



## Advanced Data Mining Tool for Analysis of Examination Results from Universities in India

Jyoti Sharma<sup>1</sup>, Sawarkar Sudhir Kumar Deoraaji<sup>2</sup>

<sup>1</sup>PHD Scholar (J.J.T University, India)

E-mail:- [jyotisharma13jan@gmail.com](mailto:jyotisharma13jan@gmail.com)

<sup>2</sup>Principal, Datta Meghe college of Engineering, Mumbai University

E-mail:- [sudhir\\_sawarkar@yahoo.com](mailto:sudhir_sawarkar@yahoo.com)

### Abstract:

*Knowledge Discovery in Databases (KDD) is a decision supporting system based on a data and data mining processing required for making critical decision in education institution. KDD can use data mining tools to discover hidden knowledge, unexpected patterns and new rules from large databases. Motivation of the Journal comes from using KDD technologies that particularly faced by academic administrators and educational departments who have various problems related to university examinations and to find the no. of educated people in India and those are educated which are related to which techniques.*

**Keywords:** KDD, OLAP, Data Warehousing, Data Mining, EMS, EDW.

### INTRODUCTION

With the available examination result data from various Universities in India, it is very difficult to answer some critical questions required for planning by the Central Education Department in India. Some of these typical queries are given below:

- Whether question paper of any particular subject was tough?
- What are the subjects having low passing % for that year? Etc.)
- What is the general pattern of examination result? (Passing % of various examinations for that session can be compared.)
- How the students have performed for the particular examination at this time compare to past?
- Which are the subjects opted by less number of students for a particular examination?
- What is the general pattern of paper setting to examination result in various Indian universities?
- To analyse the trend of average passing % for different universities?
- What is the subject wise/college wise passing percentage for a particular Examination for years from 2000 to 2009?
- How many girl students are getting opportunities for higher education in each part of India?
- For which courses there is a good response from students in each part of India?
- For which courses there is a less response from students in each part of India?

Please cite this Article as : Jyoti Sharma<sup>1</sup>, Sawarkar Sudhir Kumar Deoraaji<sup>2</sup>, Advanced Data Mining Tool for Analysis of Examination Results from Universities in India : Indian Streams Research Journal (July ; 2012)



- To determine in which region girls are able to score better % for given subjects?
- To compare average passing % of boys and girls student for subject wise and region wise?

Since necessity is the mother of all inventions, the motivation for the Journal is derived from the need to design software based on the principles of Data warehousing and Data mining that will help in answering these critical queries raised by Central Education Department of India.

### INSPIRATION

The Journal is inspired by the work of Mr Jagdish Sadhave, who used the technology of Data Mining and Knowledge Discovery in Databases to build Examination Data Warehouse (EDW) [1]. This EDW can effectively be used for mining and answering some critical questions regarding examination results of University of Mumbai. This research work done can be very effectively extended by adding more dimensions to Examination Data Warehouse and making it capable for extracting more useful information regarding all Universities in India. This information can be very useful for Central Education Departments like Ministry of Human Resource and Development (MHRD) or equivalent Ministries at various state levels for decision making.

According to Alejandro Gutierrez and Adriana Marotta [2], a data ware house is a structure that is optimized for distribution. It collects and integrates the sets of historical data from multiple operational systems and feed them to one or more data marts. It may also provide end-user access to support enterprise views of data. Knowledge specialists carefully design data warehouses, where data is extracted from these operational databases and centrally recorded. The key idea is to make available to the management the critical information that can be used for further analytical processing and decision-making.

According to Adviaans and Zantinge [3], Data Warehouse is design for strategic support and is largely built up from databases that make up the operational database. Small local data warehouse are called data marts which are configured to generate reports for specific set of users. The basic structure of data warehouse must be dependent, non-volatile, subject oriented and integrated.

According to Chaudhari and Dayal [4], a separate data warehouse is needed for special data organisation access methods and separate implementation methods are needed to support multidimensional views and various queries.

According to Wiess and Indurkha [5], the data mining step may interact with the user or a knowledge base. The interesting patterns are presented to the user, and may be stored as new knowledge in the knowledge base.

### OBJECTIVE

It is very time consuming and requires lot of human efforts to answer the queries (as raised in above section) as the data is scattered in different databases for every examination for the session (most of the examinations are conducted twice in a year). Further, the answers given manually may not be accurate. Some of the above queries can be solved by adopting new technologies but some queries the additional data is required. Solving these queries and acquiring new knowledge form examination data will help university, state government, UGC, ministry of human resource and development etc. to take correct decisions regards educational policies that will benefit the society.

