Prototype Development of a Dimensional Model of Data Warehouse for the Engineering Education Information System in West Bengal, India

Abhijit Pramanik, Somsubhra Gupta

Abstract— **An attempt had been made to develop a prototype dimensional model of data warehouse for the Engineering Education Information System in West Bengal, India using Kimball's methodology. A bus matrix has been established using fact tables and some selected dimension table. Depending on this a prototype of dimensional model based on the engineering education system has been established. Validation and testing also be adopted. This model may provide subject oriented data support for query servicing in order to smooth the progress of educational activities.**

Index Terms— **Data warehouse, higher education information system, dimensional model, bauble chart, Bus Matrix, Fact Table.**

I. INTRODUCTION

The modern education system in India has been developed by the British missionaries and the Indian social reformists [13]. Education in West Bengal is provided by both the government sector as well as the private sector. In West Bengal there are twenty two Universities (State Universities under Higher Education Department – thirteen, State Universities under other Departments – six , Deemed University (RKMVERI) – one, Central Universities – two) sixteen research Institute, more than 751 government and non-government colleges [14]. They play a great role for development of higher education. Engineering education is an important part of higher education in West Bengal. But last few years it has been seen that after completion of admission in engineering discipline so many seats are vacant. Students chose their education discipline depending on their job opportunity. During the recession top-ranked software firms used to abridge their employee strength. Lack of infrastructure and poor placement go against most private engineering colleges [16]. So a systematic approach is needed to improve the understanding, planning and operational work of institutions of engineering education in West Bengal. This education information system basically consists of very large amount of data about students, teachers, staffs, institutions etc. For fulfillment of education stake holder's variety of required information, storing of historical data is needed. Data warehousing should be a solution for this. The data warehouse is the basic part of almost every larger information system, because there is a common need for some sort of systematic services and faster presentation of prearranged data. The data which reside in the data warehouse, it should be filtered, transformed and aggregated before loading. For this reason, data warehouse helps to provide faster data browsing. During the last decade, data warehouse systems have become an essential component of modern decision support systems (Amatory and Murray, 1997). The data warehouse concept covers different aspects like architecture, physical optimization, tool support and modeling approaches. A data warehouse initiative's probability of success is greatly increased by a sound understanding of the stakeholders and their requirements. The requirements definition absolutely drives the data design for the data warehouse. Blueprint of data consists of putting in concert the data structures. a group of data elements from a data structure. The logical data design consists of grit of the various data rudiments that are required and permutation of the data elements into structures of data. The logical data design also includes establishing the relationships among the data structures. The results of the requirements gathering phase are documented in detail in the requirements definition document. This document represents the set of information package diagrams which are an essential component to form the basis for the logical data design for a data warehouse. The result of a data design process is a dimensional data model.

In this paper we have presented a brief literature review in section 2. Structure of West Bengal Engineering education system has been discussed with a block diagram (fig1) in section 3. Objective of this paper has been presented in section 4. For our research we have followed R. Kimball's methodology. There are actually four steps in this methodology. First step in this methodology is system process identification. For system identification gathering of system requirement and data profiling is needed. To gather the system requirement we have done a survey among student and guardian. The survey result has been shown in section6. Finally we have established a dimensional model of Engineering Education of West Bengal. Its limitations and future works have been discussed in section 8 and testing mechanism has been discussed in section7.

II. LITERATURE REVIEW

Higher education is learning that is provided by different organizations like universities, vocational universities, degree colleges, arts colleges, technical and medical colleges, and other institutions that reward academic degrees. Michael Boehnlein , Markus Plaha , Achim Ulbrich-vom Ende have described a successful cooperation between a company and a data warehouse research project for higher education at the University of Bamberg. Micro Strategy, as one of the leading vendors of relational online analytical processing (ROLAP)

Mr. Abhijit Pramanik,, Department of Computer Science and Engineering, Murshidabad College of Engineering and Technology, West Bengal, India.

Dr. Somsubhra Gupta, Department of Information Technology, JIS College of Engineering, Kalyani, Nadia, West Bengal, India.

solutions, is the industrial partner [3]. During the last decade, data warehouse systems have become an essential component of modern decision support systems (Anahory and Murray, 1997) [10]. The data warehouse concept covers different aspects like architecture, physical optimization, tool support and modeling approaches. André Flory, Pierre Soupirot, and Anne Tchounikine in their paper titled "A Design and implementation of a data warehouse for research administration universities" have presented an architecture of a data warehouse for research center activity [2] Nenad Milanovic, Goran Soskic, Ana Petkovic have been presents a strategy and Development of Warehouse model for Students' Nourishment Information System in Croatia [1] in their paper titled "Data Warehouse Design for Croatian Students' Nourishment Information System" Mirta Baranovid. Mirjana Madunid, Igor MekteroviC have presented data warehouse for the Higher Education information System in Croatia [4]. Veronika Stefanov have emphasized and described the relationship between the DWH and the organization with conceptual models, and to use this knowledge to support data interpretation with business metadata [5] Wilburt Juan Labio, Dallan Quass, Brad Adelberg in their paper (Physical Database Design for Data Warehouses) have studied how to select the sets of supporting the views and of indexes to materialize to minimize the down time. We call this the view index selection (VIS) problem. They present an **A*** search based solution to the problem as well as rules of thumb. They also perform additional experiments to understand the space-time tradeoff as it applies to data warehouses [6]. Song Qiang , Lingxia Liu through their paper introduces a network which is suitable for network companies and the design of a data warehouse system supporting data mining [7]**.** Joseph M. Firestone has established that that (1) any E-R model can be represented as an equivalent set of DM/star schema models; and (2) the question of whether an E-R structured data warehouse, absent associative entities, i.e. fact tables, is a viable concept given recent developments in data warehousing. [1]

III. Engineering Education System in West Bengal

Education in West Bengal is provided by mainly the government as well as the private organization. [13]. West Bengal has many reputed institutes of higher education studies. [14].All the institute follows the following hierarchal architecture of education system (fig1). It is clear from the fig1 that after passing the 10th standard students can enter in engineering discipline or they can continue study for class12. After passing class12 they can enter either in engineering discipline or in general discipline and how they can switch also presented in the following figure.

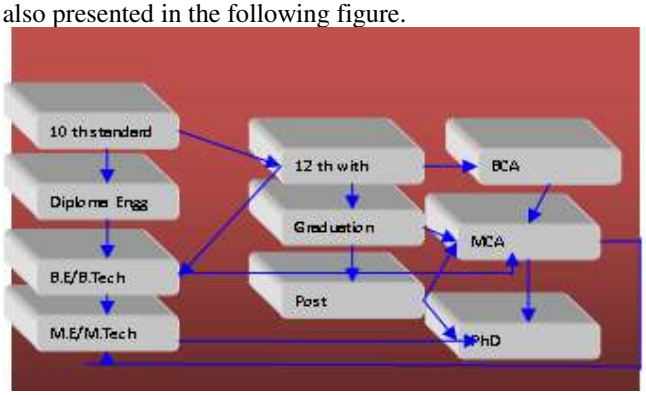

Figure1

IV. Objective

Objective of this paper is given below:

1) Understanding the basic information required by the general users (student, Guardian, administrator of different educational organization) of engineering education system of West Bengal.
2) To provide a maximum amount of ad-hoc query, logically structure the West Bengal's engineering education systems different dimensions (selected dimension only).
3) Taking initiative to create a part of prototype dimensional model of a data warehouse based on the higher education information system.

V. Methodological Aspects

Most of them are technology driven approach. To implement a prototype dimensional model for a data warehouse of West Bengal Higher education the R. Kimball's methodology [8] has been adopted which is proceed in opposite direction of technology driven approaches. According to this methodology the main four steps are:

Step1 –system process identification:
At first high level system requirements should be gathered. Then the system process should be determined and measurable event to be modeled

Step2 – Grain declaration:
After a system process chosen, on the way of emergent of a data warehouse's dimensional model, grain should be declared or the level of details in the fact table for the selected process should be measured.

Step3 – Dimensions identification:
After the establishment of a grain of the fact table dimensions are chosen effortlessly. In the third step the dimensions are settled on which are relevant to the fact table at the stated level of granularity.

Step4- Facts identification:
After identification of dimension the final steps come up in the modeling process to identify the facts or measures from the system process.

