Improving The Data Management: ETL Implementation On Data Warehouse At Indonesian Vehicle Insurance Industry

Michael Abhinaya Bagioyuwono[1], Jansen Wiratama[2*]

[1, 2] Information Systems Department, Universitas Multimedia Nusantara,
Tangerang, Banten, 15810, Indonesia
[*]Corresponding Author:
Email: jansen.wiratama@umn.ac.id

Abstract

Unknowingly, risk is an essential part of an individual's life. Every day individuals have the potential to be threatened by possibilities that can produce something that has fatal social, human, and financial consequences. Insurance can help individuals to relieve the financial burden caused by unwanted things by transferring individual losses to insurance companies. This transfer of losses will distinguish individuals from possible bankruptcy and financial security. Automotive insurance is a liability for loss or damage to motorized vehicles. This time we will take the example of an insurance company specializing in automotive insurance, namely Top Gear Insurance (TGI). This company uses customer data storage using an ordinary manual database that does not yet use a data warehouse system. This ordinary manual database makes it difficult for TGI to retrieve data for reprocessing and makes data inaccessible from anywhere, asynchronous, concise, and inefficient. There is a solution to the TGI problem of creating a data warehouse with a star schema approach for storing and processing data. The data warehouse is likely to make the data within the company more accessible, efficient, simple, and understandable so that TGI can develop its business through data analysis from the data it already has. Datawarehouse has many business advantages, such as increasing Business Intelligence, data quality and consistency, saving time, and supporting historical data analysis and queries. The data warehouse consists of datamart, OLTP, OLAP, and Star Schema. Using Mondrian as a visualization showed that TGI can get information about customer data, policies, and claims easily, quickly, and concisely. That can also help TGI create customer profiles and targeted marketing and company evaluation based on the visualization provided.

Keywords: *Data Warehouse, ETL, Mondrian, Star Schema, and Vehicle Insurance.*

I. INTRODUCTION

Without realizing it, risk is an essential part of an individual's life. Every day individuals have the potential to be threatened by possibilities that can produce something that has fatal social, human, and financial consequences That gives rise to interrelated human emotions, namely emotion, fear, and hope. Insurance can help individuals to relieve the financial burden caused by unwanted things by transferring individual losses to insurance companies. This transfer of losses will distinguish individuals from possible bankruptcy and financial security [1]. An insurance policy is an agreement between two parties, namely the party providing the insurance and the insured, which makes a partnership in which the insured will pay a certain amount of money as collateral to protect against an uncertain loss, which can mitigate the higher financial consequences of insurance, given [2]. Insurance, in general, is divided into two parts: health insurance and general insurance [3]. General insurance consists of vehicle, home, goods, and electronics insurance. Automotive insurance is a liability for loss or damage to motorized vehicles. Automotive insurance includes two guarantees, namely losses to individual vehicles and liability, namely legal responsibility to bear third-party losses related to vehicles [4].

Top Gear Insurance is an insurance company specializing in automotive insurance. Richard Hammond founded Top Gear Insurance in the early 1990s and is one of society's most trusted automotive insurance companies. This company covers motor vehicle insurance such as cars, motorcycles, buses, and trucks. This company uses customer data storage using an ordinary manual database that still needs to use a data warehouse system. This manual database makes it difficult for Top Gear Insurance to retrieve data for reprocessing and makes data inaccessible from anywhere, asynchronous, concise, and inefficient.To

overcome the problems that exist in Top Gear Insurance, it is needed to create a data warehouse with a star schema approach for storing and processing data. The data warehouse is required to make the company's data more accessible, efficient, simple, and understandable so that Top Gear Insurance can develop its business through data analysis from the data it already has.

II. STUDY OF LITERATURE

a) Insurance

Insuranciftract or agreement of compensation between two individuals, namely an identity contract, the insurer or the insurer and the party that binds the insured to compensate for the losses contained in the insurance work in a way that the insurer will be willing to cover costs incurred from the insured as long as the insured pays a premium to the insurer and the things covered are things that are still in the contract [5], [6].

b) Vehicle Insurance

Motor vehicle insurance is a contract between an individual and an insurance provider that protects the individual from situations that allow the individual to have a loss in the context of a vehicle, such as an accident or theft. The insurance company will bear all forms of responsibility in the event of an accident or unwanted things on the condition that the individual will pay the premium to the insurance company.Vehicle insurance includes three parts, namely property, damage related to vehicles or objects from individuals, liability related to individual responsibility to make compensation or payments for losses or damage experienced by other people, and medical, which will cover all medical expenses needed when sick caused by accident [7].

c) Data Warehouse

The data warehouse was discovered first by Bill Innnon in 1990 with the information warehouse. The data warehouse is subject-oriented, integrated, time-variant and unchanging data that supports management in making significant decisions within a company or organization [8]. A data warehouse can also collect data integrated with different sources to produce multidimensional data to support corporate decision-making [9]. The creation of a data warehouse can generally be classified into two categories, data-based and needs-based approaches. The data-based approach will analyze the data in the database and convert it into multidimensional data. In contrast, the needs-based approach will analyze organizational needs and create models based on the requirements. Datawarehouse has many business advantages, such as increasing Business Intelligence, data quality and consistency, saving time, and supporting historical data analysis and queries. The data warehouse consists of datamart, OLTP, OLAP, and Star Schema [10].

d) Datamarts

Datamart is part of the data warehouse, which has the contents of a small series of data warehouses that support the needs at the company level, such as departments. Datamart is different from data warehouses because datamart focuses on the needs of company users in a business, does not contain detailed data, and contains less information than data warehouses [11].

e) Online Analytical Processing (OLAP)

E.F. Codd, the inventor of the relational database, first discovered the Online Analytical Processing (OLAP) concept. OLAP technology allows analysts, managers, and executives to access data simultaneously quickly, consistently and efficiently for users [12]. OLAP was created to design and support complex analytical processes to reveal market trends and essential factors in business. Olap has several operations: Drill UP, Drill DOWN, Slicing, and Pivot.

f) Extract, Transform, and Loading (ETL)

In creating a data warehouse, combining all data from different sources is needed using Extraction - Transformation - Loading (ETL). ETL will extract data from different sources and convert the data into a form that suits your needs which will eventually be stored in the data warehouse [13].

