

# Komponen-komponen Pembangunan Data Warehouse dari Perspective Kimbal

Ade Rahmat Iskandar<sup>1</sup>, Hary Nugroho<sup>2</sup>, Asep Herman<sup>3</sup>

<sup>1,2</sup>*Telecommunication Engineering, Akademi Telkom Jakarta*

<sup>3</sup>*Universitas Kebangsaan Republik Indonesia(UKRI)*

ader@akademitelkom.ac.id

Received on dd-mm-yyyy, revised on dd-mm-yyyy, accepted on 20-01-2029

## Abstract

Data warehouse merupakan teknologi yang berorientasi subject, time variant dan non-volatile. Data warehouse merupakan disiplin ilmu untuk mentransformasikan data dari sumber sistem operasional kedalam dimensional modeling. Dalam penelitian ini menggunakan review terhadap textbook Toolkit Data warehouse Ralph Kimbal 2012 selanjutnya paparan di bagian Implementasi dijelaskan pada case study Data Warehouse Pangkalan Data Pendidikan Tinggi (forlap.ristekdikti.go.id) yang merupakan implementasi Data Warehouse Pendidikan di tingkat Dirjen Pendidikan Tinggi untuk seluruh civitas perguruan Tinggi di Indonesia.. Luaran dari penelitian ini adalah diperehnya teoritis terkait komponen-komponen arsitektur dari Data warehouse dan benchmark implementasi Data Warehouse Forlap Dikti.

**Keywords:** Data warehouse, Ralph Kimbal, Forlap, RistekDikti

## I. INTRODUCTION

Data warehouse sudah menjadi kajian yang populer sejak artikel Ralph Kimball diterbitkan edisi pertama pada The Data Warehouse Toolkit (Wiley) pada tahun 1996. Meskipun banyak perusahaan sudah mengadopsi konsep data warehouse sebelumnya, setelah publikasi Ralph Kimbal, Data Warehouse semakin banyak diimplementasikan di berbagai perusahaan atau enterprise.

Banyak industri sudah mengimplementasikan ribuan data warehouse. Volume data terus bertambah seiring dengan frekuensi peningkatan pengelolaan data yang lebih besar. Banyak vendor yang mendukung dengan menyediakan perangkat keras maupun perangkat lunak untuk pengembangan Data warehouse yang lebih baik, sehingga membuat keputusan yang lebih baik dan menghasilkan Return on Investment (ROI) yang terukur.

Sejak edisi pertama The Data Warehouse Toolkit diterbitkan, pemodelan dimensional sudah diterima secara luas sebagai teknik yang dominan untuk merepresentasikan data warehouse. Para pakar data warehouse mengakui bahwa presentasi data warehouse harus didasarkan pada kesederhanaan supaya dapat berjalan dengan sukses. Konsep kesederhanaan adalah kunci mendasar yang memungkinkan pengguna untuk memahami basis data dengan mudah dan perangkat lunak untuk menavigasi basis data secara efisien. Dalam pembangunan data warehouse harus diisngkonkan antara performasi dan perspective bisnis[1].

## II. KAJIAN TERKAIT

### A. Pemodelan Dimensional Primer

Pada bahasan pemodelan dimensional primer, data warehouse dilihat dari perspektif makro. Artikel ini dibuat untuk memberikan pemahaman konsep dalam pembangunan datawarehouse, tidak membahas hal teknis yang bersifat detail. Pada kajian ini perlu disadari perbedaan dari peran seorang ahli data warehouse dengan seorang manajer percetakan. Perbedaan ini dikaji untuk memberikan perbedaan job description dari seorang database administrator (DBA) dan pakar bisnis (MBA) atau bisnis analis.

Oleh akhir bab ini kami berharap Anda akan menghargai kebutuhan ini menjadi setengah DBA (administrator basis data) dan setengah MBA (analis bisnis) sebagai Anda menangani gudang data Anda.

