



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

Analysis of Telecommunication Data: Call Drop using Star Schema

Navin Trivedi¹, Vikas Singh², Pradeep Singh Rana³, Ravi Pratap Singh⁴, Yashwant Sinha⁵

Professor, Dept. of CSE, Mahatma Gandhi Mission College of Engineering and Technology, Noida, India.¹

B.Tech Student, Dept. of CSE, Mahatma Gandhi Mission College of Engineering And Technology, Noida, India.^{2,3,4,5}

ABSTRACT: Telecommunication industries have a very large database which is containing the information about the ongoing calls and the consumers which are currently subscribed to the telecom company. With the increase in the population their us also a increase in the users which are using the mobile phones and with the decrease in the rate of mobile phones and the tariff of the telecom services there is a fast growing users which are using the services. With the increase in the users there is also a increase in the number of call drop which are occurring in day to day life. Telecom industries have a diverse form of data which are stored in separate departmental databases and it becomes difficult to gather the information from the database.

The aim of this paper is to implement the telecommunication database using star schema, In which we will be finding the ratio of call drop which are occurring in some specific location and gather the information and where about of the occurrence of the call drop.

KEYWORDS: Data Collection, Designing of Star Schema, Data entry to Fact table and Dimension table, Adhoc Association Rule in star schema

I. INTRODUCTION

A telecommunication data warehouse contains the large number of heterogeneous database which includes information starting from the personal details of the user to the calls he/she has made and how much time he/she has consumed in chatting with the friends.

The paper includes the analysis information of our project in which we have collected the information of the user which includes bills, service line, plan, call drop reasons. In our project we have implemented telecommunication data using star schema. Star schema is basically used as its complexity is less and the performance time of the data base increases drastically as with compare to the other schemas. Our project can be split into 4 stages. Our project starts from Data Collection, Designing of Star Schema, Data entry to Fact table and Dimension table and Adhoc association rule in star schema. In our project we have focused on every aspect of the telecommunication data and we have analyzed the customer drop data for analyzing the call drop pattern and call drop problem.

II. LITERATURE SURVEY

In the telecommunication area there has been vast research going on to minimize the call drop rate in day to day life. The telecom industries are doing their best to overcome the call drop problem by increasing the number of BCS (Base Station Controller), MSC (Master Station Controller), by implementing the routing algorithm or by increasing the bandwidth of the transmission channel. In this paper we have analyzed the particular areas which are causing the maximum call drop.

III. DATA COLLECTION

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes. The goal for all data collection is to capture quality evidence that then translates to rich data analysis and allows the building of a convincing and credible

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

answer to questions that have been posed. In the project we have collected data for the customer table, bill table, rate table, service line table and the call drop table. Every table comprises of approx.1 lakh information of users.

IV. LOGICAL STRUCTURE OF DATA WAREHOUSE

Star schema the star schema is the simplest data warehouse schema. It is called a star schema because the diagram of a star schema resembles a star, with points radiating from a center. The center of the star consists of one or more fact tables and the points of the star are the dimension tables. A star schema is characterized by one or more very large fact tables that contain the primary information in the data warehouse and a number of much smaller *dimension* tables (or *lookup* tables), each of which contains information about the entries for a particular attribute in the fact table.

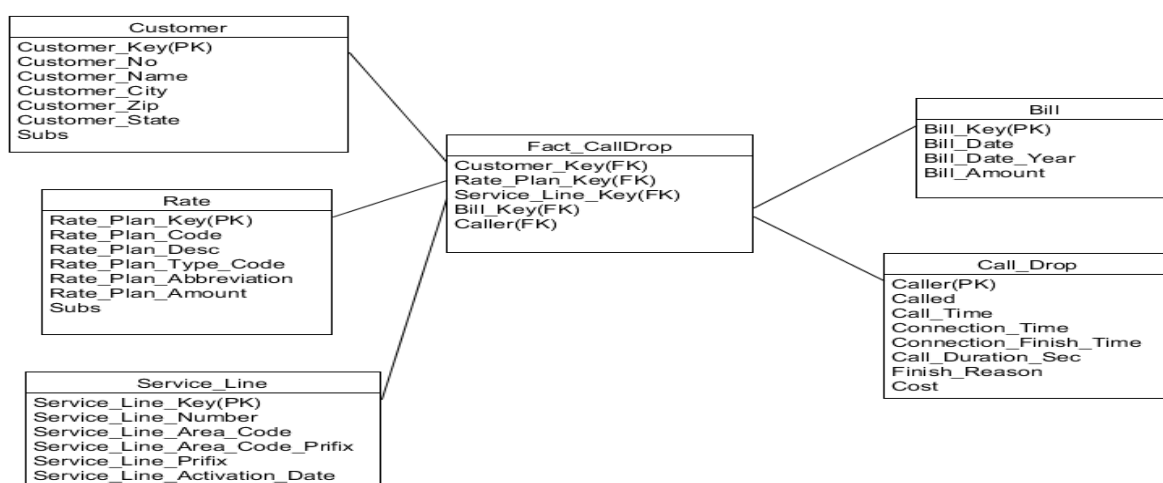


Figure3.1 Star schema of Telecommunication Database

This is the Star Schema which we have designed to analyze the telecommunication data and finding out the call drop probability. The star schema which we have created has five dimension table and one fact table. The dimension tables are customer table, rate table, service line table, bill table, call drop table and Fact Call Drop is the fact table. The customer table contains customer key as a primary key which is used to uniquely identifies a customer likewise rate table consist of a rate key, Service line table consist of a service line key, bill table consist of bill key and call drop table consist caller as a primary key.

A fact table is the central table in a star schema of a data warehouse. A fact table stores quantitative information for analysis and is often demoralized. A fact table works with dimension tables. A fact table holds the data to be analyzed, and a dimension table stores data about the ways in which the data in the fact table can be analyzed. Thus, the fact table consists of two types of columns. The foreign keys column allows joins with dimension tables, and the measures columns contain the data that is being analyzed.

V. PHYSICAL STRUCTURE

A. Customer table: This table will contain the information about the customer and the attribute it will contain are as follows customer key, number, name, city, zip code, state, subs.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

```
Run SQL Command Line

SQL> desc customer;
Name                                     Null?   Type
-----
CUSTOMER_KEY                            NOT NULL VARCHAR2(10)
CUSTOMER_NO                              NUMBER(10)
CUSTOMER_NAME                            NOT NULL CHAR(50)
CUSTOMER_CITY                            VARCHAR2(20)
CUSTOMER_ZIP                              NUMBER(6)
CUSTOMER_STATE                            CHAR(4)
SUB                                       VARCHAR2(3)

SQL>
```

B. Bill table: This table will include the information about customer bill and contains the following attribute bill key, date, year and bill amount.

```
Run SQL Command Line

SQL> desc bill;
Name                                     Null?   Type
-----
BILL_KEY                                 NOT NULL VARCHAR2(10)
BILL_DATE                                DATE
BILL_DATE_YEAR                           NUMBER(4)
AMOUNT                                    NUMBER

SQL>
```

C. Rate table: This table will contains the rate plan for the telecom companies and will contain the following attribute rate plan key, code, description, type code, abbreviation, amount, subs.

```
Run SQL Command Line

SQL> desc rate;
Name                                     Null?   Type
-----
RATE_PLAN_KEY                            NOT NULL VARCHAR2(30)
RATE_PLAN_CODE                            NOT NULL VARCHAR2(50)
RATE_PLAN_DESC                            VARCHAR2(50)
RATE_PLAN_TYPE_CODE                       NOT NULL VARCHAR2(50)
RATE_PLAN_ABBREVIATION                    VARCHAR2(40)
SUB                                       VARCHAR2(3)

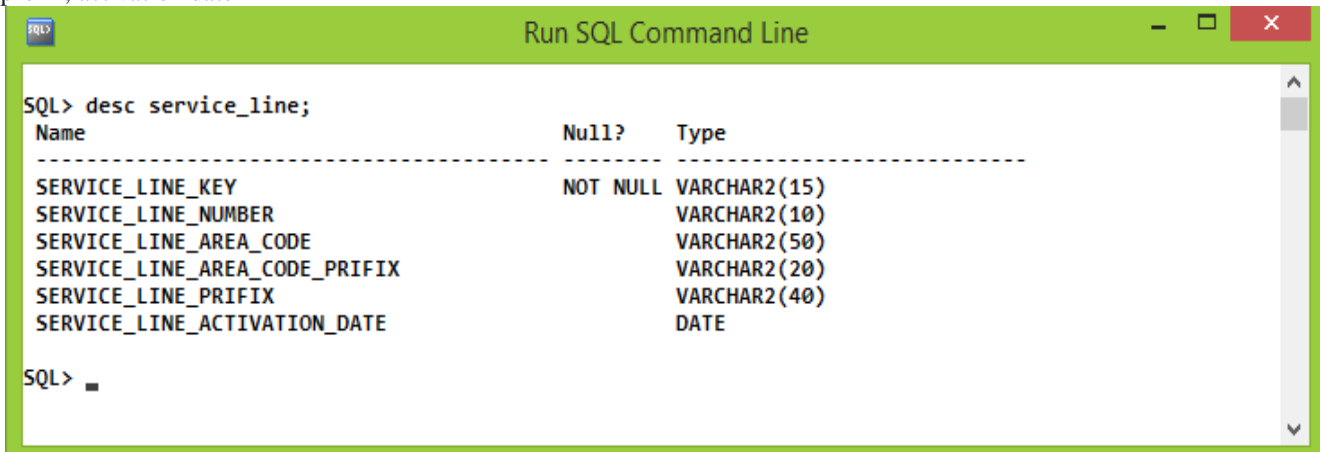
SQL>
```

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

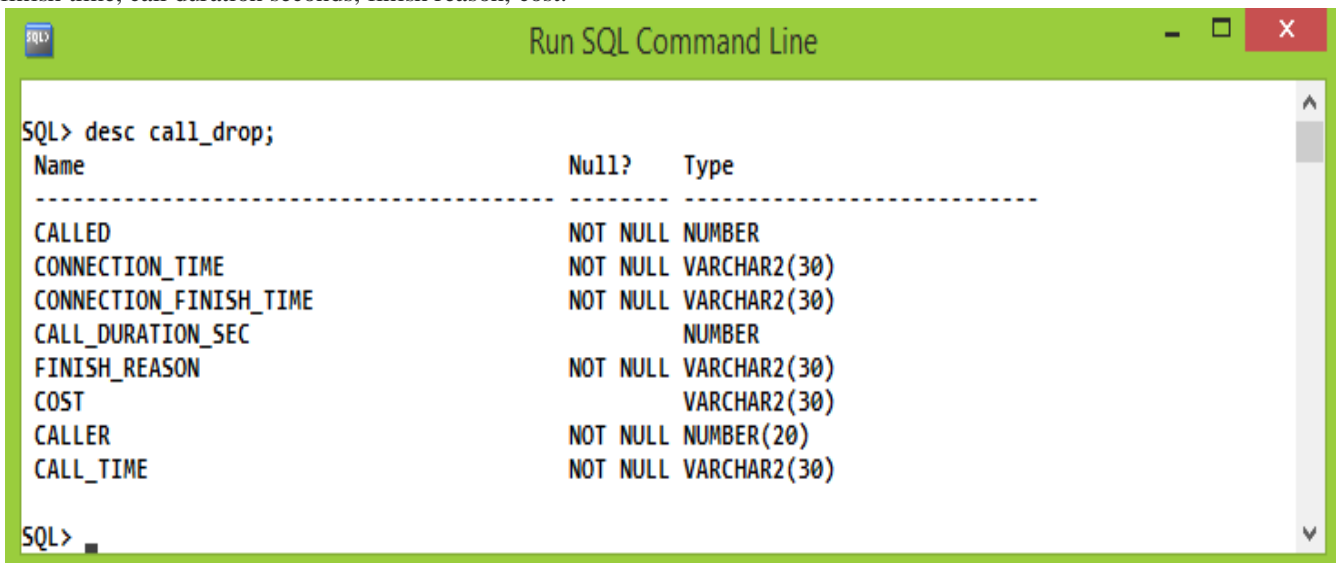
Vol. 4, Issue 5, May 2016

D. Service line table: This table contains the following attribute service line key, number, area code, area code prefix, prefix, activation date



```
Run SQL Command Line
SQL> desc service_line;
Name                                     Null?    Type
-----
SERVICE_LINE_KEY                       NOT NULL VARCHAR2(15)
SERVICE_LINE_NUMBER                     VARCHAR2(10)
SERVICE_LINE_AREA_CODE                   VARCHAR2(50)
SERVICE_LINE_AREA_CODE_PREFIX            VARCHAR2(20)
SERVICE_LINE_PREFIX                       VARCHAR2(40)
SERVICE_LINE_ACTIVATION_DATE              DATE
SQL>
```

E. Call drop table: This table will contain the following attribute caller, called, call time, connection time, connection finish time, call duration seconds, finish reason, cost.



```
Run SQL Command Line
SQL> desc call_drop;
Name                                     Null?    Type
-----
CALLED                                   NOT NULL NUMBER
CONNECTION_TIME                          NOT NULL VARCHAR2(30)
CONNECTION_FINISH_TIME                    NOT NULL VARCHAR2(30)
CALL_DURATION_SEC                          NUMBER
FINISH_REASON                              NOT NULL VARCHAR2(30)
COST                                       VARCHAR2(30)
CALLER                                    NOT NULL NUMBER(20)
CALL_TIME                                  NOT NULL VARCHAR2(30)
SQL>
```

VI. DATA ENTRY TO DIMENSION TABLE AND FACT TABLE

For the data entry purpose we have used the oracle express edition in which we have import the excel data file into the oracle database. Oracle express edition provides the feature of directly extracting the data from the excel file to the data on the oracle database tables. This intern decrease the manual effort of writing the insert query and adding it manually to the database, lot of time can be saved in importing the data directly to the data base through oracle express edition. If there is any invalid entry in the table then oracle express edition will show an error message which helps in keeping the integrity of the table as well as database.

The following screen shot clearly depict the import of excel file to oracle database bill table.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

A. Step 1

ORACLE Database Express Edition



User: SYSTEM

Home > Utilities > Data Load/Unload > Load > Load Data

Target and Method	Load Data	Cancel	Next >
Data	Load To:		
Table Properties	<input checked="" type="radio"/> Existing table		
Primary Key	<input type="radio"/> New table		
	Load From:		
	<input checked="" type="radio"/> Upload file (comma separated or tab delimited)		
	<input type="radio"/> Copy and paste (up to 30KB)		

Load Data

You can load text or spreadsheet data by either copying and pasting text or by uploading a file. To copy and paste text, the data must be less than 30KB. For files larger than 30KB, save the file in a comma delimited (CSV) or tab delimited format. Then upload the file and load the data into a new or existing table.

Language: en-us

Application Express 2.1.0.00.39
Copyright © 1999, 2008, Oracle. All rights reserved.

B. Step 2

ORACLE Database Express Edition



User: SYSTEM

Home > Utilities > Data Load/Unload > Load > Load Data

Schema	Load Data	Cancel	< Previous	Next >
Table Name	* File	Choose File	bill.txt	
File Details	* Separator	^t		
Column Mapping	Optionally Enclosed By			
	<input checked="" type="checkbox"/> First row contains column names.			
	File Character Set	Unicode UTF-8		

Globalization

Load Data

Use this page to locate the file to be uploaded.

If the first row contains column names, select **First row contains column names**.

Language: en-us

Application Express 2.1.0.00.39
Copyright © 1999, 2008, Oracle. All rights reserved.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

C. Step 3

The screenshot shows the Oracle Database Express Edition 'Load Data' interface. The 'Define Column Mapping' section displays a table with the following data:

Column Names	BILL_KEY - varchar2(10) *	BILL_DATE - date	BILL_DATE_YEAR - number	AMOUNT - number
Row 1	bill1	05-08-14	2014	998
Row 2	bill2	28-02-14	2014	255
Row 3	bill3	16-01-14	2014	816
Row 4	bill4	20-04-14	2014	401
Row 5	bill5	01-04-14	2014	644
Row 6	bill6	16-01-14	2014	103
Row 7	bill7	06-04-14	2014	605
Row 8	bill8	17-05-14	2014	955
Row 9	bill9	28-03-14	2014	651
Row 10	bill10	11-06-14	2014	117
Row 11	bill11	18-01-14	2014	361
Row 12	bill12	09-05-14	2014	553
Row 13	bill13	28-06-14	2014	903
Row 14	bill14	06-06-14	2014	575
Row 15	bill15	22-02-14	2014	818
Row 16	bill16	15-05-14	2014	615
Row 17	bill17	14-04-14	2014	497

This screen shot clearly depicts the data from the excel file is completely transferred into the bill table.

VII. ADHOC ASSOCIATION RULE IN STAR SCHEMA

Association rule is basically used to discover a relationship between the given table and the attributes which are contained in them. Association rules are used to analyze the behavior and pattern of the data which relate to each other and how actually they are related to each other. There are two terms which are used in making the association between the attribute. Support is an indication of how frequently the items appear in the database. Confidence indicates the number of times the if/then statements have been found to be true. Adhoc association provides us a methodology for generating an association between two tables in a database without being there is no existing relationship. With the help of Ad hoc association rule it is now being possible for generating a relationship between two distinct tables.

Below are the Ad hoc association queries which we have implemented in our project.

