

# **Model *Extract, Transform dan Load* (ETL) Pada *Data Warehouse* untuk Pelaporan Evaluasi Program Studi Berdasarkan Evaluasi Diri (EPSBED) Menggunakan Model *Data Warehouse* EPSBED**

## **Studi Kasus : Universitas Budi Luhur**

**Indra, Painem**

<sup>1)</sup> Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Budi Luhur

<sup>2)</sup> Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur

<sup>1,2)</sup> Jl. Raya Ciledug, Petungkang Utara, Kebayoran Lama, Jakarta Selatan 12260, Indonesia  
*e-mail* : indra@budiluhur.ac.id

### ***Abstract***

*Each of study program at university in Indonesia have to report the results of academic activities for one semester to the Directorate General of Higher Education, The Ministry of National Education of Republic of Indonesia (DIKTI) through the Coordinator of Private Higher Education (KOPERTIS). The report will be used in measuring the performance of the study program for each university in Indonesia. The reporting process is known as the Study Program Evaluation Based on Self Evaluation (EPSBED). Until now data from the processing of EPSBED not yet maximized by the executive party of Budi Luhur University to become one of the reference in the field of academic decisions. For this reason the analysis of the data warehouse of EPSBED may be one of an important component to be considered in any decision-making by the executive party of Budi Luhur University. Moreover, by using the EPSBED data warehouse the process of generating report become faster in a count of minutes because the process is automated and scheduled. Early stages in the datawarehouse EPSBED are extract, transform and load (ETL). ETL is applied using the two methods which are method of Kettle and manual methods. By ETL to datawarehouse EPSBED made the withdrawal process for the preparation of data reporting EPSBED of each program of study will be faster in a matter of minutes. In addition, the process of downloading data can be done by filtering, cleansing and standardizing of data and a data type and can be developed in the schedule for the data withdrawal in a certain period of time. automatically. Thus, reporting EPSBED can be completed quickly and precisely and academic data reported EPSBED stored in the datawarehouse.*

*Keywords* : EPSBED, data warehouse, extract, transform, load, etl, University and Majoring.

## **1. PENDAHULUAN**

### **1.1. Latar Belakang**

Sejak tahun 2002, setiap perguruan tinggi telah melaksanakan kewajibannya dalam pelaporan kinerja program studinya menggunakan struktur basis data tertentu. Struktur basis data tersebut telah dibakukan dan dikemas dalam sebuah sistem pelaporan Evaluasi Program Studi Berdasarkan Evaluasi Diri, sebagaimana telah dikenal dengan kata EPSBED oleh semua perguruan tinggi di Indonesia,

diatur dalam SK Dirjen DIKTI No. 34/DIKTI/Kep/2001.

Dalam hal proses pelaporan EPSBED, UBL harus melakukan proses *query* yang begitu banyak untuk mengambil data yang dibutuhkan pada *database* UBL. Hal ini karena struktur *database* UBL berbeda dengan struktur *database* EPSBED. Selain itu, juga harus dilakukan proses pembersihan (*cleansing*) data karena masih banyaknya data yang belum lengkap serta membutuhkan waktu yang lama.

## 1.2. Rumusan Masalah

Dari inti permasalahan ini, maka ada suatu pertanyaan mendasar yaitu “Bagaimana merancang *extract, transform* dan *load* (ETL) sebagai tahapan awal dalam penerapan data warehouse untuk mempermudah dan mempercepat proses pelaporan EPSBED ? “.

## 1.3. Tujuan Penelitian

Tujuan akhir dari penelitian ini adalah untuk:

1. Merancang ETL *data warehouse* EPSBED untuk mempermudah dan mempercepat proses pelaporan EPSBED.
2. Mengimplementasikan ETL dengan menggunakan software *kettle* (Pentaho Data Integration).

## 2. REVIEW LITERATURE

### 2.1. EPSBED

#### a. Definisi EPSBED

EPSBED merupakan media pelaporan program studi yang diselenggarakan oleh Perguruan Tinggi untuk Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan Nasional Republik Indonesia (DIKTI). Berdasarkan ketentuan dan perundangan yang berlaku, setiap program studi harus melaporkan kegiatan akademiknya setiap semester. Sejak tahun akademik 2002-2003, kegiatan pelaporan program studi tersebut menggunakan data elektronik dan aspek yang dilaporkan meliputi kelembagaan, kurikulum, dosen, mahasiswa, dan infrastruktur yang diakses oleh setiap program studi (Ilah, <http://evaluasi.dikti.go.id/>).

