

## Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho

Darmawan Subuh<sup>1)</sup>, Wita Yasman<sup>2)</sup>

<sup>1),2)</sup> Sistem Informasi, STMIK Indonesia Jakarta  
Jl. Siantar No.6 Jakarta Pusat  
Email : darmawan@stmik-indonesia.ac.id

**Abstrak** - Informasi mengenai data toko yang dibutuhkan oleh seorang pemilik toko untuk melihat kondisi penjualan. Disisi lain Toko Magnifique Clothes hanya menyediakan informasi penjualan yang berasal dari M.Access yang ada dikomputer toko tersebut dan tidak dapat menyediakan atau menyajikan informasi dengan cepat. Untuk mengatasi permasalahan tersebut dibangun sebuah data warehouse di Toko Magnifique Clothes untuk mendapatkan informasi yang cepat. Penelitian ini akan melakukan Implementasi Data Warehouse penjualan dari sumber data, proses extraction, transformation, loading (ETL) menggunakan tools Pentaho, pembuatan Starschema berupa dimensi-dimensi yang terhubung dengan fact table penjualan. Kemudian hasil data warehouse dianalisa melalui proses OLAP (On-line Analytical Processing), pembuatan cube atau schema workbench, dan pembuatan dashboard untuk Visualisasi dalam penyajian informasi yang diharapkan dari Toko Magnifique. Hasil dalam penelitian ini mencakup data penjualan yang digunakan dari tahun 2013-2017 berupa tampilan grafik atau dashboard penjualan, barang yang laku terjual, dan keuntungan yang didapat sebagai penyampaian informasi penjualan pada toko tersebut.

**Kata Kunci** : Data Warehouse, ETL, Pentaho, OLAP, cube, schema workbench, Dashboard, M.Access

### 1. Pendahuluan

Sistem informasi yang semakin pesat membawa pengaruh besar terutama pada dunia bisnis. Sistem informasi menjadi salah satu bentuk implementasi yang bisa menjawab kebutuhan perusahaan dalam menjalankan usahanya dan mengambil keputusan yang tepat bagi perusahaan. Hal ini penting mengingat persaingan bisnis yang semakin ketat, sehingga menjadikan perusahaan harus mengikuti trend zaman yang semakin inovatif untuk bisa sukses dan bertahan dalam persaingan ini.

Adanya kebutuhan perusahaan (Toko Magnifique) untuk bisa mengambil keputusan secara cepat dan tepat, mengetahui seberapa banyak keuntungan yang didapat setiap bulannya dan bisa membandingkan penghasilan setiap tahun, membuat perusahaan membutuhkan tools yang dapat membantu perusahaan untuk menganalisa dan mengelola data menjadi informasi yang berkualitas sehingga bisa digunakan untuk pengambilan keputusan. Kebutuhan perusahaan akan informasi menjadi kebutuhan pokok dalam kelangsungan hidup suatu perusahaan. Data Warehouse merupakan basis data yang dirancang untuk mengerjakan proses query, membuat laporan dan analisa. Data yang disimpan dalam sebuah data warehouse adalah data histori dari organisasi/perusahaan yang mana data tersebut tidak tersimpan secara rinci/detail.

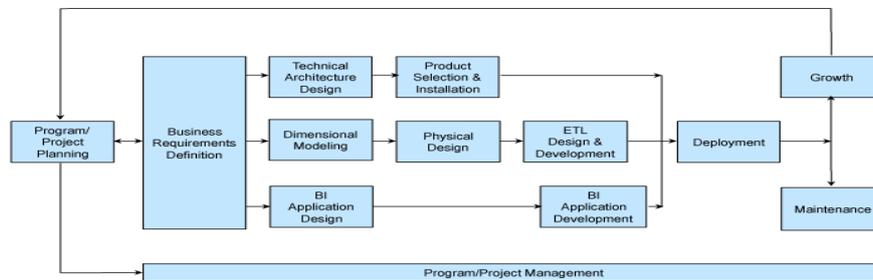
Tujuan utama dari perancangan Data Warehouse adalah untuk menyatukan data yang beragam yang berasal dari berbagai sumber ke dalam sebuah tempat penyimpanan dimana pengguna dapat dengan mudah melakukan pencarian data, menghasilkan tampilan dashboard sebagai laporan serta melakukan analisis. Merancang Data Warehouse dan Penerapannya dengan menggunakan perangkat lunak Pentaho Data Integration, Pentaho Schema Workbench dan BI Server pada Toko Magnifique Clothes.

### 2. Metodologi Penelitian

Metode Penelitian adalah cara ilmiah untuk mendapatkan data yang valid dengan tujuan dapat ditemukan, dibuktikan, dan dikembangkan suatu pengetahuan, sehingga pada gilirannya dapat digunakan untuk memahami, memecahkan dan mengantisipasi masalah dalam bidang bisnis. Metodologi penelitian ini terdiri dari metodologi pengembangan Data Warehouse dan metode pengumpulan data

## 2.1 Metodologi Pengembangan Data Warehouse

Dalam mengimplementasikan *Data Warehouse*, dapat digunakan beberapa metode yang ada. Dalam penulisan ini, metode pengembangan Data Warehouse yang dibahas adalah menggunakan pendekatan *Kimball Life Cycle* seperti yang tergambar pada Gambar 2.1. di bawah ini.



Gambar 2. 1. *Kimball Life Cycle* [1].

## 3. Landasan Teori

### 3.1 Data Warehouse

Menurut William Harvey Inmon (W.H. Inmon, 1970) atau lebih dikenal dengan sebutan Bill Inmon, merupakan salah satu tokoh penting di dalam perkembangan dunia teknologi informasi, khususnya dibidang *data warehouse*. *Data warehouse* didefinisikan sebagai sekumpulan data yang dimiliki enam buah sifat atau karakteristik berupa berorientasi subjek (*subject oriented*), terintegrasi (*integrated*), berorientasi pada proses (*process oriented*), time variant, dapat diakses dengan mudah (*accessible*), dan bersifat *non volatile* [2].

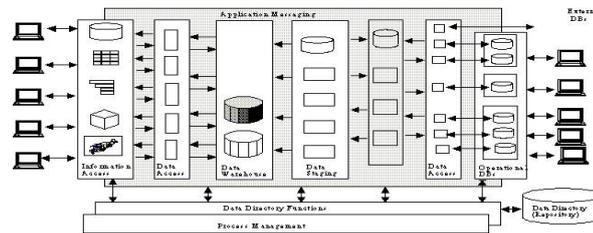
### 3.2 Karakteristik Data Warehouse

Menurut Bill Inmon (1970) terdapat enam buah karakteristik pada *data warehouse*, yang meliputi *Subject Oriented*, *Integrated*, *Process Oriented*, *Time Variant*, *Accessible* dan *Non Volatile* [2].

- 1) *Subject Oriented* adalah Data yang ditampilkan dan data yang disusun hanyalah data menurut subjek yang diperlukan untuk proses pengambilan keputusan. Data-data ini dirangkum ke dalam bentuk dimensi, yang meliputi periode waktu (*time*), riwayat (*history*), wilayah (*region*), dan lain-lain.
- 2) *Integrated* adalah *Data warehouse* dibangun dari proses integrasi berbagai sumber data (terutamanya sumber data berupa *database*) yang berasal dari berbagai aplikasi (*software*), menjadi satu kesatuan yang utuh. Karakteristik *integrated* muncul sebagai bentuk lain dari karakteristik *subject oriented*.
- 3) *Time Variant* adalah *Data warehouse* (yang dikumpulkan dari berbagai sumber data dari berbagai aplikasi) diidentifikasi dari periode waktu penyimpanannya. Atau dengan kata lain, proses penyimpanan data-data yang dikumpulkan tersebut, berdasarkan kepada waktu. Adapun informasi yang disajikan oleh data-data pada *data warehouse* tersebut, dilihat sudut pandang riwayat penyimpanan (*historical point of view*).
- 4) *Non Volatile* adalah Data-data dari berbagai sumber data yang dikumpulkan ke dalam *data warehouse*, tidak boleh mengalami manipulasi data dalam bentuk *edit*, *update* dan *delete*. Dengan kata lain, *data warehouse* mementingkan adanya *historical data* (data asli), guna menunjang proses analisis data yang dilakukan kelak.
- 5) *Process Oriented Data warehouse* dipandang sebagai sebuah proses berkesinambungan di dalam pengolahan data menjadi informasi serta pengiriman informasi tersebut. Proses menjadi orientasi dari *data warehouse* di dalam operasionalnya.
- 6) *Accessible Data warehouse* beserta dengan data-data di dalamnya, harus dapat diakses dengan mudah oleh pengguna. Pengguna dapat memperoleh data yang mana saja yang mereka butuhkan, baik keseluruhan maupun sebagian (*parsial*), sesuai dengan hak akses (*privilege*) yang diartikan oleh sistem atau pemilik sistem.

Menurut Ken Orr (1999), "*Data Warehouse Architecture (DWA)* atau arsitektur *data warehouse* merupakan sebuah cara yang dilakukan oleh arsitek *data warehouse* atau SDM yang berkaitan dengan perancangan dan desain sistem *data warehouse* di dalam sebuah organisasi, untuk mempresentasikannya ke dalam bentuk bagan/gambar/desain, yang memuat segala hal mengenai

struktur data, komunikasi, pemrosesan, dan presentasi dari komputasi *end to end* antar komputer di dalam sistem *data warehouse*” [2].



Gambar 3.1. Data Warehouse Architecture versi Ken Orr

### 3.3 OLAP (*On Line Analytical Processing*)

OLAP merupakan proses komputer yang memungkinkan pengguna dapat dengan mudah dan selektif memilih dan melihat data dari sudut pandang yang berbeda-beda. Data pada OLAP disimpan dalam basis data multidimensi. Jika pada basis data relasional terdiri dari dua dimensi, maka pada basis data multidimensi terdiri dari banyak dimensi yang dapat dipisahkan oleh OLAP menjadi beberapa *sub atribut* [2].

### 3.4 ETL (*Extract, Transform, and Loading*)

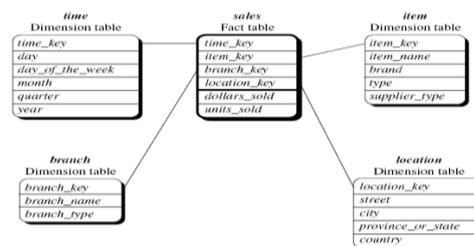
Menurut Ralph Kimball dan Joe Caserta, ETL (*Extract, Transform, and Loading*) merupakan urutan langkah di dalam pemrosesan data pada *database* (khususnya *data warehouse*), yang melibatkan proses pegekstrasian (*extraction*) data-data dari sumber-sumber datanya, mempertahankan kualitas data, menerapkan standarisasi untuk data, menyajikannya ke dalam berbagai bentuk (*transformation*), untuk kemudian di alirkan atau diteruskan (*loading*) ke *data warehouse* untuk digudangkan, dalam rangka kebutuhan untuk analisis data maupun informasi [2]

### 3.5 Dimensional Modeling

Model data yang populer untuk *data warehouse* adalah model multidimensi. Beberapa konsep pemodelan *data warehouse* pada model multidimensi yang dikenal pada umumnya adalah *star schema*, *snowflake* dan *fact constellation schema* [2].

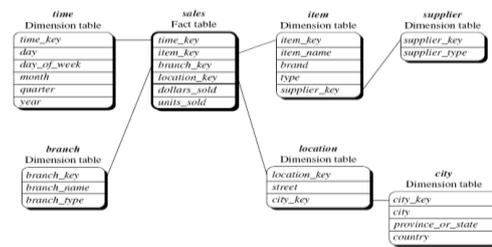
### 3.6 Star Schema (Skema Bintang)

Paradigma pemodelan yang paling umum adalah skema bintang, di mana gudang data berisi : tabel pusat (tabel fakta) yang berisi sebagian besar data tanpa redundansi, dan dimensi tabel, satu untuk setiap dimensi. Grafik skema menyerupai *starburst*, dengan tabel dimensi ditampilkan dalam pola radial di sekitar tabel fakta pusat.



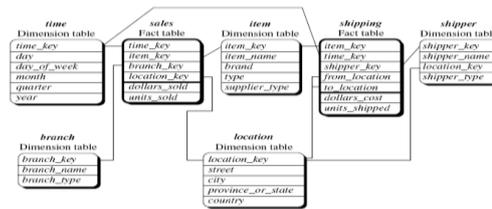
Gambar 3.2. Star Schema

1. *Snowflake Schema* Skema snowflake adalah varian dari model skema bintang, di mana beberapa tabel dimensi dinormalkan, sehingga lebih memecah data menjadi tabel tambahan. Grafik skema yang dihasilkan membentuk bentuk yang mirip dengan kepingan salju.



Gambar 3.3. Snowflake Schema

2. *Fact Constellation Schema* Aplikasi yang canggih mungkin memerlukan beberapa tabel fakta untuk dibagikan tabel dimensi. Skema semacam ini dapat dilihat sebagai kumpulan bintang, dan karenanya disebut skema galaksi atau konstelasi fakta.



Gambar 3.4. Fact Constellation Schema

#### 4 Hasil Pembahasan

Perancangan arsitektur meliputi perancangan arsitektur *logical* dan arsitektur fisik. Arsitektur *logical* berupa rancangan tahapan alur data dari sumber data sampai menjadi data pada *datawarehouse*, sedangkan arsitektur *fisik* berupa gambaran konfigurasi teknis dari data *warehouse* tersebut. Perancangan arsitektur tersebut tentunya sedapat mungkin di desain sesuai dengan kondisi yang ada.

##### 4.1 Arsitektur Logical

Arsitektur *logical* pada perancangan *data warehouse* dapat dilihat pada gambar di bawah ini.



Gambar 4.1. Arsitektur Logical

##### 4.2 Arsitektur Fisik

Arsitektur fisik pada perancangan *data warehouse* dapat dilihat pada gambar di bawah ini.



Gambar 4.2. Arsitektur Fisik

##### 4.3 Sumber Data

Sumber data toko tahun 2013 sampai 2017 memiliki 5 tabel yaitu tabel customer, tabel karyawan, tabel kategori, tabel produk, tabel penjualan header, tabel penjualan detail.

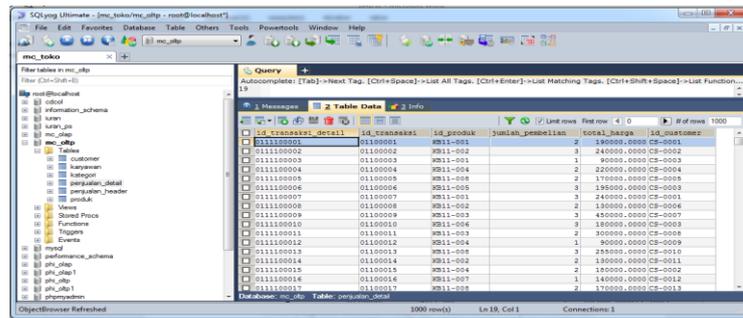
ID_transaksi	ID_customer	Kategori	tanggal_transaksi	produk
01500604	CS-0247		08/08/2015	
01500605	CS-0248		08/08/2015	
01500606	CS-0249		08/08/2015	
01500607	CS-0250		09/08/2015	
01500608	CS-0251		09/08/2015	
01500609	CS-0252		10/08/2015	
01500610	CS-0253		10/08/2015	
01500611	CS-0254		10/08/2015	
01500612	CS-0254		11/08/2015	
01500613	CS-0202		11/08/2015	
01500614	CS-0203		11/08/2015	
01500615	CS-0067		12/08/2015	
01500616	CS-0328		12/08/2015	
01500617	CS-0038		12/08/2015	
01500618	CS-0389		13/08/2015	
01500619	CS-0056		13/08/2015	
01500620	CS-0445		13/08/2015	
01500621	CS-0067		14/08/2015	
01500622	CS-0345		14/08/2015	
01500623	CS-0201		14/08/2015	
01500624	CS-0202		15/08/2015	
01500625	CS-0203		15/08/2015	
01500626	CS-0204		15/08/2015	
01500627	CS-0205		16/08/2015	
01500628	CS-0206		16/08/2015	

Gambar 4.3. Sumber Data Ms Access

4.4 Proses ETL (extract, transform, load) [3,4,5].

Proses ETL (extract, transform, load) adalah proses yang digunakan dalam mengambil, memproses dan menggabungkan data dari berbagai sumber internal dan eksternal sebelum dimasukkan ke dalam suatu data warehouse yang akan dilakukan. Proses ini dilakukan untuk menstandarisasikan data yang akan digunakan dan tahapan ini merupakan proses yang sangat penting dalam pembuatan suatu data warehouse.

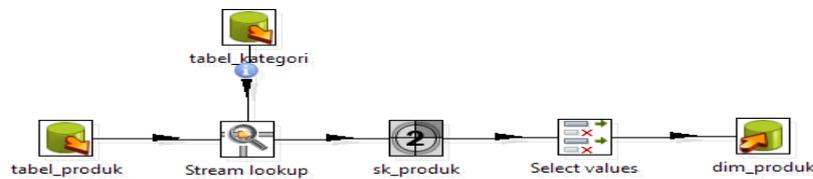
1. Proses Extract adalah proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi proses-proses selanjutnya. Melalui proses pemilihan ini dapat membaca file microsoft access lalu kemudian data yang sudah . dipilih akan disimpan di database MySQL. Untuk lebih jelasnya dapat dilihat pada gambar 4.4



Gambar 4.4. Proses Extract

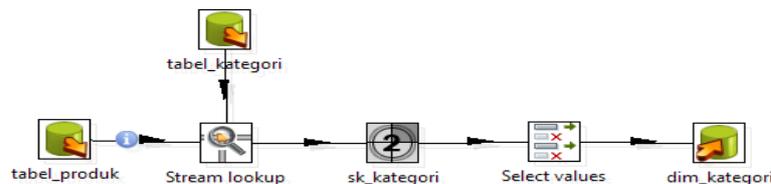
2. Proses Transformation Setelah melewati proses extract dan cleansing, proses selanjutnya adalah melakukan transformasi dari sumber database ke dalam bentuk tabel dimensi dan fact Tabel yang akan membentuk star schema.

Transformasi Dimensi *Produk*



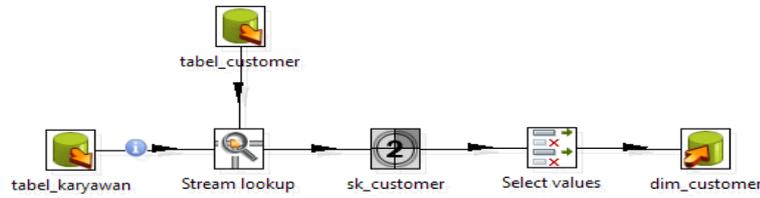
Gambar 4.5. Dimensi Produk

Transformasi Dimensi *Kategori*



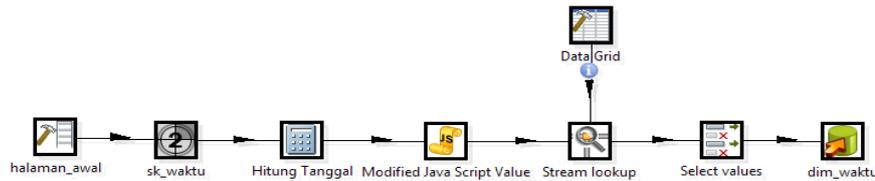
Gambar 4.6. Dimensi Kategori

- Transformasi Dimensi *Customer*



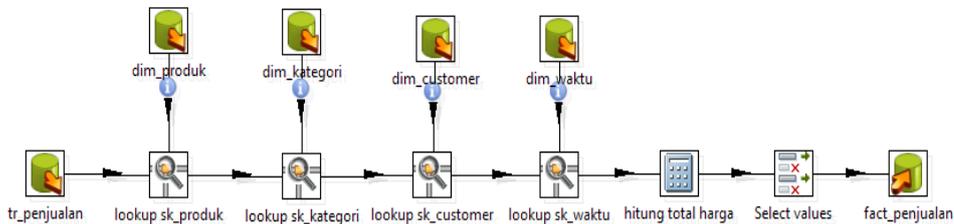
Gambar 4.7. Dimensi Customer

- Transformasi Dimensi *Waktu*



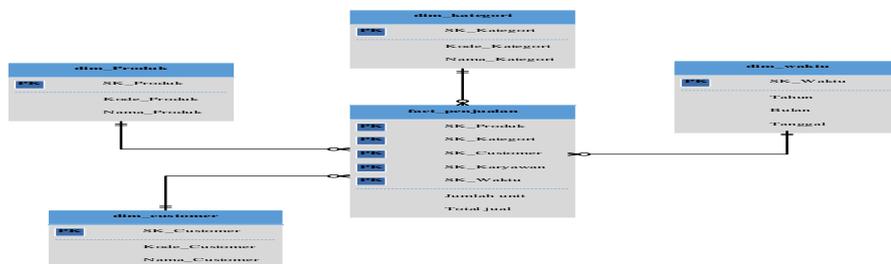
Gambar 4.8. Dimensi Waktu

3. Proses Loading Proses yang dilakukan pada tahap terakhir ini adalah proses pemuatan data (*loading*). Pada tahap ini secara garis besar data telah diseragamkan ke dalam format data yang didapatkan dari hasil transformasi ke dalam data warehouse untuk dapat diteruskan ke tatap muka aplikasi dan layanan, dimana pengguna dapat mengakses keluaran dari tahapan *loading* ini dalam bentuk data maupun informasi. Penyajian dapat dilakukan melalui laporan maupun sekumpulan data untuk kebutuhan analisis dan pengambilan keputusan. Proses *loading* dapat di lihat pada gambar. di bawah ini. [3,4,5]



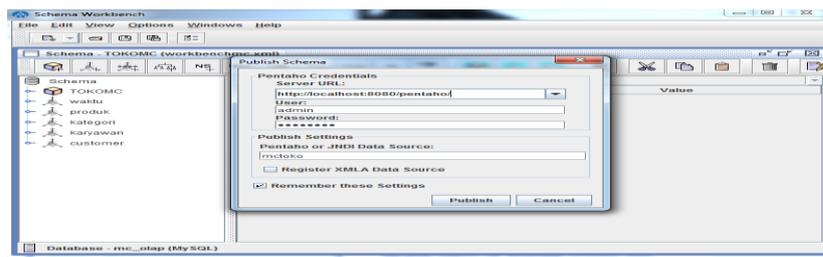
Gambar 4.9. Fact Tabel Penjualan

4.5 Skema Bintang (*Star Schema*) Berdasarkan hasil transformasi tabel-tabel dimensi yang dibuat pada perancangan data warehouse, maka dibentuk skema bintang sebagaimana terlihat pada gambar berikut:



Gambar 4.10. Star Schema

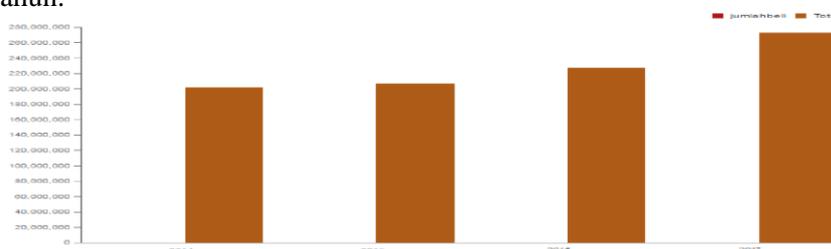
4.6 Cube Olap dengan Menggunakan Pentaho Schema Workbench. merupakan Online Analytical Processing (OLAP) engine yang menggunakan bahasa pemrograman Java. [3,4,5].



Gambar 4.11. Pentaho Workbench Olap

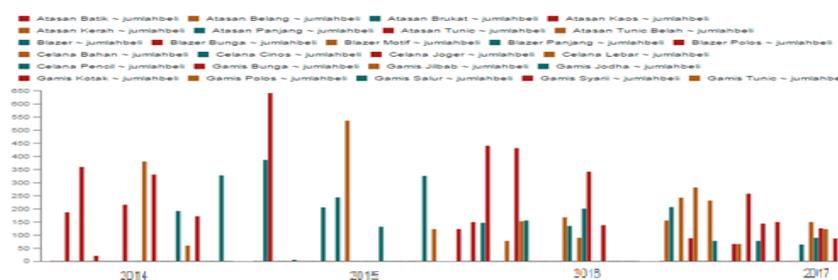
4.7 Dashboard dengan Menggunakan Pentaho BI Server /Pentaho Business Analytics merupakan engine yang menggunakan bahasa pemrograman Java adapun hasil informasi dalam bentuk charts dan grafik sehingga akan mudah di pahami oleh orang yang melihat informasi tersebut. [6].

- a. Informasi Data Penjualan dalam bentuk grafik batang (dashboard) dapat dilihat pada gambar dibawah ini yang menunjukkan peningkatan dalam total penjualan pada toko magnifique clothes selama 4 tahun.



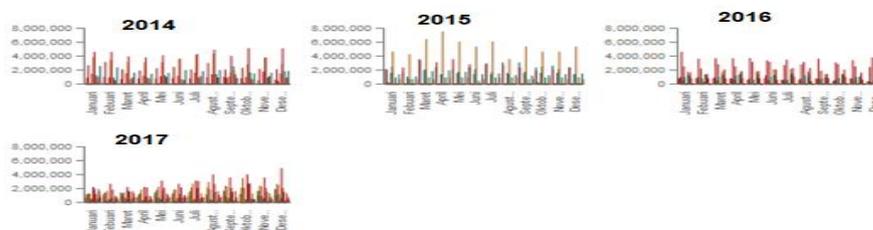
Gambar 4.12 Informasi Penjualan Tahun 2014 sd 2017

- b. Informasi yang yang banyak terjual dalam bentuk grafik batang (dashboard) dapat dilihat pada gambar dibawah ini yang menunjukkan bahwa barang yang terjual pada toko magnifique clothes selama 4 Tahun



Gambar 4.13 Informasi Barang yang Terjual Tahun 2014 sd 2017

- c. Informasi dalam bentuk grafik batang (dashboard) dapat dilihat pada gambar dibawah ini yang menunjukkan keuntungan setiap bulannya pada penjualan di toko magnifique clothes selama 4 tahun.



Gambar 4.13 Informasi Keuntungan / bulan selama tahun 2014 sd 2017

## 5 Kesimpulan dan Saran

Berdasarkan hasil penelitian yang dilakukan, maka dapat disimpulkan dan saran sebagai berikut :

- a. Dengan dibangunnya data wareouse penjualan , maka penyampaian informasi yang terkait dapat dilakukan dengan mudah dan lebih fleksibel.

- b. Data warehouse dapat digunakan oleh pemilik toko dalam melihat perkembangan keuntungan dan penjualan yang terjadi setiap minggu, bulan dan tahun, sehingga pemilik dapat melakukan analisis terhadap penyampaian informasi yang sudah disajikan dalam bentuk grafik atau dashboard.
- c. Informasi yang disajikan dari data warehouse penjualan dapat dipergunakan untuk membantu pemilik toko dalam proses evaluasi, perencanaan dan pemasaran barang yang laku terjual. Juga membantu toko dalam mencermati trend barang yang diminati konsumen setiap bulan atau tahunnya
- d. Data warehouse ini dapat dikembangkan dengan penambahan data-data baru, misal data penjualan 2018 dan seterusnya. Sehingga data warehouse yang ada menjadi tempat penyimpanan dan penyampaian informasi penjualan secara periodik
- e. Data warehouse dan dashboard aplikasi dapat dibangun berkelanjutan untuk pengembangan ke depannya.
- f. Toko Magnifique Clothes agar menambah proses lain selain data penjualan, misalnya data pembelian barang kedistributor ataupun yang lainnya.

### Daftar Pustaka

- [1]. Kimball, R, Margy R, Warren T, Joy M and Bob B 2008. *The Data Warehouse Lifecycle Toolkit*, Wiley Publishing Inc, Canada
- [2]. I. Putu Agus Eka Pratama, *Handbook Data Warehouse, Teori Dan Praktik Berbasis Open Source*". Bandung: Informatika Bandung, 2018
- [3]. Mulyana, JRP.2012. "Pentaho Solusi Open Source Membangun Data Warehouse", Andi
- [4]. Roland Bouman, Jas Van Dongen , 2009. Pentaho Solutions, Business Intelligence and Data Warehousing with Pentaho and MySQL, Wiley Publishing, Inc., Indianapolis, Indiana
- [5]. Caster C, Bouman R, Dongen JV. 2010. Pentaho Kettle Solutions: Building Open Source ETL Solutions With Pentaho Data Integration. Indianapolis (IN): Wiley.
- [6]. Diana TARNAVEANU Pentaho Business Analytics: a Business Intelligence Open Source Alternative Database Systems Journal vol. III, no. 3/2012