

Desain Model Integrasi Data Terstruktur (Heterogen) untuk Mendukung Analisis Big Data Kesehatan

Sudaryanto¹, Slamet Sudaryanto N², Fikri B³
Universitas Dian Nuswantoro
msdr8047@dsn.dinus.ac.id¹, slametalica301@dsn.dinus.ac.id²,
fikri.budiman@dsn.dinus.ac.id

Abstract

Big Data is triggered by several studies and prospects on the concepts and processes related to the data warehouse or data warehouse (DW) field. Some conclude that such Warehouse Data will disappear, then Big Data will appear as the natural evolution of the Data Warehouse. The focus of most of Big Data (BD) is very large data integration as large data source processing. The data source is viewed as a convergence area so the data warehouse is part of big data if it can be integrated. There are problems that are just as difficult as data warehouse and big data about how to normalize, integrate, and convert data from many sources into the format needed to carry out large-scale analysis and visualization tools. In previous research it has developed an approach to semi-automatically mapping multiple sources into shared domain ontologies so that they can be quickly combined. In this paper describes the approach to building and implementing integration and restructuring plans to support analysis and visualization tools in very large and diverse datasets. In writing this revises the technology features that underlie Big Data and Data Warehouse, highlighting their differences and convergence areas. Even when some differences exist, both technologies can (and should) be integrated because they aim at the same goal, namely data in exploration to support decision making. This study will explore several convergence strategies, based on common elements in both technologies. In presenting revisions of state-of-the-art in integration proposals from the point of view of the underlying objectives, methodology, architecture and technology, highlighting common elements that support both technologies that can serve as starting points for full integration and we propose a data integration model between two technologies..

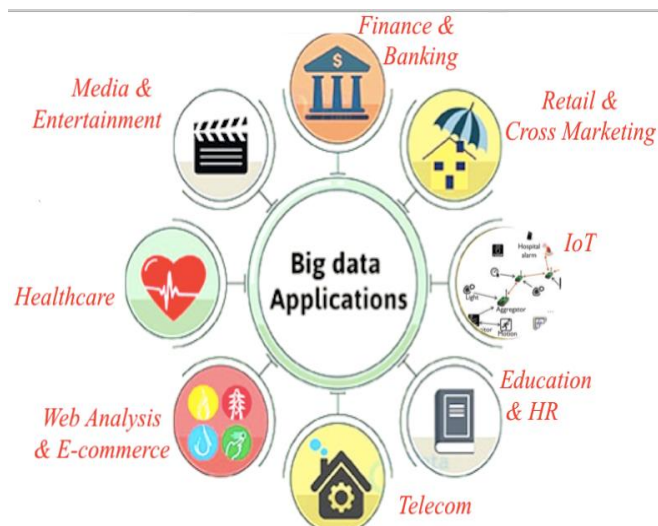
Keywords: Data Warehouse (DW), Big Data (BD), hybrid models, heterogeneous data, DSS models

1. PENDAHULUAN

Pengelolaan *Data Warehouse* (DW) atau *Big Data* (BD) sering menghabiskan lebih banyak waktu untuk mengekstrak, memformat ulang, dan mengintegrasikannya. Demikian juga untuk menjalankan program analisis dan visualisasi daripada mereka mengerjakan analisis itu sendiri. Program pembentukan ulang data semacam itu sulit untuk ditulis karena kompleksitas mereka, tetapi mereka diperlukan karena masing-masing analitik alat mengharapkan data dalam bentuk yang sangat spesifik dan untuk mendapatkan data ke dalam formulir itu biasanya membutuhkan serangkaian pembersihan, normalisasi, memformat ulang, integrasi, dan restrukturisasi operasi. Masalahnya dibuat lebih menantang dalam konteks "big data", di mana sumbernya bisa besar (jutaan hingga miliaran catatan, gigabita hingga terabyte ukuran), streaming, dan heterogen, yang berarti bahwa pengembang tidak mungkin meninjau semua data dan data persiapan harus dijalankan dengan cara yang skalabel pada skala besar ini kumpulan data. Untuk mencapai tujuan dengan cepat menerapkan analitik data dan alat visualisasi pada masalah baru membutuhkan sebuah pendekatan untuk menentukan dengan cepat rencana pembentukan ulang data yang benar dan lalu mengeksekusinya dengan cara yang skalabel. Data kesehatan selama ini belum digarap secara serius, data kesehatan masih tersebar heterogen dan terfragmentasi. Bahkan seorang

