DATA PREPARATION UNTUK ANALISA POLA DATA PEMANTAUAN CUACA DI KOTA BANDUNG

Oleh:
Ellin Natalia Panjaitan
10107351

Penguji 1:	Penguji 2:	Penguji 3:
Adam Mukharil Bachtiar, S.Kom.	Mira Kania Sabariah, S.T., M.T.	Dian Dharmayanti, S.T., M.Kom.

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS KOMPUTER INDONESIA
2013

Latar Belakang

- Kebutuhan masyarakat mengenai kondisi cuaca yang terjadi
- Kebutuhan Pusat Penelitian Informatika-Lembaga Ilmu Pengetahuan Indonesia dalam menganalisis data pemantauan cuaca

Rumusan Masalah

- Berdasarkan latar belakang masalah, yang menjadi rumusan masalah adalah bagaimana proses data preparation terhadap data pemantauan cuaca dan apakah hasil dari data preparation tersebut dapat dipakai untuk proses analisis.

Tujuan Penelitian

- Menyiapkan atau menyediakan data yang berkualitas dan konsisten.
- Menghasilkan data yang memenuhi syarat untuk digunakan dalam proses analisis.

Batasan Masalah

- Mengolah data dari tabel-tabel master yaitu tabel temperatur (suhu), kelembaban, curah hujan dan angin.
- Pendekatan analisis perangkat lunak yang digunakan adalah pemodelan analisis berbasis objek, dimana tools yang digunakan adalah use case diagram, sequence diagram, activity diagram, class diagram.
- Seluruh tools yang digunakan dalam analisis, perancangan dan pembangunan perangkat lunak adalah open source.
- Software yang digunakan dalam pembuatan aplikasi ini yaitu NetBeans 7.0.1 dengan bahasa pemrograman java.
- Menggunakan DBMS MySql untuk penyimpanan data.
- Perangkat lunak yang dibangun adalah berbasis desktop.
- Data sampel yang digunakan adalah data klimatologi tahun 2005-2009 dalam bentuk file exel.
- Sasaran pengguna perangkat lunak ini adalah peneliti di Lembaga Ilmu Pengetahuan Indonesia (LIPI).

Data *Mart*

- *Data mart* adalah suatu bagian pada *data warehouse* yang mendukung pembuatan laporan dan analisis data pada suatu unit, bagian operasi pada suatu perusahaan. *Data mart* berisi informasi yang relevan bagi user yang ingin mengambil keputusan. Banyak perusahaan yang telah menerapkan DSS (*Decision Support system*) yang telah ada di suatu data mart, sebuah data mart terintegrasi secara operasional dan sejarah data untuk aplikasi pengambilan keputusan.

- Ada empat tugas yang bisa dilakukan dengan adanya data mart menurut Williams, keempat tugas tersebut yaitu:
  - Pembuatan laporan
  - *On-Line Analytical Processing (OLAP)*
  - *Data mining*

