International Journal of Software Engineering and Its Applications
Volume 10, Number 8, August 2016

IJSEIA

ISSN: 1738-9984

SERSC
Science & Engineering Research Support soCiety

Foreword and Editorial

International Journal of Software Engineering and Its Applications

We are very happy to publish this issue of an International Journal of Software Engineering and Its Applications by Science & Engineering Research Support soCiety.

This issue contains 17 articles. Achieving such a high quality of papers would have been impossible without the huge work that was undertaken by the Editorial Board members and External Reviewers. We take this opportunity to thank them for their great support and cooperation.

The paper "A Study on Digital Signage for Facility Management based on Wireless Network" states that there is a growing demand for a system that allows users to go online and work at anytime, anywhere, as the smart phone is widely distributed and the technology is developing. In line with such a trend, this thesis has designed a system using a display with OS and wireless network to help check and manage facilities of various industries as well as producing and delivering various contents. The system is based on the Internet of Things which supports the function of inspecting and maintaining facilities while providing various contents on display devices. This also receives from a service provider and displays advertisement contents real time. Therefore in short, the system manages facilities of a building including cleaning service management and displays various contents.

In the paper "Ciphertext-Policy Attribute-Based Encryption for Access Control of Data in Cloud", in Distributed systems, the users with a certain set of attributes can only be able to access the data. At present this process can be done through a trusted server where they will store the data and there will be certain constraints on the access of the data. In this case there will be a possibility to compromise the data and so the confidentiality of the data is lost. An Attribute-Based Encryption (ABE) is an encryption scheme, where users with some attributes can decrypt ciphertexts associated with these attributes. Now this is their turn to develop a system with a more complex policy of access of the encrypted data and which can be called as Ciphertext Policy Attribute-based Encryption (CP-ABE). By using this method the information can't be traded off even through the trusted server where the information is put away. These methods are also secure against the collusion attack. In this method attributes are generally assigned in the form of access trees. The attributes are placed at the leaf nodes of this access tree. In older Attribute-based encryption strategies encrypted data is described by the attributes and policies are given to the user's keys, while in their system users credentials are described by the attributes and there will be a policy where it tells us about who should access or decrypt the data. So, this type of access method is very much closer to the Role-based attribute-based encryption

The purpose of the research entitled "A Study of Control Effect of Participation Mode in Open Collaboration Platform" is to try to find out important factors as well as individual motivation for creating co-creation and co-decision in an open collaboration platform (OCP) which attempts to deduct collective intelligence in various fields. The research was surveyed on 294 people who have had experiences on open collaboration services unconsciously or consciously. In particular, the effects of factors on participating OCP are tested under two by two matrix of the participation modes, one for level of knowledge

(general knowledge vs. professional knowledge) and the other level of the participation (simple evaluation vs. development/business) as a control variable.
The study "Pepper or Roomba? Effective Robot Design Type Based on Cultural Analysis between Korean and Japanese Users" explored the impact of robot design types in Korean and Japanese culture on product evaluation, product usefulness, and purchase intention. They conducted an experiment comparing a group of functionally designed robots having single functions each and a biologically inspired robot having multifarious functions itself. The results showed that Korean participants evaluated the functionally designed robot more positively than the biologically inspired robot on product evaluation and product usefulness whereas the evaluation was more positive toward the biologically inspired robot than the functionally designed robot for Japanese participants. While Korean participants were willing to purchase the functionally designed robot as they perceived it positively and useful than the biologically inspired robot, Japanese participants showed no significant difference on purchase intention by robot design types.

In the paper "An Efficient Approach based on Genetic Algorithm for multi-tenant Resource Allocation in SaaS Applications", in recent years, the use of cloud services has been significantly expanded. The providers of software as a service employ multi-tenant architectures to deliver services to their users. In these multi-tenant applications the resource allocation would suffer from over-utilization or under-utilization issues. Considering the significant effects of resource allocation on the service performance and cost, in this paper they have proposed an approach based on genetic algorithm for resource allocation which guarantees service quality through providing adequate resources. The proposed approach also improves system performance, meets the requirements of users and provides maximum resource efficiency.

Paper "A Study on the Analysis of De-Identification of Website of Domestic Education Institutions and Government Organizations" states that due to the recent development of smart phone, SNS has been rapidly increasing causing the amount of data from IOT explosively to be grown. Hereupon, personal information leakage has become a serious issue, and de-identification has been applied according to personal information protection guideline. However, most of the files attached on the board have not been de-identified. It is not possible to predict how data of a certain person are to be used at where. In addition, the process is much complicated and frequent that it is still not possible to prove the usage of personal information from the result of analysis even if data of a certain person are used. Hereupon, there has been an increasing fear on information holders due to the possibility of information leakage. In order to identify the risk of sensitive personal information leakage, this paper is intended to collect data from websites of domestic education institutions and government organizations, analyze how much de-identification plans for guaranteeing safety of individuals is correctly applied, and reveal the risk from sensitive personal information leakage.

In the study "Enhancing Fault Divination Accuracy Using Naïve Bayes Classifier with PYTHON and PHP", Programming Fault forecast has turned out to be most essential in programming Development uncommonly in programming Testing. The exact extrapolation of issues in conundrum can help to patch test effort, which decreases expenses and repair the nature of programming. Issue forecast model utilizing object situated measurements for code, datasets as info qualities to anticipate the issue probability by Naïve Bayes Classifier and these mock-ups have been far and wide utilized for bunching and grouping likewise exceptionally flawless eccentric to Bayesian systems for expansive range likelihood evaluation, generally in shortcoming expectation. In this paper, Naive Bayes classifier has been actualized on different consistent datasets.

Authors of the paper "Dynamic Relocation of True-False Questions Using Ready-made Arrays with Random Numbers" proposed a system to generate true-false questions using ready-made arrays instead of random functions in order to reduce the time to give true-false test set. The system consists of two phases: one is automatic question generation and the other is question selection for make question test. The first phase reads an informative sentence from a text input, and then makes false sentence using replacing keyword in the informative sentence by an antonym or adding a negative adverb to verb in the informative sentence. The second phase selects some sentences from both true sentences as the informative ones and false sentences from different informative ones, and then relocates the selected sentences using ready-made array instead of random function whenever applicant wants to take a test. The array contains fixed random numbers made by random function in advance, so the system can take random number from the array instead of generating new random number to reduce the time to take random number. Thus, the time to relocate true-false questions using the ready-made array is 1.77 times faster than one to relocate the same questions using random functions.

The paper "Design of a Smart Learning System" introduces a design of their Smart Learning which is an application running on smartphones. Smart Learning is an open, location based, context aware, and video on demand (VOD) app for education. This app is open in that a user can upload the user's own educational contents. It uses the Global Positioning System (GPS) for outdoor positioning and smartphone sensors for indoor positioning. It also guesses the user's context with sensor values and information recorded in the smartphone. Once it determines the user's context, it displays a list of learning materials that seem to be most helpful to the user, and when the user selects one out of them it starts rendering the selected one. Another uniqueness of their design is that all the functions needed in implementing the app are published as web services so that other developers can easily reuse them. Nowadays, almost everybody has a smartphone and they always carry their smartphones. Therefore, they can learn something with Smart Learning whenever and wherever the user wants to.

Paper "Hospital Record Search using RDF based Information Retrieval" states that in Today's World, the usage of Web is increased in a very fast manner. The User who is surfing the Web always wants to extract the meaningful information from the resources. As the number of user increased, the Web is diffusing day-by- day. Due to this diffusion, there is enormous lacking in retrieving the relevant information. To overcome this, Semantic Web Technologies such as Resource Description Framework along with Document Object Model are used to bring the clear answers to user queries by taking advantage of the availability of precise semantics of information in the context of the search engines. The Resource Description Framework (RDF) is defined by the World Wide Web Consortium (W3C) which is one of the enrich entity which is mainly used to encode the meaning of the data.

The study "Innovativeness or Confidence? The Effect of Consumer Innovativeness and Self-efficacy on the Acceptance and Diffusion of Innovative Technology" analyzed the effects of consumer innovativeness and self-efficacy on different age groups' acceptance and diffusion of innovative technology. For this purpose, they measured delivery application service users' innovativeness and self-efficacy, and found each age group's levels of satisfaction, intent to use again, and intent to recommend. They used this data to analyze the influence that these variables had on their satisfaction with innovative services, and the diffusion of such services. This study found that for people in their 20s and 30s, the interaction effect between consumer innovativeness and self-efficacy was not statistically significant, while for people in their 40s, there was a significant interaction effect. Higher consumer innovativeness was correlated with significantly higher

satisfaction, intent to use again, and intent to recommend. Meanwhile, higher self-efficacy was correlated with innovative services' acceptance and diffusion. Higher perceived self-efficacy for subjects in their 40s was correlated to significantly greater deviation. Additionally, a significant three-way interaction effect occurred between innovativeness, self-efficacy, and age group. An in-depth analysis and further research would be necessary to gain insight on why such deviations occurred.

In the study "Personal Health Information De-identified Performing Methods in Big Data Environments", the field that shows the most promise among the application areas of Big Data is the medical sector. They must first deal with the problems of the invasion of privacy and misuse of personal information before they can utilize personal health information. There is a method known as the de-identification of personal health information which is one of the methods that can be used to perform the tasks of the protection and the utilization of personal health information at the same time. De-identification refers to the process of making it impossible to know the identity of an individual just by the information that is revealed. The biggest problem related to de-identification is the phenomenon of re-identification. Even with information that at first cannot be used to readily identify an individual, when enough of it has accumulated in various categories, it may become possible to identify the hidden individual behind it. What makes de-identification difficult is that one cannot completely rule out the possibility of re-identification, and that the information becomes an object of much regulatory legislation if it is re-identified as personal information. Thus, it is a priority to reach an agreement among the many parties involved since they need to organize the data into categories that are actually used. This study seeks to analyze the current state of de-identification measures which are some of the protective measures for personal information for the safe utilization of Big Data in the medical sector, both in and out of Korea, and to propose implementation plans for safe de-identification. Furthermore, this study advocates an active consideration of the establishment of a "central tower" agency which will be able to carry out the much-needed continuous monitoring for re-identification as well as the assessment of the adequacy of the de-identification methods.

The aim of the study "Augmented Reality Application of Sign Language for Deaf People in Android Based on Smartphone" was to create augmented reality application that can display three-dimensional animated hand movements on a media card and implemented on the Android operating system. Modelling was begun with doing observation, in which the object of the observation was people who have a deficiency in hearing and speaking or often called deaf. The system made aimed at helping normal people in studying the letter cues when trying to communicate with the deaf. The results obtained is an application created with augmented reality and can be used by anyone having an Android-based smartphone that wants to recognize the letter cues by using letter card as the card marker. So that, when the camera is directed to the card compiled into a single word, the letter marker on the card will be detected by the application and will appear animated hand movement of each letter scanned by the camera in real-time. This is the purpose of the augmented reality on this letter recognition application, namely giving rise animation in real-time that can attract someone to study the letter cues.

The paper "Test Framework Development for Software Reliability Test using Formal Method" introduced formal method as a new method to evaluate the reliability of software that has been developed. By using a mathematical notation called formal specification, this evaluation method describes the characteristics and functions of system that includes software or hardware and proves what can be proved using a proving technique. It can give a clearer answer than other techniques because it guarantees completeness, correctness, and consistency, which are the general characteristics of math, and evaluates

reliability by mathematical technique. Since it is difficult to express a mathematical notation in a UI, a framework was developed by using it to show the measured results on a screen like other general software. This framework was named FMTF (Formal Method Test Framework) and was made to test the program of general web development. For the performance experiment of FMTF system, the parallel test, loop test, conditional branch perform test, time control test, etc. of web program were experimented. According to the measurement result, the reliability was low in the beginning of project due to many factors such as analysis design but gradually improved by error correction and debugging. By suggesting a guide on how to correct what errors, the paper could also provide a method to improve reliability.

In the study "Integration Scheme of Data Marts using Dimension Sharing Method", business is an entity that produces big data nowadays. A business scope, for example, a company will divide its activities part into a new body which will be focused on another specific field. The implementation warehouse data in company with that criteria will be favorable for the party that makes a decision since they get brief data to ease decision making which is strategic for the enterprise. Demand for the vital information creates a problem when the company has no draft of how the company condition in the future specifically in implementation the Management of Information System that will be built. The solution given is by creating a Bottom-Up Bus Data Warehouse Architecture by applying dimension sharing concept which will integrate all data marts. In realizing that solution needs an application that has the main feature that integrating data mart and does (Online Analytical Processing) as presentation information media for the end user.

The research entitled "A Study on the Causal Relationship on the Information System to Compare Hotel Information System with Hospital Information System in Korea" aims to investigate through causal relationship, how the hotel information system and hospital information system affect the strategy, employee satisfaction, work process, and results. This study aims to figure out how the service information system influences internal factors. Usually, hotels and hospitals represent the service industry. However, they believe that those two are actually quite distinct from each other. Therefore, they aim to differentiate both the hotel and hospital information system. As a result, the information system has positive influence on strategy, work process, and employee satisfaction both hotel and hospital industries.

In the paper "Noise Optimized Channel Selection Filter Design for WCDMA", a low noise 6th-order elliptic low-pass filter for WCDMA DCR is reported. For the noise optimized characteristics, noise analysis, design procedure and techniques are described for both biquad circuit and 6th-order filter. The filter is implemented in a 0.35 BiCMOS technology and has 2.3MHz cutoff frequency. The filter consumes 3.8mA current from 1.8V supply and presents 60dB stopband rejection, $27.8\,nV/\sqrt{Hz}$ input referred noise and 22.4dBm IIP3.

August 2016

Jinan Fiaidhi, Lakehead University, Canada

Editor of the August Issue on
International Journal of Software Engineering and Its Applications

Editor-in-Chief of the IJSEIA Journal

General Information of IJSEIA

Bibliographic Information

- ISSN: 1738-9984
- Publisher: **SERSC**

  Science & Engineering Research Support soCiety

Contact Information

Science & Engineering Research Support soCiety

- Head Office: 20 Virginia Court, Sandy Bay, Tasmania, Australia
- Phone no.: +61-3-9016-9027
- Email: ijseia@sersc.org

Publication and Update

- Last date of Every Month

Journal Aims

IJSEIA aims to facilitate and support research related to software engineering and its applications.

Our Journal provides a chance for academic and industry professionals to discuss recent progress in the area of software engineering and its applications.

To bridge the gap of users who do not have access to major databases where one should pay for every downloaded article; this online publication platform is open to all readers as part of our commitment to global scientific society.

Journal Topics

The topics covered by IJSEIA include the following:-

- Agent Technology
- Aspect-Orientation and Feature Interaction
- Case Studies
- Communication Systems & Networks
- Component-Based Software Engineering
- Cost Estimation
- Data Mining, Knowledge Discovery, Artificial Intelligence
- Design Patterns and Frameworks
- Economic and Financial Systems
- Embedded and Real-Time Software
- Empirical Software Engineering and Metrics
- Information Engineering
- Information Management Systems, MIS
- Intelligence
- Middleware Design Techniques
- Mobile/Wireless Computing & Applications
- Modeling with UML
- Numerical (Statistics) Analysis Development of Algorithms
- Object-Oriented Techniques
- Parallel and Distributed Computing
- Programming Languages
- Project Management
- Project Management and Maintenance
- Quality Model
- Reverse Engineering & Reengineering
- Simulation and Design Methodology for H/W and S/W System
- Software Architectures & Design
- Software Components and Reuse
- Software Development Methodology
- Software Economics & Cost Modeling
- Software Policy and Ethics
- Software Process Improvement
- Software Process Models
- Software Requirements Engineering
- Software Safety and Reliability
- Software Security

- Software Testing & Analysis
- Software Tools and Development environments
- Theory, Formal Methods and Tools
- Use Cases, Scenarios
- Visual and Multimedia Computing
- Web Engineering, & Web-Based Applications

Editorial Board/Reviewers

IJSEIA is indexed by:

- EBSCO
- ProQuest
- ULRICH
- DOAJ
- Open J-Gate
- Cabell
- SCOPUS

Editorial Secretary

- Ronnie D. Caytiles
- Regin P. Conejar

International Journal of Software Engineering and Its Applications
Vol. 10, No. 8 (2016), pp. 159-172
http://dx.doi.org/10.14257/ijseia.2016.10.8.15

Integration Scheme of Data Marts using Dimension Sharing Method

I Made Sukarsa[1], Putu Wira Buana[2], Parangkan Nurtantio Quidar[3] and Ni Wayan Wisswani[4]

[1,2,3] *Dept. of Information Technology, Fac. of Engineering, Udayana Univ., Jimbaran, Bali, Indonesia*

[4] *Dept. of Informatic Management, Bali State Polytechnic, Indonesia*

[1] *e_arsa@ymail.com*, [2] *wbhuana@gmail.com*, [3] *nurtantio.q@gmail.com*, [4] *wisswani@yahoo.com*

Abstract

Business is an entity that produces big data nowadays. A business scope, for example, a company will divide its activities part into a new body which will be focused on another specific field. The implementation warehouse data in company with that criteria will be favorable for the party that makes a decision since they get brief data to ease decision making which is strategic for the enterprise. Demand for the vital information creates a problem when the company has no draft of how the company condition in the future specifically in implementation the Management of Information System that will be built. The solution given is by creating a Bottom-Up Bus Data Warehouse Architecture by applying dimension sharing concept which will integrate all data marts. In realizing that solution needs an application that has the main feature that integrating data mart and does (Online Analytical Processing) as presentation information media for the end user.

Keywords: *Data Warehouse, Data Mart, Integration, Dimension Sharing*

1. Introduction

The development of information technology has spawned a lot of data in the fields of science, business, and government. The ability of information technology to carry out the data collection far exceeds the capacity of these technologies to analyze, summarize (summarizing), and extraction of the data into a piece of knowledge. The data warehouse is a technology that will address these problems.

Characteristics of companies that tend to grow in the division or part thereof and have growth data are relatively very high indicating the company needed a data warehouse. Bottom-Up Data Warehouse coupled with Bus Scheme for the integration and transformation of data into the data mart will be a practical solution when the company did not know the structure of future operational systems and operational systems in each section constructed independently (autonomous) [10].

There is a possibility between data marts division can be integrated with other data marts division. These events can occur if a dimension in the data mart is shareable (can be shared). Dimensions of different data marts are shareable when the dimensions are identical or similar in both the data mart [9].

A system is needed to integrate the data mart that has a dimension that is shareable [1]. The integration of data marts requires methods or variants in connecting dimensions from different data marts. Database design which used the system to be responsive so that it can be used for a variety of dimensions to be integrated, and can accommodate many dimensions from different data marts. The system can integrate the dimension using a

ISSN: 1738-9984 IJSEIA

method that has been determined, and can follow the change of data in the dimension tables in the data mart source.

2. Literature Review

The collection of theories that come from books or the internet as well as modules that support the research program.

2.1. Federated Engine MySQL

Federated Engine MySQL support Take Over Process automatically when there is no connectivity between the two pieces of the system that have previously been connected [8]. Figure 1 is an overview of the Federated Engine MySQL.

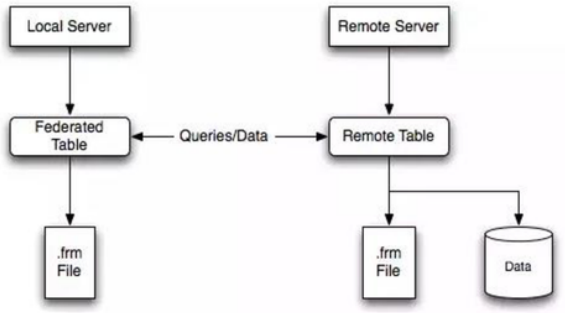

Figure 1. Overview and Table Structure of Federated Engine

Federated engine will create a clone of the table that will be taken the data on the server side then put the clone on the remote server. Remote cloning table is a table that will hold the same data as the original table. The engine will automatically take over or synchronization of the new data changes for no connectivity [8].

2.2. Data Warehouse

The data warehouse is a collection of data that have the nature of a subject-oriented, integrated, time-variant, and is fixed in the collection of data in support of management decision-making process [7]. The purpose of data warehousing is to combine core business and data from other sources in a format that facilitate reporting and decision support [14, 15]. The data warehouse is a method in the design of the database, which support the DSS (Decision Support System) and EIS (Executive Information System). Physically data warehouse is a database, but the data warehouse and database design are very different. Traditional database design using normalization in the data warehouse is not the best way. Based on the definitions described earlier, it can be concluded that the data warehouse is a database that reacts with each other can be used for query and analysis, is the orientation of the subject, integrated, time-variant, unchanged used to assist decision makers [12].

The top-down approach is done by creating a data warehouse design first overall. The manufacturing process is to determine what is the source of data and how to put it into the data warehouse has been designed in advance [7]. The approach recognizes that the development of data warehouse architecture design is the first thing to do in building a data warehouse. Bottom-Up Approach states that data warehouse is formed by data marts that have been built in advance for each division or department [9]. Data marts have been built then combined to build a data warehouse that is intact.

2.3. Data Mart

A data mart is the "retail" level data warehouse where data can be accessed directly by end-users [4]. Data retrieved from the central data warehouse to the data marts to support accurate analysis [6]. The most important requirement at this level is that the data is structured in a way that is easy for users to understand and use the data. For this reason, the dimensional modeling technique is most appropriate at this stage. It ensures that a simple data structure may be made to simplify the user query.

Dependent Data Mart is a data mart that is sourced from a data warehouse created by the Top- Down Approach [7]. Dependent Data Mart has "cleaner" characteristics than the Independent Data Mart because it comes from the data warehouse data that were already experiencing the Extract Transform Load (ETL) process. It also contains data perspective of all the existing information systems and integrated between each data marts there.

Independent Data Mart is a data mart that will be the foundation of a data warehouse built using Bottom-Up Approach [9]. Independent Data Mart has characteristics that are not yet integrated the data and information sourced from an operational system most ready. Characteristics Independent Data Mart database similar to the operating system but has a data structure that is neater and is in conformity with the purposes of the topic or type of information desired.

2.5. Dimension Sharing Integration

The dimensions of different data marts are shareable when the dimensions are identical or similar in both of the data mart [13]. Having dimension tables that are shareable is crucial because it will provide the capability of each data mart integrated with and connected to each other [10]. In particular, the integration of data marts will rely on a common variation of the dimensions of the existing used together at each data marts as a reference in the OLAP reporting will be done.

2.6. Dimension Sharing Variation

There are three variations of sharing dimension that can be done for the integration of data marts called "Dimension Sharing conformed", "Master-conformed", and "Replication-Conformed" [9]. Explanations, as well as a brief example of each of these variations, are as follows. Dimension Sharing conformed is sharing dimension with existing media dimensions that are shared between traffic data marts. Figure 2 illustrates the variation Dimension Sharing conformed as integration solutions.

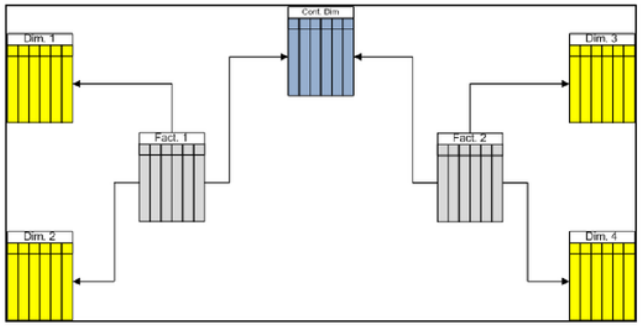

Figure 2. "Dimension Sharing Conformed" Integration

Master-conformed Dimension is the integration of data marts by using "slices" of the same data and then make a new ID from the intersecting data. Figure 3 shows the Dimension Integration Sharing by using Master-conformed.

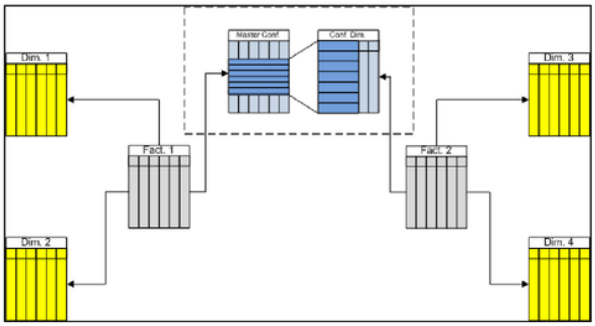

Figure 3. "Master-Conformed" Integration

Confirmed Replication Integration is the integration of data marts do the management company by determining the dimension to be integrated to be replicated in each system.

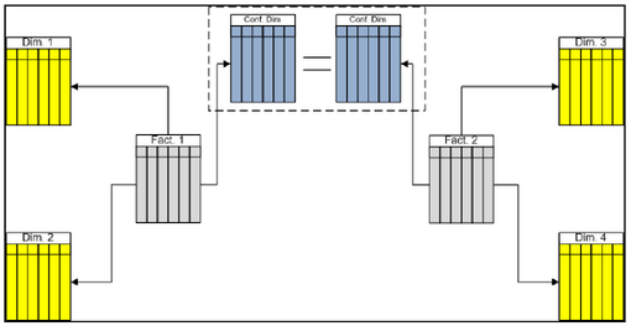

Figure 4. "Replication-Conformed" Integration

2.7. Master Data Manager

The local data marts contain atomic data taken from the source. Atomic data in question is data that becomes a master in its system [1]. The atomic data will be critical in the integration of each data mart. The scheme used is to create a middleware that can accommodate the atomic data. Middleware will act as Master Data Manager will maintain the authenticity of the data so that it complies with the perspective of management in each system. This architecture also follows bus architecture using the sharing dimension in different data marts on the local and corporate level.

The integration by using a Master Data Management will indirectly reduce the value of ownership of the data against the original owner of the data [5]. It can happen because it will be the master data object to be accessed by multiple systems. However, it will result in integration costs more "frugal" because these values focused on a Master Data Manager which is assumed to contain accurate data sourced from respective operating systems.

Integration of data in each data marts is the key factor to support the service features of the distributive process [2]. Data integration is a major prerequisite for master data exchange with across system boundaries. Therefore, the purposes of creating a consistent data distribution contained 4 Architecture Master Data Management to be a reference in establishing a Master Data Manager. An explanation of each of these architectures is as follows.

2.7.1. Central Master Data System: Master data system is set up to distribute the global master data from one system to various applications. As a result, all systems use the same master data and will create a new unique identifier of the master data is commonly called Primary Key Global (Global Primary Key) [3].

2.7.2. Leading System: Build Master Data Management with Leading System approach is to utilize the data contained in the existing system as reference data construction of master data.

2.7.3. Master Data Harmonization via Standards: This method involves the entire company a standard definition that describes the structure of the parent dataset. Global attributes are defined must have the same meaning and significance in all applications. Integration layer ensures that the master dataset is organized in the same way in every system. No mapping with this approach, so that enable data integration is to replicate tables between systems. Each system will have a table with the same attributes with other systems are interlinked.

2.7.4. Repository System: Repository approach is a cross-repository system that is applied to all of the data involved. Repository described as a table stored centrally containing core data that will meet every transactional data related systems. If an application requires customer data, Receiver System sends a request to the repository and the repository will receive an answer in the form of the desired data.

3. Schematic System Design

The design of the system integration of data marts using sharing dimension can be seen in the following figure.

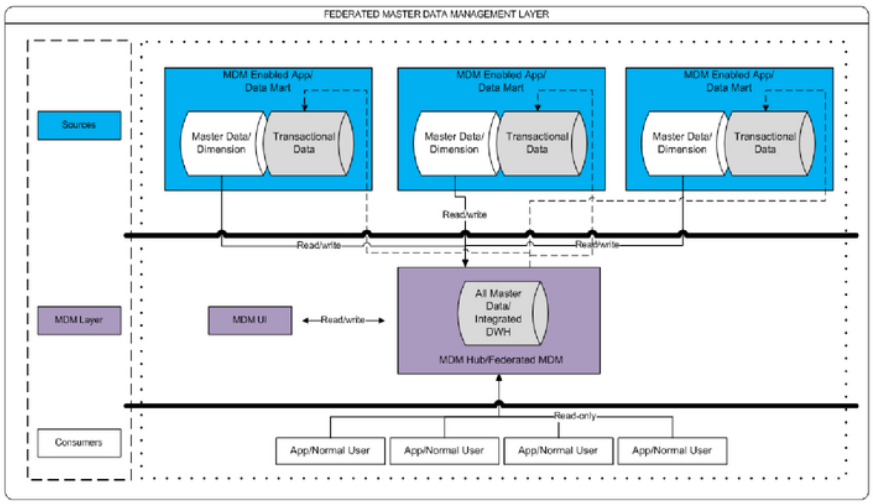

Figure 5. The Design of System Architecture

Figure 5 shows the components contained in the prototype to be built. Three layers are included in the scope of business process systems that have been created. Source layer is a layer of data sources that will be the object of integration. Source layer contains the data of origin data mart of a division within the company who want to integrate their data. In this case, the data mart must be qualified to perform the integration that has a standard field that will be an identifier.

A design scheme of the Master Data Management (MDM) Layer is done by collecting all the necessary requirements of the users and identifying the data sources of the received data from the Source Layer [11]. MDM layer is a layer that functions to collect all dimensions of good will and have been integrated. In this regard, the underlying MDM also will synchronize the data that has been integrated to maintain the consistency of the data. MDM layer also has a user interface that will connect users with dimensions MDM system that will be integrated. Users on this layer only have the right to make changes to data at the level of the connection information is used, but not at the data level.

Layers and layer source MDM must always be connected to one another to maintain data consistency. Implementation of maintenance is performed using Federated Engine and custom query. Engine Federated has the advantage of automatic take-over capabilities that synchronize data from previously disconnected connection to reconnect. Meanwhile, the custom query is useful for consistency management of data that has been integrated.

Connectedness built using an API offers a level of security which is good because the user does not need direct access to the database; it only needs to input in the form of an argument which is a flag which will provide a response in the form of data requested by the request flag is done. Use of the API will be very helpful because the API is cross-platform in the sense not depend on any particular technology implementation. Microsoft has another view of the API by stating that the Web Service has the disadvantage that the nature of "stateless". The nature of deciphering the interactions that occur between the client and server are using the Web Service as a data exchange was brief, and when no data is exchanged in a given period, the server and the client will not have the attachment or knowledge between each other.

A specific example of this statement is that if the client makes a request to the server and then the client will receive some information. There will be one time the crash will occur due to a power outage. The crash will make the server never knows that the client is no longer active. Servers require a way to keep track of what clients are doing and also to determine when the client is no longer active at a particular time. It is important because the server must know what are the activities performed by the crash client. These activities could be the addition of new data, data updates, or deletion of data. API can not take over in real-time the data changes that occur on the client. Figure 6 is an overview of the data exchange process that occurs via the API.

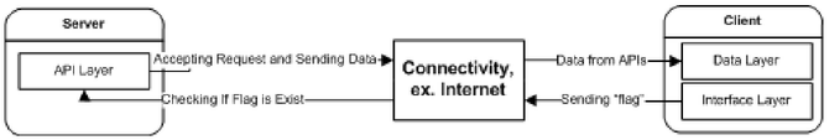

Figure 6. Data Exchange using API

Figure 7 is an overview of the integration scheme prototype integration of data marts using Dimension Sharing Method along with an explanation of each of the components and processes.

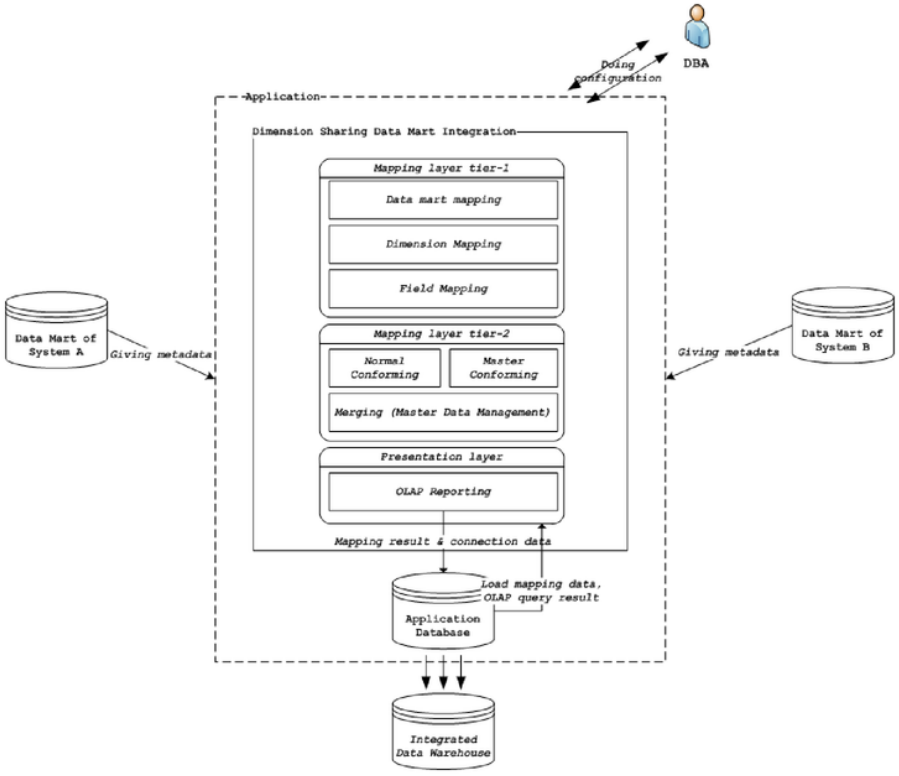

Figure 7. Overview of Data Mart Integration System

Back-end system created in which there are three layers in the architecture of the system being designed.

- ***Data Mart System A, System B:*** *The A system in data mart is one of management information systems at the agency. This type of data mart already contains a data set which is ready to get mapped with the target data from the data mart. The data mart that will be integrated has been built using DBMS platform. The data mart will provide metadata which will become a media of integration.*
- ***Mapping Layer Tier-1:*** *Mapping Layer Tier-1 is an application layer that has the features needed to map the elements within the integration starting from mapping the data mart to field-level mapping. Field mapping will be useful for reference data mapping to be performed on the next layer or coating.*
- ***Mapping Layer Tier-2****: This layer is used to the deeper stages of mapping that is up to the level of data. The mapping in this layer will be an indicator of successful integration because the Master Process in Data Management occurs to keep the similarity between the meaning of the data in each data mart.*
- ***Presentation Layer:*** *This layer will be used to display the results of the report in the desired visualization in the form of a table that contains information about the understanding data that occurs. The resulting display will be an indicator of the successful integration. Additionally, the visualization is also a necessary condition of a data warehouse.*

- ***Application Database:*** *Database application is a database that will store the data transactions occur in application integration. The stored data is in the form of mapping configuration and master data management.*
- ***Integrated Data Warehouse****: Integrated data warehouse is the result of the integration of data marts that has been done. The data warehouse will have a dimensional architecture and will be included in the category "Bottom-Up Data Warehouse" because it was built gradually based on the prepared system.*

4. Proposed System

Figure 8 is an application schema in the integration process. The integration of data marts requires access to each data mart that will be integrated. The Integrator System is in charge of integrating (Mapping) the dimensions that tend to share the two data marts by the scheme system illustrated in Figure 8.

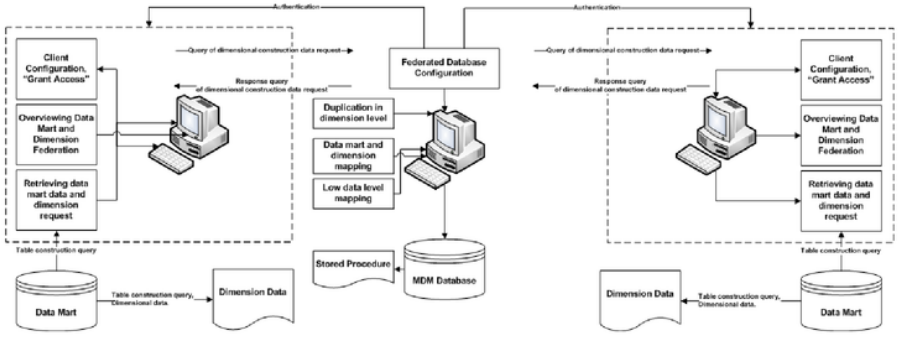

Figure 8. Overview of the Integration Process

Here is a summary of the detailed process of the integration process that will be performed by the system.

- *The Administrator on the MDM server will be logged into the Integrator System.*
- *The Administrator on the MDM server will authenticate to the client system that will perform data integration.*
- *The Administrator on the client's server will grant the access to the MDM system for duplicating the data marts and dimensions.*
- *The Administrator on the MDM server will make the process of replicating the dimensions based on the host, port, username, and password information.*
- *MDM System will give the request to the client system about the query used to build the integrated dimension.*
- *The client system provides the answer, such as the query constructs used to create the integrated dimension.*
- *The MDM system will create a duplicate dimension by utilizing the MySQL Federated with query parameter construction dimensions given.*
- *The Administrator on the MDM Server will start the process of mapping each integrated dimension.*

- *MDM system will map the dimensions with the existing data on the origin dimension by the instructions of the administrator. This stage will give a stored procedure that will maintain the consistency of the data that has been integrated.*
- *If the connection is lost, the integration of MDM system will be postponed and will resume automatically by MySQL Federated Engine when the connection is re-connect.*
- *The integration process is completed.*

The integrated data marts come from different systems. Therefore, in using the describe scheme system, the data mart can also come from different institutions. Following the design system scheme illustrated in Figure 7 and Figure 8, a system integrator is a centralized system that will manage all the data marts which connected to the system. Therefore, the service of managing the membership is needed. The membership stated is a data mart that will be integrated with other data marts, with the proviso of identical dimensions. Such services can be considered as Add Party services, which several processes are undertaken to manage the membership. The process, in general, will serve as a tool that will duplicate the dimensions of which will be integrated. Duplication is done from a database or data mart origin to the MDM system.

Here is an overview of Add Party Service contained in the system.

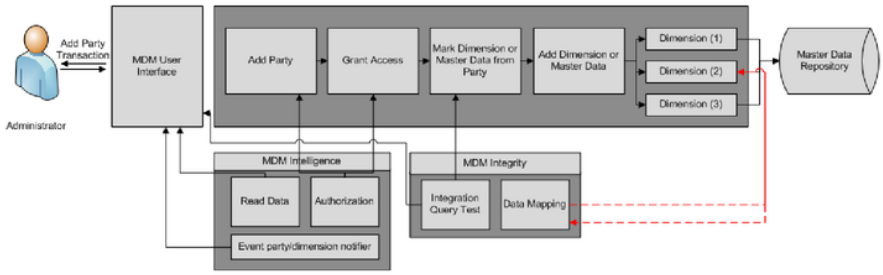

Figure 9. Add Party Service

There are several components in Add Party Service. The components are Add Party, Grant Access, Mark Dimension, Add Dimension, Master Data Repository, MDM Intelligence and MDM Integrity. Here is an explanation of each of these components.

- ***Grant Access:*** *Grant Access is the stage of granting access in Data Mart System that will integrate the data into the MDM system. It is access to the data and duplication withdrawal dimensions.*
- ***Mark Dimension:*** *Mark Dimension is the stage to mark the integrated dimensions. These dimensions are not directly added to the system as it is used to maintain the possibility of user chooses the wrong dimensions.*
- ***Add Dimension:*** *Add Dimension is the process of adding and duplicating the dimensions of the original system to the MDM system. Duplication is done by utilizing Federated Engine MySQL.*
- ***Master Data Repository:*** *Master Data Repository is MDM database system that will accommodate all dimensions that have been or will be integrated. Also, Master Data Repository will also provide all data from related dimensions for mapping purposes at the data level.*
- ***MDM Integrity:*** *MDM integrity is a module that acts as an engine of integration with the task of mapping the dimensions to the data level.*

The Implementation system is done to apply a web-based system to improve the level of cross-platform from the built prototype. The core features that have been implemented in a prototype as follows.

- Add Party
- Add Integration
- Overviewing
- Dashboard

In general, these features will work in order from the addition of integrated dimension (Add Party), integrating to dimensions (Add Integration), a feature for the overview of the integration results that occurs, and feature overview Special (Dashboard), which is useful for displaying the result of integration. A detailed explanation of each of the items that have been mentioned is as follows.

4.1. Add Party

Add Party Features is a feature that serves to perform integrated additional dimension. Extra dimensions of the system are done by utilizing MySQL Federated Engine. Figure 10 shows the user interface of Features Add Party in the system.

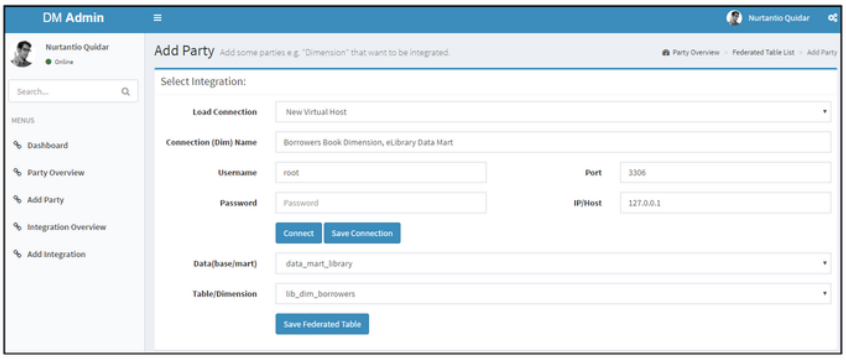

Figure 10. Add Party Feature

4.2. Add Integration

Add Integration feature is a characteristic that serves to integrate the dimension that has been taken previously by the federation data method. The integration is done by using the concept mapping to data which interrelated dimensions are integrated. Figure 11 shows the user interface of the Add Integration Features in the system.

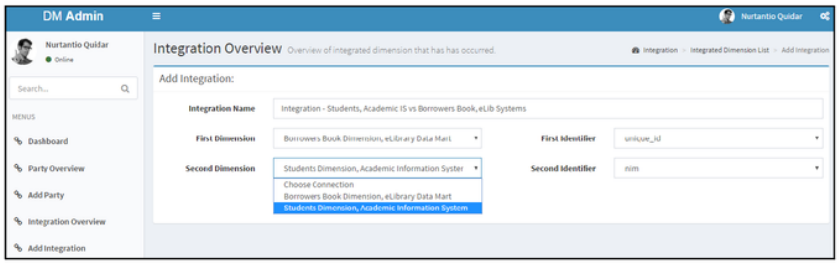

Figure 11. Add Integration Feature

Figure 11 is composed of several critical inputs as indicators of the success of the integration. "Integration Name" Input has a function to stores the name of integration, Combobox First, and Second Dimension contains data about the dimensions of which will be integrated, as well as the First and Second Combobox Identifier will perform load columns based on the selected dimensions. The field will be a key integration will do.

4.3. Overviewing

The overviewing feature is a useful feature to display all input elements that have been carried out starting from the dimension that has been added before integration into the system that has been done before. Figure 12 shows the user interface features Overviewing in the system.

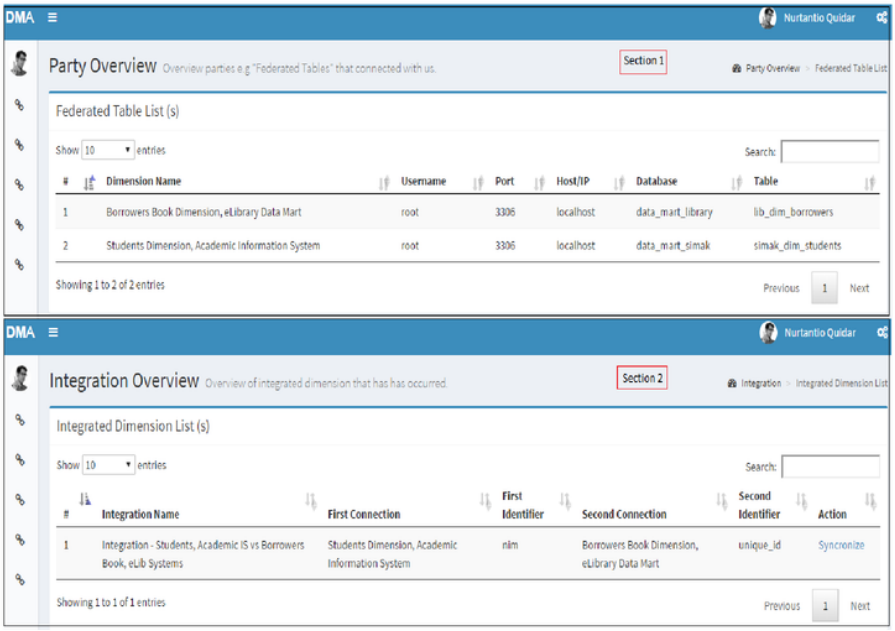

Figure 12. Overviewing Feature

Figure 12 illustrates the arrangement of the components contained in each section view that Party Overview and Integration Overview. In general, each view has a Component Search to perform a search in real-time on integration, pagination, and sorting in real-time.

4.4. Dashboard

Dashboard is the system feature which useful for displaying the result of integration that has occurred from the level of dimension to the data level. Figure 13 shows the Dashboard feature in the system.

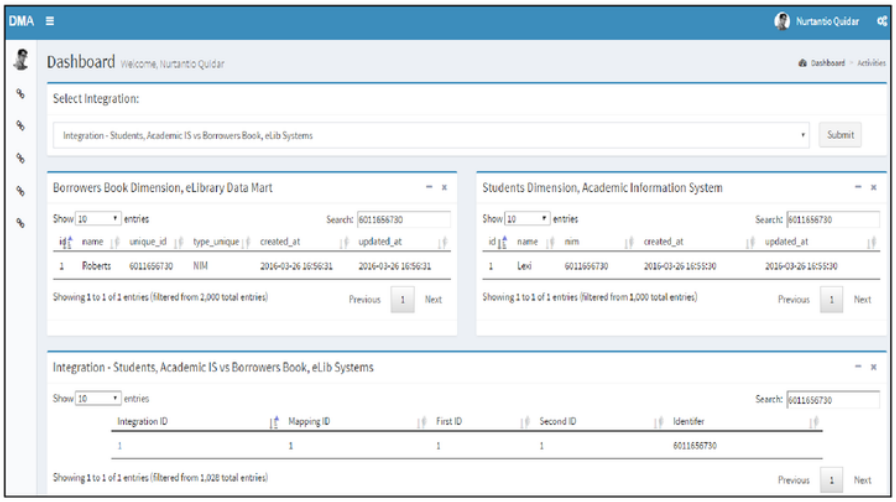

Figure 13. Dashboard Feature

Figure 13 shows the four components with their respective functions. These components include a "Select Integration Combobox", "Dimension Data Components", "Data Mapping Component". "Select Integration Combobox" in charge of loading the data integration that has been done. "Dimension Components" function related to the dimensional displays have been integrated by the selections made previously. "Data Mapping Component" in charge to show the mapping of data that occurs in two dimensions as a form of integration.

Indicators that determine the success of the integration by Figure 13 is the result of the search data sample that has been done. The "Integration - Students, Academic IS vs. Borrowers Book, eLib Systems" is "Data Mapping Component" that displays the mapping of data that has been integrated. First Column Second ID and ID on the component means that the data with ID = 1 Dimension "Borrowers Book Dimension, eLibrary Data Mart" has been linked to the data with ID = 1 Dimension "Dimension Students, Academic Information System". Indicators which became the association identifier is a column that the "Data Mapping Component".

Proof of further integration is done by doing a search of the identifier contained in the "Data Mapping Component". The "Dimension Data Components" in Figure 13 shows the data results of the data with an individual identifier that is 6011656730. These searches generate data name = Robert on the "Borrowers Book Dimension, eLibrary Data Mart" and the data name = Lexi on the "Students Dimension, Academic Information System ". It means that Data Robert and Lexi intertwined with each other and can be integrated with each other because both the data mart origin of the data is built under the same institution.

5. Conclusion

System integration is the process of merging the different subsystems and components to be larger systems. It ensures that every required function integrated subsystems. The system has integrated also means a system that has combined different functions together to work as one. It also applies to a data warehouse. Data warehouses have two general types, namely Top-Down and Bottom-Up. Integration of Bottom-Up is considered more potential to be built because the company is now generally divided into smaller parts and have a system that is autonomous.

Data marts are sourced from the autonomous system will be the object of integration that will form a Master Data Management which will collect all information. Bottom-Up

Integration can be done by applying the data federation or the shared use of dimension at every entity in the system that is considered to have relevance. The linkage that arises can be shaped in conventional perspective of an object such as students with books borrower. The linkages are formed as the Borrower book has a sub-entity as an entity that students borrow books. Sub student entity originating from the Borrower Entities with Entity Student Book will be an object that will be integrated data. This technique is called the Dimension Sharing. Dimension sharing can only be done if each entity has a unique key that is characteristic of each entity. The unique key would be a bridge between the two entities to mutually integrated.

Acknowledgment

This research was supported by the Department of Information Technology, Faculty of Engineering, Udayana University in 2016 and Karya Salemba Empat Foundation.

References

[1] B. Pandya and Dr. S. Shah, "Proposed Local Data Mart Approach for Data Warehouse Architecture", International Journal of Emerging Technology and Advanced Engineering, vol. 4, no. 2, **(2014)**, pp. 101-104.
[2] C. Loser, Dr. C. Legner and D. Gizanis, "Master Data Management For Collaborative Service Processes", Research Center for Contemporary Management, China, **(2004)**.
[3] CRM Market Watch, "Market Dynamics: Customer Data Integration Applications: Part Two", Computerwire, no. 28, **(2003)**, pp. 2-13.
[4] D. L. Moody and M. A. R. Kortink, "From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design", Symbolic Systems, Decision and Information, Portugal, **(2007)**.
[5] E. Hechler, M. Oberhofer and P. van Run, "Implementing a Transaction Hub MDM pattern using IBM InfoSphere Master Data Management Server", IBM Developer Works, New York, United States, **(2008)**.
[6] G. Bukhbinder, M. Krumenaker and A. Phillips, "Insurance Industry Decision Support: Data Marts, OLAP and Predictive Analytics", In Casualty Actuarial Society Forum, **(2005)**, pp. 171-197.
[7] W. H. Inmom, "Building the Data Warehouse, Fourth Edition", Wiley Publishing, United States of America, **(2005)**.
[8] J. Russel and R. Cohn, "MySQL Federated", Book on Demand, Miami, **(2015)**.
[9] K. Ralph and R. Margy, "The Data Warehouse Toolkit", Wiley Publishing, United States of America, **(2003)**.
[10] L. Cabibbo and R.Torlone, "On The Integration of Autonomous Data Marts", Conference on Scientific and Statistical Database Management, Santorini Island, Greece, **(2004)**.
[11] M. Paulraj and P. Sivaprakasam, "Functional Behavior Pattern for Data Mart Based on Attribute Relativity", International Journal of Computer Science Issues, vol. 4, no. 1, **(2012)**.
[12] P. Lane and V. Schupmann, "Data Warehousing Guide", Oracle Corp., California, **(2002)**.
[13] R. Chhabra and P. Pahwa, "Data Mart Designing and Integration Approaches", International Journal of Computer Science and Mobile Computing, vol. 3, no. 4, **(2014)**, pp.74 – 79.
[14] R. G. Allan and D. R. May, "Data Models for A Registrar's Data Mart", Journal of Data Warehousing, vol. 6, no. 3, **(2001)**, pp. 38-53.
[15] V. Poe, "Building A Data Warehouse for Decision Support", Prentice Hall, United States of America, **(1997)**.

Authors

I Made Sukarsa, he obtained his Master Degree in Dept. of Electrical Engineering at Gadjah Mada Univ. in 2003 He currently works as a lecturer in the Department of Information Technology University of Udayana. He has conducted almost 12 journal papers in the field of Data Warehouse, Middleware, and Information Technology Governance.

Putu Wira Buana, he obtained his Master Degree in The Science of Applied Electronics at Brawijaya University in 2007. He currently works as a lecturer in the Department of Information Technology University of Udayana. He has conducted eight journal publications in the field of Emerging Technology And Enterprise And Industry Application.

Parangkan Nurtantio Quidar, He graduated from high school in SMAN 1 South Kuta in 2012. He is currently in the process of studies at the Department of Information Technology, Udayana University, Bali, Indonesia. He has experience as a database administrator in the "Unit Sumber Daya Informasi" Udayana University. He also has the ability as a web developer and data management.

Ni Wayan Wisswani, She obtained her Master Degree in Fac. of Electrical Engineering at Udayana University. She currently works as a lecturer in the Department of Informatics Management of Bali State Polytechnic. She has conducted three journal publications in the field of Middleware.

Integration Scheme of Data Marts using Dimension Sharing Method.

ORIGINALITY REPORT

14%	9%	5%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Sungshin Women's University Student Paper	4%
2	Submitted to Study Group Australia Student Paper	4%
3	www.alexandria.unisg.ch Internet Source	2%
4	Submitted to CBA Student Paper	1%
5	Submitted to Academic Library Consortium Student Paper	1%
6	Submitted to The Nelson Mandela Africa Institution of Science and Technology Student Paper	1%
7	www.sersc.org Internet Source	1%
8	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1%

Exclude quotes	On	Exclude matches	< 1%
Exclude bibliography	On		