

PERANCANGAN DAN IMPLEMENTASI DATA WAREHOUSE MENGUNAKAN SCHEMA SNOWFLAKE UNTUK MENGETAHUI TREND PRODUKSI DAN PEMASARAN PRODUK

Novia Busiarli¹⁾, Mardhiya Hayati²⁾

^{1), 2),3)} Teknik Informatika STMIK AMIKOM Yogyakarta

Jl Ring road Utara, Condongcatur, Sleman, Yogyakarta 55281

Email : novia.busiarli@students.amikom.ac.id¹⁾, mardhiya_hayati@amikom.ac.id²⁾

Abstrak

Ketatnya persaingan menyebabkan kebutuhan penggunaan teknologi informasi meningkat. Analisa data menjadi kebutuhan penting untuk meningkatkan daya saing. Kinerja organisasi akan meningkat dan memiliki daya saing jika data terintegrasi dengan baik. Pengambilan keputusan bergaya entrepreneur cenderung mengandalkan intuisi menjadi kurang relevan ditengah persaingan yang semakin kompleks, sehingga manajemen dapat mengambil keputusan berdasarkan fakta aktual dan tidak hanya mengandalkan intuisi dan pengalaman kuantitatif saja.

Perkembangan teknologi informasi mengalami kemajuan pesat dan memberi peran besar dalam perubahan mendasar khususnya dalam manajemen sebuah organisasi. Pihak manajemen memiliki kemampuan menganalisa dan membuat strategi potensial untuk bersaing dengan pasar lain. Analisa yang tepat, benar dan lengkap menghasilkan informasi yang sangat berharga bagi sebuah organisasi dalam menanggapi trend atau progress bisnis yang sedang terjadi. Kebutuhan manajemen menganalisa data operasional dan transaksional bertujuan untuk meningkatkan kualitas dan kepuasan pelanggan. Sehingga diperlukan media penyimpanan data ringkasan dari sumber data operasional maupun sumber data transaksional berupa database data warehouse. Selanjutnya data di dalam database data warehouse diolah dan diproses dengan teknologi Online Analytical Processing (OLAP). Online Analytical Processing (OLAP) merupakan kunci business intelligence digunakan untuk menganalisa data sebagai dasar dari pengambilan keputusan atau decision support system pada sebuah organisasi. Hasil dari proses tersebut berupa laporan analisa statistik berupa grafik dan tabel serta laporan dashboard pada webportal sistem business intelligence organisasi.

Kata kunci: Database, Data Warehouse, Online Analytical Processing (OLAP), Trend.

1. Pendahuluan

Data adalah *asset* sebuah organisasi tersimpan di dalam media penyimpanan baik manual maupun digital. Data tersimpan dalam bentuk format data yang berbeda-beda, data juga terdistribusi dalam hal lokasi media

penyimpanannya. Data digital di tampung dalam sebuah basisdata (*database*) digunakan untuk membuat kebijakan, melakukan strategi manajemen, atau mengambil keputusan. Tumpukan data di *database* tidak akan berarti jika tidak diolah untuk menghasilkan sebuah informasi yang dapat menunjang proses bisnis atau untuk menghasilkan sebuah keputusan. Dasar teknologi informasi adalah sebuah *database*. Analisa data menjadi kebutuhan penting untuk meningkatkan daya saing. Sehingga diperlukan media penyimpanan data ringkasan dari sumber data operasional maupun sumber data transaksional berupa *database data warehouse*. Selanjutnya data di dalam *database data warehouse* diolah dan diproses dengan teknologi *Online Analytical Processing (OLAP)*. *Online Analytical Processing (OLAP)* merupakan kunci *business intelligence* digunakan untuk menganalisa data sebagai dasar dari pengambilan keputusan atau *decision support system* pada sebuah organisasi. Hasil dari proses tersebut berupa laporan analisa statistik berupa grafik dan tabel serta laporan *dashboard* pada *webportal* sistem *business intelligence* organisasi.