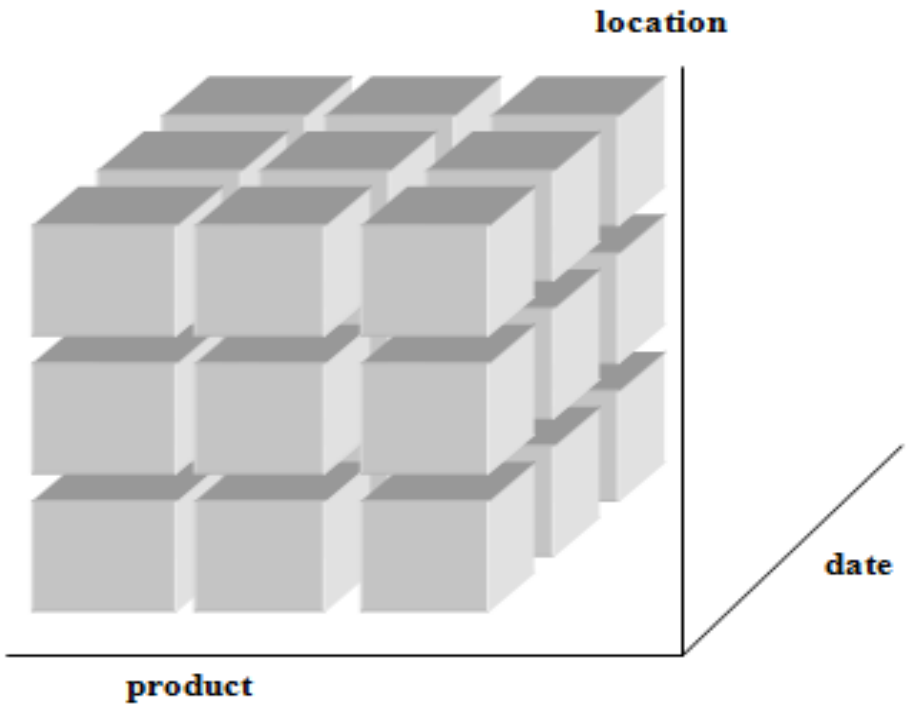
Karakteristik Data *Mart*

- Seperti yang telah dikemukakan terlebih dahulu pada sub bagian pengertian *data mart*, karakteristik yang harus dimiliki dalam sebuah *data mart* antara lain adalah *Subject Oriented*, *Integrated*, *Time Variant*, dan *Non Volatile.*

Pemodelan Data Dimensional

- Pemodelan data dimensional adalah representasi data dengan kubus multidimensional agar lebih mudah dibaca.
- Terdapat 2 macam aspek dalam pemodelan ini, yaitu ukuran (*measures*) dan dimensi (*dimension*). Ukuran adalah besaran data, sedangkan dimensi adalah konteks data atau parameter bisnis.
- Parameter ini dapat dilihat dari karakteristik seperti *who, what, when, where* dan *how* dari subjek data. Ukuran (*measures*) disimpan dalam tabel fakta (*fact table*) sedangkan dimensi disimpan dalam tabel dimensi (*dimension table*).
- Sebagai contoh misalnya untuk basis data mengenai total penjualan per bulan, maka pengukuran dapat dilakukan berdasarkan dimensi lokasi, waktu dan produk yang dijual
- Representasi dalam bentuk kubus multidimensional untuk contoh ini dapat dilihat pada gambar berikut :
  - *Fact Table*
  - *Dimension Table*
  - *Measures*

Kubus Model Dimensional



Penjelasan Kubus Model Dimensional

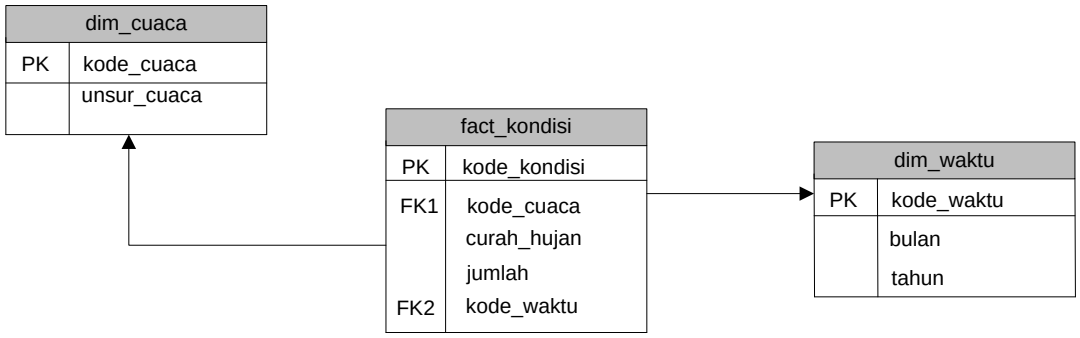
- *Fact table*
  - Tabel fakta adalah pusat dari *table star join* dimana data dengan banyak kepentingan tersimpan . Setiap *fact* biasanya merepresentasikan sebuah bisnis *item*, suatu transaksi bisnis, atau sebuah kejadian yang dapat digunakan dalam analisis bisnis atau proses bisnis.
- *Dimension table*
  - *Dimension table* atau tabel dimensi adalah tempat dimana data tambahan yang berhubungan dengan tabel fakta ditempatkan pada sebuah tabel multidimensional . Dalam dimensional model, semua data menunjukan *fact table* yang diasosiasikan dengan satu dan hanya satu member dari setiap *multiple dimensions.* Jadi dimensi menunjukan latar belakang kontekstual dari *fact.* Banyak proses analisis yang digunakan untuk menghitung dampak dari dimensi pada *fact.*
- *Measures*
  - Suatu *measures* (ukuran) adalah suatu besaran (angka numerik) atribut dari sebuah *fact*, yang menunjukan *performance* atau *behavior* (tingkah laku) dari bisnis secara relatif pada suatu dimensi. Angka atau nomor yang ditunjukan disebut dengan *variable*. Sebagai contoh ukuran dari penjualan dalam bentuk uang, besarnya penjualan, jumlah pengadaan, biaya pengadaan, banyaknya transaksi dan lainnya. Suatu ukuran dijelaskan dengan kombinasi dari member dari suatu dimensi dan diletakkan dalam *fact.* Dalam *dimensional modeling*, ada beberapa pendekatan yang digunakan untuk membuat *data mart*, yaitu:
    - Skema bintang (*star schema*)
    - Skema bola salju (*snowflake Schema*)
    - *Fact constellations*

Pendekatan Dimensional Modeling

- Dalam *dimensional modeling*, ada beberapa pendekatan yang digunakan untuk membuat *data mart*, yaitu:
  - Skema bintang (*star schema*)
  - Skema bola salju (*snowflake Schema*)
  - *Fact constellations*

Skema Bintang (*star schema*)

- Pada perancangan data mart pemantauan cuaca ini, menggunakan skema bintang.
- Skema ini mengikuti bentuk bintang, dimana terdapat satu tabel fakta (*fact table*) di pusat bintang dengan beberapa tabel dimensi (*dimensional tables*) yang mengelilinginya. Semua tabel dimensi berhubungan dengan ke tabel fakta. Tabel fakta memiliki beberapa *key* yang merupakan kunci indeks individual dalam tabel dimensi. Berikut model skema bintang :

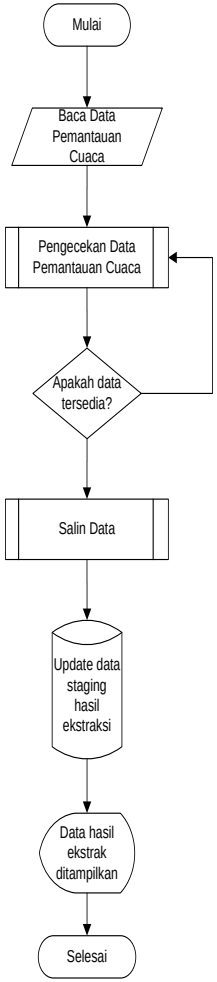


ETL(*Extract, Transform, Loading*)

- ETL merupakan proses yang sangat penting dalam *data mart*, dengan ETL inilah data dari operational dapat dimasukkan ke dalam *data mart*. ETL dapat digunakan untuk mengintegrasikan data dengan sistem yang sudah ada sebelumnya. Tujuan ETL adalah mengumpulkan, menyaring, mengolah, dan menggabungkan data-data yang relevan dari berbagai sumber untuk disimpan ke dalam *data mart*. Hasil dari proses ETL adalah dihasilkannya data yang memenuhi kriteria *data mart* seperti data yang historis, terpadu, terangkum, statis, dan memiliki struktur yang dirancang untuk keperluan proses analisis.
    a. *Extract*
    b. *Transform*
    c. *Loading*

Proses *Extract*

- Langkah pertama pada proses ETL adalah mengekstrak data dari sumber -sumber data. *Extraction* merupakan proses untuk mengidentifikasi seluruh sumber data yang relevan dan mengambil data dari sumber data tersebut. *Data mart* dapat menggabungkan data dari sumber-sumber yang berbeda dengan sistem-sistem terpisah yang menggunakan format data yang berbeda.Berikut adalah proses ekstrak pada *data mart* :



Proses *Extract* Tabel

- Proses ektrak tabel temperatur
  - Pada ektrak data disini, dilakukan proses pengambilan data dari sumber *database* operasional. Kolom yang diekstrak adalah kolom KodeTemperatur, T1, T2, T3 dan RataRata. Hasil dari ekstraksi data tabel temperatur dapat dilihat pada tabel berikut :

KodeTemperatur	T1	T2	T3	RataRata
T05001	19.6	27.2	22.2	22.2
T05002	21.3	28.5	25	24
T05003	20	28.4	26.8	23.8
T05004	21.2	29.5	24.3	24.1
T05005	21	27	24.2	23.3

Proses *Extract* Tabel

- Proses ekstrak tabel kelembaban
  - Pada ektrak data disini, dilakukan proses pengambilan data dari sumber *database* operasional. Kolom yang diekstrak adalah kolom KodeKelembaban, K1, K2, K3 dan RataRata. Hasil dari ekstraksi data tabel kelembaban dapat dilihat pada tabel berikut :

KodeKelembaban	K1	K2	K3	RataRata
K05001	91	65	91	85
K05002	91	62	75	80
K05003	87	62	67	76
K05004	93	55	81	81
K05005	91	73	87	86

Proses *Extract* Tabel

- Proses ektrak tabel angin
  - Pada ektrak data disini, dilakukan proses pengambilan data dari sumber *database* operasional. Kolom yang diekstrak adalah kolom KodeAngin, KecRataRata, ArahTerbanyak, KecTerbesar dan Arah. Hasil dari ekstraksi data tabel angin dapat dilihat pada tabel berikut :

KodeAngin	KecRataRata	ArahTerbanyak	KecTerbesar	Arah
A05001	6	W	10	N
A05002	4	W	5	W
A05003	0	C	0	C
A05004	7	W	10	W
A05005	5	W	5	W

Proses *Extract* Tabel

- Proses ekstrak tabel kondisi
  - Pada ektrak data disini, dilakukan proses pengambilan data dari sumber *database* operasional. Kolom yang diekstrak adalah kolom KodeKondisi, KodeTemperatur, KodeKelembaban, KodeAngin, CurahHujan dan Tanggal. Hasil dari ekstraksi data tabel kondisi dapat dilihat pada tabel berikut :

KodeKondisi	KodeTemperatur	KodeKelembaban	KodeAngin	CurahHujan	Tanggal
N05001	T05001	K05001	A05001	1.6	2005-01-01
N05002	T05002	K05002	A05002	2	2005-01-02
N05003	T05003	K05003	A05003	0	2005-01-03
N05004	T05004	K05004	A05004	1	2005-01-04
N05005	T05005	K05005	A05005	0	2005-01-05

Proses *Transform*

- Setelah dilakukan tahapan ekstrak, proses transformasi dilakukan untuk melakukan perubahan dan integrasi skema serta struktur yang berbeda-beda kedalam skema dan struktur yang terdefinisi dalam *data mart.* Tahapan transformasi menggunakan serangkaian aturan atau fungsi untuk mengekstrak data dari sumber dan selanjutnya akan dimasukkan ke *data mart*. Proses ini untuk menyesuaikan data apa saja yang akan di *load* ke dalam target, melakukan *join* dengan sumber data lain, melakukan agregasi, sorting dan filter.
- Fungsi transformasi yang dilakukan antara lain :
  1. Cleaning
  2. Conditioning

Proses *Cleaning*

- Pada tahap ini proses *cleaning* dilakukan untuk data- data yang tidak valid dan data yang bernilai null. Data yang yang tidak valid dan bernilai null maka data tersebut tidak diambil. Penjelasan dari proses *cleaning* adalah sebagai berikut :
- Tabel Temperatur
  - Pada tabel temperatur, kolom T1, T2 dan T3 tidak akan digunakan dalam proses transformasi, karena tidak digunakan dalam proses analisis. Hasil dari *cleaning* tabel temperatur adalah sebagai berikut :

KodeTemperatur	RataRata
T05001	22.2
T05002	24
T05003	23.8
T05004	24.1
T05005	23.3

Proses *Conditioning*

- Pada tahap ini dilakukan perubahan nama *field*, tipe data pada masing- masing tabel dan atribut dari sumber data ke target data yaitu *data mart*. Penjelasan *conditioning* pada proses transformasi adalah sebagai berikut :
- Tabel- tabel yang berada dalam sumber data akan dipilih dan diubah namanya dan dimasukkan ke dalam *database* target (*data mart*) artinya terdapat dua *database* yang berbeda yaitu db_cuaca yang menjadi sumber data dan *database* db_mart_cuaca yang menjadi *database* target.

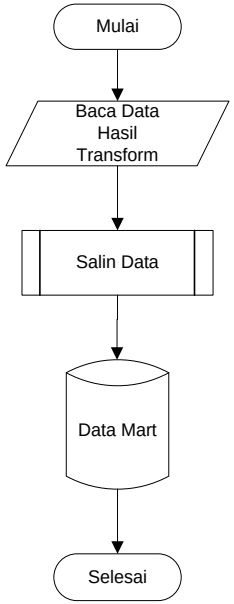
Proses *Conditioning*

- Perubahan data yang berbentuk kontinyu ke bentuk diskrit. Data sampel yang digunakan pada penelitian ini mempunyai atribut yang nilainya numerik, sedangkan data yang dibutuhkan adalah data yang nilainya nominal. Untuk itu atribut yang nilainya numerik tersebut diganti dengan atribut bernilai nominal yang menunjukkan interval nilai dengan nilai diskrit. Proses ini dikenal sebagai diskritisasi yang berisi transformasi dari *variable quantitative* kedalam variabel kualitatif. Hasil dari perubahan bentuk data kontinyu ke dalam bentuk diskrit lebih jelasnya dapat dilihat pada tabel berikut ini :

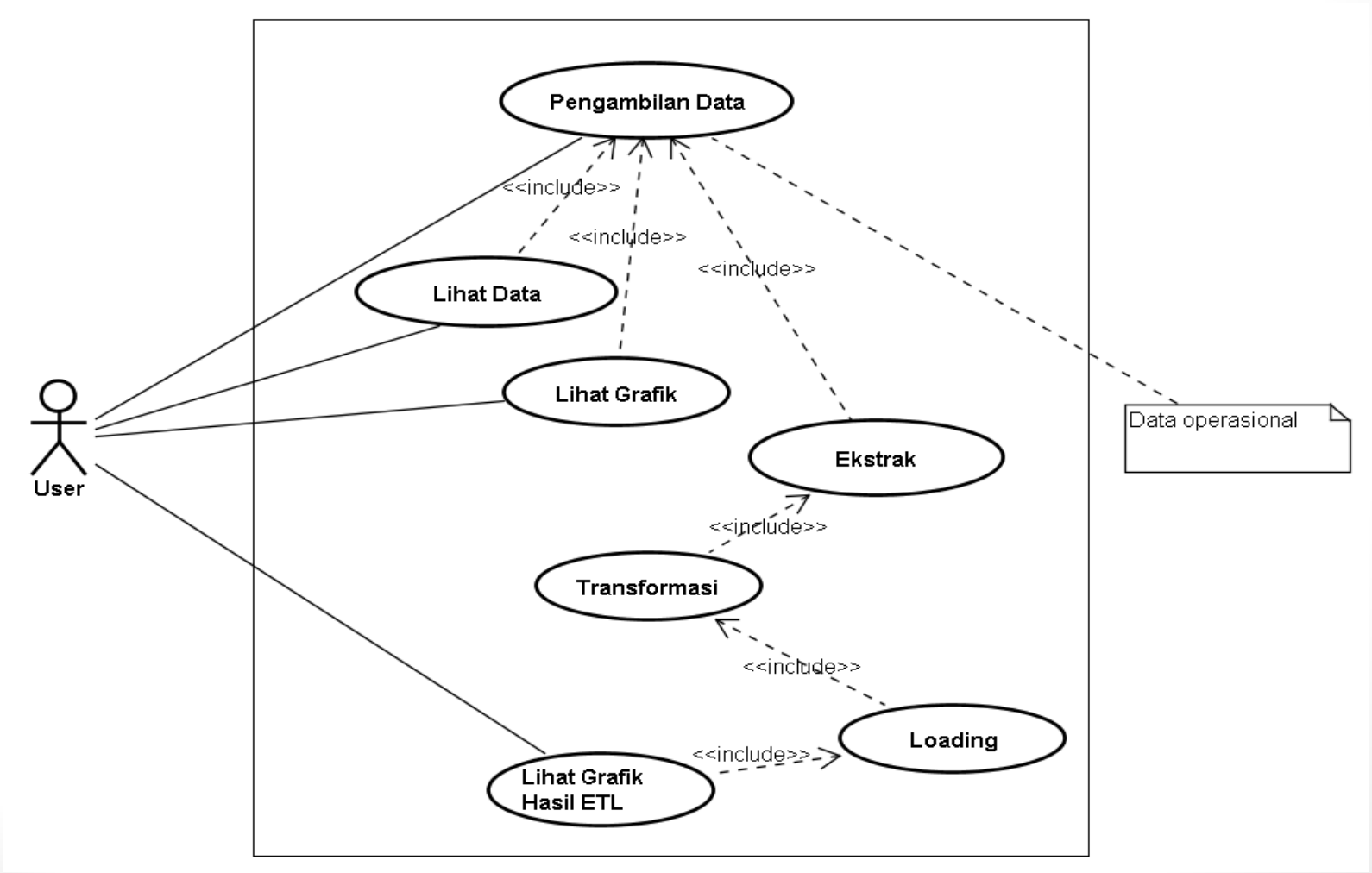
Atribut	Nilai Kontinyu	Nilai Diskrit
Temperatur	temperatur<=21.0	Temperatur sangat rendah
Temperatur	21.0<temperatur<=22.0	Temperatur rendah
Temperatur	22.0<temperatur<=23.0	Temperatur sedang
Temperatur	23.0<temperatur<=24.0	Temperatur agak tinggi
Temperatur	24.0<temperatur<=25.0	Temperatur tinggi
Temperatur	temperatur>25.0	Temperatur sangat tinggi
Kelembaban	kelembaban<=70	Kelembaban sangat rendah
Kelembaban	70<kelembaban<=75	Kelembaban rendah
Kelembaban	75<kelembaban<=80	Kelembaban sedang
Kelembaban	80<kelembaban<=85	Kelembaban agak tinggi
Kelembaban	85<kelembaban<=90	Kelembaban tinggi
Atribut	Nilai Kontinyu	Nilai Diskrit
Kelembaban	kelembaban>90	Kelembaban sangat tinggi
Kec Angin	kec.angin<=2	Kec angin sangat rendah