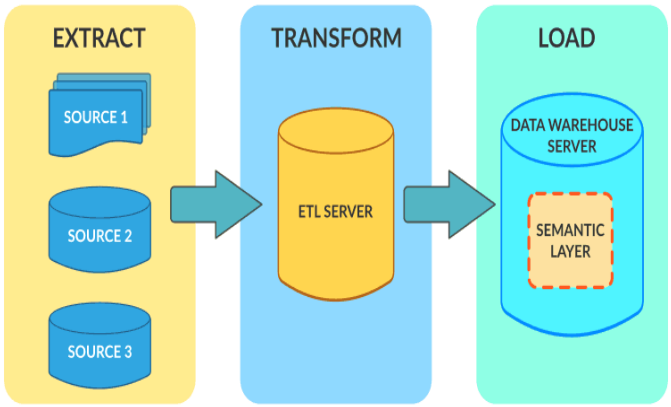

Fig 1. ETL Illustration

The use of ETL is to collect, manipulate and combine relevant data from various company sources, which will produce data that meets the criteria of a data warehouse, namely historical, integrated, encapsulated and static (cannot change) [14].

g) Star Schema

Star schema is a multi-dimensional model that provides an overview of an organizational structure database optimized for use in data warehouses and business intelligence. Star schema was first introduced by Ralph Kimball in 1990 as an efficient schema that can reduce data duplication and make data more aggregated. Data can be filtered in the data warehouse [15].The star schema consists of 2 tables, namely the fact table, which will store transactional data or data that can be measured, and the dimensional table, which will store the attributes of the data. The fact table is a table at the centre of the entire schema, and the dimension table is a supporting table that surrounds the schema [16].

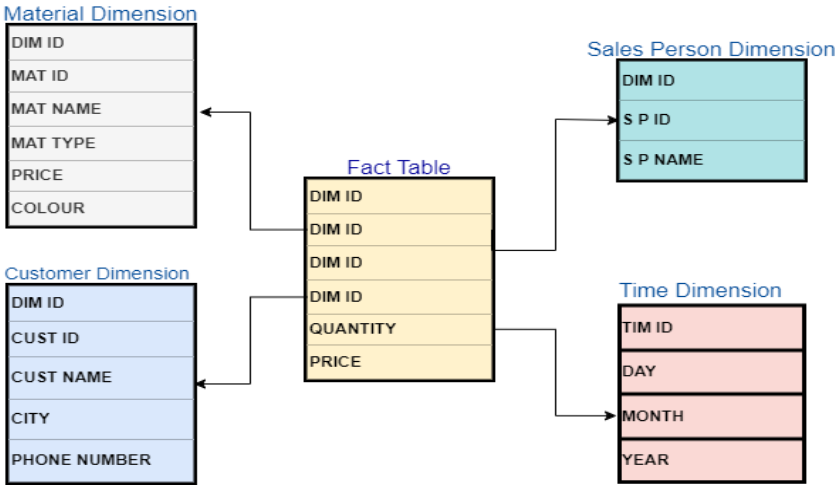

Fig 2. Star Schema Illustration

III. METHODS

Several steps must be carried out in making and designing data warehouses for companies. The first step will be the Extract, Transform, Load (ETL) Extract step to retrieve data from a different database; loadLoad is the transformation of data into appropriate data, and load produces data according to the desired output. From the ETL step that has been made, a Star Schema will be created, which contains the fact table and dimension table, and at the final stage, an output will be made in the form of a report from the existing data.

a) Data collection methods

The type of data that will be used for this research is quantitative data; this is because the data we have is primarily numerical data, which has different amounts, but in the dataset used, there is a lot of categorical data that will be used to graph the company's data output. The research will use secondary data.

Secondary data: The data we have comes from a community website that provides free data collection and usage, namely kaggle.com, with 28,000 data using the following link. https://www.kaggle.com/datasets/girishvutukuri/insurance-fraud. The Data obtained from kaggle.com is in comma-separated values (CSV) format, which will be uploaded to phpMyAdmin using the XAMPP application. The total data tables that will be used to create a data warehouse are three tables, namely train_policy, train_demographic, and train_claim.

b) Data Understanding

The data obtained for this research is customer data in the TopGear Insurance company, which consists of customer demographic data, customer claim data, and data from existing policies for each customer.

Table 1. Data Demographic

Attribute	Type	Data	Section
InsuredGender	String	Male, Female	Data Demographic
InsuredAge	Int	Numeric	
InsuredOccupation	String	Categorical	
SeverityOfIncident	String	Categorical	Data Claim
DateOfIncident	Date	Date	
TypeOfIncident	String	Categorical	
DateOfPolicyCoverage	Date	Date	Policy
InsuredRelationship	String	Categorical	

c) ETL

To create a data warehouse from TopGear Insurance, data cleaning must be carried out so that the data accessed is easier to understand and can be utilized for the company's needs later. To clean and transform data so that it is used to produce graphics for the company, the Extract – Transform – Load (ETL) process must be carried out. The tool that will be used to create an ETL from TopGear Insurance is Pentaho, also known as Pentaho Data Integration Tools (PDI). PDI will be used to create a PDI Transformation to run ETL and a PDI job to run many ETLs simultaneously in a predetermined order.

