ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

Algometric Solution for Creation of Data Warehouse

Shahid Bashir Dar[1], Ashish Sharma[2]

M. Tech Student, Dept. of CSE, BIMT, Mehli, Shimla, H.P, India[1]

Assistant Professor, Dept. of CSE, BIMT, Mehli, Shimla, H.P, India[2]

ABSTRACT: As on date when data warehouse is necessitated as an information repository, design of data warehouse is not standardized and development of warehouse is primarily treated as Extract, Transform & Load (ETL) project, just like a typical software development project. Even after years of successful development in data warehousing across the globe, there is still no standardisation on the design of data warehouse, neither there is generic design technique which can be applied for the development of warehouse irrespective of type of data. In this research paper we propose a programmatic solution for creation of data warehouse. The proposed solution creates dimension tables based on the physical structure of operations sources (OLTP) and upon successful creation of dimensions, fact table is created and accordingly populated. The proposed algorithm also automates the process of creation of a data warehouse.

KEYWORDS*:* data marts, dimension tables, fact tables.

I. INTRODUCTION

Data Warehouse is an asset for every enterprise, and it is being used within and outside the organization in order to facilitate flow of information. Data warehouse is not individual specific but is designed to entertain enterprise user with prerequisite information. Organizations across the globe are creating such information repositories for easy information access and delivery. However, even as on date when data warehouse is necessitated as information repository, design of data warehouse is not standardized and development of warehouse is primarily treated as Extract, Transform & Load (ETL) project, just like typical software development project. Even after years of successful development of data warehouse across the globe, there is still no standardisation on design of data warehouse, neither there is generic design technique which can be applied for the development of warehouse irrespective of type of data.

Majid and Muheet [1] in their paper titled, "Warehouse Creator: A Generic Enterprise Solution", proposed methodology for creation of warehouse making use of fact and dimension tables however they have not given physical architecture for creation of data warehouse In this research paper we propose programmatic solution for creation of data warehouse, the proposed solution creates dimension tables based on the physical structure of operations sources(OLTP) and upon successful creation of dimensions, fact table is created and accordingly populated.

A. Data Warehouse Design:

Online Transaction Processing systems (OLTP) can't be used for Online Analytical Processing (OLAP) neither they are used for predicting/mining as they are meant for day-to-day business operations. However data from these OLTP flows into the warehouse where it is used for Analysis Prediction and strategic decision making. Data Warehouse is primarily composed of fact table and dimension tables (referred as dimensions), these dimension tables describe the option to cut or view the data in the fact table. Number of dimension tables vary from few to numerous depending upon complexity of data resulting in data warehouse and there are no given obligations to follows as of how many dimensions should be there to have credible warehouse.

Primarily data from operational sources is to be extracted, transformed and loaded into target warehouse, however before ETL can be performed first and foremost the structure of target data warehouse is to be created. The biggest complexity lies in designing the structure of data warehouse i.e,

- Number of dimension table/s

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

- Structure of each dimension table/s
- Granularity of data
- Structure of fact table

An enterprise has numerous operation sources which vary not only in structure but also have different DBMS like Oracle, MySql, SQL server etc. thus are heterogenic in nature. Online Transaction Processing Systems (OLTP) are meant for business operations and now a day's enterprise has numerous operational sources which are geographically distant and structurally complex.

II. RELATED WORK

Bill Inmon primarily gave the idea of warehousing in 1992, however, the roots of warehousing can be traced back to 1960 when G. Mills and D. College developed in their project the concept of dimensions and facts. It was not till 2000 when people realized the essence of warehousing and emergence of its implementation. Diego Calvanese et.al [9] in 2001 presented a novel approach to data integration in a data warehouse. Their approach is based on a conceptual representation of the Data Warehouse application domain, and follows the so-called local-as-view paradigm. Watson [10] in 2002 discusses the recent developments in data warehousing. Santos et.al [11] in 2009, demonstrated that data warehouses must be able to enable continuous data integration, in order to deal with the most recent business data. Cuzzocrea et.al [12] in 2013 explores the convergence of Data Warehousing, OLAP and data-intensive Cloud Infrastructures in the context of analytics over Big Data.

Majid and Muheet [1] in their research starting from 2006 to 2014 concluded that there should be a generic warehouse tool to create warehouse of any enterprise making use of fact and dimension tables.