dokter dalam pemeriksaan seorang pasien yang sama tidak dapat mendapatkan informasi yang lengkap dan integral. Karena masing-masing rekam medis dari seorang pasien disimpan secara terpisah oleh masing-masing poliklinik atau rumah sakit, sehingga data yang harusnya memiliki kontribusi terhadap informasi yang lebih luas tidak memiliki manfaat sama sekali. Informasi adalah salah satu sumber daya paling berharga dari sebuah lembaga, dan penggunaan yang memadai untuk mendukung pengambilan keputusan telah menjadi tantangan yang semakin meningkatkan kompleksitas. Perusahaan berinvestasi dalam solusi yang memungkinkan mereka menggunakan data besar dengan cara sebaik mungkin, untuk menghasilkan strategi bisnis baru, meningkatkan layanan pelanggan, atau mengembangkan kebijakan publik, di antara banyak kegunaan lainnya seperti dalam bidang kesehatan. Saat ini volume data kesehatan yang diperlukan untuk diproses dalam suatu pengelola data kesehatan seperti dokter atau spesialis kesehatan dapat mencapai urutan *Exabyte* (Bendi dan Gautam, 2014). Hal ini menimbulkan tantangan penyimpanan dan pemrosesan yang membutuhkan solusi teknologi baru yang memungkinkan tidak hanya penyimpanan, tetapi juga memperbarui, eksploitasi efisien dan yang memiliki persyaratan data tertentu. Hal ini kadang disebut sebagai tujuh V's (Bendi dan Gautam, 2014): *Volume, Variety, Velocity, Veracity, Value, Variability and Viability* dan tiga C's (Bendi dan Gautam, 2014): *Cost, Complexity, dan Consistency*.

Mengingat keterbatasan teknik tradisional yang digunakan selama ini dan persyaratan data baru, pengelola data mengalami beberapa tantangan dalam mengelola volume data yang besar. Konsep Data Warehouse dan Big Data cenderung menyatu dan tidak mudah menemukan perbedaan di antara mereka. Sementara Data Warehouse adalah paradigma manajemen yang matang didukung oleh metodologi yang tersebar luas dan mapan (Kimball dkk, 2008). Big Data masih merupakan bidang yang sedang dikembangkan, yang berusaha untuk mengatasi aspek individual dari masalah tetapi masih tidak memiliki solusi integral. Sebagai hasil dari beberapa tinjauan, kita dapat menyimpulkan bahwa beberapa artikel menyajikan Big Data sebagai pengganti Data Warehouse, yang lain sebagai evolusi Data Warehouse (Maioreescu, 2010), beberapa mengusulkan perpanjangan Data Warehouse untuk mendukung beberapa karakteristik Big Data dan sebagian lainnya menjelajahi kemungkinan mengintegrasikan keduanya melalui mapping schema. *Mapping Schema* merupakan salah satu model integrasi untuk mendukung integrasi data yang lebih besar antara data warehouse dengan big data yang memiliki latar belakang data heterogen. Model integrasi data dimulai dari tingkat dasar sebagai pendekatan sinkronisasi *button up* antar schema local. Langkah tersebut merupakan langkah awal dengan melakukan sinkronisasi antar data heterogen di level fasilitas kesehatan (Rumah Sakit dan Puskesmas). Semua data hasil sinkronisasi dikelola dan disintegrasikan menjadi model hub & spoke data center untuk staging datawarehouse. Data sinkronisasi dan integrasi antar fasilitas kesehatan ini akan selalu berkembang sehingga membentuk *center data warehouse*. Center data warehouse merupakan koleksi data rujukan antar fasilitas kesehatan baik tingkat dasar atau primer sampai dengan tingkat lanjut. Kumpulan data ini perlu dikelola sebaik mungkin sehingga diperlukan teknologi pengelola *big data* dalam penyimpanan dan pengelolaan datanya. Teknologi apache hadoop merupakan *framework big data* yang mampu menyimpan data heterogen tanpa memandang jenis data. Apache hadoop merupakan distribute file atau bisa diartikan system file yang terdistribusi (*HDFS-Hadoop Distributed File System*). Hadoop dilengkapi dengan menggunakan model pemrograman MapReduce dalam menganalisa data dan dapat dimanfaatkan untuk menyimpan dan memaanipulasi data. Dengan begitu apache hadoop cocok untuk mendukung central data warehouse system.



Gambar 1. Area Domain Aplikasi Big Data

2. TINJAUAN PUSTAKA

2.1. Data Kesehatan

Big data dalam data kesehatan mengacu pada kondisi data kesehatan elektronik yang begitu besar dan kompleks sehingga sulit untuk dikelola dengan manajemen tradisional atau perangkat lunak dan perangkat keras yang umum (Priyanka dan Kulennavar, 2014). Beberapa data kesehatan harus terintegrasi dengan data pendukung yang lain dan dicirikan oleh kebutuhan akan ketepatan waktu, misalnya data yang dihasilkan oleh sensor biometrik wearable atau implantable dimana tekanan darah dan denyut jantung serung diperlukan untuk dikumpulkan dan dianalisis secara real time (Helm-Murtagh, 2014). Big data di bidang kesehatan atau medis sejumlah besar informasi tentang riwayat medis pasien, simptomatologi, dianosa dan tanggapan terhadap perawatan dan terapi. Teknik data mining bisa diimplementasikan untuk memperoleh pengetahuan dari data secara berurutan untuk mengidentifikasi pola baru yang menarik dari infeksi, mengontrol data atau memeriksa pelaporan praktek,. Bahkan model prediksi dapat digunakan sebagai alat deteksi mengeksploitasi rekam medis elektronik pasien terakumulasi dalam suatu wilayah tertentu (Belini dkk, 2013).

2.2. Data Warehouse

Data Warehouse (DW) muncul di tahun 80-an sebagai alternatif untuk menyimpan dan mengatur data secara terpadu dan terintegrasi, yang memungkinkan pengguna untuk melakukan analisis statistik dan intelijen bisnis. Istilah Big Data diciptakan pada tahun 1997 oleh Michael Cox dan David Ellsworth (Cox dan Ellsworth, 2015), peneliti NASA yang harus bekerja dengan kumpulan data yang umumnya sangat besar, yang membebani memori utama, cakram lokal dan kapasitas cakram jarak jauh. Mereka menyebut ini masalah Big Data. Meskipun begitu banyak referensi saat ini; Big Data tidak memiliki definisi yang ketat dan disepakati. Biasanya dikaitkan dengan penanganan data masif, diekstraksi dari sumber yang berbeda dan tanpa struktur yang telah ditentukan. Untuk beberapa penulis, Big Data tidak lebih dari kumpulan data yang ukurannya berada di luar alat database yang khas untuk menangkap, menyimpan, mengelola, dan menganalisis. Tidak seperti Data Warehouse, Big Data melampaui konsolidasi informasi karena ia digunakan terutama untuk penyimpanan dan pemrosesan jenis dan volume data apa pun dengan volume yang berpotensi tumbuh secara eksponensial. Namun demikian, apa yang disimpulkan dalam makalah ini adalah bahwa baik Data Warehouse dan Big Data memiliki tujuan akhir yang umum: eksplorasi data dengan tujuan menggambarkan situasi, perilaku, mencari pola, hubungan dan kesimpulan. Hal ini dikarenakan masing-masing merupakan sumber data yang dapat di integrasikan melalui schema maupun struktur datanya.

Data Warehouse (DW) memiliki prinsip integrasi dan konsolidasi informasi dalam struktur multidimensional yang kaku. Salah satu contohnya adalah model *snowflake* (Todman, 2001), digunakan untuk melakukan Online Analytical Processing (OLAP) menerapkan Business Intelligence (BI). Di sisi lain, Big Data tidak memiliki prinsip konsolidasi dan integrasi di bawah struktur yang telah ditentukan, itu lebih tentang penyimpanan dan manajemen volume besar data mentah (tipe, sumber dan kecepatan kedatangan heterogen (Demchenko dan Membrey, 2014), untuk dimana infrastruktur terdistribusi dan satu set perangkat keras dan perangkat lunak khusus diperlukan. Pengolahan dan analisis data menggunakan teknik-teknik canggih ilmu integrasi data, di mana konsolidasi tidak relevan, karena ini tergantung pada sifat data dan masalah khusus.

Teknologi 4GL (bahasa generasi keempat) memfasilitasi pengembangan aplikasi transaksional yang memungkinkan algoritma otomatisasi pada data terstruktur berulang. Data terstruktur (Structure Data) ditandai untuk didefinisikan dengan baik, dapat diprediksi dan ditangani dengan baik oleh infrastruktur yang diuraikan (Inmon, 2014). Perkembangan teknologi, digitalisasi, perangkat hyperconnected, dan jaringan sosial, di antara kemungkinan yang lainnya, membawa informasi yang tidak terstruktur ke lingkup perusahaan. Ini termasuk informasi dalam dokumen digital, data yang berasal dari perangkat otonom (sensor, kamera, scanner, dll.), Dan data semi-terstruktur dari situs web, media sosial, email, dll. Data tidak terstruktur (Unstructure Data) tidak dapat diprediksi dan struktur komputer yang dapat dikenali, dan dapat dibagi menjadi data yang berulang dan tidak berulang (Inmon, 2014). Data repetitif tidak terstruktur (Unstructured Repetitive Data) adalah data yang terjadi dalam banyak kesempatan pada waktunya, mungkin memiliki struktur yang sama, umumnya masif, dan tidak selalu memiliki nilai untuk analisis. Sampel atau bagian dari data ini dapat dimanfaatkan. Karena itu sifat berulang, pemrosesan algoritma rentan terhadap pengulangan dan reutilisasi. Contoh khas dari kategori ini adalah data dari sensor, di mana tujuannya adalah analisis sinyal dan untuk mana algoritma tertentu didefinisikan.

Data unrepeatitive yang tidak terstruktur memiliki struktur data yang bervariasi, yang menyiratkan bahwa algoritma tidak dapat digunakan kembali (dan tugas memprediksi atau menggambarkan strukturnya sudah kompleks). Inmon menempatkan elemen-elemen yang bersifat tekstual (yang memerlukan teknik dari Natural Language Processing dan linguistik komputasional) di dalam kategori ini (Inmon, 2014). Dari sudut pandang kami, selain teks bentuk bebas, citra, video dan audio juga berkaitan dengan kategori ini. DW tradisional lahir dengan tujuan mengintegrasikan data terstruktur yang berasal dari sumber transaksional dan menghitung dengan informasi historis yang didukung oleh analisis berbasis OLAP. Dengan datangnya tipe data baru, beberapa penulis mengusulkan DW untuk menyesuaikan arsitektur dan prosesnya, seperti yang disarankan dalam Inmon dengan DW 2.0 dan Kimball dalam peran yang berkembang dari gudang data perusahaan di era big data analytics (Kimball, 2011).

3. METODE PENELITIAN

Pada penelitian ini digunakan pendekatan metode studi literatur (*library research*) dan studi lapangan (*field research*) untuk mendesain pola sinkronisasi data dalam rangka integrasi data warehouse. Kami menangani masalah persiapan data dengan memungkinkan membuat model untuk secara interaktif dalam menentukan rencana persiapan data yang benar pada sampel data yang kecil dan kemudian melaksanakan rencana tersebut dalam lingkungan eksekusi integrasi streaming yang paralel. Untuk mencapai tujuan ini kami awali dengan membangun serangkaian operasi transformasi data yang komprehensif yang dapat mewakili berbagai macam data dunia nyata, termasuk informasi terstruktur serta data semiterstruktur, dan mengembangkan pendekatan yang mudah digunakan yang memungkinkan untuk dapat dengan cepat menentukan data yang benar dengan membentuk kembali rencana yang tidak hanya mengubah data, tetapi dapat dengan mudah merestrukturisasi output dari satu schema data menjadi input dari schema lain.

Fokus pada pekerjaan sinkronisasi ini adalah membangun pendekatan yang memungkinkan orang lain yang ingin mengintegrasikan data terstruktur atau semi terstruktur dapat menyelesaikan tugas-tugas integrasi mereka sendiri tanpa harus kebingungan mencari model integrasi. Kami membangun ide pokok untuk memudah mendefinisikan tugas integrasi, tetapi juga mengatasi tantangan tentang cara menentukan tugas transformasi data dan melakukan tugas-tugas ini dengan cara terukur.

Adapun tahapan penelaitain dapat dijelaskan dalam langkah-langkah dibawah ini- :

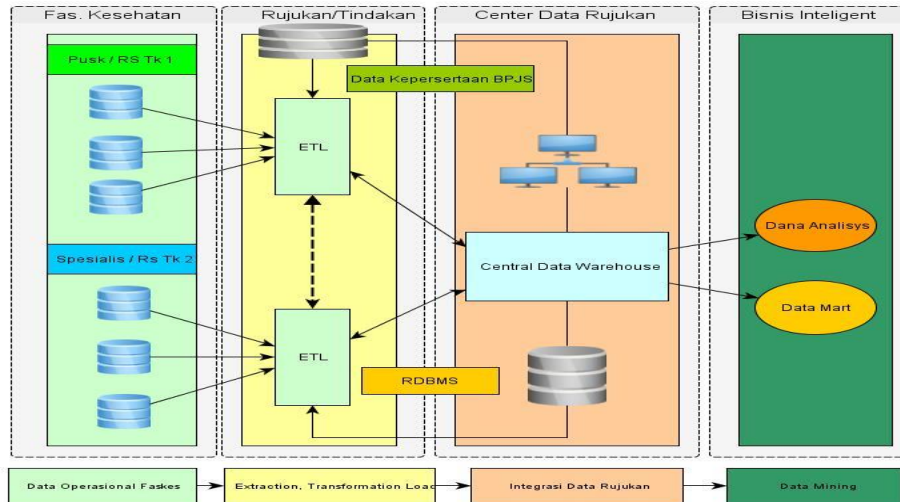
1. Studi Pendahuluan-: studi pendahuluan dilakukan dengan mengumpulkan informasi mengenai pengelolaan sistem surveilans terpadu (*SST*).
2. Observasi dan Studi Pustaka-: melakukan analisis kebutuhan dengan analisa diskriptif melalui kajian pustaka yang terkait dengan konsep sinkronisasi dan integrasi data dan keberadaan data kesehatan yang digunakan untuk surveilans data kesehatan.
3. Perumusan Masalah-: adalah merumuskan permasalahan yang akan dieksplorasi dalam rangka menemukan pola baru dalam sinkronisasi data kesehatan antar unit surveilans dengan database epidemiologi kesehatan.
4. Desain Pola Sinkronisasi dan Integrasi Data-: Pada tahapan ini dilakukan desain pola dan notasi sistem sinkronisasi dan integrasi database sebagai pola integrasi data berdasarkan diagnosis dan identifikasi masalah yang ada.
5. Pengujian-: evaluasi terhadap hasil desain dari pola notasi model relasional sinkronisasi dan integrasi dengan aturan de morgan sehingga nantinya pola sinkronisasi database dapat digunakan secara maksimal dan sesuai dengan pola baru untuk mendukung pengembangan integrasidata epidemilogi kesahatan dari lingkungan data warehous yang heterogen.
6. Hasil Pola Sinkronisasi dan Integrasi-: dalam bentuk notasi model relasional sebagai pola notasi model relasional sinkronisasi dan integrasi data. Hal ini merupakan gambaran model sistem sinkronisasi database besaar (data warehouse) berbasis proses replikasi data dari masing-masing unit surveilans yang heterogen (puskesmas, laboratorium, rumah sakit) menuju database besar (big data) epidemiologi data kesehatan.

Untuk menerapkan sinkronisasi dan integrasi data besar, dimulai dengan simulasi schema data heterogen untuk dimodelkan sebagai sumber data terstruktur, yang memungkinkan kami untuk melakukan transformasi data yang diperlukan dalam sinkronisasi, dan kemudian mengembangkan kemampuan restrukturisasi data baru dan kemudian mengeksekusi restrukturisasi ini pada dataset besar (dalam rangka integrasi). Selanjutnya akan menentukan rencana restrukturisasi (dalam rangka integrasi) pada sebagian kecil data, dan kemudian mengeksekusi dataset tersebut dalam lingkungan terdistribusi di seluruh dataset. Langkah utama dalam mendesain model sinkronisasasi adalah melakukan penterjemahan (*translation*) *schema* dan integrasi (*integration*) *schema*. Langkah utama integrasi adalah *generation*, melibatkan pembuatan *template* untuk setiap relasi yang merupakan sebagai anggota *schema* global. Melakukan analisis proses setiap tlangkah tersebut. Langkah *translation* skema lokal ke skea global dari masing-masing model data lokal dialihkan ke model data umum untuk memfasilitasi prosedur integrasi Secara sederhana bisanya *translation* dapat dilakukan sebagai berikut : hubungan hirarkies ke entitas, hubungan *network* ke entitas, hubungan relasional ke entitas. Langkah *translation* tersebut memiliki subtugas sebagai berikut : pilih model skema global, pilih metode integrasi skema.

4. HASIL PENELITIAN DAN PEMBAHASAN

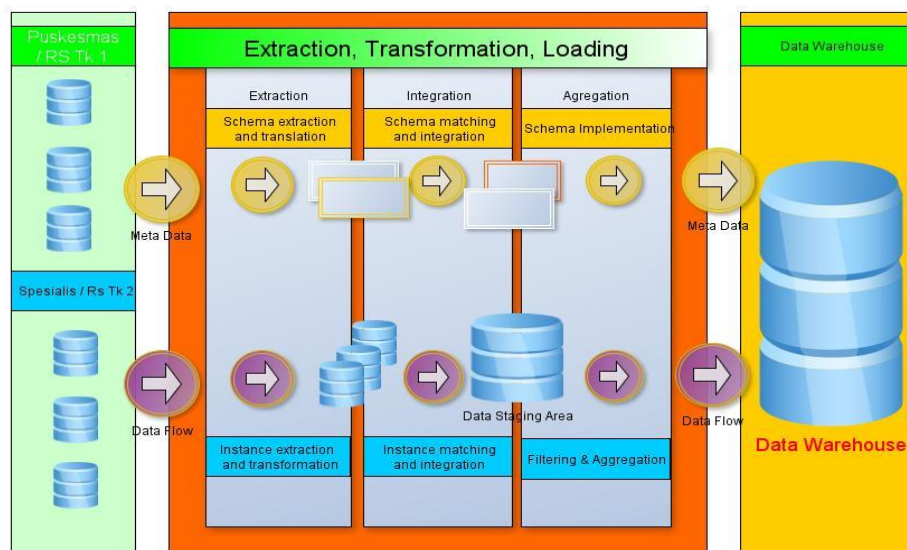
Rancangan central data warehouse (DW) yang merepresentasikan integrasi layanan rujukan antar unit fasilitas kesehatan (puskesmas, laboratorium, rumah sakit) terletak pada area geografis yang tersebar dan data (source) di peroleh dari sistem aplikasi dan dbms yang beragam (heterogen). Untuk itu perlu dirancang arsitektur central data warehouse yang

mengakomodasi sebaran data pelayanan kesehatan yang bersifat local schema dan heterogen yang merupakan sumber data dari central data warehouse atau staging area. Arsitektur dibawah ini merupakan arsitektur yang cocok mengakomodasi lokal scheme yang heterogen dapat terintegrasi pada schema global sebagai target skema dengan melalui proses *extraction, transformation dan loading* (ETL).



