

Implementasi Sistem *Data Warehouse* dan ETL (Extract, Transform, Load)

Bisnis Perumahan PT. Hidup Bumi Sejahtera

Thomas Januardy

Program Studi Sistem Informasi, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara,
Jl. Scientia Boulevard, Gading Serpong, Tangerang, Banten - 15811 Indonesia

thomas.januardy@student.umn.ac.id

Abstract—Retail is an activity that markets the end consumer to products or services for their own personal or household

Index Terms—*Data Warehouse, ETL, Pentaho, Mondrian, MySQL, Data Perumahan, Business Intelligence*

I. PENDAHULUAN

Perkembangan teknologi telah beradaptasi dan terus mengikuti perkembangan zaman, yang semakin pesat tersebar luas dan mempengaruhi seluruh aspek kehidupan manusia sehari-hari yang tidak luput dari istilah bernama ‘data’ [1]. Kebutuhan akan pengelolaan data yang efektif dan efisien dalam menjalankan aktivitas bisnis agar dapat menjadi informasi yang bernilai sebagai alat pendukung dalam pengambilan keputusan strategis dan solutif terhadap masalah perusahaan. Permasalahan yang muncul dalam bisnis tentang bagaimana cara pengelolaan data tersebut, khususnya hasil data yang berukuran besar yang tidak teratur dan tersebar di mana-mana sehingga menyulitkan perusahaan dalam mengelola maupun menganalisis untuk informasinya [2].

Adanya sistem informasi pada bisnis, dapat membantu meningkatkan produktivitas dan efisiensi dengan memungkinkan akses yang cepat dan mudah ke informasi yang diperlukan, serta dengan membantu mengelola dan mengolah informasi dengan cara yang teratur dan terstruktur. Sistem informasi bisnis juga dapat membantu meningkatkan keamanan informasi dengan menyimpan dan mengelola informasi dengan cara yang aman. Sistem informasi pada bisnis yang mencakup hampir seluruh aspek, dapat diterapkan dalam berbagai jenis bidang bisnis, termasuk perusahaan manufaktur, perusahaan jasa, perusahaan keuangan, dan perusahaan retail. Fasilitas bisnis yang dapat mencakup sistem informasi yang berfokus pada operasi bisnis seperti sistem pembelian, sistem penjualan, sistem produksi, dan sistem pengiriman, atau sistem yang lebih luas seperti sistem sumber daya manusia, sistem keuangan, dan sistem informasi pemasaran.

Implementasi nyata sebagai solusi permasalahan tersebut dengan integrasi *data warehouse* untuk membantu sebuah bisnis dalam mengambil keputusan manajemen [3]. *Data warehouse* sebagai alat penunjang dengan kemampuan menganalisis, merencanakan, dan bereaksi terhadap dinamika bisnis dengan cepat dan akurat [2]. Bagian pengaplikasian *data warehouse* dalam *business intelligence* menjadi salah satu yang terpenting merupakan proses ETL (*Extract, Transform, Load*) yang memiliki performa baik [4]. Proses dalam *data warehouse* dapat dilakukan pada Pentaho, sebagai perangkat lunak yang digunakan untuk ETL dan proses *data warehouse* lainnya [5].

Salah satunya merupakan bisnis perumahan dari perusahaan PT. Hidup Bumi Sejahtera, yang merupakan salah satu perusahaan yang bergerak pada bidang properti, berbentuk residensial berbangunan gedung terstruktur maupun rumah-rumah kontemporer. Perusahaan Hidup Bumi Sejahtera yang memiliki peranan dalam membantu pembangunan infrastruktur sebagai pembangunan arsitektur bangunan di kota-kota besar berdiri sejak tahun 1958, dinaungi dengan departemen pemerintahan, seperti Department of Energy untuk kebutuhan infrastruktur sesuai dengan tatanan lingkungan dan sumber daya bumi. Hidup Bumi Sejahtera telah melakukan implementasi sistem ERP dan pengolahan data untuk kegiatan analisis dan pengambilan keputusan yang lebih baik, yakni dengan penerapan *data warehouse*. Pada dasarnya, Hidup Bumi Sejahtera memiliki aset properti yang luas tersebar dalam beberapa wilayah sehingga terlalu sulit untuk menjaga dan mengevaluasi antar aset properti secara real-time sehingga membutuhkan suatu sistem data yang dapat langsung membawa identitas dari masing-masing properti. Data tersebut pada nantinya berguna dalam analisis, seperti pemasaran, prediksi harga, dan lain sebagainya.

Penelitian ini bertujuan untuk menyatukan dan menyimpan data perusahaan Hidup Bumi Sejahtera di sistem integrasi *data warehouse*, mulai dari data setiap bangunan dan penggunaan energinya untuk memudahkan dalam pencarian data, melakukan analisis, hingga merancang basis data bisnis.

II. TINJAUAN TEORITIS

A. Data Warehouse

Data warehouse adalah sistem penyimpanan dan pengolahan data yang digunakan untuk menyimpan dan mengelola data yang terkumpul dari berbagai sumber, seperti data transaksi, data aplikasi, dan *file-file* lainnya. *Data warehouse* biasanya digunakan untuk menyediakan data yang dibutuhkan oleh perusahaan untuk tujuan analisis dan pengambilan keputusan bisnis. *Data warehouse* terdiri dari koleksi data bersifat *subject-oriented*, *integrated*, *non-volatile*, dan *time-variant* [3]. Selain itu, *data warehouse* memiliki karakteristik ringkas, *granularity* yang rinci, dan tidak ternormalisasi serta sangat redundan [6].

