterusnya), sementara kolom diberi label dengan huruf (A, B, C, dan seterusnya).

Lokasi baris / kolom individu, seperti C3 atau B12, disebut sebagai sel (cell). Informasi dapat disimpan dengan cara yang lebih terstruktur daripada menggunakan teks biasa struktur baris / kolom juga memungkinkan data yang akan dianalisis menggunakan Formula.

Dilakukan tinjauan pustaka yang dipusatkan pada percobaan penggunaan spreadsheet dalam pengajaran SPK berorientasi terapan untuk mahasiswa peminatan SI atau ekivalennya. Tinjauan tersebut menunjukkan bahwa mahasiswa menganggap bahan yang diberikan dalam pengajaran tersebut menarik dan bermanfaat bagi mereka setelah nantinya lulus (Lucky E. Santoso, 2008).

Daftar Aplikasi Spreadsheets yang Paling Populer sebagai berikut :

1. Microsoft Excel – Offline
2. Apache OpenOffice Calc – Offline
3. LibreOffice Calc – Offline
4. Google Sheets – Online
5. Zoho Sheets – Online.

Microsoft Excel masih menguasai seluruh Dunia dalam software Spreadsheet, meskipun saingannya gratis dan memiliki fitur yang sama, namun karena kebiasaan maka pengguna sulit untuk meninggalkan aplikasi Spreadsheet buatan Microsoft ini.

Lokasi1	Jan 01	Jan 02	---	Jan 04	
Lokasi2	Jan 01	Jan 02	---	Jan 04	
Item A			---		
Item B			---		
Item C			---		

Gambar 5.12 Contoh Spreadsheet (Firdaus Solihin, 2009)

**Soal latihan :**

1. Apa yang dimaksud dengan data, database dan metadata, jelaskan !
2. Jelaskan teknik pemodelan data pada OLTP dan pada OLAP ?
3. OLAP dapat digunakan untuk melakukan 3 hal antara lain : Konsolidasi, Drilldown, Slicing & Dicing, jelaskan !
4. Jelaskan Pemodelan data multidimensi(Star schema & Snowflakes schema)?
5. Jelaskan contoh pemodelan data multidimensi (Cube & Spreadsheet) ?
6. Model pemrosesan data dari cube terdiri dari 3 macam, jelaskan !

## BAB VI

# DESAIN DATA WAREHOUSE

**P**ada Bab sebelumnya sudah dibahas tentang *data*, *database*, *metadata*, *data multidimensi*, pemodelan *data multidimensi*, *Cube* (Kubus) dan *Spreadsheet*. Pada Bab ini akan dibahas tentang Desain Data Warehouse, Aspek yang perlu dipertimbangkan (saat mendesain, saat implementasi dan saat testing).

### **6.1 Desain Data Warehouse**

Desain *data warehouse* terdiri dari dua jenis, antara lain sebagai berikut :

1. Logical Design, tahapan yang dilakukan adalah :
  - Menggunakan ER Diagram
  - Mencari objek-objek penting (entitas)
  - Mendefinisikan attribute
  - Membuat relasi

2. Physical Design, tahapan yang dilakukan adalah :
  - Merubah dari logical design menjadi objek-objek database
  - Tablespace, Table, integrity constraint, dan lain-lain.

## 6.2 Aspek yang perlu dipertimbangkan

Aspek yang perlu dipertimbangkan saat membuat *data warehouse* antara lain :

1. Saat Desain
2. Saat Implementasi
3. Saat Testing.

### 6.2.1 Saat Desain

Hal-hal yang perlu dipertimbangkan saat mendesain *data warehouse* antara lain :

1. Desain untuk Pengelolaan
  - a. Mudah di **Backup** secara Teratur
  - b. Ketika **Loading** new data
  - c. Ketika **Aggregating** new data
  - d. Ketika Melakukan Aktifitas **Pemeliharaan** Data,  
Contoh : Indexing dan Archiving
2. Desain untuk Performa

Tentukan tipe, dimana, berapa banyak ruang yang dibutuhkan untuk **indeks**.

### **6.2.2 Saat Implementasi**

Hal-hal yang perlu dipertimbangkan sebelum mengimplementasikan *data warehouse* antara lain :

1. Satu *database* atau lebih
2. Kesepakatan dalam aturan penamaan
3. Membuat *database*
4. Menentukan *skema* untuk *database*
5. Mengatur *data file* dan *tablespace*
6. Membuat *tabel fact* dan *tabel dimensi*
7. Konstrain
8. Indeks
9. Partisi
10. Membuat View
11. Keamanan.

### **6.2.3 Saat Testing**

Hal-hal yang perlu dipertimbangkan saat melakukan testing *data warehouse* antara lain :

1. Dilakukan sebelum rilis produksi.
2. Yang perlu di tes antara lain:
  - a. Waktu yang dibutuhkan untuk melakukan load data
  - b. Pembersihan data dan transformasi
  - c. Waktu respon *query*
  - d. Data *summary* yang dibutuhkan
  - e. Waktu yang dibutuhkan untuk tugas-tugas pengelolaan (manajemen).

