SOFTWARE DEVELOPMENT OF DATA WAREHOUSE WITH FUZZY DIMENSION CONCEPT ON SPACEMAN CLOTHING INDONESIA COMPANY

Ricky Azhari Pratama

Informatic Engineering – Indonesia Computer University
Dipatiukur street 112-114 Bandung
Email : coret.team@gmail.com

ABSTRACT

Spaceman Clothing Indonesia is one of the manufacturing company on convection area in Bandung, Indonesia.The design data of existing data are less effective and uncomplex for analyzing and existing report did not show expected information,so the report become to be ambigous to analyzed. Softwre development of data warehouse with OLAP as an information delivery could be solution of the existing problem. Fuzzy concept in development of data warehouse also a solusiton to ambigous fact.

Development of data warehous based on the needs of information desired by company. Data Optimization to be able to support the dynamic information and covers many important aspects through process of data staging or called ETL (extract, transfrom, loading). The fuzzy concept processed on data staging process as expansion of fact population, the ambigous fact related on fuzzy set according business needs, so it shall be show the predictive information. Olap use a pivoting method for presenting informasion based on company needs.

Based on analysis and testing, software development of data warehouse can help the owner to analyze report, multi-dimensional data modelling, data integrated, and wihh fuzzy concep application on data warehouse, the information displaying as business needs from ambigous facts, so the information is more efficient to be analyzed.

Kata Kunci : Data *Warehouse*, *Fuzzy* Data *Warehouse*, *ETL*, *OLAP*, Logika *Fuzzy*.

1. INTRODUCTION

Company Spaceman Clothing Indonesia already can be classified with technology user by applying computer for logging data. In this time, report generated in current logging is less effective for analyzing by owner because data design that less complex and not able to cover various aspects.

Data warehouse is a solution that can be used for modelling data in company so that it can be seeing from various important aspects. Application of fuzzy algorithm used to handle ambigous facts in order to select and redesign fact to more effective information.

1.1 Data Warehouse

Data warehouse is a system that presenting informations, In that system have integration and transformation of data from multiple operational systems from internal source and external sources into a relevant information for decision support in business strategy[1].

The characteristics of data warehose as follows:

a. Subject oriented
b. Integrated
c. Time variant
d. Non volatile

Data warehouse architecture is the structure that serves all components involved in data warehouse at same time[1]. The data is integrated as a unit, all the needs for the preparation of data and its storage, and direction of the presentation of information from data warehouse, so it will be produces rules, procedures, and functional to allow data warehouse work and meet the needs of the business[1]. One of architecture taht can be used is the architecture of three major areas including data acquisition, data storage, and information delivery[1].

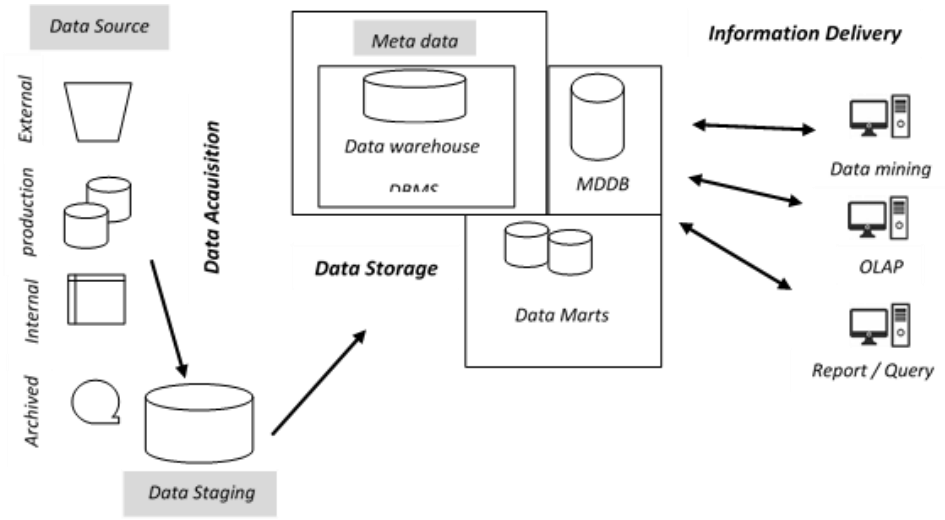


figure 1 Architecture component in *three major areas* [1]

1.1.1 ETL (Extract, Transform, Loading)

ETL is the set of functions that are done to reshape and the data into different shapes in the operational systems stored in the data warehouse as a relevant and strategic information [1].

a. Extraction
The extraction phase is identify all source of internal external sources, determine compatibility data structure if and only if data from external source, and indicate methods for extracting data[1].

b. Transform
Transform phase have function including input selection, input structure, separation of normalization and denormalization of the data source structure, agregation, convert, and solving missing value[1].

c. Loading
Loading phase is phase of initialization, define how often a group data must remain up to date in data warehouse, and determine hot to change the data will be implemented within a specific time period[1].

1.1.2 OLAP (On-Line Analytical Processing)

OLAP is one of software technology category tahat enables analysts, manager, or executives to dig data in a timely, consistent, and having interactive access to an excavation in the vastness of the informations is transformed from raw data into a fact dimensional that can be understood by user. [1].

The characteristics of OLAP as follows:

a. Multidimensional Conceptual View.
b. Transparency.
c. Accesibility.
d. Consistent Reporting Performance
e. Client/Server Architecture
f. Generic Dimensionality
g. Intuitive Data Manipulation
h. Flexible Reporting
i. Multi-User Support
j. Unlimited Dimensions and agregation levels

1.2 Fuzzy Data Warehouse

Fuzzy dimension is a dimension or fact witch contain and identify elements of fuzzy data used for analyzing[2]. Used data should be considered in order to provide a benefit towards business processes[2]. Basic information in order to define fuzzy variable is table source and target attributes, attributes type, associations, attributes, and calculations[2].

1.2.1 Konsep *Fuzzy* Data *Warehouse*

For integrating fuzzy concept into a data warehouse is the analyzation elements in data warehouse that can be classified in fuzzy where the element analyzed could be a fact in the fact table or an attribute from dimension table [3]. Domain attribute is a set values or a range possibility value of an attribute dimension or fact wich can linguistic mapped in a term that will be mapped into a set of classes in a fuzzy concept[3]. Fuzzy model of data warehouse is a combination of four elements namely table dimensions, a fact table, fuzzy membership, and fuzzy classification[3].

1.2.2 *Meta Model Fuzzy* Data *Warehouse*

A meta model define elements of the conceptualization and their relationship[3]. In fuzzy data warehouse, meta model in data warehouse concept of fuzzy integrated with each other as meta table structure[3]. Meta model of fuzzy data warehouse refers to concept of fuzzy integrated with a data warehouse where each attribute defined as target in fuzzy data warehouse model can be more than one and vice versa[3].

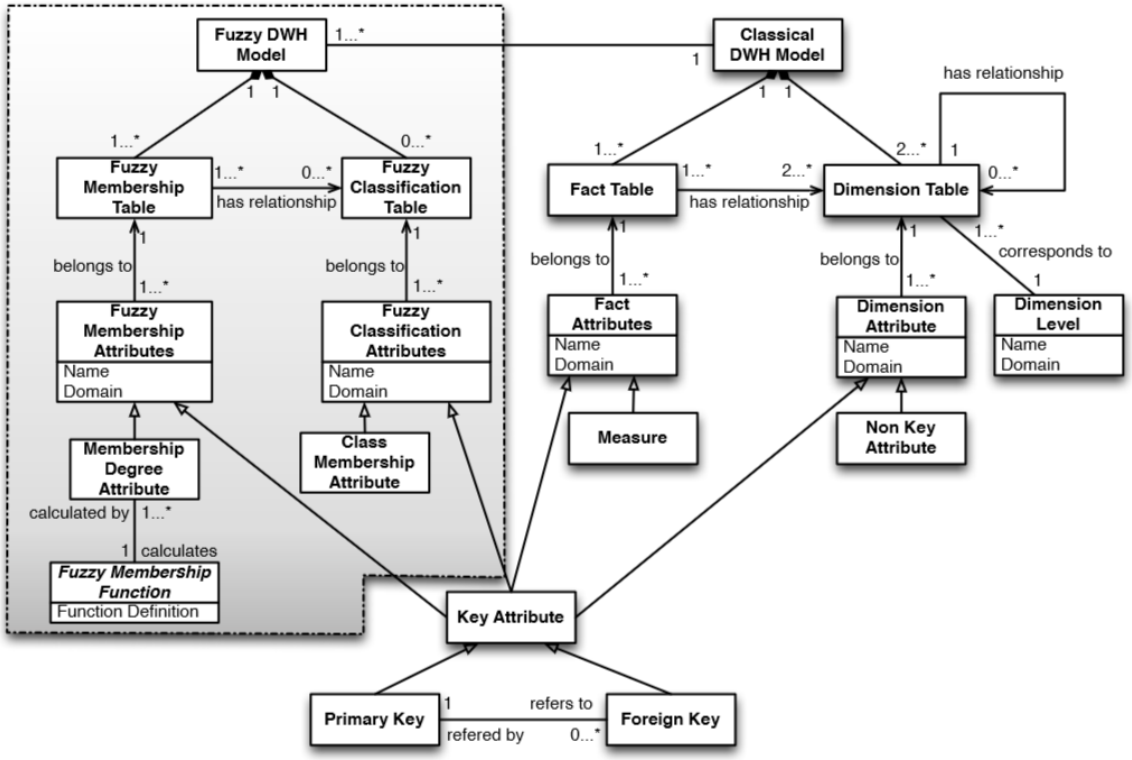


figure 2 Meta model fuzzy data warehouse [3]

2. ISI PENELITIAN

2.1 Sumber Data

Data source on this company based on an existing database. This database will be redesigned in preprpocessing phase into data warehouse. The following database used on the company.

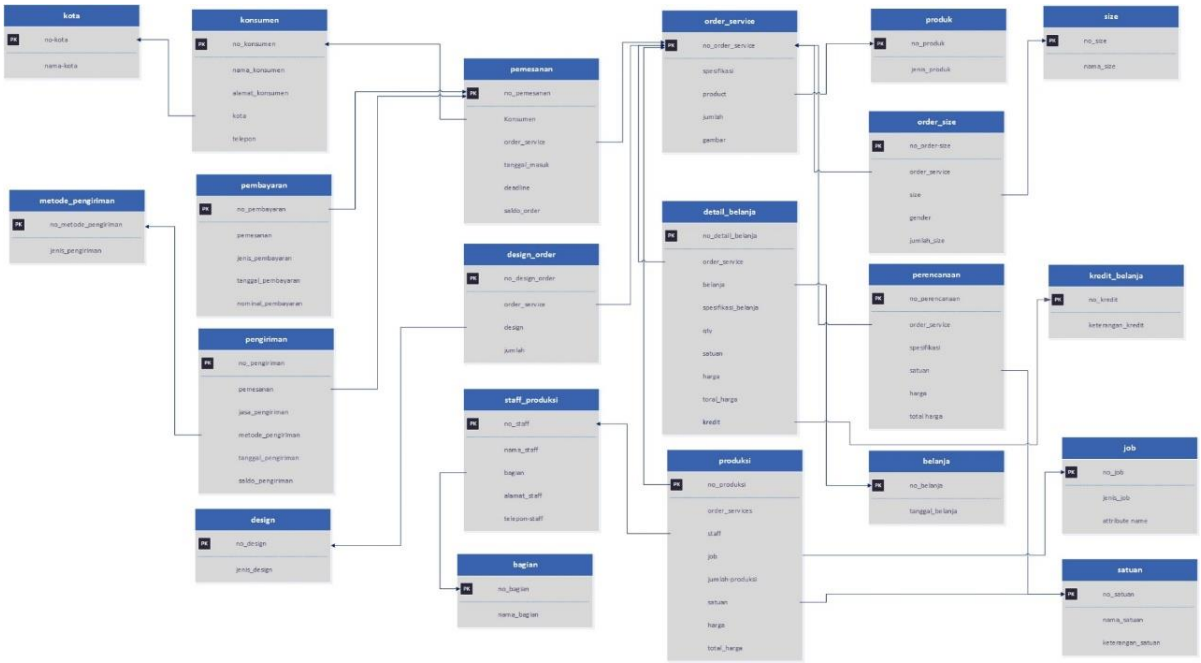


figure 3 OLTP diagram scheme in Spaceman Clothing Indonesia

2.2 Modelling Of Data Warehouse

Data warehouse development based on the exising problems in the company and become needs, The needs will be classified to get desired on strategic information needs as owner needs. There are two models that are applied to this data warehouse to be built, as follows:

a. Classical data warehouse
b. Fuzzy data warehouse

Dimension describe the business entities that support a fact so it can be in multidimensional analyzing, as follows:

a. Design dimension
   This dimension describing product design that used to product.
b. Job dimension
   This dimension describing production job on on production product phase.
c. Consumers dimension
   This dimension describing consumers profile.
d. Credit dimension
   This dimension describing type of cost in production expenditures.
e. Delivery dimension
   This dimension describing type of delivery service that used for shipping product.
f. Product dimension
   This dimension describing type of product that can be order by consumers.
g. Size dimension
   This dimension describing type of size that userd in product.
h. Staff dimension
   This dimension describing staff profile.
i. Time dimension
   This dimension describing time, event taht occur in transaction process, shipping, and expenditures.

Fact explain directly the values relating to business processes that are multidimensional in order to make it easier to process analysis. As for the fact that is used as follows:

a. Market fact
   Market fact describing fact of cost of purchasing materials and tools for the benefit of product manufacturing.
b. Payment fact
   Payment fact describing groupof payment range against deadline in fuzzy scale.
c. Product popularity fact
   Product popularity fact describing a type of product popularity that order by consumers.
d. Production fact
   Production fact describing cost fact taht used in production processes.
e. Order fact
   Order fact describing product order by consumers.
f. Product shipping fact
   Product shipping fact discribing cost of shipping product to consumers.
g. Design popularity fact
   Design popularity fact describing popularity of design taht used in product.
h. Size product fact
   Size product fact describing sumary of size that used in product..

2.3 Data Staging

Data staging or ETL processes is transition or data transfer from OLTP database into data warehouse. The control flow of data staging is as follows:

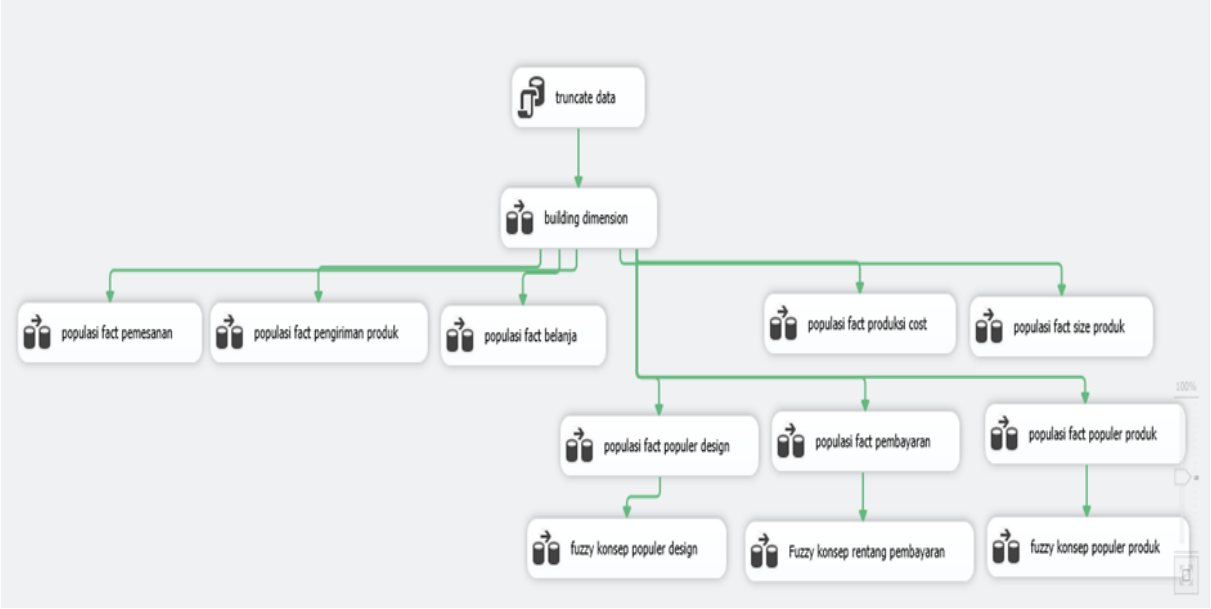


figure 4 *Control flow* data *staging*

a. Truncate data
This process is data cleaning process and regenerate DBMS data warehouse.
b. Building dimension
This process is extraction, transform dan loading data process that needed to be foundation of dimension in data warehouse from OLTP database according the needs.
c. Populate facts
this process is data look-up processs that needed as foundation of fact according with existing transactional data source that foundation of fact according the needs.
d. Fuzzy concept
This process is expansion process to the fact that meaning ambigous or inneficient to directly be used as information. So, the fact clasified to get new facts according with definingn the linguistic term.

2.4 Penerapan *Fuzzy* Data *Warehouse*

Application of fuzzy in data warehouse is aimed for classifying ambigous fact. Modeling of fuzzy concept that generate degree of membership based on attribute value of target domain classified..

2.4.1 Elemen Pengklasifikasian

Element of classification is used to determine the set of memberhip and the range of the value of domain targer value. On test case at the company toward consumer repayment in wich the target attribute is the fact of payment range where the set value range and domain membership memberships are as follows:

a. In time = -∞ - 10
b. Almost late = 0 - 20
c. Late = 16 – 30
d. Too late = 25 - ∞

As for membership function of consumer repayments are as follows:

$$\mu_a(x) = \begin{cases} x \geq R1_{rendah}, & 0 \\ x \leq 0, & 1 \\ else, & \dfrac{(R1_{rendah} - x)}{\mathrm{R1_{rendah}} - \mathrm{R1_{tinggi}}} \end{cases}$$

$$\mu_b(x) = \begin{cases} P2_{rendah} \geq x \geq R2_{rendah}, & 0 \\ x = 13, & 1 \\ P2_{rendah} < x < P2_{tinggi}, & \dfrac{(x - P2_{rendah})}{P2_{tinggi} - P2_{rendah}} \\ R2_{rendah} > x > R2_{tinggi}, & \dfrac{R2_{rendah} - x}{R2_{rendah} - R2_{tinggi}} \end{cases}$$

$$\mu_c(x) = \begin{cases} P3_{rendah} \geq x \geq R3_{rendah}, & 0 \\ x = P3_{tinggi} = R3_{tinggi}, & 1 \\ P3_{rendah} < x < P3_{tinggi}, & \dfrac{(x - P3_{rendah})}{P3_{tinggi} - P3_{rendah}} \\ R3_{rendah} > x > R3_{tinggi}, & \dfrac{R3_{tinggi} - x}{R3_{rendah} - R3_{ringgi}} \end{cases}$$

$$\mu_d(x) = \begin{cases} x \leq P4_{rendah}, & 0 \\ x \geq P4_{tinggi}, & 1 \\ else, & \dfrac{(x - P4_{rendah})}{P4_{tinggi} - P4_{rendah}} \end{cases}$$

Keterangan :
A = the fuzzy set of in time
B = the fuzzy set of almost late
C = the fuzzy set of late
D = the fuzzy set of too late
μ = Membership degree (0 – 1)

According on membership functions that has been spelled out, so it can mapping the curves of fuzzy membership according by pyment consumers.

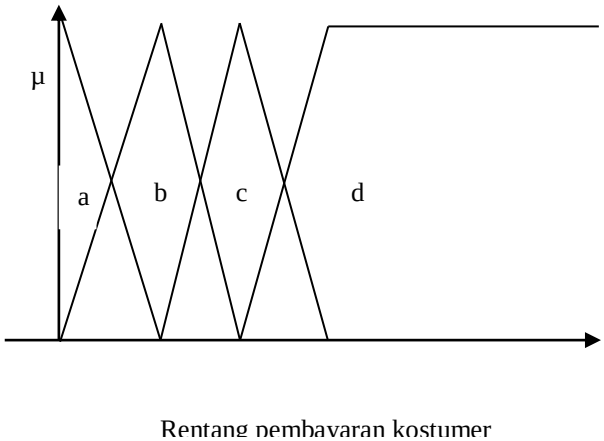


figure 5 Fuzzy set mapping curve consumer repayment

2.4.2 Modelling and fuzzy data warehouse test case

On element of classification set has been defined and function member can search for the value membership degrees, modelling in fuzzy data warehouse where the fact serving the attribute target will be given a primary key, and closely related with fuzzy table according the method [3].

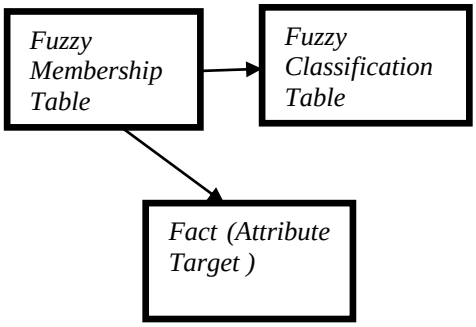


figure 6 fuzzy data warehouse scheme

In case test such as full amout of order product made by consumer, example the specified date has

crossed the line from deadline shall give rise to the range of differences dates become long span of time payments should be.

$$rentang = \begin{cases} tanggal \leq deadline, & 0 \\ else, & tanggal\ bayar - deadline \end{cases}$$

The following data range of repayment of consumer which is counted of range payment.

Table 1 Consumer payment range

No.	Consumer	Payment Range
1	John	29
2	Katrin	5
3	Inti	2

On table 1, its known range of payments made by consumers. To get memberhip degree using membership function that has been described in point of classification elements.

Table 2 Membership functions mapping

1	**Consumer**	John
	Payment Range	29
	In time Condition	Rentang ≥ 7 = True
	In time function	0
	Almost late Condition	7 ≥ Rentang ≥20 = True
	Almost late Function	0
	Late condition	16 ≥ Rentang ≥ 30 = False Rentang = 20 = false 16 < rentang < 20 = false 20 > rentang <30 = True
	Late Function	30-29/ 30 – 20 = 1/10 = 0,1
	Too late	Rentang ≤ 25 = false
	Condition	Rentang ≥ 30 = false 25 < Rentang < 30 = true
	Too late Function	29 – 25/30-25 = 4/5 = 0,8
2	**Consumer**	Katrin
	Payment Range	5
	In time Condition	Rentang ≥ 7 = false Rentang ≤ 0 = false
	In time function	10-5/10 = 0,52
	Almost late Condition	7 ≥ Rentang ≥20 = false Rentang = 13, false 0 < Rentang < 13 = true
	Almost late function	5/13-0 =7/13=0,38
	Late Condition	16 ≥ Rentang ≥ 30 = true
	Late Function	0
	Too late Condition	Rentang ≤ 25 =true
	Too late Function	0
3	**Consumer**	Inti
	Payment Range	2
	In time Condition	Rentang ≥ 7 = false Rentang ≤ 0 = false
	In time function	10-2/10 = 0,8
	Almost late Condition	7 ≥ Rentang ≥20 = false Rentang = 13, false 0 < Rentang < 13 = true
	Almost late Function	2/13-0 =7/13=0,15

	Late Condition	16 ≥ Rentang ≥ 30 = true
	Late Function	0
	Too late Condition	Rentang ≤ 25 =true

	Too late function	0

Based on results of calculation that has been describing on table 2, then the results of test case for payment of consumer is as follows :

Table 3 Mapping of consumer payment

No.	Consumer	Payment range	Intime	Almost late	Late	Too late
1	John	29	0,00%	0,00%	10,00%	80,00%
2	Katrin	5	52,00%	38,00%	0,00%	0,00%
3	Inti	2	80,00%	15,00%	0,00%	0,00%

2.5 OLAP Implementation

OLAP implementation applying reports into multidimensional models in accordance with the format of report that became the strategic needs. OLAP implementation presented a lot of data into multidimensional forms so that it becomes easier to do an analysis of the information.

For OLAP implementation in classical data warehouse can be seen in figure 7, based on need of strategic information to consumers on the service that is used in periods per month and per year.

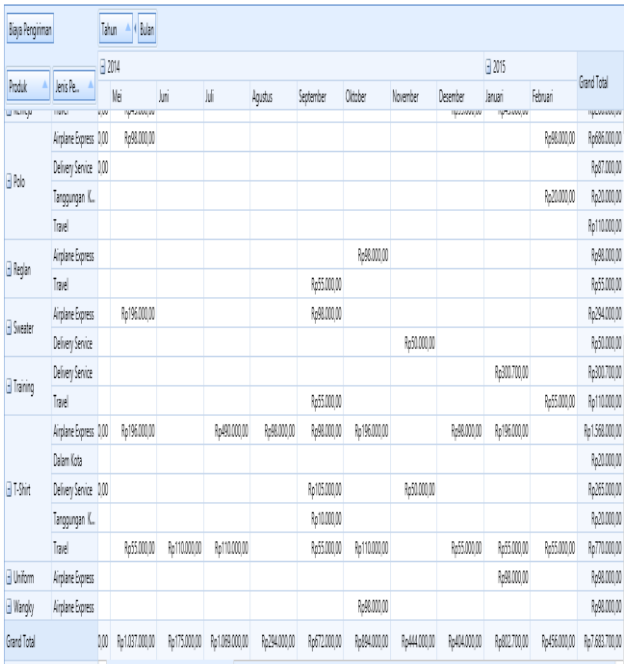


figure 7 OLAP for classical data warehouse

The application of OLAP in fuzzy data warehouse using conditional formatting that is set to the smallest value of range and the lowest, so can be easier to take decision and perform analyzing. Figure 8 is OLAP implementation of fuzzy data warehouse on strategic information needs of payment percentage against booking made by consumer in the period per year.

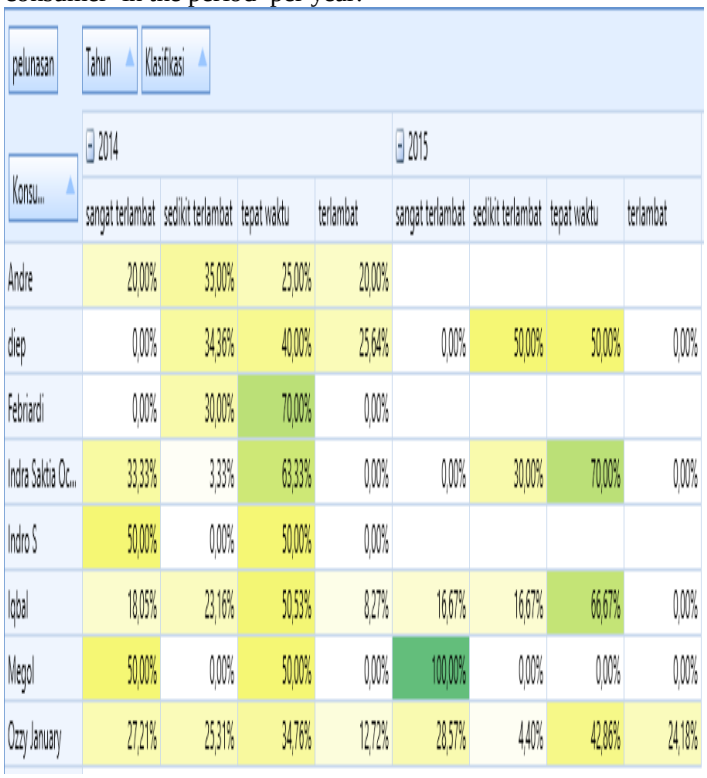


Figure 1 OLAP for fuzzy data warehouse

3. CONCLUSION

Data warehous development helped owner by data design optimalization that support for seeing from various aspects. Data on create integrated to make it easier to optimize data design of the data warehouse.

Using fuzzy concept in data warehouse helps owner to seen the facts with some classification that

has membership in accordance with the rules. So ambigous fact can be analyzing with efficiently.

4. REFERENCE

[1] P. Ponniah, Data Warehouse Fundamental, New York: John WIlley & Sons, 2001.

[2] L. Sapir, A. Schimilovici and L. Rokach, "A Methodolgy for The Design of a Fuzzy Data Warehouse," *Intelligent System,* vol. 1, no. 1, pp. 2.14 - 2.21, 2008.

[3] D. Fasel, Fuzzy Data Warehousing for Performance Measurement : Concept and Implementation, Switzerland: Springer, 2014.