Gambar 2. Arsitektur Central Data Warehouse

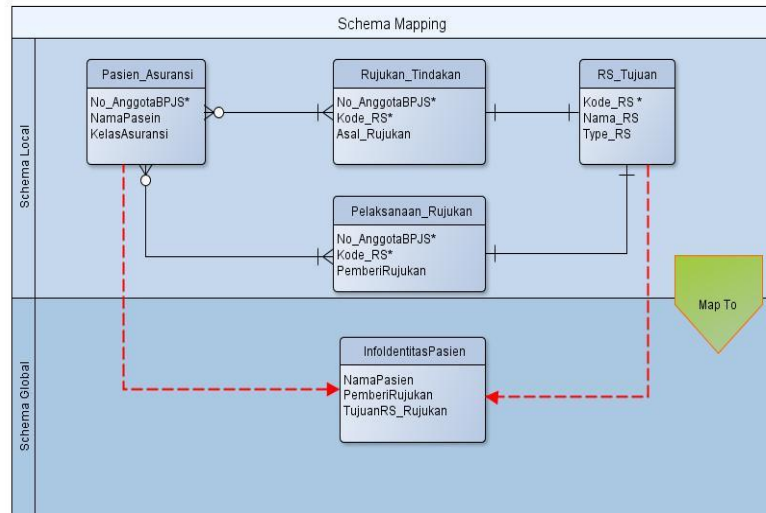
Pada gambar 2 tersebut data epidemiologi bersumber dari fasilitas kesehatan seperti puskesmas dan rumah sakit. Data epidemiologi tersebut berisi catatan pemeriksaan medis pasien termasuk kejadian dan penanganannya setiap terjadi penyakit atau kasus. Data epidemiologi ini disimpan sampai puluhan tahun oleh pihak rumah sakit atau puskesmas. Data ini merupakan big data yang bisa dijadikan data warehouse. Dengan ETL maka berbagai informasi dan pengetahuan baru dapat dihasilkan sebagai data mart yang sangat bermanfaat untuk mengambil kebijakan tertentu dibidang kesehatan. Gambar 3 dibawah ini menjelaskan proses pembentukan data warehouse sebagai big data. Data warehouse biasanya dibangun dari berbagai database yang memiliki sifat heterogen sehingga di perlukan *preprocessing* dan *data cleaning* biar tidak ada konflik antar data heterogen dalam rangka mengintegrasikannya. Salah satu cara untuk menghindari konflik schema dalam integrasi data adalah dengan *mapping* dan *matching schema*.



Gambar 3. Data Staging Process

A. Konstruksi Mapping Schema

Sebagai gambaran pembahasan dalam mapping schema, menggunakan dua buah database (kelompok database skema lokal sebagai sumber skema dan database untuk target dengan sekema global). Kami menganggap database sumber sebagai D_s yang berisi skema lokal (Source Skema - S_s) dengan beberapa relasi (n relasi $R_1 \dots R_n$). Sedangkan target database (Database Targt- D_t) yang berisi skema target (Schema Target- S_t) yang terdiri dari satu relasi target R . Dengan demikian skema mapping (M) adalah project join query yang memetakan S_s ke S_t . Untuk setiap $R_i, i \in [n]^4$, kita menunjukkan skemanya dengan $S(R_i)$ dan contohnya dengan $I(R_i)$. $S(R_i)$ adalah himpunan semua atribut dalam R_i . Demikian pula R (Relasi target) memiliki skema $S(R) = \{A_1, \dots, A_m\}$, dimana m adalah ukuran target dan $A_j (j \in [m])$ mewakili attribute ke- j dalam R . $t[A]$ singkatan dari proyeksi tuple pada attribut A . Jenis pengguna pada spreadsheet masukan dibawah skema target. Setiap sample E adalah atribut string. Kami menunjukkan basris pertama sampel dengan $tE=(E_1, \dots, E_m)$ dan menyebutnnya tuple sample. Tujuan kami untuk pencarian sample adalah untuk menemukan semua pemetaan skema yang mengubah basis data sumber menjadi target “mengandung” sample tuple.



Gambar 4. Pemetaan Skema Lokal Dengan Penggabungan ke Skema Global

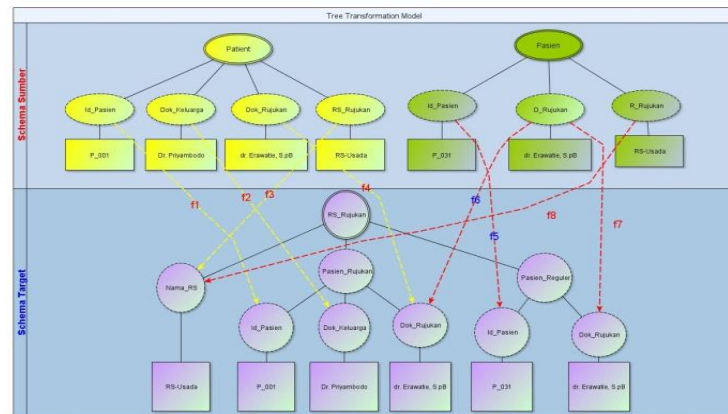
Gambar 4 diatas merupakan proses integrasi data dengan mapping schema, dimana 4 tabel schema local bis di mapping menjadi 1 schema global. Dampak data input pengguna makin tidak sama persis dengan sumbernya, karena itu dapat digunakan untuk menghasilkan pemetaan yang skema, kami memanfaatkan sample yang tidak akurat dengan membiarkannya “noisily contain” oleh beberapa instance database. Secara formal kami mendefinisikan hubungan (“noisily contain”) oleh operator biner \supseteq , yang mengembalikan nilai boolean berdasarkan model kesalahan yang diinginkan. Dengan operator ini kita katakan sampel yang mengandung sample E iff $t[A] \supseteq E$. Demikian pula kita katakan bahwa t mengandung E iff $\exists A$ s.t. $t[A] \supseteq E$. Selanjutnya diberikan $tE=(E_1, \dots, E_m)$, kita sebut T berisi tE , iff $\forall i \in [m], t[A_i] \supseteq E_i$. Akhirnya dapat ditentukan target databasenya D_t berisi tE iff $\exists t \in D_t$ s.t. berisi tE . Dengan konsep ini, kita mendefinisikan pencarian sample sebagai berikut : jika diberikan database sumber D_s dan sample tuple $tE = (E_1, \dots, E_m)$, di cari untuk menemukan semua mapping shema M sedemikian rupa sehingga hasil mapping berasal dari database sumber M (schema mapping) dari database sumber (D_s) mengandung semua tuple database sumber tE , maka setiap hasil pemetaan schema tersebut disebut pemetaan schema yang valid.

B. Proses Query Translation

Bagian ini membahas beberapa pertanyaan (Q), misalkan ada dua pasang database sumber yang berbeda strukturnya sebagai (D_s) yang berisi skema lokal (S_s) dengan beberapa relasi (n relasi $R_1 \dots R_n$), jika ada pertanyaan yang menyangkut semua skema lokal dari pasangan database sumber D_s yang berbeda. Dibutuhkan beberapa proses dari query translation,

diantaranya adalah query translation semantik, forward query translation (translation tree pattern, translation tree pattern join) dan backward query translation. Untuk setiap set kandidat V (View) dari cover yang terpilih, maka perintah umum yang digunakan untuk melakukan seleksi adalah :

Select attributes in V
From source relations in the join path for V
Where filter and join conditions from the join path



Gambar 5. Tree mapping and Translation Correspondences

Perhatikan bahwa prinsip-prinsip ini merupakan penyajian kembali prinsip data umum seperti "satu fakta di satu tempat". Bahkan dengan adanya filter ini, kami mencoba untuk menegakkan prinsip-prinsip ini untuk nilai-nilai yang dipilih oleh filter. Karena tujuan kami adalah pemetaan skema daripada desain skema, kami mengizinkan pengguna untuk mengganti asas-asas ini. Sebagai contoh, dalam mempublikasikan informasi untuk skenario "What-if", pengguna mungkin memilah produk silang sampai dia bisa mengevaluasi semua kemungkinan. Kami menggunakan prinsip-prinsip ini untuk mendorong pemetaan awal, yang menjaga, sejauh mungkin, informasi dalam sumbernya. Pengguna mencek data target saya yang berasal dari pemetaan ini dan memutuskan apakah ke kabupaten atau memodifikasi pemetaan. Sehingga proses tree mapping menjadi rujukan untuk melakuka query dengan atribut dan filter tertentu, jika di terapkan pada gambar tree mapping diatas maka perintah korespondensi tambahannya adalah :