**Soal latihan :**

1. Desain *data warehouse* terdiri dari dua jenis, jelaskan !
2. Jelaskan aspek yang perlu dipertimbangkan saat membuat *data warehouse* ?
3. Jelaskan hal-hal yang perlu dipertimbangkan saat mendesain untuk pengelolaan *data warehouse* ?
4. Jelaskan hal-hal yang perlu dipertimbangkan saat implementasi *data warehouse* ?
5. Saat melakukan testing, jelaskan hal-hal yang perlu dipertimbangkan !

## BAB VII

# MEMBANGUN DATA WAREHOUSE

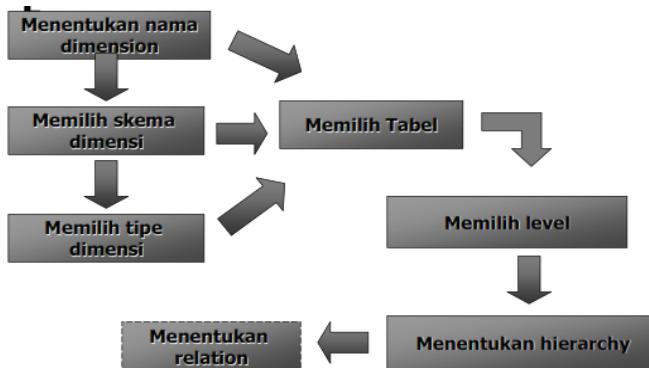
**P**ada Bab sebelumnya telah dibahas tentang Desain Data Warehouse, Aspek yang perlu dipertimbangkan (saat mendesain, saat implementasi dan saat testing), dan pada Bab ini akan dibahas tentang beberapa pendapat para pakar dalam membangun *data warehouse* dengan menggunakan langkah, metode dan tahapan, antara lain sebagai berikut :

### 7.1 Empat Langkah Membangun Data Warehouse

Menurut Firdaus Solihin (2009), ada 4 langkah membangun *data warehouse* antara lain sebagai berikut :

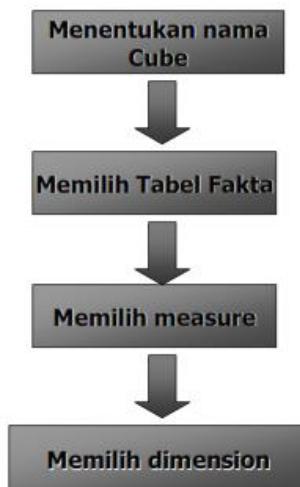
1. Desain Data Warehouse  
*Desain Logical dan Desain Physical.*
2. ETL (Extract, Transformation dan Loading)
  - a. Proses pengekstrakan data dari sumber data yang kemudian dimasukkan ke dalam *data warehouse*.
  - b. Dilakukan secara periodik untuk kebutuhan bisnis dengan analisa data yang akurat.

- c. Menggunakan Data Transformation Service (DTS),
  - d. Menggunakan SQL\* Loader.
3. Membuat Dimensi (*Create Dimension*)
- a. Dimension adalah sebuah struktur yang terbentuk dari satu atau lebih hierarki yang mengkategorisasi data.
  - b. Dimensi terbentuk dari satu atau lebih tabel. Setiap kolomnya merepresentasikan level pada hierarchy.



Gambar 7.1 Alur pembuatan dimensi (Firdaus Solihin, 2009)

4. Membuat Kubus (*Create Cube*)



Gambar 7.2 Alur pembuatan Cube (Firdaus Solihin, 2009)

## 7.2 Enam Tahap Membangun *Data Warehouse*

Menurut pendapat M. Yazdi Pusadan (2013), membangun *data warehouse* melalui beberapa tahap antara lain sebagai berikut :

1. Menentukan subyek *data warehouse*

Subyek *data warehouse* yang di buat adalah *data warehouse* yang menjadi data history untuk di analisis.

2. Mendefinisikan kebutuhan sumber data dan kebutuhan informasi

- a. Sumber Data

Kebutuhan sumber data adalah data transaksi (OLTP) yang berlangsung selama proses operasional *database*. Adapun data OLTP tersebut adalah *database* yang dikelompokkan berdasarkan mekanisme pengolahan data transaksi.

- b. Kebutuhan Informasi

Mendefinisikan kebutuhan informasi diperoleh dari hasil pengamatan dan wawancara langsung manajemen perusahaan.

3. Merancang Star Schema

Dengan memperhatikan struktur kolom yang terdapat pada setiap tabel, maka diperoleh aktivitas transaksi data yang tidak efisien. Terjadi kecenderungan pada aktivitas data yang berlangsung berulang kali. Hal ini memungkinkan terjadinya kesalahan jika tidak teliti dalam proses entry data sehingga berakibat fatal terhadap proses transaksi dan mekanisme pengambilan keputusan.

