JURNAL TEKNOLOGI INFORMASI
PROGRAM STUDI TEKNIK INFORMATIKA DAN SISTEM INFORMASI, UNIVERSITAS BUNDA MULIA
Volume 12, Nomor 2, Agustus 2016
ISSN: 1979-1496

THE DATA WAREHOUSE IMPLEMENTATION BY COMBINING KIMBALL AND INMON METHOD (Study Case In Data Warehouse Of Video Surveillance For Automated Teller Machine In Banking Industry)

Ozmar Azhari
ozumaru@yahoo.com
Sistem Informasi Universitas Bunda Mulia

ABSTRACT

Data warehouse is now the important parts of the organization, especially in Banking Area. Beside for the transaction issue the data warehouse can also be used in managing the transactional data which store by real time with the video surveillance in the Bank. Video Surveillance or known as CCTV ATM is used by Bank as the evidence when the fraud or other banking crime occurred. In this paper, the process of combining data warehouse method can also be implemented for transactional data in CCTV ATM. The Inmon approach is implemented in the development of Data Warehouse framework, while project scoping, and user requirement is based on Kimball. The transactional data first gathered via manual or automatic process via network, where the process begins with the ETL and staging to produce data mart, so data warehouse can produce the analytical report for decision making issue. Furthermore, one of Banking, which is regional development Bank, will become the model of implementation combination data warehouse related to the transactional data of CCTV ATM which gathered from the ATM.

Keywords: *CCTV ATM, Inmon, Kimball, ETL, Staging, Metadata, Data Mart, OLAP*

INTRODUCTION

Data is the most critical part in organization. Without data, the information cannot be gathered and presented as the real condition of the organization. Accordingly, data must be kept and stored in a safe way. Data must be existed when user needs it, hence the repository must be done properly and automatically. Many consequences will be got by the organization when losing the data, which leads to the financial loss, since data is also used to help the management of the organization in making the decision. Therefore, organizations need the data management called data warehouse. Data warehouse nowadays has become the most important part of business system and is the motor of the business. Organizations that have not applied data warehouse in recording their data will face a much worse problem in preparing the data for the business process. As a result, they must dig many paper documents and it will disturb the other activities that quoted from Nugroho (2013).

In banking area, data warehouse and management can also be used in storing the data transaction of Automated Teller Machine (ATM) in video surveillance. It contains the visual image of the transaction in ATM business process. The increase of crime, such as fraud via ATM and claim from the bank's customer also become the importance of data warehouse in ATM field. The data mostly are stored in repository called

Digital Video Recorder (DVR), where the data transactions are kept in hard disk on each ATM. This can become weakness to the organization, since the data can be easily deleted and overwrote when the hard disk in the DVR in status of full. Organization needs report to know that the transaction data is completed during certain period of time.
The following aspects show that the data warehouse in organizations, especially in banking field must consider to the growth of the data processing for the next five years and adapt with its IT Infrastructure, either software or hardware in that period of time. As good data management will be directly proportional to the company's profits, hence there are many implementation of the data warehouse widely used by the organizations. In order to create the data warehouse environment that is appropriate to the organization and lead it to deliver solution and financial profit, we need to understand the concept of the data warehouse and implementation of data warehouse system in field of banking. In the next study of the paper we will discuss data warehouse concept and issues appear in the implementation of each the data warehouse technique.

VIDEO SURVEILLANCE

Poole, Zhou and Abatis (2009) explained that the video surveillance technology utilized storage media and camera to capture images of objects, some of which is to use Digital Video Recorder (DVR) to record the audio and visual data; and the camera or Closed Circuit Television (CCTV) for capturing an image of an object of observation. Developments in the use of Video Surveillance DVR currently developing one of which was already in the storage medium, which according to. DVR recording media used today have changed from the use of magnetic tapes into digital storage in the hard disks and optical media.

Moreover, the obstacle in the application of the DVR is the digital storage method is when the video data stored in the hard disks damaged, due to power problems or failure to follow correct operating procedures. In playing the video data from DVR device, there are two stages using special software from the manufacturer, in which the video data can be pulled from the device using the DVR via a network cable connection, duplicate data to Compact Disc and USB flash drive device is then played using software from manufacturer.

