

Transformation of Data Warehouse Using Snowflake Scheme Method

Akhmad Dahlan

Department of Informatics Engineering
STMIK AMIKOM Yogyakarta
Yogyakarta, Indonesia
e-mail: alland@amikom.ac.id

Ferry Wahyu Wibowo

Department of Informatics Engineering
STMIK AMIKOM Yogyakarta
Yogyakarta, Indonesia
e-mail: ferry.w.wibowo@ieee.org

Abstract – Process changes transactional database into data warehouse could use some schemes, one of them is snowflake scheme. The use of snowflake scheme aims to yield many data those employ as the material of decision-making system. The different of this scheme with other schemes is splitting fact data into many dimensional data. After that, the next process is doing extracting, transformation, and loading (ETL) process. The stage of the population in ETL process has an important role because it determines the result of data those will be loaded to every dimensional data. To guarantee data those will be populated has a good integrity, so it should be done data cleansing namely erasing a noise data. The successful of data populating and loading would be yielded beneficiary of the managerial level to make a decision.

Keywords - data warehouse; ETL; managerial; populate; snowflake

I. INTRODUCTION

In the information world, the most important asset is data. The existence of data in an organization is used as material to determine policies, build strategies, and take decisions. The available data then processed into information needed by the organization. Data processing could be done in every state, e.g. in the operational database, operational application, or data warehouse technology usage. Many stored data in the operational database are ineffective if it used as material to dig managerial information. The campus library is expected to have a data warehouse so that the process of extracting information to management levels do not interfere with the process of transactions made at any time. Utilization of data warehouse is basically the process that summarizes stored data in the database system done due to a certain time period but it doesn't do at any time do. The existence of a data warehouse is also required as a space of transaction considering it will extend as increasing of students each year.

Supervisors need to get summarized data of transaction data occurred on their departments or known as transactional information system (TPS). At the management information system level, usually, there are summarizing of information resulted by supervisors to synchronize with the supervisors. At the higher level, the information should be presented in such a way to support the decision-making process or this level known as decision support systems. The highest information is necessarily available for library manager needs and the summarized information are much detail in such a way and the availability of information provided in the mechanism named by executive information systems.

At the service operational level, to do a transaction process is using transactional database systems. In this database would store all detail information for all activities and transactions. The stored data will always be dynamic as

increasing transactions in the database. Although at the information managerial level, commonly needs are aggregation results or summary of the transactional database. To present data needed by managerial level, existing of the data warehouse is very important to do. The most important thing in the process of building data warehouse is the process of modeling [1]. Because from the designed data model would be used to determine the capability of a data warehouse in providing managerial information needs.

Modeling of data warehouse could be done in such a way methods e.g. star and snowflake scheme methods. In the star scheme method, a dimensional table doesn't have sub-table or sub-dimension so the data referred as material to result information are not much. In the snowflake scheme method, every dimension table could have another dimension sub-table. This aims to minimize redundancy data. This data dimension is being information subject employed as the material in decision-making because every data dimension could be considered derivating into more detail. So the data source could be processed into huge and detail information. This scheme method is advancing method of star scheme namely method of normalization for dimension tables existed in the star scheme [2].

Based on this background, the issue of the database is still important to manage. A data model using snowflake scheme method to build a data warehouse of the library of campus XYZ has been done by [3]. In this paper focuses on the process of extraction, transformation and loading (ETL) in the data warehouse which limited on designing a data warehouse of the library of campus XYZ using snowflake scheme method, building a data warehouse that covers ETL process and cube and arranging a reporting model as a library management information systems.

II. RELATED WORKS

A. Data Warehouse

Data warehouse or also known as enterprise data warehouse is a set of subject-oriented data, integrated data, non-updateable, time-variant, historical information and support decision-making [4]. The data warehouse is made to prevent the technical and business issues in the usage of data and information to make a decision [5]. The data warehouse is also a system that aims to take and consolidate data periodically from a source system to the dimension table or storage data normalized. Generally, the data are renewed for such time not once transaction in the source system [6]. The data warehouse is used for reporting and data analysis for such knowledge.

B. Snowflake Scheme

Snowflake scheme method is a method of data modeling applied on the data warehouse. It is developed from method of star scheme. Those methods have trade-off in the implementation [7, 8]. The snowflake scheme is mixing of some data modeling of star scheme normalized. In the snowflake scheme, each dimension table may have another dimension sub-table. The data dimension has an information subject that is functioned as the material in making a decision because in each data dimension is possible to generate a description more detail. So that, the data source could be processed becomes information in a huge number and more detail. The snowflake scheme is shown in Figure 1.