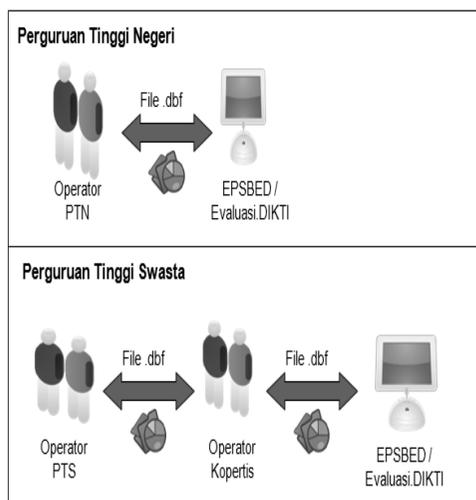
#### b. Dasar Hukum EPSBED

Berdasarkan Keputusan Dirjen DIKTI Nomor: 08/DIKTI/Kep/2002 tentang Petunjuk Teknis Keputusan Menteri Pendidikan Nasional Nomor 184/U/2001 Tentang Pedoman Pengawasan Pengendalian dan Pembinaan Program Diploma, Sarjana dan Pasca Sarjana di perguruan tinggi (termasuk ketentuan tentang Ijazah dan Transkrip). Keputusan tersebut merupakan salah satu dasar

hukum dalam pelaksanaan EPSBED disetiap perguruan tinggi di Indonesia.

#### c. Alur Kerja ESPBED.

Alur kerja EPSBED adalah urutan proses migrasi data dari *database* internal setiap perguruan tinggi ke dalam *database* EPSBED di DIKTI. Sesuai dengan acuan dari Pengembangan Pangkalan Data Pendidikan Tinggi (PDPT) maka alur kerja EPSBED dapat dijelaskan seperti Gambar 2.1.



Gambar 1 :  
Alur Kerja EPSBED (PDPT, 2010)

### 2.2. Data Warehouse.

#### a. Definisi Data Warehouse.

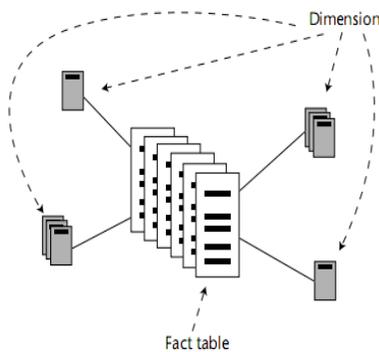
*Data warehouse* adalah suatu koleksi data yang digunakan untuk pengambilan keputusan manajemen, yang berorientasi subjek (topik), terpadu, *time variant* dan tidak mudah berubah (Inmon, 2005). Turban, Sharda and Delen (2011) menjelaskan bahwa *data warehouse* juga digunakan sebagai pusat penyimpanan data masa lalu dan data saat ini yang sangat berpotensi untuk bahan pertimbangan pihak manajer suatu organisasi.

#### b. Teknik Pemodelan *Data Warehouse*.

Pada riset ini akan digunakan model multi dimensi, dimana terdapat dua

dimensi untuk tiap *data warehouse*, yaitu:

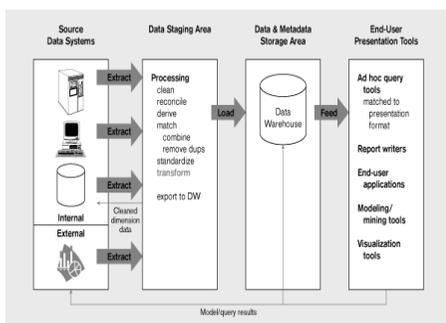
umumnya memiliki *measurement* dan *foreign key*. *Measurement* adalah *field* yang memiliki nilai numerik, digunakan untuk kepentingan pengukuran (*measure*), sedangkan *foreign key* adalah *primary key* dari dimensi yang terkait dalam perancangan tabel fakta. Teknik pemodelan data *warehouse* menggunakan pendekatan *star join*. Pendekatan *star join* ini menyerupai bentuk bintang, dimana tabel fakta terdapat di tengah dan tabel dimensi mengelilinginya. Pendekatan ini dapat dilihat pada Gambar 2.