Tujuan dari datawarehouse : untuk membuat tujan dari data warehouses yang baik, harus dibangun dengan cara mendengarkan kepada pihak-pihak manajemen bisnis terkait. Beberapa pertanyaan berikut dapat diajukan :

- “Kita memiliki data yang besar di perusahaan, tetapi kita tidak dapat mengakesnya”
- Kita harus dapat mendekomposisi (slice dan dice) dengan cara-cara tertentu
- Anda harus dapat membuatnya mudah, supaya pegawai yang terlibat dalam bisnis mendapatkan data secara langsung
- Coba tunjukan, kenapa ini penting?
- Kenapa terdapat perbedaan dari dari metriks bisnis yang sama
- Kita mengharapkan banyak orang menggunakan informasi untuk mendukung pengambilan keputusan berdasarkan fakta (fact-based).[1]

#### **Data warehouses harus membuat informasi perusahaan mudah diakses:**

Isi dari data warehouse harus dapat dimengerti. data harus bersifat intuitif dan jelas bagi pengguna bisnis, bukan hanya bagi developer. Data harus mudah dipahami menyiratkan arti dengan jelas; isi data warehouse perlu diberi label bermakna. Pengguna bisnis ingin memisahkan dan menggabungkan data di data warehouse dalam kondisi yang berkesinambungan, proses ini yang biasa disebut dengan slicing dan dicing. Alat yang mengakses data warehouse harus sederhana dan mudah digunakan. Data warehouse juga harus dapat mengembalikan hasil query kepada user dalam waktu minimum.

#### **Data warehouse harus memberikan informasi secara konsisten:**

Data pada data warehouse harus kredibel. Data harus hati-hati dikumpulkan dari berbagai sumber di sekitar organisasi, dibersihkan, kualitasnya terjamin, dan dirilis hanya jika sesuai untuk di konsumsi pengguna. Informasi dari satu proses bisnis harus sesuai dengan informasi yang lain. Dengan asumsi, Jika dua ukuran kinerja memiliki nama yang sama, maka informasi tersebut pasti memiliki hal yang sama. Sebaliknya, jika dua ukuran data yang dibandingkan tidak sama, maka mereka harus diberi label berbeda. Informasi yang konsisten berarti informasi berkualitas tinggi. Hal Ini berarti bahwa semua data diperhitungkan dan lengkap. Konsistensi tersebut akan menyiratkan bahwa definisi umum untuk isi dari data warehouse tersebut tersedia untuk pengguna.

#### **Data ware house harus adative dan resilient terhadap perubahan:**

Kita menyadari bahwa kita tidak bisa menghindari perubahan. Kebutuhan pengguna, kondisi bisnis, data, dan teknologi, semua komponen tersebut akan bergulir seiring dengan perjalanan waktu. Data warehouse harus dirancang untuk menangani perubahan yang tak terhindarkan. Perubahan pada data warehouse harus *smooth*, artinya data warehouse tidak akan membuat data yang ada atau pun dalam aplikasi menjadi tidak valid. Data dan aplikasi yang ada tidak boleh diubah atau terganggu ketika komunitas bisnis mengajukan pertanyaan baru

atau data baru ditambahkan kedalam data warehouse. Jika data di data warehouse dimodifikasi, harus dapat diperhitungkan bahwa perubahan akan dilakukan dengan dengan tepat.

**Data warehouse harus aman dan melindungi informasi :**

Informasi saat ini dibaratkan sebagai berlian bagi suatu perusahaan. Informasi membuat suatu perusahaan dapat menjadi superior dibanding dengan kompetitornya. Pada data warehouse harus dipastikan bahwa data dan informasi dikelola dengan tingkat keamanan yang baik, sehingga tidak mudah untuk di akses oleh pihak-pihak yang tidak memiliki otoritas untuk mengaksesnya.