B. Pentaho

Pentaho merupakan perangkat lunak yang berorientasi pada pekerjaan *data warehouse* (DW) dan *business intelligence* (BI). Pentaho bersifat *open-source* dan berjalan pada dengan basis Java [5]. Pentaho terdiri atas beberapa fitur program yang memiliki kegunaan dan fungsinya masing-masing dalam mengolah keperluan di bidang *data warehouse* dan menyediakan *business intelligence*. Keseluruhan program dan fitur dalam membangun aplikasi BI dan DW dirangkum dalam Pentaho BI Suite [7]. Berikut adalah beberapa aplikasi yang disediakan Pentaho [5]:

- Pentaho BI Server, berupa aplikasi dengan portal *web* yang menyediakan layanan *web service*, *workflow engine*, dan sebagai *user interface* dalam membuat laporan analisis maupun operasional di Pentaho.
- Pentaho Data Integration/Kettle, berupa aplikasi dalam melakukan proses ETL dengan elemen utama *Transformation* (sekumpulan instruksi dengan struktur input-proses-output) dan *Job* (sekumpulan instruksi untuk menjalankan transformasi).
- Pentaho Analysis/Mondrian, sebagai aplikasi pendukung sistem OLAP (*Online Analytical Processing*) dengan basis *open-source*.
- Pentaho Reporting, sebagai aplikasi *reporting*/pembuat laporan *ad hoc* yang berorientasi pada laporan operasional dan *dashboard* sederhana.
- Weka, adalah aplikasi untuk tujuan *data mining* secara *open-source* dengan basis Java yang memiliki fitur pemodelan *machine learning* yang dapat digunakan untuk melakukan generalisasi maupun pemodelan formula dari *data sampling*.

- Ctools, sebagai alat yang digunakan untuk membuat *dashboard* utama berbasis *open-source* pada Pentaho Community Edition. *Dashboard* yang interaktif dan terintegrasi pada Pentaho BI Server.

C. Online Analytical Processing (OLAP)

OLAP (*Online Analytical Processing*) adalah teknik dan teknologi yang digunakan untuk mengolah dan menganalisis data multidimensional [6]. OLAP memungkinkan pengguna untuk menjalankan kueri dan analisis data secara interaktif dari berbagai sudut pandang dan menampilkan hasilnya dalam bentuk yang mudah dipahami.

OLAP merupakan bagian dari kategori sistem *Business Intelligence* (BI), yang bertujuan untuk membantu perusahaan dalam mengambil keputusan bisnis dengan menyediakan akses ke data yang terstruktur dan mudah dipahami. OLAP memungkinkan pengguna untuk mengeksplorasi data secara dinamis, menjalankan analisis yang kompleks, dan menampilkan hasilnya dalam bentuk yang mudah dipahami seperti tabel, grafik, atau laporan.

Dengan menggunakan OLAP, perusahaan-perusahaan yang memiliki data yang kompleks dan multidimensional dan membutuhkan solusi untuk mengelola dan menganalisis data tersebut secara efektif yang dapat terbagi ke dalam beberapa sub atribut [6].

D. Extract, Transform, Load (ETL)

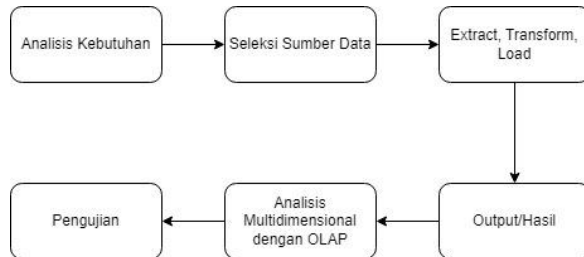
ETL (*Extract, Transform, and Load*) adalah proses yang digunakan untuk mengambil data dari sumber yang berbeda, mengubahnya sesuai dengan kebutuhan, dan memuatnya ke dalam *data warehouse*. ETL merupakan bagian penting dari sistem *Business Intelligence* (BI) yang bertujuan untuk menyediakan akses ke data yang terstruktur dan mudah dipahami bagi para pengguna untuk mengambil keputusan bisnis. Secara sederhana, maka alur ETL diawali dengan mengekstrak dan mempertahankan kualitas data dari sumber yang terseleksi, standarisasi berupa mengubahnya menjadi format baru yang terstruktur dan sesuai dengan aturan model bentuk, dan diakhiri dengan pemuatan ke dalam struktur data target [2].

III. METODOLOGI

Metode penelitian adalah langkah yang dilakukan dalam rangka untuk mengumpulkan informasi atau data, serta melakukan evaluasi pada data yang telah didapatkan tersebut. Metode penelitian memberikan suatu gambaran rancangan penelitian yang meliputi prosedur penelitian dan langkah-langkah yang harus dikerjakan, waktu penelitian, sumber data, dan dengan langkah apa data-data tersebut diperoleh dan tahap selanjutnya diolah dan dianalisis untuk mendapatkan suatu kesimpulan.

A. Alur Penelitian

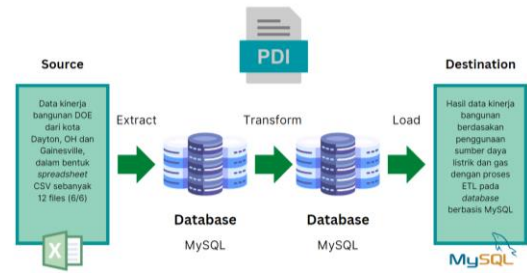
Proses dalam penelitian ini akan dijalani berdasarkan alur metodologi rekayasa dengan tahapan penelitian sedemikian rupa untuk mendapatkan hasil yang terbaik dari pengujian *data warehouse*, seperti yang ditunjukkan pada gambar di bawah ini.



Pada penelitian ini, tahap pertama pada analisis kebutuhan tentang apa yang dibutuhkan oleh bisnis dengan teknik pengumpulan data berupa laporan dan langkah strategis menggunakan sistem *data warehouse*. Analisis yang dibutuhkan perusahaan Hidup Bumi Sejahtera tentang seberapa banyak data mengenai karakteristik properti *building* beserta penggunaan energi, seperti listrik dan gasnya. Proses pengambilan sumber data adalah dengan proses ekstraksi yang dipergunakan berdasarkan laporan data dari “*DOE’s Buildings Performance Database, sample Residential Data*” dengan data bangunan properti pada kota Dayton, OH, dan Gainesville. Proses ETL yang dilakukan dengan mengubah dan memanipulasi data berdasarkan kesesuaian dengan laporan kinerja analisis data properti. Proses pengolahan output dengan proses penyuntingan dan penyesuaian lebih lanjut agar data dapat dipahami dengan mudah melalui output pada database MySQL untuk proses multidimensional. Proses analisis multidimensional yang dilakukan dengan Mondrian untuk membuat skema bintang (*star schema*), grafik dan penjabaran *data cube* lebih rinci yang digunakan untuk proses OLAP. Proses pengujian untuk menguji kesesuaian hasil yang dilakukan dengan cara mengambil sampel dari hasil output tabel fakta secara acak dan disesuaikan dengan hasil antara pemrosesan data melalui pemrosesan ETL pada Pentaho Data Integration.

B. Gambaran Umum Sistem

Gambaran umum sistem yang merupakan arsitektur dan kerangka pikir dari proses tahapan penelitian ini, yang terdiri atas analisis akan kebutuhan sistem dan bagaimana proses pengolahan data dapat menjadi analisis yang solutif. Gambaran tersebut akan dirangkum dalam bentuk ilustrasi kerangka gambar di bawah ini.



C. Analisis Kebutuhan Sistem

1) Format Data Properti

Format data dari properti yang dimiliki perusahaan Hidup Bumi Sejahtera diambil dari *DOE’s Building Performance Database* di wilayah kota Dayton, OH, dan Gainesville. Data berbentuk *file* dengan format CSV yang terdiri dari 6 *file*, masing-masing menjelaskan mengenai:

• Data Headers (metadata)

	A	B	C	D	E
1	FACILITYTYPEID				
2	RESFACIL	RESFACILITYTYPEID			
3	Single Fam	1			
4	Multifam	2			
5	Apartmen	3			
6	Mobile Hc	4			
7	Studio	5			
8	Unknown	0			
9	NULL	999			
10					
11	CENSUSDIVISIONID				
12	CENSUSDI	CENSUSDIVISIONID			
13	New Engli	1			
14	Middle At	2			
15	East Nort	3			
16	West Nort	4			
17	South Atl	5			
18	East Sou	6			
19	West Sou	7			
20	Mountain	8			
21	Pacific	9			
22	Unknown	0			
23	NULL	999			

• Building Characteristics

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	DOEAMT	FACILITYTYPEID	CENSUSDI	CENSUSDIVISIONID	GROSSFLR	FLOOR	NUMFLR	LOC	WALLTYPE	WINDOW	ATTACHED	WALLNET	ZIPCODE	CLIMATEZ	ASSET	HE	HEATING	HEATINGNET	ASSET	Cooling
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
4	3	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	4	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
7	6	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
8	7	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
9	8	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
10	9	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
12	11	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
14	13	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
15	14	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
16	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
17	16	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
18	17	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
19	18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
20	19	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
21	20	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
22	21	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
23	22	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
25	24	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
26	25	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
27	26	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
28	27	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

• Electricity Annual

	A	B	C	D	E
1	DataJamit	YEAR	AnnualElec_kWh		
2	1	2010	19703		
3	2	2010	19004		
4	3	2010	11018		
5	4	2010	9901		
6	5	2010	6614		
7	7	2010	5944		
8	8	2010	7845		
9	10	2010	6777		
10	12	2010	21842		
11	13	2010	27850		
12	14	2010	20747		
13	15	2010	1804		
14	16	2010	4916		
15	17	2010	13502		
16	18	2010	18845		
17	19	2010	11416		
18	20	2010	23527		
19	22	2010	7318		
20	23	2010	23844		
21	24	2010	5897		
22	25	2010	7738		
23	26	2010	13731		
24	27	2010	10034		

• Electricity Monthly

	A	B	C	D	E	F
1	DataJamID	RECORD_ID	YEAR	MONTH	ELECTRICITY_USE_kWh	
2	1	721649	2010	1	1048	
3	1	721651	2010	3	1069	
4	1	721650	2010	2	1012	
5	1	721658	2010	10	1681	
6	1	721652	2010	4	1025	
7	1	721659	2010	11	1156	
8	1	721653	2010	5	1276	
9	1	721656	2010	8	2669	
10	1	721655	2010	7	2583	
11	1	721654	2010	6	2330	
12	1	721657	2010	9	2512	
13	1	721660	2010	12	1342	
14	2	721672	2010	12	1591	
15	2	721661	2010	1	2518	
16	2	721670	2010	10	1196	
17	2	721665	2010	5	1325	
18	2	721662	2010	2	1390	
19	2	721667	2010	7	2350	
20	2	721669	2010	9	1721	
21	2	721663	2010	3	1421	
22	2	721664	2010	4	987	
23	2	721668	2010	8	1950	

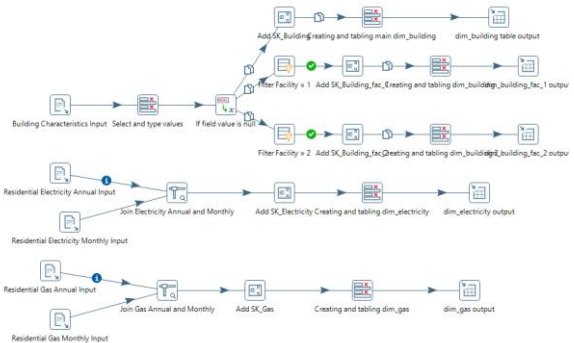
• Gas Annual

	A	B	C	D	E	F	G	H
1	DataJamID	RECORD_ID	YEAR	MONTH	ELECTRICITY_USE_kWh			
2	1	721649	2010	1	1048			
3	1	721651	2010	3	1069			
4	1	721650	2010	2	1012			
5	1	721658	2010	10	1681			
6	1	721652	2010	4	1025			
7	1	721659	2010	11	1156			
8	1	721653	2010	5	1276			
9	1	721656	2010	8	2669			
10	1	721655	2010	7	2583			
11	1	721654	2010	6	2330			
12	1	721657	2010	9	2512			
13	1	721660	2010	12	1342			
14	2	721672	2010	12	1591			
15	2	721661	2010	1	2518			
16	2	721670	2010	10	1196			
17	2	721665	2010	5	1325			
18	2	721662	2010	2	1390			
19	2	721667	2010	7	2350			
20	2	721669	2010	9	1721			
21	2	721663	2010	3	1421			
22	2	721664	2010	4	987			
23	2	721668	2010	8	1950			

• Gas Monthly

	A	B	C	D	E	F
1	DataJamID	RECORD_ID	YEAR	MONTH	Gas_USE_kBtu	
2	6	620782	2010	1	4200	
3	6	620783	2010	2	1200	
4	6	620784	2010	3	1400	
5	6	620785	2010	5	1200	
6	6	620786	2010	6	1800	
7	6	620787	2010	8	1800	
8	6	620788	2010	9	900	
9	6	620789	2010	10	900	
10	6	620790	2010	11	1300	
11	6	620791	2010	12	3700	
12	7	620792	2010	1	10800	
13	7	620793	2010	2	7600	
14	7	620794	2010	3	7100	
15	7	620795	2010	4	1200	
16	7	620796	2010	5	600	
17	7	620797	2010	6	500	
18	7	620798	2010	7	600	
19	7	620799	2010	8	400	
20	7	620800	2010	9	500	
21	7	620801	2010	10	1200	
22	7	620802	2010	11	3700	
23	7	620803	2010	12	11000	

Proses ETL (*Extract, Transform, Load*) pada penelitian di bisnis ini dilakukan dengan integrasi Pentaho, khususnya Pentaho Data Integration/Kettle. Berikut merupakan transformasi dari proses ETL pada data bisnis Hidup Bumi Sejahtera:



Proses input hanya terdiri atas 5 file yang dibutuhkan (tanpa file data headers) yang terdiri dari data *Building Characteristics*, *Electricity Annual and Monthly*, dan *Gas Annual and Monthly*. Dari proses input, khususnya pada laporan *Building Characteristics* untuk menghapus *missing values*, solusi yang dilakukan berupa identifikasi value *null* dengan penyesuaian metadata dengan kode “999” dan “5” pada *field* tertentu yang memiliki deskripsi untuk nilai *null*. Dari proses penghapusan tersebut, akan data karakteristik akan dibagi menjadi 3 macam, yakni tanpa filter, dengan filter fasilitas tipe 1, dan dengan filter fasilitas tipe 2. Pada data input *electricity* dan *gas*, karena terdapat dua bentuk file laporan yang berbeda antara *annual* dan *monthly*, maka akan diintegrasikan dengan penyesuaian *foreign key* menggunakan “DataJamID”. Setelahnya, semua proses akan berjalan dengan sama, yakni penambahan SK dan pembuatan dimensi yang sesuai dengan kebutuhan laporan. Kemudian akan diinput *values* apa saja yang nantinya akan ditampilkan pada output tabel dimensi berupa *database* di MySQL.

2) Format Laporan Properti

Format segmentasi berdasarkan kinerja data dari properti yang dimiliki perusahaan Hidup Bumi Sejahtera sebagai output dari proses ETL di basis data terdiri atas laporan-laporan berikut:

- Laporan Karakteristik Properti Bangunan
- Laporan Karakteristik Properti Bangunan dengan Fasilitas Tipe 1
- Laporan Karakteristik Properti Bangunan dengan Fasilitas Tipe 2
- Laporan Penggunaan Sumber Daya Listrik Bangunan
- Laporan Penggunaan Sumber Daya Gas Bangunan

3) Skema ETL

4) Proses Input Data

Dalam proses input data, sumber data yang digunakan berdasarkan dari penjabaran 5 file data, yakni data *Building Characteristics*, *Electricity Annual and Monthly*, dan *Gas Annual and Monthly*. Pada data *Building Characteristics*, dilakukan penyesuaian tipe data agar sesuai dengan proses transformasi *missing values* menjadi string. Beberapa *field* yang sebelumnya integer diubah menjadi string dengan penyesuaian tertentu. Berikut merupakan tampilan rinci mengenai informasi input data *Building Characteristics*.

Step name: **Building Characteristics Input**

Filename: D:\kuliah umm\sem 5\uts\data warehouse\dataset\bpd-dayton-and-gain\ Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group
1	DatalamID	Integer	#	15	0	\$.	
2	FACILITYTYPEID	String						
3	YEARCOMPLETED	String	#	15	0	\$.	
4	CENSUSDIVISIONID	String						
5	GROSSFLOORAREA	String	#	15	0	\$.	
6	NUMFLOORS	String	#	15	0	\$.	
7	WALLTYPEID	String						
8	WINDOWGLASSType	String						
9	ATTICINSTTHICKNESS	String						
10	WALLINSTHICKNESS	String						
11	ZIPCODE	Integer	#	15	0	\$.	
12	CLIMATEZONEID	String						
13	ASSET_HEATING_FUEL_ID	String	#	15	0	\$.	
14	HEATING_SYS	String	#	15	0	\$.	
15	HEATINGSYST_EFF	String						
16	HEATINGSYST_EFF_FLAG	String						
17	ASSET_COOLING_FUEL_ID	String	#	15	0	\$.	
18	COOLING_SYS	String	#	15	0	\$.	

Step name: **Residential Gas Annual Input**

Filename: D:\kuliah umm\sem 5\uts\data warehouse\dataset\bpd-dayton-and-gain\ Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim ty
1	DatalamID	Integer	#	15	0	\$.		none
2	RECORD_ID	Integer	#	15	0	\$.		none
3	AnnualGas_kBtu	Integer	#	15	0	\$.		none

• Gas Monthly

Step name: **Residential Gas Monthly Input**

Filename: D:\kuliah umm\sem 5\uts\data warehouse\dataset\bpd-dayton-and-gain\ Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	DatalamID	Integer	#	15	0	\$.		none
2	RECORD_ID	Integer	#	15	0	\$.		none
3	YEAR	Integer	#	15	0	\$.		none
4	MONTH	Integer	#	15	0	\$.		none
5	Gas_USE_kBtu	Integer	#	15	0	\$.		none

Sedangkan pada proses input data lainnya, dilakukan tanpa pengubahan dan dijalankan sebagaimana mestinya. Berikut merupakan tampilan rinci mengenai informasi input data lainnya:

• Electricity Annual

CSV file input

Step name: **Residential Electricity Annual Input**

Filename: D:\kuliah umm\sem 5\uts\data warehouse\dataset\bpd-dayton-and-gain\ Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim ty
1	DatalamID	Integer	#	15	0	\$.		none
2	YEAR	Integer	#	15	0	\$.		none
3	AnnualElec_kWh	Integer	#	15	0	\$.		none

• Electricity Monthly

Step name: **Residential Electricity Monthly Input**

Filename: D:\kuliah umm\sem 5\uts\data warehouse\dataset\bpd-dayton-and-gain\ Browse...