4. Desain ETL

Terdapat beberapa tahapan dalam desain ETL, yaitu :

- a. Pembersihan data (*cleansing*) dan mereduksi beberapa kolom pada tabel-tabel tertentu yang terdapat di *database* transaksi.

Mereduksi (*drop*) data sesuai kebutuhan informasi yang telah ditentukan.

- b. Membuat tabel baru yang berfungsi sebagai relasi (keterhubungan) antar satu tabel dengan tabel lainnya.
  - c. Melakukan proses ETL dengan menggunakan tools yang sesuai dan tepat.
5. Membuat Cube dan Dimensional Data

Pada pembuatan *cube*, *data warehouse* yang terbentuk digunakan sebagai *data source*. Tahapan pada pembuatan *cube* adalah sebagai berikut :

- a. Mendefinisikan *cube*
  - b. Menentukan *measure* yang menjadi *fact table*
  - c. Mengelompokkan beberapa tabel yang termasuk *dimension*
6. Analisis Data Warehouse

Pada analisis *data warehouse* berikut ini adalah mengolah *data warehouse* dan menyajikan data hasil analisis dalam bentuk OLAP dan pivot table :

- ◆ Konfigurasi *cube* kebutuhan informasi. OLAP dilakukan dengan mengkonfigurasi tabel-tabel yang terdapat pada *cube*, dalam hal ini tabel *measure/fact* dan tabel *dimension*. Analisis yang dilakukan dengan memperhatikan kebutuhan informasi oleh *user*.
  - ◆ View *cube* hasil analisis *data warehouse* versi OLAP Application.
  - ◆ View *cube* hasil analisis *data warehouse* versi excel application.
- a. Membuat Aplikasi OLAP Data Warehouse

Aplikasi OLAP dibuat dengan menggunakan DBMS (*Database Management System*) yang telah diproses melalui ETL. Selain

itu aplikasi OLAP menghasilkan output dalam bentuk *report* yang digunakan sebagai dokumen *blue print* analisis kinerja maupun analisis lain dalam pengambilan keputusan pihak pimpinan atau manajemen.

b. Penjelasan Implementasi Data Warehouse

Penjelasan dari tahap implementasi adalah sebagai berikut :

- ♦ Menganalisis data dan struktur tabel pada *database OLTP* yang berelasi kemudian direduksi atau menghapus tabel (*drop table*). Hal ini dilakukan dengan pertimbangan untuk normalisasi tabel sehingga tidak terjadi redundansi.
- ♦ Akibat dari relasi tabel yang banyak, penggunaan *field kunci* pada tabel makin sering. Maka dibuatlah satu tabel dengan mendaftarkan *field kunci* tabel-tabel yang ada untuk kebutuhan relasi dan meringkas *statement query* untuk akses data.
- ♦ Membentuk *database* yang baru.
- ♦ Melakukan proses ETL dari *database* yang telah dimodifikasi dengan menggunakan tool ETL.
- ♦ Membuat *cube* dalam bentuk *star schema* yang terdapat dalam *fact table*.
- ♦ Analisis data *warehouse* dengan OLAP yang menghasilkan multidimensional data dengan ditampilkan pada *pivot table* (*Cross Tabulation*).
- ♦ Membuat aplikasi OLAP yang menghasilkan informasi data berdasarkan waktu transaksi.

### 7.3 Metodologi Sembilan Tahap (Nine-step methodology)

Metode perancangan *data warehouse* menurut Kimball yang digunakan meliputi 9 (sembilan) tahap yang dikenal dengan *nine-step methodology* (Connolly dan Begg, 2005). Kesembilan tahap tersebut yaitu :

1. Pemilihan Proses

Proses mengacu pada subjek masalah dari bagian *data mart*. *Data mart* yang akan dibangun harus sesuai anggaran dan dapat menjawab masalah-masalah bisnis yang penting. Pemilihan proses ini dilakukan untuk memperjelas batasan mengenai *data warehouse* yang dibuat.

2. Pemilihan Grain

Pemilihan *grain* berarti menentukan secara tepat apa yang dipresentasikan oleh *record* pada tabel fakta.

3. Identifikasi Dari Penyampaian Dimensi

Pada tahap ini dilakukan penyesuaian dimensi dan *grain* yang ditampilkan dalam bentuk matriks.

4. Pemilihan Fakta

*Grain* dari tabel fakta menentukan fakta yang bisa digunakan.

5. Penyimpanan Pre-Calculation di Tabel Fakta

Setelah fakta-fakta dipilih, maka dilakukan pengkajian ulang untuk menentukan fakta-fakta yang dapat diterapkan untuk kalkulasi awal.

6. Memastikan Tabel Dimensi

Dalam tahap ini, kembali pada tabel dimensi dan menambahkan gambaran teks terhadap dimensi yang memungkinkan. Gambaran teks harus mudah digunakan dan dimenerti oleh *user*.

