

RANCANG BANGUN DATA WAREHOUSE DAN R BERBASIS WEB

Dian Anggraeni¹, Lutfi Ali Muharom², Alfian Futuhul Hadi³

^{1,3} Staf Pengajar Program Studi Matematika Universitas Jember

² Mahasiswa Program Studi Magister Matematika Universitas Jember

^{1,3} Email : anggraeni1602@gmail.com, afhadi@gmail.com

Abstrak

Pemanfaatan teknologi informasi sudah merupakan kebutuhan yang mendesak untuk melihat informasi dari berbagai sumber didalam pengambilan keputusan. Informasi membutuhkan suatu data yang dikelola dalam bentuk tertentu untuk memberikan arti. Perekaman data merupakan suatu aktifitas atau kejadian yang tersimpan dan membutuhkan suatu pengolahan yang baik. Data *warehouse* adalah salah satu bentuk pengolahan data yang nantinya digunakan untuk mendukung proses pengambilan keputusan tersebut. Data *warehouse* ini dimulai dari tahapan pengumpulan data, pemilihan data, perancangan data *warehouse* dan pemuatan data ke *warehouse*. Analisa data akan menggunakan program R yang terimplementasi berbasis web dan terintegrasi dengan data *warehouse*. R yang terimplementasi berbasis web akan mempermudah pengguna didalam pengoperasiannya. Perkembangan teknologi yang semakin maju, memungkinkan menjalankan R dengan browser. Dengan adanya data *warehouse* dapat dihasilkan koleksi data yang terstruktur dan integrasi dengan R dapat digunakan sebagai analisa dalam proses pengambilan keputusan.

Kata Kunci: data *warehouse*, pengambilan keputusan, R

1 Pendahuluan

Sejarah telah menunjukkan bahwa ada korelasi yang amat kuat antara kemampuan IPTEK suatu negara dengan tingkat pendapatannya per kapita. Makin tinggi kemampuan IPTEK suatu negara makin tinggi pula tingkat pendapatan per kapitanya. Perlu dicatat bahwa negara-negara kaya yang hanya memiliki 20% dari penduduk bumi, tetapi sangat tinggi kemampuan IPTEK-nya, ternyata menguasai 80% dari pendapatan global [7]

Jika kita perhatikan, Indonesia telah mengalami kenaikan pendapatan perkapita setiap tahunnya. Menurut data yang bersumber dari bps [http://www.bps.go.id/tab_sub/view.php?tabel=1&id_subyek=11¬ab=76], setiap tahunnya peningkatan pendapatan perkapita Indonesia rata-rata adalah Rp 300.309,04.

Sedangkan jika dibandingkan dengan negara-negara di ASEAN, menurut data [[http://en.wikipedia.org/wiki/List_of_countries_by_GDP_\(PPP\)_per_capita](http://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita)]. Negara Singapura masih menjadi negara dengan pendapatan perkapita tertinggi di ASEAN dengan US\$ 57,238, disusul dengan Brunei dengan pendapatan perkapita US\$ 47,200. Di urutan ketiga Malaysia dengan pendapatan perkapita US\$ 14,603. Di urutan keempat Thailand dengan pendapatan perkapita US\$ 8,643. Disusul Indonesia dengan pendapatan perkapita US\$ 4,380. Di urutan keenam Philippines dengan pendapatan perkapita US\$ 3,725. Vietnam berada di urutan selanjutnya dengan pendapatan perkapita US\$ 3,123. Di urutan kedelapan, Laos dengan pendapatan perkapita US\$ 2,435. Burma berada pada urutan ke Sembilan dengan pendapatan perkapita US\$

1,900. sedangkan Negara Kamboja menjadi negara dengan pendapatan perkapita terendah di ASEAN.

Perguruan Tinggi Sebagai salah satu pusat riset yang didalamnya terhimpun sarana dan prasarana IPTEK yang cukup mutakhir dan tersedia cukup sumber daya manusia berkualitas tinggi. Salah satu hambatan didalam penelitian / riset adalah masalah ketersediaan data, salah satu contoh adalah data sensus dan survei. Data sensus dan survei cukup sulit diakses, dengan birokrasi yang berbelit ini akan menghambat penelitian yang dilakukan oleh perguruan tinggi.

