Improving the Quality of Service: ETL Implementation on Data Warehouse at Pharmacy Industry

Fardhila Zahra Dwi Wardhani[1], Jansen Wiratama[2,*]

[1, 2] Faculty of Engineering and Informatics, Information Systems, Universitas Multimedia Nusantara, Tangerang, Banten, Indonesia
Email: [1]fardhila.wardhani@student.umn.ac.id, [2,*]jansen.wiratama@umn.ac.id
[*)] Corresponding email

Abstrak—Pelayanan Farmasi merupakan bagian penting dalam masyarakat dan memainkan peran yang tak tergantikan. Ketersediaan obat menjadi faktor kunci yang perlu diperhatikan dalam pelayanan kesehatan. Parapharma, sebagai perusahaan yang bergerak di bidang penyediaan obat-obatan, menyadari pentingnya pemantauan dan analisis data untuk mengoptimalkan ketersediaan obat. Pada proses yang berjalan, pemantauan dan analisis data terkait ketersediaan obat belum dapat berjalan secara optimal karena jumlah data yang cukup banyak dan proses pemantauan dan analisis datanya masih dikerjakan secara konvensional, sehingga pengambilan keputusan dan manajemen ketersediaan obat masih perlu dioptimalkan. Oleh karena itu, melalui penelitian ini akan diimplementasikan Data Warehouse dan Online Analytical Processing (OLAP) guna mendukung proses pengambilan keputusan dalam manajemen ketersediaan obat. Proses Extract, Transform, and Load (ETL) digunakan untuk mengintegrasikan data dari sumber-sumber berbeda ke dalam data warehouse. Data tersebut kemudian diolah dan ditransformasi agar dapat digunakan untuk analisis lebih lanjut. Dengan implementasi data warehouse dan ETL, Parapharma dapat memiliki akses yang cepat dan efisien terhadap informasi penting yang terkandung dalam data. Teknologi OLAP, khususnya menggunakan tools Mondrian, menjadi sarana bagi perusahaan dalam menggali informasi dari data warehouse dengan bantuan visualisasi yang interaktif. Berbagai visualisasi seperti informasi tentang laboratorium penyedia brand obat terbanyak, jenis obat yang paling diminati, dan rata-rata harga per-bentuk obat, dapat memberikan wawasan yang signifikan bagi keputusan strategis perusahaan. Penelitian ini menghasilkan penerapan data warehouse dan OLAP, sehingga parapharma dapat memperoleh informasi yang akurat dan terperinci mengenai permintaan obat dan tren pasar. Hal ini memungkinkan perusahaan untuk mengambil keputusan yang tepat dalam mengelola produksi obat-obatan untuk memenuhi kebutuhan apotek secara efisien. Penggunaan data warehouse dan OLAP juga memberikan dampak positif dalam meningkatkan kualitas pelayanan kepada masyarakat.

Kata Kunci: Data Warehouse, Ketersediaan Obat, OLAP, OLTP, Visualisasi Data.

Abstract— Pharmacy services are an important part of society and play an irreplaceable role. Availability of medicines is a key factor that needs to be considered in health services. Parapharma, as a company engaged in the supply of medicines, realizes the importance of monitoring and analyzing data to optimize medicine availability. In the ongoing process, monitoring and analyzing data related to pharmacy availability has yet to run optimally because the amount of data is quite large. The monitoring and data analysis process is still carried out conventionally, so decision-making and management of pharmacy availability still need to be optimized. Therefore, through this research, Data Warehouse and Online Analytical Processing (OLAP) will be implemented to support the decision-making process in managing pharmacy availability. The Extract, Transform, and Load (ETL) process integrates data from different sources into a data warehouse. The data is then processed and transformed so that it can be used for further analysis. By implementing a data warehouse and ETL, Parapharma can have fast and efficient access to important information contained in the data. OLAP technology, especially using Mondrian tools, is a means for companies to extract information from data warehouses with the help of interactive visualization. Various visualizations, such as information about laboratories providing the most popular pharmacy brands, the most in-demand types of pharmacys, and the average price per form of pharmacy, can provide significant insight into company strategic decisions. This research resulted in applying data warehouse and OLAP so that Parapharma can obtain accurate and detailed information regarding pharmacy demand and market trends. These research results allow companies to make the right decisions to manage pharmacy production to meet pharmacy needs efficiently. The use of data warehouse and OLAP also positively impacts improving the quality of service to the community.

Keywords: Data Warehouse, Data Visualization, Medicine Availability, OLAP, OLTP.

1. INTRODUCTION

As time progresses, supplies also grow. Not only in terms of food and drink supplies but medicine too. Medicines are biological products that influence or evaluate physiological systems to establish a diagnosis, prevention, cure, and recovery for humans. Medicine is an irreplaceable component of community service. In the Decree of the Minister of Health, there is a statement saying that there are at least 923 medicine forms in stock, and currently, the medicine supply circulating in the community has various forms of production, namely patents and generics [1]. Patent medicines are medicines that have gone through a research process by the first manufacturer to discover the compound or active substance of the medicine.

The data warehouse is very important for companies that store large amounts of data. A data warehouse allows companies to access and analyze all data. The data warehouse has components that extract, transform, and

load data. This component is usually called ETL (Extract, Transform, Load), which can collect, read, and transfer raw data to various data sources. In a data warehouse, the ETL process is part of a process that takes a long time [3]. Several prior studies have shown that adopting information technology helps optimize company operations, which in turn affects the quality of service provided by the organization [4]. Using the correct technologies can also make decision-making more optimal and objective [5]. The benefits of adopting information technology have also been shown to make businesses more productive and efficient in terms of time, cost, and effort, as well as increase data security [6].

Meanwhile, generic medicines have expired patent periods and can be marketed by other traders or pharmacies. However, in choosing the use of medicines, one does not only look at the medicine from the validity of its name but there are several other aspects, such as the dosage of the medicine, the physical form of the medicine, the medicine content, the type of medicine, the benefits of the medicine, the price of the medicine and many others. Therefore, in the effort to provide health services, one important component that needs to be considered is the availability of medicines. Several important aspects need to be considered to achieve the right target, such as complete types, sufficient quantities, guaranteed efficacy, safety, effectiveness, and quality [2]. Parapharma is one of the businesses engaged in the supply of medicines. The name para pharma is taken from 2 words: Para, which means healthy, and Pharma, which means medicine provider. If taken as a whole, Parapharma is a medicine provider expected to help consumers get back to health. This company provides various types, brands, and detailed information about medicines, prices, and so on that are needed for medicine distribution in every pharmacy that needs medicine. In the process of producing medicines, Parapharma will enter the laboratory where the medicines are made, as well as information about medicines in the database. Due to the increasing amount of data regarding medicines, Parapharma needs a data warehouse developer to design a data warehouse so that it can accommodate large amounts of data and transform data so that the data can be used for analysis purposes and retrieve important information stored in these data. With the existence of a data warehouse in this company, Parapharma will be supported in the decision-making process, especially in processing the availability of medicine production available in this company.

2. LITERATURE REVIEW

2.1 Data Marts

A data mart is a subject-oriented database that is a partitioned segment of a company's data warehouse. Data mart is a limited-use system used to analyze information on the right area of a particular production sector. Data marts play an important role in a business because they can speed up business processes with easy access to relevant information in data warehouses for quite a long time, several days, no, even more. Data marts are an important component of a data warehouse focused on certain lines of business, such as sales, marketing, human resources, and so on [7].

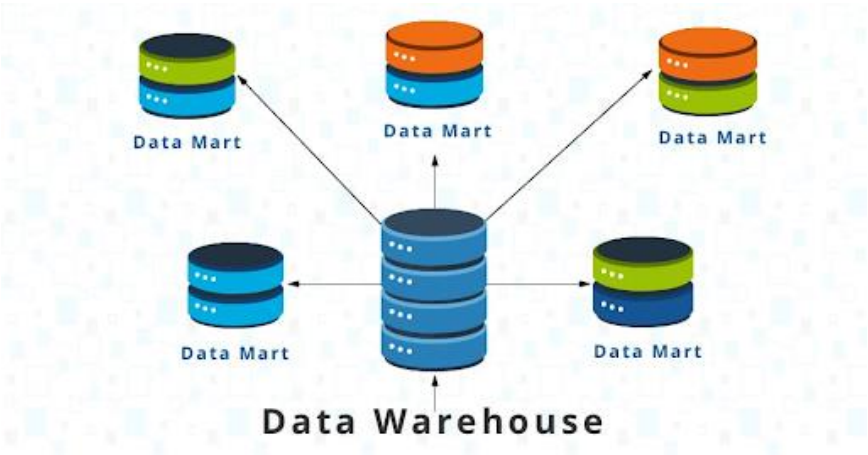


Fig 1. Data Marts Illustrations

2.2 Online Transaction Processing (OLTP)

Online Transaction Processing (OLTP) is a process-oriented system for processing transactions directly through network-connected devices. The online transaction process usually involves inserting, updating, and deleting small portions of data in data storage to collect, process data, and secure these transactions. All transactions that occur between consumers, suppliers or partners will be tracked and updated on the OLTP database. This data is very important for businesses which will later be used to be analyzed for data-based decision making [8].

2.3 Online Analytical Processing (OLAP)

Online Analytical Processing (OLAP) is a technology that manages large databases and supports complex analytical processes. OLAP is a method for quickly presenting answers from dimensional analysis processes in the form of designs of applications and technologies that can collect, store, and process multidimensional data for analysis. OLAP will closely relate to master data whose nature does not change frequently. OLAP technology will facilitate companies in making decisions based on stored data [9].

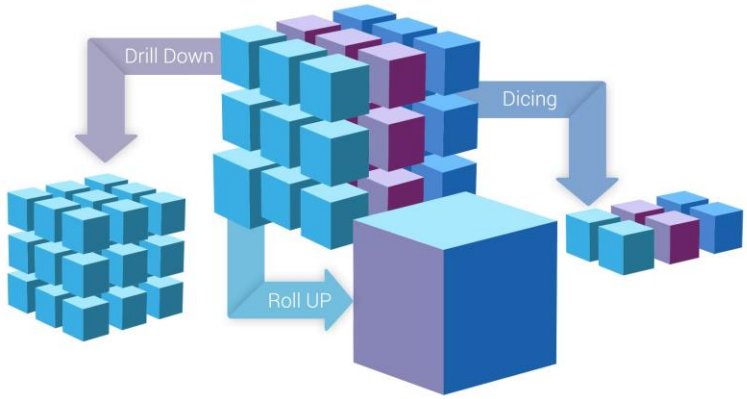


Fig 2. OLAP Cube

2.4 Metadata

Metadata is data that is used to represent other data. Metadata is data that is summarized to represent the user regarding detailed data. Metadata is a directory that helps system decision-makers find the data warehouse. Metadata has a very important role in the data warehouse, including acting as a directory that helps the system to find the contents of the data warehouse, helps map data, and helps summarize between detailed data and summarized data, overall metadata plays an important role in loading a function [10].

2.5 Extract, Transform, and Load (ETL)

Extract, Transform, and Load or ETL is a data integration process that combines data from several sources into one coherent data store that is processed to a data warehouse or other target system. ETL will clean and manage data for monthly reporting purposes and improve back-end processes. Companies commonly use ETL to extract data from old systems, clean or transform it to improve data quality and build consistency, and load data into the target database [11].

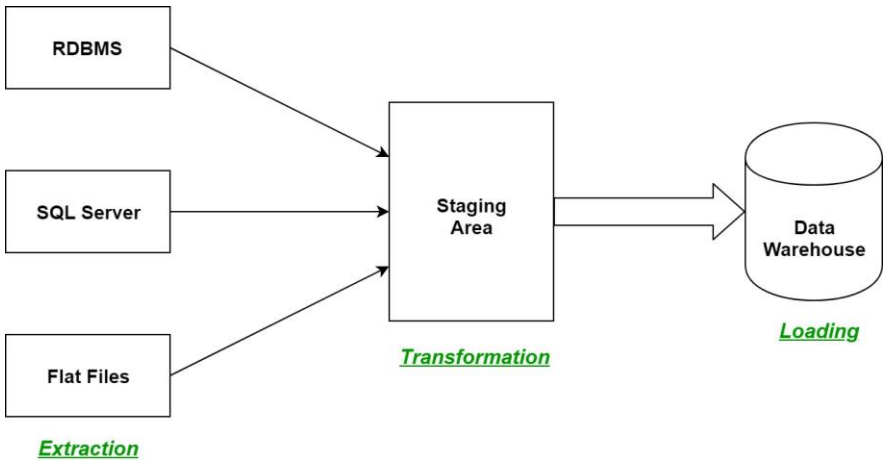


Fig 3. ETL Process

2.6 Star Schema

Star Schema is a dimensional model consisting of a fact table located at the center and dimensional tables located around the central table. Most fact tables in the star schema are the third normalized form of the database, while the dimensional table is the second normalized form. The fact table contains all IDs from the dimensional table or properties as a calculation component. The dimensional table contains descriptive information from all measurements in the fact table. An illustration of the overall dimensions of this model can be seen in Figure 3.

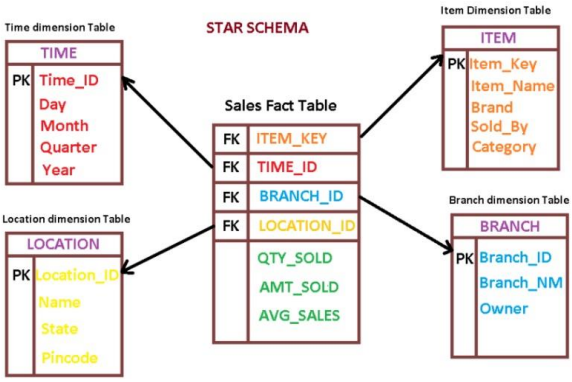


Fig 4. Star Schema Illustration

This dimensional model has advantages and disadvantages in its use. The advantage of the star schema is that it is easy to understand because of its simple structure and relatively easy maintenance. While the drawback is that its performance is somewhat poor compared to other schemas, the dimension tables owned by the star schema are far more numerous. The Star schema has no levels or subdimensions [12].

3. METHODOLOGY

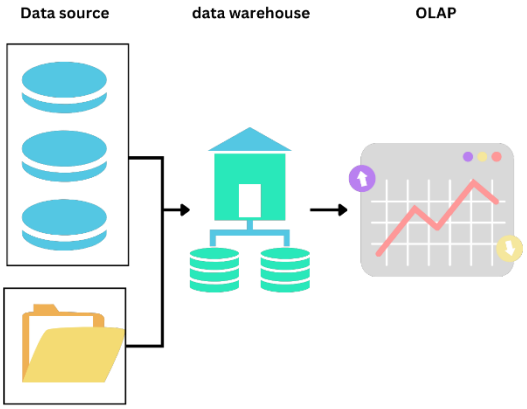


Fig 5. Data Processing flow

Several stages are used to design and create a data warehouse at the Parapharma company. The steps to be carried out are summarized in Figure 5. Data will be collected to start processing raw data. Then enter the data warehouse stage, and the data will be processed by the data integration process, namely ETL using Pentaho. After that, the data will be processed using Mondrian as a tool.

A. Data Collecting

The method used for data collection is by taking data via the internet, which is in the form of secondary data provided by the Parapharma company [13]. The data used has a .csv format; this format will make it easier to transfer data to the database. This data has three tables, namely manufacturer, medicine, and dosage_form. Overall, this dataset has a total of 13934 rows and 17 columns.

B. Data Understanding

The data used in this research is data regarding the production of medicines in the form of types and detailed information about a medicine in the Parapharma company. Figure 6 shows the dimension tables that will be used to design the data warehouse [14].

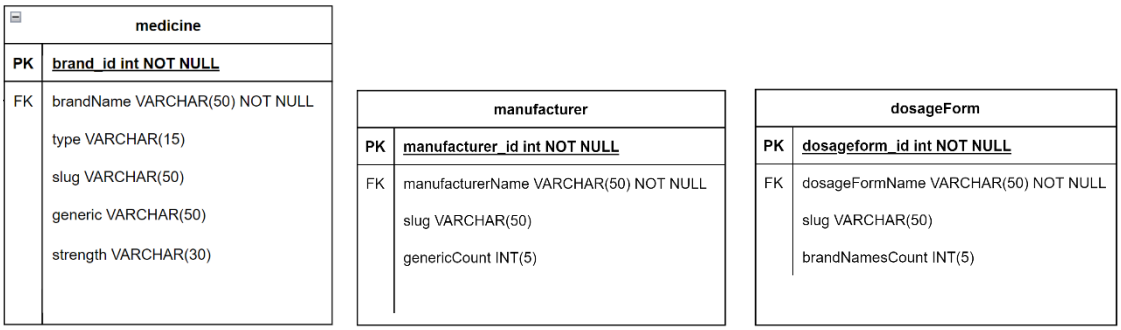


Fig 6. Medicine, Manufacturer, and Dosage Form Dimensional tables

C. ETL Process

All data that will be stored in the data warehouse must be in a neat and structured condition so that it can be easily accessed and processed into a source of information for the future. Therefore, the data must go through the Extract, Transform, and Load (ETL) process. Pentaho Data Integration (PDI) will be used as a tool in this process. The stages of the ETL process will be carried out, as shown in Figure 7.

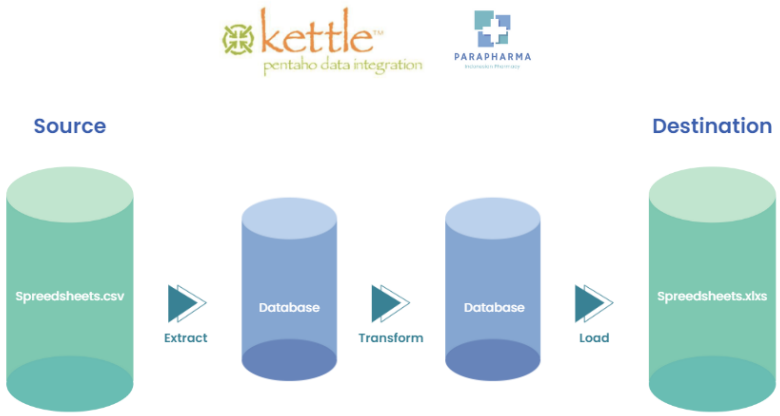


Fig 7. Proposed ETL Process

Extract is the stage to retrieve the data that has been provided. In this case, the data is in the .csv format, so in the PDI transformation, the input step CSV file will be used as the step to retrieve the data. Transform is the stage for transforming data into data that is neater, structured, and easy to process. In PDI transformation, the step if the value is null will be used to fill in data indicated as null, the step select value to select the column you want to use, and the step sort rows to sort rows. When the data has passed through these various stages, the data is structured and can be used for the next stage. Load is the final stage for entering transformed data into the data warehouse. This stage will use the output step text file with the .csv extension so that the resulting file is a file with the .csv format. This stage also includes creating a star schema containing fact tables and dimension tables for later use in the Pentaho Schema Workbench. Fact tables are created using step stream lookup and select values for all dimension tables. Next, UML will be generated from PDI Transformation. Figure 7 is a star schema containing fact tables and dimension tables that have been made.

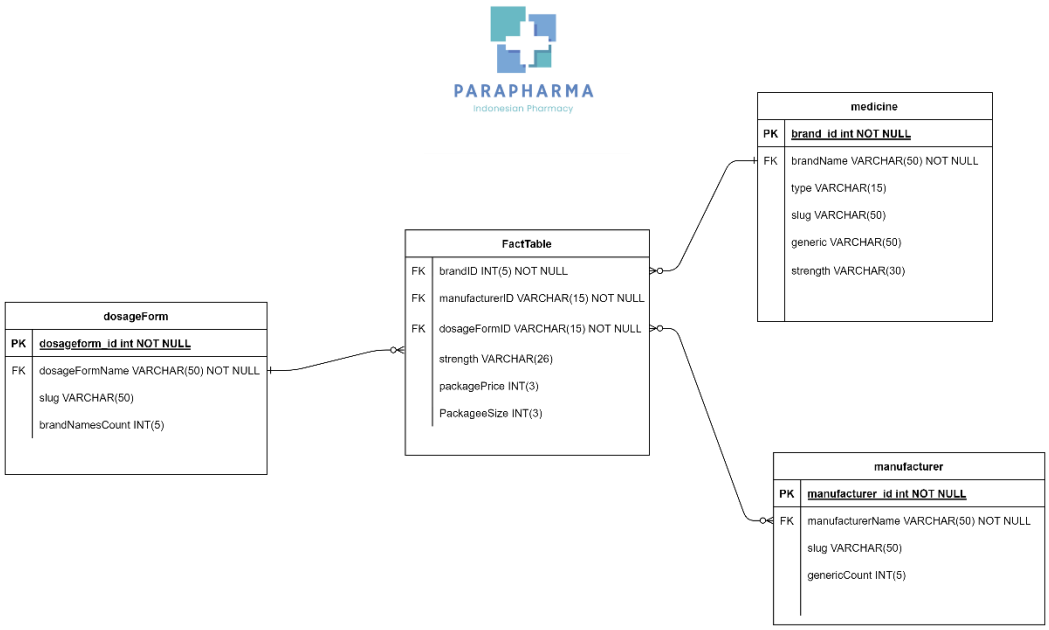


Fig 8. Star Schema on Parapharma

D. Mondrian

Mondrian is an Online Analytics Processing (OLAP) engine for executing queries written in the MDX language, reading data from relational databases, and presenting the results in a multidimensional format via the Java API. Mondrian requires XML for the analysis process to generate the XML in the Pentaho Schema Workbench. To use Mondrian, you need Tomcat, which allows you to read Java programming-based applications [15]. Figure 3.8 is a display of the Mondrian index of the Parapharma company.

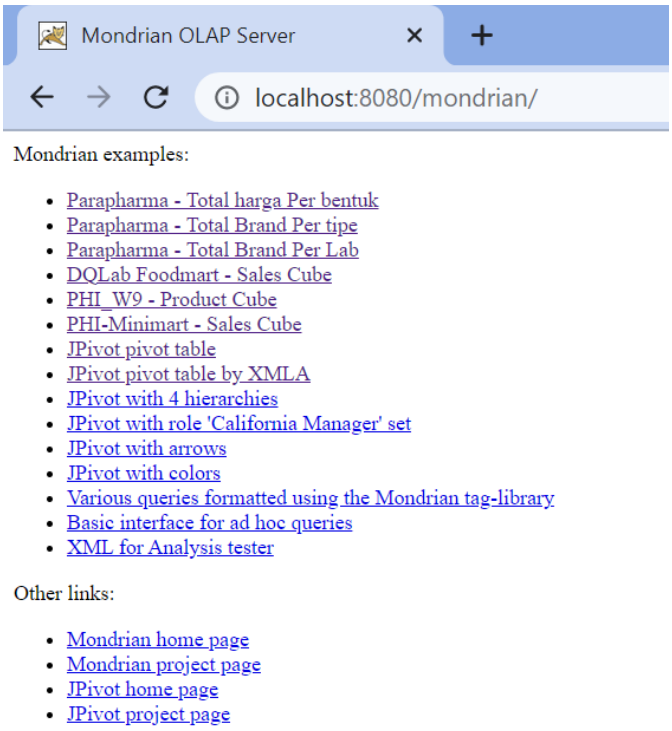


Fig 9. Mondrian Interface on Parapharma

4. RESULTS AND DISCUSSIONS

Drawing upon the aforementioned methodology, the comprehensive procedure will yield organized data and visual representations that can facilitate the formulation of plans or decisions for enterprises operating in the Parapharma industry. The subsequent section provides a comprehensive elucidation of each individual step.

3.1 Extract, Transformation, and Load (ETL) Process

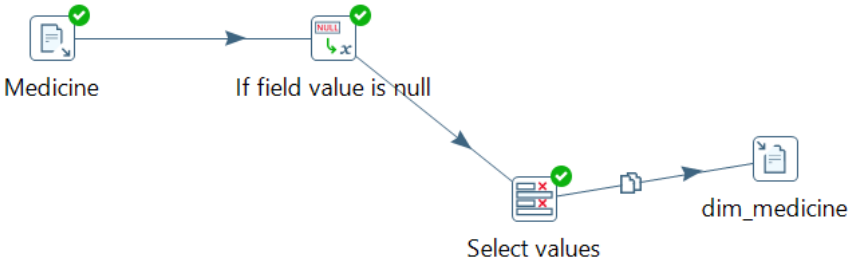


Fig 10. ETL Medicine Dimension Table

Figure 10 is one of the ETL processes to extract and transform data for table medicine. Data is extracted using step CSV file input. Then, after the data has been successfully read, a lot of null data is found in the strength field, as shown in Figure 11.

Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

	dosage form	generic	strength
	Tablet	Garlitab	<null>
	Capsule	Nigella Sativa [Black Seed Oil]	500 mg
	Chewable Tablet	Vitamin E [Alpha Tocopherol Acetate]	200 mg
	Capsule	Tabkheer	<null>
	Syrup	Herbal cough syrup [Tulsi]	<null>
	Tablet	Ambar Momiyaee	<null>
	Syrup	Herbal cough syrup [Adhatoda Vasica]	<null>
	Syrup	Herbal cough syrup [Adhatoda Vasica]	<null>
t	Oral Powder	Probiotic Combination [4 Billion]	4 billion/sachet
	Capsule	Probiotic Combination [4 Billion]	4 billion
	Syrup	Mustakarista [Cyperus, Zingiber, Piper nigrum]	<null>
	Capsule	Spirulina	450 mg
ision	Oral Suspension	Emblic Myrobalan + Aswaqandha + Grape	<null>

Fig 11. Medicince table before transformed

As a result of the null data in the strength field, the transformation cannot work properly. Therefore, if the field value is null, the step is used to fill in the null data. At this step, null data is filled with NA, so the data will still be read as a string.

Execution Results

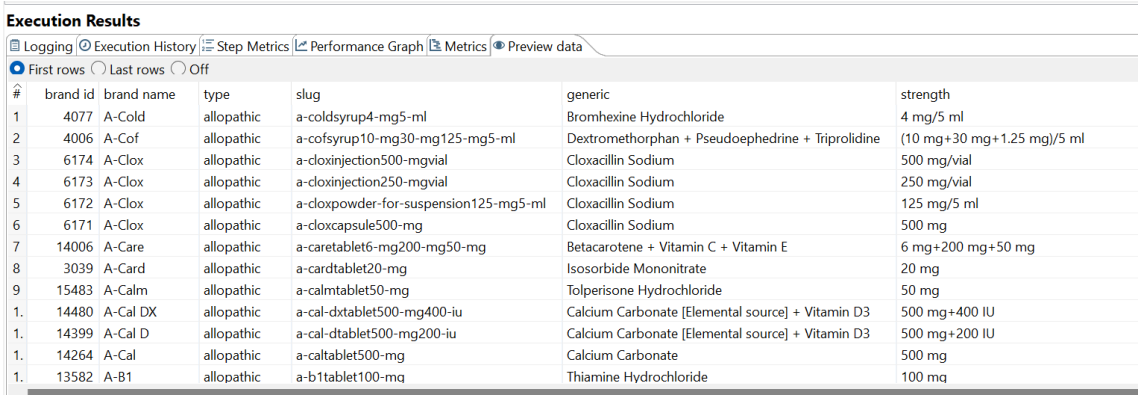
Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

ıg	dosage form	generic	strength
satablet	Tablet	Garlitab	NA
ksircapsule500-mg	Capsule	Nigella Sativa [Black Seed Oil]	500 mg
a-echewable-tablet200-mg	Chewable Tablet	Vitamin E [Alpha Tocopherol Acetate]	200 mg
erdcapsule	Capsule	Tabkheer	NA
odicsyrup	Syrup	Herbal cough syrup [Tulsi]	NA
odictablet	Tablet	Ambar Momiyaee	NA
ovassyrup	Syrup	Herbal cough syrup [Adhatoda Vasica]	NA
olefsyrup	Syrup	Herbal cough syrup [Adhatoda Vasica]	NA
teriaoral-powder4-billionsachet	Oral Powder	Probiotic Combination [4 Billion]	4 billion/sachet
teriacapsule4-billion	Capsule	Probiotic Combination [4 Billion]	4 billion
minasyrup	Syrup	Mustakarista [Cyperus, Zingiber, Piper nigrum]	NA
mes-spirulinacapsule450-mg	Capsule	Spirulina	450 mg
mes-chyabanprashoral-suspension	Oral Suspension	Emblic Myrobalan + Aswaqandha + Grape	NA

Fig 12. After using step If field value is null

After that, the step select value is used to select the fields for the medicine table; the fields used are brandID, brandName, type, slug, generic, and strength, with results like in Figure 13.



Execution Results

Logging | Execution History | Step Metrics | Performance Graph | Metrics | Preview data

First rows | Last rows | Off

#	brand id	brand name	type	slug	generic	strength
1	4077	A-Cold	allopathic	a-coldsyrup4-mg5-ml	Bromhexine Hydrochloride	4 mg/5 ml
2	4006	A-Cof	allopathic	a-cofsyrup10-mg30-mg125-mg5-ml	Dextromethorphan + Pseudoephedrine + Triprolidine	(10 mg+30 mg+1.25 mg)/5 ml
3	6174	A-Clox	allopathic	a-cloxinjection500-mgvial	Cloxacillin Sodium	500 mg/vial
4	6173	A-Clox	allopathic	a-cloxinjection250-mgvial	Cloxacillin Sodium	250 mg/vial
5	6172	A-Clox	allopathic	a-cloxpowder-for-suspension125-mg5-ml	Cloxacillin Sodium	125 mg/5 ml
6	6171	A-Clox	allopathic	a-cloxcapsule500-mg	Cloxacillin Sodium	500 mg
7	14006	A-Care	allopathic	a-caretablet6-mg200-mg50-mg	Betacarotene + Vitamin C + Vitamin E	6 mg+200 mg+50 mg
8	3039	A-Card	allopathic	a-cardtablet20-mg	Isosorbide Mononitrate	20 mg
9	15483	A-Calm	allopathic	a-calmtablet50-mg	Tolperisone Hydrochloride	50 mg
1.	14480	A-Cal DX	allopathic	a-cal-dxtablet500-mg400-iu	Calcium Carbonate [Elemental source] + Vitamin D3	500 mg+400 IU
1.	14399	A-Cal D	allopathic	a-cal-dtablet500-mg200-iu	Calcium Carbonate [Elemental source] + Vitamin D3	500 mg+200 IU
1.	14264	A-Cal	allopathic	a-caltablet500-mg	Calcium Carbonate	500 mg
1.	13582	A-B1	allopathic	a-b1tablet100-mg	Thiamine Hydrochloride	100 mg

Fig 13. Tranformation results on Medicine dimension table

After the data has been successfully transformed, the data will be loaded with a step txt output file to save the data back in .csv form and saved with the format name dim_medicine.

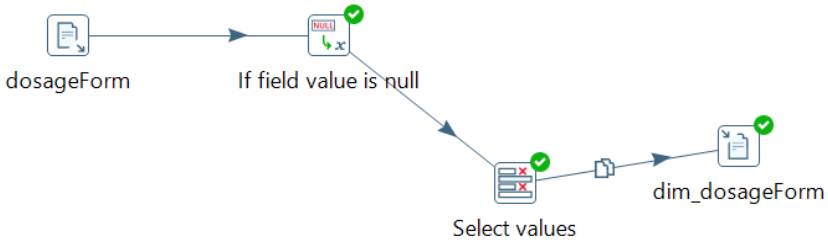


Fig 14. ETL Dosage Form dimension table

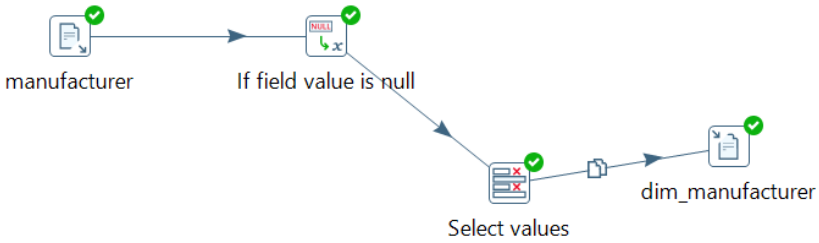


Fig 15. ETL Manufacturer dimension table

Figure 14 and Figure 15 is the ETL process on DosageForm and Manufacturer Dimension Table. The three Dimension Tables have the same problem: null data and unused fields. Therefore, use the step if the field value is null for all fields indicated as null. After that, field selection is carried out using the step select value. The final stage is saving data in .csv format to make it easier to import into the database. This storage uses a step txt output file with the .csv extension format.

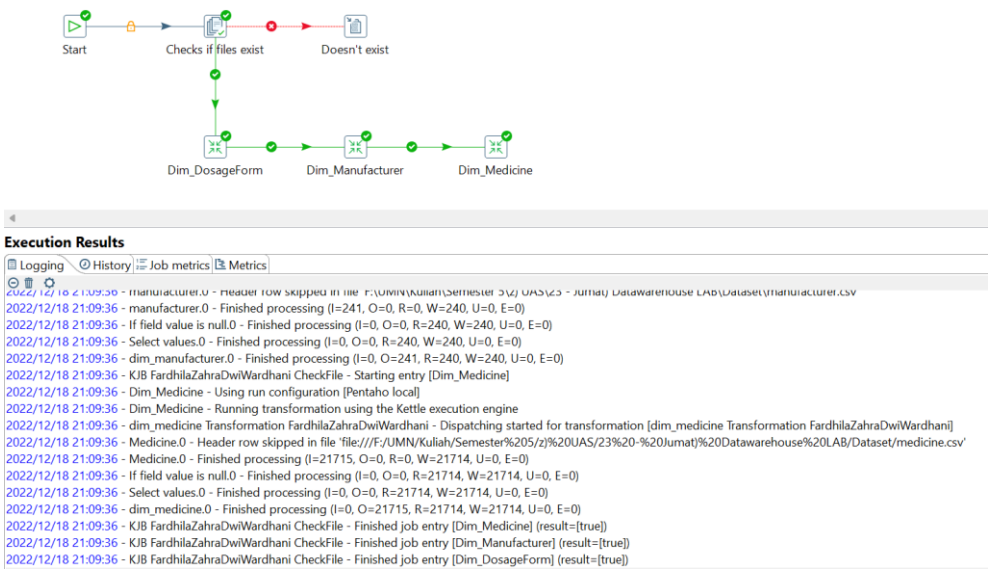


Fig 16. PDI Jobs runs the Transformation dimension tables

Figure 16 shows the Jobs process used to run all the Dim_DosageForm, Dim_Manufacturer, and Dim_Medicine transformations that were made previously. In this job, the Check if files exist step will be used to see whether the data file exists; if the data is not found, it will display a log that the file was not found using the write-to-log step.

	A	B	C	D	E
24	64	Vaginal Cream	vaginal-cream-64	11	
25	62	Transdermal Patch	transdermal-patch-62	6	
26	126	Topical Suspension	topical-suspension-126	1	
27	39	Topical Spray	topical-spray-39	20	
28	93	Topical Solution	topical-solution-93	20	
29	168	Topical Powder	topical-powder-168	10	
30	138	Topical Gel	topical-gel-138	30	
31	37	Tablet (Sustained Release)	tablet-sustained-release-37	127	
32	160	Tablet (Prolonged Release)	tablet-prolonged-release-160	2	
33	42	Tablet (Modified Release)	tablet-modified-release-42	55	
34	154	Tablet (Immediate Release)	tablet-immediate-release-154	1	
35	41	Tablet (Extended Release)	tablet-extended-release-41	247	
36	46	Tablet (Enteric Coated)	tablet-enteric-coated-46	308	
37	132	Tablet (Delayed Release)	tablet-delayed-release-132	122	
38	43	Tablet (Controlled Release)	tablet-controlled-release-43	65	
39	1	Tablet	tablet-1	9368	
40	3	Syrup	syrup-3	1010	
41	96	Surgical Scrub	surgical-scrub-96	2	
42	4	Suppository	suppository-4	64	
43	152	Sublingual Tablet	sublingual-tablet-152	4	
44	146	Sprinkle Capsule	sprinkle-capsule-146	3	
45	48	Solution for Inhalation	solution-for-inhalation-48	9	
46	85	Solution	solution-85	37	
47	125	Shampoo	shampoo-125	14	

Fig 17. Output from DosageForm Dimension Table processed by ETL

	A	B	C	D
1	manufacturer id	manufacturer name	slug	generics count
2	2	ACI Limited	aci-limited-2	352
3	3	ACME Laboratories Ltd.	acme-laboratories-ltd-3	410
4	103	Ad-din Pharmaceuticals Ltd.	ad-din-pharmaceuticals-ltd-103	70
5	114	Aexim Pharmaceuticals Ltd.	aexim-pharmaceuticals-ltd-114	34
6	292	Al-Madina Pharmaceuticals Ltd.	al-madina-pharmaceuticals-ltd-292	49
7	4	Alco Pharma Ltd.	alco-pharma-ltd-4	106
8	304	Alien Pharma	alien-pharma-304	3
9	245	Alkad Laboratories	alkad-laboratories-245	13
10	115	Allied Pharmaceuticals Ltd.	allied-pharmaceuticals-ltd-115	7
11	5	Ambee Pharmaceuticals Ltd.	ambee-pharmaceuticals-ltd-5	68
12	6	Amico Laboratories Ltd.	amico-laboratories-ltd-6	87
13	7	Amulet Pharmaceuticals Ltd.	amulet-pharmaceuticals-ltd-7	35
14	128	APC Pharma Ltd.	apc-pharma-ltd-128	27
15	8	Apex Pharmaceuticals Ltd.	apex-pharmaceuticals-ltd-8	77
16	9	Apollo Pharmaceutical Ltd.	apollo-pharmaceutical-ltd-9	37
17	258	AqVida bangladesh	aqvida-bangladesh-258	5
18	289	Arges Life Science Limited	arges-life-science-limited-289	18
19	10	Aristopharma Ltd.	aristopharma-ltd-10	312
20	81	Asiatic Laboratories Ltd.	asiatic-laboratories-ltd-81	93
21	11	Astra Biopharmaceuticals Ltd.	astra-biopharmaceuticals-ltd-11	54
22	246	Aztec Pharmaceuticals Ltd.	aztec-pharmaceuticals-ltd-246	14
23	13	Beacon Pharmaceuticals Ltd.	beacon-pharmaceuticals-ltd-13	237

Fig 18. Output from Manufacturer Dimension Table processed by ETL.

	A	B	C	D	E	F
1	brand id	brand name	type	slug	generic	strength
2	4077	A-Cold	allopathic	a-coldsyrup4-mg5-ml	Bromhexine Hydrochloride	4 mg/5 ml
3	4006	A-Cof	allopathic	a-cofsyrup10-mg30-m	Dextromethorphan + Pseudoephedi	(10 mg+30 mg+1.25 mg)/5 ml
4	6174	A-Clox	allopathic	a-cloxinjection500-mg	Cloxacillin Sodium	500 mg/vial
5	6173	A-Clox	allopathic	a-cloxinjection250-mg	Cloxacillin Sodium	250 mg/vial
6	6172	A-Clox	allopathic	a-cloxpowder-for-susp	Cloxacillin Sodium	125 mg/5 ml
7	6171	A-Clox	allopathic	a-cloxcapsule500-mg	Cloxacillin Sodium	500 mg
8	14006	A-Care	allopathic	a-caretablet6-mg200-n	Betacarotene + Vitamin C + Vitamin	6 mg+200 mg+50 mg
9	3039	A-Card	allopathic	a-cardtablet20-mg	Isosorbide Mononitrate	20 mg
10	15483	A-Calm	allopathic	a-calmtablet50-mg	Tolperisone Hydrochloride	50 mg
11	14480	A-Cal DX	allopathic	a-cal-dxtablet500-mg4	Calcium Carbonate [Elemental sour	500 mg+400 IU
12	14399	A-Cal D	allopathic	a-cal-dtablet500-mg20	Calcium Carbonate [Elemental sour	500 mg+200 IU
13	14264	A-Cal	allopathic	a-caltablet500-mg	Calcium Carbonate	500 mg
14	13582	A-B1	allopathic	a-b1tablet100-mg	Thiamine Hydrochloride	100 mg
15	25154	5X	allopathic	5xtablet30-mg	Ulipristal Acetate [For emergency co	30 mg
16	105	5-Fluril	allopathic	5-fluriliv-injection-or-ir	Fluorouracil	25 mg/ml
17	27661	Arubin	herbal	arubincapsule500-mg	Nabayas Louha Herbal Haematinic	500 mg
18	29715	Arkid	herbal	arkidsyrup	Aravindasav	NA
19	29714	Arjunacard	herbal	arjunacardsyrup	Arjunarista	NA
20	30450	Apelin	herbal	apelinsyrup	Malus sylvestris	NA
21	29646	Antiligo	herbal	antiligocapsule	Barsina	NA
22	31923	Anseng	herbal	ansengcapsule500-mg	Panax Ginseng	500 mg

Fig 19. Output from Medicine Dimension Table processed by ETL

Figures 17, 18, and 19 are the Outputs from DosageForm, Manufacturer, and Medicine Dimension tables that have gone through the ETL process.

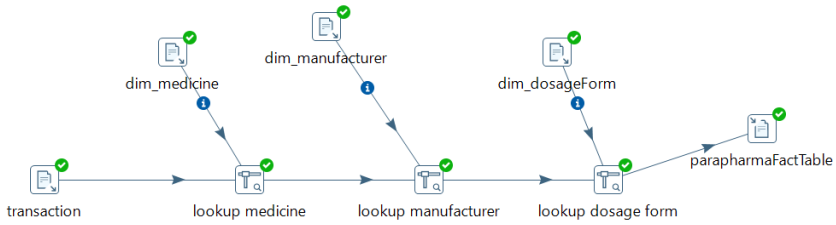


Fig 20. Create the fact table for Parapharma

Figure 20 is the ETL process for creating a fact table. At this stage, the implementation of the star schema at Parapharma companies is carried out. This stage merges several tables based on ID. Dim_mendicine, Dim_manufacturer, and Dim_dosageForm Dimension tables use the step Stream Lookup Value based on ID, and from each table, the fields you want to include in the fact table are selected. For the output of this transformation process, use the txt file output step to save the results into a file with the .csv extension. Figure 21 is the result of the ETL process for creating a Fact Table; this process contains foreign keys from the three dimension tables and several fields of type number so that they can be used later.

	A	B	C	D	E	F	G	H	I
1	brand id	manufacturer id	dosage form id	strength	package price	package size	type	dosage form	manufacturer name
2	4077	2	38	4 mg/5 ml	595	50	allopathic	Syrup	ACI Limited
3	4006	3	109	(10 mg+30 mg+1.25 mg)/5 ml	479	80	allopathic	Syrup	ACME Laboratories Ltd.
4	6174	103	80	500 mg/vial	510	30	allopathic	Injection	Ad-din Pharmaceuticals Ltd.
5	6173	114	60	250 mg/vial	422	100	allopathic	Injection	Aexim Pharmaceuticals Ltd.
6	6172	292	82	125 mg/5 ml	308	100	allopathic	Powder for Suspension	Al-Madina Pharmaceuticals Ltd.
7	6171	4	2	500 mg	586	40	allopathic	Capsule	Alco Pharma Ltd.
8	14006	304	179	6 mg+200 mg+50 mg	710	30	allopathic	Tablet	Alien Pharma
9	3039	245	12	20 mg	365	80	allopathic	Tablet	Alkad Laboratories
10	15483	115	14	50 mg	285	10	allopathic	Tablet	Allied Pharmaceuticals Ltd.
11	14480	5	136	500 mg+400 IU	727	40	allopathic	Tablet	Ambee Pharmaceuticals Ltd.
12	14399	6	105	500 mg+200 IU	298	100	allopathic	Tablet	Amico Laboratories Ltd.
13	14264	7	68	500 mg	488	80	allopathic	Tablet	Amulet Pharmaceuticals Ltd.
14	13582	128	147	100 mg	215	80	allopathic	Tablet	APC Pharma Ltd.
15	25154	8	69	30 mg	746	70	allopathic	Tablet	Apex Pharmaceuticals Ltd.
16	105	9	94	25 mg/ml	243	90	allopathic	IV Injection or Infusion	Apollo Pharmaceutical Ltd.
17	27661	258	6	500 mg	589	80	herbal	Capsule	AqVida bangladesh
18	29715	289	140		583	10	herbal	Syrup	Arges Life Science Limited
19	29714	10	169		518	100	herbal	Syrup	Aristopharma Ltd.
20	30450	81	86		708	50	herbal	Syrup	Asiatic Laboratories Ltd.
21	29646	11	113		467	90	herbal	Capsule	Astra Biopharmaceuticals Ltd.
22	31923	246	65	500 mg	430	20	herbal	Capsule	Aztec Pharmaceuticals Ltd.
23	29754	13	112	300 mg	612	50	herbal	Capsule	Beacon Pharmaceuticals Ltd.
24	29712	110	64		285	60	herbal	Syrup	Belsen Pharmaceuticals Ltd.
25	30288	169	62		641	70	herbal	Syrup	Bengal drugs Ltd.
26	29661	107	126	(1.66 ml+3.3 ml)/5 ml	557	80	herbal	Scalp Solution	Benham Pharmaceuticals Ltd.
27	27660	14	39	(3.03 ml+0.12 gm)/5 ml	442	60	herbal	Syrup	Beximco Pharmaceuticals Ltd.
28	29645	102	93		371	70	herbal	Syrup	Biogen Pharmaceuticals Ltd.
29	28006	15	168	60 mg	402	30	herbal	Capsule	Biopharma Laboratories Ltd.
30	29711	191	138	345 mg	292	90	herbal	Tablet	BOTS Pvt. Limited
31	29644	16	37		724	40	herbal	Syrup	Bristol Pharmaceuticals Ltd.
32	29643	336	160		332	20	herbal	Tablet	Bronson Laboratories (BD) Ltd.
33	29868	273	42	500 mg	208	60	herbal	Capsule	Centeon Pharma Ltd.
34	13947	171	154	200 mg	327	70	herbal	Chewable Tablet	Central Pharmaceuticals Ltd.
35	29642	18	41		593	40	herbal	Capsule	Chemist Laboratories Ltd.
36	29753	174	46		543	100	herbal	Syrup	City Overseas Ltd.
37	29752	19	132		647	20	herbal	Tablet	Concord Pharmaceuticals Ltd.
38	27659	153	43		728	40	herbal	Syrup	Cosmic Pharma Ltd.

pharapharma fact table

Fig 21. Fact table (output from ETL)

3.2 Schema Workbench

Schema Workbench is used in this research to blind them with XML format. Using the workbench schema is necessary to establish a connection to the database containing the data files resulting from the ETL processing that has been done before. For transferring data, you can use PDI or directly use PHPMyAdmin. After connecting the database to the workbench schema, a multidimensional database schema will be created, as shown in Figure 22. In the schema, there is a cube and a fact table; in 3-dimensional, there is a hierarchy containing dimension tables and levels in the form of fields from each table. For the cube, a measure is made to make calculations on a numeric field. When the schema has been created based on the star schema that was created before, the schema will be saved in XML format.

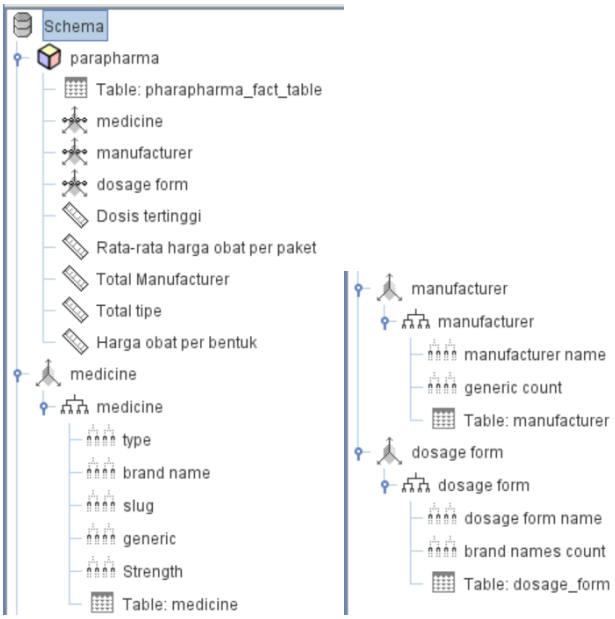


Fig 22. Schema Workbench for Parapharma

3.3 Mondrian

Mondrian is a tool that will be used to run the schema that was created previously in the schema workbench. To display images using MDX queries on .jsp files to generate information based on available data. The following is a diagram created based on data from an MDX query.

	Measures
manufacturer	Total Manufacturer
Abbott Laboratories	27,667
ACI Limited	4,077
ACM laboratoire dermatologique	31,394
ACME Laboratories Ltd.	4,006
Ad-din Pharmaceuticals Ltd.	6,174
Adienne Pharma, Switzerland	33,118
ADM Protexin Ltd	31,091
Aexim Pharmaceuticals Ltd.	6,173
Al-Madina Pharmaceuticals Ltd.	6,172
Alco Pharma Ltd.	6,171
Alien Pharma	14,006
Alkad Laboratories	3,039
Allergan, Inc.	29,730
Allied Pharmaceuticals Ltd.	15,483
Alpha Therapeutic Corporation	29,858
Ambee Pharmaceuticals Ltd.	14,480
Amico Laboratories Ltd.	14,399
Amulet Pharmaceuticals Ltd.	14,264
APC Pharma Ltd.	13,582
Apex Pharmaceuticals Ltd.	25,154
Apollo Pharmaceutical Ltd.	105
AqVida bangladesh	27,661
AqVida GmbH	29,636
Arges Life Science Limited	29,715

	Measures
manufacturer	Total Manufacturer
Aristopharma Ltd.	29,714
Asiatic Laboratories Ltd.	30,450
Astra Biopharmaceuticals Ltd.	29,646
AstraZeneca pharmaceuticals	29,676
Aztec Pharmaceuticals Ltd.	31,923
Baxter	30,311
Beacon Pharmaceuticals Ltd.	29,754
Belsen Pharmaceuticals Ltd.	29,712
Bengal drugs Ltd.	30,288
Benham Pharmaceuticals Ltd.	29,661
Beximco Pharmaceuticals Ltd.	27,660
Biogen Pharmaceuticals Ltd.	29,645
Biopharma Laboratories Ltd.	28,006
Biotest	21,163
Boehringer Ingelheim	27,684
BOTS Pvt. Limited	29,711
Bristol Myers Squibb	13,971
Bristol Pharmaceuticals Ltd.	29,644
Bronson Laboratories (BD) Ltd.	29,643
C.B. Fleet Company, USA	33,154

Fig 23. Sorting total brand by manufacturer

Fig 24. Visualize total brand from manufacturer.

Figure 23 and Figure 24 are tables and visualizations of the total for each medicine brand produced by different laboratories. From this visualization, we can see that each lab produces, on average, more than 10 thousand medicines from various medicine brands.

	Measures
medicine	Total tipe
allopathic	345,826,110
herbal	9,641,516

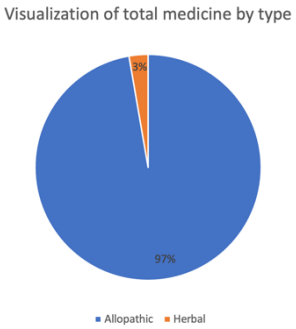


Fig 25. Pie chart and detailed information about medicine type

Figure 25 displays the total medicines by type. The visualization illustrates that the production of modern medicines is more common than herbal medicines. The visualization proves that the use of modern medicine is much more than the use of herbal medicine.

	Measures
dosage form	Harga obat per bentuk
Bolus Tablet	710
Capsule	586
Capsule (Controlled Release)	308
Capsule (Delayed Release)	422
Capsule (Extended Release)	510
Capsule (Modified Release)	479
Capsule (Sustained Release)	595
Capsule (Timed Release)	469
Chewable Tablet	383
Chewing Gum Tablet	722
Cream	416
Dental Gel	526
Dialysis Solution	462
Dispersible Tablet	438
Ear Drop	291
Effervescent Granules	450
Effervescent Powder	465
Effervescent Tablet	278
Emulsion for infusion	247
Eye Capsule	249

	Measures
dosage form	Harga obat per bentuk
Flash Tablet	295
Gel	589
Hand Rub	243
IM Injection	746
IM/IA Injection	215
IM/IV Injection	488
IM/SC Injection	298
Inhalation Capsule	727
Inhaler	285
Injection	365
Intra-articular Injection	226
Intracameral Injection	613
Intraspinal Injection	440
Intratracheal Suspension	385
Intravitreal Injection	586

Fig 26. Average medicine price

Fig 27. Average price by the medicine type

Figures 26 and 27 show the average price for each form of medicine. This visualization shows that the average price of medicines varies widely, but overall, only a few average medicine prices reach 700. Most of the medicines have prices in the range of 200-500.

5. CONCLUSION

The research succeeded in implementing a data warehouse at a pharmacy company so that the available data was more structured and information on drug availability was more detailed. ETL implementation with Star Schema and queries can be displayed on the Online Analytical Process (OLAP), which uses Mondrian to produce data in various visual forms, significantly impacting Parapharma companies in supporting decision-making and drug availability management. Through the implementation of a data warehouse and OLAP, Parapharma Companies can see which laboratories produce many brands of drugs, what types of drugs are most in demand and the average price per form of drug. So that in the future, pharma can adjust the amount of drug production according to the pharmacy's needs.

ACKNOWLEDGEMENT

We extend our heartfelt gratitude to Universitas Multimedia Nusantara for their invaluable support, which played a pivotal role in the successful completion of this research endeavor. Their substantial contribution was instrumental in achieving our objectives, and we are deeply grateful for their unwavering assistance.

REFERENCES

[1] F. Yusuf, A. Farmasi, Y. Tenaga, P. Arjuna, and T. Samosir, "STUDI PERBANDINGAN OBAT GENERIK DAN OBAT DENGAN NAMA DAGANG," *J. FARMANESIA*, vol. 3, no. 1, pp. 5–10, Nov. 2016, Accessed: Dec. 17, 2022. [Online]. Available: http://114.7.97.221/index.php/2/article/view/21

[2] A. Chan, D. Syamsul, M. Farmasi Fakultas Farmasi dan Kesehatan, I. Kesehatan Helvetia, and D. Farmasi Fakultas Farmasi dan Kesehatan, "Gambaran Persepsi Masyarakat Tentang Obat Generik dan Obat Merek Dagang di Daerah Pasar Lam Ateuk Aceh Besar," *J. Dunia Farm.*, vol. 3, no. 2, pp. 91–99, Oct. 2019, Accessed: Dec. 17, 2022. [Online]. Available: http://ejournal.helvetia.ac.id/index.php/jdf/article/view/4478

[3] A. A. Yulianto, "Extract Transform Load (ETL) Process in Distributed Database Academic Data Warehouse," *Aptikom J. Comput. Sci. Inf. Technol.*, vol. 4, no. 2, pp. 61–68, Jul. 2019, doi: 10.11591/APTIKOM.J.CSIT.36.

[4] M. C. Ramadhan, J. Wiratama, and A. A. Permana, "A Prototype Model on Development Of Web-Based Decision Support System For Employee Performance Assessments With Simple Additive," vol. 10, no. 1, pp. 25–32, 2023, doi: 10.30656/jsii.v10i1.6137.

[5] F. Imanuel and F. Tobing, "Implementation of Simple Additive Weighting on Decision Support System for Accoustic Guitar with Web-Based (Study Case: Chaniago Sport)", Ultima InfoSys : Jurnal Ilmu Sistem Informasi, vol. 13, no. 2, pp. 100-109, Jan. 2023.

[6] S. Suryasari, J. Wiratama, and R. Desanti, "The Development of Web-based Sales Reporting Information Systems using Rapid Application Development Method", Ultima InfoSys : Jurnal Ilmu Sistem Informasi, vol. 13, no. 2, pp. 110-116, Jan. 2023.

[7] I. A. Najm *et al.*, "OLAP Mining with Educational Data Mart to Predict Students' Performance," *Informatica*, vol. 46, no. 5, pp. 11–19, Mar. 2022, doi: 10.31449/INF.V46I5.3853.

[8] Oracle, "What is Online Transaction Processing (OLTP) | Oracle," *Oracle*. https://www.oracle.com/database/what-is-oltp/ (accessed Dec. 18, 2022).

[9] H. Sulistiani, S. Setiawansyah, and D. Darwis, "Penerapan Metode Agile untuk Pengembangan Online Analytical Processing (OLAP) pada Data Penjualan (Studi Kasus: CV Adilia Lestari)," *J. CoreIT*, vol. 1, no. 1, pp. 1–1, 2020, Accessed: Dec. 18, 2022. [Online]. Available: https://garuda.kemdikbud.go.id/documents/detail/2018816

[10] R. B. Allen, "Metadata for Social Science Datasets Related papers," in *Rich Search and Discovery for Research Datasets: Building the Next Generation of Scholarly Infrastructure*, 2020.

[11] Y.-J. Chen *et al.*, "A Comparative Analysis of Extract, Transformation and Loading (ETL) Process," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 306, no. 1, p. 012066, Feb. 2018, doi: 10.1088/1757-899X/306/1/012066.

[12] Y. Indarta, D. Irfan, M. Muksir, W. Simatupang, and F. Ranuharja, "Analisis dan Perancangan Database Menggunakan Model Konseptual Data Warehouse Sistem Manajemen Transaksi Toko Online Haransaf," *EDUKATIF J. ILMU Pendidik.*, vol. 3, no. 6, pp. 1–20, Sep. 2021, doi: 10.31004/EDUKATIF.V3I6.1477.

[13] P. Aritonang, M. Johan, and I. Prasetiawan, "Aspect-Based Sentiment Analysis on Application Review using Convolutional Neural Network", Ultima InfoSys : Jurnal Ilmu Sistem Informasi, vol. 13, no. 1, pp. 54-61, Aug. 2022.

[14] K. Nagorny, S. Scholze, A. W. Colombo and J. B. Oliveira, "A DIN Spec 91345 RAMI 4.0 Compliant Data Pipelining Model: An Approach to Support Data Understanding and Data Acquisition in Smart Manufacturing Environments," in IEEE Access, vol. 8, pp. 223114-223129, 2020, doi: 10.1109/ACCESS.2020.3045111.

[15] D. Hammouche, M. Loukam, K. Atif and K. W. Hidouci, "Fuzzy MDX queries for taking into account the ambiguity in querying the baccalaureate data warehouse," 2017 4th International Conference on Control, Decision and Information Technologies (CoDIT), Barcelona, Spain, 2017, pp. 0179-0183, doi: 10.1109/CoDIT.2017.8102587.