iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 2 **Issue:** XI **Month of publication:** November 2014

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

A Comprehensive Study on Data Warehouse, OLAP and OLTP Technology

Namrata[1], Dr. Saket Bihari Singh[2]
Research scholar (PhD), Professor
Computer Science, Magadh University, Gaya, Bihar

Abstract: Data warehouse is an integrated, subject oriented, time variant and non-volatile collection of data used for decision making. It is a type of database for decision making which is separately maintained from the operational database. Data warehouse provides a very important tool OLAP (online analytical processing) for more intensive decision support. The main characteristics of OLAP are: advanced database support, analysis of multidimensional data, support client-server architecture and easy to use end user interface. So, data warehouse and OLAP are very important elements for decision making and knowledge discovery process. In this paper, we discussed about data warehouse, OLAP and OLTP and also described data warehouse model and some back end tools.

Keywords: Data Warehouse, OLAP, Data Mining, Knowledge Discovery, Data Warehouse model, OLTP.

I. INTRODUCTION

Data warehouse has been developed to meet particular need, it is a collection of data [1]; data warehouse is differ from traditional databases because it has some unique characteristics that help in decision making [2]. There is no any perfect and single definition of data warehouse so it can be described in many ways [1, 3].

- It is a system used for data analysis and reporting.
- A database that integrate information from more than one source and provides a central view.
- Support data processing by providing a platform of historical data for analysis.

Data warehouse enable executive, manager or analyst (knowledge worker) to make faster and better decision [3], it provides consolidated and summarized data which is more important than individual records and detailed. Data warehouse technology includes: on-line analytical processing (OLAP) [2, 4], data integration, data cleaning and analysis techniques with various functionalities such as consolidation and summarization ability to view data from different angles [3].

A. Design Of Data Warehouse

In this section, we discussed four different views for the design of data warehouse.

1) *Top down view:* This type of view allows selection of the relevant data that is necessary for the data warehouse [5].
2) *Data source view:* Displays the data that is being stored, captured and managed by operational systems [5].
3) *Data warehouse view:* It consists of fact tables and dimension tables.
4) *Business query view:* This view sees the perspectives of data in the data warehouse from the user's point of view.

B. Data Warehouse Back-End Tools

1) *Data extraction:* receive data from heterogeneous and external sources.
2) *Data cleaning:* find errors in the data and remove them [1,3,5].
3) *Data transformation:* change the format of data from host or legacy format to warehouse format [1,5].
4) *Load:* summarize, sort, consolidate, check integrity, compute views and build partitions and indices [1,5].
5) *Refresh:* generate the updates from the external sources to the data warehouse [1].

C. Data warehouse architecture

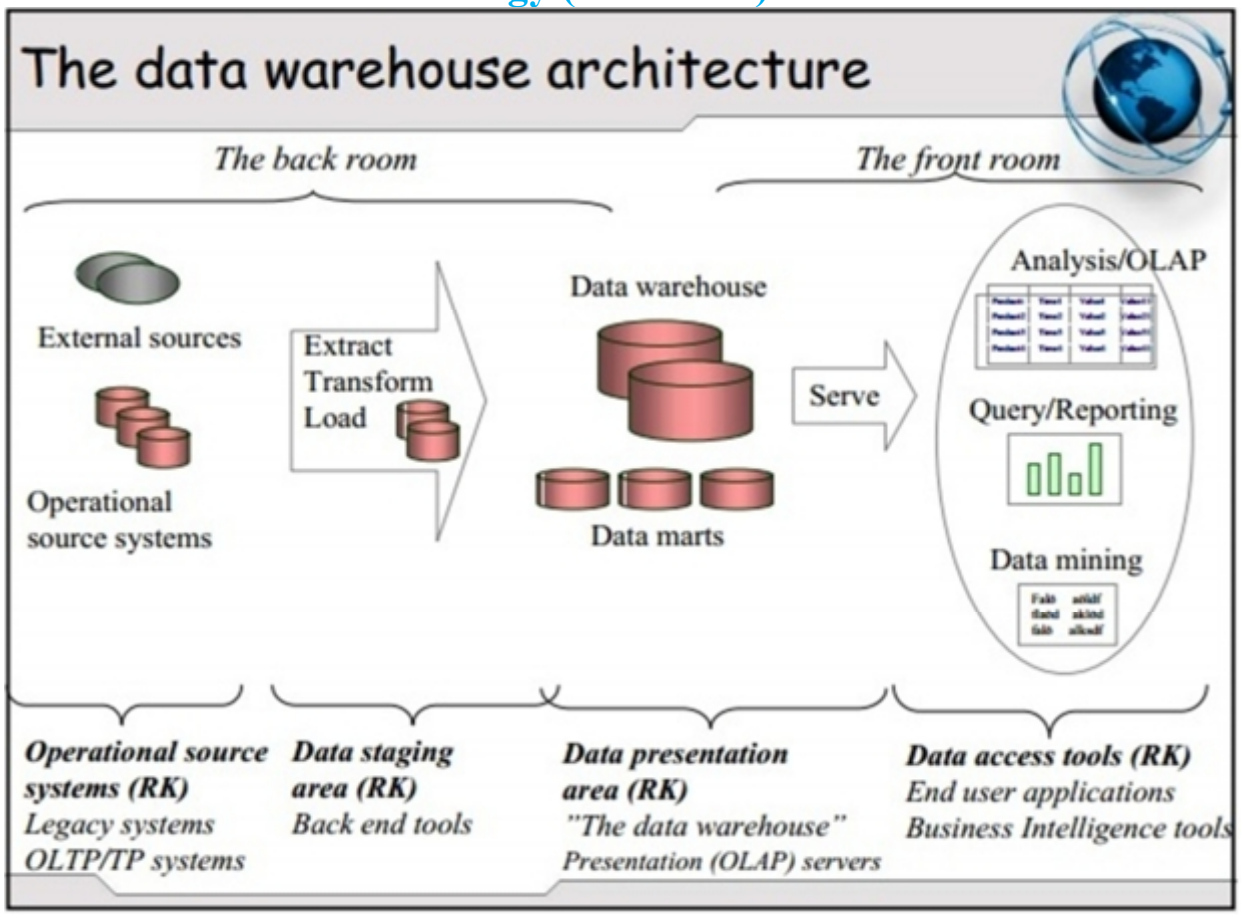


Figure1. Architecture of Data Warehouse [6]

1) Operational source systems:
   a) Maintain small historical data.
   b) Source system is not allowed to query in unexpected and broad ways [6].
   c) High level of availability and performance.
2) Data staging area: This is the most important part of data warehouse architecture [6].
   It includes:
   a) Extraction: It means reading the source information and copy the data which is important for the data warehouse into data staging area for further modification [5, 6].
   b) Transformation: conversion of data from host format to warehouse format by specifying transformation rule. Integrate data from multiple external sources [1,5,6].
3) The presentation area is referred as a sequence of combined data marts. It is a flexible set of data which is possible to extract from operational source. Data mart is a dimensional model for data warehouse. Each data marts contain 5-15 dimensional tables [6]. The data available in the presentation area should be stored, presented and accessed in a particular model (dimensional model).
4) Data access tool is end user application where different types of tasks are performed: Analysis/OLAP, Query/Reporting tools and Data Mining [5,6].

D. Characteristics of data warehouse

These are some distinctive characteristics of data warehouse [1,3,5,6]:

1) Multi user support
2) Multidimensional view

3) Unrestricted cross_dimensional operation
4) Support client-server architecture
5) Accessibility
6) Transparency
7) Consistent throughput(reporting performance)
8) Unlimited aggregation and dimension level

E. OLAP (On-Line Analytical Processing)

OLAP tool is provided by data warehouse for more intensive decision support system where data is stored in multidimensional model [1,2,3]. It is basically used to describe the analysis of more complex information from data warehouse [7]. There are four main characteristics of OLAP technique, these are:

- Support advanced database
- Support analysis of multidimensional data
- Support client-server platform
- Easy to use end-user interface

OLAP provides summarized data and perform rich computation; On-Line Analytical Processing is generally characterized by low volume of transactions [3,7]. This tool is widely used by data-mining technique.

1) *Functionalities of On-Line Analytical Processing (OLAP) tools:*
   a) Drill_down: decreasing the aggregation level [3,4].
   b) Drill_up or Consolidation: increasing the aggregation level.
   c) Pivoting: changing of row and column from rows to columns and columns to rows [5].
   d) Ranking: sorting.
   e) Slicing and dicing: able to see database from different angles [3,5].
2) *OLAP Client/Server Architecture:*

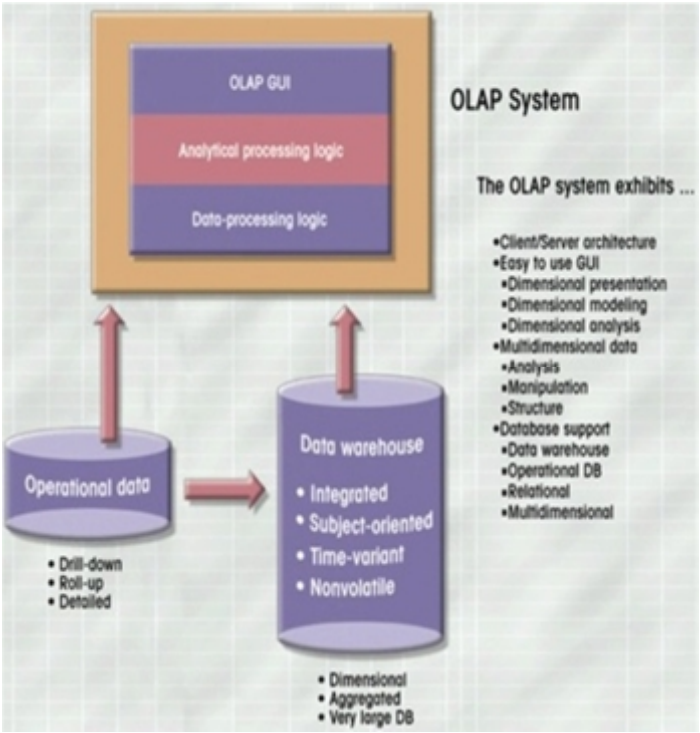


Figure2. OLAP Client/Server Architecture [5]

Traditional database uses OLTP (On-Line Transaction Processing) technique which performs insert, update and delete operation. This technique is used by airlines, bank and other business organization where end user wants to access all information online. The main difference between OLAP and OLTP is that the OLAP tool helps in analyzing of data [7], while OLTP tool gives source information to data warehouse. The processing of OLTP is faster than OLAP and also maintains integrity where multiple users access the database [3,7,8].

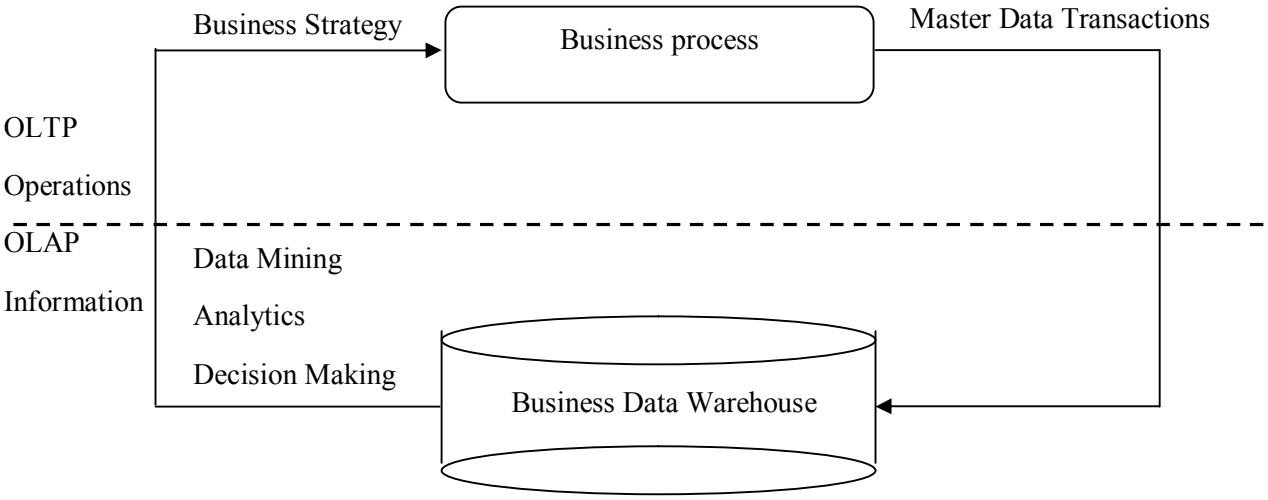


Figure3. Working of OLAP and OLTP

3) OLAP vs OLTP:

	OLAP	**OLTP**
Users	Knowledge worker	IT Professional, Clerk etc.
Database Design	Subject oriented	Application oriented
Function	Decision support	Day-to-day operation
Data	Summarized, consolidated, multidimensional	Updated detailed, current data, flat relational isolated
Access	Lots of scans	Read/write, index on primary key
Usage	Ad-hoc	Repetitive
#users	100	1000
Database size	100Gb-TB	100MB-GB
Metric	Response, query throughput	Transaction throughput

Table 1: On-Line Analytical Processing vs. On-Line Transaction processing

F. Result/Outputof OLAP Tool

1) Graphical Result/Output of OLAP Tool:

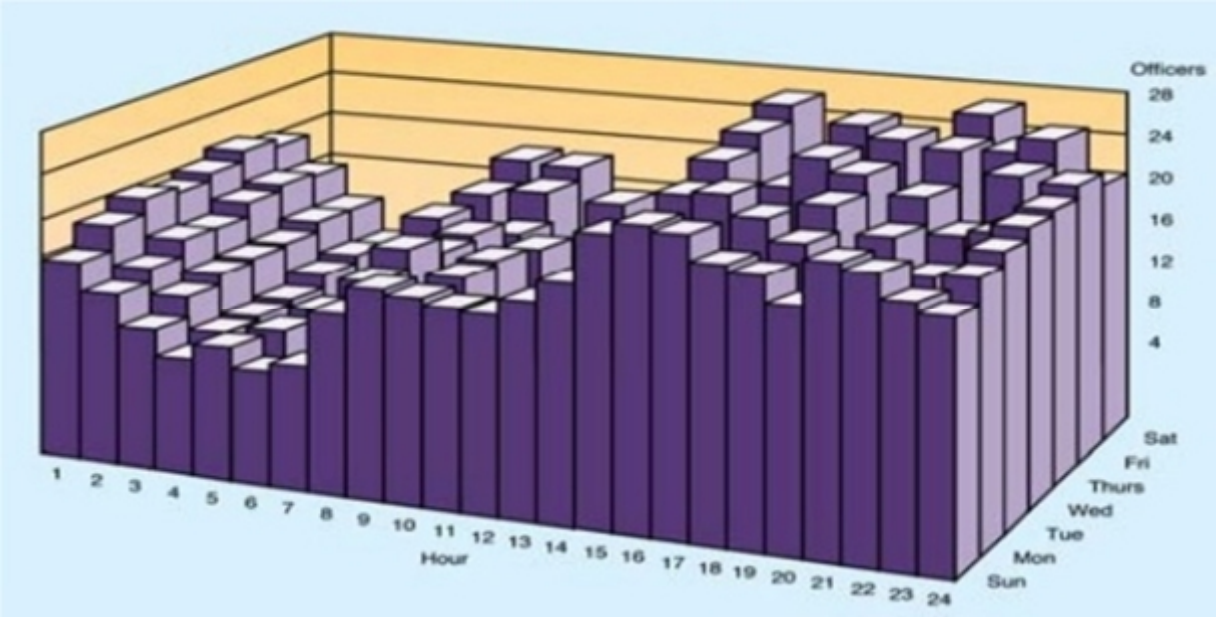


Figure4. Graphical Result of OLAP tool [6]

2) Spreadsheet Result/Output of OLAP tool:

Product month office
Product group quarter region

Column headers (join constraints)

Column headers (Application constraints)

Answer set representing focal event

Product Group	Region	First quarter- 2013
A	ABC	1245
A	XYZ	34534
B	ABC	45543
B	XYZ	34533

Row headers

II. CONCLUSION

This paper presents the complete overview of data warehouse, OLAP and OLTP technique. In this paper, we discussed on different parameters of data warehouse and OLAP such as: Architecture, Characteristics, Function and back-end tool used in data warehouse technique. Data warehouse provides two different types of tools known as OLAP and OLTP, OLTP is used by traditional database whereas OLAP is used for decision making. The database of OLAP is subject oriented whereas database of OLTP is application oriented. From the above comprehensive study we found that query processing of OLTP is faster than OLAP but OLAP is beneficial in complex query processing.

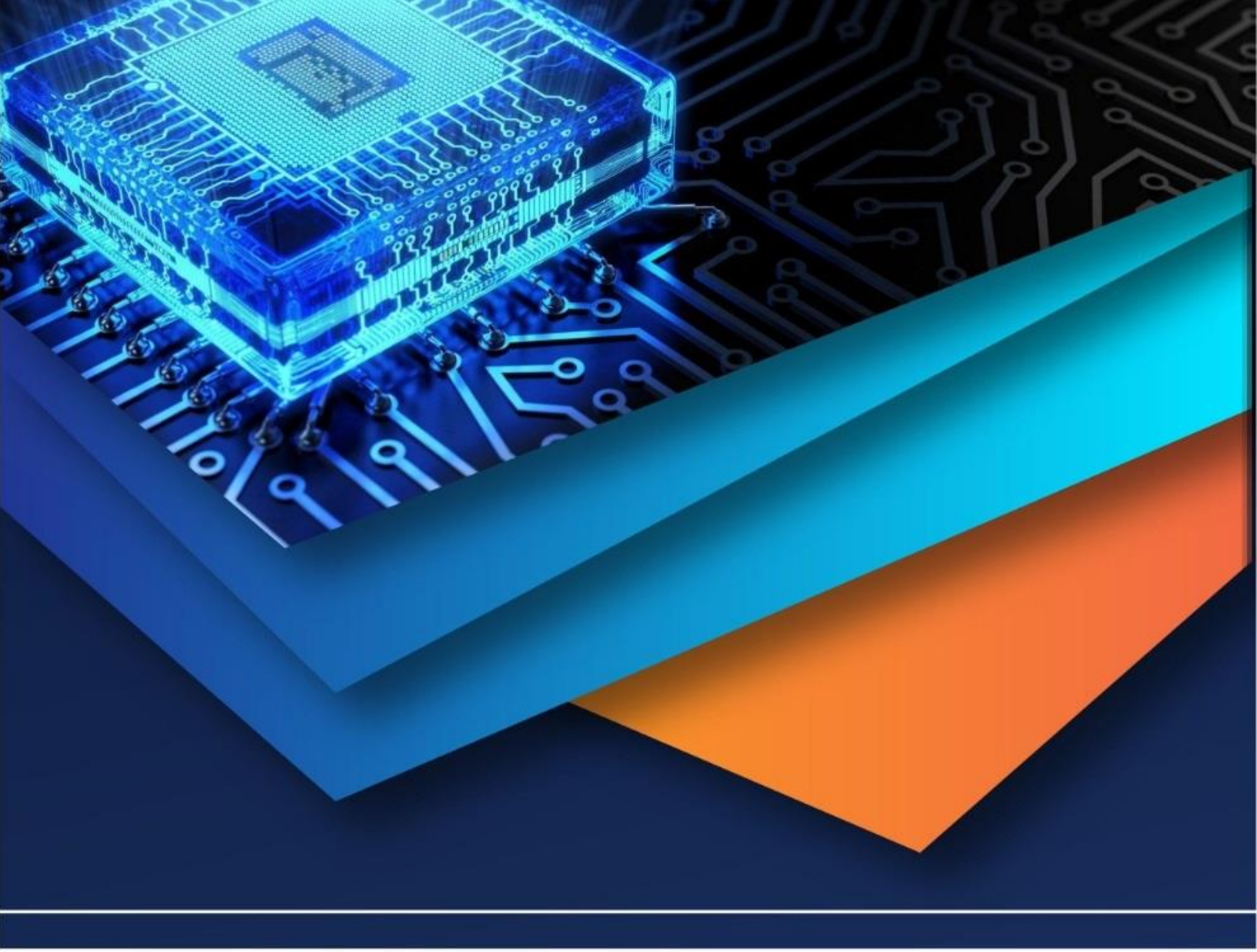
III. ACKNOWLEDGEMENT

I would like to express my deepest appreciation to all those who provided me the possibility to complete this report. I would like to thank My Guide Mr. Saket Bihari Singh, my family and A special gratitude I give to My friends Md Asif Mushtaque and Mr Atuar

Rahman, whose contribution in stimulating suggestions and encouragement, helped me to coordinate my project especially in writing this report.

REFERENCES

[1] Surajit Chaudhuri and Umeshwar Dayal, "An Overview of Data Warehousing and OLAP Technology", Appears in ACM Sigmod Record, March 1997.
[2] Wided Oueslati and and Jalel Akaichi, "A survey on data warehouse Evolution", International Journal of Database Management Systems (IJDMS), Vol.2, No.4, November 2010.