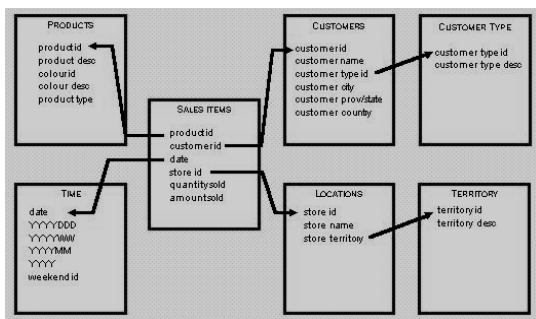


Figure 1. Snowflake Scheme.

Some advantages and disadvantages of the use of snowflake scheme method could be written as bellows [9],

Advantages:

1. Dimension table process divided into several sub-dimension that could cause stored media efficiency. This is very useful because data warehouse would contain a lot of data so fewer storage media could help data warehouse management process.
2. A lot of sub-dimension could result in a lot of information. The higher point of view, the higher information will be gotten as needed.

Disadvantages:

1. Query process would be complicated because it consists of many inter-correlated sub-dimension.

2. Loading data process should be managed and considered to prevent normal update or anomaly insert.

C. ETL

Extraction, transformation, and loading (ETL) are a process to pull a data out from the source system and place it into the data warehouse [10]. In the process of managing a data, the data should be filtered and fixed in order to get a good data quality. The invalid data should be cleaned. In the ETL processes, approaching of traditional is retrieving a data from a data source, and putting it in a staging area, thus transforming and loading it into a data warehouse [6].

III. METHODOLOGY AND RESULTS

The methodologies in this section are defining subjects of data warehouse, defining needs and data collection, making a design of snowflake scheme and designing a process of ETL, making a cube and dimensional for data warehouse, analyzing data warehouse using reporting, and testing on the data warehouse. The process could be depicted as shown in Figure 2.

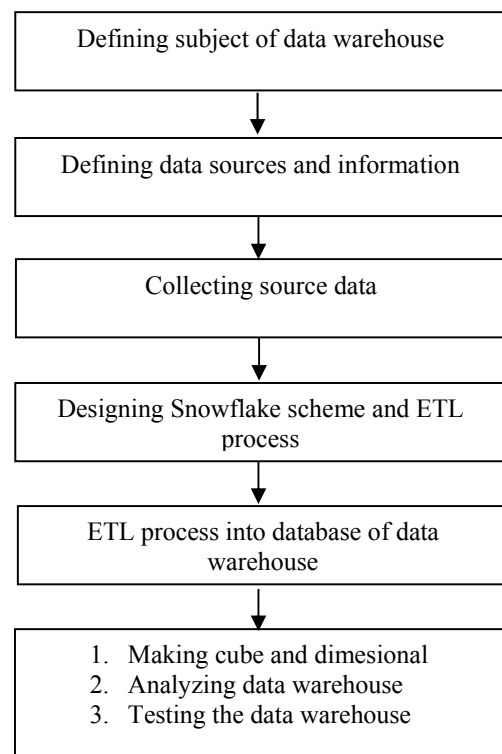


Figure 2. Methodology Processes.

A. Define Data Subject of Warehouse

A characteristic of the data warehouse is subject-oriented, so the first level in making data warehouse is defining its subject. In the planning stage, the main subject in the library database is a process of book circulation that student could

borrow books. This process stage describes the flow of borrowing and returning books by students.

B. Define Data Source and Information Requirement

This process stage defines all data sources which would be implemented in the data warehouse and defines information requirements those have to be resulted by the data warehouse via reporting analysis to encourage a management decision. The data source that would be processed to be used in the data warehouse obtained from the library transactional database at such periods. In this case, the obtained data are library transactional data for ten years. The data collection method that has been done are doing an observation to the transactional database and studying documents required by library management consisting of inter-table relation in the database of campus XYZ library, data at every table in the library’s database, and kinds of report and decision those needed by library management.

Book table of database consists of data of book details based on book titles interpreted by a column of the international standard book number (ISBN) as a primary key. The book table of the database is shown in Figure 3.

Column Name	Data Type	Allow Nulls
isbn	nvarchar(15)	<input type="checkbox"/>
judul	nvarchar(200)	<input checked="" type="checkbox"/>
jenis	nvarchar(20)	<input checked="" type="checkbox"/>
th_terbit	int	<input checked="" type="checkbox"/>
edisi	nvarchar(2)	<input checked="" type="checkbox"/>
bahasa	nvarchar(10)	<input checked="" type="checkbox"/>
sampul	image	<input checked="" type="checkbox"/>
sinopsis	nvarchar(MAX)	<input checked="" type="checkbox"/>
jumlah	decimal(18, 0)	<input checked="" type="checkbox"/>
jumlahada	decimal(18, 0)	<input checked="" type="checkbox"/>
jurnal	nvarchar(1)	<input checked="" type="checkbox"/>
tgl	datetime	<input checked="" type="checkbox"/>
tgl_edit	datetime	<input checked="" type="checkbox"/>
is_akreditasi	bit	<input type="checkbox"/>
akreditasi	nvarchar(MAX)	<input checked="" type="checkbox"/>

Figure 3. The Design of dbo_BUKU Table.

Different with book table, book data table contains specific data for each book copy because each book title could have some number of book copy. The design of book data table is shown in Figure 4.

Column Name	Data Type	Allow Nulls
kd_buku	nvarchar(25)	<input type="checkbox"/>
isbn	nvarchar(15)	<input checked="" type="checkbox"/>
tgl	datetime	<input checked="" type="checkbox"/>
lokasi	nvarchar(5)	<input checked="" type="checkbox"/>
sumber	smallint	<input checked="" type="checkbox"/>
status	nvarchar(1)	<input checked="" type="checkbox"/>
TGL_EDIT	datetime	<input checked="" type="checkbox"/>

Figure 4. The Design of dbo_DATA_BUKU Table.

Database of book type table consists of data of categories of material contains type of each book title. The design of book type table is shown in Figure 5.

Column Name	Data Type	Allow Nulls
no_jenis	nvarchar(20)	<input type="checkbox"/>
njenis	nvarchar(MAX)	<input checked="" type="checkbox"/>

Figure 5. The Design of dbo_JENIS Table.

Publisher table consists of book publisher data. The design of publisher table is shown in Figure 6.

Column Name	Data Type	Allow Nulls
kd_terbit	decimal(18, 0)	<input checked="" type="checkbox"/>
npenerbit	nvarchar(100)	<input checked="" type="checkbox"/>
alamat	nvarchar(100)	<input checked="" type="checkbox"/>

Figure 6. The Design of dbo_PENERBIT Table.

Each book author data is stored in author table. The design of author table is shown in Figure 7.

Column Name	Data Type	Allow Nulls
kd_karang	decimal(5, 0)	<input type="checkbox"/>
namad	nvarchar(100)	<input checked="" type="checkbox"/>
namab	nvarchar(100)	<input checked="" type="checkbox"/>

Figure 7. The Design of dbo_PENGARANG Table.

There are two tables of book borrowing. Firstly, table of book borrowing is especially used to store borrowing data detail done by every student (see Figure 8). Secondly, the table that is used to store borrowing data detail done by lecturers and staffs (see Figure 9).

Column Name	Data Type	Allow Nulls
kd_pinjam	decimal(18, 0)	<input type="checkbox"/>
kd_buku	nvarchar(25)	<input type="checkbox"/>
denda	money	<input checked="" type="checkbox"/>

Figure 8. The Design of dbo_PINJAM_MAHASISWA Table.

Column Name	Data Type	Allow Nulls
kd_pinjam	decimal(18, 0)	<input type="checkbox"/>
kd_buku	nvarchar(25)	<input type="checkbox"/>
tgl_kembali	datetime	<input checked="" type="checkbox"/>

Figure 9. The Design of dbo_PINJAM_DOSEN Table.

There are also two tables of circulation. Firstly, the table that is used to store borrowing data done by students (see Figure 10). Secondly, the table that is used to store borrowing data done by lecturers or staffs (see Figure 11).

Column Name	Data Type	Allow Nulls
kd_pinjam	decimal(18, 0)	<input type="checkbox"/>
nim	nvarchar(10)	<input type="checkbox"/>
tgl_pinjam	datetime	<input checked="" type="checkbox"/>
tgl_kembali	datetime	<input checked="" type="checkbox"/>
disc_denda	money	<input checked="" type="checkbox"/>
tgl_server	datetime	<input checked="" type="checkbox"/>

Figure 10. The Design of dbo_SIRKULASI1 Table.

Column Name	Data Type	Allow Nulls
kode_pnj	decimal(18, 0)	<input type="checkbox"/>
kd_dsn	nvarchar(20)	<input checked="" type="checkbox"/>
tgl_pinjam	datetime	<input checked="" type="checkbox"/>
tgl_server	datetime	<input checked="" type="checkbox"/>

Figure 11. The Design of dbo_SIRKULASI2 Table.

The table that contains source data of library collection providing is shown in Figure 12.

Column Name	Data Type	Allow Nulls
kd_sum	smallint	<input type="checkbox"/>
nnumber	nvarchar(50)	<input checked="" type="checkbox"/>

Figure 12. The Design of dbo_SUMBER Table.

Student table used as research materials consists of column of student ID number (NPM), Name, and sex. This table is shown in Figure 13.

Column Name	Data Type	Allow Nulls
NPM	nvarchar(10)	<input type="checkbox"/>
NAMA	nvarchar(50)	<input checked="" type="checkbox"/>
JENIS	nvarchar(1)	<input checked="" type="checkbox"/>

Figure 13. The Design of dbo_MAHASISWA Table.

Publish table consists of institution data which are publishing books. This table is shown in Figure 14.

Column Name	Data Type	Allow Nulls
id_terbit	decimal(18, 0)	<input type="checkbox"/>
isbn	nvarchar(15)	<input checked="" type="checkbox"/>

Figure 14. The Design of dbo_TERBIT Table.

The last table is author table which consists of author data which are publishing books. This table is shown in Figure 15.

Column Name	Data Type	Allow Nulls
id_karang	decimal(5, 0)	<input type="checkbox"/>
isbn	nvarchar(15)	<input checked="" type="checkbox"/>

Figure 15. The Design of dbo_KARANG Table.

The relation tables of library database are shown in Figure 16.

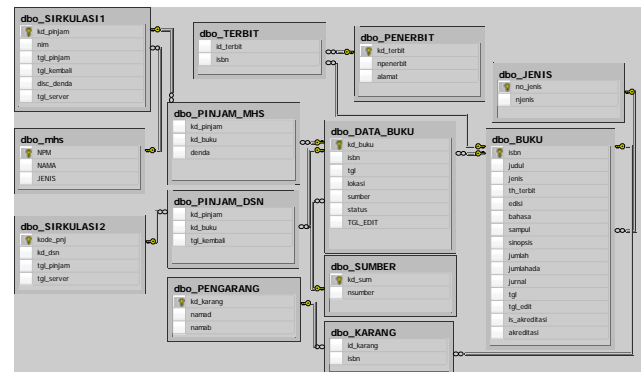


Figure 16. The Relation Tables of Library Database.

The requirements of information in the library those are written as follows,

- Analysis of books data which is borrowed by students
- Analysis of circulation related with the multidimensional tables
- Data students who are most frequently borrowed books based on a multidimensional tables
- Information about types of books which are most frequently borrowed
- Information on the author names of the books which are most frequently borrowed
- Information on the publisher names of the books which are most frequently borrowed
- Time series analysis or circulation process trends

C. Design a Snowflake Scheme and ETL Processes

This stage is making a dimensional data model that forms snowflake scheme and a design of ETL processes. Data warehouse is using dimensional data model. One of the methods is snowflake scheme. The snowflake scheme has 2 parts, namely fact table and dimension table which could have other dimension tables. After snowflake scheme has been built, so the next process is designing extraction, transformation and loading (ETL) process from the operational database to the data warehouse. The main subject in the library warehouse is the circulation of borrowing books for students that would be the main of the scheme. In this scheme, the name of the fact table that was implemented is Fact_StudentLibrary whereas the other data those have been implemented into dimension and sub-dimension tables are the tables related to the fact table. There are 3 dimensional tables, namely:

- Table dim_time (time_dimensional) has 2 sub-dimensional tables, i.e. table day_lookup and table month_lookup
- Table dim_databuku has 2 sub-dimensional tables, i.e. table dim_sumber and table dim_buku. Table dim_buku has 3 sub-dimensional, i.e. table dim_pengarang, Table dim_penerbit, and table dim_jenis.

3. Table `dim_mahasiswa`
The snowflake scheme design model is shown in Figure 17.

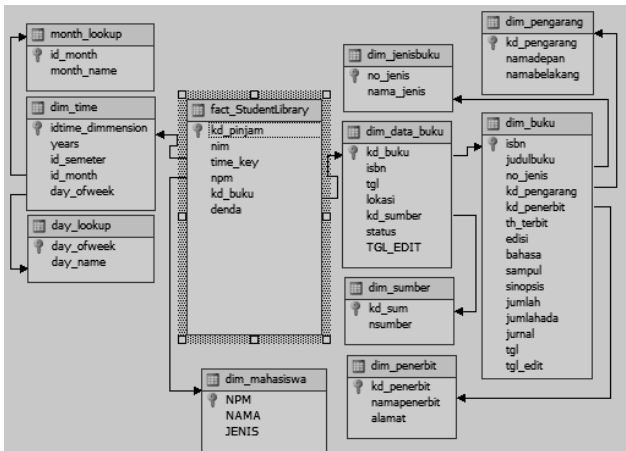


Figure 17. The Snowflake Scheme of Data Warehouse.

D. ETL Processes In Data Warehouse Database

Periodically ETL process extracts data from the source system and transforms into a general format and then load it into data store target, data warehouse or data mart. ETL is very important to integrate data and data warehousing. Depict of ETL process is shown in Figure 18.

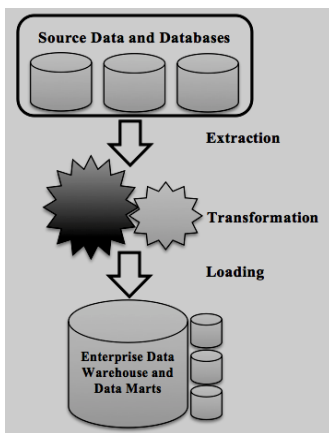


Figure 18. ETL Process.

In the processes of ETL from the library database to the data warehouse database, there are a few things must be done. Generally divided into 2 stages i.e.

1) Omitting a Noise Data

It is necessary to omit the noise data which exist in the source database or online transaction processing (OLTP) database before the ETL processes has been done. This process is necessarily to be done in the case of maintain the integrity of data which would be processed. When the database analysis processes have been done, it would be found a lot of noise data existed in the source database. Example for this case, there were 58 rows in the table of

`dbo_KARANG` which was inconsistent. It could be seen from ISBN data in the table of `dbo_KARANG`, although ISBN data didn't exist in the reference table namely table of `dbo_BUKU`. Omitting of process proof of noise data is shown in Figure 19.

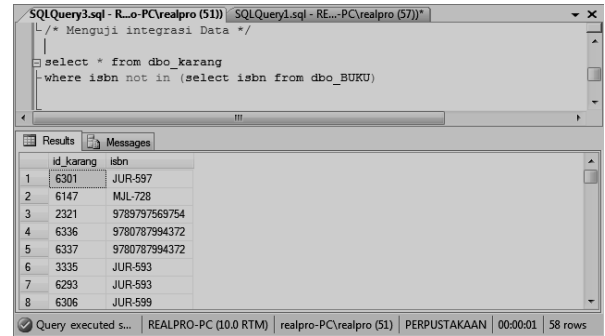


Figure 19. Testing on Data Consistency.

2) Process of Data Population

The extracting data process from the library operational database to data warehouse database applied a population method. This process contains method in populating data become dimension and fact tables. The design of data population process from the operational database of dimension and fact tables is shown in Figure 20. The design of dimension tables defining dimension table population flows for book table, author table, publisher table, source table, type table, and student table has been done via OLTP database.

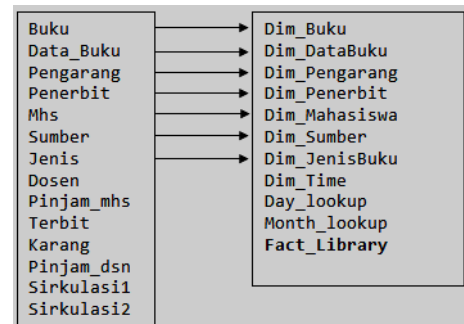


Figure 20. Data Population Process.

After the mapping process for all the tables has been done, the data transformation services (DTS) process would be run until all data is distributed. The DTS process indication has been done is shown in Figure 21.

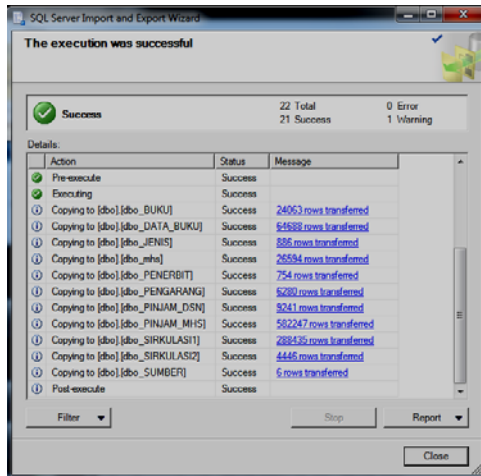


Figure 21. An Indication of Successful DTS Process.

Thus, it would run the periodic loading process at each dimension table. The periodic loading process was conducted periodically based on the time required for developing data in OLTP database. These processes were using structured query language (SQL) transact commands in the database of the data warehouse. The commands of the transact SQL are shown as follows,

```
/* 1- Data Population of dim_jenisbuku */
insert into dim_jenisbuku
select * from DBPerpus.dbo.dbo_JENIS
where DBPerpus.dbo.dbo_JENIS.no_jenis <>
dim_jenisbuku.no_jenis
```

```
/* 2- Data Population of dim_penerbit */
insert into dim_penerbit
select * from DBPerpus.dbo.dbo_PENERBIT
where DBPerpus.dbo.dbo_PENERBIT.kd_terbit <>
dim_penerbit.kd_penerbit
```

```
/* 3- Data Population of dim_Pengarang */
insert into dim_pengarang
select * from DBPerpus.dbo.dbo_PENGARANG
where DBPerpus.dbo.dbo_PENGARANG.kd_karang <>
dim_pengarang.kd_pengarang
```

```
/* 4- Data Population of dim_mahasiswa*/
insert into dim_mahasiswa
select * from DBPerpus.dbo.dbo_mhs
where DBPerpus.dbo.dbo_mhs.NPM <>
dim_mahasiswa.NPM
```

```
/* 5- Data Population of dim_Sumber */
insert into dim_sumber
select * from DBPerpus.dbo.dbo_SUMBER
where DBPerpus.dbo.dbo_SUMBER.kd_sum <>
dim_sumber.kd_sum
```

```
/* 6- Data Population of dim_buku*/
insert into dim_buku
select [isbn],[judul],[jenis]
```

```
,[th_terbit],[edisi]
,[bahasa],[sampul]
,[sinopsis],[jumlah]
,[jumlahada],[jurnal]
,[tgl],[tgl_edit]
FROM DBPerpus.dbo.dbo_BUKU
Where DBPerpus.dbo.dbo_BUKU.isbn <> dim_buku.isbn
```

```
/* 7- Data Population of dim_data_buku*/
insert into dim_data_buku
select * from DBPerpus.dbo.dbo_DATA_BUKU
where DBPerpus.dbo.dbo_DATA_BUKU.kd_buku <>
dim_data_buku.kd_buku
```

E. Making Cube and Dimensional for Data Warehouse

A multidimensional database is a form of a database which data stored in cells and position every cell determined by a number of dimensions called as the hierarchy. Every cell is an event and value of dimension that shows when and where the event occurs [6]. The popular conceptual model used for front-end tools, database design, and query for online analytical processing (OLAP) is multidimensional in a data warehouse. In multidimensional data model contains numerical size set used for analysis objects. Figure 22 shows depict of popular multidimensional database known as a cube with 3 dimensions of time, product, and customer. The combination of three dimensions depicted as a cell. Each cell performs the business event. The value of that dimension shows when and where the event occurs. Cell A consists of 3 values of revenue, cost and profit [6].

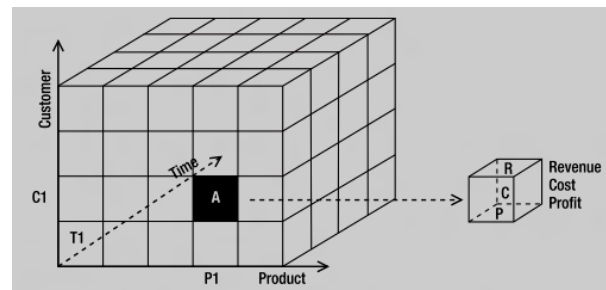


Figure 22. Three Dimension of Multidimensional Database [6].

IV. DISCUSSIONS

Analysis for the data warehouse is applied in reporting of the business requirement. The reporting has been generated using commands of transact SQL. Some of the reports could be generated as follows:


A. Report of The Borrowing Number Yearly

This report consists of information about the management of the library related to the number of borrowing books per year as shown in Figure 23.

```
--/* REPORT OF THE NUMBER OF BORROWING YEARLY */
SELECT dim_time.Year,
```

```

COUNT(kd_pinjam) AS jml_peminjam
FROM fact_StudentLibrary JOIN dim_time
ON fact_StudentLibrary.time_key =
dim_time.DateKey
GROUP BY dim_time.Year ORDER BY dim_time.Year
DESC
    
```



Year	jml_peminjam
2013	2330
2012	31207
2011	39330
2010	39477
2009	48422
2008	61162
2007	71305
2006	74766
2005	61680
2004	46995
2003	12385
2002	36
2001	24

Figure 23. Report of The Number of Borrowing Books Per Year.

B. Report of Best Seller Books

This report consists of information about the book titles and author’s names, in this case the books are selected in the category of best seller books. By using this report, the library management can determine the policy in the process of providing books. The commands of SQL transact for this case are shown as follows,

```

/* REPORT OF BEST SELLER BOOKS */
SELECT
dim_buku.isbn,
dim_buku.judul,
dim_pengarang.namadepan + dim_pengarang.namadepan
AS Pengarang,
COUNT(dim_buku.isbn) AS jml_dipinjam
FROM dim_buku
INNER JOIN
fact_StudentLibrary ON dim_buku.isbn =
fact_StudentLibrary.isbn
INNER JOIN
dim_pengarang ON fact_StudentLibrary.kd_karang =
dim_pengarang.kd_karang
GROUP BY dim_buku.isbn, dim_buku.judul,
dim_pengarang.namadepan,
dim_pengarang.namabelakang
ORDER BY jml_dipinjam DESC
    
```

C. Report of Days At Most Borrowers

This report consists of information on the busiest days within borrowing process. From this report, the library management may anticipate the moments when the most crowded process of circulation, so the librarian may serve them better. Example for this case is the determination of the time schedule of librarian shift, so the librarian should

prepare to serve when in the busiest day because a lot of circulation process. The commands of SQL transact for this are shown as follows,

```

/* REPORT OF BUSIEST DAYS */
SELECT dim_date.DayofWeek AS hari,
COUNT(fact_StudentLibrary.kd_pinjam) AS
jml_pinjam
FROM fact_StudentLibrary INNER JOIN
dim_time ON fact_StudentLibrary.time_key =
dim_time.idtime_dimension
GROUP BY dim_time.DayofWeek
ORDER BY jml_pinjam DESC
    
```

