ANALISIS DATA *MART* UNTUK VIRTUALISASI *TRAFFIC* MONITORING INSIDEN DI ID-SIRTII/CC

Mira Kania Sabariah, Fiska Mekar Kustiani
Universitas Komputer Indonesia (UNIKOM)
Jl. Dipati Ukur No. 112-116, Bandung 40132
Email : mira_ljuan@yahoo.com, vizz91agasi@gmail.com

ABSTRAK

Id-SIRTII/CC (*Indonesia Security Incident Response Team on Internet Infrastructure/Coordination Center*) merupakan pertahanan pertama untuk melindungi para pengguna di Indonesia dari kejahatan yang mungkin terjadi di dunia maya berupa sebuah tim yang bertugas untuk menjaga keamanan informasi data di Indonesia, terutama dari serangan *cracker* yang banyak menyerang *web* dan server internet di Indonesia. Berdasarkan hasil wawancara terhadap staf *Laboratorium Data Mining Deputy of Researh and Development* Id-SIRTII/CC, untuk mengolah insiden web yang telah disusupi oleh peretas, staf mencari informasi di situs Zone-h (http://www.zone-h.org dan Indonesian Defacer (http://indonesiandefacer.org). Data tersebut digunakan untuk *traffic* monitoring insiden. Adapun tujuan dilakukan *traffic* monitoring adalah untuk mempermudah melihat website berdomain Indonesia yang telah disusupi oleh peretas. Saat ini, proses penyajian data masih dilakukan secara manual dan belum tervirtualisasi dengan baik, sehingga pada saat proses monitoring mengalami kesulitan saat akan melakukan analisis insiden dari berbagai kondisi.

Pada penelitian ini dilakukan pembangunan piranti lunak untuk memonitoring insiden web yang terjadi di Id-SIRTII/CC. Tujuannya adalah untuk memudahkan melakukan pemantauan website berdomain Indonesia yang disusupi oleh peretas dan peringatan terhadap ancaman serta gangguan pada website yang berdomain Indonesia. Untuk pengolahan data insiden menggunakan teknik *data mart*.

Dari hasil penelitian dapat disimpulkan bahwa virtualisasi data *traffic* monitoring insiden dapat mempermudah staf untuk melakukan pemantauan website berdomain Indonesia yang disusupi oleh peretas dan peringatan terhadap ancaman serta gangguan pada website yang berdomain Indonesia. Selain itu, teknik *data mart* dapat digunakan sebagai mengolah data insiden.

Kata kunci : *virtualisasi data, traffic, monitoring, insiden, data mart.*

1. PENDAHULUAN

Id-SIRTII/CC (*Indonesia Security Incident Response Team on Internet Infrastructure/Coordination Center*) merupakan pertahanan pertama untuk melindungi para pengguna di Indonesia dari kejahatan yang mungkin terjadi di dunia maya berupa sebuah tim yang bertugas untuk menjaga keamanan informasi data di Indonesia, terutama dari serangan *cracker* iseng yang banyak menyerang web dan server internet di Indonesia.

Berdasarkan wawancara dengan staf *Laboratorium Data Mining Deputy of Research and Development* Id-SIRTII/CC pada tanggal 10 September 2012, kasus atau insiden yang menimpa sistem informasi dan teknologi pendukung pemilu 2004 di Indonesia membuka mata masyarakat umumnya dan Id-SIRTII/CC khususnya, akan besarnya ancaman keamanan yang dapat menimpa berbagai sistem berskala nasional apapun yang ada di tanah air. Bila eksploitasi tersebut diabaikan dan terjadi pada obyek vital yang ada di Indonesia, seperti pada sistem pembayaran nasional, sistem distribusi listrik, sistem persenjataan militer, sistem pelabuhan udara, dan lain sebagainya, akan melumpuhkan semua aktifitas yang sedang dilakukan.

