International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-10, Issue-1, January 2022
https://doi.org/10.55524/ijircst.2022.10.1.21
Article ID IRP221017, Pages 117-121
www.ijircst.org

Approaches of Data Warehousing and Their Applications: A Review

Mrinal Paliwal [1], and Pankaj Saraswat [2]

[1, 2]Assistant Professor, Department of Computer Science Engineering, Sanskriti University, Mathura, Uttar Pradesh

Correspondence should be addressed to Mrinal Paliwal; mrinalpaliwal.cse@sanskriti.edu.in

ABSTRACT: A data warehouse, DW in short is a huge repository of corporate data that is employed to aid an organization's decision-making. The data warehouse idea has been around throughout eighties, while it was created to assist in the transformation of data from just enabling activities to fueling judgment assistance capabilities that disclose business insight. The huge volume of data in data stores originates from a variety of sources, including interior services like branding, selling, and treasury, customer-facing services, and outsourced systems, besides several. On a scientific basis, a DW gathers data from various apps and platforms on a regular basis; the data is then formatted and imported to match the data currently in the storehouse. This generated content is stored in the DW so that decision makers may access it. The frequency with which data pulls happen, how data is organized, and so on will vary relying on the needs of the company. The procedure of mining data from a basic system or excavating information from a huge quantity of data is known as data warehousing. It is generally known as ETL, which stands for extract, transform, and load. This paper discusses the following topics: an overview of Datawarehouses, different Datawarehouse design approaches and their benefits and drawbacks, different sorts of pulling out techniques in Datawarehouses, characteristics of Datawarehouses, dissimilar doles of data warehousing, unalike components used in DW, and data warehousing usages.

KEYWORDS: Datawarehouse Applications, Datawarehouse Characteristics, Datawarehouse Components, Datawarehouse Design, Datawarehouse, Extraction Methods.

I. INTRODUCTION

Datawarehouses are a useful tool that allows you to store and analyze data, which can help you make better pronouncements in relation with the business. It's also crucial to be certain that the right information is released, and that it's simple to find for those in charge of making choices [1,2]. Data warehousing is a subject-oriented, integrated, time-variant, and non-volatile data gathering that assists management in making decisions. