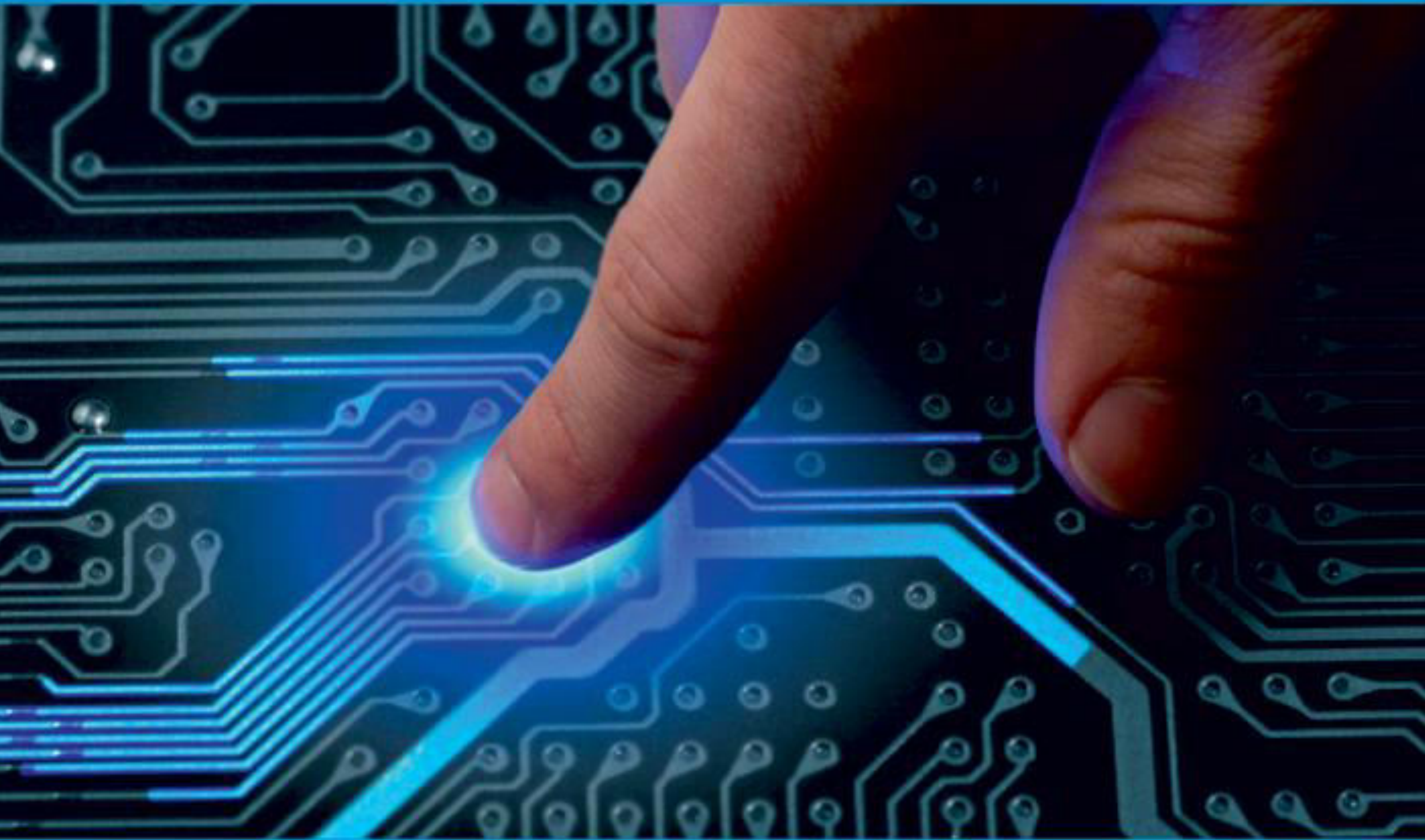
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 12, December 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.625

© 2024 IJIRCCE | Volume 12, Issue 12, December 2024| DOI: 10.15680/IJIRCCE.2024.1212102

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data-Driven Business Intelligence: A Comprehensive Review of Data Mining and Data Warehousing

Sandhyarani Ramakant Shinde, Prof. Mote Ashwini Gopinath

Research Scholar, Department of Computer Science & Engineering, (Master of Technology in Computer Science & Engineering), Vidya Vikas Pratishthan Institute of Engineering and Technology, Solapur, Maharashtra, India

Associate Professor, Department of Computer Science & Engineering, Vidya Vikas Pratishthan Institute of Engineering and Technology, Solapur, Maharashtra, India

ABSTRACT: A comprehensive review of literature on data warehouse implementations reveals a significant shift in expert opinions over time. Notably, fewer authors now emphasize the importance of a clear business purpose before embarking on a data warehouse project. Data warehouses (DW) are crucial to Business Intelligence (BI) applications, integrating disparate data sources, primarily from transnational databases. However, recent BI studies indicate that data is no longer confined to structured formats or databases. Instead, it can also be sourced from unstructured sources, enhancing decision-makers' analysis. Effective management of this diverse data is vital for informed decision-making. This paper's case study review supports the concept of strategic alignment. Moreover, mapping the experiences to the configuration school of strategic management provides valuable insights into the degree of success achieved.

