

Proses *Extraction, Transformation, and Loading* Pada Pemodelan *Data Warehouse* PO. Sumber Alam Kutoarjo

Agustinus Fritz Wijaya¹, Antonius Teddy Sugiarto²

*Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jalan Diponegoro No. 52 – 60, Salatiga, Jawa Tengah
agustinus.wijaya@staff.uksw.edu*

*Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana
Jalan Diponegoro No. 52 – 60, Salatiga, Jawa Tengah
a.teddy.sugiarto@gmail.com*

Abstract— *Information technology used by most companies otobus in warehousing systems that manage all the information of spare parts buses. Data from recording activity at the warehouse of spare parts PO. Sumber Alam from year to year will continue to grow. Transaction data must be transferred to a data warehouse using extraction, transformation, and loading data that can process a variety of data into one unified system. The design of the PO. Sumber Alam data warehouse model is using the star schema approach. This approach has one fact table contained in the center of a star with several dimension tables surrounding it. Formed from the fact table data warehouse design in the barn of PO. Sumber Alam which is a transaction request spare parts from the garage to the barn to the needs of the bus fleet improvements.*

Intisari— *Teknologi informasi dimanfaatkan oleh sebagian besar perusahaan otobus dalam sistem pergudangan yang mengatur segala informasi suku cadang bus. Data dari kegiatan pencatatan suku cadang di bagian gudang PO. Sumber Alam dari tahun ke tahun akan terus bertambah. Data-data transaksi tersebut harus dipindahkan ke sebuah data warehouse melalui proses ekstraksi, transformasi, dan load data yang dapat mengolah berbagai data menjadi satu kesatuan sistem. Perancangan model data warehouse pada PO. Sumber Alam ini menggunakan pendekatan star schema. Pendekatan ini memiliki satu tabel fakta yang terdapat di pusat bintang dengan beberapa tabel dimensi yang mengelilinginya. Tabel fakta yang terbentuk dari perancangan data warehouse di bagian gudang PO. Sumber Alam yang merupakan transaksi permintaan suku cadang dari bagian bengkel ke bagian gudang guna kebutuhan perbaikan armada bus.*

Kata Kunci— *Data Warehouse, Ekstraksi, Transformasi, Loading, Star Schema, Multidimensi, Suku Cadang.*

I. PENDAHULUAN

Penggunaan teknologi informasi sangat diperlukan untuk dapat menyelesaikan berbagai macam permasalahan yang dihadapi oleh sebagian besar perusahaan dalam menjalankan proses bisnisnya. Salah satu aktivitas bisnis yang dilakukan di perusahaan otobus adalah sistem pergudangan yang memiliki fungsi yang sangat sentral karena sistem pergudangan mengatur segala informasi yang berhubungan dengan suku cadang bus. PO. Sumber Alam Kutoarjo yang bergerak di bidang perusahaan otobus

tergolong ke dalam perusahaan berskala besar. Perusahaan ini memiliki 300 armada yang terbagi ke dalam 3 jenis kategori angkutan, yaitu angkutan kota, angkutan antar kota, dan angkutan pariwisata. Oleh karena itu, PO. Sumber Alam ingin lebih memberikan pelayanan kepada para penumpangnya yaitu dalam hal kecepatan pelayanan dalam persediaan armada bus. Persediaan armada bus berkaitan erat dengan bagaimana perawatan setiap armada bus secara berkala. Proses perawatan ini juga berhubungan dengan penggunaan suku cadang yang terdapat di bagian gudang. Sehingga, sebuah sistem informasi yang terintegrasi diperlukan untuk mempercepat kegiatan pelayanan bagi para penumpang. Data dari kegiatan pencatatan suku cadang di bagian gudang PO. Sumber Alam dari tahun ke tahun akan terus bertambah dan membuat ukuran penyimpanan data menjadi semakin besar. Data yang diolah tidak hanya data yang berformat *database*, tetapi juga data dalam format lain misalnya data teks. Supaya tidak membebani pemrosesan transaksi yang terjadi di bagian gudang, maka data-data transaksi tersebut harus dipindahkan ke sebuah *data warehouse* yang dapat mengolah berbagai data menjadi satu kesatuan sistem.

Sebuah *data warehouse* dapat memberikan laporan-laporan yang bersifat dinamis karena dapat dilihat dari berbagai dimensi dan memiliki kemampuan untuk diperinci lebih detail maupun diringkas sesuai kebutuhan pengguna. Tanpa menggunakan *data warehouse*, maka laporan-laporan yang akan dihasilkan bersifat statis sesuai dengan output aplikasi berdasarkan permintaan pengguna sistem [1]. Kelebihan lain adalah *data warehouse* juga dapat menyediakan *dashboard* bagi para pimpinan PO. Sumber Alam guna memantau bagaimana arus suku cadang yang terjadi di bagian gudang. Dibutuhkan sebuah sistem otomatisasi yang dapat mencatat jumlah penggunaan suku cadang dan jumlah persediaan yang terdapat di bagian gudang pada PO. Sumber Alam secara cepat dan tepat. Sistem yang dibangun akan menggunakan teknologi *data warehouse* dimana semua data yang berkaitan dengan suku cadang akan saling terintegrasi ke dalam sebuah data

sehingga memudahkan dalam pencarian, pemasukan, pembaruan, analisis dan pelaporan data. Data yang diolah tidak hanya data yang berformat *database*, tetapi juga data dalam format lain misalnya data yang berformat teks maupun yang berformat *spreadsheet*. Untuk dapat mengintegrasikan berbagai format data yang tersebar di setiap sumber data yang ada, maka dibutuhkan proses *extraction, transformation, dan loading* (ETL) pada *data warehouse* yang dibangun. Proses tersebut akan mempercepat proses *query* pada saat data akan digunakan dan akan meningkatkan kinerja dari *data warehouse*. Sehingga, harus dilakukan analisa kinerja dari proses ETL pada pemodelan *data warehouse* yang diterapkan pada PO. Sumber Alam guna membantu perusahaan dalam melakukan manajemen sistem informasi pergudangan.

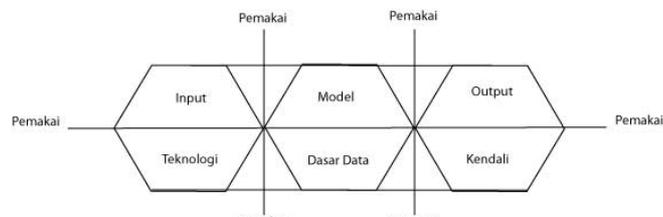
II. TINJAUAN PUSTAKA

Penelitian terdahulu mengenai bagaimana kinerja *extraction, transformation, dan loading* pada *data warehouse* pernah dilakukan dengan judul “*Improve Performace of Extract, Transform, and Load (ETL) in Data Warehouse*”. Penelitian tersebut membahas *extract, transform, dan load* (ETL) adalah inti dari proses integrasi data yang terkait dengan *data warehouse*. ETL mengambil data dari sebuah sumber data yang dipilih, kemudian mengubahnya menjadi format baru sesuai dengan aturan bisnis, dan kemudian menjadikan ke dalam struktur data tujuan. Untuk dapat mengelola aturan dan proses bisnis untuk meningkatkan keseragaman sumber data dan volume data yang besar yang diproses oleh ETL, maka harus diakomodasi dan dikelola kinerja dan biaya bagi pengguna *data warehouse*. ETL merupakan kunci utama dari proses dalam *data warehouse* untuk membawa semua data bersama-sama dalam lingkungan yang homogen. Fungsi ETL akan membentuk kembali data yang relevan dari sumber data ke dalam informasi yang dapat digunakan di dalam *data warehouse*. Tanpa fungsi ini, maka tidak akan diperoleh informasi yang strategis. Jika proses ETL ini tidak dilakukan, maka tidak akan ada integrasi data dan permintaan data yang dapat diperoleh dari *data warehouse*. Penelitian tersebut membahas bagaimana meningkatkan kecepatan proses ETL data dalam *data warehouse* dengan dukungan dari *cache query* yang dapat mengurangi waktu respon dan meningkatkan kinerja *data warehouse* [2].