## 7. Pemilihan Durasi Database

Pemilihan durasi data histori yang dimiliki oleh rumah sakit dapat dilakukan sesuai dengan kebutuhan informasi. Umumnya semakin banyak data yang dipindahkan ke dalam *data warehouse* semakin lengkap pula informasi yang bisa dihasilkan. Perlu diperhatikan pula tingkat durasi yang dimiliki oleh data histori dengan memperhatikan isi dan format data yang ada. Jangan sampai data yang dipindahkan merupakan data sampah yang tidak bermanfaat sama sekali.

## 8. Melacak Perubahan Dari Dimensi Secara Perlahan

Mengamati perubahan dari dimensi pada tabel dimensi dapat dilakukan dengan tiga cara, yaitu mengganti secara langsung pada tabel dimensi, membentuk *record* baru untuk setiap perubahan baru dan perubahan data yang membentuk kolom baru yang berbeda.

## 9. Penentuan Prioritas dan Model Query

Mempertimbangkan pengaruh dari rancangan fisik, seperti penyortiran urutan tabel fakta pada disk dan keberadaan dari penyimpanan awal ringkasan (*summaries*) atau penjumlahan (*aggregate*).

### **Soal latihan :**

Lakukan penelitian dengan memilih metode atau tahapan yang sesuai dengan obyek penelitian yang anda tentukan.

## BAB VIII

# IMPLEMENTASI DATA WAREHOUSE

**P**ada Bab sebelumnya telah dibahas tentang beberapa pendapat para pakar dalam membangun *data warehouse* dengan menggunakan langkah, metode dan tahapan, dan pada Bab ini akan dibahas tentang mengimplementasikan semua teori yang telah dibahas sebelumnya.

### 8.1 Implementasi

Pada pemaparan berikut digunakan data transaksi nilai akademik mahasiswa dari hasil penelitian Neni Purwati dan Hariyanto Wibowo. Berikut implementasi data warehouse menggunakan metode *nine-step* Kimball, ada 9 tahapan yang akan dilakukan agar *data warehouse* sesuai dengan kebutuhan :

#### 1. Pemilihan Proses

Dalam tahap ini, ditentukan pada proses bisnis apa *data warehouse* akan digunakan. Berdasarkan data mahasiswa yang telah dinyatakan lulus ujian skripsi atau tugas akhir, kemudian dilakukan rekapitulasi oleh BAAK dan dilakukan musyawarah pada rapat akademik untuk menentukan mahasiswa yang layak

diyudicium dengan diterbitkannya surat keputusan rektor. Adapun proses-proses yang telah dipilih yaitu proses alumni.

## 2. Pemilihan Grain

Pemilihan *grain* berarti menentukan secara tepat apa yang dipresentasikan oleh *record* pada tabel fakta. Adapun grain dalam rancangan *data warehouse* mahasiswa alumni adalah NPM, Nama, Tempat\_Lahir, Tanggal\_Lahir, Jenis\_Kelamin, Jurusan, IPK, Tanggal\_Lulus.

## 3. Identifikasi Dan Penyesuaian Dimensi

Dalam tahap ini, ditentukan dan dibuat set dimensi yang dibutuhkan untuk menjawab seluruh pertanyaan yang diajukan pada tabel fakta. Adapun tabel dimensi yang diperlukan adalah dimensi mahasiswa, dimensi Jurusan dan dimensi nilai, dimensi waktu.

## 4. Pemilihan Fakta

Grain dari tabel fakta menentukan fakta yang bisa digunakan. Jumlah tabel fakta yang dibuat berdasarkan jumlah proses yang telah dipilih yaitu tabel fakta alumni.

## 5. Penyimpanan Pre-Kalkulasi di Tabel Fakta

Setelah fakta-fakta dipilih, maka dilakukan pengkajian ulang untuk menentukan apakah ada fakta-fakta yang dapat diterapkan untuk kalkulasi awal. Pre-kalkulasi yang akan disimpan dalam tabel fakta proses alumni meliputi : Jumlah lulusan reguler, yang merupakan jumlah dari mahasiswa yang telah lulus.

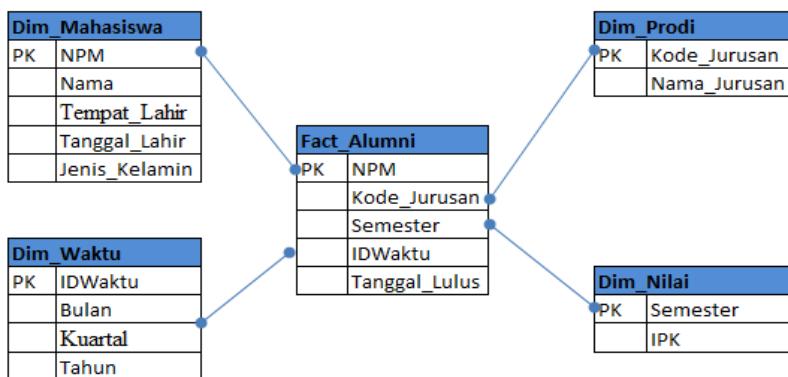
## 6. Memastikan Tabel Dimensi

Dalam tahap ini, akan diberikan deskripsi pada tabel dimensi agar dapat menjelaskan dengan mudah kepada *user* dan mudah dimengerti oleh *user* mengenai dimensi tersebut.

- a. Tabel dimensi Mahasiswa : NPM, Nama, Tempat\_Lahir, Tanggal\_Lahir, Jenis\_Kelamin.

- b. Tabel dimensi Jurusan : Kode\_Jurusan, Nama\_Jurusan
- c. Tabel dimensi Nilai : Semester, IPK (Index Prestasi Kumulatif).
- d. Tabel dimensi Waktu : Bulan, Kuartal, dan Tahun.

Skema yang digunakan dalam penelitian perancangan *data warehouse* ini adalah skema bintang (Star Schema). Seperti yang terdapat dalam tabel fakta alumni dikelilingi oleh tabel dimensi mahasiswa, tabel dimensi prodi (program studi) dan tabel dimensi nilai dapat dilihat pada gambar sebagai berikut :



Gambar 8.1 Star Schema (Neni P & Hariyanto W, 2017)

## 7. Pemilihan Durasi Database

Pemilihan durasi data histori yang dimiliki oleh Informatics and Business Institute (IBI) Darmajaya yaitu Database Ada Sejak Tahun 2007, Data Alumni Sejak Tahun 2001-2016, Data Dalam *Data Warehouse* 15 Tahun.

## 8. Melacak Perubahan Dari Dimensi Secara Perlahan

Untuk mengantisipasi adanya perubahan atribut data yang mungkin terjadi pada database asal misalnya, seperti bertambahnya nomor induk mahasiswa pindahan atau transfer.

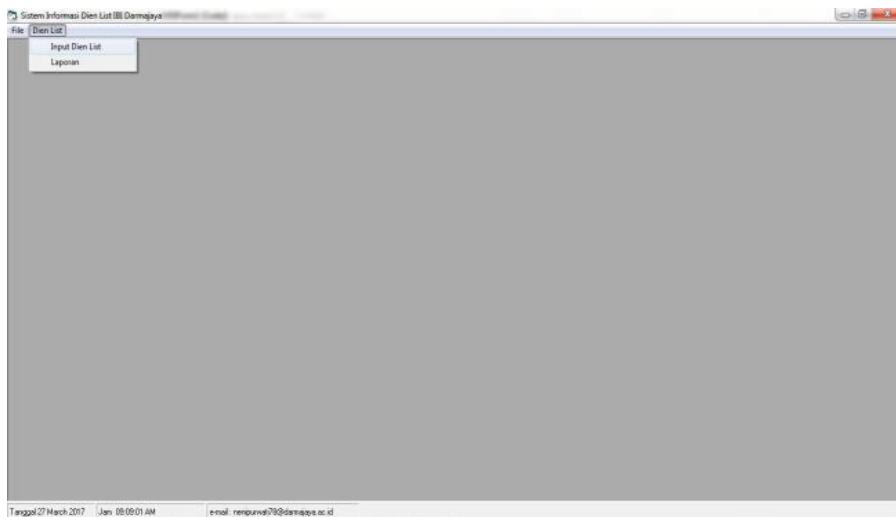
## 9. Penentuan Prioritas dan Model Query

Langkah ini berhubungan dengan *physical* terutama pengurutan data dan pencarian data melalui *indexing*. Pada *Data Warehouse* ini telah diurutkan berdasarkan *primary key* dari setiap tabel dan juga secara otomatis terindeks berdasarkan *primary key* tersebut.

### 8.2 Tampilan Program

#### 8.2.1 Tampilan Menu Program

Tampilan Menu Program *data warehouse* yang telah dibangun untuk dapat menampilkan Menu Dean List IPS(Indeks Prestasi Semester) dengan memilih dari dua sub menu Input Dean List dan Laporan pada program berikut :



Gambar 8.2 Tampilan Program Import IPS (Neni P & Hariyanto W, 2017)

#### 8.2.2 Tampilan Sub Menu Input Dean List

Pada sub menu Input Dean List ini dapat dilakukan proses pencarian IPS yang konsisten dari semester awal hingga semester akhir (lulus) mengikuti ketentuan Dekan untuk jumlah standar IPKnya menggunakan formula berikut :

$$\text{IPSx} = \text{IPS} \geq \text{IPSk}$$

Keterangan :

IPSx = Indeks Prestasi Semester dari semester awal hingga semester akhir (sesuai masa studi)

IPS = Indeks Prestasi Semester

IPSk = IPS Ketentuan

Adapun hasil data yang dicari untuk konsistensi IPS dengan menginputkan Semester, Tahun Akademik dan IPSk maka dapat dilihat pada gambar berikut :

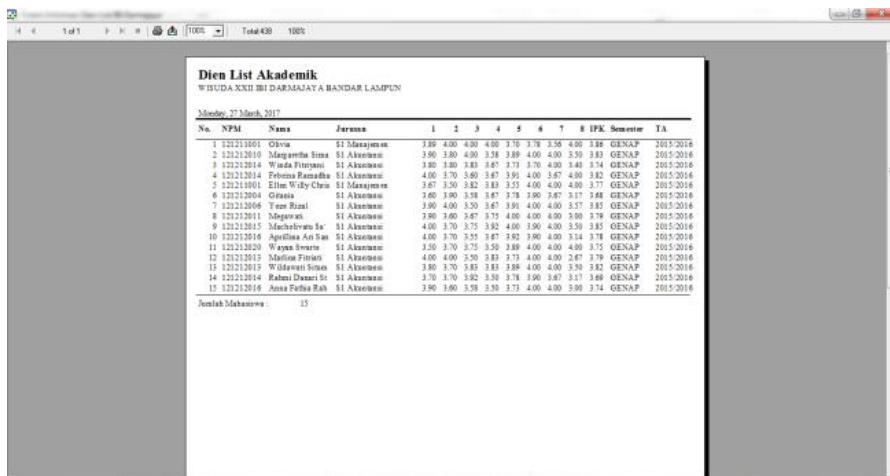
DAFTAR PERKuliahan GENAP (DARMAJAYA BANDAR LAMPUNG) SEMESTER GENAP						
NIM	nama	kodenrmatkul	semester	TA	Semester	IPS
121.2110018	Oleka	0211	Matematika	GENAP	2015/2016	3.05
121.2120014	Margaretha Suciawati	0212	Alkuntansi	GENAP	2015/2016	3.9
121.2120016	Yulianti Pajaya	0212	Alkuntansi	GENAP	2015/2016	3.8
121.2120019	Ferdiansyah	0212	Alkuntansi	GENAP	2015/2016	3.8
121.2110057	Ellen Willy Chintas	0211	Matematika	GENAP	2015/2016	3.67
121.2120044	Ukens	0212	Alkuntansi	GENAP	2015/2016	3.6
121.2120087	Yossi Rizal	0212	Alkuntansi	GENAP	2015/2016	3.9
121.2120119	Megawati	0212	Alkuntansi	GENAP	2015/2016	3.9
121.2120134	Mucholikra Gita Adhi	0212	Alkuntansi	GENAP	2015/2016	4
121.2120180	Apollina Aisy Sandy	0212	Alkuntansi	GENAP	2015/2016	3.7
121.2120205	Yulianti	0212	Alkuntansi	GENAP	2015/2016	3.6
121.2120319	Mulyana Rizal	0212	Alkuntansi	GENAP	2015/2016	4
121.2120332	Widjatmi Sutarmeng	0212	Alkuntansi	GENAP	2015/2016	3.8
121.2120414	Rahmat Darmi Shafiqah	0212	Alkuntansi	GENAP	2015/2016	3.7
121.2120181	Annisa Faizha Rahmat	0212	Alkuntansi	GENAP	2015/2016	3.9

Gambar 8.3 Tampilan Data Konsistensi IPS (Neni P & Hariyanto W, 2017)

### 8.2.3 Tampilan Sub Menu Laporan

Pada sub menu Laporan ini merupakan tampilan/view hasil inputan pencarian data dan dapat dilakukan pencetakan data.

Berikut sub menu Laporan dapat dilihat pada gambar berikut :



No.	NPM	Nama	Jurusan	1	2	3	4	5	6	7	8	IPK	Semester	TA	
1	1212131001	Oktiva	S1 Magister	3.99	4.00	4.20	4.00	3.70	3.78	3.56	4.00	3.84	GENAP	2015-2016	
2	1212131010	Margaretha Sinta	S1 Akademisi	3.90	3.80	4.00	3.59	3.89	4.00	4.00	3.59	3.81	GENAP	2015-2016	
3	1212132014	Winda Fitriyani	S1 Akademisi	3.80	3.80	3.83	3.87	3.73	3.76	4.00	3.40	3.74	GENAP	2015-2016	
4	1212132015	Wulan Faranda	S1 Akademisi	3.80	3.80	3.83	3.87	3.73	3.76	4.00	3.40	3.74	GENAP	2015-2016	
5	1212131001	Ellen Widy Chira	S1 Magister	3.67	3.50	3.82	3.83	3.53	4.00	4.00	4.00	3.77	GENAP	2015-2016	
6	1212132004	Gitaanu	S1 Akademisi	3.60	3.80	3.58	3.87	3.78	3.80	3.87	3.17	3.68	GENAP	2015-2016	
7	1212132006	Tens Rizal	S1 Akademisi	3.90	4.00	3.20	3.87	3.91	4.00	4.00	3.77	3.82	GENAP	2015-2016	
8	1212132007	Yudha Rizal	S1 Akademisi	3.80	3.80	3.83	3.87	3.73	3.76	3.80	3.40	3.74	GENAP	2015-2016	
9	1212132012	Machayuwan Is'	S1 Akademisi	4.00	3.70	3.75	3.92	4.00	3.90	4.00	3.59	3.85	GENAP	2015-2016	
10	1212132016	Agustina Ani San	S1 Akademisi	4.00	3.70	3.55	3.87	3.92	3.96	4.00	3.14	3.78	GENAP	2015-2016	
11	1212132020	Wayan Sri	S1 Akademisi	3.80	3.80	3.75	3.89	3.89	4.00	4.00	3.40	3.72	GENAP	2015-2016	
12	1212132021	Wulan Gunawati	S1 Akademisi	4.00	4.00	3.80	3.87	3.89	4.00	4.00	3.40	3.87	GENAP	2015-2016	
13	1212132013	Widawanu Sinten	S1 Akademisi	3.80	3.70	3.83	3.83	3.88	4.00	4.00	3.59	3.82	GENAP	2015-2016	
14	1212132014	Rahmi Diantri Sri	S1 Akademisi	3.70	3.70	3.92	3.59	3.78	3.90	3.87	3.17	3.69	GENAP	2015-2016	
15	1212132018	Anna Febha Rah	S1 Akademisi	3.90	3.80	3.58	3.59	3.73	4.00	4.00	3.80	3.74	GENAP	2015-2016	

