International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 17 No. 1 Aug. 2015, pp. 149-154

http://www.ijisr.issr-journals.org/

A framework of data warehouse and data mining technique to evaluate teachers performance in University of Agriculture, Faisalabad

Mehwash Rafiq[1] and Sehrish Rafiq[2]

[1]Department of Computer science,
University of Agriculture,
Faisalabad, Pakistan

[2]Institute of Business Management Science,
University of Agriculture,
Faisalabad, Pakistan

ABSTRACT: Data is a vital asset of an organization and every organization tries its best to store large amount of data. Important decisions are made on the base of stored data. University of Agriculture, Faisalabad is the popular university of the Pakistan. UAF is using Oracle database to store the university employees and students record. This database can store data upto Gbs. Data in this database cannot be analyzed to find hidden information or patterns mean ad hoc queries are not supported. Only few tens of records can be accessed at the same time. Data warehouse stores very large amount of historical data. It stores data up to Tbs. Millions of records can be accessed at the same time. Different data mining techniques explore data warehouse to find hidden patterns and determines what was happening, what is happening and what will happen. Data warehouse supports ad hoc queries. Organizations can make proactive and knowledge driven decisions for more profit and improvements in business. We proposed a framework of Data warehouse for the University of Agriculture, Faisalabad, Pakistan. This Data warehouse stores the large data of teachers of UAF. This data is analyzed for evaluating the teachers performance. This data is used by data warehouse to make prediction about teachers. This warehouse helps not only to administration but also to teachers about their performance and weaknesses.

KEYWORDS: Data warehouse with data mining, Performance evaluation with Data warehouse, Decision making with data warehouse, uses of Data mining.

1 INTRODUCTION

The life of an organization depends on its data. The organization will succeed as much as strong data it has .The progress of an organization can be measured as it gets information from stored data. In decision making this stored data can be used. Trends are determined by using this stored data as history.

University of Agriculture, Faisalabad is not only the famous university of Pakistan but also famous at international level. This university has six faculties: faculty of Agriculture, faculty of veterinary science, faculty of sciences, faculty of animal husbandry, faculty of agricultural engineering and technology and faculty of social sciences. Then each faculty is subdivided into many departments. University has more than 500 teachers and half of them are p.hds. University also hires many visiting staff for improving education standard.

University database stores only bio data of teachers like name father name address qualification. This data cannot be used as history to take important decision and to evaluate teacher performance. University is using Oracle database to store the teachers data. This database is only used for record keeping. Database has many limitations. This database can store data upto Gbs. But university should store data of many years as history. This database stores daily operational data but does not

store long term informational requirement data. University cannot use this database to determine which teacher is performing best and to whom university should hire for next semester. This Database stores simple and short transactions. In database user can view data only as flat file. User can access few tens of records at a time. Indexing and hashing on primary key takes a lot of time. Data in database cannot be analyzed.

In this research a new approach is used to store teachers data in a way that it can be used in decision making or evaluation can be made on this data. For this purpose the concept of data warehousing is used in this research. We proposed a framework of Data warehouse and data mining technique for this university. Data warehouse is used to store large amount of historical data. Data of years can be stored in this database. Decisions can be made on the basis of stored data. Organizations use data warehouse to promote their business. From this stored historical data organization can take important decision.

An insatiable need of the companies in this brutally competitive marketplace is Information. The key to maintain competitive advantage is to understanding the customers's requirements, and the approach in which they want to accept your products or services. Well-known queries and analysis tools can be used to grasp competitive edges, to access crucial data in an appropriate, well-organized manner is significant. Touching and distribution of data all over an organization, between all departments, offices and business followers has equal value [1].

1.1 DATA WAREHOUSE

Data warehouse stores data for years. A relational catalog more willingly than for transaction processing designed for query and investigation is known as data warehouse. Historical data resultant from transaction data is usually stored in this database, but data from additional springs can also be included. Analysis workload is separated from transaction workload and consolidate data from several sources can be gained by organization.

It is analyzed that an immense pact of time spending is concerned in data warehouse system and in its assembling process and daily operation problems often occur, one major concern takes place in the effort of taking out, making over and putting together data in data warehouse from business database. This is because data is received from different types of sources and business applications varieties. A variety of data extraction curriculum supports the functions of purifying the data, processing of the data and usual procedures, and multiple data sources can be handled. In sort to grasp valuable preservation and supervision to data quality Automatic data extraction is demanded through this we let the data warehouse to precisely, securely and constantly mine data from the database and managerial staff can use this for analysis, when it is renovated into the regular data [2].

An extraction, transportation, transformation, and loading (ETL) solutions are included in a data warehouse, adding together to a relational database. And other applications to gather the data and make it available to uses such as an online analytical processing (OLAP) engine, client analysis tools, are provided.

It is presented that in the joining process by reinstating subject-oriented data mutually from data sources data warehouse is measured useful. The job of bringing back data from equipped information system side to data warehouse side throughout data warehousing is carried out by ETL tool. ETL tool performs many key functions including: (1) the classification and withdrawal of data from associated data sources, (2) the enhancement of data on the derivation of business rules, (3) the adaptation of the data into an incorporated format, and (4) the collection of the data to the data warehouse and data mart [3].

This data can be analyzed by the use of different techniques of data mining to predict the future based on the stored historical data. User can perform ad hoc queries on the data stored in data warehouse. Major difference of Data warehouse over Database is that data in data warehouse meets the four characteristics (i) subject oriented (ii) time variant (iii) integrated (iv) non volatile

1.1.1 SUBJECT ORIENTED

To help analyze data in efficient way data warehouses are designed. Data warehouse can be designed to focus on a single subject. For illustration, for discovering further about organization's sales data, a warehouse can be built that concentrates only on sales. Through this warehouse, useful information about sale can be query like "Who was the top customer last year for specific item? Mean a data warehouse by subject issue can be defined, sales in mentioned scenario, the data warehouse is made subject oriented.

1.1.2 INTEGRATED

An intimately associated issue to subject orientation is integration. Data comes in data warehouses from different sources or data marts. In these sources data may be stored in different way or format. Data is taken from these different sources in different formats and put into central data warehouse by resolving different tribulations as identification divergence and contradictions among components of measure. When these conflicts are resolves mean consistency is achieved, that data is said to be integrated.

1.1.3 NONVOLATILE

As it is shown from the word nonvolatile, it means that, once entry of data into the warehouse is made, it is supposed to not be change. Because data warehouse has to store large amount of historical data this is commonsense and the rationale to a warehouse, analysis is allowed to determine what has occurred and why has occur and when has occur.

1.1.4 TIME VARIANT

In industry with the intention of discovering inclinations a bulky amount of data is looked-for. Data warehouse stores data over a long period of time so that important decision can be taken based on this stored data. As match up to **online transaction processing** systems this is very much unusual, historical data demands to be moved in annals for recital necessities. Intended by the term time variant is that change over time is the focus of data warehouse.

1.2 INTRODUCTION TO DATA MINING TECHNIQUES

The technique with reference to dealing out data and categorizing patterns and trends in that information is data mining so that decision or judgment can be made. To detection relevant patterns in a database, to predict future trends analysis is made by defined approaches and algorithms on current and historical data. Data mining reads databases for hidden patterns and to make prediction about future trends and behavior. Proactive, knowledge-driven decisions can be made by organizations by using these tools and organization can answer to the questions that were previously too time sustained to steadfastness.

Data Mining can be described as a technique for taking out the "sense" holded in information to let the understanding needed by a user to formulate a "right" choice. Another definition could be providing the exact information, in the precise structure, at the accurate time, so as to allow the manager to proficiently and successfully perform his duties. Data Mining allows a computer to absorb the stable flow of data produced by the computerized sensors and monitors of the stand, and then extract meaningful information from these contents [4].

1.2.1 DATA

Any facts, truth, numbers, or text that can be practiced by a computer are data. Today, enormous and budding amounts of data are accumulated by organizations in diverse formats and different databases.

1.2.2 INFORMATION

Information is provided by the prototype, associations, or relationships among all this *data*. For example, Information on which products are selling and when can be yielded as a result of analysis of trade point of sale transaction data.

1.2.3 KNOWLEDGE

Knowledge is gained from historical patterns and future trends by converting the information into accurate blueprint. For example, in illumination of promotional labors to supply knowledge of consumer buying manners review information on retail supermarket sales can be examined. Thus, it could be determined by manufacturer or retailer that which items are most vulnerable to promotional efforts.

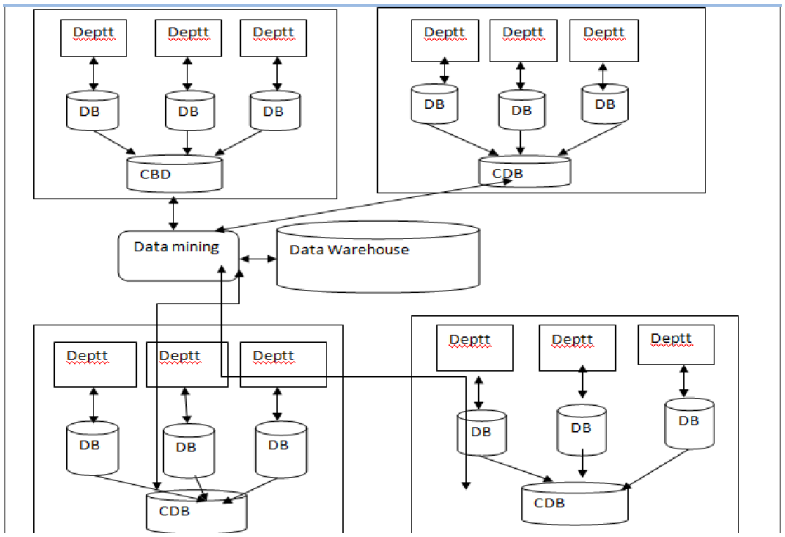
It is stated that the information and knowledge turns into the most precious resources of electric power companies due to the changeable market. The Use of advanced information technology to assimilate presented software and hardware assets, permit the recruits and tools in most excellent operation status, perfectly understanding existing production position

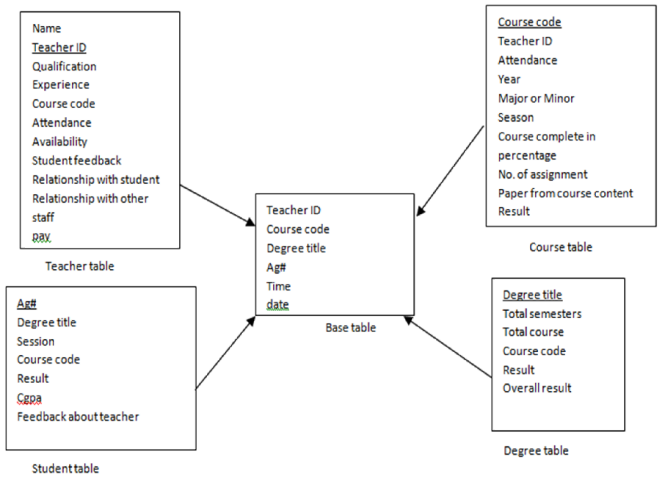
and the predict market stipulate, has become a imperative technique for the electric power enterprise to improve the industry rivalry [5].

It is explained that to take out hidden or masked information from bulky databases a set of robotic procedures are used called data mining. The phrase data mining in large databases withdraws nontrivial of convincing, unstated, potentially serviceable and eventually logical information and the use of the current computing devices is carried out for achieving significant data. From diverse sectors successful appliances in data mining have been accounted in the preceding few decades. Invaluable knowledge that was unknown before can be exposed in the existing database using these mining techniques. For example date profiles of the customers and their procure are kept up to date, a list of available products is maintained with quantities and prices [6].

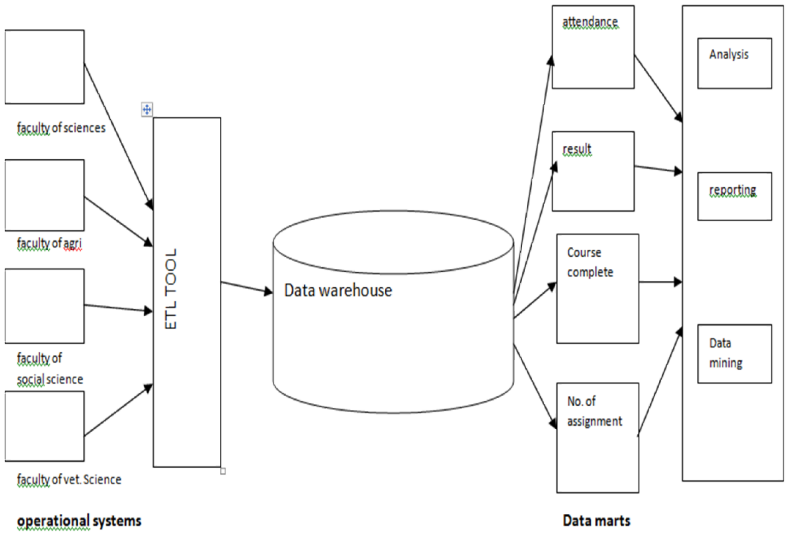
2 PROPOSED FRAMEWORK

We proposed a framework of data warehouse and data mining technique for this university. According to this framework each department stores data of its teachers mean every department has a database. Then the data of these departments is stored in concerned faculty. Mean the data of computer science department, Botany department and chemistry department is stored in the centralized database of faculty of sciences. And the data of anatomy department and pathology department is centralized in the faculty of veterinary sciences database. Then this centralized data of all faculties again centralized at data center. Mean all these faculties then stores data in data warehouse of data center.





Structure of DW tables



3 METHOD

A deep study of data warehouse and data mining technique is conducted. The current system of university of Agriculture Faisalabad is studied and expectations from the new system are analyzed. SPSS is used to evaluate the result. Questionnaire of 20 questions is prepared. A survey is conducted in three main universities of Faisalabad National textile university Faisalabad, GC university Faisalabad and Agriculture university, Faisalabad. Questionnaires are filled by the lecturers of computer departments of these universities. Following attributes are used as parameters to evaluate teacher performance: Attendance, percentage of course, no of assignments, availability besides lectures, result, relationship with students, relationship with other staff, student feedback, no of jobs done.

Using the information on websites, a larger and extra knotty data warehouse is built origination the standpoint of modification on data warehouse. Each site on the Web can be used as data source and every data establishment is diverse reason is that different formats are used to develop these websites and these websites use different types of data, and every site has different information and tissue. From the perspective of multidimensional database, A Web data environment has immense and mixed data warehouse background, based on two-dimensional flat face data cooperatively time dimension and link dimension use to comprise this. The center of attention on conventional data warehouse is data mining which is much more easier than the Web data mining technology. Consequently, the amalgamation of miscellaneous data by the side of with the sites should be premeditated. At these sites only the data are integrated, and an uniform view is provided to users, to accomplish what it is compulsory from the massive data possessions is possible [7].

4 RESULTS AND DISCUSSION

Mean, weighted score and rank is calculated for each question.

Data is consolidated from heterogeneous sources in Data warehouse (DW) of all enterprise to support enterprise wide decision support system (DSS) which consists of dashboard, reporting, and analyzing. At the presentation layer the high level overview of the data warehouse together with the DSS is shown. To support dynamic load allocation based on demand, the DW is implemented in the Cloud Infrastructure. Their respective standalone applications are used to obtain these data sources which are not deployed in the cloud infrastructure. In addition, presentation layer is supported in both client-server platform as well as web-based platform for reporting and analysis. Various selected format such as pie chart and other format in the interactive dashboard are used to present OLAP. A large number of data has been accumulated from heterogeneous data sources from multiple applications, in the healthcare sector in which we are building the data warehouse. Through the ETL process, these data needs to be populated into DW. To support the needs of presenting in both ad-hoc reports and statistical reports we have also implemented the data model and OLAP [8].

From this stat some parameters got high rank these parameters are as follow:

The use of data warehouse and data mining technique can improve teachers performance got 16 rank.

The use of data warehouse with data mining technique is sufficient for performance evaluation got 14 rank.

Data warehouse and data mining technique in this scenario can improve the level of education got 15 rank.

Collectively by these three factor we got 31% success.

Reliability between questions is calculated and .642 reliability is found between questions which shows questions are highly reliable.

Scale: ALL VARIABLES

Case Processing Summary

		N	%
Cases	Valid	20	100.0
	Excluded[a]	0	.0
	Total	20	100.0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
.642	20

And if regression is calculated for these questions good results are obtained which are shown in diagram below.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.922[a]	.850	.593	.54780

REFERENCES

[1] Y.zhu and S. Liu, "Data updating and query in real-time data warehouse system," *international journal on Computer science and software engineering,* Vol. 5, no.4, pp. 1295-1297, 2008.

[2] T. wu, "ETL Function Realization of Data Warehouse System Based on SSIS Platform," *International journal on Database Technology and Applications,* vol.2, no.5, pp. 1-4,2010.

[3] H. jumaily and p. Martinez, "Deriving active mechanisms for relational databases using Model-Driven Architecture," *Journal of Systems and Software*, vol.81, no.12, pp. 2299-2314, 2008.

[4] K. tyagi and B.Sharma, "Data mining tools and techniques to manage the textile quality control data for strategic decision making," *International journal of computer application, vol.13, no.4, pp.0075-8887, 2011.*

[5] S. Farooq, "software testing-goals, principals and limitations, verification and validation," *international journal of computer application*, vol.6, no.9, pp.7-12, 2010.

[6] M. Doller and H. Kosh, "The MPEG-7 multimedia database system," *Journal of systems and software*, vol.81, no.9, pp. 1559-1580, 2008.

[7] S. Ali and H. Jave, " A Technique for handling the SQL Aggregate functions Over Encrypted Data," *Global Journal of computer Science and Technology, vol.*52, no. 6, pp.415-444, 2010.

[8] R. Akerkar, "International multiconference on computer science and information USA." *International journal of computer science and applications*, vol.4, no.1, pp.1-6, 2007.