**Data warehouse harus dapat menjadi pondasi untuk menghasilkan keputusan yang tepat:**

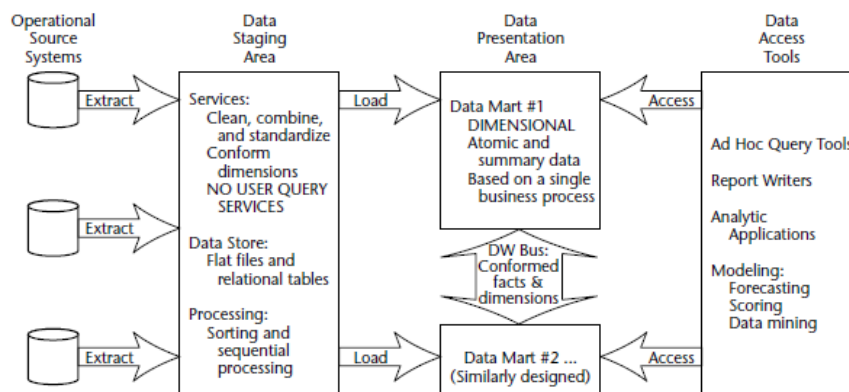
Data warehouse harus memiliki data yang tepat untuk mendukung proses pengambilan keputusan yang akan diambil. Poin terpenting dari data warehouse adalah : keputusan yang dibuat setelah data warehouse dapat disajikan dengan bukti-buktinya. Keputusan ini memberikan dampak dan nilai bisnis yang dapat didistribusikan kepada data warehouse.

**Komunitas bisnis harus menerima data warehouse untuk mendapatkan sistem yang baik:**

Komunitas bisnis meskipun sudah membuat sistem informasi atau aplikasi untuk mengelolah data sebelumnya, mereka harus melakukan transisi untuk menerika data warehouse. Data warehouse memiliki konsep yang berbeda dibandingkan dengan pengeloan basis data biasa. Sesuah implementasi data warehouse harus dilakukan upaya training paling tidak dalm kurun waktu enam bulan sampai proses acceptance test dilakukan.[1]

**B. Komponen Data Warehouse**

Pada data warhouse memiliki beberapa komponen yang berbeda dengan komponen di basis data. Data warehouse memiliki tahapan penting yang dikenal dengan Extract, Transform dan Load (ETL) sebelum diperoleh data yang akurat, untuk proses pengambilan keputusan. Berikut ini adalah komponen-komponen dari data warehouse :



Gambar 1-1 Komponen-komponen Data Warehouse[1]

### **Operational Source System:**

Prioritas utama dari sistem sumber operasional adalah memproses kinerja dan ketersediaan data inputan. Data saat ini dapat diperoleh dalam berbagai format tidak hanya inputan database saja, sumber data lain seperti data dengan format spreadsheet, txt, XML, CSV atau pun bentuk format lainnya dapat menjadi inputan untuk dikelola lebih lanjut oleh sistem data warehouse.

Pada sistem sumber tersebut dapat memberikan informasi dari historical data. Setiap sistem sumber biasanya mengadopsi konsep “stovepipe system” dimana sistem secara efisiensi untuk mengelola data bersama seperti produk, pelanggan, geografis atau kalender dari sistem sumber lain dari perusahaan tersebut. Jika sistem sumber dikelola dengan baik seperti menggunakan enterprise application integration (EAI), maka akan membuat tugas-tugas rancangan data warehouse semakin mudah.

### **Data Staging Area :**

Data Staging Area pada data warehouse merupakan area penyimpanan dan satu set proses yang biasa disebut extract-transform-load (ETL). Data staging area berada diantara sistem sumber operasional dan penyajian data. Pada staging area, para profesional berada pada tempat ini, kemampuan Database Administrator (DBA) memiliki peran tambahan sebagai profesional yang memahami konsep management data atau data scientist, pahaman terhadap Structured Query Language (SQL) dan No SQL diperlukan pada data warehouse.[1]

Pada penelitian lain, Ade Rahmat[2] menjelaskan, bahwa pada tahapan Extract, transform dan Load (ETL) merupakan proses atau tahapan dalam Online Analytical Processing (OLAP) menggunakan tool open Source Pentaho Data Integration. Dalam staging Area, bisa mengadopsi tools untuk melakukan tahapan-tahapan ETL tersebut menggunakan tool open Source seperti Pentaho Data Integration, maupun menggunakan tools licence seperti Microsoft SSAS dan Oracle Data warehouse.