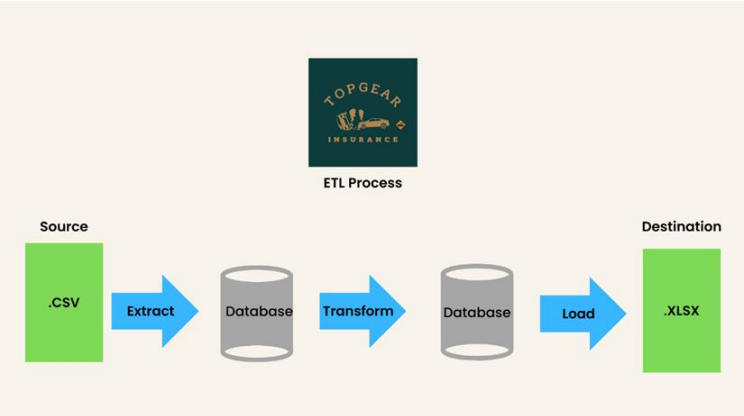

Fig 3. ETL process from TopGear insurance

Extract is the stage of selecting and retrieving data in a company database, which has many formats, for example, .csv, .xlsx, or accurate live data, which is fetched into the database at regular intervals. At the TopGear Insurance transformation stage, data will be collected using Microsoft CSV Input, taken from individual computer files or files from the PHP Myadmin database on the company server.Transform is the stage where data will be changed into more structured data where the data will be normalized. At this stage, the data will be copied, converted according to needs, and reformatted according to the needs of the company database. The quality of the data in the database will be maintained and maintained so that it is not damaged during the transformation process. Load is the final process where data changed by performing ETL will be loaded into the data warehouse. In PDI transformation, the data will be loaded into XLSX format. From the ETL created, a star schema will be made using PDI transformation, forming a fact table whose results will be used to create a workbench schema.

d) Schema Workbench

Schema workbench will be used to create results from star schema designs using UML diagrams. The results are in Figure 4. Created a fact table and dimension table containing train_claim, train_demographic, and train policy. The data taken is stored in the phpMyAdmin database, made with the Xampp application's help.

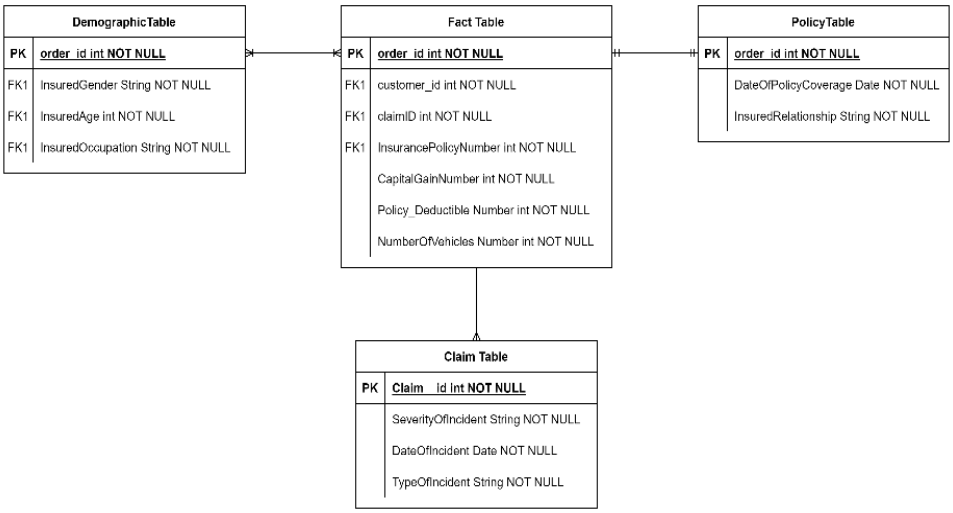

Fig 4. Class Diagram of TopGear Insurance

e) Mondrian

Mondrian will display BI in Online Analytical Processing (OLAP). Mondrian will display a visualization of the query measure carried out in the schema workbench to be communicated to the user in the browser.

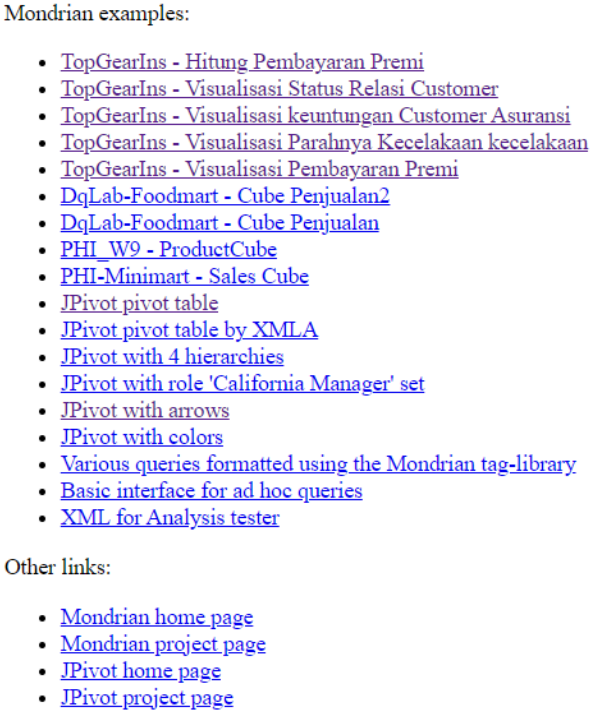

Fig 5. Results from Modrian's creation

IV. RESULTS AND DISCUSSION

In this study, several processes will be carried out to get the desired final result, namely the star schema. First, ETL will be carried out to select data and produce more structured and understandable data. Then, a star schema will be created using Pentaho, which will later be made as a workbench schema to deliver business intelligence visualizations using Mondrian.