Dalam hal ini, jika terjadi insiden web yang telah disusupi oleh peretas, staf *Laboratorium Data Mining Deputy of Research and Development* Id-SIRTII/CC mencari informasi di situs Zone-h (http://www.zone-h.org) dan Indonesian Defacer (http://indonesiandefacer.org). Data tersebut digunakan untuk *traffic monitoring* insiden. Adapun tujuan dilakukan *traffic monitoring* adalah untuk mempermudah melihat website berdomain Indonesia yang telah disusupi oleh peretas. Saat ini, proses penyajian data masih dilakukan secara manual dan belum tervirtualisasi dengan baik, sehingga pada saat proses *monitoring* mengalami kesulitan saat akan melakukan analisis insiden dari berbagai kondisi.

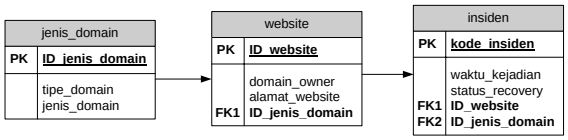
Dari permasalahan tersebut, maka *Laboratorium Data Mining Deputy of Research and Development* di Id-SIRTII/CC membutuhkan alat bantu kerja berupa piranti lunak untuk memonitoring insiden web berbasis protokol internet di Indonesia yang nantinya dapat menyimpan rekaman transaksi (*log file*). Sedangkan tujuan yang akan dicapai dalam

penelitian ini adalah memudahkan virtualisasi terhadap monitoring website berdomain Indonesia yang disusupi oleh peretas dan peringatan terhadap ancaman serta gangguan pada website yang berdomain Indonesia.

2. ANALISIS DATA *MART*

2.1 Analisis Sumber Data

Tahap awal yang dilakukan pada saat akan membuat sebuah data mart adalah melakukan analisis terhadap sumber data. Data masukan merupakan data yang disajikan dalam bentuk format Excell yang kemudian diubah ke dalam bentuk data relasional. Pada gambar 1 adalah skema relasi dari data masukan awal berupa Excell. Data tersebut berasal dari *Laboratorium Data Mining Deputy of Research and Development* Id-SIRTII/CC saat ini, berikut skema relasi yang digunakan :



Gambar 1. Skema Relasi Data Sumber

2.2 Analisis Arsitektur *Data Mart*

Arsitektur data menyediakan kerangka dengan mengidentifikasikan dan memahami bagaimana data akan dipindahkan melalui sistem dan digunakan untuk analisis. Arsitektur data untuk *data mart* mempunyai komponen utama yaitu *read-only database*.

Arsitektur yang akan digunakan adalah *Two – Layer Architecture*. Arsitektur ini terdiri dari 4 lapisan aliran data, yaitu :

1. Lapisan pertama adalah *source layer*. Pada lapisan ini, data masih berupa operasional data. Data operasional yang akan digunakan pada pembangunan *data mart* kali ini sudah berupa data *logic* yang ada di *database server*.
2. Lapisan kedua adalah *data staging*. Pada lapisan ini, data operasional akan diekstrak (lebih dikenal dengan proses *ETL*) ke dalam *data mart*.
3. Lapisan ketiga adalah *data mart layer*. Informasi akan disimpan pada sebuah penyimpanan *logic* yang tersentralisasi, yaitu *data mart*. *Data mart* dapat diakses secara langsung, dan juga bisa digunakan untuk keperluan *Online Analysis Processing* (OLAP).
4. Lapisan keempat adalah *analysis*. Analisis disini nantinya akan menggunakan OLAP.

Secara Umum terdiri dari empat bagian yaitu Sumber data, Data *Staging*, Penyimpanan Data, dan Analisis. Penjelasan dari tiap-tiap bagian adalah sebagai berikut :

1. *Source layer*

Berasal dari *Database* Operasional (*Online transaction Processing*(OLTP)) dengan format Ms Ms Excel yang diubah kedalam sebuah Skema Relasi yang terlihat pada gambar 1.