Extract-transform-loading (ETL) merupakan bagian dari pertanggung jawaban dalam proses meng-extraksi data dari sumber-sumber yang bervariasi, seperti proses cleansing (pembersihan), kustomisasi, pemormatan ulang, dan proses memasukan kedalam suatu datawarehouse. Pembangunan proses ETL merupakan suatu tahap yang paling besar dalam suatu pembangunan datawarehouse, pembangunan data warehouse merupakan suatu project yang rumit, memerlukan waktu yang luas, serta dana sumber daya yang besar[3].

### **Data Presentation :**

Area presentasi data merupakan tempat mengatur, menyimpan dan menyediakan data untuk permintaan langsung oleh pengguna, penulis laporan, dan aplikasi analitik lainnya. Pada sistem sumber operasional merupakan area terbatas hanya untuk para profesional (data scientist) sedangkan pada presentation area merupakan tempat eksplorasi para komunitas bisnis. Luaran dari presentasi area, bisa berupa report, dashboard maupun dinamic chart lainnya.

Industry berasumsi bahwa dimensional modeling merupakan teknik yang paling banyak digunakan dalam mengirimkan data kepada user-user data warehouse. Pada pemahannya pakar industri sebelumnya, mengasumsikan bahwa dimensional modeling merupakan karakteristik dari 3NF dalam basis data, dimana proses redundansi dihilangkan untuk mendapatkan tabel-tabel yang baik. Jika presentatition area berdasarkan suatu database relasional, maka dimentional modeling mengacu pada tabel-tabel yang diacu kedalam suatu skema star dalam Data warehouse, presentation pada dimentional modeling mengacu pada teknologi Online Analytical Processing (OLAP).

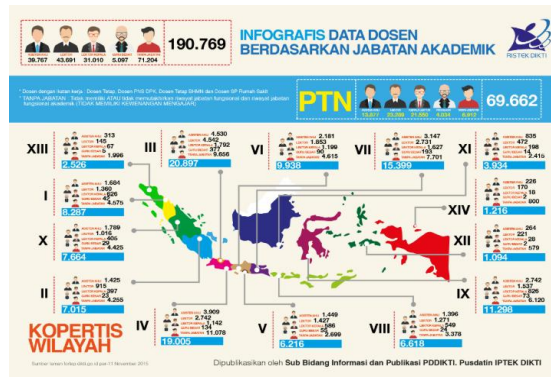
### **Data Access Tools:**

Komponen utama terakhir dari lingkungan data warehouse adalah alat-alat akses data. Pada area ini, query merupakan kemampuan utama dalam mengeksplorasi data. Aplikasi-aplikasi analitic, proses pemodelan

dengan forecasting, scoring, data mining yang akan mengeksplorasi data lebih jauh dari hasil data warehouse tersebut.

### III. Implementasi Data Warehouse (Forlap.Ristekdikti.go.id)

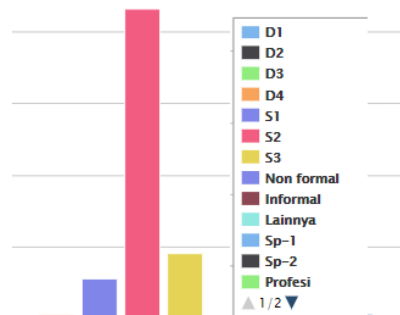
Implementasi Data warehouse dewasa ini sudah banyak diimplementasikan di beberapa perusahaan skala enterprise. Beberapa bidang yang sangat penting dalam mengimplementasikan teknologi data warehouse diantaranya adalah bidang retail, kesehatan, perbankan dan pendidikan. Dari penelitian penulis sebelumnya Indonesia sudah menerapkan Implementasi Data warehouse untuk sistem Pangkalan Data Pendidikan Tinggi. Berikut adalah data total perguruan tinggi yang diambil dari Data warehouse forlap.ristekdikti.go.id :