$f1 : pasien(id_pasien) \rightarrow rs_rujukan(id_pasien)$
 $f2 : pasien(dok_keluarga) \rightarrow rs_rujukan(dok_keluarga)$
 $f3 : pasien(dok_rujukan) \rightarrow rs_rujukan(do_rujukan)$
 $f4 : pasien(rs_rujukan) \rightarrow rs_rujukan(rs_rujukan)$
 $f5 : pasien(id_pasien) \rightarrow rs_rujukan(id_pasien)$
 $f6 : pasien(d_rujukan) \rightarrow rs_rujukan(dok_rujukan)$
 $f7 : pasien(dok_rujukan) \rightarrow rs_rujukan(dok_rujukan)$
 $f8 : pasien(r_rujukan) \rightarrow rs_rujukan(nama_rs)$

```
select pt.id_pasien, pt.dok_keluarga, pt.dok_rujukan, ps.id_pasien, ps.nama_rs
from patient pt, pasien ps
where pt.id_pasien=ps.id_pasien
union all
select null as dok_keluarga, ps.id_pasien, ps.dok_rujukan*pt.d_rujukan, null as pasien_reguler
from patient pt, pasien ps
where pt.id_pasien=ps.id_pasien
```

```
create view t_rujukan (id_pasien, dok_keluarga, nama_kasus, nama_rs) as
select f1(s1, id_pasien, dok_keluarga), f2 (s2,dok_rujukan, fast_rujukan, rs_rujukan)
from s1,s2
where s1.id_pasien=s2.id_pasien
union
select f2(s2, id_pasien, null, rs_rujukan)
```

5. KESIMPULAN DAN SARAN

Dengan peningkatan Big Data pada bidang kesehatan maka bisa dimanfaatkan untuk melakukan analisis dan data mining dibidang kesehatan. Big Data dibangun berdasarkan dari data yang diperoleh dari keseluruhan proses diagnosis dan pengobatan setiap kasus. Data besar bersifat analitik dapat digunakan untuk pemodelan prediktif, diskriptif, maupun analisis model lain yang bersifat *knowledge discovery*. Dengan Big data dan model analisis tersebut maka banyak manfaat yang dapat di hasilkan dalam bidang kesehatan seperti, pencegahan penyakit, mengurangi kesalahan medis, menghindari mal praktek, Dibidang obat-obatan big data juga dapat digunakan untuk memprediksi kebutuhan obat, distribusi obat maupun peningkatan perawatan berbasis nilai. Namun demikian big data memiliki tantangan seperti mengkonsolidasikan data, melakukan segmentasi data, penggabungan data heterogen, menstrukturkan data, pengindeksan, penyimpanan dan keamanan data.

6. DAFTAR PUSTAKA

- Bellini, P., M. d. Claudio, P. Nesi and N. Rauch, 2013. Taxonomy and Review of Big Data Solutions Navigation. In: Big Data Computing, Akerkar, R. (Ed.), Chapman and Hall/CRC, pp: 57-101.
- C. Todman, Designing A Data Warehouse: Supporting Customer Relationship Management. 2001. D. L. W.H. Inmon, Data Architecture: A Primer for the Data Scientist: Big Data, Data Warehouse and Data Vault. Amsterdam, Boston: Elsevier, 2014.
- G. N. W.H. Inmon, Derek Strauss, DW 2.0: The Architecture for the Next Generation of Data Warehousing (Morgan Kaufman Series in Data Management Systems) (): : Books. Burlington, USA: Morgan Kaufmann Publishers Inc., 2008.
- M. Cox and D. Ellsworth, "Application-Controlled Demand Paging for Out-of-Core Visualization," 1997. [Online]. Available: <http://www.nas.nasa.gov/assets/pdf/techreports/1997/nas-97-010.pdf>. [Accessed: 09-Juni-2018].
- Helm-Murtagh, S.C., 2014. Use of Big Data by Blue Cross and Blue Shield of North Carolina. NCMJ, 75: 195-197.
- P. Bedi, V. Jindal, and A. Gautam, "Beginning with Big Data Simplified," 2014.
- Priyanka, K. and N. Kulennavar, 2014. A survey on big data analytics in health care. Int. J. Comput. Sci. Inform. Technologies, 5: 5865-5868.
- R. Kimball, M. Ross, W. Thorthwaite, B. Becker, and M. J, The Data Warehouse Lifecycle Toolkit, 2nd Edition. 2008.
- R. Kimball, "The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics," Kimball Gr., 2011.
- T. Maiorescu, "General Information on Business Intelligence," pp. 294-297, 2010.
- Thusoo, A., Shao, Z., Anthony, S., Borthakur, D., Jain, N., Sen Sarma, J., Murthy, R. and Liu, H., 2010, June. Data warehousing and analytics infrastructure at facebook. In Proceedings of the 2010 ACM SIGMOD International Conference on Management of data (pp. 1013-1020). ACM.
- Y. Demchenko, C. De Laat, and P. Membrey, "Defining Architecture Components of the Big Data Ecosystem," Collab. Technol. Syst. (CTS), 2014 Int. Conf., pp. 104-112, 2014.
- Warnars, Spits. "Perbandingan Penggunaan Database OLTP dan Data Ware-house." *Creative Communication and Innovative Technology (CCIT)* 8: 83-100.