a) ETL Process

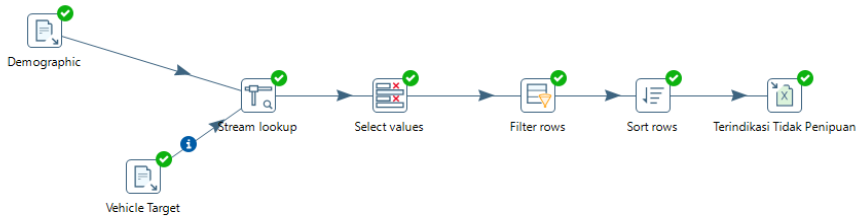

Fig 6. Demographic Distribution Indication of No Fraud

Figure 6 is the process for sorting and filtering rows, which will group indications of customers who do not commit insurance fraud to get money deliberately. First, data will be taken from CSV, which will take two tables from the local database on the computer, namely demographic data and vehicle target data, which contains indications of whether the customer has signs of fraud. Then, the data you want to retrieve will be separated using stream lookup. Then, it will select the values used for the table. After that, row filtering will be carried out for indications of fraud, which will be sorted based on customer ID and saved as a .xlsx output file.

#	CustomerID	Umur Customer	Jenis Kelamin Customer	Education Level Customer	InsuredOccupation	InsuredHobbies	CapitalGains	CapitalLoss	Indikasi Penipuan
1	Cust10000	35	MALE	JD	armed-forces	movies	56700	-48500	N
2	Cust10001	36	MALE	JD	tech-support	cross-fit	70600	-48500	N
3	Cust10002	33	MALE	JD	armed-forces	polo	66400	-63700	N
4	Cust10003	36	MALE	JD	armed-forces	polo	47900	-73400	N
5	Cust10004	29	FEMALE	High School	exec-managerial	dancing	0	-41500	N
6	Cust10005	28	FEMALE	High School	exec-managerial	dancing	0	-41500	N
7	Cust10006	57	MALE	Masters	adm-clerical	sleeping	67400	0	N
8	Cust10007	49	MALE	Masters	adm-clerical	sleeping	67400	0	N
9	Cust10009	27	FEMALE	High School	handlers-cleaners	camping	56400	-32800	N
10	Cust10011	41	FEMALE	MD	prof-specialty	golf	58100	-65600	N
11	Cust10012	36	FEMALE	Associate	craft-repair	paintball	0	0	N
12	Cust10013	36	FEMALE	Associate	craft-repair	base-jumping	0	0	N
13	Cust10014	59	MALE	College	other-service	movies	30400	0	N
14	Cust10016	44	FEMALE	PhD	priv-house-serv	yachting	0	0	N
15	Cust10017	44	FEMALE	PhD	priv-house-serv	basketball	0	-53300	N
16	Cust10018	33	MALE	High School	prof-specialty	polo	0	-29200	N
17	Cust10019	33	MALE	MD	prof-specialty	polo	0	-29200	N
18	Cust10021	40	FEMALE	Masters	exec-managerial	dancing	0	0	N
19	Cust10022	44	FEMALE	Associate	adm-clerical	hiking	73000	-37900	N
20	Cust10023	39	FEMALE	Associate	tech-support	hiking	38700	-37900	N
21	Cust10024	42	MALE	High School	exec-managerial	polo	69400	-41000	N
22	Cust10025	44	MALE	High School	exec-managerial	hiking	69400	0	N
23	Cust10026	60	FEMALE	High School	protective-serv	bungie-jumping	0	0	N
24	Cust10027	57	MALE	High School	protective-serv	exercise	0	-22400	N
25	Cust10028	45	FEMALE	PhD	farming-fishing	board-games	0	0	N
26	Cust10029	44	MALE	PhD	protective-serv	dancing	39000	0	N
27	Cust10030	35	FEMALE	Masters	sales	exercise	37800	0	N
28	Cust10031	38	MALE	Masters	sales	exercise	51500	0	N
29	Cust10032	53	MALE	College	tech-support	sleeping	45700	-65400	N
30	Cust10034	44	FEMALE	Masters	sales	camping	0	-56800	N
31	Cust10035	42	FEMALE	Masters	sales	reading	71300	-70300	N
32	Cust10037	29	FEMALE	Associate	farming-fishing	golf	0	-18600	N
33	Cust10040	40	MALE	JD	transport-moving	sleeping	0	0	N
34	Cust10041	40	MALE	JD	transport-moving	sleeping	0	0	N
35	Cust10042	41	FEMALE	Associate	craft-repair	skydiving	81800	0	N
36	Cust10043	41	FEMALE	Associate	craft-repair	sleeping	81800	0	N
37	Cust10044	52	FEMALE	Masters	exec-managerial	base-jumping	64800	-24300	N
38	Cust10047	27	FEMALE	Masters	protective-serv	board-games	0	0	N
39	Cust10049	42	MALE	High School	prof-specialty	base-jumping	0	0	N
40	Cust1005	44	MALE	Associate	sales	board-games	66000	-46000	N
41	Cust10051	38	FEMALE	College	farming-fishing	skydiving	0	-62500	N