III. CASE STUDY

University has numerous processes going on simultaneously like Accounts, HRM, Examination, Library etc, however in this case study we will consider only one such process; University Examination System.

University Examination System is itself composed of numerous sub processes like

i. Registration-records: Student registration details at the time of admission. Every student admitted to university has to go through the process of registration.
ii. Enrollment: Every registered student is enrolled for a specific course, student may be enrolled with more than one course but will be registered only once.
iii. Transit- records: Student marks he/she has appeared for each subject/paper.
iv. Result: For every enrolled course result will be declared

All the data present in operational sources do not make it into warehouse, but selective data become part of data warehouse. Warehouse designers carefully choose source (database/table) from OLTP which will be extracted and become part of warehouse. The table discussed above are graphically represented below (structure & data), the sources which will be part of warehouse.

These sources are heterogeneous and are spread across organization (geographically distant) and are accessed via remote host using IP address. These sources only have read access for the creation for warehouse and no data modification/updating can be performed on these data sets.

Data set in consideration here will result in data mart for the university examination system. Data Marts are individual components and refer to specific process of the system like sales, finance, examination etc. Thus data mart is subset of a warehouse and these data marts can be integrated to create Data warehouse or central data warehouse. The purpose of data mart is for the specific process of the system while as central data warehouse is massive in size and caters entire enterprise.

Graphical presentation of these sources (structure & data) is shown below

i. Database Registration : Table Master

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

```
MariaDB [registration]> desc Master;
+-------+-------------+------+-----+---------+-------+
| Field | Type        | Null | Key | Default | Extra |
+-------+-------------+------+-----+---------+-------+
| regno | varchar(20) | NO   | PRI |         |       |
| name  | varchar(35) | YES  |     | NULL    |       |
| fname | varchar(35) | YES  |     | NULL    |       |
+-------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)

MariaDB [registration]>
```

Figure 1: Structure of Master Table

```
Database changed
MariaDB [registration]> select * from Master;
+----------------+--------------------------------+-----------------------------
-+
| regno          | name                           | fname
 |
+----------------+--------------------------------+-----------------------------
-+
| 10034-PC-2006  | SAZIYAH MAHBOOBA               | MOHMAD SYED SHAH
 |
| 11632-AW-2004  | NARGIS JAN                     | MOHD AKRAM GANIE
 |
| 12497-PC-2007  | IRSHAD NAZIR SHAH              | NAZIR AHMAD SHAH
 |
| 12499-PC-2007  | ZAHOOR AHMAD TEELI             | ABDUL REHMAN TEELI
 |
| 12500-PC-2007  | MOHD AMIN TEELI                | ABDUL GANI TEELI
 |
| 12501-PC-2007  | OYAIS JAN                      | NAZIR AHMAD KAW
 |
| 12503-PC-2007  | AB WAHID MIR                   | AB GANI MIR
 |
| 12508-PC-2007  | SHEERAZ AHMAD MIR              | AB GANI MIR
 |
| 12509-PC-2007  | AJAZ AHMAD BHAT                | AB KHALIQ BHAT
 |
| 12513-PC-2007  | GAZENFAR GULL                  | GH MOHD DAR
 |
| 12514-PC-2007  | YOUNOUS GANI WANI              | ABDUL GANI WANI
 |
| 12515-PC-2007  | JAVEED AHMAD DAR               | ASSADULLAH DAR
 |
| 12523-PC-2007  | MUDASIR MAQBOOL                | MOHAMMAD MAQBOOL SOFI
 |
| 12525-PC-2007  | ROUF FAYAZ                     | FAYAZ AHMAD TEELI
 |
```

Figure 2: Data of Master Table

ii. Database enrollment : Table Enrollment_12

```
MariaDB [enrollment]> desc Enrollment_12;
+--------+-------------+------+-----+---------+-------+
| Field  | Type        | Null | Key | Default | Extra |
+--------+-------------+------+-----+---------+-------+
| regno  | varchar(15) | NO   | PRI |         |       |
| rollno | int(9)      | NO   | PRI | 0       |       |
| sub1   | varchar(4)  | YES  |     | NULL    |       |
| sub2   | varchar(4)  | YES  |     | NULL    |       |
| sub3   | varchar(4)  | YES  |     | NULL    |       |
| sub4   | varchar(4)  | YES  |     | NULL    |       |
| year   | varchar(4)  | YES  |     | NULL    |       |
+--------+-------------+------+-----+---------+-------+
7 rows in set (0.00 sec)

MariaDB [enrollment]>
```

Figure 3: Structure of Enrollment_12 Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

```
MariaDB [enrollment]> select * from Enrollment_12;
+-----------------+--------+------+------+------+------+------+
| regno           | rollno | sub1 | sub2 | sub3 | sub4 | year |
+-----------------+--------+------+------+------+------+------+
| 10034-PC-2006   |   2676 | GE   | CH   | BO   | ZO   | UG3  |
| 11632-AW-2004   |   2540 | GE   | ZO   | BO   | CH   | UG3  |
| 12497-PC-2007   |   2617 | GE   | CH   | BO   | ZO   | UG3  |
| 12499-PC-2007   |   2680 | GE   | CH   | BO   | ZO   | UG3  |
| 12500-PC-2007   |   2681 | GE   | CH   | BO   | ZO   | UG3  |
| 12501-PC-2007   |   2704 | GE   | CH   | BO   | ZO   | UG3  |
| 12503-PC-2007   |   2641 | GE   | CH   | BO   | ZO   | UG3  |
| 12508-PC-2007   |   2701 | GE   | CH   | BO   | ZO   | UG3  |
| 12509-PC-2007   |   2595 | GE   | CH   | BO   | ZO   | UG3  |
| 12513-PC-2007   |   2685 | GE   | CH   | BO   | ZO   | UG3  |
| 12514-PC-2007   |   2683 | GE   | CH   | BO   | ZO   | UG3  |
| 12515-PC-2007   |   2618 | GE   | CH   | BO   | ZO   | UG3  |
| 12523-PC-2007   |   2657 | GE   | CH   | BO   | ZO   | UG3  |
| 12525-PC-2007   |   2711 | GE   | CH   | BO   | ZO   | UG3  |
| 12526-PC-2007   |   2649 | GE   | CH   | BO   | ZO   | UG3  |
| 12527-PC-2007   |   2628 | GE   | CH   | BO   | ZO   | UG3  |
| 12528-PC-2007   |   2646 | GE   | CH   | BO   | ZO   | UG3  |
| 12529-PC-2007   |   2643 | GE   | CH   | BO   | ZO   | UG3  |
| 12534-PC-2007   |   2596 | GE   | CH   | BO   | ZO   | UG3  |
| 12548-PC-2007   |   2710 | GE   | CH   | BO   | ZO   | UG3  |
| 12571-PC-2007   |   2607 | GE   | CH   | BO   | ZO   | UG3  |
| 12572-PC-2007   |   2674 | GE   | CH   | BO   | ZO   | UG3  |
| 12575-PC-2007   |   2688 | GE   | CH   | BO   | ZO   | UG3  |
| 12576-PC-2007   |   2598 | GE   | CH   | BO   | ZO   | UG3  |
| 12579-PC-2007   |   2599 | GE   | CH   | BO   | ZO   | UG3  |
| 12583-PC-2007   |   2605 | GE   | CH   | BO   | ZO   | UG3  |
| 12585-PC-2007   |   2717 | GE   | CH   | BO   | ZO   | UG3  |
| 12588-PC-2007   |   2694 | GE   | CH   | BO   | ZO   | UG3  |
| 12589-PC-2007   |   2698 | GE   | CH   | BO   | ZO   | UG3  |
| 12590-PC-2007   |   2620 | GE   | CH   | BO   | ZO   | UG3  |
| 12591-PC-2007   |   2626 | GE   | CH   | BO   | ZO   | UG3  |
| 12592-PC-2007   |   2671 | GE   | CH   | BO   | ZO   | UG3  |
| 12593-PC-2007   |   2706 | GE   | CH   | BO   | ZO   | UG3  |
```

Figure 4: Data of Enrollment_12 Table

iii. Database Marks : Table Xmarks

```
MariaDB [Marks]> desc Xmarks;
+--------+-------------+------+-----+---------+-------+
| Field  | Type        | Null | Key | Default | Extra |
+--------+-------------+------+-----+---------+-------+
| rollno | varchar(20) | NO   | PRI |         |       |
| sub    | varchar(10) | NO   | PRI |         |       |
| paper  | char(1)     | NO   | PRI |         |       |
| marks  | int(5)      | YES  |     | NULL    |       |
+--------+-------------+------+-----+---------+-------+
4 rows in set (0.00 sec)

MariaDB [Marks]>
```

Figure 5: Structure of Xmarks Table

```
MariaDB [Marks]> select * from Xmarks
+--------+-----+-------+-------+
| rollno | sub | paper | marks |
+--------+-----+-------+-------+
| 1      | BO  | A     |    24 |
| 1      | BO  | B     |    26 |
| 1      | CH  | A     |    19 |
| 1      | CH  | B     |    12 |
| 1      | GE  | A     |    41 |
| 1      | GE  | B     |    39 |
| 1      | ZO  | A     |    18 |
| 1      | ZO  | B     |    18 |
| 2      | BO  | A     |    23 |
| 2      | BO  | B     |    29 |
| 2      | CH  | A     |    24 |
| 2      | CH  | B     |    15 |
| 2      | GE  | A     |    39 |
| 2      | GE  | B     |    55 |
| 2      | ZO  | A     |    18 |
| 2      | ZO  | B     |    12 |
| 3      | BO  | A     |    14 |
| 3      | BO  | B     |    24 |
```

Figure 6: Data of Xmarks Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

iv. Database uresult : Table Result

```
MariaDB [uresult]> desc Result;
+--------+-------------+------+-----+---------+-------+
| Field  | Type        | Null | Key | Default | Extra |
+--------+-------------+------+-----+---------+-------+
| regno  | varchar(15) | NO   | PRI |         |       |
| rollno | int(7)      | NO   | PRI | 0       |       |
| result | varchar(15) | YES  |     | NULL    |       |
+--------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

Figure 7: Structure of Result Table

```
MariaDB [uresult]> select * from Result;
+-----------------+--------+--------------+
| regno           | rollno | result       |
+-----------------+--------+--------------+
| 10034-PC-2006   |   2676 | PASS/374     |
| 11632-AW-2004   |   2540 | PASS/361     |
| 12497-PC-2007   |   2617 | PASS/395     |
| 12499-PC-2007   |   2680 | PASS/395     |
| 12500-PC-2007   |   2681 | PASS/381     |
| 12501-PC-2007   |   2704 | PASS/365     |
| 12503-PC-2007   |   2641 | PASS/317     |
| 12508-PC-2007   |   2701 | PASS/393     |
| 12509-PC-2007   |   2595 | PASS/344     |
| 12513-PC-2007   |   2685 | PASS/411     |
| 12514-PC-2007   |   2683 | PASS/378     |
| 12515-PC-2007   |   2618 | PASS/419     |
| 12523-PC-2007   |   2657 | PASS/387     |
| 12525-PC-2007   |   2711 | PASS/365     |
| 12526-PC-2007   |   2649 | PASS/421     |
| 12527-PC-2007   |   2628 | PASS/388     |
| 12528-PC-2007   |   2646 | PASS/393     |
| 12529-PC-2007   |   2643 | PASS/340     |
| 12534-PC-2007   |   2596 | PASS/375     |
| 12548-PC-2007   |   2710 | PASS/394     |
| 12571-PC-2007   |   2607 | PASS/400     |
| 12572-PC-2007   |   2674 | PASS/430     |
| 12575-PC-2007   |   2688 | PASS/397     |
| 12576-PC-2007   |   2598 | PASS/412     |
| 12579-PC-2007   |   2599 | PASS/400     |
| 12583-PC-2007   |   2605 | PASS/416     |
| 12585-PC-2007   |   2717 | PASS/392     |
| 12588-PC-2007   |   2694 | PASS/352     |
| 12589-PC-2007   |   2698 | PASS/388     |
| 12590-PC-2007   |   2620 | PASS/393     |
| 12591-PC-2007   |   2626 | PASS/375     |
| 12592-PC-2007   |   2671 | PASS/371     |
| 12593-PC-2007   |   2706 | PASS/381     |
```

Figure 8: Data of Result Table

IV. PROPOSED ALGORITHMIC SOLUTION

We propose automated warehouse creation algorithm, and then implement the same using Linux and Java. The solution makes certain assumptions as;

i. Operational Sources have been identified.
ii. Operational Sources can be remotely accessed via IP address.
iii. Have read rights of the operational sources.
iv. Operational Sources are consistent.

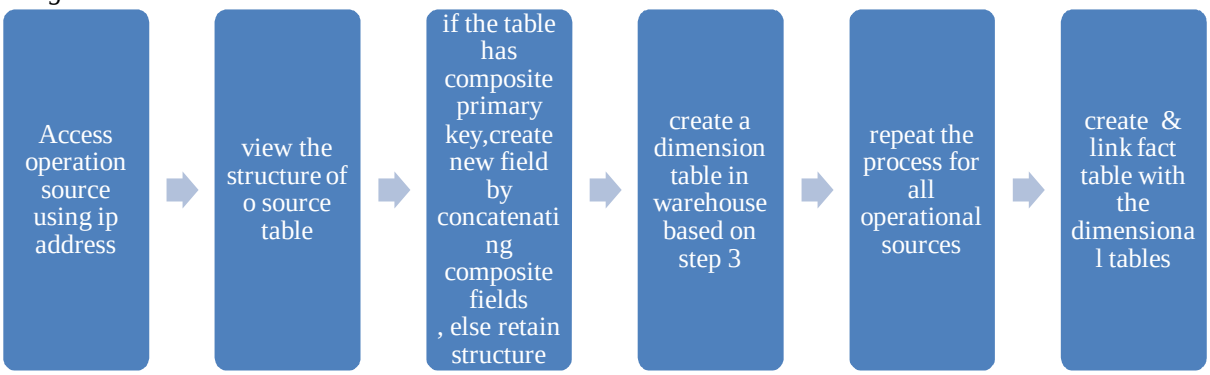
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

A. Algorithm:



B. Elaboration:

i. Establish JDBC connection with operation Source.
ii. View structure of the table which has been identified e.g;
"Database enrollment: Table Enrollment_12 "
iii. If the table has composite primary key, create new primary key e.g;
Since Enrollment_12 has composite primary key (regno,rollno)
New primary key will be created i.e, regnorollno and data in this filed will be like
10034-PC-2006-2676-
iv. New dimension table will be created in warehouse.
v. Create fact table maintaining primary key foreign key relationship with the dimensional tables.

C. Results:

Data warehouse based on algorithm is created with following tables & structures;

```
MariaDB [fuok]> show tables;
+----------------+
| Tables_in_fuok |
+----------------+
| Enrollment_12  |
| Master         |
| Result         |
| Xmarks         |
| fuok_fact      |
+----------------+
5 rows in set (0.00 sec)
```

Figure 9: Tables in Warehouse

The Created Warehouse retains the names of the source tables however, with variant structure based on presence or absence of composite primary keys. The structure of data warehouse tables along with data is shown below.

i. Master :

```
MariaDB [fuok]> desc Master;
+-------+-------------+------+-----+---------+-------+
| Field | Type        | Null | Key | Default | Extra |
+-------+-------------+------+-----+---------+-------+
| regno | varchar(20) | NO   | PRI |         |       |
| name  | varchar(35) | YES  |     | NULL    |       |
| fname | varchar(35) | YES  |     | NULL    |       |
+-------+-------------+------+-----+---------+-------+
3 rows in set (0.00 sec)
```

Figure 10: Structure of Master Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

```
MariaDB [fuok]> select * from Master;
+----------------+---------------------------------+-------------------------------+
| regno          | name                            | fname                         |
+----------------+---------------------------------+-------------------------------+
| 10034-PC-2006  | SAZIYAH MAHBOOBA                | MOHMAD SYED SHAH              |
| 11632-AW-2004  | NARGIS JAN                      | MOHD AKRAM GANIE              |
| 12497-PC-2007  | IRSHAD NAZIR SHAH               | NAZIR AHMAD SHAH              |
| 12499-PC-2007  | ZAHOOR AHMAD TEELI              | ABDUL REHMAN TEELI            |
| 12500-PC-2007  | MOHD AMIN TEELI                 | ABDUL GANI TEELI              |
| 12501-PC-2007  | OYAIS JAN                       | NAZIR AHMAD KAW               |
| 12503-PC-2007  | AB WAHID MIR                    | AB GANI MIR                   |
| 12508-PC-2007  | SHEERAZ AHMAD MIR               | AB GANI MIR                   |
| 12509-PC-2007  | AJAZ AHMAD BHAT                 | AB KHALIQ BHAT                |
| 12513-PC-2007  | GAZENFAR GULL                   | GH MOHD DAR                   |
| 12514-PC-2007  | YOUNOUS GANI WANI               | ABDUL GANI WANI               |
| 12515-PC-2007  | JAVEED AHMAD DAR                | ASSADULLAH DAR                |
| 12523-PC-2007  | MUDASIR MAQBOOL                 | MOHAMMAD MAQBOOL SOFI         |
| 12525-PC-2007  | ROUF FAYAZ                      | FAYAZ AHMAD TEELI             |
```

Figure 11: Data of Master Table

ii. Enrollment_12:

```
MariaDB [fuok]> desc Enrollment_12;
+-------------+-------------+------+-----+---------+-------+
| Field       | Type        | Null | Key | Default | Extra |
+-------------+-------------+------+-----+---------+-------+
| regno       | varchar(15) | YES  |     | NULL    |       |
| rollno      | int(9)      | YES  |     | NULL    |       |
| sub1        | varchar(4)  | YES  |     | NULL    |       |
| sub2        | varchar(4)  | YES  |     | NULL    |       |
| sub3        | varchar(4)  | YES  |     | NULL    |       |
| sub4        | varchar(4)  | YES  |     | NULL    |       |
| year        | varchar(4)  | YES  |     | NULL    |       |
| regnorollno | varchar(50) | NO   | PRI |         |       |
+-------------+-------------+------+-----+---------+-------+
8 rows in set (0.09 sec)
```

Figure 12: Structure of Enrollment_12 Table

```
MariaDB [fuok]> select * from Enrollment_12;
+----------------+--------+------+------+------+------+------+-----------------------+
| regno          | rollno | sub1 | sub2 | sub3 | sub4 | year | regnorollno           |
+----------------+--------+------+------+------+------+------+-----------------------+
| 10034-PC-2006  |   2676 | GE   | CH   | BO   | ZO   | UG3  | 10034-PC-2006-2676-   |
| 11632-AW-2004  |   2540 | GE   | ZO   | BO   | CH   | UG3  | 11632-AW-2004-2540-   |
| 12497-PC-2007  |   2617 | GE   | CH   | BO   | ZO   | UG3  | 12497-PC-2007-2617-   |
| 12499-PC-2007  |   2600 | GE   | CH   | BO   | ZO   | UG3  | 12499-PC-2007-2600-   |
| 12500-PC-2007  |   2681 | GE   | CH   | BO   | ZO   | UG3  | 12500-PC-2007-2681-   |
| 12501-PC-2007  |   2704 | GE   | CH   | BO   | ZO   | UG3  | 12501-PC-2007-2704-   |
| 12503-PC-2007  |   2641 | GE   | CH   | BO   | ZO   | UG3  | 12503-PC-2007-2641-   |
| 12508-PC-2007  |   2701 | GE   | CH   | BO   | ZO   | UG3  | 12508-PC-2007-2701-   |
| 12509-PC-2007  |   2595 | GE   | CH   | BO   | ZO   | UG3  | 12509-PC-2007-2595-   |
| 12513-PC-2007  |   2685 | GE   | CH   | BO   | ZO   | UG3  | 12513-PC-2007-2685-   |
| 12514-PC-2007  |   2683 | GE   | CH   | BO   | ZO   | UG3  | 12514-PC-2007-2683-   |
| 12515-PC-2007  |   2618 | GE   | CH   | BO   | ZO   | UG3  | 12515-PC-2007-2618-   |
| 12523-PC-2007  |   2657 | GE   | CH   | BO   | ZO   | UG3  | 12523-PC-2007-2657-   |
| 12525-PC-2007  |   2711 | GE   | CH   | BO   | ZO   | UG3  | 12525-PC-2007-2711-   |
| 12526-PC-2007  |   2649 | GE   | CH   | BO   | ZO   | UG3  | 12526-PC-2007-2649-   |
| 12527-PC-2007  |   2620 | GE   | CH   | BO   | ZO   | UG3  | 12527-PC-2007-2620-   |
| 12528-PC-2007  |   2646 | GE   | CH   | BO   | ZO   | UG3  | 12528-PC-2007-2646-   |
| 12529-PC-2007  |   2643 | GE   | CH   | BO   | ZO   | UG3  | 12529-PC-2007-2643-   |
| 12534-PC-2007  |   2596 | GE   | CH   | BO   | ZO   | UG3  | 12534-PC-2007-2596-   |
```

Figure 13: Data of Enrollment_12 Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

iii. Xmarks:

```
MariaDB [fuok]> desc Xmarks;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| rollno         | varchar(20) | YES  |     | NULL    |       |
| sub            | varchar(10) | YES  |     | NULL    |       |
| paper          | char(1)     | YES  |     | NULL    |       |
| marks          | int(5)      | YES  |     | NULL    |       |
| rollnosubpaper | varchar(50) | NO   | PRI |         |       |
+----------------+-------------+------+-----+---------+-------+
5 rows in set (0.00 sec)
```

Figure 14: Structure of Xmarks Table

rollno	sub	paper	marks	rollnosubpaper
1	BO	A	24	1-BO-A-
1	BO	B	26	1-BO-B-
1	CH	A	19	1-CH-A-
1	CH	B	12	1-CH-B-
1	GE	A	41	1-GE-A-
1	GE	B	39	1-GE-B-
1	ZO	A	18	1-ZO-A-
1	ZO	B	18	1-ZO-B-
2	BO	A	23	2-BO-A-
2	BO	B	29	2-BO-B-
2	CH	A	24	2-CH-A-
2	CH	B	15	2-CH-B-
2	GE	A	39	2-GE-A-
2	GE	B	55	2-GE-B-
2	ZO	A	18	2-ZO-A-
2	ZO	B	12	2-ZO-B-
3	BO	A	14	3-BO-A-
3	BO	B	24	3-BO-B-
3	CH	A	21	3-CH-A-

Figure 15: Data of Xmarks Table

iv. Result:

```
MariaDB [fuok]> desc Result;
+---------------+-------------+------+-----+---------+-------+
| Field         | Type        | Null | Key | Default | Extra |
+---------------+-------------+------+-----+---------+-------+
| regno         | varchar(15) | YES  |     | NULL    |       |
| rollno        | int(7)      | YES  |     | NULL    |       |
| result        | varchar(15) | YES  |     | NULL    |       |
| regnorollno1  | varchar(50) | NO   | PRI |         |       |
+---------------+-------------+------+-----+---------+-------+
4 rows in set (0.00 sec)
```

Figure 16: Structure of Result Table

```
MarialDB [fuok]> select * from Result;
+-----------------+----------+-------------+-----------------------+
| regno           | rollno   | result      | regnorollno1          |
+-----------------+----------+-------------+-----------------------+
| 10034-PC-2006   |    2676  | PASS/374    | 10034-PC-2006-2676-   |
| 11632-AW-2004   |    2540  | PASS/361    | 11632-AW-2004-2540-   |
| 12497-PC-2007   |    2617  | PASS/395    | 12497-PC-2007-2617-   |
| 12499-PC-2007   |    2680  | PASS/395    | 12499-PC-2007-2680-   |
| 12500-PC-2007   |    2681  | PASS/381    | 12500-PC-2007-2681-   |
```

Figure 17: Data of Result Table

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

v. fuok_fact:

```
MariaDB [fuok]> desc fuok_fact;
+----------------+-------------+------+-----+---------+-------+
| Field          | Type        | Null | Key | Default | Extra |
+----------------+-------------+------+-----+---------+-------+
| regnorollno    | varchar(50) | YES  | MUL | NULL    |       |
| regno          | varchar(20) | YES  | MUL | NULL    |       |
| regnorollno1   | varchar(50) | YES  | MUL | NULL    |       |
| rollnosubpaper | varchar(50) | YES  | MUL | NULL    |       |
+----------------+-------------+------+-----+---------+-------+
4 rows in set (0.00 sec)
```

Figure 18: Structure of fuok_fact Table

```
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-GE-A-
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-GE-B-
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-IT-A-
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-IT-B-
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-MA-A-
| 50188-W-2007-965- | 50188-W-2007 | 50188-W-2007-965- | 965-MA-B-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-CH-A-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-CH-B-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-GE-A-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-GE-B-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-MA-A-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-MA-B-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-PH-A-
| 50197-W-2007-966- | 50197-W-2007 | 50197-W-2007-966- | 966-PH-B-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-CH-A-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-CH-B-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-GE-A-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-GE-B-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-MA-A-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-MA-B-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-PH-A-
| 50198-W-2007-967- | 50198-W-2007 | 50198-W-2007-967- | 967-PH-B-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-CH-A-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-CH-B-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-GE-A-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-GE-B-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-MA-A-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-MA-B-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-PH-A-
| 50199-W-2007-968- | 50199-W-2007 | 50199-W-2007-968- | 968-PH-B-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-CH-A-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-CH-B-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-GE-A-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-GE-B-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-MA-A-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-MA-B-
| 50201-W-2007-969- | 50201-W-2007 | 50201-W-2007-969- | 969-PH-A-
```

Figure 19: Data of fuok_fact Table

V. CONCLUSION AND FUTURE WORK

With almost all databases designed using third normal form Algometric solution is proposed for design of data warehouse using a university data set. Algorithm was implemented in java on Linux platform (Fedora). The algorithm focuses completely on design of dimension tables and resultant fact table. Resultant warehouse was tested for basic OLAP operations and gave satisfactory results. Researchers/engineers are encouraged to test it on different and large data sets so that algorithm (if required) can be accordingly modified/enhanced.

REFERENCES

1. Majid Zaman, Muheet Ahmed Butt,"Warehouse Creator: A Generic Enterprise Solution",International Journal Of Engineering And Science, Vol.2, Issue 12 (May 2013), Pp 65-68.
2. Shaweta," Critical Need of the Data Warehouse for an Educational Institution and Its Challenges", International Journal of Computer Science and Information Technologies,Vol. 5 (3) , 2014, 4556-4559.
3. Salvatore T. March, and Alan R. Hevner. "Integrated decision support systems: A data warehousing perspective." Decision Support Systems 43.3 (2007): 1031-1043.
4. Thomsen, Christian, Torben Bach Pedersen, and Wolfgang Lehner. "RiTE: Providing on-demand data for right-time data warehousing." Data Engineering, 2008. ICDE 2008. IEEE 24th International Conference on. IEEE, 2008.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 9, September 2015

5. Ricardo Santos Jorge, and Jorge Bernardino. "Optimizing data warehouse loading procedures for enabling useful-time data warehousing." Proceedings of the 2009 International Database Engineering & Applications Symposium. ACM, 2009.
6. Ashish Thusoo, Borthakur, Raghotham Murthy, Zheng Shao, Namit Jain, Hao Liu, Suresh Anthony, Joydeep Sen Sarma , "Data warehousing and analytics infrastructure at facebook." Proceedings of the 2010 ACM SIGMOD International Conference on Management of data. ACM, 2010.
7. Arun Sen, K. Ramamurthy, and Atish P. Sinha. "A model of data warehousing process maturity." Software Engineering, IEEE Transactions on 38.2 (2012): 336-353.
8. Kamil Bajda-Pawlikowski, Daniel J. Abadi, Avi Silberschatz, Erik Paulson "Efficient processing of data warehousing queries in a split execution environment." Proceedings of the 2011 ACM SIGMOD International Conference on Management of data. ACM, 2011.
9. Calvanese, Giuseppe De Giacomo, Maurizio Lenzerini, Daniele Nardi, and Riccardo Rosati, "Data integration in data warehousing." International Journal of Cooperative Information Systems 10.03 (2001): 237-271.
10. Hugh J. Watson, "Recent developments in data warehousing." Communications of the Association for Information Systems 8.1 (2002): 1.
11. Ricardo Santos Jorge, and Jorge Bernardino. "Optimizing data warehouse loading procedures for enabling useful-time data warehousing." Proceedings of the 2009 International Database Engineering & Applications Symposium. ACM, 2009.
12. Cuzzocrea, Alfredo. "Analytics over big data: Exploring the convergence of datawarehousing, olap and data-intensive cloud infrastructures." 2013 IEEE 37th Annual Computer Software and Applications Conference. IEEE, 2013.

BIOGRAPHY

Shahid Bashir Dar is the student of MTECH department of CSE, BIMT Shimla(HP) India. He has done his BTECH in CSE from IUST Kashmir. His research interests are Data warehouse, Data and network Security.

Ashish Sharma is working as an Assistant Professor in the department of CSE , BIMT Shimla(HP). His research interests are data and cloud security, databases etc.