# Implementation of Data Warehouse Technology in Academic Data Management

Ubash Yadav[1]
Nepal Open University
ubash@nou.edu.np

Hira Lal Bhandari[2]
Nepal Open University
drw.moon@gmail.com

Sushil Poudel[3]
Nepal Open University
sushil@nou.edu.np

**Abstract:-** In the current time of digital flux of data, existing techniques of data warehouse cannot handle the complexity and analyzing processes of getting meaningful information. Extending an existing data warehouse solution requires the adoption of big data technology. Academic institutions have already with plenty of data, so with the day by day incremental academic data in the institutions comes with the need of big data technology. Academia requires top level management of data so that the tools are used to extract the information from records. Those information obtained support in the decision-making process at the top level management. In this article, we explore how data ware house technology has been implemented to manage academic databases and finds out the best one techniques used for data ware house. At the end of the paper, we suggest the best implementation techniques used in the warehouse in this current time.

**Keywords:-** *Academic Data , Data Warehouse, RTDW, ETL.*

## I. INTRODUCTION

The term data warehousing refers to the practice of storing corporate data in a second location, usually remote from existing production systems. This site will be a place where you can report requests and make inquiries. This data is similar to real operational data [1]. A data warehouse is always a physically separate data store, transformed from application data in production. The data warehouse is physically separate from the application data, but both can be on the same computer.

Data warehousing is the process of building and using a data warehouse. Data warehouses are created by integrating data from multiple disparate sources to support analytical reporting, structured and ad-hoc queries, and decision-making. Data warehousing includes data cleansing, data integration, and data integration.

Data warehouses are optimized for analyzing large amounts of data rather than the performance speed of individual transactions. This is the basic approach you can use to meet your organization's reporting obligations and query statistical analysis [2].

Academia operates in gradually complex and competitive environment. To stay up with national and international economic, political, and social developments, it must compete with other institutions. Additionally, a number of interest groups anticipate the university to quickly come up with answers that are appropriate for their demands. Higher education must assess the large data sources produced to overcome this situation in order to make the appropriate judgments required to handle these quick changes. The majority of colleges spend a lot of money on information technology.

### A. Data warehouse components

With the help of the diagram [3] below, we will learn more about the architecture and components of a data warehouse:
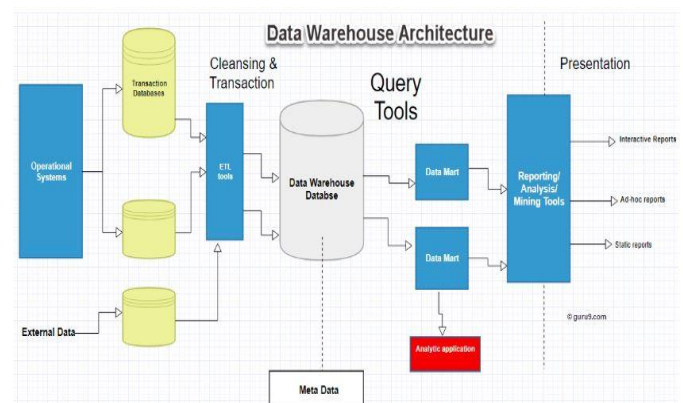


Fig 1 Data Warehouse Architecture

A data warehouse is based on an RDBMS server, a central information store surrounded by several key data warehouse components that make the entire environment functional, manageable, and accessible.

### B. Data Warehouse Models

➤ *Enterprise warehouse [4]*

An Enterprise Data Warehouse (EDW) is a relational data warehouse that contains a company's business data, including information about its customers. EDW enables data analysis that can provide actionable insights. Like all data warehouses, an EDW collects and aggregates data from multiple sources and acts as a repository for most or all enterprise data, enabling broad access and analysis.

➤ *Data Mart [5]*

A data mart is a subset of a data warehouse focused on a specific business area, department, or subject area. Data marts make specific data available to defined groups of users so that they can quickly access key insights without wasting time searching through your data warehouse, will be For example, many companies may have data marts that serve specific departments within the company, such as finance, sales, and marketing.

➢ *Virtual warehouse [6]*

The virtual warehouse is another term for the data warehouse. A data warehouse is a computer tool designed to simplify decision-making in business management. Collect and display point-in-time business data to create a snapshot of the state of your business at that point in time. Virtual warehouses often collect data from various sources.

*C. Statement of the Problem*

This research explores different hurdles and difficulties faced during data management using data warehouse management in the academic institutions on the literature review basis. Data management in the academia is become major problems in this current time of digital data flux which are associated with academic purpose. This research study has the major focus on why data generated from academic institutions are raising problems in problem in education.

➢ *Other problem areas:*

Mistakes in information can prompt detailing issues. The mail may not reach its destination if the address information is incorrect.

*D. Objectives of the research*

This paper has performed review of different recently published research papers. The paper has explored about data ware house and associated technologies for better understanding the concept regarding storage techniques with ware house concept. In this way, this paper has the following research objective:

- To find out techniques used in the data ware house technology in academic data management.
- To understand the importance of data ware house technology in the academic context of academic data management.

## II. REVIEW OF LITERATURES

This paper reviews previous research articles, conference papers and dissertations. This work performs study on the review bases. Previously published research papers and articles are sources of study. We explore, discuss and analyze the previous studies so that major techniques used in data warehousing are traced. This way we find the best techniques used in the data warehousing.

Data warehouse technology, most popular deep learning methods proposed by Bo Zhao[1,2] and Yanjin Liu[3] in [7] i.e. HWPCNNs are used to allow optimal outcomes in the management of physical education quality. The quality is better enhanced with the comparison to benchmarks and then redefined for higher levels of precision.

The paper [8] by Leo Willyanto Santoso* and Yulia have proposed Hadoop as big data analytic tools to be implemented for data ingestion/ staging. This paper has proposed the system so as to handle the unstructured data which is not possible with traditional data warehouse technology using extract load and transfer (ETL) process. The system has been improved to handle several data quality issues such as duplicate data, inconsistent data and garbage data. Hadoop has been implemented to enhance relational database management system (RDBMS) as data ingestion/ staging tool as well as for data management and data presentation platform.

The paper [1] by A F Neamah also has proposed Hadoop as big data analytic tools to be implemented for data ingestion/ staging. This paper has proposed the system so as to handle the unstructured data which is not possible with traditional data warehouse technology using extract load and transfer (ETL) process. The system has been improved to handle several data quality issues such as duplicate data, inconsistent data and garbage data. Hadoop has been implemented to enhance relational database management system (RDBMS) as data ingestion/ staging tool as well as for data management and data presentation platform.

The research article [9] by Sonia Ordoñez Salinas and Alba Consuelo Nieto Lemus have proposed the model of integrating data warehouse and big data techniques to better enhance the performance of traditional data warehouse. This paper also uses the NoSQL database technology to overcome the lacks of SQL database. We present a revision of the state-of-the-art in integration proposals from the point of view of the purpose, methodology, architecture and underlying technology, highlighting the common elements that support both technologies that may serve as a starting point for full integration and we propose a proposal of integration between the two technologies.

## III. METHODOLOGY

The major work of the study is to find out the best techniques in data warehouse to manage academic data. Literature review is the main method of performing comparative study to find out the most preferred technique from recently used data warehouse techniques.

➢ *Data Description*

Data used in this research are papers, article, conference paper and digital mediums such as website and different survey portals. Sources used to perform the research are recently published materials and the survey.

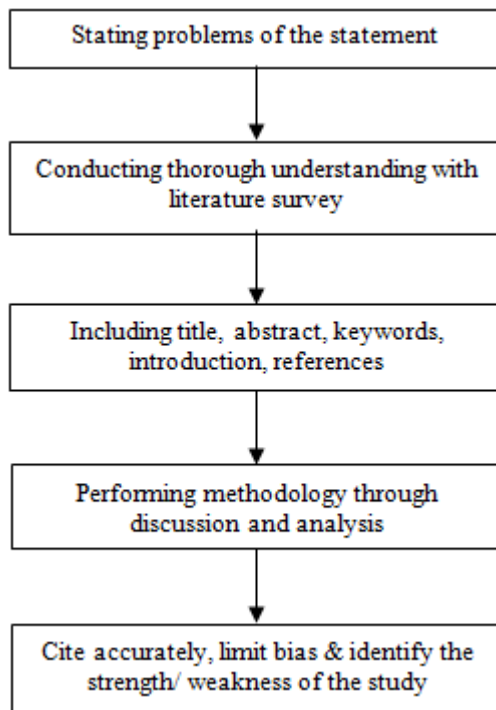The method includes different stages for performing the research task that has been presented below [10]:

Fig 1 Data Warehouse Architecture

The paper basically contains five stages to perform the research task. Firstly introduction section briefly talks about data warehouse, its importance and pinpoints the problems indicated in the different studies. In the Review of literature section, we present the reviews of relevant papers so as to enhance the differences between the techniques used in the data warehouse technology. In the abstract, we present the summary form of the article and focus the task performed. And the major section of the article is methodology which talks about how research has been performed to find out the facts and get the result of the work.

## IV. DISCUSSION

Data warehousing can play the vital role in making decision in every institutions. The data generated day by day is very complex so it requires better improvement and assessment of technology in managing several forms of data in the form of structured, semi-structure and unstructured. Here, in this research we have performed the review of literatures to find out recently used techniques in proper data management in the academia. The importance of exploration and analysis of data has significantly raised the importance in decision making process for the institutions.

The process of decision making is being so much complicated in context of huge production of data day by day. This has been the very complicated because of its

## V. CONCLUSION

After the study of previously performed researches, we come to the thorough understanding the concept of data warehouse and its importance in this huge generation of academic data in several forms like structured, unstructured and semi-structured data. Data warehousing is the most important in decision making process so as to direct the institution on right track for future advancement. Majorly, we have found out different four techniques used in data warehousing to management academic data. They are:

- Deep learning methods HWPCNNs
- Big data tool i.e. Hadoop
- Integration of data warehouse and big data techniques

This way, we come to the conclusion that traditional data warehousing techniques are insufficient in handling real time data and as well as heterogeneous forms of data generated in academic institutions to better achieve the efficacy in managing those data.

## REFERENCES

[1]. A. F. Neamah, "Adoption of Data Warehouse in University Management: Wasit University Case Study," *J. Phys. Conf. Ser.*, vol. 1860, no. 1, 2021, doi: 10.1088/1742-6596/1860/1/012027.

[2]. J. Kembro and A. Norrman, "The transformation from manual to smart warehousing: an exploratory study with Swedish retailers," *Int. J. Logist. Manag.*, vol. 33, no. 5, pp. 107–135, 2022, doi: 10.1108/IJLM-11-2021-0525.

[3]. "Data Warehouse Architecture, Components & Diagram Concepts." https://www.guru99.com/data-warehouse-architecture.html (accessed Jun. 15, 2023).

[4]. "What is an Enterprise Data Warehouse (EDW)? | Snowflake." https://www.snowflake.com/guides/what-enterprise-data-warehouse (accessed Jun. 15, 2023).

[5]. "What is a Data Mart | IBM." https://www.ibm.com/topics/data-mart (accessed Jun. 15, 2023).

[6]. "What Is a Virtual Warehouse? | Techwalla." https://www.techwalla.com/articles/what-is-a-virtual-warehouse (accessed Jun. 15, 2023).

[7]. B. Zhao and Y. Liu, "Application of Data Warehouse Technology Based on Neural Network in Physical Education Quality Management," *Math. Probl. Eng.*, vol. 2022, 2022, doi: 10.1155/2022/4456885.

[8]. L. W. Santoso and Yulia, "Data Warehouse with Big Data Technology for Higher Education," *Procedia Comput. Sci.*, vol. 124, pp. 93–99, 2017, doi: 10.1016/j.procs.2017.12.134.

[9]. F. Engineering, "D w b d," vol. 9, no. 2, pp. 1–17, 2017, doi: 10.5121/ijcsit.2017.9201.

[10]. "How to Write a Scientific Review Article - Enago Academy." https://www.enago.com/academy/how-to-write-a-scientific-review-article/ (accessed Jun. 14, 2023).