Delimiter: , Insert TAB

Enclosure: "

NIO buffer size: 50000

Lazy conversion? ☒

Header row present? ☒

Add filename to result ☐

The row number field name (optional):

Running in parallel? ☐

New line possible in fields? ☐

Format: mixed

File encoding:

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group
1	DatalamID	Integer	#	15	0	\$.	
2	RECORD_ID	Integer	#	15	0	\$.	
3	YEAR	Integer	#	15	0	\$.	
4	MONTH	Integer	#	15	0	\$.	
5	ELECTRICITY_USE_kWh	Integer	#	15	0	\$.	

• Gas Annual

5) Proses Menghilangkan Missing Values

Proses menghilangkan *missing values* yang hanya dilakukan pada transformasi data *Building Characteristics* adalah dengan me-replace semua nilai *missing values* menjadi *null* sesuai yang tertera pada keterangan pada file *Data Headers*, yakni dengan kode “999” dan “5” pada *field* tertentu. Berikut merupakan tampilan rinci mengenai proses *replace* tersebut:

Step name: **If field value is null**

Replace Null for all fields

Replace by value: None

Set empty string? ☐

Mask (Date):

Select fields ☒

Select value type ☐

#	Type	Replace by value	Conversion mask (Date)	Set empty string?

#	Field	Replace by value	Conversion mask (Date)	Set empty st
1	FACILITYTYPEID	999		N
2	YEARCOMPLETED	999		N
3	CENSUSDIVISIONID	999		N
4	GROSSFLOORAREA	0		N
5	NUMFLOORS	999		N
6	WALLTYPEID	999		N
7	WINDOWGLASSType	5		N
8	ATTICINSTTHICKNESS	999		N
9	WALLINSTHICKNESS	999		N
10	CLIMATEZONEID	999		N
11	ASSET_HEATING_FUEL_ID	999		N
12	HEATING_SYS	999		N
13	HEATINGSYST EFF	999		N

6) Proses Filter Data (Facility Type)

Proses ekstrak dilakukan dengan cara filter data sesuai dengan kebutuhan yang dibutuhkan. Salah satu proses yang membutuhkan filter untuk memenuhi tujuan bisnis adalah pada keterangan karakteristik bangunan, yakni pada tipe fasilitas

apa yang dipergunakan, apakah *single family*, *apartment*, *multifamily*, *mobile home*, *studio*, dan lain sebagainya. Tetapi karena ketersediaan data yang hanya memiliki *values* antara 1 dan 2, yang di mana masing-masing merupakan “*single family*” dan “*multifamily*”, maka proses filter hanya terdiri dari dua nilai tersebut. Berikut merupakan gambaran *detail* filter pada data dengan tipe fasilitas 1 yang begitu juga dengan tipe fasilitas 2:

Step name: **Filter Facility = 1**

Send 'true' data to step: Add SK_Building_fac_1

Send 'false' data to step:

The condition:

☐ FACILITYTYPEID = 1 (String)

Tahap selanjutnya dengan penambahan *sequence* guna membuat identifikasi pada setiap dimensi output-nya, yang terdiri dari data *Building Characteristics*, *Electricity Annual and Monthly*, dan *Gas Annual and Monthly*. Berikut merupakan tampilan rinci mengenai salah satu proses penambahan *sequence* pada data karakteristik bangunan dengan fasilitas bertipe 1:

Step name: **Add SK_Building_fac_1**

Name of value: SK_Building_fac_1

Use a database to generate the sequence

Use DB to get sequence? ☐

Connection: Hidup Bumi Sejahtera

Schema name:

Sequence name: SEQ_

Use a transformation counter to generate the sequence

Use counter to calculate sequence? ☒

Counter name (optional):

Start at value: 1

Increment by: 1

Maximum value: 999999999

OK Cancel

7) Proses Transform

Tahap *transform* pada proses transformasi ini dilakukan dengan menyeleksi *field values* apa saja yang akan dipergunakan dan ditampilkan pada output di *database* MySQL nantinya. Pemilihan *values* tersebut yang menjadi kolom sebagai acuan data dari *field* target pada proses *load data warehouse* yang nantinya akan diambil dari output *database* yang terdapat pada koneksi MySQL. Berikut merupakan salah satu tampilan rinci dari data karakteristik bangunan dengan fasilitas bertipe 1 mengenai informasi penyeleksian *values* serta koneksi terhadap *database* agar dapat menjadi wadah output di MySQL:

Step name: **Creating and tabling dim_building**

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	SK_Building_fac_1			
2	DataId			
3	FACILITYTYPEID			
4	YEARCOMPLETED			
5	CENSUSDIVISIONID			
6	GROSSFLOORAREA			
7	NUMFLOORS			
8	WALLTYPEID			
9	WINDOWGLASSTYPE			
10	ATTICINSTHICKNESS			
11	WALLINSTHICKNESS			
12	ZIPCODE			
13	CLIMATEZONEID			
14	ASSET_HEATING_FUEL_ID			
15	HEATING_SYS			
16	HEATINGSYST_EFF			
17	HEATINGSYST_EFF_FLAG			
18	ASSET_COOLING_FUEL_ID			
19	COOLING_SYS			
20	COOLINGSYST_EFF			
21	COOLINGSYST_EFF_FLAG			

Get fields to select Edit Mapping

Include unnormalized fields: ordered by name

Step name: **dim_building_fac_1_output**

Connection: Hidup Bumi Sejahtera

Target schema:

Target table: dim_building_fac_1

Commit size: 1000

Truncate table: ☐

Ignore insert errors: ☐

Specify database fields: ☐

Main options: Database fields

Partition data over tables: ☐

Partitioning field:

Partition data per month: ☒

Partition data per day: ☐

Use batch update for inserts: ☒

Is the name of the table defined in a field? ☐

Field that contains name of table:

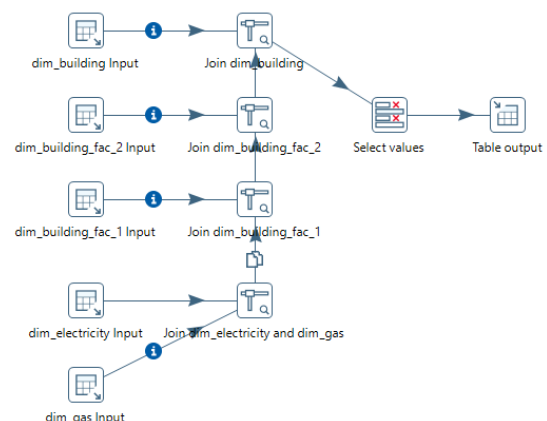
Store the tablename field: ☒

Return auto-generated key: ☐

Name of auto-generated key field:

8) Proses Tabel Fakta

Tahap pembuatan tabel fakta yang merupakan tabel inti yang berisi relasi-relasi dari sekian tabel dimensi ke dalam satu tabel konklusi yang berisi informasi dari sebuah *database*. Pada perusahaan ini, tabel fakta terdiri dari seluruh *sequence* yang terdapat pada tabel dimensi beserta informasi penggunaan daya energi listrik dan gas secara *annual* untuk mengetahui *record* dari setiap *building*-nya. Berikut merupakan tampilan proses transformasi pada tabel fakta: fact_residential, yang mengambil *sequence* sebagai *key* dalam tabel fakta beserta penyeleksian dan output yang dihasilkan pada transformasi tabel fakta ini.



Step name: **Select values**

Select & Alter Remove Meta-data

Fields:

#	Fieldname	Rename to	Length	Precision
1	SK_Electricity			
2	SK_Gas			
3	SK_Building			
4	SK_Building_fac_1			
5	SK_Building_fac_2			
6	DatalamID			
7	YEAR			
8	AnnualElectricity_kWh_perYear			
9	AnnualGas_kBtu_perYear			

Get fields to select Edit Mapping

Step name: **Table output**

Connection: **Hidup Bumi Sejahtera** Edit... New... Wizard...

Target schema: Browse...

Target table: **fact_residential** Browse...

Commit size: 1000

Truncate table ☐

Ignore insert errors ☐

Specify database fields ☐

IV. HASIL DAN PEMBAHASAN

A. Star Schema

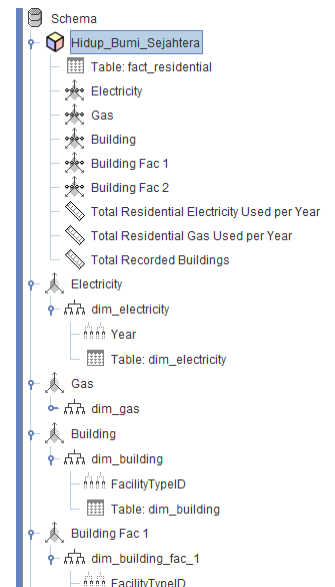
Hasil dari pengolahan ETL dari sekian data mengenai properti dan bangunan perusahaan Hidup Bumi Sejahtera yang menggunakan *tools* Pentaho berupa *database* dengan tabel dimensi dan tabel fakta sesuai dengan laporan yang ada dan menjadi kebutuhan perusahaan. Salah satunya merupakan *schema* yang merupakan wujud dari integrasi persatuan antar tabel dimensi dan fakta pada setiap data yang dimiliki perusahaan. *Schema* yang digunakan pada perusahaan Hidup Bumi Sejahtera adalah *schema* berbentuk *star/bintang*, yang menggabungkan data dengan 5 *file* tersebut menjadi setiap tabelnya: *dim_building*, *dim_building_fac_1*, *dim_building_fac_2*, *dim_electricity*, *dim_gas*, ke tabel fakta *fact_residential* berdasarkan *stream lookup* dari setiap SK tabel dimensional. Berikut merupakan tampilan setiap tabel dalam *database* dan visualisasi *star schema* pada Pentaho:

Table	Action	Rows	Type	Collation	Size	Overh
dim_building	Browse Structure Search Insert Empty Drop	~65,648	InnoDB	utf8mb4_general_ci	8.5 MiB	
dim_building_fac_1	Browse Structure Search Insert Empty Drop	~65,961	InnoDB	utf8mb4_general_ci	8.5 MiB	
dim_building_fac_2	Browse Structure Search Insert Empty Drop	465	InnoDB	utf8mb4_general_ci	88.0 KiB	
dim_electricity	Browse Structure Search Insert Empty Drop	~1,262,428	InnoDB	utf8mb4_general_ci	188.6 MiB	
dim_gas	Browse Structure Search Insert Empty Drop	~257,725	InnoDB	utf8mb4_general_ci	28.5 MiB	
fact_residential	Browse Structure Search Insert Empty Drop	~3,142,952	InnoDB	utf8mb4_general_ci	141.7 MiB	
6 tables	Sum	~3,797,163	InnoDB	utf8mb4_general_ci	280.0 MiB	

B. OLAP (Schema Workbench)

Proses OLAP yang menggunakan *tools* dari Pentaho Schema Workbench dan juga Mondrian adalah sebagai wujud menciptakan tabel multidimensional berdasarkan acuan *star-schema* yang telah dibuat. Schema Workbench memungkinkan basis data untuk membuat *cube* yang berisi dari data-data dimensi berstruktur hierarki beserta setiap *level* dan