D. Report of Authors of The Best-Seller Books

This report gives information on names of the author of best-selling where books authored often borrowed by library patrons. The SQL commands for this shown as follows,

```

/*REPORT OF BEST SELLING AUTHORS */
SELECT dim_pengarang.namadepan +
dim_pengarang.namabelakang AS pengarang,
COUNT(fact_StudentLibrary.kd_pinjam) AS
jml_pinjam
FROM dim_buku
INNER JOIN
fact_StudentLibrary ON dim_buku.isbn =
fact_StudentLibrary.isbn
INNER JOIN
dim_pengarang ON fact_StudentLibrary.kd_karang =
dim_pengarang.kd_karang
GROUP BY dim_pengarang.namadepan,
dim_pengarang.namabelakang
ORDER BY jml_pinjam DESC
    
```

E. Report of Publishers of The Best-Seller Books

This report consists of information on the publisher names of books and more often borrowed books by library patrons. The commands of SQL transact for this is shown as follows,

```

--/* REPORT OF BEST SELLING PUBLISHER */
SELECT dim_penerbit.namapenerbit as [Nama
Penerbit],
COUNT(fact_StudentLibrary.kd_pinjam) AS [Jumlah
Peminjaman]
FROM fact_StudentLibrary INNER JOIN
dim_penerbit ON fact_StudentLibrary.kd_penerbit=
dim_penerbit.kd_penerbit
GROUP BY dim_penerbit.namapenerbit
ORDER BY [Jumlah Peminjaman] DESC
    
```

Depict of the report each publisher is shown in Figure 24.

	Nama Penerbit	Jumlah Peminjaman
1	andi offset	179590
2	elex media komputindo	105923
3	Infomatika	25885
4	Andi - Amikom Yogyakarta	22924
5	Gramedia	15450
6	Erlangga	10230
7	Prenhallindo	8329
8	Gava Media	7568
9	Salemba Infotek	6798
10	Mizan	6445
11	BPFE	6042
12	Graha Ilmu	5670
13	UPT Penerbitan AMIKOM	5044
14	Wahana Komputer	3897
15	Microsoft	3201
16	Salemba Empat	3077

Figure 24. Report of Each Publisher.

F. Testing on The Data Warehouse

In building a data warehouse, the testing process is required. In this paper, there are two methods has been implemented, namely the ETL testing and functional testing. The reasons for using those are the ETL process spent 60% up to 80% in the process of building a data warehouse and to check that the data warehouse is built in conformity with the business requirement requested by the library management.

The ETL process is ensuring the loading process has been done entered to the appropriate tables and columns and at the right time. The process carried out is to match if the number of query result records in the operational database is equal to the number of records resulted from the ETL process. The ETL process for the fact table is the most important because it will be roled as main of the data warehouse. The SQL command for the ETL process could be seen as follows,

```

INSERT INTO dbo.fact_StudentLibrary
SELECT dbo_SIRKULASI1.kd_pinjam,
[DWPerpus].dbo.datedim.datekey, dbo_mhs.NPM,
dbo_DATA_BUKU.kd_buku,
dbo_PENGARANG.kd_karang, dbo_PENERBIT.kd_terbit,
dbo_PINJAM_MHS.denda
from dbo_PENGARANG
join dbo_KARANG on
dbo_PENGARANG.kd_karang=dbo_KARANG.id_karang
join dbo_BUKU on dbo_BUKU.isbn=dbo_KARANG.isbn
join dbo_TERBIT on dbo_BUKU.isbn=dbo_TERBIT.isbn
join dbo_PENERBIT on
dbo_PENERBIT.kd_terbit=dbo_TERBIT.id_terbit
join dbo_DATA_BUKU on
dbo_BUKU.isbn=dbo_DATA_BUKU.isbn
join dbo_JENIS on
dbo_JENIS.no_jenis=dbo_BUKU.jenis
join dbo_SUMBER on
dbo_SUMBER.kd_sum=dbo_DATA_BUKU.sumber
join dbo_PINJAM_MHS on
dbo_PINJAM_MHS.kd_buku=dbo_DATA_BUKU.kd_buku
    
```

```

join dbo_SIRKULASI1 on
dbo_PINJAM_MHS.kd_pinjam=dbo_SIRKULASI1.kd_pinjam
join dbo_mhs on dbo_mhs.npm=dbo_SIRKULASI1.nim
join [DWPerpus].dbo.datedim on
[DWPerpus].dbo.datedim.actualdate=dbo_SIRKULASI1.
tgl_pinjam
WHERE dbo_mhs.npm != 'xx.xx.xxxx'
ORDER BY dbo_SIRKULASI1.kd_pinjam
go
    
```

The success of a fact table of the ETL process is shown in Figure 25. It shows populating success about 489,119 rows of data.

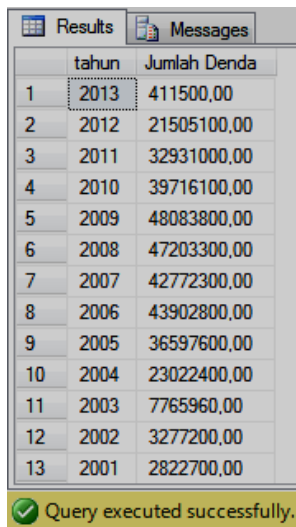
Figure 25. The Success of The ETL Testing.

The functional testing is to check whether the data warehouse has been built in accordance with a business requirement or not. The step for this is analyzing the data warehouse with the report models and the results are matched with business requirements at the stage of defining needs. Example for the report model of loan amount each month is shown in Figure 26.

	Bulan	Jumlah Peminjaman
1	Maret	54444
2	April	51606
3	Januari	51111
4	Desember	48856
5	Mei	47648
6	Oktober	46312
7	November	40951
8	Juni	37183
9	September	34885
10	Februari	33692
11	Juli	22587
12	Agustus	19844

Figure 26. The Report Model of Loan Amount Each Month.

From Figure 26, the managerial party could draw the conclusion that in the months approaching graduation time, many lending transactions occur. While in the months near the turn of the academic year, there were slight lending transactions. Another example for report model of the number of admissions in fines each year is shown in Figure 27.



	tahun	Jumlah Denda
1	2013	411500,00
2	2012	21505100,00
3	2011	32931000,00
4	2010	39716100,00
5	2009	48083800,00
6	2008	47203300,00
7	2007	42772300,00
8	2006	43902800,00
9	2005	36597600,00
10	2004	23022400,00
11	2003	7765960,00
12	2002	3277200,00
13	2001	2822700,00

Query executed successfully.

Figure 27. The Report Model of the number of admissions in fines each year.

The SQL commands for this is written as follows,

```
--/* REPORT OF ADMISSIONS IN FINES EACH YEAR */
SELECT dim_time.Year AS tahun,
SUM(fact_StudentLibrary.denda) AS [Jumlah Denda]
FROM fact_StudentLibrary INNER JOIN
dim_time ON fact_StudentLibrary.time_key =
dim_time.DateKey
GROUP BY dim_time.Year
ORDER BY dim_time.Year DESC
```

V. CONCLUSIONS

SnowFlake scheme has an advantage of efficiency in the use of storage media. Data modeling and transactional database system will ease in the data warehouse arrangement. Extract, transform, and loading (ETL) process has been done within 2 stages, namely data integrity testing and population process. In testing has been found data integrity issues because there was field foreign key that was inconsistency or null valued. So, to arrange data warehouse ought to eliminate noise data. In the ETL process, the

amount of 489,119 data rows has been successfully populated into fact data. Reporting model has been arranged based on information needs using transact SQL commands and also employed as functional testing.

Determining database object as source data should be concerned. In the next project, data warehouse development using Business Intelligence process will help to do analysis more over. And from that analysis could be used to build analysis service.

ACKNOWLEDGMENT

We thank STMIK AMIKOM Yogyakarta which has given us to support our paper research.

REFERENCES

- [1] A.R. Patel, "Data Modeling Techniques For Data Warehouse," ZENITH, International Journal of Multidisciplinary Research, Vol. 2, Issue 2, February 2012, pp. 240-246.
- [2] P. Ponniah, "Data Warehouse Fundamentals: A Comprehensive Guide for IT Professional," J. Wiley, New York.
- [3] A. Dahlan, and F. W. Wibowo, "Design of Library Data Warehouse Using Snowflake Scheme Method," Proceedings of IEEE 7th International Conference on Intelligent Systems, Modelling and Simulation, 2016.
- [4] W. H. Inmon, "Building the Data Warehouse," Fourth Edition, J. Wiley, New York.
- [5] J. Han and M. Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann, San Fransisco.
- [6] V. Rainardi, "Building A Data Warehouse with Examples in SQL Server," Apress, New York.
- [7] K. Srivastava, S. Srivastava, A. Sharma, and A. Pandey, "Comparison of Star Schema and Snow Flake Schema Using Telecommunication Database," International Journal of Latest Trends in Engineering, Science and Technology (IJLTEST), Vol. 1, Issue 5, May-June 2014, pp. 1-9.
- [8] M. H. Peyravi, "A Schema Selection Framework for Data Warehouse Design," International Journal of Machine Learning and Computing, Vol. 2, No. 3, June 2012, pp. 222-225.
- [9] G. Garani and S. Helmer, "Integrating Star and Snowflake Schemas in Data Warehouses," International Journal of Data Warehousing and Mining, Vol. 8, Issue 4, 2012, pp. 22-40.
- [10] F. S. Esmail Ali, "A Survey of Real-Time Data Warehouse and ETL," International Scientific Journal of Management Infromation Systems, Vol. 9, No. 3, 2014, pp. 3-9.