Gambar 2 :

Pendekatan *Star join* model data multidimensional ([Inmon 2005], 360)

**c. Arsitektur Data Warehouse Pada Gambar 3 terdapat Arsitektur data warehouse dapat digambarkan :**



Gambar 3: Arsitektur *Data Warehouse* (Hoffer et al., 2007)

Dari Gambar 3, data *warehouse* terbagi menjadi empat bagian yaitu:

**1. Source Data System**

*Data sources* diperoleh dari berbagai transaksi dan produksi perusahaan hasil

tabel fakta dan tabel dimensi. Tabel fakta

operasional aplikasi yang dijalankan tiap harinya. Data masih bersifat transaksional biasa atau masih berupa data mentah.

**2. Data Staging Area**

Sebelum masuk kedalam tahap ini, data diestrak terlebih dahulu untuk dimasukkan kedalam tahap *data staging area*. Pada tahap ini data di *cleansing*, *reconcile*, *match*, *standardize* agar data bersih dari cacat atau lebih dikenal dengan nama *transform*.

**3. Data & Metadata Storage**

Setelah data dibersihkan maka data dimasukkan (*load*) kedalam data warehouse. data pada data warehouse ini dapat digunakan sebagai bahan dalam menentukan kebijakan (*decision support*) dalam berbagai masalah pihak eksekutif.

**4. End User Presentation Tools**

Pada tahap akhir ini adalah pengembangan dari *data warehouse* yang sudah ada. Salah satu dari penggunaan *data warehouse* ini adalah untuk digunakan menjadi suatu *business intelligence*.

**d. Tools ETL Data Warehouse**

Untuk perancangan data *warehouse* menggunakan *Pentaho Schema Workbench*. Sedangkan untuk implementasi *Online Analytical Processing (OLAP)* menggunakan JPivot yang sudah terintegrasi dengan Pentho Bi Server. Untuk ETL menggunakan *Kettle (Pentaho Data Integration)*. Ketiga *tools* ini dapat diunduh melalui sebuah situs <http://www.sourceforge.net>.

**3. METODOLOGI PENELITIAN**

**3.1. Analisa Kebutuhan Informasi**

Kebutuhan informasi inilah yang dijadikan dasar untuk melakukan pengumpulan data pada tahapan berikutnya.

**3.2. Perancangan Data Warehouse**

Pada tahapan ini dilakukan perancangan *data warehouse* menggunakan model *star schema*. Dari hasil rancangan model *star schema* ini didapatkan sebuah tabel fakta yang

diharapkan dapat mendukung pelaporan EPSBED.

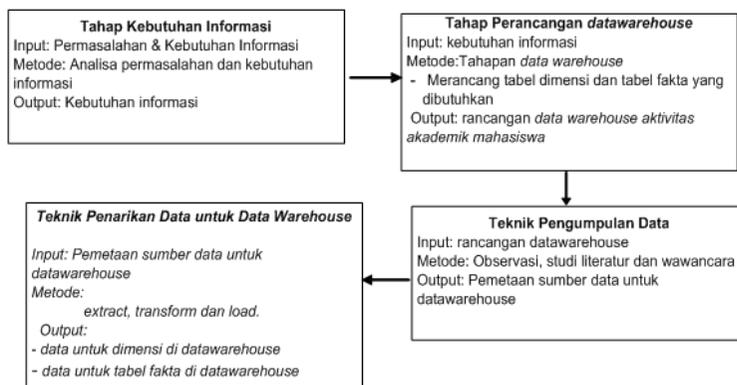
**3.3. Teknik Pengumpulan Data**

Pada tahap ini dilakukan proses pengumpulan data dengan teknik observasi, studi literatur dan wawancara dengan pihak terkait. Wawancara dilakukan dengan pihak Kepala Biro Teknologi UBL dan Ketua Program Studi Teknik Informatika, Sistem Informatasi dan Diploma 3 Manajemen Informatika. Hasil dari tahap ini akan didapatkan sebuah data transaksi (OLTP) untuk digunakan dalam perancangan *data warehouse*.

**3.4. Teknik Penarikan Data untuk Data Warehouse**

Pada tahap ini data diekstrak dari *database* transaksional, kemudian dilakukan proses *cleansing* untuk menghilangkan data yang *empty* atau redundan. Setelah data di *cleansing* maka data ditransformasikan dengan melakukan pendefinisian *view relational* pada tabel dalam sumber data. Setelah proses transformasi maka dilakukan proses *loading data* untuk memasukkan data ke dalam *data warehouse*.