Berdasarkan permasalahan yang telah dipaparkan diatas dapat dirumuskan permasalahan sebagai berikut: Bagaimana merancang *Data Warehouse* untuk mengetahui *trend* produksi dan pemasaran produk dengan skema *snowflake* sebagai metode perancangan untuk menghasilkan sebuah informasi yang dipakai oleh pihak manajemen dalam menyusun strategi bisnis dikemudian hari. permasalahan akan dibatasi pada hal-hal yang berkaitan dengan perancangan data *warehouse* ini sebagai berikut:

1. Data ditransformasikan dalam format yang berbeda-beda dari sumber *database* yang berbeda-beda ke *database data warehouse*.
2. Data diproses melalui proses *Extract Transform Load (ETL)* untuk disimpan ke *database data warehouse*.
3. Analisa data direpresentasikan menggunakan teknologi *Online Analytical Processing (OLAP)* dalam bentuk laporan analisa statistik berupa pivot tabel dan pivot grafik.
4. Skema perancangan *Data Warehouse* adalah *Snowflake*.

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengetahui proses transformasi data dari sumber yang berbeda melalui mekanisme *Extract Transform*

Load (ETL) dalam merepresentasikan analisa data menggunakan teknologi *Online Analytical Processing (OLAP)*.

2. Menambah pengetahuan dan wawasan tentang *Database*, *Data Warehouse* dan proses yang terlibat di dalamnya.

Penelitian ini dilakukan melalui beberapa tahapan:

1. Tahap Identifikasi Masalah.
Melalui proses studi *literature* dan observasi untuk menghasilkan identifikasi masalah yang merupakan langkah awal dalam menganalisa proses bisnis.
2. Tahap Analisa Masalah.
Proses yang dilalui berupa membuat rumusan masalah untuk menghasilkan rumusan permasalahan yang terjadi saat ini kemudian dilanjutkan dengan mengumpulkan data sumber dengan teknik observasi serta mendefinisikan sumber kebutuhan.
3. Tahap Perancangan *Data Warehouse*.
Proses yang dilakukan membuat rancangan / arsitektur data *warehouse*, menentukan skema / permodelan data *warehouse* dengan melihat subjek kebutuhan, melakukan proses *Extract Transform Load (ETL)* dengan format beberapa sumber data yang berbeda-beda kemudian merancang analisa grafik dan *reporting* untuk kebutuhan bisnis.
4. Tahap Pengujian.
Proses yang dilakukan melalui pengujian perancangan data *warehouse* dengan tranformasi format data sumber yang berbeda-beda, pemisahan *database* antara *database* transaksional *Online Transactional Processing (OLTP)* dengan *database* data *warehouse* serta mengimplementasikan teknologi *multidimensional query (OLAP)*.
5. Tahap Kesimpulan dan Saran.
Penarikan kesimpulan dari penelitian yang sudah dilakukan. Kemudian evaluasi sistem untuk saran dan pengembangan sistem di masa depan.

Wilaksana (2013), meneliti tentang Pengembangan *Data Warehouse* Untuk Pengelolaan Tidak Lanjut Hasil Pemeriksaan BPK RI Menggunakan Skema *Star*. Hasil penelitiannya adalah data yang termasuk kedalam *table* fakta merupakan data hasil pemeriksaan yang terdiri dari temuan pemeriksaan. Sedangkan dimensi yang digunakan adalah dimensi tahun anggaran, tahun pemeriksaan, jenis temuan, entitas pemeriksaan dan objek pemeriksaan. Manajer dapat menentukan status pemantauan TLHP dan menentukan susunan tim pemeriksa berdasarkan karakter masing-masing entitas pemeriksaan. Dari penelitian yang sudah dilakukan masih terdapat kelemahan saat proses *Extract Transform Load (ETL)* yang dilakukan yaitu proses *extract* dan *transform* hanya dari 1 format file (.xls) yang hanya dengan proses *upload* akan di *load* ke *database* dan belum bisa mengetahui *error* dari *upload* data[1].

Yuris (2012), meneliti tentang Perancangan dan Pengembangan *Data Warehouse Marketing* Sebagai Alat Bantu Pengambilan Keputusan Pemasaran pada PT Serambi Ilmu Semesta. Hasil penelitiannya adalah

database Online Transactional Processing (OLTP) yang tersedia, rancangan *Dimensional modeling* membentuk *star schema* dengan beberapa kategori dihasilkan aplikasi *Online Analytical Processing (OLAP)* dengan kombinasi laporan dengan tingkat *granularity* yang berbeda-beda. Pengembangan dari penelitian berikutnya diharapkan untuk menambah data *mining* dan ditampilkan dengan pemanfaatan *dashboard*, dengan fitur *marketing mix* akses *end user* terhadap aplikasi dapat dikelompokkan lebih detail lagi dan penambahan data *measures* pada aplikasi *Online Analytical Processing (OLAP)* untuk tampilan laporannya[2].
Pada penelitian ini akan dilakukan pengembangan dari penelitian sebelumnya dengan menggunakan beberapa format dokumen / file yang berbeda untuk proses *extraction*, diantaranya: *SQL file*, *sheet file*, *text file*, *XML file*, *Raw data*. Untuk merepresentasikan laporan hasil *execute multidimensional query* dalam bentuk analisa berupa *pivot tabel* dan *pivot grafik*.

Menurut Pusadan (2013), *Data warehouse* merupakan kumpulan data yang berorientasi subjek, terintegrasi, tidak dapat di *update*, memiliki dimensi waktu yang digunakan untuk mendukung proses manajemen pengambilan keputusan dan kecerdasan bisnis[3]. *Data warehouse* memiliki karakteristik sebagai berikut:

- a. Berorientasi Subjek
Data diorganisasi oleh subjek detail (misal berdasarkan pelanggan, jenis kebijakan, dan klaim dalam perusahaan asuransi), yang berisi informasi yang relevan untuk mendukung keputusan. *Data warehouse* berbeda dengan *database* operasional. Pada umumnya, *database* operasional mempunyai sebuah orientasi produk untuk menangani transaksi yang memperbarui *database*.
- b. Terintegrasi
Data yang terintegrasi mengatasi inkonsistensi dan menyediakan istilah yang seragam di organisasi keseluruhan, juga format waktu dan data yang bervariasi.
- c. *Time Variant*
Data tidak menyimpan status saat ini. Mereka disimpan untuk 5 atau sepuluh tahun atau lebih dan digunakan untuk *trend*, peramalan, dan perbandingan. Ada kualitas sementara pada sebuah data *warehouse*. Waktu adalah dimensi penting yang harus didukung oleh semua data *warehouse*. Data untuk analisis dari berbagai sumber berisi sebagai poin waktu (misal harian, mingguan, bulanan).
- d. *Non volatile*
Data adalah *read only*, mereka tidak bisa dirubah atau dibarui. Data usang dibuang, dan perubahan akan direkam sebagai data baru. Ini memungkinkan data *warehouse* untuk disesuaikan hampir secara eksklusif untuk akses data.
- e. *Granularity*
Pada sistem operasional data dibuat secara *real-time* sehingga untuk mendapatkan informasi langsung dilakukan proses *query*. Pada data *warehouse* proses

menganalisis harus memperhatikan *level-of-detail*, semakin detail *granularity*-nya semakin baik.

Menurut Ponniah (2001), *Snowflake* skema adalah salah satu permodelan data untuk data *warehouse*. Metode *snowflake schema* adalah pengembangan dari *star schema*, yaitu metode normalisasi untuk tabel dimensi yang ada dalam *star schema*. *snowflake* skema adalah gabungan dari beberapa permodelan data *star* skema yang dinormalisasi. Sedangkan menurut ,Dalam *snowflake schema*, setiap tabel dimensi dapat memiliki sub-tabel *dimensi* lagi[4]. Hal ini bertujuan untuk meminimalkan data yang berlebihan (*redundancy* data).

Menurut Arifin (2013), Adapun kelemahan dari penggunaan metode *snowflake* skema adalah[5]:

1. Proses *query* akan lebih rumit karena terdiri dari banyak sub dimensi yang saling terhubung.
2. Proses *loading* data harus dikelola dan mendapat perhatian lebih untuk menghindari terjadinya nomali *update* maupun *unomali insert*.

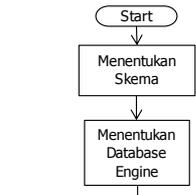
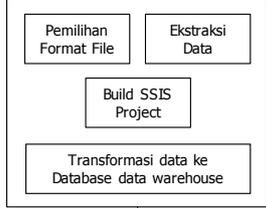
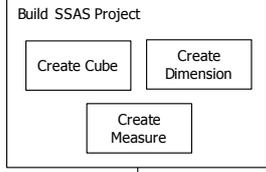
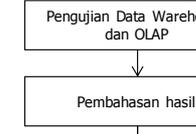
2. Pembahasan

Dalam hal ini, Penulis menggunakan metode analisa *PIECES*. Karena metode ini lebih tepat penggunaannya dalam hal analisa data yang bersifat *kuantitatif*. Metode ini digunakan dalam hal analisa untuk mengidentifikasi masalah yang ditemukan berdasarkan kriteria-kriteria berikut:

- a. *Performance*
 Mengintegrasikan 5 *format* sumber data yang berbeda-beda. Membuat *Cube*, *Dimension*, *measure* dan *Multidimensional query*.
- b. *Information*
 Menggunakan *SQL Server Integration Service (SSIS)* sebagai *tool ETL* untuk mendeteksi *error / noise*. Menggunakan *SQL Server Reporting Service (SSRS)* sebagai *tool* untuk membuat laporan bentuk *pivot* tabel, grafik, dan *dashboard* 3 dimensi.
- c. *Economic*
 Membuat sistem yang menampilkan laporan untuk mengurangi penggunaan kertas.
- d. *Control*
 Merancang data *Warehouse* untuk level *Manager* keatas.
- e. *Efficiency*
 Menggunakan *multidimensional query* untuk menampilkan data berdasarkan dimensi waktu..
- f. *Service*
 Membuat sistem yang merepresentasikan hasil analisa dalam bentuk *pivot* tabel, grafik dan *dashboard*.

Peneliti akan menerangkan alur / diagram kerja dari perancangan data *warehouse* seperti tabel 1.

Tabel 1. Tahapan Perancangan Data Warehouse

Kegiatan	Proses	Output
Perancangan Arsitektur	 <pre> graph TD Start([Start]) --> MenentukanSkema[Menentukan Skema] MenentukanSkema --> MenentukanDatabaseEngine[Menentukan Database Engine] </pre>	Snowflake Skema OLTP : MySQL OLAP : SQL Server 2012
Extraction dan Transformation	 <pre> graph TD subgraph Box1 [] direction LR A[Pemilihan Format File] B[Ekstraksi Data] end Box1 --> C[Build SSIS Project] C --> D[Transformasi data ke Database data warehouse] </pre>	Data Staging dan Database Data Warehouse
Loading Process	 <pre> graph TD A[Build SSAS Project] --> B[Create Cube] A --> C[Create Dimension] B --> D[Create Measure] C --> D </pre>	OLAP Multidimensional Query
Pengujian	 <pre> graph TD A[Pengujian Data Warehouse dan OLAP] --> B[Pembahasan hasil] </pre>	Report : tabel, grafik dan dashboard
Kesimpulan dan Saran	 <pre> graph TD A[Kesimpulan] B[Saran untuk pengembangan] </pre>	Kesimpulan dan Saran

Perancangan *database data warehouse* ini terpisah oleh *Database Engine* yang berbeda. Di dalam penelitian ini, peneliti menggunakan *database Mysql* sebagai *database Online Transaction Processing (OLTP)*, sedangkan untuk *database Online Analytical Processing (OLAP)* menggunakan *SQL Server 2012* dan dibantu oleh *tools* lain seperti *Microsoft Visual Studio Business Intelligent* untuk *build SSIS, SSAS dan SSRS project*.

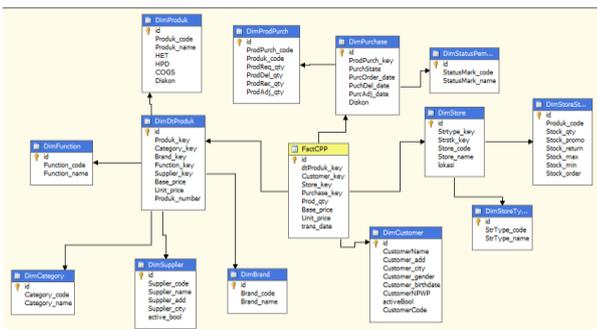
Untuk *Source* data, peneliti menggunakan beberapa *format* data yang berbeda untuk proses *extraction*, diantaranya: *SQL file, sheet file, text file, XML file, Raw data*. Peneliti juga merancang skema *Snowflake* pada *database data warehouse* karena dapat merelasikan dimensi yang lebih banyak sehingga banyak informasi yang dapat dihasilkan.

Salah satu *source* data yang digunakan didalam penelitian ini adalah data transaksional sistem informasi yang sudah berjalan atau yang disebut data *primer database Online Transaction Processing (OLTP)*. Peneliti juga merepresentasikan hasil *execute multidimensional query* dalam bentuk analisa laporan berupa grafik, tabel serta *dashboard*. Hasil analisa akan digunakan pihak manajemen untuk kebutuhan bisnis lebih lanjut.

2.1. Rancangan Skema Snowflake

Skema yang dirancang berbentuk *snowflake* karena dengan menggunakan skema ini, kita dapat mendapatkan informasi rinci dari dimensi yang terhubung ke table fakta. Adapun rancangan dari skema *snowflake* ini terdiri dari beberapa dimensi diantaranya: Dimensi Produk, Dimensi Penawaran, Dimensi Customer, dimensi waktu dan dimensi pemasaran. Dari dimensi produk dapat dijabarkan lagi sub dimensi yaitu: dimensi jenis barang, dimensi kategori, dimensi fungsi, dimensi Brand, Dimensi Supplier.

Untuk Dimensi penawaran memiliki sub dimensi yaitu: dimensi datapenawaran dan dimensi status penawaran, sedangkan untuk dimensi store memiliki sub dimensi stok toko dan dimensi type store. Sehingga jika digambarkan akan terlihat seperti gambar 1.

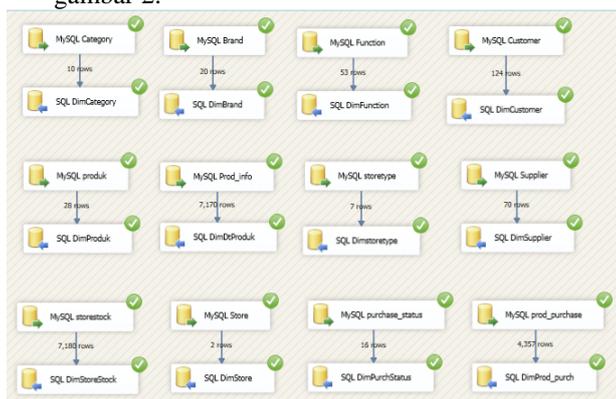


Gambar 1. Skema Snowflake table fakta Pemasaran

2.2. Proses ETL (Extraction, Transformation dan Loading)

1. Extraction

Extraction adalah suatu proses yang mengidentifikasi seluruh sumber data yang relevan dan kemudian mengambil data dari sumber-sumber data tersebut. Pengujian yang dilakukan dalam proses *extraction* ini akan terlihat pada gambar 2.