[3] P.Kavitha and Dr.G.N.K.Suresh babu, "A Survey of Data Warehouse and OLAP Technology", International Journal of Latest Trends in Engineering and Technology (IJLTET), Vol. 3 Issue 1 September 2013.
[4] Arun Sen and Atish P. Sinha, "A Comparison of Data Warehousing Methodologies", communications of the ACM March 2005/Vol. 48, No. 3.
[5] "Data Mining: Concepts and Techniques.pptx", November 5, 2014.
[6] Erik Perjons, "DW Architecture and Lifecycle.pdf".
[7] Oscar Romero and Alberto Abelló, "A survey of Multidimensional Modeling Methodologies", International Journal of Data Warehousing & Mining, 5(2), 1-23, April-June 2009.
[8] Hasso Plattner, "A Common Database Approach for OLTP and OLAP Using an In-Memory Column Database", ACM 978-1-60558-551-2/09/06, SIGMOD'09, June 29–July 2, 2009.
[9] Benedikt K¨ampgen, Sean O'Riain and Andreas Harth, "Interacting with Statistical Linked Data via OLAP Operations", Proceedings of Interacting with Linked Data (ILD 2012), 9th Extended Semantic Web Conference, Heraklion, Greece, 28-05-2012.
[10] Charu C. Aggarwal and Philip S. Yu, "A Survey of Uncertain Data Algorithms and Applications", IEEE transactions on knowledge and data engineering, vol. 21, no. 5, may 2009.
[11] Pushpal Desai and Desai Apurva, "The study on Data Warehouse and Data Mining for Birth registration System", International Journal of Computer Applications, 2011.
[12] pushpal Desai, "A survey of Data Warehouse implementation for E-Governance systems", Indian Journal of Applied research, Vol. 4 Issue-7, July-2014.
[13] Meenakshi Arora and Anjana Gosain, "Schema Evaluation for Data Warehouse: A Survey", International Journal of Computer Applications, 2011.
[14] Vina Lomte and Saloni Shah, ""The Survey paper on Importance of Integration of Knowledge Management and Bussiness Intelligence", International Journal of Science and Research, Vol-3, Issue-6, June-2014.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY