# Design General Hospital Data Warehouse Base on Nine Step Methodology

Ida Bagus Leo Mahadya Suta[1*], I Gusti Ngurah Agung Surya Mahendra[2], and Yanu Prapto Sudarmojo[3]

[1,2]Department of Electrical and Computer Engineering, Post Graduate Program, Udayana University

[3]Department of Electrical and Computer Engineering, Udayana University

*gusleo.bali@gmail.com

**Abstract** Data warehouse is a method used to hold and analyze data in large number. The analysis can be viewed from several different dimensions or points of view. The application of data warehouse technology can overcome the problems and needs of executive at Genral Hospital in conducting analysis of large amounts of patient data, especially data inpatient and outpatient care. The data warehouse design method is done by applying 9 step (Nine-Step Methodology) used by Kimball. The 9 stages are the selection process, selection, grain, identification and adjustment, fact selection, pre-calculation storage in the fact table, ensure dimension tables, select database durations, track changes from dimensions slowly, determination of priority and query model. The results of this study are in the form of schema design and data simulation warehouse by displaying inpatient and outpatient data based on time dimensions, patients reference, diagnosis, patient status, room, type of patient, condition of return, and district in the form of a table and graph.

*Index Terms*—Data Warehouse, patient data, nine-step methodology, Schema.

## I. INTRODUCTION[1]

Data warehouse is a collection of data obtained from various sources used to support the decision making process management within the company [1].

The ability of the data warehouse to accommodate large amounts of data, and display data in a multidimensional manner so that it can support analysis process, and decision making make the data warehouse available used in several organizations, companies and several government agencies.

General hospital has used information technology to assist its operational activities. Along with its development, more and more complex patient data is also owned, especially inpatient and outpatient data. The data needs to be processed and reported to the executive to find out the development of the hospital for a certain period of time[2]-[3]. Meanwhile, to make the report, it takes longer to collect information and process data stored in the operational database. Therefore, there is a need for data processing tools or applications that can support the analysis process of these patient data.

Based on the explanation above, the Data warehouse can be a tool to facilitate the executive in analyzing patient data as a consideration in decision making.

## II. LITERATURE REVIEW

### A. Data Warehouse

Data warehouse is a data collection that is subject oriented, integrated, time-variant, and non-volatile in supporting the decision making process. [4].

There are four characteristics of a data warehouse [1]:
1. Subject-oriented (Subject Oriented)
2. Integrated (Integrated)
3. Non-Volatile
4. Time-Variant

### B. Online Analytic Processing (OLAP)

OLAP is a device that describe technology using multidimensional visualization of data for the purpose of accelerating analysis. In the OLAP data model, information is conceptually described as a cube, which consists of descriptive categories (dimensions) and quantitative values (measures) [5].
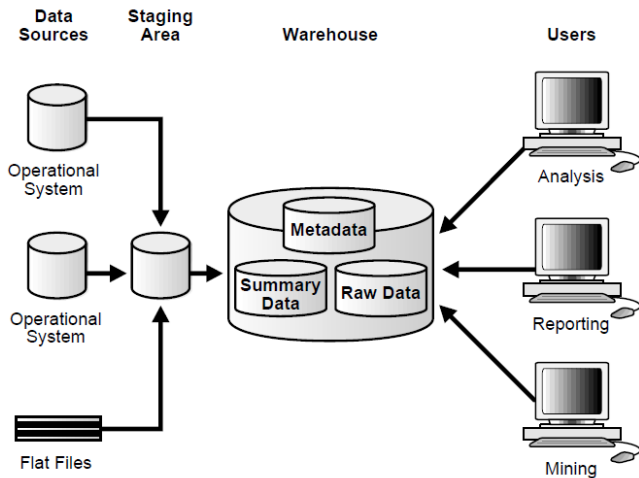
## C. Architecture Data Warehouse



Fig 1. Architecture Data Warehouse with Staging Area

## D. Model Dimensional

The OLAP dimensional model is composed of fact tables and dimensions that have schema standards in the form of schemes [1]. The scheme includes:
1. Star Scheme
2. Snowflake Scheme
3. Starflake Scheme

## E. Data Warehouse Design Method

The method of research carried out is described in the framework of the research as in Figure 2. The process carried out is data collection, compilation that is the analysis of data requirements relating to inpatient and outpatient data and the last design [6]-[7].
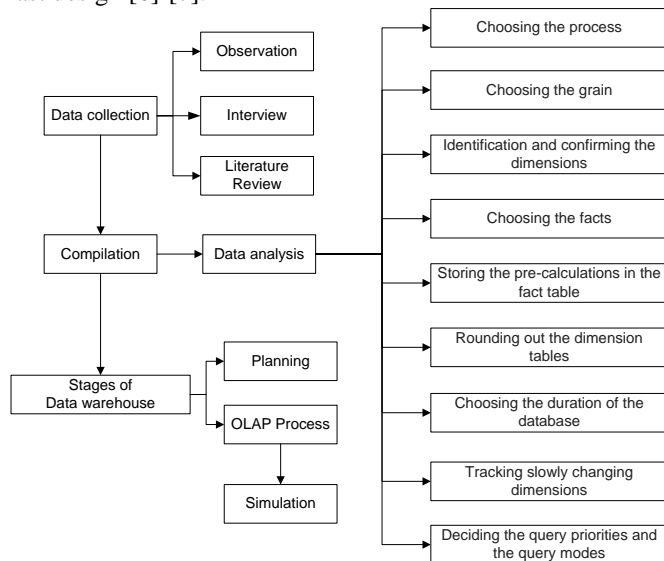


Fig 2. Data Warehouse Design Method

Nine-step methodology
1. Choosing the process
2. Choosing the grain
3. Identification and confirming the dimensions
4. Choosing the facts
5. Storing the pre-calculations in the fact table
6. Rounding out the dimension tables
7. Choosing the duration of the database
8. Tracking slowly changing dimensions
9. Deciding the query priorities and the query modes

## III. RESULT AND DISCUSSION

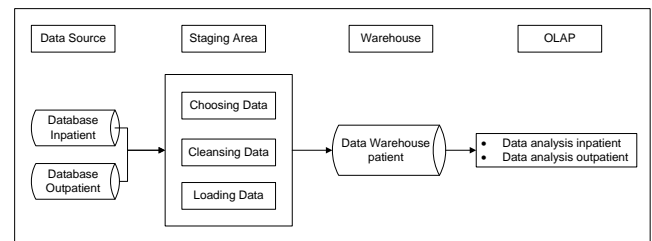### A. Design of Data Warehouse Architecture



Fig 3. Design of Data Warehouse Architecture

### B. Stages of Data Warehouse

1. Design of Data Warehouse

   a. Choosing the process

   Based on the scope and information needed by the executive of General Hospital, the selection process will be used to design the data warehouse. The process chosen is the process of hospitalization and outpatient care.

   b. Choosing the grain

   Grain is a process for determining what the record will describe in the fact sheet. Here are the grain available in the data warehouse design.

   1) Inpatient

   The analysis that will be carried out on the inpatient process includes the number of inpatients based on patient sex, sub-district, city, type of patient, origin of referral, room name, room section, service class, admission date, diagnosis, patient status, and condition return.

   2) Outpatient

   The analysis that will be carried out in the outpatient process includes the number of outpatients based on patient sex, sub-district, city type of patient, origin of referral, name of room, section of rehabilitation, class of service, date of admission, diagnosis, and patient status.

   c. Identification and confirming the dimensions

   In this stage, the activity that will be carried out is to build a set of dimensions to set the context for asking questions about facts in the fact table. The following are the dimensions chosen for each fact table:

   1) Inpatient

   Dimensions to be used in the analysis hospitalization is the dimension of time, patient,

referral, diagnosis, patient status, room, type of patient and home conditions, and sub-districts.

2) Outpatient
The dimensions to be used in the outpatient analysis are the dimensions of time, patient, reference, diagnosis, patient status, room, type of patient and sub-district.

d. Choosing the facts
At this stage will be the selection of facts to be used on the Medical Record (MR) MR inpatients and MR outpatients as fact tables. Here is the fact of fact to use.

1) Facts MR Inpatient
The facts chosen for use in the inpatient MR fact table consist of TimeID, PatientID, ReferralID, DiagnosisID, PatientStatusID, RoomID, PatientTypeID, SubDisctrictID and CheckoutConditionID. The following are the attributes in the inpatient MR fact table.

Table 1 Facts MR Inpatient

| Column | Data Type | Length |
|---|---|---|
| TimeID | Datetime | |
| PatientID | Number | 20 |
| ReferralID | Number | 2 |
| DiagnosisID | Number | 7 |
| PatientStatusID | Number | 2 |
| RoomID | Number | 3 |
| PatientTypeID | Number | 2 |
| SubDisctrictID | Number | 2 |
| CheckoutConditionID | Number | 4 |

2) Facts MR Outpatient
The facts chosen for use in the outpatient MR fact table consist of TimeID, PatientID, ReferralID, DiagnosisID, PatientStatusID, RoomID, PatientTypeID, and SubDisctrictID. The following are the attributes in the outpatient MR fact table.

Table 2 Facts MR Outpatient

| Column | Data Type | Length |
|---|---|---|
| TimeID | Datetime | |
| PatientID | Number | 20 |
| ReferralID | Number | 2 |
| DiagnosisID | Number | 7 |
| PatientStatusID | Number | 2 |
| RoomID | Number | 3 |
| PatientTypeID | Number | 2 |
| SubDisctrictID | Number | 2 |

e. Storing the pre-calculations in the fact table
The initial calculations contained in the fact sheet are as follows:

1) Facts Inpatient
The fact calculation of hospitalization is the number of inpatients based on the patient's sex, city, sub-district, type of patient, origin of the room, room name, service class,

admission date, diagnosis, patient status, return condition and total patient stay at hospital.

2) Facts Outpatient
The fact outpatient calculation is the number of outpatients based on patient sex, city, sub-district, type of patient, origin of referral, room name, room section, outpatient admission date, diagnosis, patient status and total outpatient.

f. Rounding out the dimension tables
This stage is adding information to each dimension to make it easier to use and understand by the user.

Table 3. Explanation for each dimension

| Dimension | Field | Notes |
|---|---|---|
| Time | Date Month Year | Analysis of the number of patients can be seen per date, per month, and per year. |
| Patient | Sex | Analysis of the number of patients can be seen by sex |
| Referral | ReferralFrom | Analysis of the number of patients can be seen based on the origin of the referral |
| Diagnosis | Diagnosis | Analysis of the number of patients can be seen based on diagnosis |
| Patient Status | PatientStatus | Analysis of the number of patients can be seen based on patient status |
| Room | RoomName RoomType | Analysis of the number of patients can be seen based on room name and room type |
| Patient Type | PatientType | Analysis of the number of patients can be seen based on patient type |
| Checkout Condition | CheckoutCond | Analysis of the number of patients can be seen based on checkout condition |
| Sub-district | SubDistrictName | Analysis of the number of patients can be seen based on sub-district |
| City | CityName | Analysis of the number of patients can be seen based on city |

g. Choosing the duration of the database
The duration of the database that will be entered into the data warehouse is for one year from 2017 to 2018.

h. Tracking slowly changing dimensions
Attributes from dimension tables do not always have a fixed value or are relatively static. Changes in attribute values can occur for a long time. The three basic types of changing attributes in dimensions are rewriting attributes, creating new records and creating new columns.

The type of change used in the dimensional attributes of a hospital data warehouse is to create a new record. Some of these attributes can be seen in table 14.

Table 4. Dimension Change

| Dimension | Available Attributes Change |
|---|---|
| Dim_Room | Room_Name |
| Dim_Referral | Referral_From |
| Dim_Status | Patient Status |
| Dim_Patient_Type | Patient_Type |
| Dim_Diagnosis | Diganosis_Name |
| Dim_SubDistrict | SubDistrict |
| Dim_CheckoutCondition | Checkout_Condition |

    i.  Deciding the query priorities and the query modes
       In this phase will be discussed about OLAP process design. The discussion will take into account the influence on physical design, such as summaries and summaries.

2. OLAP Process

    The OLAP process is carried out by creating a 'Hospital' database as a source of data using the database management tool. Then creating a database with the name 'DW Hospital' to store 'Hospital' data sources by using the OLAP tool can connect with data sources. And designing the data warehouse forms a scheme.

3. Simulation

    Simulation is done by running the design results by displaying some information as follows:

    a.  Information about the number of patients admitted based on the sex of the patient, sub-district, city, type of patient, origin of the referral, room name, part of the room, date of entry (month and year), diagnosis, patient status and return condition.

    b.  Information on the number of outpatients based on patient sex, sub-district, city, type of patient, origin of the referral, room name, part of the room, date of entry (month and year), diagnosis and patient status.

*C. Results*

    After creating database, next step is OLAP processing to create fact table and dimensions table.
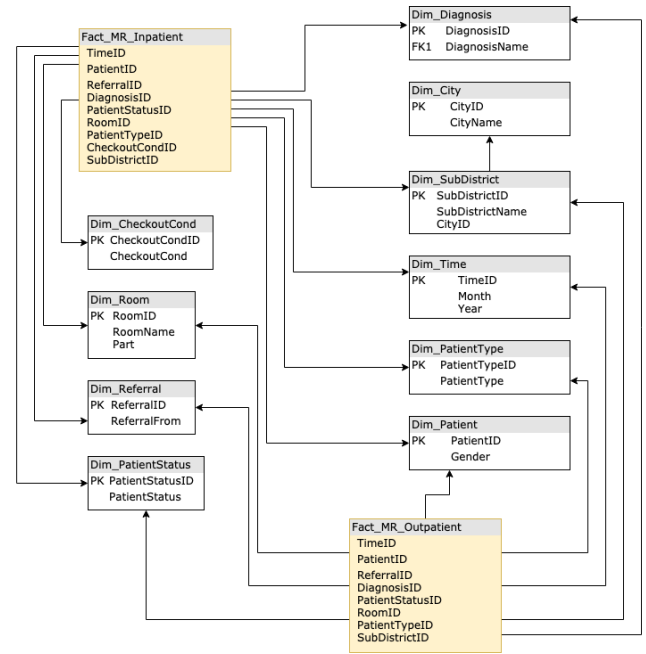


Fig 4. Snowflake Schema Data Warehouse

    Some information that can be simulated based on a scheme that has been designed is as follows:

Table 5. Information on the number of inpatients per patient dimension by gender.

| Row Labels | MR Inpatient Count |
|---|---|
| **M** | 357 |
| **F** | 347 |
| Grand Total | **704** |

Table 6. Information on the number of outpatients per time dimension based on months and years.

| MR Outpatient | Year | | |
|---|---|---|---|
| **Month** | **2017** | **2018** | **Total** |
| **Aug-17** | 52 | | 52 |
| **Apr-17** | 31 | | 31 |
| **Dec-17** | 145 | | 145 |
| **Feb-17** | 32 | | 32 |
| **Jan-17** | 30 | | 30 |
| **Jan-18** | | 87 | 87 |
| **Jul-17** | 36 | | 36 |
| **Jun-17** | 46 | | 46 |
| **Mar-17** | 41 | | 41 |
| **May-17** | 39 | | 39 |
| **Nov-17** | 145 | | 145 |
| **Oct-17** | 125 | | 125 |
| **Sep-17** | 74 | | 74 |
| **Grand Total** | **796** | **87** | **883** |

In addition to the information form of patient data that has been shown previously, analysis of patient data that has been formed into multidimensional data can also be carried out based on the following functions.

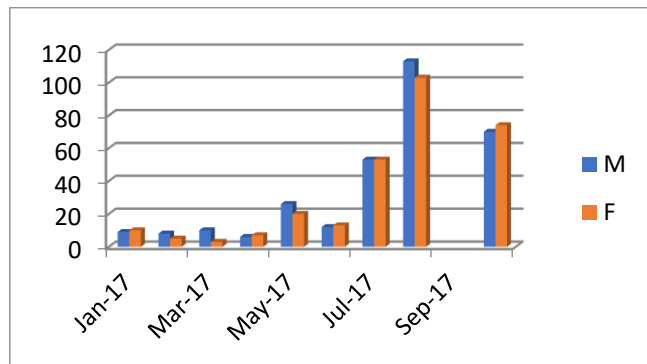| MR Inpatient Count | Gender | | |
|---|---|---|---|
| Date | M | F | Total |
| Jan-17 | 8 | 9 | 17 |
| Feb-17 | 4 | 9 | 13 |
| Mar-17 | 7 | 6 | 13 |
| Apr-17 | 9 | 10 | 19 |
| May-17 | 8 | 5 | 13 |
| Jun-17 | 10 | 3 | 13 |
| Jul-17 | 6 | 7 | 13 |
| Aug-17 | 26 | 20 | 46 |
| Sep-17 | 12 | 13 | 25 |
| Oct-17 | 53 | 53 | 106 |
| Nov-17 | 113 | 103 | 216 |
| Dec-17 | 70 | 74 | 144 |
| Grand Total | 326 | 312 | 638 |



Fig 5. Information Patient with Chart Bar

## IV. CONCLUSIONS

Based on the research that has been carried out at the General Hospital it can be concluded that the design of the data warehouse can be made in the form of a Starflake scheme generated through the OLAP method. Data analysis has been made in the form of simulations for the delivery of information or statistical reports relating to patient data to the executive of the General Hospital. Some data analysis simulations are generated, namely displaying information or reports on inpatient data based on patient sex, sub-district and city, type of patient, origin of referral, room name and part of the room, month and year, diagnosis, patient status, and return condition. Make a display of information or reports on outpatient data based on the sex of the patient, sub-district and city, type of patient, origin of reference, room name and part of the room, month and year, diagnosis, and patient status.

In further research, it can be carried out by applying a data mining system so that the analysis process can be carried

out in more depth and patterned based on existing approaches in data mining systems.

REFERENCES

[1] Inmon, W.H. 2002. *Building The Data Warehouse*.Canada: John Wiley &Sons, Inc.

[2] Muninjaya, A.A Gde. (2004). Manajemen Kesehatan. Edisi 2. Kedokteran EGC, Jakarta. Siregar, Charles J.P. dan Amalia, Lia. (2003).

[3] Kennebeck, S., Timm, N., Spooner, A. and Farrell, M. (2012) 'Impact of electronic health record implementation on patient flow metrics in a pediatric emergency department', *Journal of American Medical Informatics Association*, Vol. 19, No. 3, pp.443–447.

[4] Connolly, T., & Begg, C. (2005), *Database Systems: A Practical Approach to Design, Implementation, and Management*. USA : Addison-Wesley

[5] Lane, P. (2002). Oracle91 Data Warehousing Guide. Oracle Corporation.

[6] Hwang, Mark I., Hongjiang. (2008). A Structural Model of Data Warehousing Success

[7] Inmon, W.H. (2005). Building the Data Warehouse, 4th Edition. Wiley Publishing, Inc., Indianapolis, Indiana.