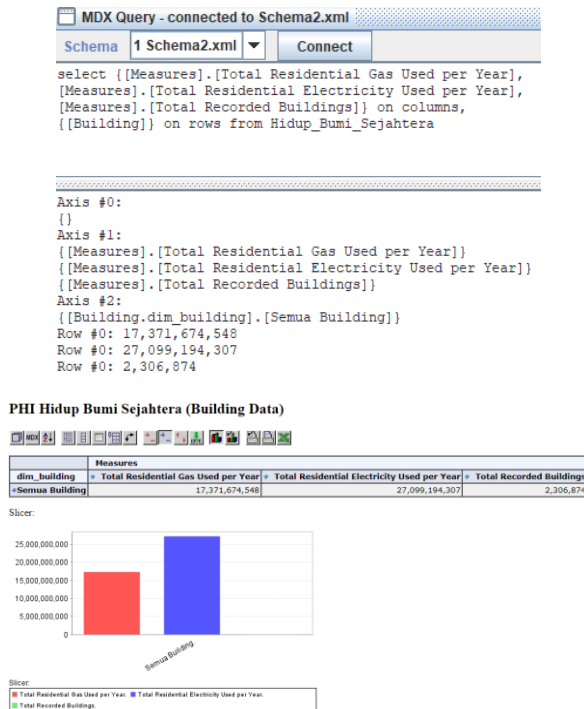
tabel sumbernya. Berikut merupakan ilustrasi dari tahap pengerjaan menggunakan Schema Workbench.



Proses diselesaikan menggunakan koneksi *database* MySQL pada *database* yang menjadi output dari proses ETL sebelumnya. Dalam Schema Workbench, data-data dari basis data tersebut memungkinkan untuk dimasukkan ke dalam sebuah *cube* dimensi berisi dengan dimensi lainnya yang memiliki kemampuan mengolah data dimensi lainnya, yang disebut sebagai multidimensional. Pengolahan tersebut dapat dilakukan dengan fitur kueri bernama “MDQ Query” yang menghasilkan output dari pengolahan data multidimensional. Tidak lupa sebelum membuat kueri, diperlukan suatu *measures* yang menjadi acuan dari dimensi data lainnya dalam menghitung data yang terdapat pada *database*, seperti pada kasus perusahaan ini yang membuat *measures* berupa *total recorded buildings* dan *total residential electricity and gas used per year*. Berikut merupakan gambaran rinci mengenai *measures* salah satunya yang *electricity used per year*.

Attribute	
name	Total Residential Electricity Used per Year
description	
aggregator	sum
column	AnnualElectricity_kWh_perYear
formatString	
datatype	
formatter	
caption	
visible	<input checked="" type="checkbox"/>

Setelah ada *measures* yang dapat digunakan, maka perusahaan dapat merumuskan kueri yang sesuai dengan kebutuhannya, salah satunya adalah dengan menghitung jumlah total penggunaan energi listrik dan gas pada setiap *building* yang terdapat pada datanya. Berikut merupakan tampilan mengenai kueri dan hasil output yang diberikan.



Dapat terlihat bahwa pada struktur dan visualisasi dari hasil output data multidimensional yang telah dilakukan bahwa daya penggunaan energi elektrik yang mencapai 27 juta kWh pada setiap residensialnya, diikuti dengan 17 juta kBTu setiap tahunnya. Dijelaskan juga bahwa dari sekian banyak properti, total yang tercatat dan terdaftar oleh sistem dihitung dengan perhitungan matematis.

C. Pengujian

Pengujian yang dilakukan adalah dengan menyesuaikan *database*, terutama dengan output yang dihasilkan pada proses ETL dan juga transformasi lainnya. Hal tersebut juga diikuti dengan perubahan yang terjadi jika implementasi pada *cube* untuk data multidimensional dilakukan, maka butuh beberapa penyesuaian lebih lanjut agar data dapat diimprovisasi sedemikian rupa untuk menghasilkan hasil yang lebih baik lagi. Berikut merupakan hasil akhir dari *database* terutama tabel fakta *fact_residential* sebagai bahan pengujian dari metode penelitian ini.

SK_Electricity	SK_Gas	SK_Building_fac_1	SK_Building_fac_2	DataItemID	YEAR	AnnualElectricity_kWh_perYear	AnnualGas_kBtu_perYear	SK_Building
1	NULL	1	NULL	1	2010	19703	NULL	1
2	NULL	1	NULL	1	2010	19703	NULL	1
3	NULL	1	NULL	1	2010	19703	NULL	1
4	NULL	1	NULL	1	2010	19703	NULL	1
5	NULL	1	NULL	1	2010	19703	NULL	1
6	NULL	1	NULL	1	2010	19703	NULL	1
7	NULL	1	NULL	1	2010	19703	NULL	1
8	NULL	1	NULL	1	2010	19703	NULL	1
9	NULL	1	NULL	1	2010	19703	NULL	1
10	NULL	1	NULL	1	2010	19703	NULL	1
11	NULL	1	NULL	1	2010	19703	NULL	1
12	NULL	1	NULL	1	2010	19703	NULL	1
13	NULL	2	NULL	2	2010	19004	NULL	2
14	NULL	2	NULL	2	2010	19004	NULL	2
15	NULL	2	NULL	2	2010	19004	NULL	2
16	NULL	2	NULL	2	2010	19004	NULL	2
17	NULL	2	NULL	2	2010	19004	NULL	2
18	NULL	2	NULL	2	2010	19004	NULL	2
19	NULL	2	NULL	2	2010	19004	NULL	2
20	NULL	2	NULL	2	2010	19004	NULL	2
21	NULL	2	NULL	2	2010	19004	NULL	2

SK_Electricity	SK_Gas	SK_Building_fac_1	SK_Building_fac_2	DataItemID	YEAR	AnnualElectricity_kWh_perYear	AnnualGas_kBtu_perYear	SK_Building
344967	64376	29237	NULL	29382	2010	6616	24800	29382
344968	64376	29237	NULL	29382	2010	6616	24800	29382
344969	64376	29237	NULL	29382	2010	6616	24800	29382
344970	64376	29237	NULL	29382	2010	6616	24800	29382
344971	64376	29237	NULL	29382	2010	6616	24800	29382
344972	64376	29237	NULL	29382	2010	6616	24800	29382
344973	64376	29237	NULL	29382	2010	6616	24800	29382
344974	64376	29237	NULL	29382	2010	6616	24800	29382
344975	64376	29237	NULL	29382	2010	6616	24800	29382
344976	64376	29237	NULL	29382	2010	6616	24800	29382
344977	64376	29237	NULL	29382	2010	6616	24800	29382
72829	NULL	6028	NULL	6183	2010	9793	NULL	6183
72830	NULL	6028	NULL	6183	2010	9793	NULL	6183
72831	NULL	6028	NULL	6183	2010	9793	NULL	6183
72832	NULL	6028	NULL	6183	2010	9793	NULL	6183
72833	NULL	6028	NULL	6183	2010	9793	NULL	6183
72834	NULL	6028	NULL	6183	2010	9793	NULL	6183
72835	NULL	6029	NULL	6184	2010	8028	NULL	6184
72836	NULL	6029	NULL	6184	2010	8028	NULL	6184
72837	NULL	6029	NULL	6184	2010	8028	NULL	6184
72838	NULL	6029	NULL	6184	2010	8028	NULL	6184
72839	NULL	6029	NULL	6184	2010	8028	NULL	6184
72840	NULL	6029	NULL	6184	2010	8028	NULL	6184
72841	NULL	6029	NULL	6184	2010	8028	NULL	6184
72842	NULL	6029	NULL	6184	2010	8028	NULL	6184
72843	NULL	6029	NULL	6184	2010	8028	NULL	6184
72844	NULL	6029	NULL	6184	2010	8028	NULL	6184
72845	NULL	6029	NULL	6184	2010	8028	NULL	6184
72846	NULL	6029	NULL	6184	2010	8028	NULL	6184
72847	NULL	6030	NULL	6185	2010	16718	NULL	6185
72848	NULL	6030	NULL	6185	2010	16718	NULL	6185
72849	NULL	6030	NULL	6185	2010	16718	NULL	6185
72850	NULL	6030	NULL	6185	2010	16718	NULL	6185
72851	NULL	6030	NULL	6185	2010	16718	NULL	6185

V. CONCLUSION (KESIMPULAN)

Berdasarkan analisis data dan pembahasan yang telah dipaparkan, menjadi informasi bernilai bagi perusahaan Hidup Bumi Sejahtera untuk dapat mengetahui total penggunaan sumber daya setiap bangunan di residensialnya, termasuk identifikasi setiap propertinya yang memiliki karakteristik berbeda-beda. Insight yang diberikan pada hasil output tersebut tentang penggunaan keseluruhan tingkat konsumsi sumber daya listrik dan gas pada penerapannya di jenis-jenis karakteristik propertinya.

Penggunaan *star schema* dalam mengisi variabel dan aspek saja yang dibutuhkan untuk analisis dengan menggunakan integrasi data, Pentaho, menghasilkan sebuah tabel fakta, yang merupakan hasil record pada data yang ada dari pengumpulan kolektif tabel dimensi sebagai dukungan informasi. Hal ini juga menjadi acuan bahwa memang perlu adanya perbaikan dan peningkatan kinerja pembangunan pada sebagian aset properti yang dimiliki perusahaan Hidup Bumi Sejahtera untuk peningkatan kualitas barang properti yang lebih baik lagi, terlebih mencukupi standar yang diberikan oleh Department of Energy yang berlaku di wilayah tersebut.

REFERENSI

- [1] A. Pujianto, A. Mulyati, and R. Novaria, "Pemanfaatan Big Data Dan Perlindungan Privasi Konsumen Di Era Ekonomi Digital," *Maj. Ilm. Bijak*, vol. 15, no. 2, pp. 127–137, 2018, doi: 10.31334/bijak.v15i2.201.
- [2] I. P. W. Prasetya and I. N. H. Kurniawan, "Implementasi ETL (Extract, Transform, Load) pada Data warehouse Penjualan Menggunakan Tools Pentaho," *TIERS Inf. Technol. J.*, vol. 2, no. 1, pp. 1–8, 2021, doi: 10.38043/tiers.v2i1.2844.
- [3] K. Khotimah, "Perancangan Dan Implementasi Data Warehouse Untuk Mendukung Sistem Akademik (Studi Kasus Pada STKIP Muhammadiyah Kotabumi)," *J. Teknol. Inf. Magister Darmajaya*, vol. 2, no. 01, pp. 94–107, 2016.
- [4] M. Hilman and D. Djamaludin, "Analisis Faktor Optimalisasi Proses Etl Pada Data Warehouse Sebagai

- Pendukung Pengambilan Keputusan Management Dengan Business Intelligence,” *Fakt. Exacta*, vol. 11, no. 1, p. 24, 2018, doi: 10.30998/faktorexacta.v11i1.2325.
- [5] R. Wahjoe Witjaksono, M. Wiyogo, and P. N. Wicaksono, “Perancangan Aplikasi Business Intelligence Pada Sistem Informasi Distribusi PT Pertamina Lubricant Menggunakan Pentaho R. Wahjoe Witjaksono, Mardiyanto Wiyogo, Prima Nanda Wicaksono (hal,” *J. Rekayasa Sist. Ind.*, vol. 2, pp. 12–18, 2015.
- [6] G. A. Udayana, I. M. Y. Mahendra, I. K. A. Sukawirasa, G. D. Dimastawan Saputra, and I. B. M. Mahendra, “Implementasi Data Warehouse Dan Penerapannya Pada PHI-Minimart Dengan Menggunakan Tools Pentaho dan Power BI,” *JELIKU (Jurnal Elektron. Ilmu Komput. Udayana)*, vol. 10, no. 1, p. 163, 2021, doi: 10.24843/jlk.2021.v10.i01.p19.
- [7] N. Hidayati, “Pentaho Sebagai Solusi Masalah,” *J. Transform.*, pp. 86–94, 2012.