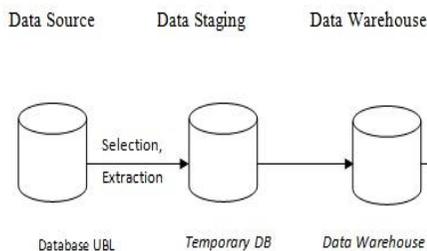
Dari keempat perancangan diatas dijabarkan dalam bentuk sebuah alur diagram pada Gambar 4 :



Gambar 4. Tahapan Metodologi Penelitian

**3.5. Arsitektur Data Warehouse**

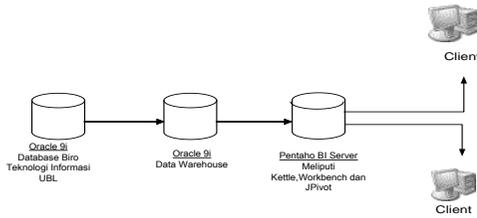
**3.5.1. Arsitektur Logik Data Warehouse.** Pada Gambar 5 berisi penjelasan tentang arsitektur logik pada *data warehouse* UBL untuk kebutuhan pelaporan EPSBED.



Gambar 5. Arsitektur Logik Data Warehouse EPSBED UBL

*Data source* merupakan sumber data dari keseluruhan proses transaksi akademik di UBL. *Data source* ini menggunakan *software Oracle 9i licence*. Pada tahap pertama ini dilakukan proses pemilihan tabel yang dibutuhkan dalam perancangan *data warehouse* sesuai dengan dimensi dan tabel fakta yang ada, proses ini disebut dengan *selection*. Tabel yang sudah ditentukan selanjutnya dilakukan proses ekstraksi data untuk memetakan data dari setiap tabel yang dibutuhkan untuk dimasukkan ke dalam *data warehouse*, proses ini lebih dikenal dengan nama *extraction*.

**3.5.2. Arsitektur Fisik Data Warehouse EPSBED UBL.**



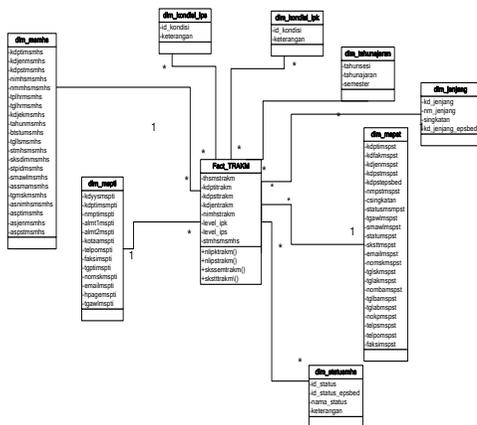
Gambar 6. Arsitektur Fisik Data Warehouse EPSBED UBL

Database operasional UBL menggunakan Oracle 9i dengan SID: SISTEM. Sedangkan database yang digunakan untuk data warehouse menggunakan Oracle 9i dengan SID: SIDIKTI. Untuk tools dalam proses ETL digunakan Pentaho Data Integration (Kettle), untuk cube menggunakan Pentaho Schema Workbench dan OLAP menggunakan Pentaho Analysis Service.

**a. Perancangan Data Warehouse**

Didalam menerapkan data warehouse EPSBED berisi beberapa tabel fakta, diantaranya adalah tabel fakta aktivitas akademik mahasiswa.

**b. Perancangan Tabel Fakta Aktivitas Akademik Mahasiswa (FACT\_TRAKM)**



Gambar 7. Perancangan Tabel Fakta Aktivitas Akademik Mahasiswa

Tabel fakta FACT\_TRAKM merupakan tabel fakta yang digunakan untuk menghasilkan laporan sebaran IPK, sebaran IPS dan jumlah SKS mahasiswa selama melakukan studi di setiap program studi serta digunakan untuk *generate* laporan status mahasiswa. Tabel fakta FACT\_TRAKM berisi *measurement* dan *foreign key*. *Measurement* adalah *field* bertipe *numeric* yang dijadikan pengukuran pada tabel fakta. *foreign key* adalah *primary key* pada dimensi yang terkait dalam perancangan tabel fakta. Deskripsi dari *measurement* dan *foreign key* pada tabel fakta FACT\_TRAKM.