Gambar 8.4 Tampilan Laporan Dean List (Neni P & Hariyanto W, 2017)

## DAFTAR PUSTAKA

---

Astria Yumalia , Richardus Eko Indrajit. 2017. Penerapan Konsep Business Intelligence Untuk Percepatan Penyelesaian Perkara Pada Panmud Perdata Khusus Mahkamah Agung Ri, *Ikraith-Informatika*, Vol. 1, No. 2, November 2017

Connolly Thomas, Carolyne Begg. 2005. *Database System : A Practical Approach to Design, Implementation, and Management*. Fourth Edition. English : Addison Wesley.

Connolly Thomas, Carolyn Begg. 2010. *Database Systems: A Practical Approach to Design, Implementation and Management*. 5th Edition. New Jersey, English : Pearson Education.

Curko, K., Varga, M., 2008. The Review of the Role of Business Intelligence in Business Engineering, Department of Business Computing University of Zagreb, Croatia.

David, Edward. 2000. "Business Intelligence, Have We Forgotten the Basics", white paper from [www.bitpipe.com](http://www.bitpipe.com).

DJ. Power. 2002. A Brief History of Decision Support Systems, [DSSResources.com](http://DSSResources.com).

Humphries, Hawkins, Dy. 1999. Sumber : *Data Warehouse Architecture and Implementation*. Prentice Hall.

<https://library.binus.ac.id/eColls/eThesisdoc/Bab2HTML/2011100638SIHCBAB2/page6.html>

Deliana et al, 2009. Diakses pada tanggal 4 Oktober 2017 pukul 16.30 WIB.

<http://www.academia.edu/5354417/Makalah-data-warehouse>  
Diakses pada tanggal 4 Oktober 2017 pukul 17.42 WIB.

<https://sis.binus.ac.id/2016/12/15/arsitektur-business-intelligence/>  
Diakses Tanggal 10 Maret 2017 Pukul 13.49 WIB

Imelda, S.T, M.T. 2013. Business Intellegence. 111- 122

Inmon WH. 2005. Building the Data Warehouse. Indiana : Willey.

Kamber, H. J., 2006. Data Mining:Concepts and Techniques. San Francisco: Morgan Kaufmann Publisher.

Kimbal Raplh, Ross Margy. 2013. The Data Warehouse Toolkit. Indiana : Willey.

Kimbal Raplh, Caserta Joe. 2004. The Data Warehouse ETL Toolkit. Indiana : Willey.

Laudon, Kenneth C, Jane P. 2008, "Essen-tials of Business Information Sys-tems. 7th Edition". Pearson Prentice Hall.

Lita Alfriany Ndoloe. 2012. Sistem Informasi Lulusan Dengan Metode Online Analitycal processing (OLAP) Pada Politeknik Negeri Kupang, Jurnal Sistem Informasi Bisnis.

Mike Steadman. 2003. The Value of BI for Association Executives, Association Xpertise Inc.

Moss, Larissa T., and Shaku Atre. 2003. Business Intelligence Roadmap: The Complete Lifecycle for Decision-Support Applications. Boston, MA: Addison-Wesley.

Mulyana JRP. 2014. Pentaho : Solusi Open Source untuk Membangun Data Warehouse. Yogyakarta : Andi Offset.

Multimatics. 2015. Buku Ajar : Pelatihan CIBIA(Certified International Business Intelligence Associate). Jakarta : Multimatics.

Purwati Neni, Wibowo Hariyanto. 2017. Pemanfaatan Data Warehouse Untuk Menentukan Penghargaan Dean List Akademik Pada Alumni. Bandar Lampung : Explore.

Nazwirman. 2014. Tahapan Pelaksanaan Sistem Intelijen Bisnis Dalam Perusahaan, Jurnal Lentera ICT Vol.2 No.1, November 2014/ISSN 2338-3143

Poe, Vidette. 1998. Building Data Warehouse for Decision Support. Edisi-2. Prentice Hall.

Ponniah Paulraj. 2010. Data Warehousing Fundamentals For IT Professionals. Indiana : Willey.

Power, D.J. 2002. "Decision support systems: Concepts and resources for managers". United State of America: Green Wood Publishing Group.

Pusadan Mohammad Yazdi. 2013. Rancang Bangun Data Warehouse. Yogyakarta : Graha Ilmu.

Rainer and Cegielski. 2011. Introduction to Information Systems. 3rd Edition. USA: Wiley.

Rajagukguk, Ronald. 2008. (e-book) "Quick Intro to Microsoft Office Performance Point Server 2007". ITB Bandung: MIC.

Ricky Akbar, Azizi Soniawan, Rafel Dinur, Jovi Adrian, Rafki Azim, Afdhal Zikri. 2017. Implementasi Business Intelligence untuk Menganalisis Data Persalinan Anak di Klinik Ani Padang dengan Menggunakan Aplikasi Tableau Public, JOIN Volume 2 No. 1, Juni 2017/ISSN 2527-9165

Santoso Lucky E. 2008. Penggunaan Spreadsheet Dalam Pengajaran Sistem Pendukung Keputusan. Jakarta : JIK (Jurnal Ilmu Komputer).

Solihin Firdaus. 2009. Bahan Ajar : Data Warehouse. Madura : Universitas Trunojoyo. [www.fsolihin.co.cc](http://www.fsolihin.co.cc)

Steve and Nancy Williams. 2007. The Profit Impact of Business Intelligence . Getting to Green, DM Review.

Thia, F. 2011. Apa Itu Business Intelligence?,  
<http://www.kampusbi.com/artikel/apa-itu-business-intelligence>,  
diakses tanggal 28 Mei 2011

Turban, E., Sharda, R., Delen, D., & King, D. 2011. Introduction To Business Intelligence. *Business Intelligence: A Managerial Approach*, 3–18. [Https://Doi.Org/10.1300/J155v11n01\\_01](Https://Doi.Org/10.1300/J155v11n01_01)

Vercellis, C. 2009. *Business Intelligence: Data Mining And Optimization For Decision Making*. Wiley

Von Halle, Barbara. 2001. *Business Rules Applied: Building Better Systems Using the Business Rules Approach*. New York: John Wiley & Sons.

## TENTANG PENULIS

---



Neni Purwati lahir di Metro, 1 Januari 1978. Pendidikan SD, SMP dan SMA ditempuh di Tulang Bawang Lampung. Menyelesaikan pendidikan S1 di Jurusan Sistem Informasi, STMIK Teknokrat Lampung tahun 2006. Pendidikan S2 diselesaikan pada program Pasca-sarjana di Program Studi Magister Teknik Informatika, Institut Informatika dan Bisnis Darmajaya Lampung tahun 2013. Mulai mengajar sejak tahun 2007 hingga sekarang di Jurusan Sistem Informasi Program Studi S1, dan menjadi Dosen Profesional sejak tahun 2016. Buku ini adalah buku pertama yang ditulis.



# DATA WAREHOUSE

Buku ini akan membahas tentang data warehouse dalam beberapa Bab antara lain:

- Bab I : Membahas pengertian dan latar belakang dikembangkannya Data Warehouse
- Bab II : Membahas tentang keterkaitan antara Business Intelligent dengan Data Warehouse .
- Bab III : Membahas, tentang hal-hal yang berkaitan dengan Data Warehouse (Karakteristik Data Warehouse, Konsep Dasar Data Warehouse, Tujuan Data Warehouse, Mekanisme Data Warehouse, Anatomi data Warehouse, Arsitektur Data Warehouse, Proses dalam Data Warehouse)
- Bab IV : Menjelaskan tentang infrastuktur (Operasional dan Fisik) karena membangun data warehouse harus didukung oleh Infrastuktur yang baik agar hasilnya maksimal.
- Bab V : Menjelaskan tentang data, database, metadata, data multidimensi, pemodelan data multidimensi, Cube (Kubus) Spreadsheet.
- Bab VI : Menjelaskan tentang Desain Data Warehouse, Aspek yang perlu dipertimbangkan (saat mendesain, saat implementasi dan saat testing).
- Bab VII : Membahas tentang berapa pendapat para pakar dalam membangun data warehouse dengan menggunakan langkah, metode dan tahapan.
- Bab VIII : Membahas tentang mengimplementasikan semua teori yang telah dibahas sebelumnya.



Kampus IIB DARMAJAYA  
Jl. Zainal Abidin Pagar Alam No 93,  
Bandar Lampung 35142, INDONESIA

ISBN: 978-602-51690-2-1