Penelitian lain yang berjudul “*Data Warehouse Approach to Build a Decision-Support Platform for Orthopedics Based on Clinical and Academic Requirements*”. Dalam penelitian ini dilakukan studi kasus pada dunia medis yaitu bagaimana membangun sebuah sistem pendukung keputusan yang menggunakan *database* khusus dalam hal ini *data warehouse* guna meningkatkan mutu pelayanan kepada pasien secara berkelanjutan. *Data warehouse* dianggap menawarkan fungsi yang lebih canggih daripada menggunakan *database* konvensional untuk sistem pendukung keputusan. Namun, menerapkan *data warehouse* mungkin menghadapi banyak hambatan, termasuk biaya mahal dan tenaga yang besar. Pada penelitian sebelumnya dalam penelitian ini, peneliti

telah membangun sebuah *database* ortopedi, yang berfungsi untuk mengumpulkan data pasien sejak tahun 2002. Sedangkan sistem baru yang dikembangkan ini dibangun menggunakan *database* khusus yaitu *data warehouse*. Fungsi utama adalah untuk menghasilkan informasi yang digunakan untuk mendukung proses pengambilan keputusan dalam aspek akademis dan klinis. Para pengguna sistem ini akan menganggap sistem yang jauh lebih efektif daripada *database* konvensional yang hanya menyimpan data pasien saja. Penelitian ini dapat berguna sebagai salah satu pilihan untuk meng-*upgrade* sistem dari *database* konvensional menjadi *database* khusus dan sebuah langkah maju untuk tujuan dari peningkatan kualitas kepada pasien secara berkelanjutan dimana data yang diperoleh dapat dengan mudah dianalisis sehingga pengguna dari sistem ini dapat segera mengambil keputusan terhadap masalah pasien ortopedi [3].

Sistem informasi adalah suatu sistem di dalam suatu organisasi yang mempertemukan kebutuhan pengolahan transaksi harian, mendukung operasi, bersifat manajerial dan kegiatan strategi dari suatu organisasi yang menyediakan pihak luar tertentu dengan laporan – laporan yang diperlukan. Sistem informasi terdiri dari komponen – komponen yang disebut dengan istilah blok bangunan (*building block*), yaitu blok masukan (*input block*), blok model (*model block*), blok luaran (*output block*), blok teknologi (*technology block*), blok basis data (*database block*), dan blok kendali (*controls block*). Sebagai suatu sistem keenam blok tersebut masing – masing saling berinteraksi satu dengan lainnya membentuk satu kesatuan untuk mencapai sasarannya [4]. Blok sistem informasi yang saling berinteraksi dapat digambarkan pada Gambar 1.



Gambar 1. Blok Sistem Informasi [4]

Sarana pergudangan dapat dijelaskan sebagai cara untuk mendapatkan manfaat ekonomis dengan cara melaksanakan cara produksi paling ekonomis di sumber suplai dan memenuhi kebutuhan pemakai akhir [5]. Secara garis besar, tujuan dari suatu fasilitas pergudangan dapat diuraikan sebagai berikut: (1) Membantu pengadaan material yang dibutuhkan, menurut spesifikasi yang benar, dalam jumlah yang paling ekonomis dan dengan harga yang paling rendah. (2) Memudahkan dan mempercepat pencapaian ke daerah gudang dan materi yang disimpan di dalamnya, untuk menyelesaikan proses-proses penerimaan, distribusi dan pengelolaan inventaris secara tepat. (3) Memelihara kondisi material simpanan agar tetap dalam kondisi baik sampai material tersebut diperlukan, dengan mencegah kerusakan atau penurunan kualitas akibat penanganan fisik atau kondisi

lingkungan.

Suatu perusahaan biasanya menggunakan *database* untuk pemrosesan transaksi pada suatu perusahaan. Pemrosesan transaksi tersebut dikenal dengan OLTP (*OnLine Transaction Processing*). Selain *database* tersebut, ada juga *database* yang digunakan untuk mendukung proses pengambilan keputusan (*decision making*), yaitu *data warehouse*. *Data warehouse* atau gudang data adalah *database* yang menyimpan data saat ini dan data pada masa lalu yang berasal dari berbagai sistem operasional dan sumber lain. Yang perlu dipahami adalah *data warehouse* tidak digunakan untuk pemrosesan transaksi [6].

Data warehouse sebenarnya adalah suatu basis data. *Data warehouse* dibedakan dengan basis data operasional (basis data yang digunakan pada sistem operasional) antara lain berdasarkan rentang waktu data yang ditanganinya. Rentang waktu basis data operasional adalah dari sekarang sampai dengan 60-90 hari, sedangkan rentang waktu *data warehouse* yaitu 5 sampai 10 tahun. Dengan kata lain, *data warehouse* menampung data sejarah yang cukup panjang sedangkan basis data operasional hanya memiliki sejarah yang pendek. *Data warehouse* adalah koleksi data yang bersifat *subject-oriented*, terintegrasi, *time-variant*, dan *non-volatile* yang digunakan untuk mendukung proses pengambilan keputusan yang strategis untuk perusahaan. *Data warehouse* merupakan salah satu konsep penyediaan solusi ke organisasi, dimana memiliki database yang distrukturkan secara khusus untuk dilakukan proses *query* dan analisis [7].

TABEL I
AKTIVITAS PROSES SUMBER DATA DAN DATA WAREHOUSE [8]

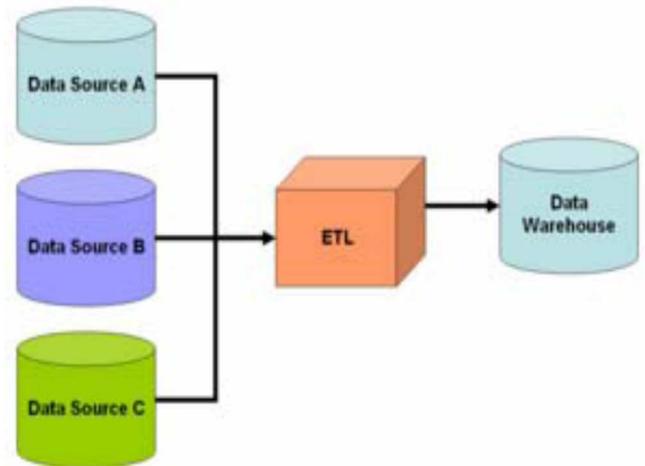
Proses	Keterangan
Pengambilan (<i>Extraction</i>)	Memperoleh data dari basis data operasional dan sumber eksternal
Pembersihan (<i>Cleaning</i>)	Memperkecil kesalahan-kesalahan dan sedapat mungkin mengisi informasi yang hilang
Transformasi (<i>Transformation</i>)	Menyamakan perbedaan-perbedaan makna (semantik) maupun kaidah (sintaks). Hal ini diperlukan mengingat bahwa beberapa DBMS bisa saja menggunakan nama yang berbeda untuk menyatakan data yang sama.
Pemuatan (<i>Loading</i>)	Melakukan pembuatan pandangan (<i>view</i>) pengurutan, dan pembangkitan informasi ringkasan. Untuk mengefisienkan pengaksesan data, data juga diindeks.
Penyegaran (<i>Refreshing</i>)	Melaksanakan pemutakhiran data berdasarkan data pada sumber data. Hal ini dilakukan karena data warehouse dan sumber data tidak terhubung secara online.

Sumber data bagi *data warehouse* adalah data internal yang terdapat pada perusahaan (basis data operasional) dan data eksternal yang berasal dari sumber-sumber di luar perusahaan. Oleh karena itu tidak mengherankan jika ukuran *data warehouse* jauh lebih besar. Ukurannya dapat berkisar antara *gigabyte* dan *terabyte*. Data yang ada pada *data warehouse* dapat diakses dan dianalisis melalui perangkat seperti OLAP, *data mining*, paket visualisasi informasi, paket statistik, dan pembangkit laporan. Proses yang terjadi antara sumber data dan *data warehouse* disebut proses ETL

(*Extraction, Transformation, Loading*) proses ETL tersebut dapat berupa pengambilan, pembersihan, transformasi, pemuatan, dan penyegaran [8]. Aktivitas masing-masing proses dapat dilihat pada Tabel I.

Dalam pengelolaan *data warehouse* sering ditemui istilah *data mart*, yaitu bagian dari *data warehouse* yang berisi rangkaian kecil *data warehouse* yang mendukung kebutuhan pada tingkat tertentu dari suatu perusahaan seperti departemen. Adapun yang dimaksud dengan *data mart* adalah *data warehouse* yang mendukung kebutuhan pada tingkat departemen atau fungsi bisnis tertentu dalam perusahaan [9]. Karakteristik yang membedakan *data mart* dan *data warehouse* adalah sebagai berikut:

- *Data mart* memfokuskan hanya pada kebutuhan-kebutuhan pemakai terkait dalam sebuah departemen atau fungsi bisnis.
- *Data mart* biasanya tidak mengandung data operasional yang rinci seperti pada *data warehouse*.
- *Data mart* hanya mengandung sedikit informasi dibandingkan dengan *data warehouse*. *Data mart* lebih mudah dipahami dan dinavigasi.



Gambar 2. Arsitektur Data Warehouse

Dalam pembuatan *data warehouse*, dilakukan beberapa langkah yang ada, antara lain [10]: (1) *Data Extraction*, fungsi ini biasanya berhadapan dengan bermacam *data source*, dan menggunakan teknik yang sesuai dengan setiap *data source*. Sumber data mungkin berasal dari *source machine* yang berbeda dalam format data yang berbeda pula. (2) *Data Transformation*, fungsi ini melibatkan berbagai bentuk dalam mengkombinasikan bagian dari data yang berasal dari sumber yang berbeda. Kombinasi data dilakukan dari sumber *record* tunggal, atau dapat juga dilakukan dari elemen data yang berelasi dengan banyak sumber *record*. Proses *cleaning* mungkin dilakukan dalam *data transformation*, dimana proses *cleaning* memiliki fungsi untuk melakukan koreksi terhadap kesalahan pengejaan, atau untuk melakukan eliminasi terhadap duplikat data. (3) *Data Loading*, setelah selesai melakukan desain dan konstruksi dari *data warehouse* dan aplikasi digunakan untuk pertama kalinya, akan dilakukan pengisian awal data ke dalam media penyimpanan data warehouse. Dalam pengisian awal,

dilakukan pemindahan data dalam jumlah yang besar. Gambar 2 menjelaskan arsitektur sederhana *Data Warehouse*.

Proses perancangan arsitektur *data warehouse* adalah proses yang sangat rumit, untuk memulainya harus dimulai dengan mendefinisikan kebutuhan dari pengguna dan menentukan data apa saja yang dibutuhkan. Perancangan arsitektur *data warehouse* dapat dibagi ke dalam dua bagian yaitu arsitektur *logical* dan arsitektur fisik. Arsitektur *logical* adalah rancangan tahapan alur data dari sumber data sampai ke *data warehouse*, sedangkan arsitektur fisik adalah gambaran teknis dari konfigurasi yang akan diterapkan pada *data warehouse*. Perancangan *data warehouse* konseptual adalah sebuah proses yang disajikan dalam bentuk model multidimensi. Salah satu pendekatan yang diambil oleh komunitas riset basis data tersebut adalah pendekatan yang berdasarkan model ER (*Entity Relationship*) yang diperluas atau ditransformasi ke dalam model multidimensi. Beberapa penelitian telah dilaksanakan untuk pengembangan metodologi perancangan model *data warehouse* konseptual menggunakan model ER. Model ER ini saat ini disebut dengan pendekatan skema bintang atau *star schema* yang mengadopsi model relasi setiap entitas di dalam perancangan model *data warehouse*. *Star schema* merupakan paradigma *modeling* yang paling banyak digunakan dalam perancangan *Data Warehouse* dimana di dalamnya mengandung antara lain sebuah tabel pusat yang besar tanpa adanya data *redundancy* di dalamnya, yang biasa disebut dengan tabel fakta. Selain itu, di dalam *star schema* juga mengandung satu set tabel yang lebih kecil, yang biasa disebut dengan tabel dimensi. Berikut adalah beberapa karakteristik utama dari *star schema* antara lain:

- Pusat dari *star schema* adalah tabel fakta.
- Tabel fakta berisi indikator-indikator kinerja pokok.
- Obyek-obyek informasi dan waktu adalah kunci utama tabel fakta.
- Tabel-tabel yang ada di sekeliling tabel fakta adalah tabel dimensi.
- Tabel dimensi berisi data mengenai obyek-obyek informasi atau waktu.
- Tabel fakta dan tabel dimensi direlasikan dengan *key* yang ada.
- *Star schema* diimplementasikan menggunakan teknologi *relational database*.

Snowflake schema atau skema bola salju merupakan perluasan dari skema bintang dengan tambahan berupa beberapa tabel dimensi yang tidak terhubung secara langsung ke tabel fakta, melainkan melalui tabel dimensi lainnya.

OLAP (*OnLine Analytical Processing*) adalah jenis perangkat lunak yang digunakan untuk melakukan permintaan terhadap data dalam bentuk yang kompleks dan bersifat sementara serta sewaktu-waktu. OLAP memanipulasi dan menganalisis data bervolume besar dari berbagai perspektif (multidimensi). Oleh karena itu OLAP seringkali disebut analisis data multidimensi. Tujuan OLAP adalah menggunakan informasi dalam sebuah basis data

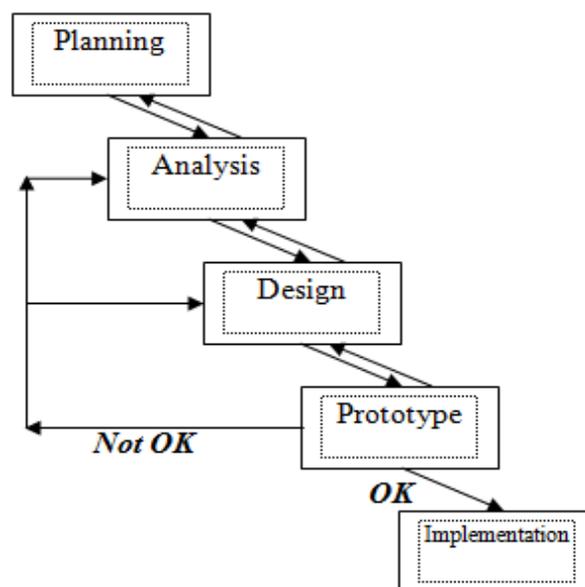
(*data warehouse*) untuk memandu keputusan-keputusan yang strategis. OLAP dapat digunakan untuk melakukan konsolidasi, *drill-down*, dan *slicing and dicing*.

- Konsolidasi melibatkan pengelompokkan data. Sebagai contoh kantor-kantor cabang dapat dikelompokkan menurut kota atau bahkan propinsi. Kadangkala istilah *rollup* digunakan untuk menyatakan konsolidasi.
- *Drill-down* adalah suatu bentuk yang merupakan kebalikan dari konsolidasi, yang memungkinkan data yang ringkas dijabarkan menjadi data yang lebih detail.
- *Slicing and dicing* atau *pivoting* menjabarkan pada kemampuan untuk melihat data dari berbagai sudut pandang.

Sistem OLAP pada masa awal menggunakan larik multidimensi di dalam memori untuk menyimpan data kubus. Sistem seperti ini disebut MOLAP (*Multidimensional OLAP*). Pada perkembangan selanjutnya, data disimpan dalam bentuk basis data relasional. Sistem OLAP seperti ini dikenal dengan sebutan ROLAP (*Relational OLAP*), selain MOLAP dan ROLAP, terdapat pula sistem yang dinamakan *Hybrid OLAP* (HOLAP), yaitu sistem OLAP yang menyimpan beberapa ringkasan dalam memori dan menyimpan basis data dan ringkasan-ringkasan yang lain dalam basis data relasional.

III. METODOLOGI PENELITIAN

Metode penelitian yang digunakan oleh penulis dalam merancang *data warehouse* untuk PO. Sumber Alam, menggunakan metode pengembangan *Decision Support Systems* (DSS). Adapun siklus dari metode pengembangan DSS tersebut seperti terlihat pada Gambar 3.

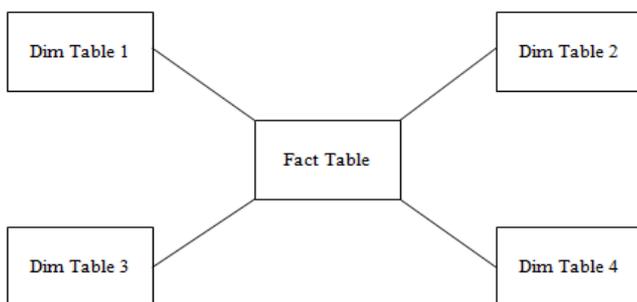


Gambar 3. Metode Pengembangan DSS

Metode pengembangan DSS hampir sama dengan metode pengembangan perangkat lunak (*Software Development Life Cycle*) pada umumnya seperti yang terlihat pada Gambar 3. Perbedaannya adalah pada DSS lebih menekankan pada tahap *Prototyping*. *Prototyping* ditekankan karena dalam

pengembangan DSS, interaksi antara pengembang sistem dengan pengguna sangat intensif sehingga diperlukan suatu pendekatan yang dapat mengkomunikasikan dengan baik hasil yang dibuat oleh pengembang dengan kebutuhan yang diperlukan oleh pengguna. Adapun penjelasan dari tahapan dalam siklus pengembangan DSS yang dilakukan oleh penulis akan dijabarkan di bawah ini.

Perancangan model *data warehouse* pada PO. Sumber Alam ini menggunakan pendekatan *star schema* atau skema bintang. Pendekatan ini memiliki satu tabel fakta (*fact table*) yang terdapat di pusat bintang dengan beberapa tabel dimensi (*dimensional tables*) yang mengelilinginya. Semua tabel dimensi tersebut secara langsung terhubung ke tabel fakta. Gambar 4 akan menjelaskan bagaimana bentuk keterhubungan antara tabel fakta dengan beberapa tabel dimensi.



Gambar 4. Pendekatan *Star Schema* Perancangan *Data Warehouse* PO. Sumber Alam

Tabel fakta berisi *field* yang menunjukkan fakta dari suatu subjek, yang diidentifikasi bahwa dengan *field* ini kita bisa mengukur suatu aktivitas dalam subjek tertentu. Dalam kasus PO. Sumber Alam, maka subjeknya adalah pergudangan. *Field* dari tabel fakta ini bisa berisi total suku cadang yang dikeluarkan, total biaya perbaikan setiap armada, dan lain-lain. Misalnya dalam tabel fakta menyebutkan suku cadang yang dikeluarkan, maka untuk melihat detail keterangan mengenai suku cadang tersebut dapat melihat dalam tabel dimensi suku cadang yang berhubungan dengan tabel fakta tersebut. Penggunaan *star schema* dipilih karena proses *query* yang terjadi lebih ringan dibandingkan dengan pendekatan *snowflake* schema. Tabel dimensi yang akan dibuat dalam perancangan model data warehouse ini, antara lain adalah:

- Tabel dimensi waktu;
- Tabel dimensi suku cadang;
- Tabel dimensi perbaikan; dan
- Tabel dimensi permintaan perbaikan.

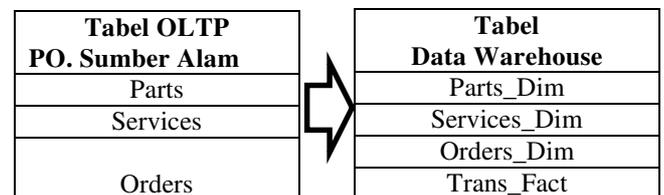
Sedangkan tabel fakta yang terbentuk dari perancangan model *data warehouse* menggunakan pendekatan *star schema* ini adalah tabel yang berhubungan dengan transaksi perbaikan yang menggunakan suku cadang dari mulai permintaan di bagian gudang sampai digunakan oleh bagian bengkel guna perbaikan armada PO. Sumber Alam.

IV. HASIL DAN ANALISIS

Analisis terhadap kinerja proses *extraction*, *transformation*, dan *loading* pada perancangan model *data warehouse* pada Bagian Gudang di PO. Sumber Alam diawali dengan melakukan analisis kebutuhan data guna merancang sistem *data warehouse* yang akan dibangun. Kebutuhan data tersebut antara lain adalah kebutuhan tabel *on-line transaction* (OLTP) dan tabel *data warehouse*. *Data warehouse* yang dibangun di Bagian Gudang PO. Sumber Alam dirancang dengan menggunakan pendekatan *Star Schema*. Tabel-tabel yang dirancang dalam *data warehouse* yang dapat memenuhi untuk kebutuhan data pada Bagian Gudang PO. Sumber Alam, dapat disusun antara lain:

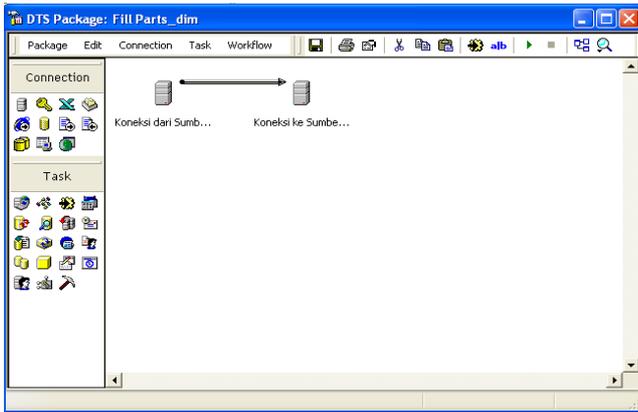
- **Parts_Dim**, berisi detail informasi mengenai suku cadang yang dipakai oleh setiap armada.
- **Services_Dim**, berisi detail informasi mengenai jenis perbaikan setiap armada.
- **Orders_Dim**, berisi detail informasi mengenai jumlah dan jenis permintaan suku cadang yang akan digunakan untuk perbaikan setiap armada.
- **Time_Dim**, berisi detail informasi mengenai waktu yang dipilih di dalam setiap transaksi yang terjadi di Bagian Gudang.
- **Trans_Fact**, berisi detail informasi mengenai nilai dan jumlah transaksi di Bagian Gudang.