Fig 7. The results of the KTR demographic division are not fraudulent

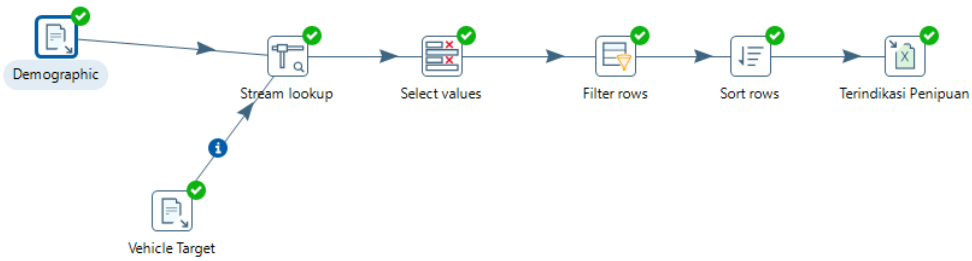

Fig 8. Demographic Distribution of Indications of Fraud

Figure 8 is the process for sorting and filtering rows, which will group indications of customers who commit insurance fraud to get money deliberately. First, data will be taken from CSV, which will take two tables from the local database on the computer, namely demographic data—and vehicle targets, which contain indications of whether the customer has signs of fraud. Then, the data you want to retrieve will be separated using stream lookup. Then, it will select the values used for the table. After that, row filtering will be carried out for indications of fraud, which will be sorted based on customer ID and saved as a .xlsx output file.

#	CustomerID	Umur Customer	Jenis Kelamin Customer	Education Level Customer	InsuredOccupation	InsuredHobbies	CapitalGains	CapitalLoss	Indikasi Penipuan
1	Cust1001	48	MALE	MD	craft-repair	sleeping	53300	0	Y
2	Cust1004	41	FEMALE	PhD	armed-forces	board-games	48900	-62400	Y
3	Cust10046	32	FEMALE	Masters	machine-op-inspct	base-jumping	0	0	Y
4	Cust1006	39	FEMALE	PhD	tech-support	bungie-jumping	0	0	Y
5	Cust10073	27	FEMALE	High School	prof-specialty	kayaking	0	0	Y
6	Cust10084	33	MALE	Masters	handlers-cleaners	exercise	0	0	Y
7	Cust10112	41	FEMALE	College	tech-support	chess	0	0	Y
8	Cust10127	41	FEMALE	JD	transport-moving	sleeping	0	-67400	Y
9	Cust10129	44	FEMALE	PhD	priv-house-serv	yachting	83200	0	Y
10	Cust10131	31	MALE	College	priv-house-serv	exercise	67900	0	Y
11	Cust1015	38	FEMALE	College	machine-op-inspct	board-games	41300	-55500	Y
12	Cust10189	38	FEMALE	High School	farming-fishing	yachting	53000	0	Y
13	Cust1023	55	MALE	High School	prof-specialty	paintball	72700	-68200	Y
14	Cust10239	33	FEMALE	PhD	handlers-cleaners	kayaking	0	-40900	Y
15	Cust1026	35	MALE	Masters	sales	polo	0	0	Y
16	Cust10268	27	FEMALE	JD	other-service	skydiving	31500	0	Y
17	Cust1028	34	MALE	JD	exec-managerial	chess	31000	-30200	Y
18	Cust10289	43	MALE	College	craft-repair	skydiving	23300	0	Y
19	Cust10316	35	FEMALE	High School	transport-moving	board-games	0	0	Y
20	Cust10344	42	MALE	College	craft-repair	exercise	0	0	Y
21	Cust10352	41	MALE	PhD	priv-house-serv	reading	50000	-28700	Y
22	Cust10356	48	MALE	PhD	machine-op-inspct	polo	0	0	Y
23	Cust1036	33	MALE	High School	craft-repair	reading	53300	-49200	Y
24	Cust10369	35	MALE	MD	priv-house-serv	polo	0	-55800	Y
25	Cust1037	28	FEMALE	Masters	protective-serv	camping	0	0	Y
26	Cust10390	36	MALE	College	craft-repair	bungie-jumping	0	-26400	Y
27	Cust1040	37	MALE	JD	tech-support	video-games	48500	0	Y
28	Cust10405	44	FEMALE	Masters	protective-serv	board-games	0	0	Y
29	Cust10407	52	MALE	Masters	transport-moving	kayaking	0	-24400	Y
30	Cust1041	26	MALE	Associate	tech-support	kayaking	0	-55700	Y
31	Cust1042	34	MALE	Associate	transport-moving	chess	0	-24100	Y
32	Cust10433	47	MALE	Masters	sales	sleeping	0	0	Y
33	Cust10448	47	MALE	Associate	adm-clerical	video-games	0	0	Y
34	Cust10474	35	MALE	Masters	machine-op-inspct	basketball	82100	0	Y
35	Cust1048	47	MALE	College	other-service	kayaking	0	0	Y
36	Cust10492	37	MALE	Masters	machine-op-inspct	sleeping	0	-39800	Y
37	Cust10510	31	MALE	JD	other-service	base-jumping	55400	-53700	Y
38	Cust10617	36	FEMALE	MD	protective-serv	camping	49600	-49200	Y
39	Cust10623	31	MALE	College	machine-op-inspct	video-games	51000	0	Y
40	Cust10625	57	MALE	JD	prof-specialty	camping	67600	0	Y
41	Cust1064	42	MALE	JD	transport-moving	video-games	0	-49000	Y

Fig 9. KTR results of the demographic division of fraud

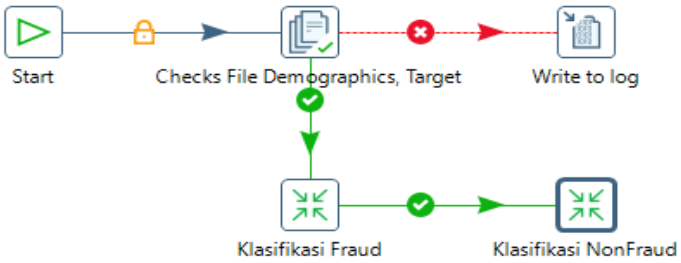

Fig 10. ETL Process For the division of fraud classification

In Figure 10, PDI jobs will run to carry out transformations from customers who have indications of fraud and those who do not. This job has a demographic and target file check before running the job. If there is no file, it will be written to the log.

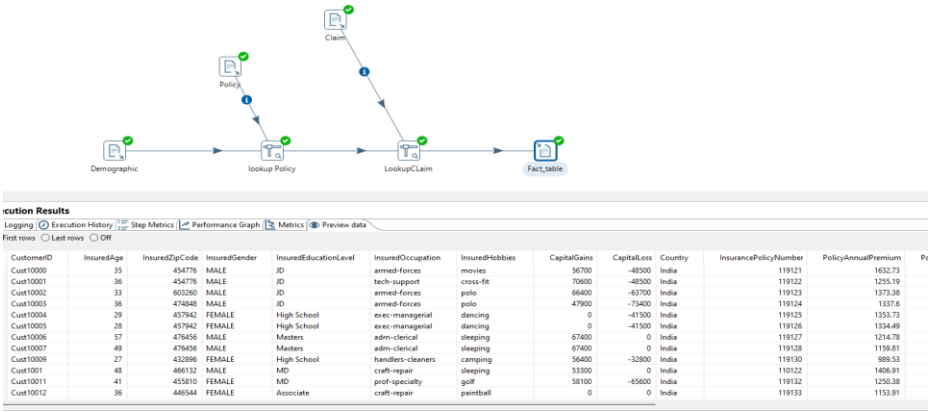

Fig 11. Creating a Star Schema

Figure 11 is an image of creating a star schema from Top Gear Insurance using Pentaho Data Integration. Star Schema will be made to combine several tables, which will become a dimension table, and the main table will produce a fact table. First, each table combined to create a star schema will be called into Pentaho using the CSV File Input. First, demographic tables will be combined with policies using stream lookup; the data collected is on insurance policy numbers in the form of foreign keys, policy annual premiums, and policy deductibles. Then, it will be combined with the claims table, which is combined with the stream lookup value, to retrieve the Claim ID, date of incident, number of vehicles, and authorities contacted. The results of all these combinations will form the fact table of Top Gear Insurance.

	ClaimID	InsurancePolicyNumber	CapitalGains	CapitalLoss	PolicyAnnualPremium	Policy_Deductible	NumberOfVehicles	AuthoritiesContacted
2	22	119121	56700	-48500	16327	1000	3	Police
3	56	119122	70600	-48500	12552	1000	3	Police
4	9	119123	66400	-63700	13734	617	1	Other
5	84	119124	47900	-73400	13376	722	1	Other
6	10	119125	0	-41500	13537	500	1	Fire
7	13	119126	0	-41500	13345	500	1	Fire
8	84	119127	67400	0	12148	512	1	Ambulance
9	49	119128	67400	0	11598	877	1	Police
10	79	119130	56400	-32800	9895	2000	3	Ambulance
11	88	110122	53300	0	14069	1000	1	Police
12	79	119132	58100	-65600	12504	500	1	Fire
13	27	119133	0	0	11539	511	3	Other
14	21	119134	0	0	11748	1044	3	Other
15	53	119135	30400	0	12264	1000	3	Police
16	55	119137	0	0	11079	500	1	Other
17	88	119138	0	-53300	10976	500	1	Other
18	2	119139	0	-29200	13382	1193	1	Police
19	88	119140	0	-29200	1338	1201	1	Police
20	7	119142	0	0	12728	2000	1	None
21	52	119143	73000	-37900	11949	1620	1	Ambulance
22	42	119144	38700	-37900	1408	745	1	Ambulance
23	6	119145	69400	-41000	1283	500	3	Other
24	37	119146	69400	0	13828	500	4	Other
25	100	119147	0	0	13393	1000	2	Other
26	14	119148	0	-22400	13647	1000	3	Fire
27	43	119149	0	0	13891	1000	2	Police
28	84	119150	39000	0	13925	1000	2	Police
29	51	119151	37800	0	11673	2000	1	Fire
30	91	119152	51500	0	13095	2000	1	Fire
31	13	119153	45700	-65400	7403	2000	4	Police
32	89	119155	0	-56800	10173	1000	3	Other
33	71	119156	71300	-70300	12211	1000	3	Ambulance
34	75	119158	0	-18600	12092	654	1	Ambulance
35	1	110125	48900	-62400	14157	2000	1	Police
36	29	119161	0	0	13284	842	3	Police
37	79	119162	0	0	12424	741	3	Ambulance

Fig 12. Contents of the Star Schema

Figure 12 displays the results of the fact table created using Pentaho. In the Schema Workbench, a data warehouse will be created, which will later be used in the model for data visualization. The first thing to do in the workbench is to retrieve data that Etl has done into the TopGearInsurance database; this data is stored in phpMyAdmin, assisted by XAMPP. After that, you can create a multidimensional database. Making Cube has three dimensions; each has a hierarchy containing tables and levels. In the cube section, a dimension usage table will be made, and calculations will be made based on the numerical data in the fact table; this is illustrated in Figure 13.

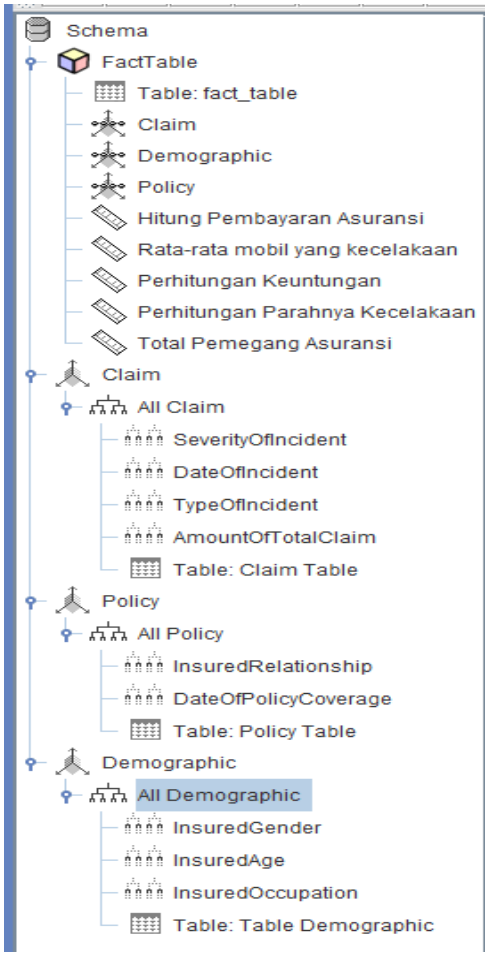

Fig 13. Results from the Workbench schema

The previously created Schemaworkbench will be used in Modrian, using MDX queries to generate tables. From the results of the created workbench, there were tables of the results from Mondrian.

	Measures
All Demographic	Hitung Pembayaran Asuransi
FEMALE	19,516,128.1
MALE	16,649,319.5
NA	38,969

Fig 14. Comparison of Total Premium Payments for Men and Women

	Measures
All Claim	Rata-rata mobil yang kecelakaan
Major Damage	4,041,006
Minor Damage	5,485,281
Total Loss	4,335,131
Trivial Damage	1,339,764

Fig 15. Total Accident Damage Based on Type

	Measures
All Demographic	Perhitungan Keuntungan
FEMALE	364,074,400
MALE	300,411,600
NA	661,600

Fig 16. Company Profit Calculations

	Measures
All Demographic	Total Insurance Holders
FEMALE	15,644
MALE	13,162
NA	30

Fig 17. Total Holders / Users of TopGearInsurance services

	Measures
All Policy	Calculation of Insurance holder
Husband	5,002
Not-in-family	5,222
Other-relative	5,153
Own-child	5,242
Unmarried	3,993
Wife	4,224

Fig 18. Demographic status of insurance holders

Figure 14 is a visualization that depicts insurance payments based on premiums obtained from all Top Gear Insurance customers based on male and female gender. Figure 15 shows the damage to the customer's car when the accident occurred. Damage is divided into three parts: Major Damage, Minor Damage, Total Loss, and trivial damage. Figure 16 is a visualization that depicts the total calculation of profits obtained from Top Gear Insurance customers, with a total profit of around 664 million US Dollars. Figure 17 is a visualization of data on total insurance customers at TopGear Insurance with a total of 15,664 and 13,162 female insurance holders, which indicates that insurance users with female gender use Top Gear Insurance more than men. Figure 18 is a visualization of the relationship between Top Gear insurance holders. From the visualization, insurance holders who have children have the highest number, with 5,242.

V. CONCLUSION

The conclusion drawn from this research is that implementing a data warehouse within a company can help companies make decisions by providing essential and understandable information. Using Mondrian as a visualization showed that Top Gear Insurance could easily, quickly, and concisely get information about customer data, policies, and claims. This can help Top Gear Insurance to create customer profiles and targeted marketing and company evaluation based on the visualization provided.

VI. ACKNOWLEDGMENTS

We would like to acknowledge the invaluable support received from the Big Data Laboratory of the Information Systems Department at Universitas Multimedia Nusantara. Their contribution played a significant role in the successful completion of this research. We extend our sincere gratitude to our colleagues from the Information Systems Department, whose insightful input and expertise greatly assisted us throughout the research process. Their valuable contributions helped shape the direction of this study and enhance its overall quality.

REFERENCES

[1] E. Grant, "The Social and Economic Value of Insurance," *Geneva Assoc.*, 2012.

[2] A. Gepp, J. H. Wilson, K. Kumar, and S. Bhattacharya, "A Comparative Analysis of Decision Trees Vis-`a-vis Other Computational Data Mining Techniques in Automotive Insurance Fraud Detection," ***J. Data Sci.,*** vol. 10, no. 3, pp. 537–561, 2021, doi: 10.6339/jds.201207_10(3).0010.

[3] S. S. Weedige, H. Ouyang, Y. Gao, and Y. Liu, "Decision making in personal insurance: Impact of insurance literacy," *Sustain.*, vol. 11, no. 23, pp. 1–24, 2019, doi: 10.3390/su11236795.

[4] I. Purnama Batubara and R. Syahriza, "Analisis Klaim Asuransi Kendaraan Bermotor pada Pt Asuransi Jasindo Kantor Cabang Medan," ***J. Soc. Res.,*** vol. 1, no. 9, pp. 1026–1031, 2022, doi: 10.55324/josr.v1i9.62.

