Efficiency Analysis of Data Warehouse Implementation for Forestry Industry Reports

Yusuf Wibisono*, Mohamad Adiyanto, Sigit Sigalayan
Mulia University, Balikpapan, Indonesia

*wibisono@universitasmulia.ac.id

Abstract. As one of the environmentally sensitive areas, the forestry industry required fast and accurate data processing for decision making, and data warehouses could be used to meet those needs. Decision-makers need aggregate, historical, and multi-dimensional data. Provided a data warehouse to present aggregate data could improve efficiency in presenting data. In this study, the data warehouse was built using the fact constellation scheme and the ETL process using query commands in MySQL. The level of efficiency was measured by comparing the number of tables, record length, number of records, and total bytes needed for each report. The efficiency of the report presentation was also compared with OLTP data and from the data warehouse. The test results in data management obtained an efficiency level of 56% for the number of tables, 145% for the length of records, 15,833% for the number of records, and 48,846% for the total bytes. On average, the level of efficiency in data management was 15,720%, while the efficiency of report presentation speed was 852%. This study showed that the use of data warehouses was very efficient in managing aggregate data for the forestry industry.

1. Introduction

Forest industry management requires serious attention because it is a source of state revenue but on the other hand, it is also very sensitive to environmental problems such as deforestation [1]. To address the issue of forestry, one of the initiatives that the government can take is to improve the forest governance system by implementing a forest reporting system supported by an information system covering all operational activities within the forest area [2].

To improve forestry industry management, the government has launched the SIPUHH (Sistem Informasi Penatausahaan Hasil Hutan) in 2016. SIPUHH is a web-based information system, which serves to record forestry company activities related to forest management products. The SIPUHH implementation produces very large data because it records all wood movements from the forest to the processing industry.

A large amount of data collected in the implementation of information systems in forestry [3] can be a very useful asset for leaders in planning and decision making. It is important to consider the use of data warehouses in supporting decision making [4]. The data warehouse has many advantages because it is structurally made to facilitate data analysis, while OLTP databases are designed to handle transactions [5].

Data warehouses proved to be more reliable and efficient than OLTP for implementation in the education sector. Based on research conducted at the university [6], the implementation of the data warehouse can provide an average efficiency of 461,801.84%. However, studies that have been
conducted also show that the implementation of the data warehouse in the government sector is still lacking when compared to the private sector [7].

2. Method
A descriptive quantitative method was used in this study and focused on measuring how efficiently data warehouses are implemented in forest industry reports compared to OLTP databases.

In general, people might know theoretically that the data warehouse is more efficient, but this research will get an accurate amount of efficiency with examples of applications in the forestry industry report.

The steps of this study were as follows:
1. Identified and limited only to 6 reports.
2. Designed the OLTP logic model used to create the reports.
3. Made hypercubes for reports and identified the needs of OLTP tables and data warehouses.
4. Created a table relationship for the data warehouse identified.
5. Implemented a data warehouse using MySQL.
6. Created ETL designs for processes from OLTP to the data warehouse.
7. Executed ETL to process from OLTP to the data warehouse.
8. Executed queries in OLTP to generate reports.
9. Executed queries in the data warehouse to generate reports.
10. Measured the efficiency comparison between the OLTP database and the data warehouse for all reports with parameters: Number of Tables, Record Length, Number of Records, Total Byte.
11. Measured the comparison of the execution time efficiency for each report, by comparing the reports generated from the OLTP database and reports created using the data warehouse.
12. Measuring the level of efficiency gained, using the formula: \( \frac{a-b}{b} \times 100\% \) where \( a = \) old data and \( b = \) new data

The data warehouse was constructed based on the needs of top managers who want to view multi-dimensional reports. If the needs of the report were 3-dimensional, then the shape could be seen as in Figure 1.

![Figure 1. The 3 Dimensional Model](image-url)

In Figure 1 we could see that the data will be presented in aggregate with 3 points of view or 3 dimensions: Species, Year, and Month. For example, based on the cube in Figure 1 we could obtain information about the Acacia Crasicarpa species from 2014 to 2017 with each seen for January, February, March, April, and May. Of course, we could also read data for other species that we need. If the manager wanted to add one more dimension, such as the Company dimension, then we use the hypercube model.
2.1. Report Identification

Based on the research stage, the first step was to identify 6 selected reports, namely:

1. Wood Production Volume Per Species Per Year Per Month.
2. Wood Production Volume Per Company Per Species Per Year Per Month.
3. PSDH Payments Per Company Per Bank Per Year Per Month.
4. The volume of Wood Transportation Per Company Per LHP Number Per Year Per Month.
5. PSDH Payments Per Business Type Per Company Per Year Per Month.
6. PSDH Payments Per Province Per District Per Year Per Month.

2.2. Hypercube Design

Furthermore, from 6 predetermined reports, a hypercube design was made for each report. The hypercube design helped simplify how reports are formed from data in the database. In Figure 3 we could see the hypercube design for Report 1 (Wood Production Volume Per Species Per Year Per Month).

![Hypercube for Report 1](image)

In Figure 2, aggregate production data would be viewed from 3 dimensions, namely Species, Year and Month. To produce information as shown in Figure 3 we need an OLTP table, namely PRODUKSI (13 columns) and SPECIES (3 columns). If using a data warehouse, only 1 fact table is needed (WPTSPECIES, 5 columns) and 1 dimension table (WSPECIES, 2 columns).

Next, the hypercube model was carried out for the entire report. Overall, the information obtained from the hypercube model was as shown in Table 1.

| Report   | Dimension                  | OLTP Tables                        | Data Warehouse        |
|----------|----------------------------|------------------------------------|-----------------------|
| Report 1 | Species, Year, Month       | PRODUKSI, SPECIES                  | WPTSPECIES, WSPECIES  |
| Report 2 | Company, Species, Year, Month | PRODUKSI, PT, SPECIES            | WPTSPECIES, WPT, WSPECIES |
| Report 3 | Company, Bank, Year, Month | PSDH, LHP, PT, BANK               | WPTBANK, WPT, WBANK   |
| Report 4 | Company, Year, LHP Number, Bank | PRODUKSI, LHP, PT, IZIN         | WPTLHP, WPT          |
| Report 5 | Company, Business Type, Year, Month | PSDH, LHP, PT, IZIN           | WPTBANK, WPT          |
| Report 6 | Year, Month, Province, Region | PSDH, LHP, PROVINSI, KABUPATEN    | WPTBANK, WWILAYAH     |

2.3. Table Relationship of Data Warehouse

Based on the hypercube design for 6 types of reports, then we could see the table relationship of the data warehouse as shown in Figure 3.
From Figure 3, the data warehouse model was a fact constellation scheme shown by 3 fact tables, namely WPTLHP, WPTSPEcies, WPTBANK that share the use of the WPT dimension table.

2.4. ETL Process

Once the OLTP tables were identified and the tables in the data warehouse are created, the next step is to prepare the ETL (Extraction, Transformation, Loading). In this case, ETL was made using a small program that will be executed at the end of the transaction date. All aggregate data, namely timber production, timber transportation, and payments at OLTP will be calculated from the beginning of the month to the present and stored in the fact table in the data warehouse. For dimension tables, a control table was created to see if there are changes in OLTP. If there is a new value, it would be saved. If not, the process would not be executed.

2.5. Execution and Measurement

The next step was the execution of ETL to process data from OLTP to the data warehouse. After the fact table and dimension table in the data warehouse contain data, then execute the query to generate reports both on OLTP and data warehouse. The tables needed in each report would be compared in terms of the number of tables, length of records, number of records, and total bytes. Also, the time taken for report execution would be compared between OLTP and the data warehouse.

3. Results and Discussion

Overall, the OLTP tables needed to produce 6 reports were shown in Table 2

| Table Name | Record Length | Number of Records | Total Bytes |
|------------|---------------|------------------|-------------|
| PRODUKSI   | 67            | 78,560           | 5,247,452   |
| PSDH       | 68            | 3,403            | 231,376     |
| LHP        | 57            | 3,403            | 195,148     |
| SPECIES    | 30            | 4                | 120         |
| BANK       | 36            | 11               | 392         |
| PT         | 41            | 119              | 4,844       |
| IZIN       | 51            | 6                | 304         |
| PROVINSI   | 32            | 34               | 1,100       |
| KABUPATEN  | 36            | 8                | 288         |
|            | 418           | 85,548           | 5,681,024   |

Whereas tables are needed for 6 reports if using the data warehouse as shown in Table 3.
Furthermore, the level of efficiency was measured using the formula:

\[
\text{% Efficiency} = \left( \frac{a - b}{b} \right) \times 100\%, \text{ where } a = \text{old data, } b = \text{new data}
\]

The results of efficiency measurements from OLTP and data warehouse were shown in Table 3.

### Table 3. List of Data Warehouse Tables Required to Generate Reports

| Table Name     | Record Length | Number of Records | Total Bytes |
|----------------|---------------|-------------------|-------------|
| WPTSPECIES     | 18            | 145               | 2,610       |
| WPTLHP         | 49            | 649               | 31,912      |
| WPTBANK        | 20            | 653               | 13,060      |
| WPT            | 44            | 119               | 5,268       |
| WBANK          | 20            | 11                | 220         |
| WSPECIES       | 26            | 4                 | 104         |
| WWILAYAH       | 48            | 8                 | 384         |
|                | 225           | 1,589             | 53,558      |

### Table 4. Measurement Results of Efficiency Levels in All Reports

| Object          | Number of Tables | Record Length | Number of Records | Total Bytes |
|-----------------|------------------|---------------|-------------------|-------------|
| OLTP            | 9                | 418           | 85,548            | 5,681,024   |
| Data Warehouse  | 7                | 225           | 1,589             | 53,558      |

Efficiency measures were applied to each table that was needed to produce the report. In this case, the table could be used more than once in different reports. The results of measuring the level of efficiency of each report were shown in Table 4.

### Table 5. Measurement Result of Efficiency Levels in Each Report

![Measurement Result of Efficiency Levels in Each Report](image)

From detailed measurements, it could be seen that the use of data warehouse gets an efficiency level for 56% of tables, 145% of record length, 15833% of record counts and 46846% of total records.

In addition, measurements were also taken of the time needed in the execution of the report. The results were shown in Table 5.
Table 5. Efficiency Rate of Execution Time for All Reports

| Source Data          | Execution time for each report (seconds) |
|----------------------|-----------------------------------------|
|                      | #1  | #2  | #3  | #4  | #5  | #6  |
| OLTP                 | 2.169 | 4.075 | 0.253 | 3.759 | 0.267 | 0.391 |
| Data Warehouse       | 0.156 | 0.171 | 0.195 | 0.255 | 0.212 | 0.189 |
| Efficiency           | 1,290% | 2,283% | 30% | 1,374% | 26% | 107% |

From the efficiency level for execution time in each report, then the average was calculated and the result is 852%. A comparison of the execution speed was seen more clearly as shown in Figure 4.

![Figure 4. Time Consumption of Report Execution](image)

4. Conclusions

Based on efficiency measurements, it was concluded that the data warehouse provides a high level of efficiency in the number of records, record length, and total bytes in storage, which makes data management easier. Furthermore, report execution from the data warehouse is faster than OLTP. This will be useful in the speed of decision making.

5. Reference

[1] S. Chakravarty, S. K. Ghosh, C. P. Suresh, a. . DEy, and G. Shukla, “Deforestation: Causes, Effects and Control Strategies, Global Perspectives on Sustainable Forest Management,” Sustain. For. Manag., pp. 3–29, 2012.

[2] M. Simanjuntak, A. Daryanto, and T. Rusolono, “FORESTRY COMPANY SATISFACTION TOWARDS ONLINE INFORMATION SYSTEM IMPLEMENTATION OF FOREST PRODUCT MANAGEMENT (SI-PUHH),” Indones. J. Bus. Entrep., 2016.

[3] S. W. Woudenberg, B. L. Conkling, B. M. O’Connell, E. B. LaPoint, J. A. Turner, and K. L. Waddell, “The forest inventory and analysis database: Database description and users manual version 4.0 for phase 2,” USDA For. Serv. - Gen. Tech. Rep. RMRS-GTR, 2010.

[4] Mir S A et al., “DECISION SUPPORT SYSTEMS IN A GLOBAL AGRICULTURAL PERSPECTIVE - A COMPREHENSIVE REVIEW,” Int. J. Agric. Sci., 2015.

[5] A. Parekh, “Introduction on Data Warehouse with OLTP and OLAP,” Int. J. Eng. Comput. Sci., 2013.

[6] H. L. H. S. Warnars, “Desain model data warehouse dengan contoh kasus perguruan tinggi,” J. Ind. Eng. Manag. Syst. (ISSN 1979-1720), 2010.

[7] M. Bilal, U. Sheikh, B. Raza, and Q. Javaid, “Application of Data Warehouse in Real Life: State-of-the-art Survey from User Preferences’ Perspective,” Int. J. Adv. Comput. Sci. Appl., 2016.