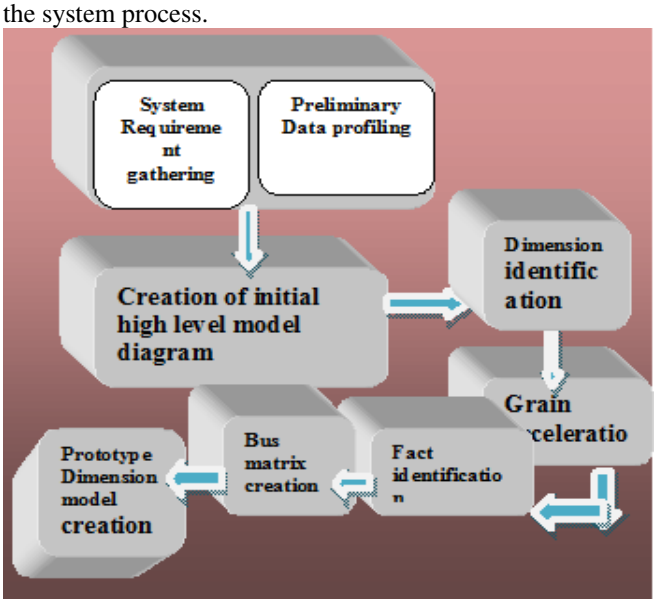

Figure 2

VI. Results and Findings

A. System process identification

For gathering system requirements and preliminary data profiling we have done a questionnaire survey which has been conducted among 1000 participant at the time of JEE counseling at nodal center and helpdesk at Murshidabad district's engineering colleges.

Participant may be either the candidate and/or their respective guardian. The topic presented at the time of survey and the number of participant raised they are interested about the respective topic(s) are listed in table below. The selection of the topic has been framed by the authors experience at previous counseling and different prescribed norms of the affiliating organization like AICTE or the university (in this case WBUT). Depending on the categories of question the set of question has been divided in the following sub set of questions and Response among 1000 student and their respective guardians have been collected as follows:

Sl. no	Topic covered	Number of student /guardian response
1	syllabus	361
2	Admission processes	954
3	Evaluation mechanism	256
4	Time takes university for the declaration of results (Days)	432
5	Student centric academic activities (Technical Fest, workshops, competitions, industrial training, etc.)?	421
6	placement activities	792
7	Scholarship	532

Table 1

The set of question in the table1 are related affiliating university. Maximum number of student and guardian was response for admission procedure.

Sl no.	Topic related to institution	Number of student / guardian response
1	Infrastructural facilities	978
1.1	Classroom	786
1.2	Laboratory equipment	564
1.3	Workshop equipment	562
1.5	Internet	789
1.6	Sports	324
1.7	Medical	190
1.8	Canteen	785
2	Principal/Director leadership qualities	246
3	Total number of faculty	807
4	Total number of staff	502
5	Opportunities provided by the institutions for the overall development of the students	876
6	Level of students seeking admission at the institute	654
7	Organization and opportunity for participation of students in Tech. Feast/ Seminars/ workshop	256
8	Facilities like computer centers, labs, workshops, etc. are opened beyond working hours	125

Table 2

The set of question in the table2 are related to institution. Maximum number of student and guardian was response for Infrastructural facilities.

Sl no.	Topic covered	Number of student / guardian response
1	Number of companies coming for compassing	987
2	Types of company	956
3	Training	765
4	The package offered	987
5	Accreditation of company	521
6	Place of posting	452
7	Type of job	978

Table 3

Table3 represents question about job assistance provided by the institution. Maximum number of question asked by student and guardian was about "Number of companies coming for compassing "

Sl no.	Topic covered	Number of student / guardian response
1	Text Books	865
2	Reference books	678
3	Journals	532
4	Seating Capacity	765
5	Internet Facility	674
6	Member of online journals	212

Table 4

Table4 represents question about Library assistance provided by the institution. Maximum number of response by student and guardian was about "Text Books ".

Depending on the required information about student and guardian the following process has been chosen to design the first set of logical structures:

- Process for Information about syllabus
- Process for Information about admission procedure
- Process for information about scholarship
- Process for information about affiliating university
- Process for information about students
- Process for information about classroom facility
- Process for information about laboratory
- Process for information about library
- Process for information about hostel facility
- Process for information about internet facility
- Process for information about sports facility
- Process for information about medical facility
- Process for information about knowledge development program
- Process for information about result
- Process for information about training
- Process for information about placement

- Process for information about contact person
- Process for information about fee structure
- Process for information about project work
- Process for information about management.
- Process for information about evaluation mechanism
- Process for information about location
- Process for information about working hours
- Process for information about faculties
- Process for information about staff

B. Grain Defining

To declare the grain, the following high level dimensional model has been created:

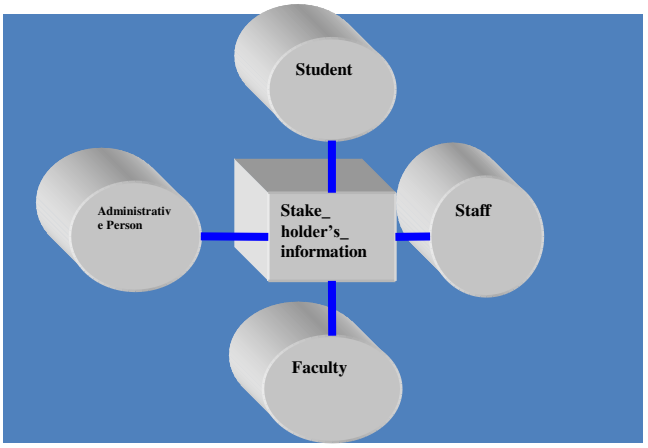

Figure 3

Figure3 represents a bauble chart of four dimensions student, staff, faculty, administrative person with the fact table stake_holder's_information. Here grain of the fact table is one row for each truncation.

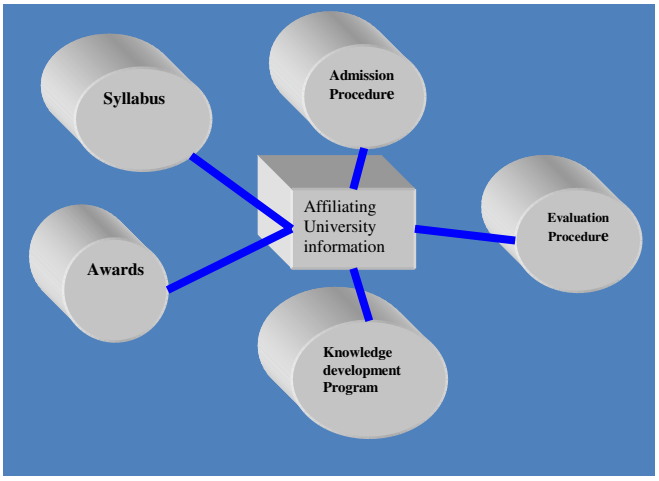

Figure4

Figure4 represents a bauble chart of five dimensions syllabus, admission_procedure, evaluation_procedure, Knoledge_ development_ program, Awards with the fact table affiliating_university_information. Here grain of the fact table is one row for each truncation.

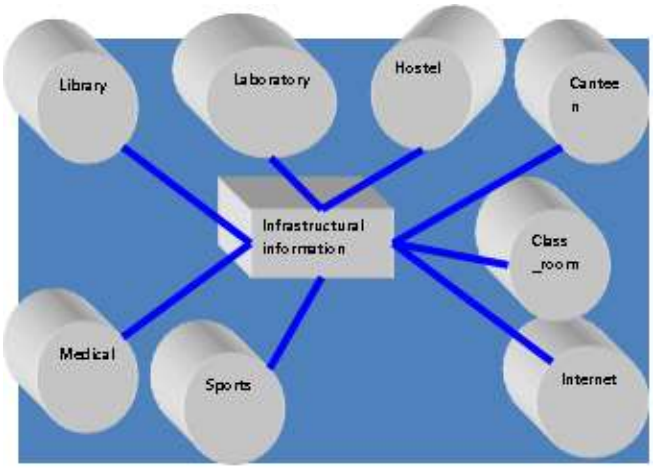

Figure 5

Figure 5 represents a bauble chart of ten dimensions library, laboratory, hostel, canteen, classroom , internet, electricity, sports, medical, and water_supply with the fact table infrastructural information. . Here grain of the fact table is one row for each truncation.

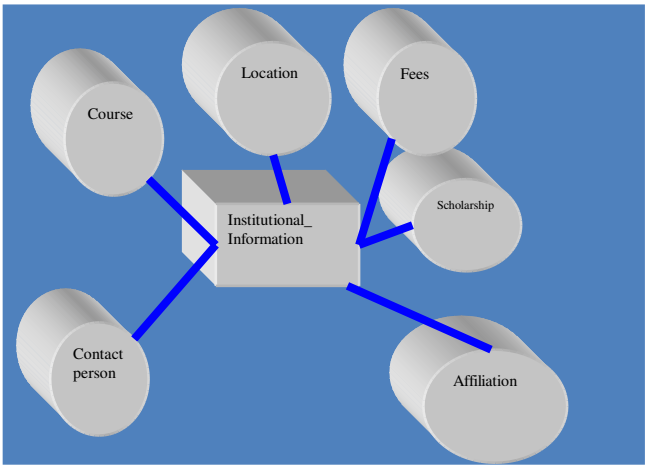

Figure 6

Figure 6 represents a bauble chart of dimensions course, location, fees, scholarship, contact_person with the fact table institutional_information. . Here grain of the fact table = multiple row for each traction.

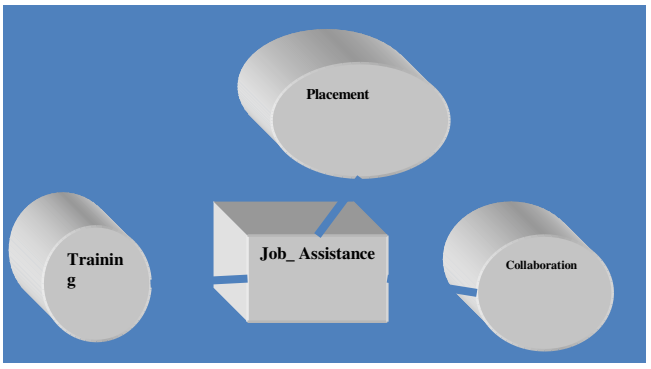

Figure 7

Figure7 represents a bauble chart of dimensions Training, placement and collaboration with the fact table job assistance. . Here grain of the fact table is multiple rows for each truncation.

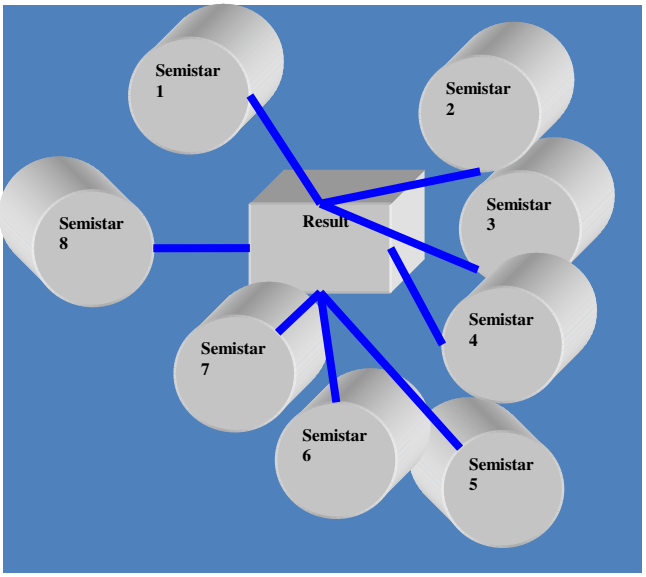

Figure 8

Figure 8 represents a bauble chart of fact table result with the different semester dimension. . Here grain of the fact table = one row for each traction.

C. Identify the Dimension

After creation of high level model diagram authors have defined dimension and its attributes/Metrics list which is given in the table6. Here authors have considered only few dimensions depending on the basic information required by the stake holders which are given below:

1. System stakeholder
2. Institution
3. Affiliating university
4. Job assistance
5. Infrastructure
6. Result

D. Fact identification

To identify the facts for required process authors have developed the detailed bus matrix which is given in the table6. Depending on detailed bus matrix we have set attributes of the fact table which are given below by table 5:

Fact table	Attributes
System stake holder	Staffinfo, Faculty_Id,U_roll,Aid
Institution	Coursename, campus, feecourse, schinfo, affinfo, coninfo
Affiliating university	Awinfo, Scode, Acourse, knoinfo, Ecourse
Job assistance	Palyear, trayear, collid
Infrastructure	Ldip, nameoflab, name of hostel, winfo, minfo, einfo, iinfo, sinfo, cldepname, caninfo
Result	Sem1,sem2,sem3,sem4,sem5,sem6,sem7,sem8 uroll

Table 5

Depending on the fact table and bus matrix we have designed the following star schema:

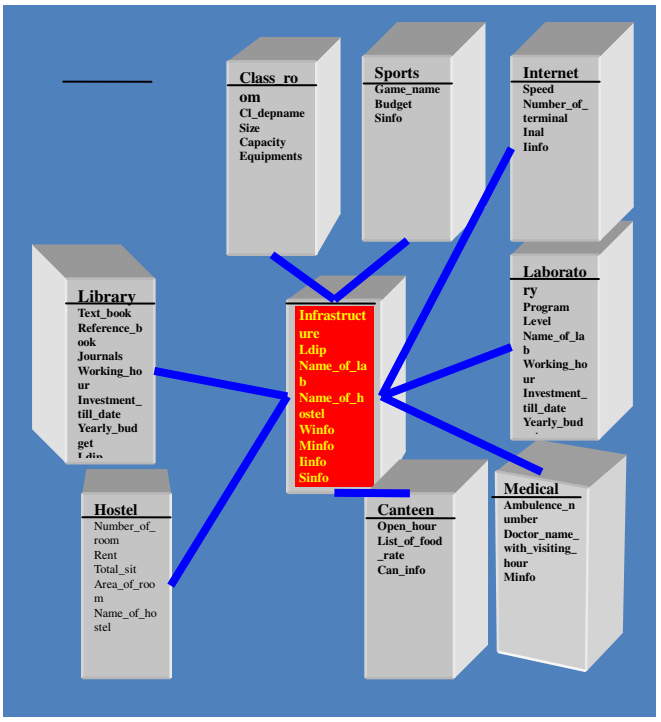

Figure 9

It is clear from the above star schema (figure8) that the fact table infrascture connected with the dimension table Electricity, Class_room, Sports, Internet, Laboratory, Medical, Canteen, Water_supply, Hostel, and Library.

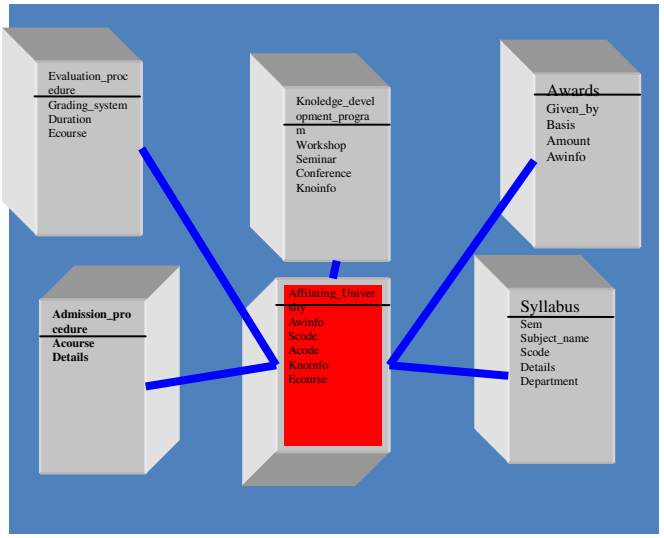

Figure 10

Figure 9 represents star schema for the fact table Affilating_university which connected with the dimension tables Evaluation_procedure, Admission_procedure, Knoledge_development_program, Awards, and Syllabus.

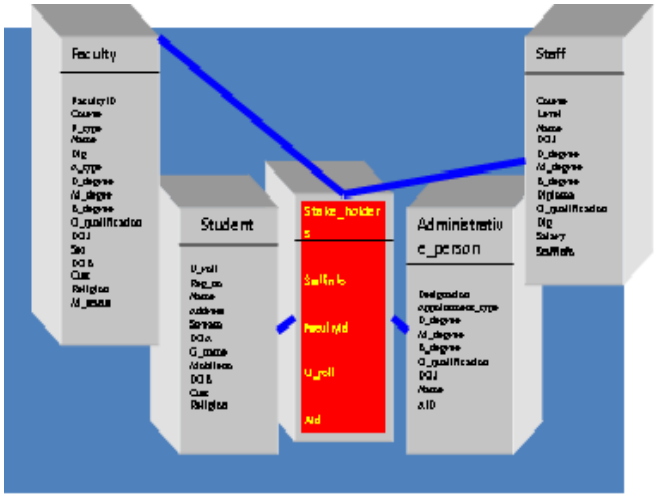

Figure 11

It is clear from the above star schema (figure10) that the fact table Stake_holders connected with the dimension table's student, Administrative_person, Faculty, Staff.

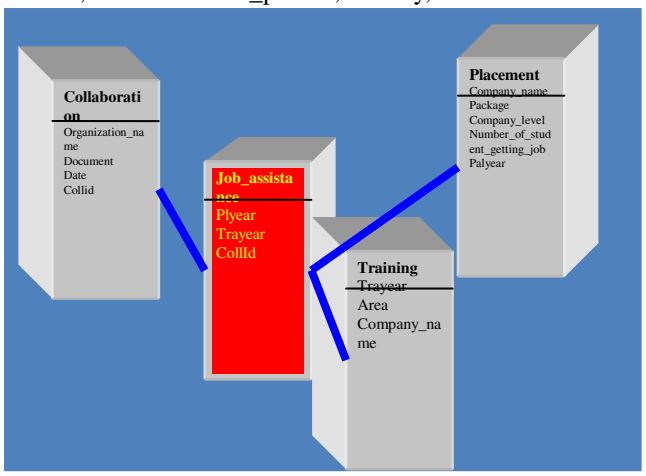

Figure 12

Figure11 represents star schema for the fact table Job_assistance which is connected with collaboration, Placement and Training.

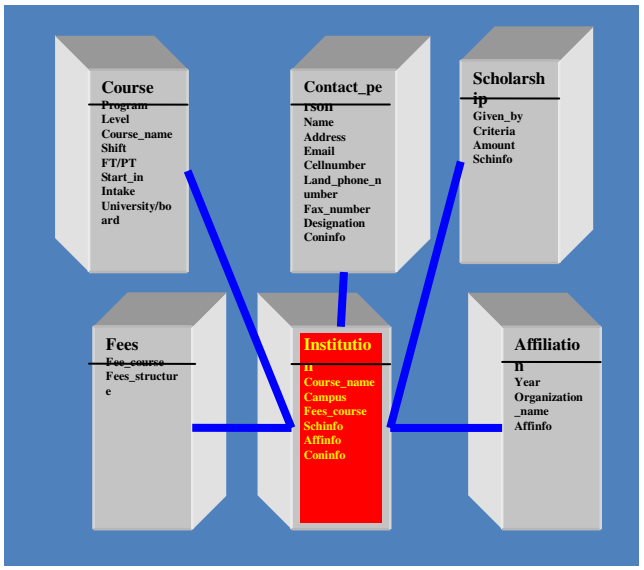

Figure 13

It is clear from the above star schema (figure12) that the fact table Institution connected with the dimension table Fees, Course, Contact_persion, Affiliation, and Scholarship.

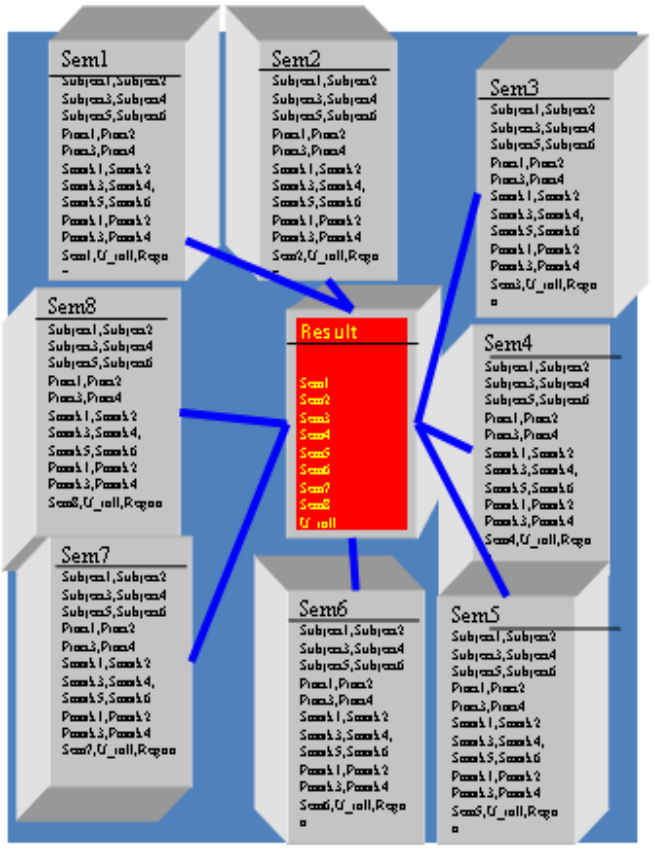

Figure 14

Figure13 represents star schema for the fact table Result which is connected with the different dimension tables (Sem1, Sem2, Sem8).

After integration the above star schema the following part of the prototype dimensional model of the data warehouse for Engineering Education Information System has been created:

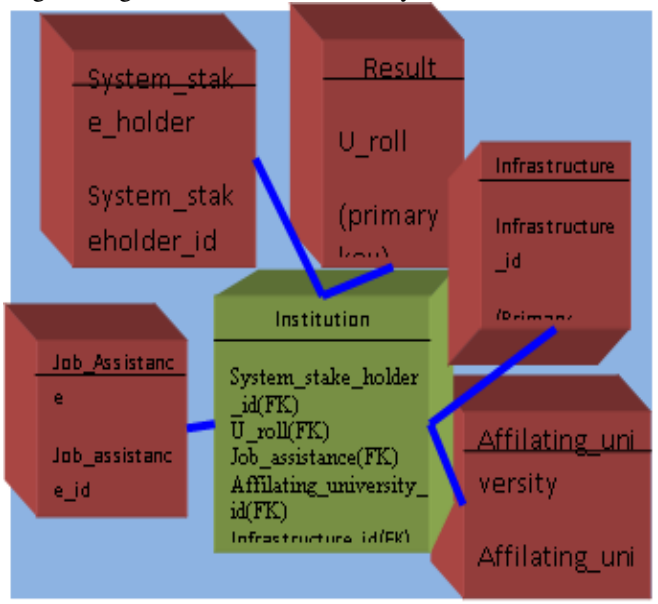

Figure 15

VII. TESTING

After creation of dimensional model authors has involved in reviewing the model and validation segment. To review the model authors were trying to get a response from the stakeholder of the education system. To check the validation we have mate three types of people:

- Source system developers and DBAs (Data Base Administrator) who marked errors in the model and authors have removed some of this.
- Some administrative person and industrial person who were not directly involved in the model development phase.
- Finally, the general user community (student, guardian, teacher etc.).

VIII. CONCLUSION

To partially fulfill the design of dimensional model of engineering education system of West Bengal we had conducted a questioner survey among student and their respective guardian. Depending on the survey result some process has been selected. To declare the grain six bubble chart (high level dimension model) has been drown. From these six dimensions (Stake holder, Institution, Affiliating University, Job assistance, infrastructure and result) has been selected. To identify the facts for required process a detailed bus matrix and fact table has been established and depending on these bus matrix six star schemas has been established. Integrating these star schemas prototype dimensional model has been established. That is the end result is a partially developed prototype dimensional model of engineering Education information System in West Bengal. It has been tested against the education stake holders need for refining the model. This prototype model may helps developer to take an initiative for designing of data warehouse based on engineering Education information System in West Bengal.

IX. LIMITATION

The limitations of the study are shown in the first stage of the methodology that is requirements gathering. As authors meet only few numbers of students and guardians and consider only their required information, so it is not possible to give 100% fulfillment of designing of the dimensional model of engineering education system. Here the authors have considered only a few dimensions and few types of system processes, but the stakeholder of education system required more different types of information. To review and check the validation authors have spend few times and have collected a less number of feedbacks. Only some dimensions have been considered at the time of creation of invented prototype dimensional model. In feature more information about education system should be analyzed for creation of strong dimensional model based on Engineering Education information System and depending on this model full data warehouse should be implemented.

ACKNOWLEDGMENT

The authors acknowledge their respective Institution for kind help and support for the study and research.

REFERENCES

[1] Nenad Milanovic, Goran Soskic, Ana Petkovic , Data Warehouse Design for Croatian Students' Nourishment Information System , June 22-25, 2009, Proceedings of the ITI 2009 31st Int. Conf. On Information Technology Interfaces.
[2] André Flory, Pierre Soupirot, and Anne Tchounikine , A Design and implementation of a data warehouse for research administration universities .
[3] Boehnlein M., Plaha M., Ulbrich-vom Ende A. Case Study - Building A Data Warehouse For Higher Education In The Course Of Microstrategy's University Program,
[4] Mirta Baranovid. Mirjana Madunid, Igor MekteroviC, Data Warehouse as a Part of the Higher Education Information System in Croatia , June 16-1 9, 2003, 25th Int. Conf. Lnformafion Technology lnterfaces /TI 2003, , Cavtat, Croatia
[5] Veronika Stefanov, Bridging the Gap between Data Warehouses and Organizations, 2002, This research has been funded by the Austrian Federal Ministry for Education, Science, and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.
[6] Wilburt Juan Labio, Dallan Quass, Brad Adelberg , Physical Database Design for Data Warehouses , 1997,1063-6382/97 $10.00 0 1997 IEEE
[7] Song Qiang , Lingxia Liu, Research in the Design of a New Data Warehouse System, 2009, 978-1-4244-4520-2/09/$25.00 ©2009 IEEE
[8] R. Kimball: The Data Warehouse toolkit. Wiley Computer Publishing. 1996
[9] J. Widom: Research Problems in Data Warehousing. Proceedings of the 1995 International Conference on Information and Knowledge
[10] ANAHORY, S. And MURRAY, D. (1997). Data Warehousing in the Real World. Addison-Wesley, Harlow.
[11] C.S.R. Prabhu, Data Warehousing, second Edition, Prentice-Hall of India
[12] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy, B. Becker. The Data Warehouse Lifecycle Toolkit, Second edition.
[13] http://en.wikipedia.org/wiki/Education_in_West_Benga
[14] Annual report 2010-2011, DEPARTMENT OF HIGHER EDUCATION, West Bengal
[15] Joseph M. Firestone, Dimensional Modeling and E-R Modeling In The Data Warehouse. file: ///E|/FrontPage webs/Content /EISWEB/DMERDW.html
[16] Articals at Business Standard Feb03,2013.
[17] Dr. B.K.Sthapak, "Globalisation Of Indian Engineering Education Through The Washington Accord ".
[18] Daniel L. Moody, Mark A.R. Kortink, " From Enterprise Models to Dimensional Models: A Methodology for Data Warehouse and Data Mart Design ",http://sunsite.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-28/
[19] Mickey Yost and Jack Nealon, " Using A Dimensional Data Warehouse To Standardize Survey And Census Metadata ".
[20] Ronald Gage Allan, David R. May, Data Models For A Registrar's Data Mart, This paper was presented at CUMREC 2000, the Higher Education Administrative Technology conference.