2. *Data Staging*

Pada bagian data *staging*, dilakukan proses berikut:

- Pemilihan sumber data yaitu database *Ms Excel* yg nantinya akan diolah menjadi data *mart*.
- Pengecekan database Ms Excel layak atau tidak untuk dijadikan data *mart.*
- Setelah dinyatakan layak maka dilakukan ekstraksi data dengan mengambil sumber data yaitu database Ms Excel.
- Proses selanjutnya adalah *transform.* Tabel-tabel yang sudah di*ekstrak* nantinya dipilih lagi sesuai strategi bisnis yang ditentukan, kemudian dilakukan juga pemilihan kolom, pengubahan nama dan pengambilan field tanggal menjadi dimensi waktu.
- Proses selanjutnya adalah *Load*. Semua proses *transform* tadi menghasilkan sedikit tabel dengan kolom yang penting. Load disini memasukan berfungsi untuk memasukan data hasil *transform* di data *stagging* ke data *mart.* Cara lain bisa juga proses *transform* dan *Load* disatukan sehingga nantinya begitu tabel-tabel sudah dipilih, maka langsung di *load* ke data *mart.*

3. Penyimpanan Data

Hasil dari *ETL*, akan disimpan ke dalam Data *mart.* nantinya akan digunakan untuk proses analisis.

4. Analisis

Dari Data *mart* yang dibuat dapat dilakukan total jumlah insiden web berdasarkan teknik *roll-up* dan *drill-down.*

2.3 Analisis ETL

Proses yang dilakukan pertama kali dalam data *preprocessing* adalah proses Ekstrak. Sebagai tahap awal dalam proses ekstrak adalah melakukan *import* data dari sumber data berupa *file* Ms. Excel ke dalam basis data relasional yang dibuat sebagai tahap awal untuk memudahkan proses selanjutnya.

Proses selanjutnya setelah ekstrak adalah proses *transformation*, dimana dalam tahap *transformation* ini dilakukan proses *Cleaning* dan *Conditioning*. Berikut ini adalah tahapannya :

1. *Cleaning*

Tahap ini untuk membersihkan/meningkatkan kualitas data. Pada proses ini tabel-tabel dan kolom yang bernilai null dan tidak digunakan untuk proses selanjutnya tidak akan diambil.

Penjelasan dari proses *cleaning* adalah terlihat pada tabel 1, dimana pada proses ini status_recovery

yang merupakan salah satu *field* pada tabel insiden dihilangkan ketika proses *transformation*.

Tabel 1. Ilustrasi *Cleaning* Tabel Insiden

Nama Field	Tipe Data	Panjang Data	Kunci	Keterangan
kode_insiden	*int*		PK	*not null*
waktu_kejadian	*datetime*			*not null*
status_recovery	*varchar*	20		*not null*
ID_website	*int*		FK	*not null*
ID_jenis_domain	*int*		FK	*not null*

Nama Field	Tipe Data	Panjang Data	Kunci	Keterangan
kode_insiden	*int*		PK	*not null*
waktu_kejadian	*datetime*			*not null*
ID_website	*int*		FK	*not null*
ID_jenis_domain	*int*		FK	*not null*

2. *Conditioning*

Proses *conditioning* dilakukan dengan pemilihan tabel dan atribute dari sumber data ke target data (*data mart*). Penjelasan dari *conditioning* pada proses transformasi adalah sebagai berikut :

a. Tabel-tabel yang berada dalam sumber data akan dipilih dan diubah namanya dan dimasukkan kedalam database target (*data mart*). Perlu diperhatikan bahwa *database* yang menjadi sumber data (*data source*) berbeda dengan *database* target (*data mart*) artinya terdapat 2 *database* yaitu DBInsiden yang menjadi sumber data, dan DMInsiden yang menjadi target data.
b. Tabel Insiden merupakan Tabel Fakta dalam DMInsiden, sedangkan tabel Jenis Domain dan Tabel Website merupakan Tabel Dimensi dalam DMInsiden.
c. ID_website, ID_waktu, ID_jenis_domain sebagai FK dari fakta insiden. Untuk lebih jelas mengenai FK pada tabel fakta insiden, akan dijelaskan dengan tabel 2.