Dimensi yang terkait dengan perancangan tabel fakta aktivitas akademik (FACT\_TRAKM) adalah dimensi mahasiswa (DIM\_MSMHS), dimensi perguruan tinggi (DIM\_MSPTI), dimensi kondisi IPK (DIM\_KONDISI\_IPK), dimensi kondisi IPS (DIM\_KONDISI\_IPS), dimensi tahun ajaran (DIM\_TAHUNAJARAN), dimensi jenjang studi (DIM\_JENJANG), dimensi program studi (DIM\_MSPST) dan dimensi status mahasiswa (DIM\_STATUS\_MHS).

**4. PEMBAHASAN PENELITIAN.**

Setelah dilakukan perancangan tabel fakta dan tabel dimensi. Tahapan berikutnya adalah dilakukan proses *extract, transform* dan *load* (ETL) untuk mendapatkan sebuah data yang valid yang disimpan dalam *data warehouse*. Proses ETL ini menggunakan dua metode yaitu metode Kettle dan metode manual. Kedua metode ini dapat dijelaskan sebagai berikut :

**4.1. Metode Kettle**

Metode Kettle ini adalah sebuah metode untuk penarikan data yang dilakukan dengan menggunakan software Kettle (Pentaho Data Integration). Metode Kettle digunakan karena data yang ditarik lebih dari 20 *record* bahkan mencapai jutaan *record*. Dalam penerapan metode ini masih menggunakan tahapan *extract, transform* dan *load*.



manual dalam penarikan datanya. Sebagai contoh adalah dimensi jenis kelamin. Untuk mengisikan data kedalam dimensi jenis kelamin ini karena hanya berisi dua *record* yaitu Laki-Laki dan Perempuan, maka untuk mengisikan datanya cukup memasukkan secara manual.

## 5. KESIMPULAN

Berdasarkan penelitian yang sudah dilakukan sebelumnya, dapat disimpulkan beberapa hal sebagai berikut :

1. Implementasi *data warehouse* di UBL dapat membantu mengatasi masalah dalam menyelesaikan pelaporan EPSBED dalam waktu yang cepat. Sebelum implementasi *data warehouse*, semua proses pelaporan EPSBED dilakukan dengan menggunakan *query* untuk *extract*, *transform* dan *load* dalam pengumpulan data. Waktu yang dibutuhkan dalam menyelesaikan pelaporan dengan *query* tersebut biasanya membutuhkan waktu satu bulan. Dengan adanya *data warehouse* maka keseluruhan data EPSBED yang akan dilaporkan ke DIKTI sudah melalui tahap *extract*, *transform* dan *load* dengan menggunakan *Kettle* sehingga data yang disajikan dalam aplikasi EPSBED lebih cepat dan sudah berupa data yang valid. Waktu yang dibutuhkan dengan adanya *data warehouse* inipun cukup dua jam dalam menyelesaikan pelaporan EPSBED.

2. Proses pengolahan pelaporan EPSBED dilakukan secara otomatis dan dapat dibuat *schedule* dengan menggunakan komponen *job* pada *Kettle*, Sehingga mempermudah dan mempercepat proses kerja dari tim menyusun pelaporan EPSBED.

3. Hasil dari pengolahan data warehouse EPSBED ini dapat dijadikan salah satu bahan pertimbangan oleh pihak eksekutif dalam menentukan kebijakannya. Informasi yang disajikan berupa laporan sebaran IPK, laporan sebaran IPS, laporan status dan kelulusan mahasiswa, laporan jumlah dosen tetap aktif dan jumlah dosen tetap berdasarkan pendidikan terakhir setiap program studi.

## DAFTAR PUSTAKA

- DIKTI. (2010). *Pengembangan Pangkalan Data Pendidikan Tinggi*. May, 12 2011. Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan Nasional Republik Indonesia. [http://bapsi.ub.ac.id/documents/Paparan\\_PDPT\\_Dikti\\_Hery.ppt](http://bapsi.ub.ac.id/documents/Paparan_PDPT_Dikti_Hery.ppt)
- Inmon, W.H.(2005). *Building The Data warehouse*. New York: John Wiley and Sons, Inc.
- Efraim Turban et al. (2007). *Decision Support and Business Intelligent System*. Pearson.
- Ilah. (2010). *Evaluasi Program Studi Berdasarkan Evaluasi Diri (EPSBED)*. May, 10 2011. Direktorat Jenderal Pendidikan Tinggi, Kementerian Pendidikan Nasional Republik Indonesia. <http://evaluasi.dikti.go.id>
- Roldan, Maria Carina. (2010). *Pentaho 3.2 Data Integration Beginner's Guide*. Birmingham : Pack Publishing Ltd.