A. `select c.customer_no,d.connection_time,c.customer_city,d.finish_reason from fact_calldrop F inner JOIN customer c ON f.customer_key=c.customer_key inner join call_drop d ON f.caller=d.caller where c.customer_city='gnnoida' and d.finish_reason='NETWORK_ERROR';`

```

CUSTOMER_NO CONNECTION_TIME          CUSTOMER_CITY
-----
FINISH_REASON
-----
7867773368 02-06-14 11:39          gn noida
NETWORK_ERROR

7906299942 05-06-14 6:03          gn noida
NETWORK_ERROR

9911876616 02-06-14 10:12          gn noida
NETWORK_ERROR

CUSTOMER_NO CONNECTION_TIME          CUSTOMER_CITY
-----
FINISH_REASON
-----
8794888407 03-06-14 16:13          gn noida
NETWORK_ERROR

1582 rows selected.

```

In this we have found out the customer which are present in greater noida and whose call finished reason is network error which means that a call drop is occurred.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

B. select c.customer_no, d.connection_time, c.customer_city, d.finish_reason from fact_calldrop F inner JOIN customer c ON f.customer_key=c.customer_key inner join call_drop d ONf.caller=d.callerwhere c.customer_city='noida' and d.finish_reason='NETWORK_ERROR';

```
Run SQL Command Line
```

CUSTOMER_NO	CONNECTION_TIME	CUSTOMER_CITY	FINISH_REASON
8402351245	03-06-14 15:27	noida	NETWORK_ERROR
852783613	05-06-14 4:55	noida	NETWORK_ERROR
9881955276	04-06-14 0:39	noida	NETWORK_ERROR
9189589636	01-06-14 17:20	noida	NETWORK_ERROR
8032697295	01-06-14 23:46	noida	NETWORK_ERROR
8263439040	06-06-14 13:30	noida	NETWORK_ERROR

3891 rows selected.

In this we have found out the customer which are present in noida and whose call finished reason is network error which means that a call drop is occurred.

C. select c.customer_no, d.connection_time, c.customer_city, d.finish_reason from fact_calldrop F inner JOIN customer c ONf.customer_key=c.customer_key inner join call_drop d ONf.caller=d.callerwhere c.customer_city='gzb' and d.finish_reason='NETWORK_ERROR';

```
Run SQL Command Line
```

8858402030	06-06-14 0:18	gzb	
9405768578	04-06-14 0:56	gzb	
8372391636	05-06-14 13:35	gzb	
9565977374	04-06-14 12:06	gzb	
8206676312	08-06-14 3:23	gzb	
9563765388	03-06-14 15:02	gzb	
9536370128	08-06-14 20:15	gzb	
8569684507	08-06-14 21:52	gzb	
8665474399	01-06-14 7:39	gzb	
8691333821	02-06-14 16:51	gzb	
9687253412	05-06-14 1:12	gzb	
CUSTOMER_NO	CONNECTION_TIME	CUSTOMER_CITY	FINISH_REASON
9156551853	04-06-14 5:50	gzb	
9505346579	04-06-14 5:15	gzb	
8574637257	07-06-14 15:52	gzb	
9008350679	08-06-14 10:32	gzb	
9716199730	05-06-14 13:41	gzb	
9204538897	05-06-14 4:06	gzb	
9112759051	07-06-14 3:19	gzb	
8714900808	06-06-14 22:07	gzb	
8416412894	01-06-14 7:37	gzb	

1593 rows selected.

In this we have found out the customer which are present in Ghaziabad and whose call finished reason is network error which means that a call drop is occurred.

D. select c.customer_name, c.customer_no, c.customer_city, c.subs from customer c inner join fact_calldrop f ON f.customer_key=c.customer_key inner join call_drop d on f.caller =d.caller where c.sub='3g' and d.finish_reason='NETWORK_ERROR';



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

```
Run SQL Command Line
CUSTOMER_NAME ----- CUSTOMER_NO
CUSTOMER_CITY SUB
Cauvery 7906299942
gn noida 3g
Chandalini 8552439685
gbz 3g
Deepabali 8794888407
gn noida 3g
CUSTOMER_NAME ----- CUSTOMER_NO
CUSTOMER_CITY SUB
Dhanya 8416412894
gzb 3g
Dhriti 9977309599
gbz 3g
3104 rows selected.
```

In query shows the people who are using 3g and call finished reason is network error irrespective of the location of the user.

E. select c.customer_name, c.customer_no, c.customer_city from customer c inner join fact_calldrop f ON f.customer_key=c.customer_keyinner join call_drop d on f.caller =d.caller where c.sub='2g' and d.finish_reason='NETWORK_ERROR';

```
Run SQL Command Line
CUSTOMER_NAME ----- CUSTOMER_NO
CUSTOMER_CITY SUB
Divya 8093174524
gzb 2g
Draupadi 7866341995
noida 2g
Dristi 9987018076
gn noida 2g
CUSTOMER_NAME ----- CUSTOMER_NO
CUSTOMER_CITY SUB
Dulari 8442324588
noida 2g
67060 rows selected.
SQL>
```

In query shows the people who are using 2g and call finished reason is network error irrespective of the location of the user.

VIII. CONCLUSION

The conclusions of the project yield that the call drop occurred in these areas are more than the TRAI permitted limit. We have survey a approx. of 1 lakh customers and then we have analyzed the data given by them. In Noida 3891 customer are face the problem of Network error. In Greater Noida 1582 customer are face the problem of Network error. In Ghaziabad 1593 customer are face the problem of Network error. In Noida 3.83% people are facing the problem of call drop. In Greater Noida 1.58% people are facing the problem of call drop. In Ghaziabad 1.583% people are facing the problem of call drop. In approx. 1 Lakh customer data 67060 customers are still using 2g network. In approx 1 Lakh customer data 33171 customers are using 3g network. In 6369 customer are facing the call drop problem which is using 2g. In 3104 customer are facing the call drop problem which is using 3g.

IX. ACKNOWLEDGMENT

I am highly obliged to Mr. Navin Trivedi for his guidance, support, supervision and help during the research work and I also thank to friends and college management for supporting us. The authors would also be grateful to the reviewers for their valuable comments and suggestions in order to improve the quality of the paper.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 5, May 2016

REFERENCES

1. The data warehouse toolkit: the complete guide to dimensional modeling by Ralph Kimball, Margy Ross. — 2nd ed. (ISBN 0-471-20024-7).
2. Watson, H. J., Annino, D. A., and Wixom, B. H. Current Practices in Data Warehousing. Information Systems Management. 18 (1), (2001).
3. Agrawal R., Imielinski T. and A. Swami. Mining Association Rules between Sets of Items in Large Databases. Proceeding of ACM SIGMOD International Conference. (1993), 207-216.
4. SvetlozarNestorov and Ad-Hoc Association-Rule Mining within the Data Warehouse Proceedings of the 36th Hawaii International Conference on System Sciences - 2003
5. Data Mining: Concepts and Techniques by Jiawei Han, Micheline Kamber, Jian Pei-2nd edition (ISBN-9780080475585)
6. Agrawal R. and Srikant R. Fast Algorithms for Mining Association Rules. Proceeding of International Conference On Very Large Databases VLDB. (1994), 487-499.
7. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A. Verkamo. Fast Discovery of Association Rules. Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press (1996).
8. Berry, M. and Linoff, G. Data Mining Techniques for Marketing, Sales and Customer Support. Wiley (1997).
9. Chaudhri, S. and Dayal, U. An overview of Data Warehousing and OLAP Technology. ACM SIGMOD Record 26 (1), (1997), 65-74.
10. Hilderman, R.J., Carter, C.L., Hamilton, H.J., and Cercone, N. Mining Association Rules from Market Basket Data Using Share Measures and Characterized Itemsets. International Journal of Artificial Intelligence Tools. 7 (2), (1998), 189-220.
11. Inmon, W. H. Building the Data Warehouse. Wiley (1996).
12. Kimball, R., Reeves, L., Ross, M., and Thornthwhite, W. The Data Warehouse Lifecycle Toolkit. Wiley (1998).
13. <http://searchbusinessanalytics.techtarget.com/definition/association-rules-in-data-mining>.
14. S. Sarawagi, S. Thomas, R. Agrawal. Integrating Mining with Relational Database Systems: Alternatives and Implications. Proceedings of ACM SIGMOD Conference (1998), 343-354
15. S. Tsur, J. Ullman, S. Abiteboul, C. Clifton, R. Motwani, S. Nestorov, A. Rosenthal. Query Flocks: A Generalization of Association-Rule Mining. Proceedings of ACM SIGMOD Conference, (1998), 1-12.
16. Wang, K., He Y., and Han J. Mining Frequent Itemsets Using Support Constraints. Proceedings of International Conference on Very Large Databases VLDB, (2000), 43-52