[5] Lubis, H., & Pratama, K., Safrida, S. (2022). HR related antecedes to sustainability reporting in Indonesian public listed firm: The mediating role of employee committeemen. *Cuadernos de Economía, 45*(128), 87-97.

[6] Purwanto, "Pembaruan Definisi Asuransi dalam Sistem Hukum di Indonesia (Insurance Definition Renewal in Law System in Indonesia)," *Risal. Huk. Fak. Huk. Risal. Huk. Unmul*, vol. 2, no. 2, pp. 87–93, 2006, [Online]. Available:https://webcache.googleusercontent.com/search?q=cache:UulmGj3VXHAJ:https://e-journal.fh.unmul.ac.id/index.php/risalah/article/download/130/80/+&cd=11&hl=id&ct=clnk&gl=id

[7] T. Riasanow, L. Jäntgen, S. Hermes, M. Böhm, and H. Krcmar, "Core, intertwined, and ecosystem-specific clusters in platform ecosystems: analyzing similarities in the digital transformation of the automotive, blockchain, financial, insurance and IIoT industry," *Electron. Mark.*, vol. 31, no. 1, pp. 89–104, 2021, doi: 10.1007/s12525-020-00407-6.

[8] Simamora, R. N. H., & Elviani, S. (2022). Carbon emission disclosure in Indonesia: Viewed from the aspect of board of directors, managerial ownership, and audit committee. ***Journal of Contemporary Accounting,*** 1-9.

[9] N. Nizamuddin and A. Abugabah, "Blockchain for automotive: An insight towards the IPFS blockchain-based auto insurance sector," **Int. J. Electr. Comput. Eng.,** vol. 11, no. 3, pp. 2443–2456, 2021, doi: 10.11591/ijece.v11i3.pp2443-2456.

[10] B. Aadil, L. Kzaz, A. Ait Wakrime, and A. Sekkaki, "Linking context to data warehouse design," ***Int. J. Adv. Comput. Sci. Appl.***, vol. 10, no. 1, pp. 11–20, 2019, doi: 10.14569/IJACSA.2019.0100102.

[11] M. AlMeghari, S. Taha, H. Elmahdy, and X. Shen, "A proposed authentication and group-key distribution model for data warehouse signature, DWS framework," ***Egypt. Informatics J.***, vol. 22, no. 3, pp. 245–255, 2021, doi: 10.1016/j.eij.2020.09.002.

[12] E. Saddad, A. El-Bastawissy, H. M. O. Mokhtar, and M. Hazman, "Lake data warehouse architecture for big data solutions," ***Int. J. Adv. Comput. Sci. Appl.***, vol. 11, no. 8, pp. 417–424, 2020, doi: 10.14569/IJACSA.2020.0110854.

[13] Kurnianingsih, H. T., & Rahayu, S. (2020). Financial Performance Assessed From Economic Value Edded (EVA) and Market Value Added (MVA) Cases in the Cosmetics Sub-Sektor and Household Needs Listed On the BEI. Budapest International Research and Critics Institute, 3(4), 3179-3184.

[14] J. Siahaan, W. Wella, and R. I. Desanti, "Apakah Youtuber Indonesia Kena Bully Netizen?," ***Ultim. InfoSys J. Ilmu Sist. Inf.,*** vol. 11, no. 2, pp. 130–134, 2020, doi: 10.31937/si.v11i2.1764.

[15] A. Vatresia, A. Johar, F. P. Utama, and S. Iryani, "Automated Data Integration of Biodiversity with OLAP and OLTP," *Sisforma*, vol. 7, no. 2, pp. 80–89, 2020, doi: 10.24167/sisforma.v7i2.2817.

[16] S.M.F.Ali and R.Wrembel,"From conceptual design to performance optimization of ETL workflows:current state of research and open problems,"***VLDB J.,***vol.26,no.6, pp.777–801, 2017, doi: 10.1007/s00778-017-0477-2.

[17] Rahayu, S., Zufrizal, Z., Astuty, W., & Triastuti, H. (2020). Keputusan Hedging dan Faktor-Faktor yang Mempengaruhinya pada Perusahaan Pertambangan yang Terdaftar di Indeks Saham Syariah Indonesia (ISSI). Institut Penelitian dan Kritikus Internasional Budapest-Jurnal (BIRCI-Journal) , 3 (4), 2662-2676.

[18] I. Zaelani, "Implementasi Data Mart Terhadap Sistem Penjualan Pada Perusahaan Bidang Distributor Di Pt. Eigen Trimathema," ***J. Penelit. Mhs. Tek. dan Ilmu Komput.***, vol. 1, no. 2, pp. 95–103, 2021, doi: 10.34010/jupiter.v1i2.7309.

[19] M. M. Amin, A. Sutrisman, and Y. Dwitayanti, "Development of Star-Schema Model for Lecturer Performance in Research Activities," ***Int. J. Adv. Comput. Sci. Appl.,*** vol. 12, no. 9, pp. 74–80, 2021, doi: 10.14569/IJACSA.2021.0120909.

[20] Rahayu, S., & Riana, Z. (2020). Determinants of Fraud Pentagon Theory Perspective and Its Effects on Fraudulent Financial Statement in Mining Companies Which Is Listed In Indonesia Stock Exchange. ***Budapest International Research and Critics Institute-Journal (BIRCI-Journal),*** 3(3), 1995-2010.

[21] W. Suharso, A. Fardiansa, Y. Munarko, and H. Wibowo, "Implementasi Star Schema Pada Studi Kasus Perpustakaan Berskala Universitas," ***SINTECH (Science Inf. Technol. J.***, vol. 4, no. 1, pp. 1–11, 2021, doi: 10.31598/sintechjournal.v4i1.446.