

# Design and Implementation of Data Cube for Khasra Khatauni in the Region of Uttar Pradesh

**Avinash Sonkar**

Department of Computer science & Engineering  
B. B. Ambedkar University, (A Central University)  
Lucknow, (U. P.), India

**Dr. Raj Shree**

Department of Information Technology  
B. B. Ambedkar University, (A Central University)  
Lucknow, (U. P.), India

## Abstract—

**O**n-Line Analytical Processing system (OLAP) has offered a strong locus on the interdependent optimization of the data and consistently arranges the inclusive efficiency for showing the data and performs arbitrary information. Data sets can be perform in the form of table, where each row represented by the objects and each column represented by the attributes. Data cube are express the multidimensional data with entirely desirable result. The 3-D data cube express the various attributes perfectly controlled by the help of objects. Normally, a data cube is observation of analytical terminology as a cross-index. In this research paper, wordsmith have represent a system of OLAP data cube to examine the ratio of area holder comparing with male and female, to recognize the entity, which is extremely preferred by the client. This research represents an approach of OLAP data cube and pivot table additionally optimize the co-relation technique that can perform a strong interconnection among the attributes of the data. Graph and table are constructing for the sample database of Khasra khatauni system.

**Keywords—** Data Cube, DSS, OLAP, Correlation Coefficient, pivot table, OOMD.

## I. INTRODUCTION

CUBE is representing a database in a 3-dimensional or multidimensional form for storing a large amount of data by the concept of olap technology. Olap cube technology design to object oriented database. Strength of OLAP & Data Mining as well as enhance model for banking system has been discussed in [1]. The methodology of OLAP data cube & pivot table is described for vehicle insurance policy with the help of correlation technology and show the strongest relationship among the attribute of the data and design the data with the assist of table and graph discussed in [2]. Modeling of the data and design the architecture of a data warehouse for uniquely access the data in the data warehouse has been discussed in this paper. And also discussed about many issues like logical model, conceptual model design the methodology for the data warehouse and able to exchange the data use the informational data as well as security of the data. [3]. MR-CUBE in which find the subset of constant or genuine measure by the help of holistic measures identifying the interesting cube group for that queries act as a set of class in decision support system. And enhance the performance for accessing the data in the OLAP CUBE. As well as handle the large amount of database and apply the parallel solution techniques with communication efficient partitioning technique has been discussed in [4]. The object oriented database is designed along with OLAP cube with sequence diagram, star schema, and UML diagram. That is quickly access the unique data in the database by the help of supporting query and enhances the accessibility of the data from the database for Indian Post Services that's why the productivity of the data is exclusive & very rapid [5]. To convene the learning process by providing the acquired knowledge and strength to each and every learners. And there are providing such tools to the teachers they can easily analyze the individual performance of the students will increase has been discussed [6]. By using semantic cube model approach we try to overcome from the problem of duplication of data as well as enhance the performance of query integrity & establish the strong relationship between the data information in the data cube have been discussed [7]. And focus on handling the huge amount of data in the database and complicated queries which are used in data warehouse. By using clustering techniques to provide the exact result of the complicated queries that enhance the performance of the queries for finding the interesting result. [8]. Discuss about enhance the performance of old complex queries by two times better result than the old one for better outcome of the queries an algorithms is proposed called as PDIC. This is the combination of three different strategies [9]. And to provide the information about the multimedia data set's by using powerful tool's & techniques and also discusses the concept of multimedia data mining, multimedia architecture, application & models in [10]. Enhance model of OLAP queries which strictly support a new data warehouse based on graph cube. Graph cube gives power for decision making support in large multidimensional network with the use of graph cube in a multidimensional network enhance the performance of OLAP queries are given in [11]. XML data cube model handles the acquired information in digital documental form. This also use an open source computing framework that process a huge data this is called apache hadoop by using these tools & techniques the efficiency of XML data is granular. As well as it also used optimization algorithms for the query performance of analysis [12]. And to optimizing power for parallel data cube is constructed by using ROLAP techniques. Enhances the output of the queries and analyzes system and algorithms that access an effective data in a few times. And also increase the processing power for access the data with the processor have been discussed [13]. This technology manages the multidimensional data cube according to handheld devices & keeping in mind that the data is being

compressed & should be according to the ratio of the device & view the information in two dimensional data cube. It also enhances the performance for accessibility of the informational data [14]. OLAP technology design the enhance model of query processing according to the operator for cube. This shows the huge data should be in a compressed & generalized form by using N – space. This SQL standard fits in the SQL database which improves the result of each query. At this point of time huge amount of data can be managed [15]. Design and maintenance of OLAP cube optimize the reduction of maintenance cost and queries time by using greedy algorithms. It is used to analysis the performance of maintenance of the cube and increases the access data quickly [16]. OLAP is one of the modern techniques which are used in the database technology OLAP technology that is related with the MD database & relational databases. This can be seen by interrelationship between the related model & the multidimensional attributes of the data by using MOLAP & ROLAP techniques in the data cube & designed the growing up model of relational database [17]. And analyze the performance evaluation of data warehouse environment. & propose the enhance model for accessing the informational data in the data warehouse by using OLAP tool [18]. In this research paper, there are consider the parameters of three dimensional databases for Khasra khatauni and selected database is represented by the three dimensional data cube. And pivot table are generated for OLAP data cube. Because A pivot table are dynamic tool for the data summarization. As well as OLAP cube is design by using OLAP tool.

## II. METHODOLOGY

### A. Olap Data Cube and Pivot Table

OLAP data cube has been created for analysis purpose retrieved from olap database. Most of the data in data storage are designed for on-line transaction processing. These olap cube techniques retrieve the information from the database and also clarify the efficiency of information. As the data stored in such a way so that it makes easy to retrieve and make correct reporting. OLAP data cube is divided into dimensions and measures. Dimensions are representing a standard parameter's and measures are representing a sub-parameter.

A pivot table is also represented in dimension and measures in Microsoft excel work sheet. Which allow representing the data from the data cube in summarized form. Pivot table uses olap cube to interact between user and the data storage. The information can be gathered from the pivot subset of filtering, grouping, subtotaling, sorting that you want.

In this sample table, data is described for Khasra khatauni. AREA\_IN\_HECTARE, NUMBER\_KHATA\_KHATAUNI AND ACCOUNT\_HOLDER are representing into dimensions and measures. Each and every NUMBER\_KHATA\_KHATAUNI contains ACCOUNT\_HOLDER name and AREA\_IN\_HECTARE. Khasra khatauni is represented in increasing order and account holder is sorted in gender. The male is represented by '1' and female is represented by '2'. The grant total of AREA\_IN\_HECTARE and NUMBER\_KHATA\_KHATAUNI is shown in pivot table.

Table 1: A pivot table of OLAP DATA CUBE

ROW LABELS	SUM OF AREA IN HECTARE	SUM OF NUMBER KHATA KHATAUNI
FEMALE	7.978	1943
MALE	22.8765	2486
GRAND TOTAL	30.8545	4429

### B. Correlation Technique

To define a correlation between two random variables A and B, this is the set of data

$$A = \{a_1, \dots, a_n\}$$

And

$$B = \{b_1, \dots, b_n\}$$

Usually, the linear correlation coefficient  $L_c$  is defined as

$$L_c = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2} \sqrt{\sum_i (b_i - \bar{b})^2}}$$

Where  $\bar{a}$  and  $\bar{b}$  are the usual means for defined values of a and b respectively. When the linear coefficient is known, the coefficient  $L_c$  is one in summarized strength. Since, there is no optimized way to compute the  $L_c$  result in hypothesis form, where the variable and b are not mutually correlated. We cannot decide whether the correlation is significant or not. Since,  $L_c$  is a weak static.

To overcome from this problem the use rank static. It has a probability distribution function which allows for calculation of significant correlation. In this paper the rank correlation has not detail theory, It is of because rank test is based on sorting algorithms that contains huge amount of data set. In place of correlation test the research contain sign test. Because it is more simple to use and very easy to implement.

Let us consider the following sample table 2, based on 3-D olap cube representation for Khasra khatauni. The table consists following data as show in table 2. The data table is AREA\_IN\_HECTARE, NUMBER\_KHATA\_KHATAUNI AND ACCOUNT\_HOLDER, and can be connected for the N number of khata Khatauni.

Table 2: Sample Data from Khasra khatauni

AREA_IN_HECTERE	number khata khatauni	account holder
3.731	2	male
1.245	29	male
0.963	76	male
0.733	104	female
2.529	173	female
0.759	240	female
1.012	241	female
0.6	242	female
0.036	243	female
1.966	272	female
1.265	280	male

### III. IMPLEMENTATION

#### A. Data Cube Design

Transform a sample detain three dimensional Data Cube which are represent in dimensions and figures. Dimensions are represented by the AREA\_IN\_HECTARE, NUMBER\_KHATA\_KHATAUNI and ACCOUNT\_HOLDER. These are the selected parameter of OLAP database for designing the data cube. Shown in table 2. These sample data are creating an olap data cube.

A sample data of the data cube are shown by the parameter. And the data cube attributes are shown in figure 1. Data cube can be used to access the informational data from the Khasra khatauni database. The main objective of this research is to access the informational data quickly and easily from the data cube. Establish the strong relationship among attributes. It refers to decision support system for retrieve the data. Queries have been performed for fetching the required information.

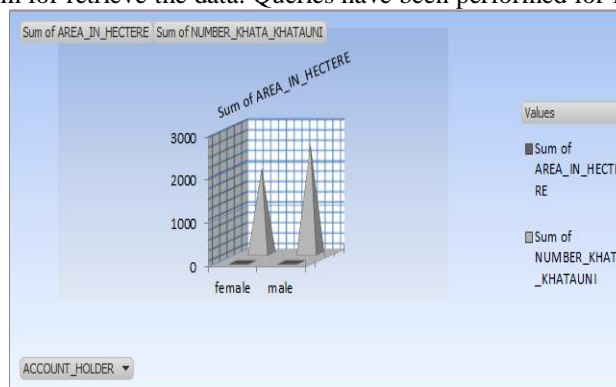


Fig. 1 View of Sample OLAP Data Cube

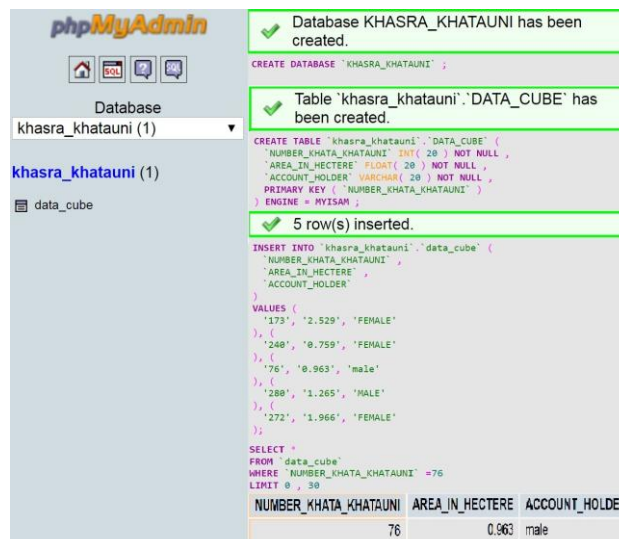


Fig. 2 Data Cube Queries

**B. Correlation Values**

The sample data of Khasra khatauni is shown below in the following tables.

Table 3: Sample Table from Khasra khatauni

Area	khata No.	Gender	Acc. Holder No. (A)	Khata No. (B)	(AB)	A <sup>2</sup>	B <sup>2</sup>
2.529	173	FEMALE	2	1	2	4	1
0.759	240	FEMALE	2	2	4	4	4
1.012	241	FEMALE	2	3	6	4	9
0.6	242	FEMALE	2	4	8	4	16
0.733	104	FEMALE	2	5	10	4	25
0.089	178	FEMALE	2	6	12	4	36
1.966	272	FEMALE	2	7	14	4	49
0.036	243	FEMALE	2	8	16	4	64
0.218	8	FEMALE	2	9	18	4	81
0.036	248	FEMALE	2	10	20	4	100
0.963	76	MALE	1	11	11	1	121
1.245	29	MALE	1	12	12	1	144
1.265	280	MALE	1	13	13	1	169
3.731	2	MALE	1	14	14	1	196
0.367	234	MALE	1	15	15	1	225
2.035	7	MALE	1	16	16	1	256
0.74	172	MALE	1	17	17	1	289
2.868	176	MALE	1	18	18	1	324
0.443	3	MALE	1	19	19	1	361
0.49	4	MALE	1	20	20	1	400
0.202	5	MALE	1	21	21	1	441
0.253	6	MALE	1	22	22	1	484
0.114	9	MALE	1	23	23	1	529
0.052	10	MALE	1	24	24	1	576
0.31	11	MALE	1	25	25	1	625
0.01	12	MALE	1	26	26	1	676
0.589	43	MALE	1	27	27	1	729
1.763	171	MALE	1	28	28	1	784
1.957	174	MALE	1	29	29	1	841
1.67	175	MALE	1	30	30	1	900
0.152	177	MALE	1	31	31	1	961
0.258	235	MALE	1	32	32	1	1024
0.413	236	MALE	1	33	33	1	1089
0.2205	237	MALE	1	34	34	1	1156
<b>30.0885</b>	<b>4433</b>	<b>Gender</b>	<b>44</b>	<b>595</b>	<b>650</b>	<b>64</b>	<b>13685</b>

To identify the correlation values of the data of Khasra khatauni. And put the values of A and B in the equation (1) and calculate the values of Lc (correlation coefficient).

**Coefficient Correlation (Lc) = 0.789542**

**IV. CONCLUSIONS**

From the above work, it can be analyzed that the OLAP cube technology is versatile to store huge amount of the data, design cube based olap database, implementation and can search complex queries & records with in a part of second. Data cube can store huge amount of information data of Khasra khatauni which can be accessed by administrator /user. Data cube are propose the enhance model for storing large capacity of the data. In which the data are in numeric, text, audio /video form. In this research we observe the data in the reference of gender. In which the ratio of male & female account holder are compared using olap data cube technique. The future scope is to increase the women empowerment in society through the analysis of present work.

**REFERENCES**

- [1] Dr. Harsh Dev., Suman Kumar Mishra, “Design of Data Cube and Mining for Online Banking System” International Journal of Computer Application, Volume 30, Issue no. 3, September 2011, ISSN (0975-8887).
- [2] Narander Kumar, Vishal Verma, Vipin Saxena, “Data Cube Representation for Vehicle Insurance Policy System” International Journal of Computer Application, Volume 58, Issue no. 1, November 2012, ISSN (0975-8887).
- [3] Stefano Rizzi, Alberto Abello, Jens Lechtenborger, Juan Trujillo, “Research in Data Warehouse Modeling and Design: Dead or Alive?” Proceeding DOLAP ‘06’ proceeding of the 9th ACM international workshop on datawarehousing and OLAP page 3-10, 2006, ISBN: 1-59593-530-4.

- [4] kawhale Rohit Kumar, Sarita Patil, “Data cube Materialization Using Map Reduce” International Journal of Innovative Research in Computer and Communication Engineering, Volume 2, Issue no. 11, November 2014, ISSN (2320-9801).
- [5] Narander Kumar, Vishal Verma, Vipin Saxena, “OLAP cube representation for Object-Oriented Database” International Journal of Software Engineering & Applications, Volume 53, Issue no. 2, March 2012.
- [6] Ade G. Ola, Vishal Verma, Xue Bai, “Modeling of Student’s Profile and Learning Chronicle with Data cube” Research in High Education Journal, Volume 24, Issue no. 1, and August 2014.
- [7] Shi-Ming Huang, Tung-Hsiang Chou, Jia-Lang Seng, “Data Warehouse Enhancement: A Semantic Cube Model Approach” Journal of information science, Volume 64, 13 December 2006.
- [8] Feng Yu, Wang Shan, “Compressed Data cube for Approximate OLAP Query Processing “Journal of Computer Science & Technology, Volume 17, Issue no.5, September 2012.
- [9] Shichao Zhang, Rifeng Wang, Zhi Jin, “Strategies for Complex Data Cube Queries” Springer Science + Business Media, LLC 2008, Dec1, 2009, DOI: 10.1007/s10489-008-0130-2, ISSN (1573-7497), pp. - 332.
- [10] Dr. S. Vijayarani, Ms. A. Sakila, “Multidimensional Mining Research – An Overview” International Journal of Computer Graphics & Animation, Volume 5, Issue no. 1, January 2015.
- [11] peoxiang Zhao, Xiaolei Li, Dong Xin, Jiawei Han “Graph Cube: On Warehousing and OLAP Multidimensional Networks” DOI: 10.1145/1989323.1989413 Conference: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, Athens, Greece, June 12-16, 2011.
- [12] Bo Wang, Hao Gui, Mark Roantree, Martin F. O’connor2 “Data cube Computational model with Hadoop Map Reduce“, 2010.
- [13] Frank Dehne, Todd Eavis, Andrew Rau-Chaplin, “The cgmCUBE Project: Optimizing
- [14] Parallel data cube generation for ROLAP”, Springer Science + Business Media, Inc. Jan 2006.
- [15] Alfredo Cuzzocrea, Filippo Furfaro, Domenico Sacca, “Enabling OLAP in mobile environments via intelligent data cube compression techniques” Springer Science + Business Media, Inc. 2006, 31 May 2008, ISSN 0925-9902 Online ISSN 1573-7675, October 2009, Volume 33, Issue 2, pp. 95-143.
- [16] Usama Fayyad, Jim Gray at all, “Data cube: A Relational Aggregation Operator Generalizing Group-By, Cross-Tab, And Sub-Totals”, Kluwer Academic, Vol no.9, Issue no.4, November 1996, ISSN (1384-5810) pp.29-53.
- [17] Mohammed J. Zaki, Frank Dehne at all, “Parallelizing the Data Cube”, Journal of data mining and knowledge discovery, Issue no.1, January 2012, ISSN (1384-5810), Vol no. 2.
- [18] Panos Vassiliadis, “Modeling Multidimensional Databases, Cubes and Cube Operations”, General Serretariat of Research and Technology, Volume25, Issue no.2, July 1998, ISSN (1099-3371) pp.14-25.
- [19] Manya Sethi, “Data Warehousing And OLAP Technology” International Journal of Engineering Research and Application, Volume 2, Issue no. 2, April 2012, ISSN (2248-9622) pp.955-960