Tabel 2. Tabel Fakta Insiden

kode_ insiden	ID_ website	ID_ waktu	ID_ jenis_ domain	jumlah_ insiden
1	1	01	5	50
2	2	02	5	40
3	3	03	5	30
4	4	04	5	20

d. *Field* waktu_kejadian pada tabel insiden akan menjadi tabel dim_waktu (dengan field berupa ID_waktu, bulan, tahun) karena ketika proses analisis, data yang dibutuhkan bisa dianalisis lebih dalam berdasarkan *range* waktu yang diinginkan. Untuk lebih jelasnya mengenai tanggal dari tabel insiden yang dijadikan dimensi waktu akan dijelaskan dengan tabel 3.

Tabel 3. Tabel field waktu_kejadian pada tabel insiden menjadi dim_waktu

kode_ insiden	ID_ website	waktu_ kejadian	jumlah_ insiden
1	1	3/4/2012 0:00	50
2	2	31/3/2012 0:00	40
3	3	1/4/2012 0:00	30
4	4	2/4/2012 0:00	20
5	5	26/3/2012 0:00	60
6	6	2/4/2012 0:00	10

ID_ Waktu	Bulan	Tahun
01	4	2012
02	3	2012
03	4	2012

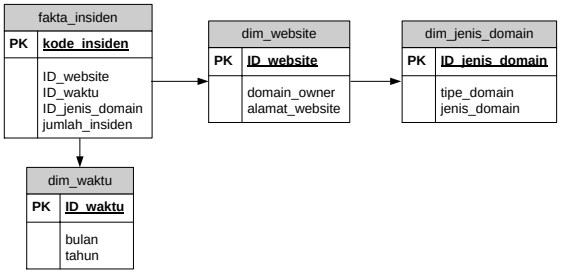
Pada proses *transformation*, data yang sudah melalui tahap *conditioning* dan *cleaning* di *load* kedalam data *mart*.

3. *Load*

Fase *Load* merupakan tahapan yang berfungsi untuk memasukkan data ke dalam target akhir yaitu data *mart* . Pada proses ini data OLTP insiden yang sudah dibaca dan diubah formatnya akan disimpan pada *data mart*. Dalam tahap ini, nantinya pihak *Laboratorium Data Mining* Id-SIRTII/CC akan melakukan *update* berkala.

2.4 Analisis Skema Data Mart

Berdasarkan hasil analisis dan kebutuhan dari *Laboratorium Data Mining* Id-SIRTII/CC, model skema data dimensional yang dibuat nantinya adalah *Snowflake*. Model ini dipilih karena dianggap sesuai untuk kebutuhan informasi yang ingin dianalisis yaitu melakukan analisis insiden *web*. Berdasarkan kebutuhan tersebut maka terbentuklah 1 tabel fakta dan 3 buah tabel dimensi. Dari 3 tabel dimensi yang terbentuk salah satu tabel dimensi yaitu tabel dimensi jenis_domain tidak berhubungan langsung dengan tabel fakta insiden, namun berhubungan dengan tabel dimensi website. Hal tersebut disebabkan karena diperlukannya penurunan dimensi website berdasarkan jenis domain sebagai salah satu kebutuhan analisis berdasarkan jenis domain. Dari kebutuhan tersebut maka *Snowflake Schema* tepat digunakan untuk skema Data *Mart* yang dibangun. Skema Data Mart yang dibuat dapat dilihat pada gambar 2.



Gambar 2. Tabel *Snowflake Schema Data Mart* Insiden

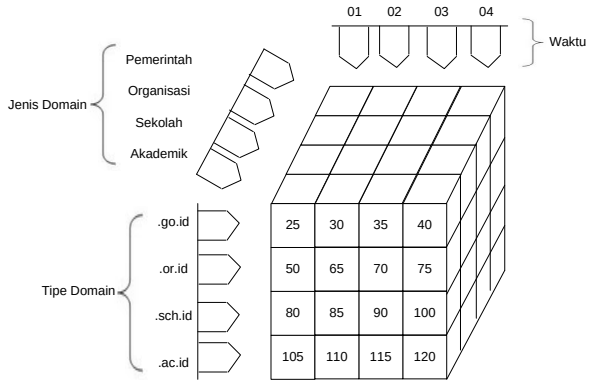
2.5 Pemodelan Data Dimensional

Untuk kebutuhan virtualisasi, digunakan teknik data dimensional dan metode OLAP berupa fakta_insiden dan beberapa tabel dimensi yaitu : dim_website, dim_jenis_domain dan dim_waktu. Tabel 4 dibawah ini merupakan ilustrasi dari Data Dimensional Data *Mart* Insiden.