Gambar : Sebaran Dosen berdasarkan jabatan akademik  
(Sumber : [www.forlap.ristekdikti.go.id](http://www.forlap.ristekdikti.go.id), 2019)

Implementasi Data warehouse yang diterapkan dalam Sistem berbasis web forlap.ristekdikti.go.id sudah diimplementasikan di Indonesia sejak tahun 2013 lalu. Dari penelitian penulis sebelumnya, selama transisi dari tahun 2009 sampai tahun 2012 masih mengalami kendala yang signifikan. Dari tahun 2015 sampai saat ini data warehouse (forlap risktek dikti) sudah memiliki tingkat kematangan yang lebih baik. Luaran dari Data warehouse (Forlap Risktek Dikti) diantaranya dapat diketahui jumlah Perguruan Tinggi baik perguruan tinggi negeri (PTN) atau perguruan tinggi swasta (PTS) di Indonesia secara realtime, diketahui jumlah Dosen, jumlah Mahasiswa, jumlah Dosen dengan tingkat Kepangkatan Akademik dari (Asisten ahli, Lektor, Lektor Kepala sampai Guru besar).

Berikut adalah luaran dari Data warehouse (Forlap.ristekdikti.go.id) untuk total Jumlah Dosen di Indonesia berdasarkan jumlah pendidikan :



Gambar : Jumlah dosen berdasarkan tingkat pendidikan  
(Sumber : [www.forlap.ristekdikti.go.id](http://www.forlap.ristekdikti.go.id), 2019)

Jumlah dosen secara realtime dapat ditampilkan secara realtime yang sangat bermanfaat bagi para decision maker di masing-masing perguruan tinggi dan juga pihak Dirjen Pendidikan Tinggi (DIKTI). Dari data tersebut, diperoleh informasi bahwa jumlah tingkat pendidikan tinggi di Indonesia adalah dengan pendidikan S2 atau magister, tingkat S3 masih sedikit dan masih terdapat Dosen yang masih berpendidikan S1.

#### **IV. Hasil dan Diskusi**

Dari kajian yang sudah diimplementasikan, Data warehouse dapat memberikan manfaat yang sangat baik terutama untuk pengolahan data secara historical, dan berorientasi subjek. Pada case Data Warehouse pendidikan tinggi di Indonesia berbasis web bases (Forlap.ristekdikti.go.id) sudah dapat men-tracer semua data perguruan tinggi secara real time, pada tahun 2017 dapat diketahui terdapat 100an lebih perguruan tinggi tidak sehat dengan perolehan data secara realtime dari data warehouse (forlap.ristekdikti) tersebut. Untuk kajian akademisi di level mahasiswa bisa dipersiapkan lab Data Science sehingga dapat digunakan secara terintegrasi kedepannya dari disiplin ilmu (Data warehouse, Data Mining, Machine learning, Big Data dan Data Science).

## REFERENCES

- [1] R. Kimball and M. Ross, *The Datawarehouse Toolkit -The complete guide to dimention modeling*. PAck ublisheing, 2010.
- [2] A. R. Iskandar, "Rancang Bangun Online Analytical Processing ( OLAP ) Classic Model Data," in *Seminar Nasional Informatika dan Aplikasinya*, 2019.
- [3] S. H. A. El-sappagh, A. Hamed, E. Bastawissy, and A. M. Ahmed, "ORIGINAL ARTICLE A proposed model for data warehouse ETL processes," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 23, no. 2, pp. 91–104, 2011.