KEY WORDS: Business Intelligence (BI), Data Warehousing, Data Analysis, Transnational data, Data Mining, Data Warehouse, OLAP, OLTP, Data Warehouse, Data Mining, Vendors, System Design, Environment, Data Marts, Data Knowledge Discovery, Staging Layer.

I. INTRODUCTION

In late 1997, the author initiated a research study to establish "Best Practices for Implementing a Data Warehouse." This study aimed to elucidate the experiences of a bank's data warehouse project, which ultimately resulted in implementation failure. Data warehousing is defined as the process of gathering and storing data in a managed database, characterized by subject-oriented and integrated data, time variance, and no volatility, ultimately supporting decision-making (Inman, 1993) Data warehouses enable online analytical processing (OLAP) capabilities, which differ significantly from online transaction processing (OLTP) requirements typically supported by operational databases. Data Mining plays a crucial role in extracting valuable insights from vast amounts of data. It involves the study of algorithms and computational paradigms that enable computers to identify patterns, make predictions, and improve performance through data interaction. Machine learning, a closely related field, focuses on developing computer systems that can enhance their performance in a specific domain through experience. Data warehousing and data mining are essential components of modern businesses, sharing similarities in concepts and systems. Both play critical roles in driving organizational value by leveraging data in diverse ways.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

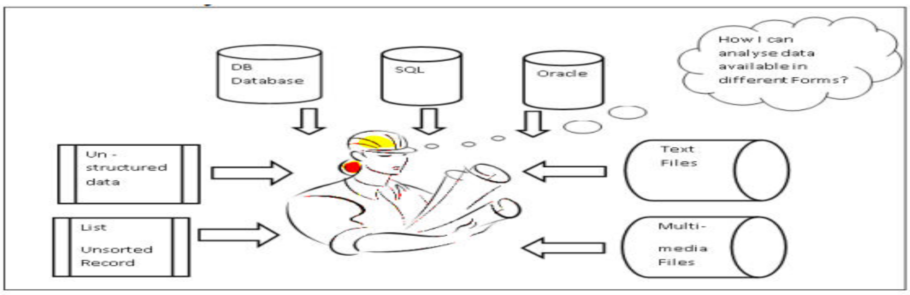


Figure 1: Problem in Decision Making

II. LITERATURE REVIEW

A data warehouse is a specialized database application system designed to consolidate and analyze data from multiple sources. Its effectiveness is directly tied to the quality of data input, and extracting valuable insights depends on the accuracy and relevance of the data stored within. Data Warehousing is a flow process used to gather and handle structured and unstructured data from multiple sources into a centralized repository to operate actionable business decisions. With all of your data in one place, it becomes easier to perform analysis, reporting and discover meaningful insights at completely different combination levels.

A data warehouse setting includes extraction, transformation, and loading (ETL) resolution, an online analytical processing (OLAP) engine, consumer analysis tools, and different applications that manage the method of gathering data and delivering it to business. The term data warehouse life-cycle is used to indicate the steps a data warehouse system goes through between when it is built.

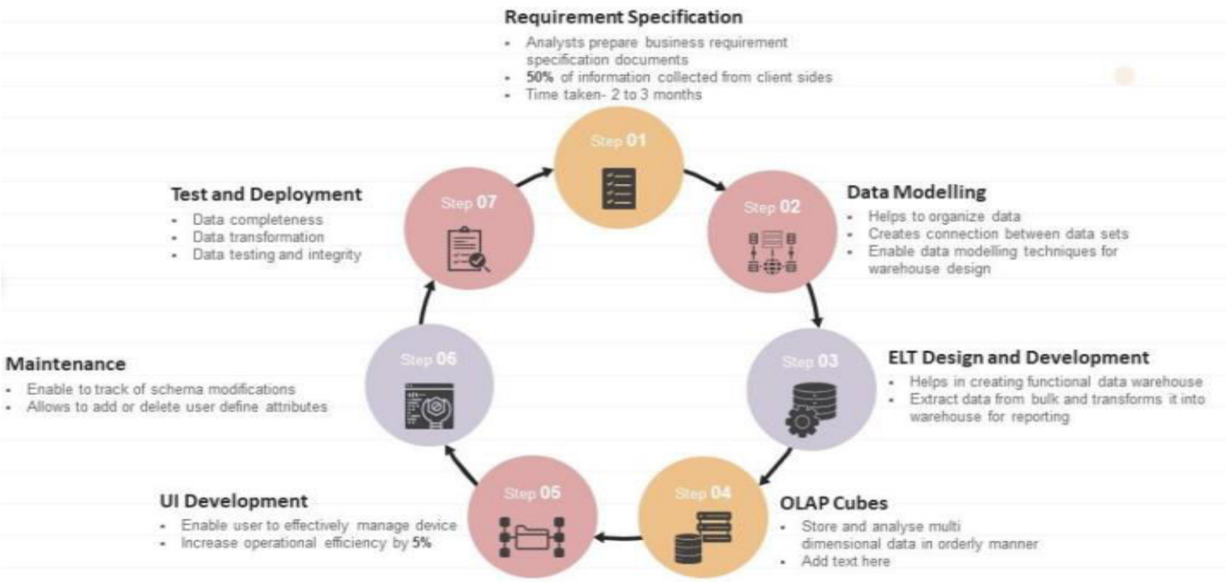


Figure 2: Data Warehouse Life Cycle

III. RELATED WORK

Data warehousing empowers businesses to transform data into actionable insights, enabling informed decision-making and a competitive edge. By modeling and analyzing data in the warehouse, organizations can uncover opportunities to

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

drive growth and improvement. Data in the warehouse is organized by subject, rather than application, ensuring that only relevant information is stored for decision support. Historical data is collected over time, facilitating comparisons, trend analysis, and forecasting. While data is not updated in real-time, it is regularly migrated from operational systems during scheduled maintenance windows to minimize performance impacts. Some key sources of important information about a financial institution include:

1. The customers - what they think, what they want, how they see the bank or the financial institution as a provider of service both materially and psychologically.
2. The employees - what they know, their perceptions about the bank or the financial institution
3. The legacy systems
4. The actual data, information and knowledge that flows through the bank
5. The business environment.

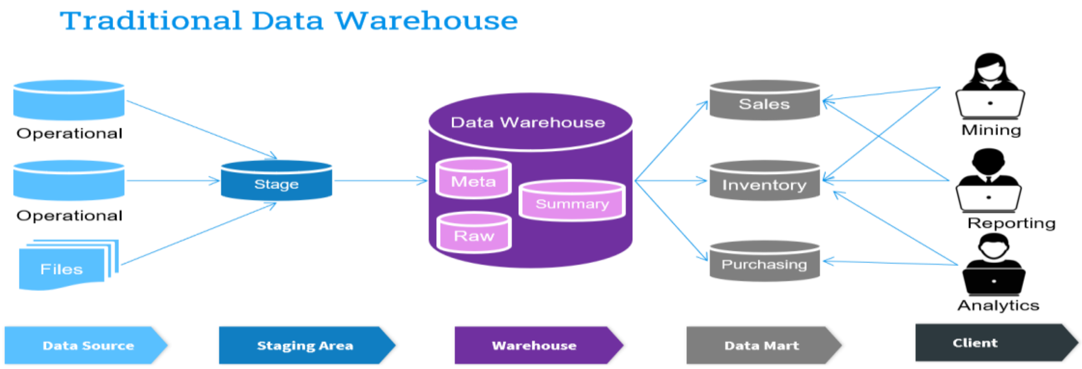


Figure 3: Traditional Data Warehouse

IV. DATA WAREHOUSE

A Data Warehouse (DW) is a centralized repository that stores data from various sources in a single location, making it easier to access and analyze.
data warehouse (DW) should have the following
characteristics:

- Subject-Oriented: Organized around specific business subjects or processes.
- Integrated: Combines data from multiple sources.
- Time-Variant: Stores historical data.
- Non-Volatile: Data is not updated in real-time.