Salah satu langkah yang perlu diambil adalah dengan membangun suatu data *warehouse* tersendiri yang berfungsi untuk menampung data dan angka yang dihasilkan oleh sensus dan survei. Kemudahan dalam mengakses data ini akan menjadikan penelitian terutama dalam bidang statistika akan semakin berkembang. Diharapkan hasil-hasil penelitian ini menjadi salah satu acuan pemerintah maupun swasta didalam mengambil kebijakan sehingga memberikan dampak baik langsung maupun tidak langsung pada pertumbuhan ekonomi.

Data *warehouse* adalah koleksi data yang bersifat *subjectoriented*, terintegrasi, *time-variant*, dan *non-volatile* yang digunakan untuk mendukung proses pengambilan keputusan yang strategis organisasi [9]. Data yang ada pada perguruan tinggi yang begitu besar dan banyak membutuhkan alokasi tempat penyimpanan akan terbantu dan menjadi efisien dengan adanya data *warehouse*[6].

R adalah salah satu dari program sumber terbuka yang dapat diunduh gratis di <http://cran.r-project.org/>, yang mana situs resminya ada di <http://www.r-project.org/>. R pertama kali diciptakan oleh Ross Ihaka dan Robert Gentleman (- nama R berasal dari huruf pertama nama depan kedua orang ini), mereka adalah Statistawan asal Selandia Baru, sedangkan kini R dikembangkan sebagai upaya kolaborasi pakar-pakar Statistawan dan komputer di seluruh dunia. R diluncurkan pertama kali tahun 1997, dan versi terakhir ketika tulisan ini dibuat adalah 3.1.0. Bagi sebagian pengguna yang terbiasa mengguna GUI dengan sistem menu, R juga menyediakan banyak GUI yang berbasis sistem menu, antara lain R Studio, Tinn-R, R Commander dan banyak lagi yang lainnya, dan dapat diunduh gratis juga.

R Studio adalah salah satu GUI untuk R, salah satu keunggulan pada *rstudio* ini adalah dapat dijalankan pada browser. Sehingga dengan dijalankan diatas browser, maka pengguna tidak memerlukan lagi instalasi R, kecuali paket(*package*) pemrograman sesuai dengan kebutuhan pengguna.

2 Metode Penelitian

2.1 Perancangan Arsitektur

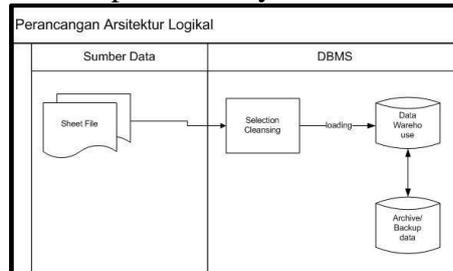
Proses perancangan arsitektur data warehouse adalah proses yang sangat rumit, untuk memulainya kita harus mendefinisikan kebutuhan dari pengguna. Pada tahapan ini terdiri dari perancangan arsitektur logikal dan perancangan arsitektur fisik. Data yang akan digunakan sebagai data awal dalam proses perancangan ini adalah data SUSENAS BPS Indonesia [2].

1. Perancangan Arsitektur Logikal

Pada perancangan arsitektur logikal pada data warehouse, sumber data yang akan digunakan adalah sumber data yang berasal dari kegiatan SUSENAS BPS. Data yang digunakan atau diperoleh untuk data SUSENAS tahun 1997 – 2012 adalah

berupa format *.csv, dimana didalam data tersebut terdapat data berupa kode administratif wilayah (mulai dari propinsi sampai dengan desa) yang dijadikan sampel dan data berupa nilai angka dari survey yang dilakukan.

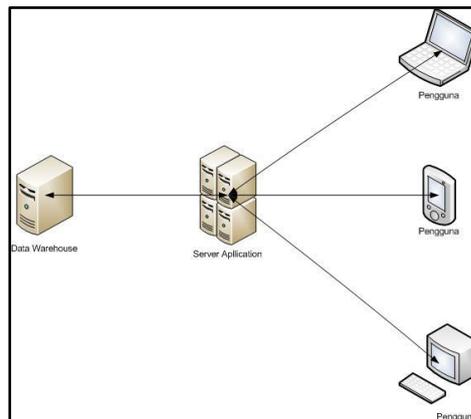
Dari sumber tersebut, kemudian dilakukan proses seleksi yaitu proses pemilihan data yang diperlukan dari sumber tersebut, karena bisa jadi dalam pembuatan data warehouse tidak semua data akan digunakan. Proses yang akan dilakukan selanjutnya adalah proses *cleansing* yaitu proses pembersihan data dan proses transformasi. Proses terakhir yang dilakukan adalah proses loading yaitu proses memasukkan data hasil dari tahap sebelumnya ke dalam data warehouse.



Gambar 1. Perancangan Arsitektur Logikal

2. Perancangan Arsitektur Fisik

Arsitektur Fisik data warehouse yang didapat data SUSENAS BPS akan terlihat sebagai berikut :



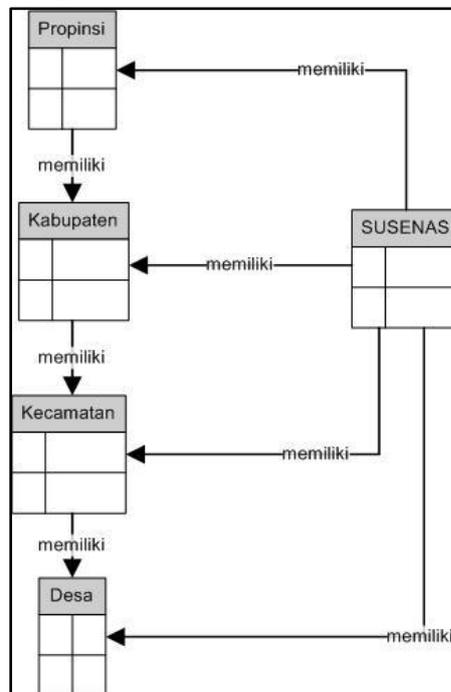
Gambar 2. Perancangan Arsitektur Fisik

2.2 Pemodelan Data

Skema yang digunakan untuk pemodelan data adalah dengan menggunakan relationship schema. Penggunaan skema ini memungkinkan proses query yang lebih ringan dan memudahkan penjelajahan terhadap data dimensinya. Tabel tabel yang dihasilkan antara lain:

- Tabel Propinsi
Tabel ini berisi data propinsi dan kode propinsi yang ada di Indonesia
- Tabel Kabupaten
Tabel ini berisi data kabupaten pada setiap propinsi
- Tabel Kecamatan
Tabel ini berisi data kecamatan pada setiap kabupaten

- Tabel Desa
Tabel ini berisi data desa/kelurahan pada setiap kecamatan
- Tabel SUSENAS
Tabel yang berisi catatan survey terhadap sampel yang diambil pada setiap desa, kecamatan, kabupaten dan propinsi yang ada berdasarkan variabel yang menjadi variabel survey pada SUSENAS



Gambar 3. Skema Data

2.3 Implementasi R server

Ada beberapa alasan mengapa aplikasi berbasis web, antara lain :

1. Kita dapat menjalankan aplikasi berbasis web dimanapun kapanpun tanpa harus melakukan penginstalan.
2. Terkait dengan isu lisensi (hak cipta), kita tidak memerlukan lisensi ketika menggunakan web-based application, sebab lisensi telah menjadi tanggung jawab dari web penyedia aplikasi.
3. Dapat dijalankan di system operasi manapun. Tidak peduli apakah kita menggunakan linux, windows, aplikasi berbasis web dapat dijalankan asalkan kita memiliki browser dan akses internet.
4. Dapat diakses lewat banyak media seperti: *computer*, *handheld* dan *handphone* yang sudah sesuai dengan standard WAP.
5. Tidak perlu spesifikasi computer yang tinggi untuk menggunakan aplikasi berbasis web ini, sebab di beberapa kasus, sebagian besar proses dilakukan di web server penyedia aplikasi berbasis web in.

Server yang menjadi uji coba installasi R, ini adalah server blade dengan 8 inti prosessor dengan kecepatan masing-masing prosessor adalah xxx dan menggunakan

memori sebesar 4Gb. Sistem Operasi yang digunakan adalah linux server 12.04 Precise. Langkah-langkah instalasi adalah sebagai berikut:

1. Instalasi OS Linux server 12.04
2. Edit pada source.list arahkan pada mirror cran. Pada uji coba ini cran diarahkan pada mirror unej.ac.id

```
## r installation package from mirror.unej.ac.id
deb http://cran.unej.ac.id/bin/linux/ubuntu precise/
deb http://cran.unej.ac.id/raring-backports main restricted universe
```

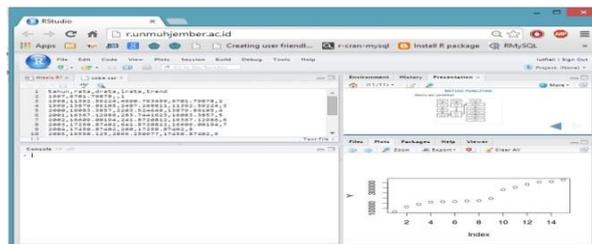
3. Kemudian install paket R sistem. Gunakan perintah berikut :
sudo apt-get install r-base-dev
4. Instalasi paket R pada ubuntu membutuhkan kunci yang dibuat oleh "Michael Rutter". Gunakan perintah berikut:
sudo apt-key adv --keyserver keyserver.ubuntu.com --recv-keys E084DAB9
5. Langkah berikutnya adalah instalasi Rstudio server. Rstudio server ini adalah sebuah server yang berfungsi menjalankan program R lewat browser. Gunakan perintah berikut (OS 64 bit):

```
$ sudo apt-get install gdebi-core
$ sudo apt-get install libapparmor1 # Required only for Ubuntu, not Debian
$ wget http://download2.rstudio.org/rstudio-server-0.98.1091-amd64.deb
$ sudo gdebi rstudio-server-0.98.1091-amd64.deb
```

6. Langkah berikutnya adalah pada konfigurasi server. Hal ini diperlukan karena pada setting default dari Rserver menggunakan port <8787>. Agar lebih mudah mengakses maka kita gunakan port <80> dan address dari 0.0.0.0. Penggunaan alamat (address) 0.0.0.0 ini memungkinkan akses R dari berbagai IP. Gunakan perintah berikut :
/etc/rstudio/rserver.conf
www-port=8787 -> ubah menjadi 80
www-address=127.0.0.1 -> ubah menjadi 0.0.0.0
7. Selesai

3 Hasil dan Pembahasan

Data warehouse harus memiliki kapabilitas dalam menyediakan informasi yang lengkap pada pengguna, oleh karena itu diperlukan perangkat antarmuka yang menyediakan informasi kepada pengguna. Presentasi kepada pengguna dikembangkan sebuah sistem berbasis PHP, DBMS yang digunakan sebagai data warehouse platform adalah MySQL. Selain itu pengguna dari R juga mempunyai kemampuan untuk berinteraksi atau berkoneksi secara langsung dengan data warehouse, karena antara sistem R dan data warehouse dapat berkomunikasi melalui paket RMySQL.



Gambar 4. Aplikasi R berbasis web

Kemampuan dari data warehouse dan sistem R ini adalah :

1. Kemampuan untuk menampilkan data dengan tingkat rincian yang lebih rendah
2. Kemampuan membuat query sendiri sesuai dengan kebutuhan
3. Kemampuan berintegrasi dengan program R
4. Kemampuan untuk ekspor dalam format excel atau pdf.

4 Kesimpulan

Rancang bangun data warehouse ini mampu berintegrasi dengan R. R dapat diakses dari segala komputer yang terhubung dengan jaringan (internet) karena dibangun menggunakan platform web. Penelitian ini nantinya dapat dilanjutkan dengan memanfaatkan metode statistika untuk menggali informasi dari data menjadi knowledge yang dapat digunakan sebagai pendukung keputusan.

Daftar Pustaka

- [1] Bishop, C.M. 1995. *Neural Networks for Pattern Recognition*. New York: Oxford University Press Inc.
- [2] Data SUSENAS 2000 - 2010, <www.ada.edu>. Diunduh pada 25 - 28 Maret 2014
- [3] FAO. 2012. *The State of Food Insecurity in The World 2012*. Rome
- [4] J. Supranto, 2000, *Statistik Teori dan Aplikasi*, Erlangga, Jakarta 3.
- [5] Kementan. *Laporan Kinerja Kementerian Pertanian Tahun 2011*. Jakarta: Kementan, 2012.
- [6] Mukhlis F., Bayu Adhi T., *Rancang Bangun Data Warehouse Untuk Menunjang Evaluasi Akademik Di Fakultas., Prosiding Konferensi Nasional Teknologi Informasi dan Aplikasinya (KNTIA)*, Oktober 2011
- [7] Prayoto, *Peranan Perguruan Tinggi Dalam Pengembangan IPTEK, Seminar Nasional Dies Natalis 45 UGM*, 2008
- [8] Pusdatin Kementan. *Statistik Kosumsi Pangan Tahun 2012*. Jakarta, Kementan 2012.
- [9] Radityo A.N., Johan T., Tony J, H., *Aplikasi Data Warehouse untuk Analisa Penjualan Mobil Berbasis Multidimensional Modeling (MDM) dan Star Schema Design*, *Jurnal Teknologi Informasi -AITI*, Vol 5 No 2, Agustus 2008