Perancangan *data warehouse* ini terlebih dahulu melakukan pemetaan tabel *on-line transaction* dengan tabel *data warehouse* yang dibangun. Pemetaan tersebut dapat dilihat pada Gambar 5 berikut ini.



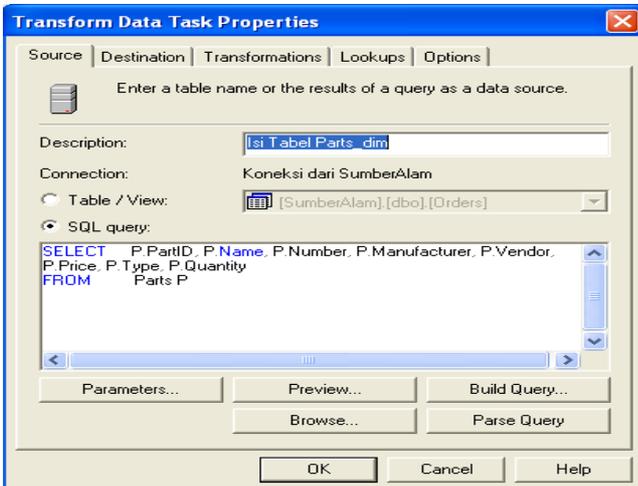
Gambar 5. Pemetaan Pengambilan *Field*

Pemetaan pengambilan *field* pada Gambar 5 akan diimplementasikan menggunakan pernyataan sql (*sql statement*) untuk pembangunan tabelnya. Tabel *data warehouse* yang telah dibuat, kemudian akan diisi dengan melakukan *data transformation*. Adapun tabel *data warehouse* yang diisi akan dibagi menjadi tiga (3) jenis, yaitu pengisian tabel dimensi non-waktu, tabel waktu, dan tabel fakta. Pembuatan *data transformation* dimulai dengan membuat paket baru di basis data PO. Sumber Alam. Pembuatan *data transformation* ini menggunakan bantuan *tools* Microsoft SQL Server 2000. Kemudian setelah membuat paket *data transformation*, pada panel di bagian kiri *Connection*, dipilih *server* yang akan digunakan kemudian dibawa ke panel di bagian kanan. Proses ini terlihat seperti pada Gambar 6.



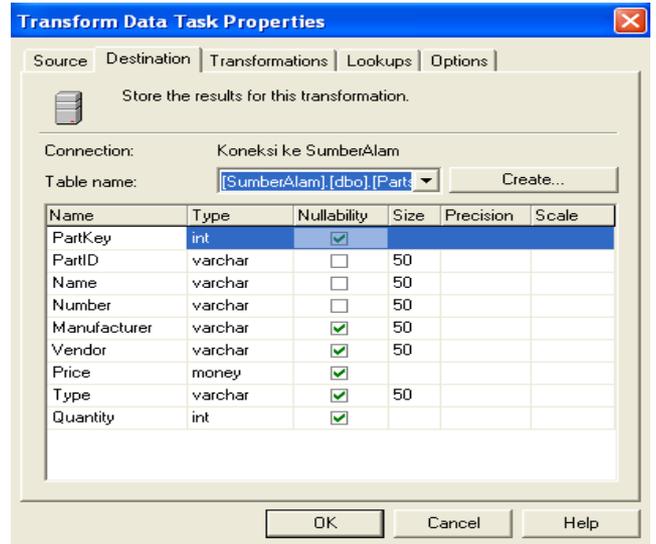
Gambar 6. Data Transformation Services

Kemudian pada proses transformasi tersebut akan diisi informasi yang dibutuhkan. Informasi yang diisi antara lain adalah koneksi dari Sumber Alam yang menunjukkan koneksi asal, nama *Server* di mana basis data Sumber Alam berada, dan nama basis data. Langkah selanjutnya adalah menambahkan proses transformasinya, yaitu dari panel bagian kiri dipilih yang menunjukkan *Transformation Data Task*, dan dipindahkan ke panel di bagian kanan, kemudian ditunjukkan mana yang merupakan *Select Source Connection* yang dipilih koneksi dari *Server* Sumber Alam dan *Select Destination Connection* dipilih koneksi yang berasal dari Sumber Alam, dan hasilnya adalah koneksi dari Sumber Alam yang mengarah ke koneksi ke Sumber Alam. Selanjutnya adalah diatur transformasi apa yang akan berlaku di *Transformation Data Task* ini. Pertama kali adalah mengisi informasi pada *Tab Source* seperti terlihat pada Gambar 7 berikut ini.



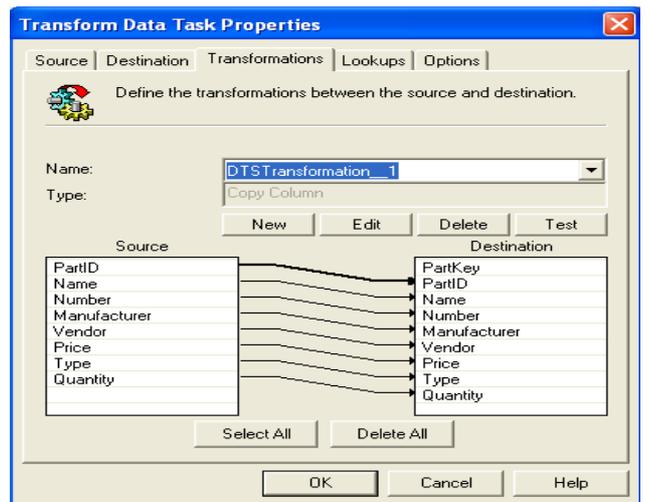
Gambar 7. Informasi Tab Source Transformation Data Task

Berikutnya adalah akan diisi koneksi untuk tujuan transformasi data yang diambil dari tabel asal Sumber Alam. Penyimpanan informasi ini dilakukan pada *Tab Destination* seperti pada Gambar 8.



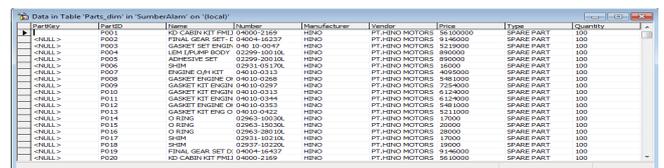
Gambar 8. Informasi Tab Destination Data Transformation Task

Setelah informasi-informasi diisi ke dalam masing-masing *Tab Source* dan *Tab Destination*, maka kemudian akan dilihat bagaimana pemetaan transformasi data yang berasal dari tabel asal menuju ke tabel *data warehouse*. Pemetaan tersebut terkait dengan *field-field* (kolom-kolom) yang terdapat di masing-masing tabel yang terlihat pada Gambar 9.



Gambar 9. Pemetaan Field pada Tab Transformations Data Transformation Task

Setelah proses transformasi selesai dilakukan, kemudian hasil dari transformasi adalah pemuatan data (*loading*) dengan melakukan eksekusi paket transformasi tersebut. Hasil eksekusi paket transformasi atau pemuatan data (*loading*) suku cadang yaitu seperti pada Gambar 10.