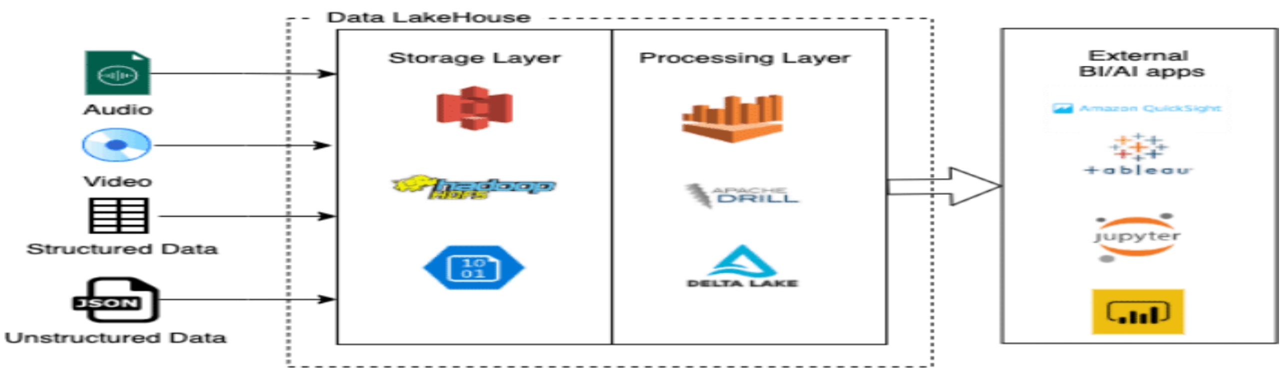


Figure 4: Data Warehouse **architecture**

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The data warehouse architecture provides a structured approach to collecting, transforming, and storing data from multiple sources, enabling organizations to make informed decisions through data analysis.

The major steps for data warehouse implementation are:

a) **Subject definition:** It is determining which subjects will be created and populated in the data warehouse.

b) **Data capture:** The core of data capture is Data Replication, is defined as 'a set of techniques that provides comprehensive support for copying and transforming data from source to target location in a managed, consistent and well-understood manner'.

c) **Data transformation:** It is used to convert and summarize operational data into a consistent, business-oriented format.

d) **Metadata management:** To access to the data warehouse, it is necessary to maintain some form of data, which describes the data warehouse. This data is called metadata. It masks the complexities of the technology of a Data Warehouse from the users. It acts as a critical aid for navigating the data warehouse.

e) Loading the warehouse: This is the periodic loading of static snapshots from the online transaction-processing eenvironment gives the data warehouse its time-variant quality

Benefits of Implementing a Data Warehouse in Banking

- A Data Warehouse enables banks to analyze customer data, transaction history, and market trends to identify potential risks and opportunities.
- A Data Warehouse helps banks identify and prevent fraudulent activities by analyzing transaction data, customer behavior, and market trends.
- A Data Warehouse provides banks with a competitive edge by enabling them to respond quickly to changing market conditions, customer needs, and regulatory requirements.
- A Data Warehouse helps a bank gain valuable customer insights, including:

- Buying habits and patterns
- Financial behavior and preferences

V. DATA MINING

Data Mining (DM) combines database technology and artificial intelligence to extract valuable insights from large datasets, enabling informed decision-making. It serves as a decision support system, helping users make better decisions. The exponential growth of data has created a challenge for humans to manually analyze and extract valuable insights.

This highlights the importance of Data Mining (DM) techniques, which enable the extraction of valuable insights from large datasets. DM is applied across various domains to facilitate informed decision-making and drive business success. For example, DM can be used for marketing purposes. It can help by giving useful information about the best media and time to publish an advertisement which would help to increase the sales of a product.

A. Data Mining Usage

With the exponential growth of data, manual analysis becomes increasingly challenging. This emphasizes the need for Data Mining (DM) techniques to uncover valuable insights.

B. Data Mining Process

The data mining process is complex and iterative, involving feedback loops that require repetition of certain steps. As per figure the process consists of several interconnected steps. Notably, these steps may be repeated, and in some cases, it may be necessary to restart the entire process from the beginning.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

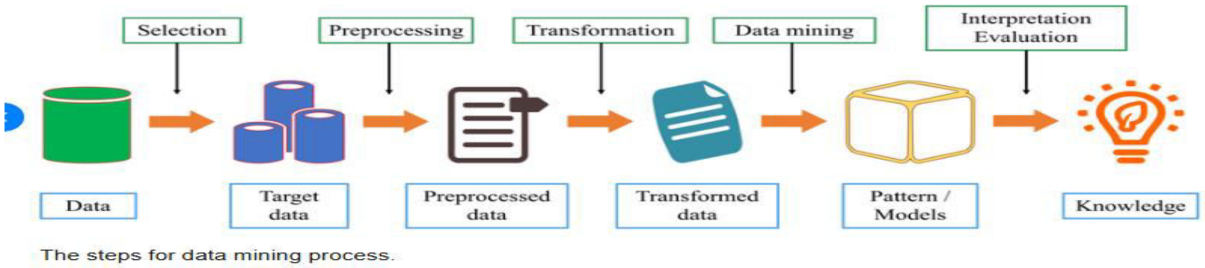


Figure 5: Steps of Data Mining Process

The Data Mining process is a complex, iterative cycle that often involves multiple feedback loops. As users apply various techniques, they may encounter issues such as:
- Poor data quality, requiring additional cleaning or prepossessing
- Ineffective techniques, necessitating the exploration of alternative methods
- Unexpected results, prompting further refinement of the analysis

These feedback loops enable users to refine their approach, ensuring that the insights gained are accurate, reliable, and meaning-full .The data mining process involves six steps:
1) Problem definition
2) Data Preparation
3) Data Exploration
4) Modeling
5) Evaluation
6) Deployment

VI. EXPLANATION/DISCUSSION OF MODEL

1.Goal of Data Mining Process:
The Data Mining process is not a simple function, as it often involves a variety of feedback loops since while applying a particular technique, the user may determine that the selected data is of poor quality or that the applied techniques did not produce the results of the expected quality.

2. Problem Definition:
A data-mining project starts with the understanding of the bbusiness problem. Data mining experts, business experts, and domain experts work closely together to define the project objectives and the requirements from a business perspective. The project objective is then translated into a data mining problem definition. In the problem definition phase, data.

3. Data Investigation:
Domain experts understand the meaning of the metadata. They collect, describe, and explore the data. They also identify quality problems of the data. A frequent exchange with the data mining experts and the business experts from the problem definition phase is vital. In the data exploration phase, traditional data analysis tools, for example, statistics are used.

4. Data Integration:
Domain experts build the data model for the modeling process. They collect, cleanse, and format the data because some of the mining functions accept data only in a certain format. They also create new derived attributes, for example, an average value. In the data preparation phase, data is tweaked multiple times in no prescribed order

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.625| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

5. Data Modeling:

Data mining experts select and apply various mining functions because you can use different mining functions for the same type of data mining problem. Some of the mining functions require specific data types. The data mining experts must assess each model. In the modeling phase, a frequent exchange with the domain experts from the data preparation phase is required.

6. Evaluation:

Data mining experts evaluate the model. If the model does not satisfy their expectations, they go back to the modeling phase and rebuild the model by changing its parameters until optimal values are achieved.

VII. CONCLUSION

In today's data-driven world, the enormous volume of data necessitates the use of data warehousing and data mining. A **data warehouse** serves as a centralized repository for subject-oriented, integrated, time-sensitive, and non-volatile data collected from various sources, such as operational databases. To ensure faster performance, data warehousing organizes this data using a specialized architecture, which includes **fact tables** and **dimension tables**. Unlike the conventional modeling of operational databases, which uses **entity-relationship diagrams**, data warehouses are modeled using **dimensional modeling** techniques, such as **star**, **snowflake**, or **galaxy schemas**. **Data mining** has become an essential tool for extracting valuable insights from large datasets, including those from the internet. This process involves six key phases: **problem definition**, **data preparation**, **data exploration**, **modeling**, **evaluation**, and **deployment**. Data mining is an **iterative process**, meaning feedback loops often occur between these phases, and the entire process may need to be repeated to refine results. These iterations are crucial for providing better answers that support improved decision-making, ultimately helping users make more informed and effective decisions. The iterations are needed in the mining process in order to provide better answers which will be used by the users to make better decisions.

REFERENCES

1. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), 1135-1144.
2. Bhavani, T., Data Mining: Technologies, Techniques, Tools and Trends,
3. Plotnikova, V., Dumas, M., and Milani, F. "Adaptation of data mining methodologies: A systematic literature review." PeerJ Computer Science, 2020.peerj.com
4. Ralph Kimball, the Data Warehouse ETL Toolkit, Wiley India Pvt Ltd., 2006
5. Fang Y. A DSS assistant model for college counselors based on data mining. In: Proceedings - 2017 International Conference on Smart Grid and Electrical Automation, ICSGEA 2017. 2017
6. https://sonra.io/reference-enterprise-data-architecture-the-ultimate-guide-to-architecting-data-warehouse-date lakehouse
7. https://www.geeksforgeeks.org/data-warehouse-development-life-cycle-model/
8. https://www.slideteam.net/data-warehouse-development-lifecycle-model.html
9. https://estuary.dev/traditional-data-warehouse/
10. Xiaoyan ,"Data Mining Based Algorithm for Traffic Network Flow Forecasting" , IEEE, 2003.
11. .Han and M.Kamber, "Data Mining: Concepts and Techniques," Morgan Kaufmann Publishers, San Francisco, CA, 2006. ISBN: 1-55860-489-8
12. Data mining and knowledge discovery: making sense out of data.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE

SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING