p-ISSN: 2301-5373
e-ISSN: 2654-5101

Implementasi Data Warehouse Dan Penerapannya Pada PHI-Minimart Dengan Menggunakan Tools Pentaho dan Power BI

I Kadek Anom Sukawirasa [1)], I Gede Acintia Udayana[2)], I Made Yoga Mahendra[3)], Gde Deva Dimastawan Saputra[4)], Ida Bagus Made Mahendra[5)]

[a]Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Udayana
Jl. Raya Kampus Unud, Indonesia
sukawirasaanom@gmail.com, acintiaudayana74@gmail.com,
yoga.mahendra2000@gmail.com, devadimastawan@gmail.com,
ibm.mahendra@unud.ac.id

Abstract

A shop owner needs an information system to see sales conditions. However, PHI Minimart only provides sales information that is stored in a database so that it cannot provide or present information quickly. To solve this problem, a data warehouse was built at PHI Mart to get fast information. This research will carry out the implementation of data warehouse sales from data sources, the extraction, transformation, loading (ETL) process using Pentaho tools, making Starschema in the form of branch dimensions, product dimensions, employee dimensions, time dimensions connected to the sales fact table, dashboard using BI power. Then the results of the data warehouse are analyzed through the OLAP (On-line Analytical Processing) process, making a cube or schema workbench, and making a dashboard for visualization in presenting the expected information from PHI Minimart. The results in this study include sales data used in 2008 in the form of sales graphs or dashboards, goods sold, and total sales of each branch, which are obtained as the delivery of sales information at the store.

Keywords: Data Warehouse, ETL, Pentaho, OLAP, Pentaho, Power BI

Abstrak

Seorang pemilik toko membutuhkan sebuah sistem informasi untuk melihat kondisi penjualan Namun pada PHI Minimart hanya menyediakan informasi penjualan yang disimpan dalam database sehingga tidak dapat menyediakan atau menyajikan informasi dengan cepat. Untuk mengatasi permasalahan tersebut dibangun sebuah *data warehouse* di PHI Mart 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 cabang, dimensi produk, dimensi karyawan, dimensi waktu yang terhubung dengan *fact table* penjualan, *dashboard* menggunakan power BI. 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 PHI Minimart. Hasil dalam penelitian ini mencangkup data penjualan yang digunakan tahun 2008 berupa tampilan grafik atau *dashboard* penjualan , barang yang laku terjual, dan total penjualan tiap cabang, yang didapat sebagai penyampaian informasi penjualan pada toko tersebut.

Kata Kunci : *Data Warehouse*, ETL, Pentaho, OLAP, Pentaho, Power BI

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 (PHI Minimart) 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 tesebut 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 PHI Minimart. .

2. METODOLOGI PENELITIAN

2.1 Metode Pengumpulan Data

Metodologi perancangan *data Warehouse* yang digunakan dalam penelitian ini menggunakan data PHI-Minimart versi 0.1 yang bersumber dari *(*phi_minimart_mysql_v0.1.tar.gz*).* Dimana data tersebut merupakan data penjualan PHI Mart dari tahun 2018 januari hingga desember.

3. Landasan Teori

3.1 Power BI

Power BI merupakan aplikasi pembuatan *Business Intelligence* yang dimiliki oleh *Microsoft* yang mampu menampilkan visualisasi data, memungkinkan membuat *query*, koneksi data, dan laporan (Ronald, 2008). *Power BI* mudah digunakan karena *user interface* yang dimiliki mirip dengan aplikasi *Microsoft Office*. *Power BI* dapat mengolah data dengan detail dan menampilkannya kedalam bentuk yang lebih interaktif. *Power BI* dapat di akses dari 3 *platform*, yaitu *platform desktop* yang dapat di-*install* di komputer atau laptop, *platform web* yang bisa diakses melalui *web browser*, dan *platform mobile* yang dapat digunakan di *smartphone*.

Power BI mampu membantu menganalisis data dan membantu pengambilan keputusan bisnis harian di perusahaan, sehingga memungkinkan pengguna *Power BI* untuk mengambil data yang berguna dan digunakan untuk memecahkan masalah. *Power BI* memiliki kemampuan analisis mencakup *Predictive analytics*, *Data Visualizations*, *R integration*, dan *Data Analysis Expression*. Ada 3 konsep sistem kerja dalam aplikasi *Power BI*, yaitu *Datasets*, *Report,* dan *Dashboard* [1]

3.2 Pentaho Data Integration (Kettle)

Pentaho Data Integration (PDI) atau Kettle adalah utilities ETL open source di bawah Pentaho Corp. Amerika. Kettle terdiri dari 4 aplikasi yang dijalankan melalui shell atau batch script yang berkaitan, yaitu (http://pentaho.phi-integration.com/kettle) [2]:

- Spoon, yaitu aplikasi grafis berbasis swing yang digunakan untuk merancang file skema job dan transformation
- Pan, yaitu script yang digunakan untuk menjalankan file skema transformation melalui terminal / command line

p-ISSN: 2301-5373
e-ISSN: 2654-5101

- Kitchen, yaitu script yang digunakan untuk menjalankan file skema job melalui terminal / command line
- Carte, yaitu temporary web server yang digunakan untuk mengeksekusi job/transformation secara cluster atau parallel.

Saat ini Kettle merupakan utilitas ETL yang sangat popular dengan beberapa fitur sebagai berikut:

1. Memiliki utilitas grafik yang dapat digunakan merancang control flow umum maupun data flow (aliran data).
2. Multi platform - karena dikembangkan di atas Java yang notabene berjalan di banyak platform sistem operasi.
3. Bersifat concurrent, dalam arti row-row data diambil oleh suatu step dan diserahkan ke step lain secara parallel.
4. Scalable - dapat beradaptasi dengan penambahan kapasitas memori RAM atau pun storage (scale up) dan dapat node komputer / cluster (scale out).
5. Koleksi step transformation dan job yang cukup banyak
6. Extensible, kita dapat membuat step transformation dan job baru dengan sistem plugin.
7. Dukungan luas berbagai produk database yang terkenal di pasaran baik itu proprietary maupun free open source seperti Oracle, SQL Server, MySQL, PostgreSQL dan lain sebagainya.

3.3 Data *Warehouse*

Menurut Inmondan Richard D.H (1994) data *Warehouse* adalah koleksi data yang mempunyai sifat berorientasi subjek, terintegrasi, time-variant, dan bersifat tetap dari kumpulan data dalam mendukung proses pengambilan keputusan manajemen.

Menurut Vidette Poe (1997), data *Warehouse* merupakan database yang bersifat analisis dan read only yang digunakan sebagai fondasi dari sistem penunjang keputusan. Data *Warehouse* merupakan database relasional yang didesain lebih kepada *query* dan analisis daripada proses transaksi, biasanya mengandung *history* data dari proses transaksi dan bisa juga data dari sumber lainnya. Data *Warehouse* adalah kumpulan data-data logik yang terpisah dengan database operasional dan merupakan suatu ringkasan. [3]

Karakteristik *data Warehouse* terdiri dari:

1) Berorientasi subyek

   Data Warehouse adalah tempat penyimpanan berdasakan subyek bukan berdasakan aplikasi. Subyek merupakan bagian dari suatu perusahaan. Contoh subyek pada perusahaan manufaktur adalah penjualan, konsumen, inventori, daln lain sebagainya. Gambar dibawah ini merupakan perbedaan mengenai *data Warehouse* dan *database* operasional.

2) Data yang terintegrasi

   Sumber data yang ada dalam *data Warehouse* tidak hanya berasal dari *database* operasional (*internal source*) tetapi juga berasal dari data diluar sistem (*external source*). Data pada sumber berbeda dapat di-*encode* dengan cara yang berbeda.

3) *Nonvolatile*

   Data dalam *database* operasional akan secara berkala atau periodik dipindahkan kedalam *data Warehouse* sesuai dengan jadwal yang sudah ditentukan. Misal perhari, perminggu, perbulan, dan lain sebagainya. Sekali masuk ke dalam *data Warehouse*, data adalah *read- only*.

4) Time-Variant

Sistem operasional mengandung data yang bernilai sekarang sedangkan data dalam *data Warehouse* mengandung data tidak hanya data terkini tetapi juga data *history* yang akan digunakan dalam analisis dan pengambilan keputusan. Waktu adalah dimensi penting yang harus didukung oleh semua *data Warehouse*. Data untuk analisis dari berbagai sumber berisi berbagai nilai waktu, misalkan harian, mingguan, dan bulanan.

5) Ringkas

Jika diperlukan, data operasional dikumpulkan ke dalam ringkasan-ringkasan.

6) Granularity

Pada sistem operasional data dibuat secara *real-time* sehingga untuk mendapatkan informasi langsung dilakukan proses *query*. Pada data *Warehouse* pada menganalisis harus memperhatikan *level-of-detail* misalkan perhari, ringkasan perbulan, ringkasan per- tiga-bulan.

7) Tidak ternormalisasi

Data di dalam sebuah *data Warehouse* biasanya tidak ternormalisasi dan sangat redundan.

3.4 OLAP

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. [4].

3.5 ETL

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 pengekstrasian (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 kebutuan untuk analisis data maupun informasi [5]

3.6 Dimensional Modeling

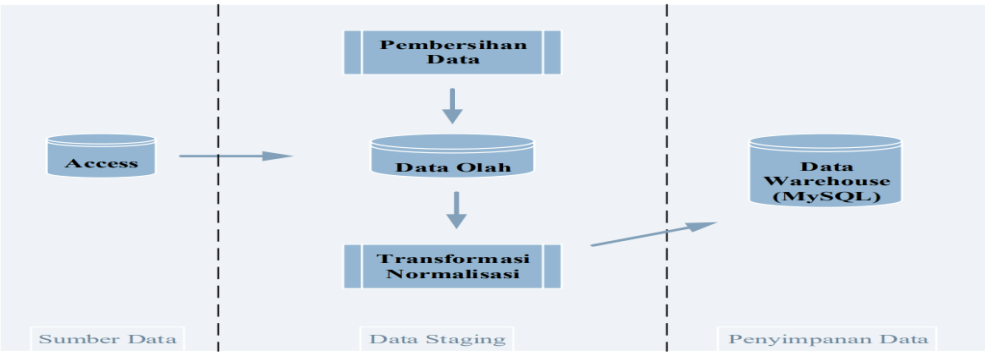
Model data yang populer untuk data warehouse adalah model multidimensi. Beberapa konsep pemodelan data warehouse pada model multidimensi yang dikenal pada umumnya adalah star schema, snowflake dan fact constellation schema [4]

4. HASIL DAN PEMBAHASAN

4.1 Arsitektur *Logical*

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

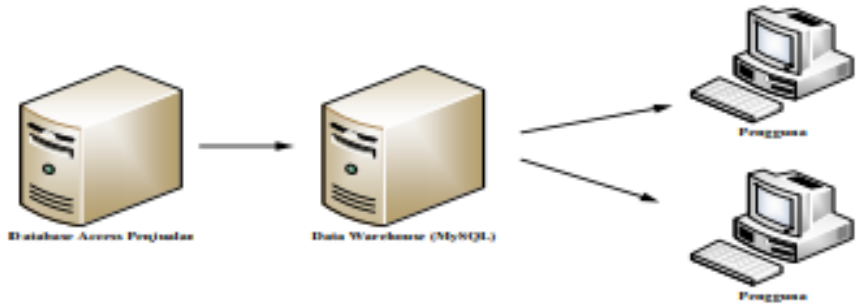
p-ISSN: 2301-5373
e-ISSN: 2654-5101



Gambar 1. *Arsitektur Logical*

4.2 Arsitektur *Fisik*

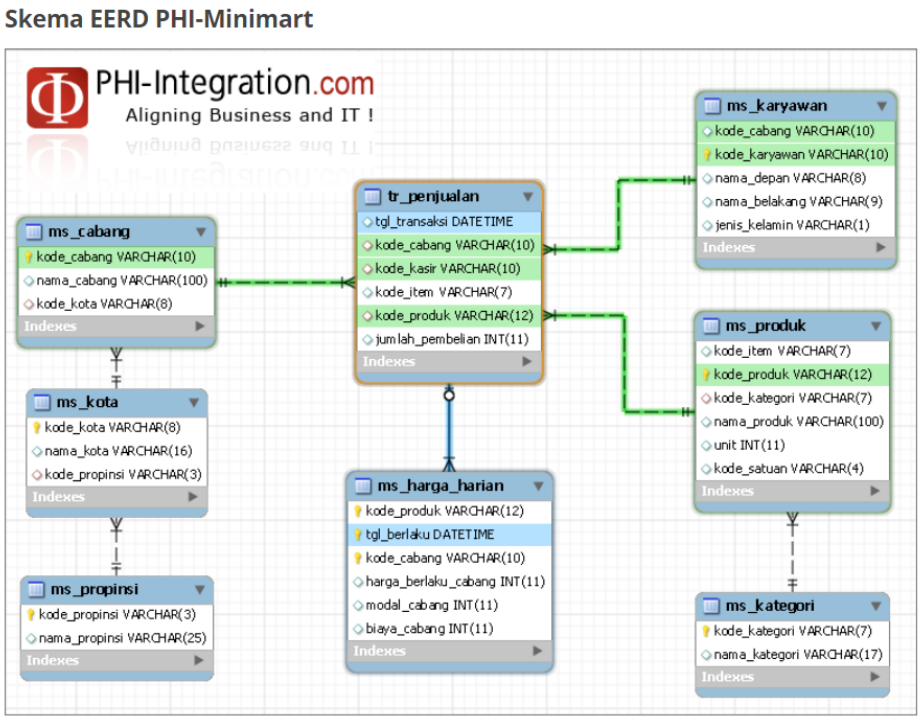
Arsitektur fisik pada perancangan *data warehouse* pada penelilitian ini dapat dilihat pada gambar dibawah ini.



Gambar 2*. Arsitektur Fisik*

4.3 Sumber Data

Sumber data toko yang digunakan pada tahun 2008 yang terdapat 8 tabel yaitu ms_kota, ms_kategori, tr_penjualan, ms_propinsi, ms_harga_harian, ms_karyawan, ms_produk, ms_cabang.

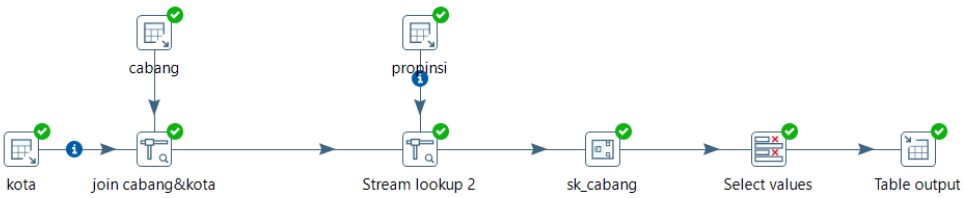


Gambar 3. *Skema ERD PHI-Minimart*

4.4 Proses ETL (*extract, transform, load*)

Proses ETL (*extract, transform, load*) merupakan 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.

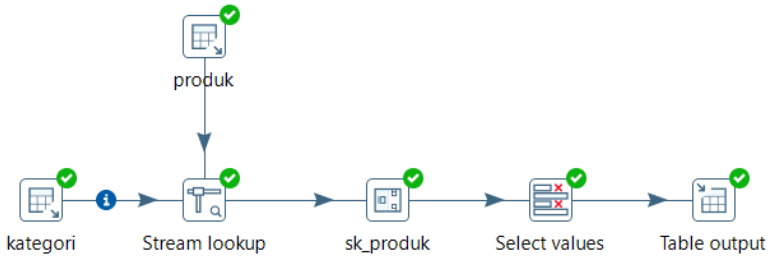
1. Proses Extract adalah proses yang diperlukan untuk terhubung dengan beragam sumber data, dan membuat data tersebut tersedia bagi proses-poses selanjutnya. Melalui proses pemilihan ini dapat membaca file microsoft access lalu kemudian data yang sudah dipilih akan disimpan di database MySql.
2. Proses Tranformation 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 Cabang



Gambar 4*. Transformasi Dimensi Cabang*

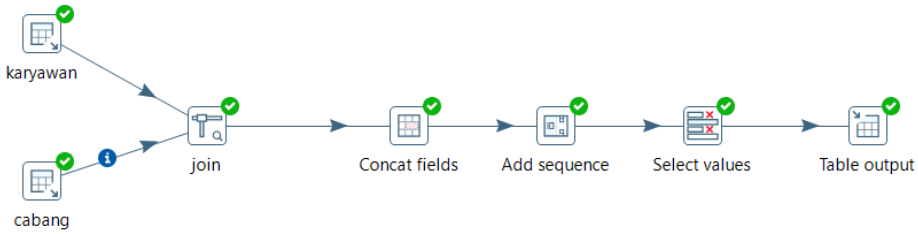
- Transformasi Dimensi Produk

p-ISSN: 2301-5373
e-ISSN: 2654-5101



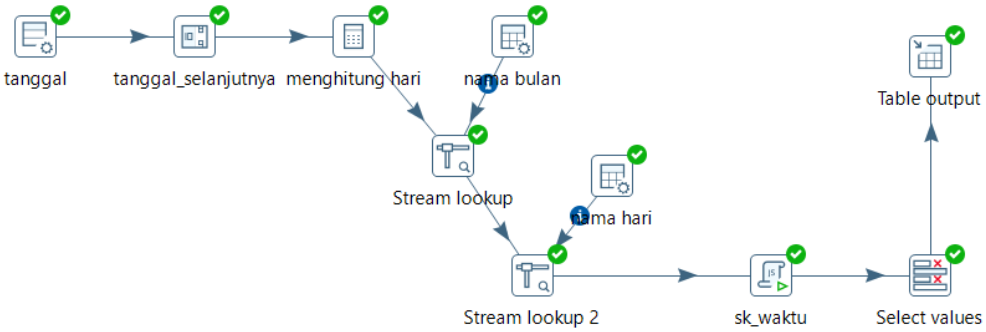
Gambar 5. *Transformasi Dimensi Produk*

- Tranformasi Dimensi Karyawan



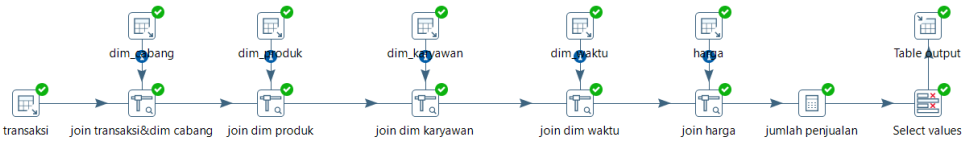
Gambar 6. *Tranformasi Dimensi Karyawan*

- Tranformasi Dimensi Waktu



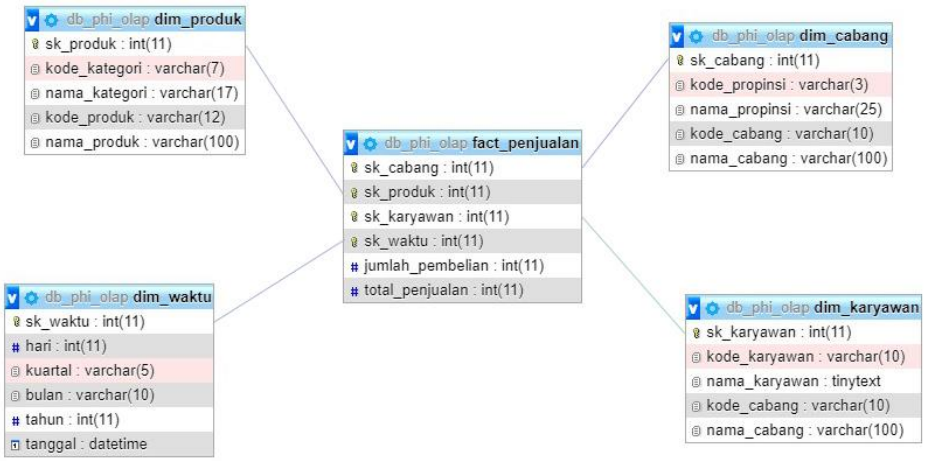
Gambar 7. *Tranformasi Dimensi Waktu*

3. Proses *Loading* merupakan tahap terakhir ini pada 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.



Gambar 8. *Proses Loading*

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:

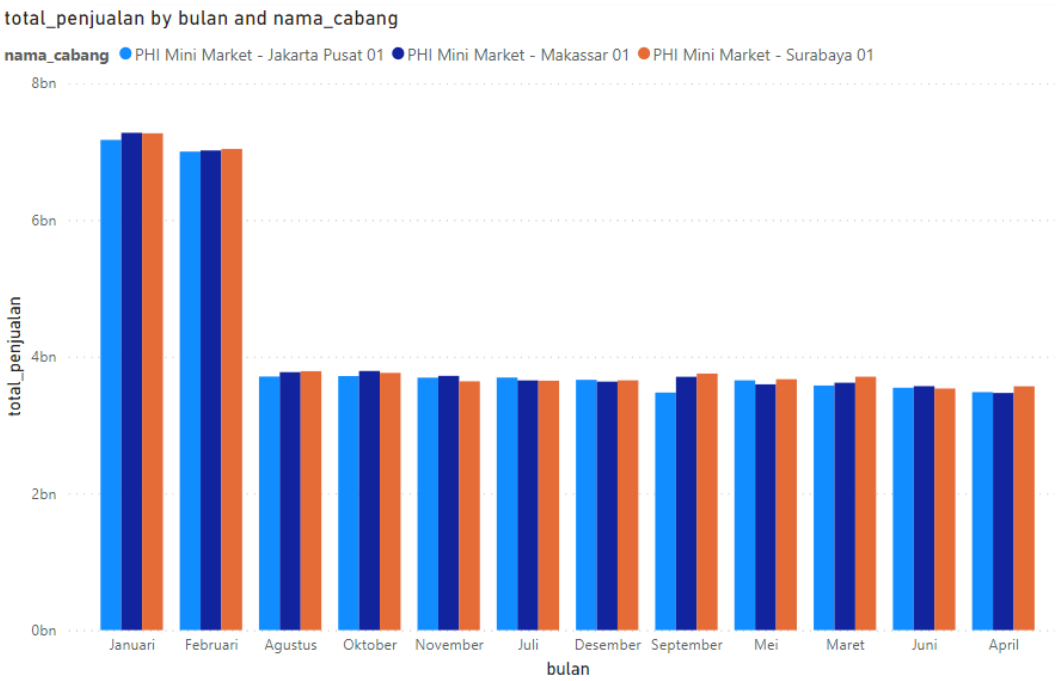


Gambar 9. *Perancangan Datawarehouse*

4.6 Cube Olap dengan Menggunakan Pentaho Schema Workbench. merupakan *Online Analytical Processing* (OLAP) *engine* yang menggunakan bahasa pemograman Java.

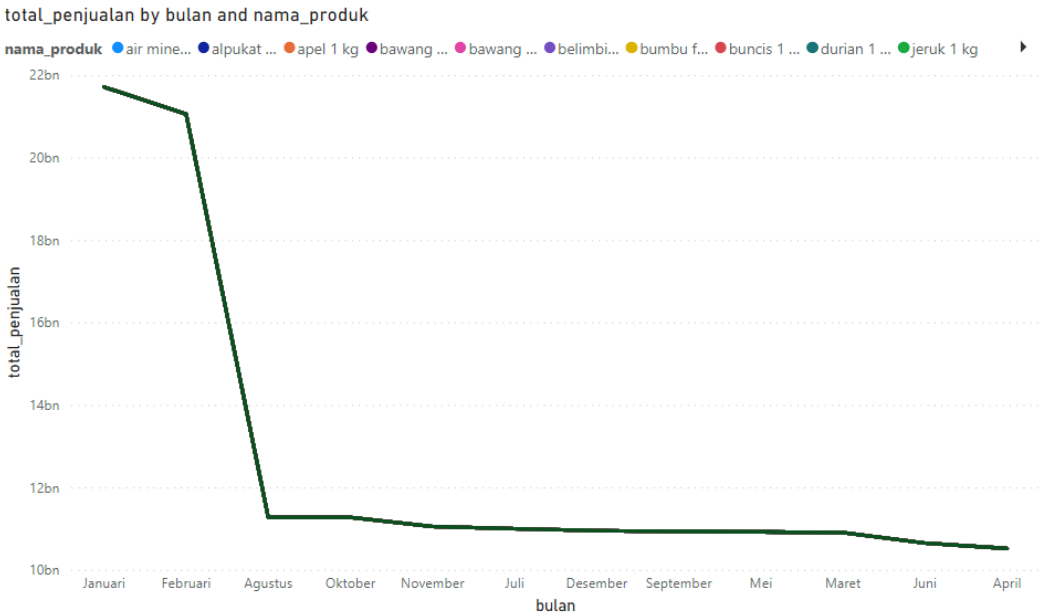
4.7 Dashboard dengan menggunakan Power BI, karena Power BI memiliki kemampuan untuk memproses set data yang besar yang diperlukan untuk Business Intelligence dan analisis data yang besar

p-ISSN: 2301-5373
e-ISSN: 2654-5101



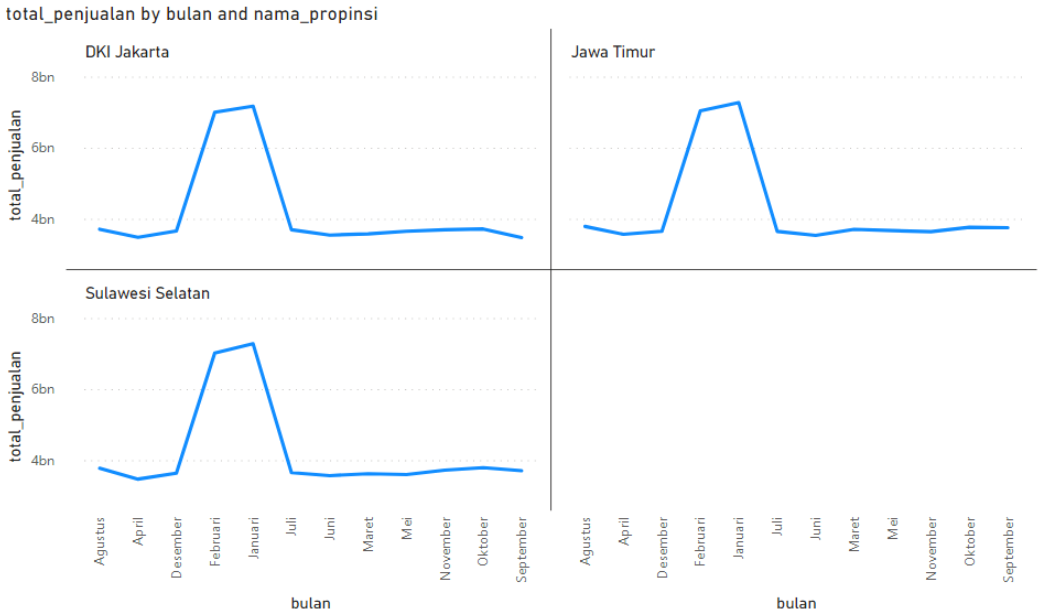
Gambar 10. *Total Penjualan dari bulan dan nama cabang*

Pada gambar diatas terdapat diagram batang yang diurutkan dari total penjualan yang berisi bulan dan nama cabang, Penjualan tertinggi terdapat pada bulan Januari dan Februari dengan penjualan melebihi 6 Juta pada setiap cabangnya.



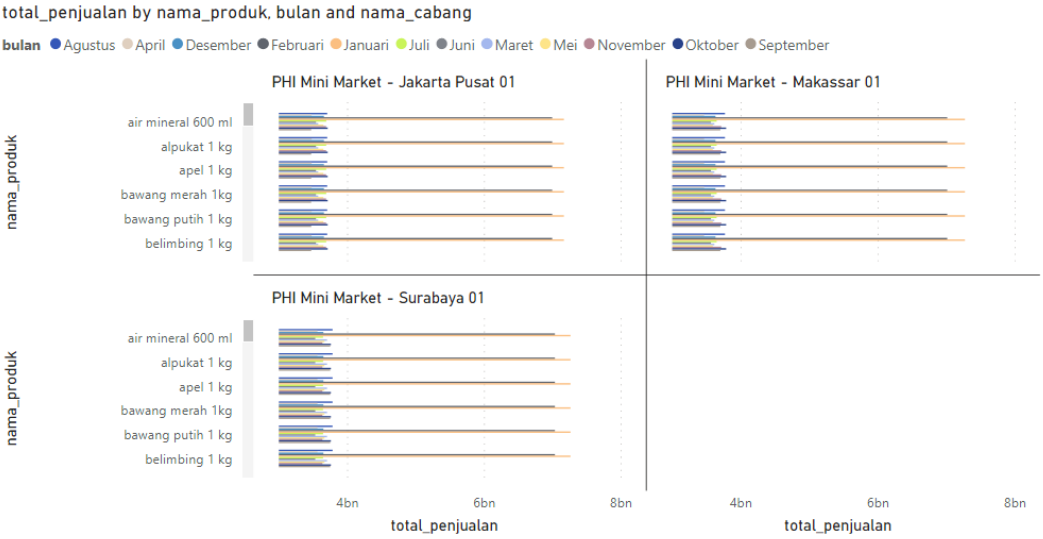
Gambar 11. *Total penjualan dari bulan dan nama produk*

Pada gambar diatas terdapat grafik yang diurutkan dari total penjualan yang berisi bulan dan nama produk, Penjualan tertinggi terdapat pada bulan Januari dan Februari dengan penjualan hampir 22 Juta pada produknya.



Gambar 12. *Total penjualan dari bulan dan nama propinsi*

Pada gambar diatas terdapat grafik yang diurutkan dari total penjualan yang berisi bulan dan nama provinsi, Penjualan tertinggi terdapat pada bulan Januari dan Februari dengan penjualan hampir 7 Juta pada tiap provinsinya.



Gambar 13. *Dashboard dengan menggunakan Power BI*

Pada gambar diatas terdapat dashboard dengan Power BI berisi grafik yang diurutkan dari total penjualan yang berisi nama produk dan nama cabang,

5. Kesimpulan dan Saran

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

a. Dengan dibangunnya data warehouse penjualan, maka penyampaian informasi yang terkait dapat dilakukan dengan mudah dan lebih fleksibel.
b. Data warehouse dapat digunakan oleh pemilik toko dalam melihat

p-ISSN: 2301-5373
e-ISSN: 2654-5101

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 2008 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. PHI-Minimart agar menambah proses lain selain data penjualan, misalnya data pembelian barang kedistributor ataupun yang lainnya.

Daftar Pustaka

[1] M. Silvana, R. Akbar, and Derisma, “Pengembangan Model Business Intelligence Manajemen Rumah Sakit untuk Peningkatan Mutu Pelayanan,” *J. Chem. Inf. Model.*, vol. 4, no. 1, pp. 125–129, 2017.

[2] R. Y. Astuty, “Analisis Data Cube Menggunakan Multiway Array Aggregation For Full Cube Computation,” UNIVERSITAS SANATA DHARMA YOGYAKARTA, 2012.

[3] A. Andri and B. Tujni, “Perancangan Data Warehouse Perpustakaan,” *J. Ilm. Matrik*, no. 3, pp. 101–110, 2017.

[4] S. Darmawan and Furkon, “PERANCANGAN DATAWAREHOUSE DENGAN MENGGUNAKAN TOOLS PENTAHO DAN TABLEAU PADA DATA LAYANAN ANTAR JEMPUT IZIN BERMOTOR (AJIB) DI DINAS PM DAN PTSP PROVINSI DKI JAKARTA,” *Semin. Nas. Teknol. Inf. dan Komun.*, vol. Vol 1 No 1, pp. 573–586, 2018.

[5] D. Subuh and W. Yasman, “Implementasi Data Warehouse Dan Penerapannya Pada Toko Magnifique Clothes Dengan Menggunakan Tools Pentaho,” *Pros. SENIATI*, pp. 29–36, 2019.

This page is intentionally left blank