In order to assist the organization of losing the data transaction, the image of video surveillance can also be kept in storage system via network technology where DVR capture and send the output of recording process per transaction per day. Instead, the integrated system and data warehouse needs to be built to manage the data transaction, so that the data transaction can be managed properly. This issue is not discussed in any the reference paper. Therefore, in the next section will be discussed how to managed the CCTV data using the data warehouse.

DATA WAREHOUSE

Padhy and Panigrahi (2012) mentioned that the data warehouse is a repository of an organization's electronically stored data and are design to facilitate reporting and analysis. It can be called as a computer system to archive and analyze the historical data of an organization such as data of sales, salary and other information. Data warehouse is built to separate the historical data from the transactional data, is more static, and the data is collected for business analysis.

Meanwhile, as quoted from Guerra and Andrews (2013), the data warehouse concept is deceptively simple, where data is extracted periodically from the applications that support business

processes and copied onto special dedicated computers. Data are mostly gathered and input from the operational structure of the organization to either computer or server. There it can be validated, reformatted, reorganized, summarized, restructured, and supplemented with data from other sources. The data in data warehouse can be queried and selected based on the requirements of the organization and the output of this process is report for the management to help them in making the decisions. Basically, the data warehouse concept is based on the concept that the management of the organization needs tools and data to analyze and store data in order to help the organization in making the decision.

However, Nugroho (2013) explained that the data warehouse has some characteristics that can be the identity of the services. The characteristics of data warehouse are Subject Oriented or being oriented on the subject, Time-variant, means that the data changing is traced and recorded so that the report can be made by showing the time of the changing, Non Volatile, means that the recorded data cannot be changed, and the last characteristic is integrated in which the warehouse includes all operational data of an organization which is stored consistently. For addition, Guerra and Andrews (2013) explained that the implementation of data warehouse typically reside on computers dedicated to this function; run on a database management system (DBMS) such as those from Oracle, Microsoft, or IBM; retain data for long periods of time; consolidate data obtained from many sources; are built around a carefully designed data; model that transforms production data from a high speed data entry design to one that supports; high speed retrieval.

DATA WAREHOUSE'S STRUCTURE

The data warehouse structures consist of ETL, staging, metadata and analyzing tools (OLAP).

ETL

Extract, Transform, and Load (ETL) is one structure in data warehouses, where data from a variety of forms and formats are processed in such a way into a standard form that is stated by the developer so that the data can be used in data processing. This structure is started from the process of extracting the relevant data in business process at the input phase. The process then continues by transforming data from the previous process to the format of data warehouse and building the keys of the queries. The last process is load the data that has been transformed to the proper format into data warehouse and build the aggregates key of the queries.

Kocharl and Chhillar (2012) explained that the warehousing system has the responsibility to provide proper data cleaning technique to clean the dirty data which occurs in the applications. Also, the cleaned data has to be transformed and to be loaded properly. Moreover, Rizzi (2016), showed that the research on data warehouse self-maintainability and independence has shown that ETL structure can present how to set up the data warehouse in such a way that the maintenance processes can be simplified and made more efficient by avoiding maintenance queries.

Staging

Nugroho (2013) mentioned that in the staging process, the data, which are originated from OLTP database, will be copied to the database of data warehouse through a staging area. Staging area is a temporary area to select data being copied from OLTP database. The data which is copied by data totally has good

quality. Stage is the former place to evaluate the data source from the data warehouse.
Staging also can be defined as a place and process of data standardization occurred. The data that extracted from the OLTP database are gathered into temporary repository, called as staging area. In the staging area, data warehouse system can transform the data into one type as developer stated for the system development.

Metadata
Guerra and Andrews (2013) explained that the data warehouse uses metadata repository in integrating the component. Metadata is used to explain the detail information of the data in data warehouse. It can also assist the integration and transformation of the data in the data warehouse. Metadata is defined as "data about data." It can include something as simple as an average.

While Sen and Sinha (2005) showed that the data warehouses can be used to create and store a great deal of metadata of potentially great value. The metadata contains the basic information of the data, such as source of data, time where the data was created, data model for databases, and transformation rules that converts source data into target data. Data warehouse metadata includes definitions of conformed dimensions and conformed facts, data cleansing specifications, DBMS load scripts, data transform runtime logs, and other types of metadata. Besides, Ramesh (2012) explained that the metadata may become the media that act as repository which stores each and every characteristic of the data items loaded in the data warehouse.

OLAP
Nugroho (2013) explained that OLAP in the data warehouse is a category of software technology that enables the analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the user. It also can be defined as process that is used to request the data in complex form and analyze the big volume of data which is provided in the data warehouse.

OLAP can become a specific method to analyze data in repository for the reporting purpose as requested by the user. To achieve the purpose data are made into specific formats by grouping the data.

DATA WAREHOUSE METHODOLOGY

As introduced by Sen and Sinha (2005), the data warehouse implementation and technique it selves, consisted of two main parts, which are the Inmon and Kimbal. Inmon emphasized the importance of cross-functional slices of data drawn from multiple sources to support a diversity of needs; the foundation of his subject-oriented design was an enterprise data model. While Kimball introduced the notion of dimensional modeling, which addresses the gap between relational databases and multidimensional databases needed for Online Analytical Processing (OLAP) tasks. These different definitions and concepts gave rise to an array of data warehousing methodologies and technologies, which we survey here and provide useful guidelines for future adopters.

Furthermore, based on Ariyachandra and Watson (2011), also mentioned that Inmon advocates the hub-and-spoke architecture (e.g., the Corporate Information Factory), while Kimball promotes the data mart bus architecture

with conformed dimensions. There are other architecture alternatives, but these two options are fundamentally different approaches, and each has strong advocates.

Ramadhan and Soepriadi (2013) distinguished these two methods that Inmon suggests that for storing the data in a normalized form (known as relational modeling) and data marts are granules and appropriated to the requirements of the model built from the data normalized models and this is defined as an enterprise data warehouse. Inmon methodology is usually managed by using the top down methodology, which fit to the big enterprise scale of the data warehouse. Meanwhile, the Kimball emphasizes on Star Schema as a data model in presentation layers which refer to the Data Warehouse Bus Architecture. The star schema is built by extracting directly the source system or by data which are kept in staging area. Kimbal defines the Enterprise Data Warehouse as composite of staging area and presentation layer. This means that the methodology can be known as bottom up approach. Thus, the methodology fits for small scale companies.

The combination between Inmon and Kimball can also occur when the organization decide to choose this process. The process divides into several steps. The process begins by doing standardization, which can be done by extracting data from the raw data then transform into standard form of data. The output of the process is the statistical report which contains summarization of detail data gathered before and the quality assurance process occurs in this phase. The process can describe in the figure 1.

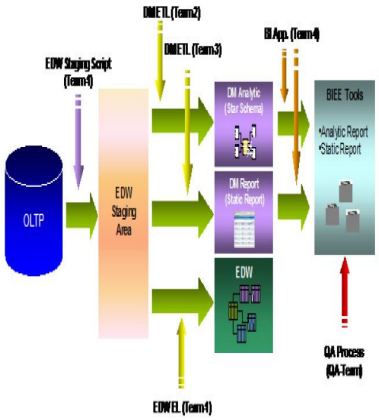

Figure 1. First Phase of Combination Phase

Process continues by accomplishing direct connection from enterprise data warehouse to data mart, the process is implemented into one form of process below:

Data Source -> Staging-> EDW-> DM-> Report

After this process, the User Acceptance Test (UAT) is done to get the accurate report, especially to watch the effect of adding data when the system already implemented in the organization. The process can be described as the figure 2.

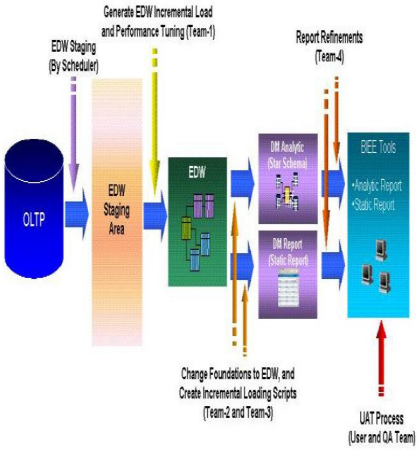

Figure 2. Second Phase of Combination Method

The last process is followed by deploying the data to production server. In this phase the developer has ensured that the data can produce reports which contain analytical data to help organization making decision for the business that is operated, as explained with the figure 3.

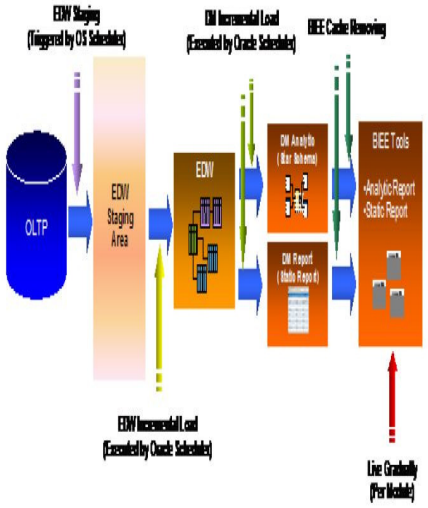

Figure 3. Third Phase of Combination Method

The key success factor in implementing data warehouse methodology in the organization is not only in selecting the technique of data warehouse, however understanding and practicing the development methodology of the data warehouse and application of the best practice of the requirement in the organization. The system analyst must know well the timeframe of the project, the environment of the organization and the development in future time of the data warehouse before deciding the methodology that will be used in the organization or even combining those two methodologies become the approach that suit to the organization's requirements.

Nguyen (2011) mentioned that once designed data warehouse that is user friendly, it will next want to look at operational efficiency. Once the data warehouse has been created, it should be able to carry out operations quickly. In addition to this, it should not have errors or other technical problems. When errors or technical problems do occur, they should be simple to fix. Another thing that will want to look at is the cost involved with supporting the system and will want to keep these costs low as much as possible.

In the next chapter of the paper, the implementation case of data warehouse related to CCTV ATM will try to be discussed of how the data warehouse will improve the management of transactional data in the DVR of the Bank's ATM.

STUDY CASE

Ramadhan and Soepriadi (2011) mentioned that the implementation strategy can be accomplished by identifying the main module and business process and then defined those two things as Data Mart which already classified to each business process. Likewise, the static and analytic report must already be classified to each business process. The next step is that each module will be finished in parallel by using different teams.

The case will describe about the Regional Development Bank in one of the provinces in Indonesia, where in conducting business process in ATM processing, the management has been pushed by the government to record any transactional data which happen in the Bank's ATM either branches or standalone ATM. Bank owns over two hundred ATMs which spread all over the province. Therefore, as management push the IT division to implement the video surveillance in the ATM room, starting in 2009 the Bank began to pair the DVR inside the ATM in every location of the Bank. The data are collected by using the manual method, where hard disk of DVR devices, which

contain data transaction (which called the transaction hard disk), are unplugged from the device and replaced with the new hard disk inside the device. The hard disk can save the data on each ATM for 2 months. Hence, per 3 months the engineer must take the data with manual method as mentioned before.

Procedure was made and engineers must either get data from the DVR in ATMs, they also have to store the hard disk manually in repository case. When the complaint from the customer came, the engineer of IT division must find the hard disk manually and get to related transaction. Since the regulation of the management stated that the transactional data must be kept within five-year period of time, data warehouse must be created for storing the transactional data of the ATM, so that the user can easily maintain and help the user in making decision to manage the ATM and prepare backup devices to cover when the ATM stop to operate because of the device error or when electrical failure occur at the ATM. The IT division then created the step of process in order to accomplish the mission.

The IT division is required to consider that the CCTV ATM of the Bank contains the critical data transactions of the bank's customers, so that the IT division has to ensure that the transaction data must always exist, unless there is error on the environment, such as electrical failure, which is the main threat of video surveillance system. The hard disk transaction data in the hard disk can be defined as staging area. The data are then transferred from the hard disk to storage system which is defined as EDW. Based on the paper that learned before, the implementation of creating data warehouse for transactional data of CCTV ATM will combine between Inmon and Kimball approach.

The combination of the method followed by using the framework of Data Warehouse which origin from Inmon, however the methods for calculating project scope, presentation layer and user process requirements is done by using Kimball. The processes are defined into several steps. First, to recall the transaction data, the engineer needs to Extract and Load the data from staging area to data warehouse. Since transaction data have already fit to the system requirements, the engineer can only extract transaction data manually from hard disk to the storage by separating the data into folders arranged by time of incident. Since transactional data must be classified by the time of incident, this process must be done accurately otherwise the process of finding data that will be occurred in the next phase would be unstable. In the other words, the engineer has to ensure in doing transfer data from the DVR to storage, the metadata that are gathered from the DVR accurately classify before the data transferred or collected to the storage system, so the identity of the data can be kept to for the system's process.

The next process is continued by creating the Extract, Transformation and Load (ETL) process. In this process the engineer can analyze the requirement for establishing the statistic report. The statistic report contains information of transactional data in each ATM which existed in the Bank, including standalone ATM or branch ATM. Furthermore, the statistic report is needed to inform the user in which data is fully or only partial because of the electricity issue or the error of DVR.

The ATM ID could be the key that can represent many things about the data and make relation between the data to the transaction of the customer. The ATM ID also can describe the existence of the data of an ATM. The attributes that

should mention in the system consists of ATM ID, Name, Address, Transaction date, Record start, Record end. The attribute can be obtained from the metadata of CCTV ATM, which gathered from the process where the engineer transforms the data from hard disk to storage system or called as ETL process, which mention before. Those attributes can be described in the table 1.

Table 1. Data Attribute Table

No	Attribute
1	ATM_ID*
2	ATM_NAME
3	ATM_ADDRESS
4	TRANSACTION_DATE
5	RECORD_START
6	RECORD_END

After the statistic report is created, the engineer will then create the Quality Assurance (QA) process to ensure that the transactional data already able to proceed to next step of the data warehouse process. In this phase, after data stored to the server, the engineer should recheck the data transaction of CCTV ATM already complied with the system they made, so that can ease the next process of the system. The next phase in this process is to make a direct connection from Data Warehouse to the Data Mart, so the circuit has been implemented in a single unified process. Then the user business and its QA team should conduct UAT process in order to obtain accurate reports, mainly to see the effect of the increase of data when go live.

The on line process gradually occurs for EDW, DM and reports (called as elements) which has been done in the QA stage. Process of deploying the elements to the production server is also done in this phase. The engineers either from the QA teams and developers needs to perform maintenance process after the process accomplish, so that a stable product is obtained. In the other words, the process stated above shows that the OLAP process occur and as the result, the analytical report should be rechecked by the engineers to ensure the system already comply to the user requirement.

By implementing the combination data warehouse technique, whenever customer came to the Bank, complained about transaction, customer only needs to mention the date and the Branch name which customer did the transaction. Since the metadata of related transactional data is kept in the storage, the IT Division can easily find the date of the transaction by using the ATM ID to the data warehouse and data mart. By the ATM ID, the Bank can also recheck the transaction of the customer to the bank system whether the same to the complaint or instead, the transaction never happened. The image of the CCTV ATM is then showed to the customer as the evidence of customer's complaint. Periodically every two months of period, the system can release either analytical or statistical report by using the ATM ID, the report should contain the existence of the data transaction, and the cause when the data loss occured to the ATM. The management can examine the report that is released by the system and make analytical decision of the action that should be implemented in order to make full existence of the data for the next period, for example at the place where electrical failure appeared the management can instruct the Bank to provide electrical backup. However, the condition of the recent case still has weaknesses where the system does not comply with the data warehouse process, where ETL process manually accomplished. Thus, in the next section of the best practice in integrating transactional data (data warehouse) with CCTV ATM environment will be proposed.

IMPROVEMENT CONDITION

Based on the study case that already described on previous section, the weakness on implementing data base still appear especially in the ETL process, where data are selected by the engineer manually. Therefore, in order to eliminate the manual process, the transactional data of the CCTV ATM must grab automatically per periodic by using network infrastructure or in the other words using bandwidth via router or modem in the ATM, where time of extracting data are scheduled per day. In addition, The DVR must be equipped by LAN Card so that the procedure can be implemented. Furthermore, the IT Division should improve the system in such a way that the system it selves can extract the data automatically from hard disk of the DVR to system storage. The system must set the schedule of extracting transactional data during closing time so that the process may not interrupt the overall ATM connection which may cause bandwidth latency.

After data are gathered to the storage, the staging system occurred. In this phase, the system must ensure in the quality assurance test phase that the transactional data which origin from the DVR comply to the prerequisite of the system, where recording of transactional data must exist at least ninety percent or twenty-one hours and contain metadata of transactional data. This condition is adapted from the environment where there is ten percent tolerance of the downtime caused by electrical failure. DVR will release the alarm when the system failure occurred. The system will still extract the transactional data and alarm released from the DVR to data warehouse when the data existence below the system requirements. However, the alarm will store in data warehouse as downtime description.

The data then will be rechecked in the user acceptance test (UAT) phase. The result of the attribute transactional data that is transformed from staging process will create the new transformed queries which may include attributes in the table 2.

Table 2. ETL Table

No	Attribute
1	ATM_ID*
2	ATM_NAME
3	ATM_ADDRESS
4	TRANSACTION_DATE
5	RECORD_START
6	RECORD_END
7	ALARM_WARNING

Data mart that is created from the process will be processed into the next phase by using tools that can generate the analytical final report from the tool. The report can represent the uptime of CCTV, the statistical graph of alarm, and many more which can assist management to decide the requirement to fulfill the uptime such as provision of backup unit, UPS, and other devices that can support the uptime, so that data existence can be guaranteed within five-year period of time.

CONCLUSION

The use of video surveillance is now directly related to the raise of fraud and other crime in organization business process, especially in Bank industry. In order to give solution to the organization, the data warehouse then implemented as the requirement of business, especially in ATM Area.

In order to enhance the requirement of data that must be kept within five years and overcome the sophisticated problem in finding the data, the data warehouse is implemented. The combination of data warehouse (Inmon and Kimball) is selected where the framework of Data Warehouse origins from Inmon, however the methods for calculating project scope, presentation layer and

user process requirements is accomplished by Kimball.

The process begins by processing the data standardization, to ensure the data in the same format without eliminating the source of the data. The data then move to the storage system. In the other words, the process is also known as ETL process, where the data is moved from staging area to repository area (storage). The IT Division also could utilize the network infrastructure in grabbing CCTV ATM data where the data is extracted automatically from the DVR to storage using network ATM's connection. The data then transferred to the staging area then process using the combination of Inmon and Kimbal method. Likewise, to the previous way of combination where the ETL process is conducted to Data warehouse then produces data mart and by the tools can create analytical report.

The key of transactional data of CCTV ATM is ATM ID, where becomes the main key to describe and connect to other attributes, such as ATM Name, Address, transaction date, record start and end. The next process is to do the quality assurance process in order to produce the statistical reports. After quality assurance accomplish the engineer must again recheck the report by User Acceptance Test (UAT). The last process is OLAP, where the data warehouse produces the analytical report and may assist the bank industry to make decision regarding the ATM transactional data issues within five years as directed by the regulator.

REFERENCES

[1] Ariyachandra, Thilini and Watson, Hugh J., 2011, Which Data Warehouse Architecture Is Most Successful; Business Intelligent Journal, pp. 4-6.

[2] Guerra, Joseph and Andrews, David**,** 2013, Why Do You Need A Data Warehouse; Rapid Decision.

[3] Kocharl, Barjesh and Chhillar, Rajender**,** 2012, An Effective Data Warehousing System for RFID Using Novel Data Cleaning, Data Transformation, and Loading Techniques; The International Arab Journal of Information Technology

[4] Nguyen, Phuc V, 2011, *Using Data Warehouse To Support Building Strategy Or Forecast Business Tend*

[5] Nugroho, Didik, et.al., 2013, Design of Data Warehouse System to Support the Quality Management of Information Technology Based School; IJCSI.

[6] Padhy, Neelamadhab and Panigrahi, Rasmita, 2012, Data Warehousing and its OLAP; MRDM Technology for Decision Support in Business Organizations of 21st.

[7] Poole, N.R., Zhou, Q. and Abatis, P., 2009, Analysis of CCTV digital video recorder hard disk storage system; Science Direct, pp. 85-92.

[8] Ramadhan, Hasnur and Soepriadi, Agus, 2011, Studi Kasus Pembangunan Enterprise Data Warehouse/Business Intelligence (EDW/BI) di Perusahaan Multi Finance Nasional; Seminar Nasional Aplikasi Teknologi Informasi, pp. 95-100.

[9] Ramesh Babu Palepu; Dr K V Sambasiva Rao, 2012, Meta Data Quality Control Architetecture in Data Warehousing*;* International Journal of Computer Science

[10] Rizzi, Stefano, et al., et al, 2016, Research in Data Warehouse Modeling and Design: Dead or Alive?

[11] Sen, Arun and Sinha, Atish P, 2005, A Comparison Of Data Warehousing Methodologies.

.