The number of examinations and number of subjects being examined are increasing over period of time. According to Jagdish Sadhave [1], in 1980, University of Mumbai was conducting 205 examinations while in 2002 it conducted 1092 examinations. Also there is 63.38% increase in total number of student enrolled in all faculties in 2002 as compared to 1980. Over a period of time, the educational environment has also changed. Fundamental changes are influencing the way the university visualizes and plans future activities. Following changes have been observed over a period of time:

- Student behaviour pattern
- Increase in student population region wise and universities
- Increase in number of colleges
- Starting new courses
- Student is under sport quota

Impacts of social situation for e.g. in 7th semester of B.E. (Computer Engineering) examination, no students are opting subjects like system analysis and programming. Instead, almost all the students are opting for the following subjects:



- Database management system,
- Advanced computer techniques,
- Artificial intelligence

Also the passing percentage is increasing year wise. Thus, the student's interest is changing. One has to analyse for these aspects on real data. Based on the analysis, university may have to enforce the changes in syllabus of the examination and take appropriate corrective keeping the academic and financial perspectives.

Like many other organizations, Central Education Department is now beginning to view their accumulated data resources from various Universities as an important asset to take the correct decisions. The new technology of Knowledge Discovery in Databases (KDD) in databases and data mining can be effectively used to improve the effectiveness of data driven decision-making. Keeping these aspects in view the software Examination Management System (EMS) has been developed.

Examination Management System (EMS) has the various advantages like: efficient access, saving, cost, labor, easy to use, Summarized or detailed analysis, visual insights, helping management to take correct decisions, presenting clear status of the various colleges, various universities finding hidden information/trends etc. To sum up, Examination Management System (EMS) is useful for mining the Examination data Warehouse (EDW).

### ANALYSIS

According to Adviaans and Zantinge [3], data warehouse is that portion of an overall Architected Data Environment that serves as the single integrated source of data for processing and generating information. A subset of a data warehouse that supports the requirements of particular department or business function. A data mart focuses on only the requirements of users associated with one department or business function. A data warehouse is designed especially as the Decision Supporting System (DSS). Therefore, only the data that is needed for decision support is extracted from the operational data and stored in warehouse is extracted from the operational data stored in the warehouse.

The data warehouse is a central store of data that has been extracted from the operational data. The data store in a warehouse is subject oriented and of historic bearing. Thus, data warehouses tend to contain extremely large data sets. According to Chaudhari and Dayal [4], Separate implementation methods are needed to support multidimensional views and various queries.

In a world that is becoming truly networked, business, management, society and culture, are undergoing revolutionary changes. We have grown accustomed to rapidly increasing computer-processing power and relatively stable costs. The widespread availability of relatively inexpensive and powerful computing has led to another revolution: the storage of massive amounts of data in electronic form. More and more organizations are moving data for decision support to centralized resources known as a data warehouse.

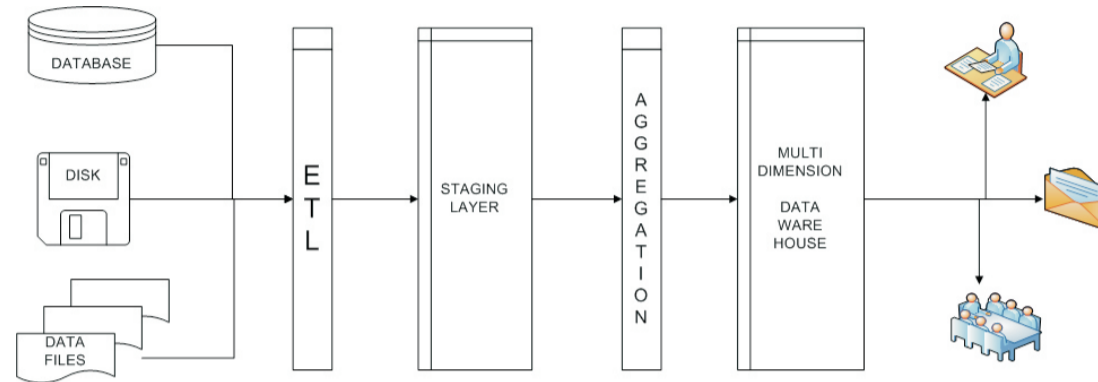
A data structure that is optimized for distribution. It collects and integrated sets of historical data from multiple operational systems and feed them to one or more data marts. It may also provide end-user access to support enterprise views of data. Knowledge specialists carefully design data warehouses, where data is extracted from these operational databases and centrally recorded. the key idea is to make available to the management the critical information that can be used for further analytical processing and decision-making.

Designing of data warehouse requires specialist's knowledge of data design because the data model consists of data needed by users who wants to access it at high speed. Thus, the design for data warehouse can be completely different from that of the operational database (OLTP).

The university Examination system has many operational database files. After creating a central education department data model for the data warehouse, we have to design specific data management environments. If there are a number of databases needed for the data operations, we have to copy this information to the data warehouse. For quick response to user's queries, data warehouse requires high-speed machines and wide variety of optimization processes. This Journal provides Four Tier Architecture of Data Warehouse as detailed below:

- Integration of Data of various examinations.
- Staging layer of data warehouse to contain detailed fact.
- Warehouse on database Server(MS SQL Server 9.0)
- Clients will have driven front end Data mining, Query and Reporting.





The Journal proposes the star schema and Dimensional model to be used for design of Examination Data Warehouse (EDW). EDW consists of mainly following nine tables:

•FACT\_EXAM\_RESULT: this is a fact table containing foreign keys EXAM\_KEY, TIME\_KEY, GENDER\_KEY, INST\_KEY, LOC\_KEY, RESULT\_KEY, SPORT\_KEY obtained by students.

The various dimensional used are given below:

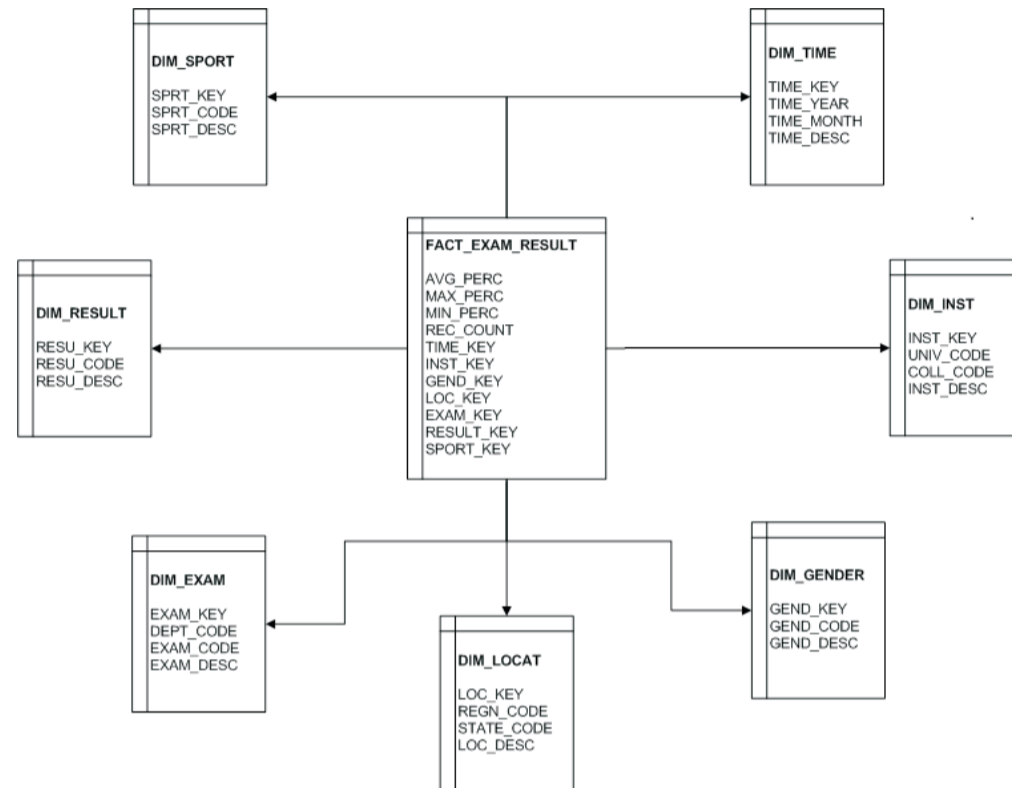
- DIM\_EXAM: contains EXAM\_KEY which is a primary key, DEPT\_CODE, EXAM\_CODE, EXAM\_DESC for example-BSC, BCM, CO8, etc.
- DIM\_TIME: contains TIME\_KEY which is a primary key, TIME\_MONTH (APR, OCT), TIME\_YEAR (1998, 2009) TIME\_DESC.
- DIM\_GENDER: contains GEND\_KEY which is primary key GEND\_CODE, GEND\_DESC.
- DIM\_INST: contains INST\_KEY which is a primary key, UNIV\_CODE, COLL\_CODE, INST\_DESC are two level of this dimension.
- DIM\_LOCAT: contains LOC\_KEY which is a primary key, REGN\_CODE, STATE\_CODE, LOC\_DESC are two levels.
- DIM\_SPORT: contains SPRT\_KEY which is a primary key, SPRT\_CODE, SPRT\_DESC are flags.
- DIM\_RESULT: contains RESU\_KEY which is a primary key, RESU\_CODE, RESC\_DESC.

The following summaries are pre calculated and stored to reduce the response time for frequently asked queries.

- AVG\_PERC: will contains the average of total percentage achieved by students for the given examination, examination time, gender, institution, location, result and sport quota.
- MAX\_PERC: will contains the maximum of total percentage achieved by students for the given examination, examination time, gender, institution, location, result and sport quota.
- MIN\_PERC: will contains the minimum of total percentage achieved by students for the given examination, examination time, gender, institution, location, result and sport quota.
- REC\_COUNT: will contains the result count of total percentage achieved by students for the given examination, examination time, gender, institution, location, result and sport quota.

A typical multi-dimensional model of EDW is shown in following figure.





### KNOWLEDGE DISCOVERIES IN DATABASES (KDD)

As per requirements the KDD process consists of the following steps:

- Cost Benefit Analysis
- Study of process & Requirements
- Data preparation consists of Data selection, Data Pre-processing and Data transformation
- Data Mining
- Analysis of Results and Knowledge Assimilation

The Journal discusses the proportional efforts required at each step of KDD process in our setup. The cost-benefit analysis is essential in the first step to determine the industrial feasibility of the project. The second step helps in choosing the databases that are relevant to satisfy the user requirements.

The third step i.e. Data preparation/transformation in EDW requires up to about 50% of the effort of the entire project. This step is the most resource consuming step in the KDD process that comprises the following three phases: Data Selection, Data pre-processing and Data Transformation. The fourth step of Data Mining requires the efficient software and the last step of Analysis of Results and Knowledge Assimilation tries to find out any interesting, valid and actionable findings.

### CONCLUSION

The Data warehousing and Data mining concepts are the foundation of the proposed EMS system. The appropriate data warehouse and data mining techniques will be used for designing a multi-dimensional EMS.

EMS with proper data modeling of EDW will also ensure an easy maintenance and customization to the system. Thus the proposed research work is aimed to design a multi-dimensional data warehousing environment which can transform the raw data from various Universities in India into Information.



This information can be used for future planning and provisioning of new colleges, Universities and new subjects by Central Education Department of India.

#### REFERENCES

- [1]Jagdish Sadhave, Examination Data Warehouse for Analysing Exam Results for University of Mumbai, PHD Thesis, Department of Computer Science, University of Mumbai (TKS 7139).
- [2]Alejandro Gutierrez, Adriana Marotta, An Overview of Data Warehouse Design Approaches and Techniques, Instituto de Computación, Facultad de Ingeniería, Universidad de la República, Montevideo, Uruguay, October 2000.
- [3]Pieter Adriaans, Dolf Zantinge, Data Mining, Publisher: Addison-Wesley Professional (June 16, 1996), ISBN: 978-0201403800.
- [4]Chaudhuri, S., Dayal, U.: An overview of Data Warehousing and OLAP technology. SIG-MOD Record 26(1), 65-74(1997).
- [5]Wiess, S.M. and Indurkha, N. Predictive data mining, San Francisco: Morgan Kaufmann (1998).
- [6]Christopher Ralph Westphal, Teresa Blaxton, Data mining solutions: methods and tools for solving real-world problems, Edition illustrated, Publisher Wiley, 1998 Original from the University of Michigan Digitized 20 Nov 2007.
- [7]Barquin, R., and H. Edelstein (Eds.). Building, Using and Managing the Warehouse, Prentice Hall (1997).
- [8]Building the Data Warehouse (3rd Edition) by W. H. Inmon (Author) Kimball - R. Kimball. The Data Warehouse Lifecycle Toolkit. J. Wiley & Sons, Inc.