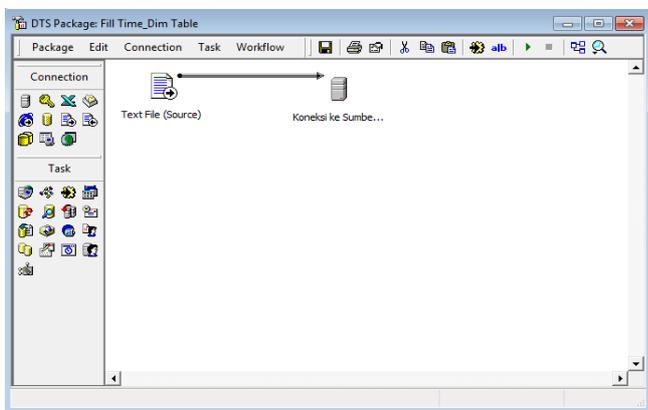
Gambar 10. Hasil Loading Tabel Parts_Dim

Pengisian tabel dimensi waktu dilakukan dengan terlebih dahulu membuat sumber data yang berasal dari *file excel* dalam format (.csv). Adapun pembuatan *file excel* tersebut seperti pada Gambar 11 di bawah ini.

A	B	C	D	E	F	G	H	I	J	K
TheDate	DayOfWei	Month	Year	Quarter	DayOfYea	Holiday	WeekEnd	YearMont	WeekOfYear	
01-Jan-11	1	Jan	2011	Q1						
02-Jan-11	2	Feb	2011	Q1						
03-Jan-11	3	Mar	2011	Q1						
04-Jan-11	4	Apr	2011	Q1						
05-Jan-11	5	May	2011	Q1						
06-Jan-11	6	Jun	2011	Q1						
07-Jan-11	7	Jul	2011	Q1						
08-Jan-11	8	Aug	2011	Q1						
09-Jan-11	9	Sep	2011	Q1						
10-Jan-11	10	Okt	2011	Q1						
11-Jan-11	11	Nov	2011	Q1						
12-Jan-11	12	Des	2011	Q1						

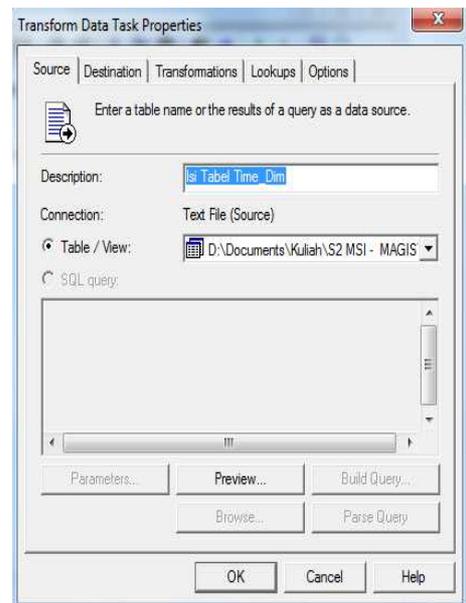
Gambar 11. Data Source Dimensi Waktu dalam Format Excel

Kemudian selanjutnya adalah membuat paket transformasi antara sumber data dengan tabel *data warehouse*. Paket transformasi ini akan menghubungkan sumber data dari *file excel* ke dalam tabel di *data warehouse*. Proses transformasi tersebut seperti pada Gambar 12 di bawah ini.



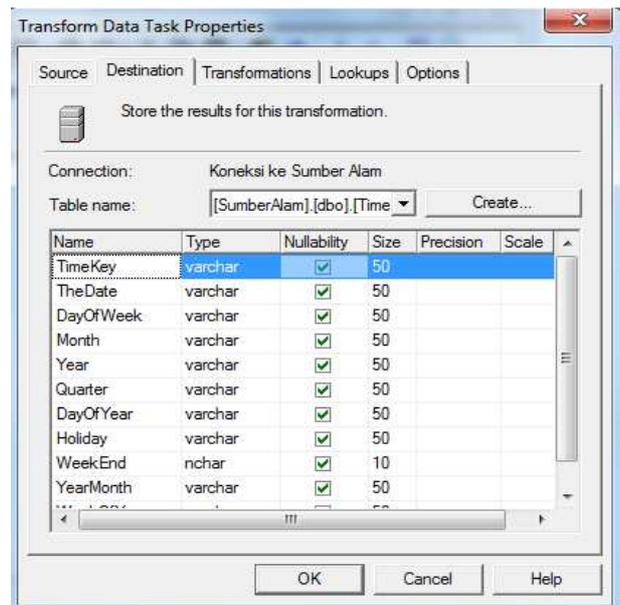
Gambar 12. Data Transformation Services

Kemudian pada proses transformasi tersebut akan diisi informasi yang dibutuhkan. Informasi yang diisi antara lain adalah koneksi dari Sumber Alam yang menunjukkan koneksi asal, nama *Server* di mana basis data Sumber Alam berada, dan nama basis data. Langkah selanjutnya adalah menambahkan proses transformasinya, yaitu dari panel bagian kiri dipilih yang menunjukkan *Transformation Data Task*, dan dipindahkan ke panel di bagian kanan, kemudian ditunjukkan mana yang merupakan *Select Source Connection* yang dipilih koneksi dari *Server* Sumber Alam dan *Select Destination Connection* dipilih koneksi yang berasal dari Sumber Alam, dan hasilnya adalah koneksi dari Sumber Alam yang mengarah ke koneksi ke Sumber Alam. Selanjutnya adalah diatur transformasi apa yang akan berlaku di *Transformation Data Task* ini. Pertama kali adalah mengisi informasi pada *Tab Source* seperti terlihat pada Gambar 13 berikut ini.



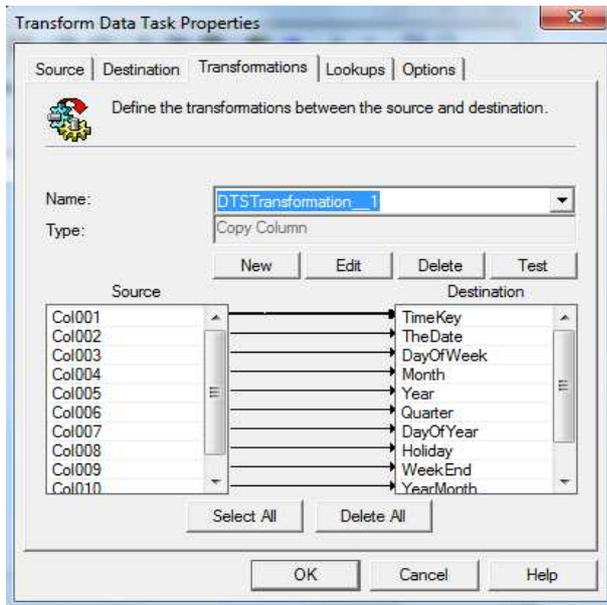
Gambar 13. Informasi Tab Source Transformation Data Task

Berikutnya adalah akan diisi koneksi untuk tujuan transformasi data yang diambil dari tabel asal Sumber Alam. Penyimpanan informasi ini dilakukan pada *Tab Destination* seperti pada Gambar 14.



Gambar 14. Informasi Tab Destination Data Transformation Task

Setelah informasi-informasi diisi ke dalam masing-masing *Tab Source* dan *Tab Destination*, maka kemudian akan dilihat bagaimana pemetaan transformasi data yang berasal dari tabel asal menuju ke tabel *data warehouse*. Pemetaan tersebut terkait dengan *field-field* (kolom-kolom) yang terdapat di masing-masing tabel yang terlihat pada Gambar 15.



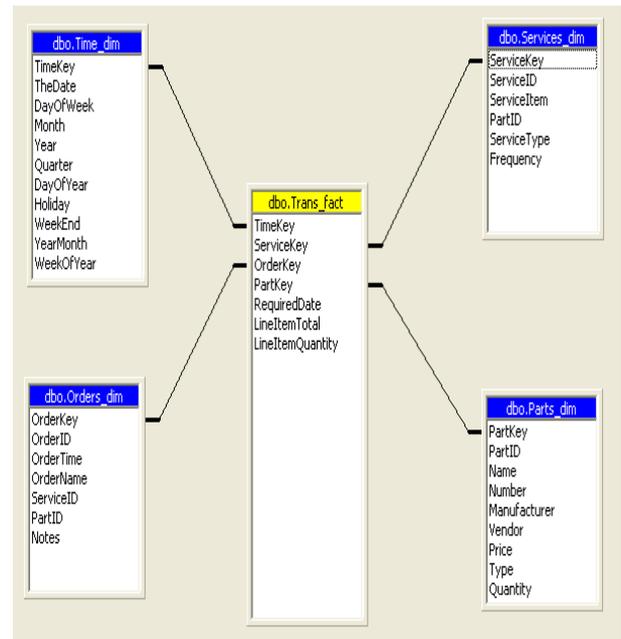
Gambar 15. Pemetaan Field pada Tab Transformations Data Transformation Task

Setelah proses transformasi selesai dilakukan, kemudian hasil dari transformasi adalah pemuatan data (*loading*) dengan melakukan eksekusi paket transformasi tersebut. Hasil eksekusi paket transformasi atau pemuatan data (*loading*) dimensi waktu yaitu seperti pada Gambar 16.

TimeKey	TheDate	DayOfWeek	Month	Year	Quarter	DayOfYear	Holiday	WeekEnd	YearMonth	YearMonth
2011-01-01	2011-01-01	Mon	Jan	2011	Q1	1			2011-01	2011-01
2011-01-02	2011-01-02	Tue	Jan	2011	Q1	2			2011-01	2011-01
2011-01-03	2011-01-03	Wed	Jan	2011	Q1	3			2011-01	2011-01
2011-01-04	2011-01-04	Thu	Jan	2011	Q1	4			2011-01	2011-01
2011-01-05	2011-01-05	Fri	Jan	2011	Q1	5			2011-01	2011-01
2011-01-06	2011-01-06	Sat	Jan	2011	Q1	6			2011-01	2011-01
2011-01-07	2011-01-07	Sun	Jan	2011	Q1	7			2011-01	2011-01
2011-01-08	2011-01-08	Mon	Jan	2011	Q1	8			2011-01	2011-01
2011-01-09	2011-01-09	Tue	Jan	2011	Q1	9			2011-01	2011-01
2011-01-10	2011-01-10	Wed	Jan	2011	Q1	10			2011-01	2011-01
2011-01-11	2011-01-11	Thu	Jan	2011	Q1	11			2011-01	2011-01
2011-01-12	2011-01-12	Fri	Jan	2011	Q1	12			2011-01	2011-01
2011-01-13	2011-01-13	Sat	Jan	2011	Q1	13			2011-01	2011-01
2011-01-14	2011-01-14	Sun	Jan	2011	Q1	14			2011-01	2011-01
2011-01-15	2011-01-15	Mon	Feb	2011	Q1	15			2011-02	2011-02
2011-01-16	2011-01-16	Tue	Feb	2011	Q1	16			2011-02	2011-02
2011-01-17	2011-01-17	Wed	Feb	2011	Q1	17			2011-02	2011-02
2011-01-18	2011-01-18	Thu	Feb	2011	Q1	18			2011-02	2011-02
2011-01-19	2011-01-19	Fri	Feb	2011	Q1	19			2011-02	2011-02
2011-01-20	2011-01-20	Sat	Feb	2011	Q1	20			2011-02	2011-02
2011-01-21	2011-01-21	Sun	Feb	2011	Q1	21			2011-02	2011-02
2011-01-22	2011-01-22	Mon	Mar	2011	Q1	22			2011-03	2011-03
2011-01-23	2011-01-23	Tue	Mar	2011	Q1	23			2011-03	2011-03
2011-01-24	2011-01-24	Wed	Mar	2011	Q1	24			2011-03	2011-03
2011-01-25	2011-01-25	Thu	Mar	2011	Q1	25			2011-03	2011-03
2011-01-26	2011-01-26	Fri	Mar	2011	Q1	26			2011-03	2011-03
2011-01-27	2011-01-27	Sat	Mar	2011	Q1	27			2011-03	2011-03
2011-01-28	2011-01-28	Sun	Mar	2011	Q1	28			2011-03	2011-03
2011-01-29	2011-01-29	Mon	Apr	2011	Q2	29			2011-04	2011-04
2011-01-30	2011-01-30	Tue	Apr	2011	Q2	30			2011-04	2011-04

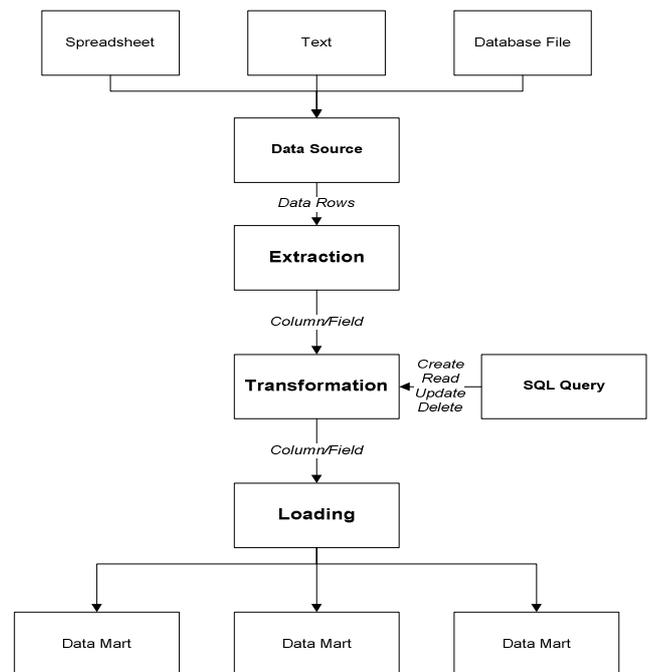
Gambar 16. Hasil Loading Tabel Time_Dim

Data warehouse yang telah dirancang pada tahapan sebelumnya akan direpresentasikan menggunakan *SQL Server OLAP Manager* dengan tujuan membantu analisis bagi PO. Sumber Alam mengenai jumlah pemakaian suku cadang untuk setiap armada di setiap proses perbaikan. Setelah membuat empat (4) tabel dimensi di tahapan sebelumnya, maka tabel-tabel tersebut akan didaftarkan ke dalam *OLAP Manager* dengan menerapkan pendekatan *star schema*. Perancangan model multidimensi *data warehouse* yang telah dibuat dapat dilihat pada Gambar 18 di bawah ini yang menunjukkan bagaimana keterhubungan antara tabel dimensi waktu (*Time_Dim*), dimensi perbaikan (*Services_Dim*), dimensi suku cadang (*Parts_Dim*), dan dimensi pemesanan (*Orders_Dim*) terhadap tabel fakta transaksi (*Trans_Fact*).



Gambar 17. Diagram Model Data Warehouse Menggunakan Star Schema

Skema yang digunakan untuk membangun pemodelan *data warehouse* adalah *star schema* dimana terdapat satu tabel fakta dan beberapa tabel dimensi. Penggunaan pendekatan *star schema* memungkinkan proses *query* yang lebih ringan dan memudahkan dalam penjelajahan terhadap data data dimensinya. Tabel fakta yang terbentuk dari perancangan *data warehouse* di bagian gudang PO. Sumber Alam ini merupakan tabel yang berhubungan dengan transaksi permintaan barang dalam hal ini yaitu suku cadang dari bagian bengkel ke bagian gudang guna kebutuhan perbaikan armada bus. Proses ETL yang terjadi pada pemodelan *data warehouse* PO. Sumber Alam dapat digambarkan seperti pada Gambar 18 berikut ini.



Gambar 18. Proses Extraction, Transformation, dan Loading Pada Pemodelan Data Warehouse PO. Sumber Alam

Proses ekstraksi yang terjadi saat pengambilan data dari sumber data eksternal dilakukan dengan proses pembentukan koneksi asal (sumber data) dan koneksi tujuan (data target). Pada proses ini berbagai sumber data selain format basis data dapat diambil datanya. Pada kasus ini, format sumber data yang ada PO. Sumber Alam antara lain dapat dilihat pada Tabel II.

TABEL II
SUMBER DATA EKSTERNAL PO. SUMBER ALAM

No.	Nama Sumber Data	Format Data
1	Parts	Database (.ldf)
2	Orders	Database (.ldf)
3	Services	Database (.ldf)
4	Time	Excel (.csv)

Proses berikutnya adalah melakukan transformasi data dimana pada proses ini data yang berasal dari sumber data eksternal yang terdapat pada Tabel II yang memiliki format dan tipe data berbeda akan diseragamkan menjadi format yang homogen ke dalam tabel *data warehouse* (tabel dimensi). *Data warehouse* yang dibangun memiliki kemampuan penyediaan informasi yang digunakan untuk melakukan analisis data terhadap kegiatan transaksi yang terjadi di bagian gudang dalam hal permintaan data oleh bagian bengkel. Kemampuan penyediaan informasi dari *data warehouse* ini digunakan untuk melakukan analisa terhadap permintaan suku cadang yang terjadi di bagian gudang PO. Sumber Alam. Kemampuan penyediaan informasi ini antara lain: (1) Penyediaan informasi mengenai jenis dan jumlah suku cadang setiap dimensi waktu. (2) Penyediaan informasi mengenai jumlah permintaan suku cadang oleh bagian bengkel setiap dimensi waktu. (3) Penyediaan informasi mengenai jenis perbaikan dan penggunaan suku cadang setiap dimensi waktu. Adapun ketersediaan informasi suku cadang tersebut dapat dilihat dari hasil multidimensi *data warehouse* pada penelitian ini seperti terlihat pada Gambar 19 berikut ini.

Gambar 19. Informasi Suku Cadang di PO. Sumber Alam

Sedangkan ketersediaan informasi mengenai jenis dan jumlah pemesanan permintaan suku cadang dapat dilihat dari hasil multidimensi *data warehouse* pada penelitian ini seperti terlihat pada Gambar 20 berikut ini.

Gambar 20. Informasi Permintaan Suku Cadang di PO. Sumber Alam

Ketersediaan informasi mengenai jenis perbaikan armada bus PO. Sumber Alam dapat dilihat dari hasil multidimensi *data warehouse* pada penelitian ini seperti terlihat pada Gambar 21 berikut ini.

Gambar 21. Informasi Jenis Perbaikan di PO. Sumber Alam

V. KESIMPULAN

Berdasarkan hasil pembahasan pada bab sebelumnya terhadap perancangan model *data warehouse* guna manajemen sistem informasi pergudangan PO. Sumber Alam, maka dapat ditarik beberapa kesimpulan, antara lain: berdasarkan penelitian pada Bagian Gudang di PO. Sumber Alam, terlihat pada perusahaan tersebut membutuhkan suatu sistem yang mampu mengolah data barang yang terdapat maupun yang digunakan untuk aktivitas operasional setiap hari seperti permintaan suku cadang untuk perbaikan armada dan pencatatan penggunaan suku cadang tersebut yang masih manual sampai saat ini sehingga perlu suatu sistem yang dapat menampung data-data tersebut yaitu dengan menggunakan *data warehouse*. Analisis kinerja proses ETL pada perancangan model *data warehouse* yang akan diimplementasikan pada PO. Sumber Alam dihasilkan bahwa proses *query* yang terjadi dengan menggunakan pendekatan *star schema* dimana dibuat empat (4) tabel dimensi, antara lain: dimensi waktu (Time_Dim), dimensi suku cadang (Parts_Dim), dimensi perbaikan (Services_Dim), dan dimensi permintaan (Orders_Dim) dapat dilakukan dengan eksekusi yang cepat menggunakan *tools Transformation Data Task* dimana sumber data asal akan dihubungkan ke dalam masing-masing data tujuan yaitu tabel dimensi *data warehouse*. Sedangkan untuk mengintegrasikan ke-empat tabel dimensi tersebut, dibangunlah suatu tabel fakta yaitu fakta transaksi (Trans_Fact). Model *data warehouse* yang telah dibangun kemudian dapat digunakan untuk membangun sistem informasi manajemen pergudangan yang nantinya akan dibuat *interface* untuk masing-masing kebutuhan informasi yang diinginkan oleh PO. Sumber Alam.

DAFTAR PUSTAKA

- [1] Ma, X., Yao-Jan Wu, Yinhai Wang, 2014, *E-Science Transportation Platform for Data Sharing, Visualization, Modeling, and Analysis*, Transportation Research Record: Journal of the Transportation Research Board Vol. 2215.
- [2] Gour, V., S. Sarangdevot, Govind S., and Anand S., 2010, *Improve Performance of Extract, Transform, and Load (ETL) in Data Warehouse*, International Journal on Computer Science and Engineering Vol. 02 No. 03 pp. 786-789.
- [3] Lin, S. H., Yuan-Chii G. L., and Chien-Yeh H., 2010, *Data Warehouse Approach to Build a Decision-Support Platform for Orthopedics Based on Clinical and Academic Requirements*, International Journal of Bio-Science and Bio-Technology Vol. 57 pp. 89-96.
- [4] Inmon, W. H., 2002, *Building The Data Warehouse 3rd ed.*, New York: John Wiley & Sons.

- [5] Chou, Jui-Sheng; Tseng, Hsien-Cheng, 2011, *Establishing Expert System for Prediction Based on the Project-Oriented Data Warehouse*, Expert Systems With Applications Journal Vol. 38 pp. 640-651.
- [6] Kerkri, E. M., C. Quantin, et. al., 2001, *An Approach for Integrating Heterogeneous Information Sources in a Medical Data Warehouse*, J Med Syst Vol. 25 Issue 3 pp. 167-176.
- [7] Roelofs, E., dkk., 2013, *Benefits of a Clinical Data Warehouse with Data Mining Tools to Collect Data for a Radiotherapy Trial*, Radiotherapy and Oncology Journal Vol. 108 Issue 1 pp. 174-179.
- [8] Sultan, F., Abdul Aziz, 2010, *Deal Strategy to Improve Datawarehouse Performance*, International Journal on Computer Science and Engineering Vol. 02 No. 02 pp. 409-415.
- [9] Ponniah, P., 2001, *Data Warehousing Fundamentals*, Wiley-Interscience Publication.
- [10] Glorio, O., dkk., 2011, *A Personalization Process for Spatial Data Warehouse Development*, Decision Support System Journal Vol. 52 Issue 4 pp. 884-898.