Tabel 4. Ilustrasi Data Dimensional Data *Mart* Insiden

Tipe Domain	Jenis Domain	Waktu	Jumlah Insiden
.go.id	Pemerintah	01	50
.or.id	Organisasi	02	40
.sch.id	Sekolah	03	30
.ac.id	Akademik	04	20

Berikut ini adalah gambar 3 merupakan ilustrasi Data *Cube* Insiden dari tabel 4.



Gambar 3. Data *Cube* Insiden

Proses analisis akan dilakukan untuk menganalisa *traffic* insiden web yaitu dengan menggunakan OLAP, Nilai *atribute* seperti tanggal memiliki nilai yang menyatakan tahun, bulan, juga hari. Seringkali pengkategorian ini dapat diorganisasikan sebagai pohon hirarki. Struktur hirarki ini memunculkan operasi *Roll-up* dan *Drill-down*. Contohnya pada jenis instansi yang merupakan data *multidimensional* (fakta insiden), dapat ditentukan agregasi insiden untuk jumlah per hari dalam 1 bulan dan memisahkan total insiden per 3 bulan ke dalam total insiden bulanan.

Operator *roll up* menyebabkan peningkatan agregasi data dan menghapus level data yang lebih detail dari sebuah hirarki. Untuk data insiden, kita dapat mengumpulkan data (*Roll up*) website dari seluruh tanggal (harian) dalam satu bulan. Sebagai contoh, untuk menampilkan jumlah insiden per bulan. *Roll up* dapat menampilkan informasi website yang sering tersusupi berdasarkan periode total 3 bulanan. Hal ini dapat terlihat pada Tabel 5 dan 6 berikut ini :

Tabel 5. Tabel Data Multidimensional

Tipe Domain	Jenis Domain	Jumlah Insiden		
		April	Mei	Juni
.ac.id	Akademik	116	28	86
.web.id	Personal	96	95	58
.co.id	Perusahaan	185	63	92
.sch.id	Sekolah	390	57	179
.go.id	Pemerintah	63	75	163
.or.id	Organisasi	58	32	34
.mil.id	Militer	4	10	5
.net.id	Network	0	0	2

Menjadi :

Tabel 6. Tabel Hasil *Roll Up*

Tipe Domain	Jenis Domain	Jumlah Insiden
.ac.id	Akademik	227
.web.id	Personal	249
.co.id	Perusahaan	339
.sch.id	Sekolah	626
.go.id	Pemerintah	301
.or.id	Organisasi	124
.mil.id	Militer	19
.net.id	Network	2

3. PENUTUP

Berdasarkan hasil pengujian menggunakan OLAP bahwa dapat disimpulkan Data *Mart* yang dibuat dapat membantu memudahkan virtualisasi terhadap monitoring website berdomain Indonesia yang disusupi oleh peretas dan peringatan terhadap ancaman serta gangguan pada website yang berdomain Indonesia.

UCAPAN TERIMA KASIH

Terimakasih kami ucapkan kepada lembaga Id-SIRTII, khususnya *Laboratorium Data Mining Deputy of Researh and Development* Id-SIRTII/C yang telah bekerjasama sehingga penelitiannya ini dapat berjalan dengan baik.

DAFTAR PUSTAKA

[1] Rainardi, Vincent. 2008. "*Building a Data Warehouse with Examples in SQL Server*". New York:Springer.
[2] Santosa, Budi. 2007. *Data Mining : Teknik Pemanfaatan Data untuk Keperluan Bisnis Teori & Aplikasi.* Yogyakarta:Graha Ilmu.
[3] Sutanta, Edhy. 2011. *Basis Data dalam Tinjauan Konseptual*. Yogyakarta:Andi.
[4] Sommerville, Ian. 2001. "*Software Engineering*". 6th. Addison Wesley.