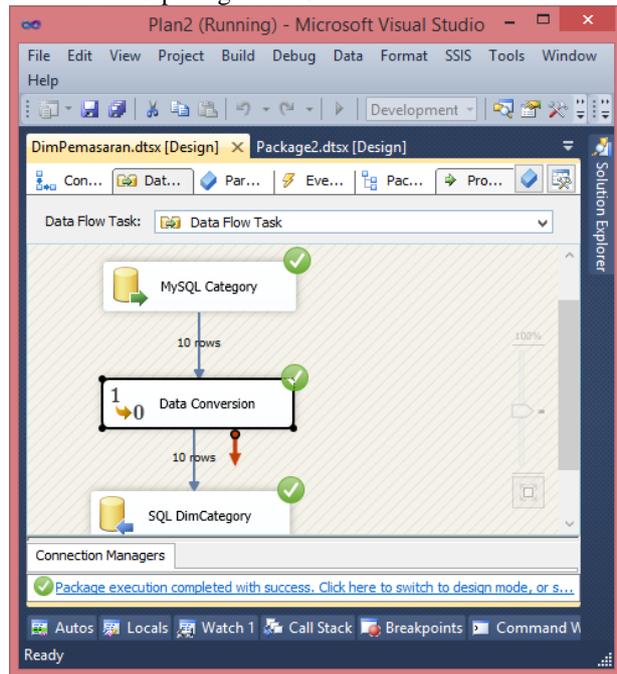


Gambar 2. Pengujian pada proses Extraction

2. Transformation

Transformation adalah suatu proses yang memiliki peran dalam melakukan perubahan dan integrasi skema data serta struktur yang telah didefinisikan sebelumnya oleh data *warehouse*. Pengujian yang

dilakukan dalam proses *transformation* ini akan terlihat seperti gambar 3.



Gambar 3. Pengujian pada proses Transformation

3. Loading

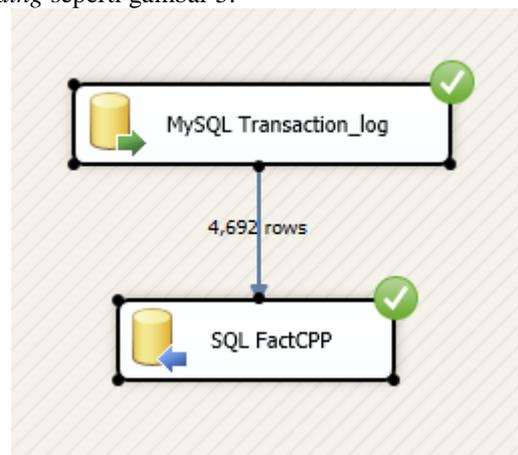
Loading adalah suatu proses pemindahan data secara fisik dari sistem operasional ke dalam data *warehouse*.

Sebelum proses loading, kita akan mengetahui berapa *record* data yang tersedia seperti gambar 4.



Gambar 4. Hasil Query pada database OLTP

Jumlah *record* tersebut akan kita lihat pada proses *loading* seperti gambar 5.



Gambar 5. Pengujian pada proses Loading

Kemudian akan kita lihat *record* data pada database *OLAP* seperti gambar 6.

id	dtProduk_key	Customer_key	Store_key	Purchase_key	Prod_qty	Base_price
1	57	57	1	NULL	1	130470.00
2	57	57	1	NULL	1	120470.00

Gambar 6. Hasil query pada database OLAP

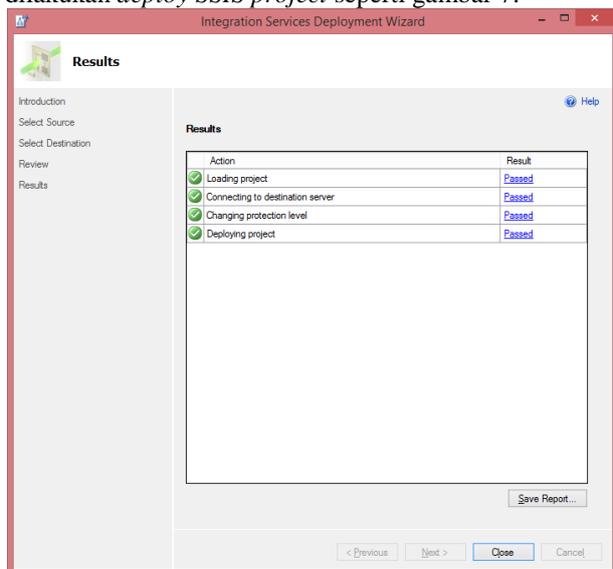
2.3. Rancangan dan Implementasi Cube, Dimension dan Measure

Cube berisi kumpulan banyak data yang telah disatukan (di agregasi) sehingga mempercepat query. Adapun bagian-bagian dari cube adalah:

- a. Storage location
 Lokasi penyimpanan cube data dapat disimpan di sejumlah server analisis menggunakan didistribusikan baru, dipartisi fasilitas kubus.
- b. Action
 Action adalah set operasi yang telah didefinisikan di muka. Sebagai pengguna, dapat memanggil cube atau hanya digunakan untuk keperluan analisis.
- c. Dimension
 Dimension adalah bagian dari cube yang menggambarkan suatu kategori data, yang digunakan sebagai dasar untuk merangkum data. Dimension mengacu kepada kategori yang digunakan untuk mengatur measure.
- d. Measure
 Measure adalah bagian dari cube yang berisi kumpulan data yang akan dirangkum, pada umumnya data yang tergolong dalam measure ini tidak memiliki hirarki atau jenjang, dan dapat dilakukan proses matematis.

2.4. Deploy SSIS Project

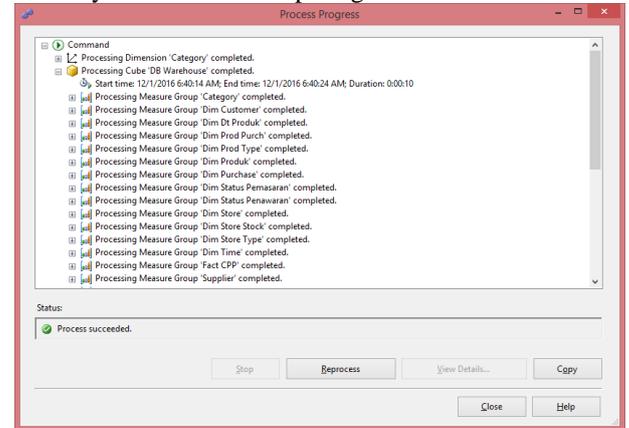
SSIS project merupakan package project dalam melakukan prose ETL (Extraction, Transformation, Loading). Setelah proses ETL (Extraction, Transformation, Loading) berjalan dengan baik maka dilakukan deploy SSIS project seperti gambar 7.



Gambar 7. Hasil Deploy SSIS Project

2.5. Deploy Cube pada SSAS Project

SSAS project merupakan package project dalam mendeploy cube. Dimana SSAS package terdiri atas data source, data source view, cube, dimension, dan hierarki. Untuk membuild SSAS project harus mendeploy Cube dan dinyatakan berhasil seperti gambar 8.



Gambar 8. Hasil Deploy Cube pada SSAS Project

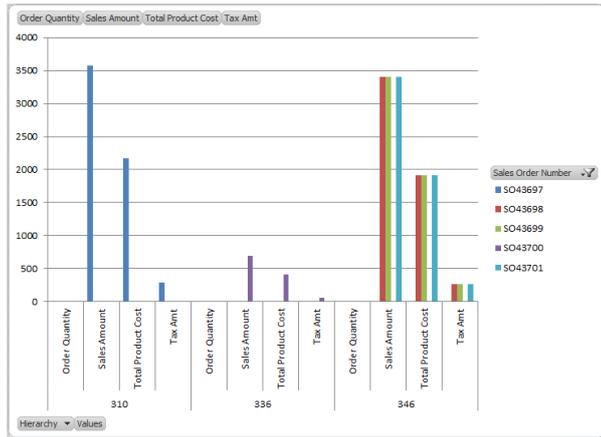
2.6. Pivot Table dan Pivot Chart

Hasil dari deploy cube adalah data cube yang dapat dibuat analisa laporan dalam bentuk pivot table, dari cube bisa di filter data yang akan di tampilkan seperti gambar 9.

Row Labels	SO43697	SO43698	SO43699	SO43700	SO43701	Grand Total
310						
Order Quantity	1					1
Sales Amount	3578.27					3578.27
Total Product Cost	2171.2942					2171.2942
Tax Amt	286.2616					286.2616
336						
Order Quantity				1		1
Sales Amount				699.0982		699.0982
Total Product Cost				413.1463		413.1463
Tax Amt				55.9279		55.9279
346						
Order Quantity	1	1			1	3
Sales Amount	3399.99	3399.99		3399.99		10199.97
Total Product Cost	1912.1544	1912.1544		1912.1544		5736.4632
Tax Amt	271.9992	271.9992		271.9992		815.9976
Total Order Quantity	1	1	1	1	1	5
Total Sales Amount	3578.27	3399.99	3399.99	699.0982	3399.99	14477.3382
Total Total Product Cost	2171.2942	1912.1544	1912.1544	413.1463	1912.1544	8320.9037
Total Tax Amt	286.2616	271.9992	271.9992	55.9279	271.9992	1158.1871

Gambar 9. Hasil Analisa Pivot Table

Sedangkan untuk merepresentasikan data ke dalam bentuk Pivot Chart dengan tools pivot grafik sehingga dari pivot table akan tergambar seperti gambar 10.



Gambar 10. Hasil Analisa Pivot Grafik

Mardhiya Hayati, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Teknik Industri Universitas Ahmad Dahlan, lulus tahun 2003. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Teknik Informatika STMIK AMIKOM Yogyakarta, lulus tahun 2013. Saat ini menjadi Dosen tetap di STMIK AMIKOM Yogyakarta.

3. Kesimpulan

Data *warehouse* dapat dibangun dari sumber data dengan format dokumen / *file* yang berbeda-beda, tidak hanya berasal dari *file SQL / csv* saja. Proses *extraction* dilakukan dengan pemilihan *column* yang akan dipetakan ke *database OLAP* dengan menggunakan *connection manager* untuk mengkonfigurasi *connection string source* datanya. Sedangkan proses *transformation* adalah bagaimana mengolah, merubah dan mentransformasikan data yang sama seperti *type* data, ukuran panjang data. Proses *loading* yaitu memindahkan data secara tepat ke *table* yang telah ditentukan dan tidak boleh bernilai *null* dan memastikan jumlah *record* data yang masuk ke *database OLAP*. Analisa laporan dalam bentuk *pivot table* dan *pivot grafik* didapat dari hasil rancangan skema *snowflake* yang dapat menormalisasikan *table-table* dimensi dalam bentuk *hierarki* dalam sebuah *cube*.

Daftar Pustaka

- [1] Wilaksana, A, *Pengembangan Data Warehouse untuk Pengelolaan Tindak Lanjut Hasil Pemeriksaan BPK RI*, Yogyakarta, 2013.
- [2] Yuris, R, *Perancangan dan Implementasi Data Warehouse Manajemen Resiko Kredit pada Bank BPD DIY*, Yogyakarta, 2012.
- [3] Pusadan, M. Y, *Rancang Bangun Data Warehouse*, Yogyakarta, Graha Ilmu, 2013.
- [4] Ponniah, P, *Data Warehouse Fundamentals: a Comprehensive Guide for IT profesional*. New York: John Wilay end Sons, 2001..
- [5] Arifin; Kusrini; Emha Taufiq Luthfi. (2014). Penerapan dan Implementasi Data Warehouse Pemasaran pad CV. Diandra Primamitra Media Yogyakarta. *Seminar Nasional Teknologi Informasi dan Multimedia*, 2.04-79.

Biodata Penulis

Novia Busiarli, memperoleh gelar Ahli Madya Komputer (AMd), Jurusan Teknik Informatika Politeknik Negeri Padang, lulus tahun 2011. Saat ini sedang melanjutkan pendidikan ke jenjang Strata-1 di STMIK AMIKOM Yogyakarta.