Kec Angin	2< kec.angin <=4	Kec angin rendah
Kec Angin	4< kec.angin <=6	Kec angin sedang
Kec Angin	6< kec.angin <=8	Kec angin agak tinggi
Kec Angin	8< kec.angin <=10	Kec angin tinggi
Kec Angin	kec.angin >10	Kec angin sangat tinggi
Curah Hujan	Curah hujan=0	Tidak hujan
Curah Hujan	0<Curah hujan<=5	Curah hujan sangat rendah
Curah Hujan	5<Curah hujan<=20	Curah hujan rendah
Curah Hujan	20<Curah hujan<=50	Curah hujan sedang
Curah Hujan	50<Curah hujan<=100	Curah hujan tinggi
Curah Hujan	Curah hujan>100	Curah hujan sangat tinggi

Proses Loading

- Mendistribusikan data hasil transformasi



Use Case


Pengambilan Data
<<include>>
<<include>>
<<include>>
Lihat Data
Lihat Grafik
Ekstrak
<<include>>
Transformasi
<<include>>
Loading
<<include>>
Lihat Grafik
Hasil ETL
User
Data operasional

Kesimpulan dan Saran

- Kesimpulan yang dapat ditarik setelah melakukan analisis, perancangan, pembangunan dan pengujian terhadap sistem simulasi data preparation adalah sebagai berikut :
  - Analisis data preparation yang dilakukan pada penelitian ini dapat menyediakan data yang berkualitas dan konsisten.
  - Dari hasil data preparation yang dilakukan ternyata data sudah memenuhi syarat untuk digunakan dalam proses analisis

- Saran yang diajukan dalam pengembangan analisis *data preparation* ini adalah melakukan penelitian lanjutan untuk menemukan metode analisis yang tepat pada *data mart* ini.

DEMO PROGRAM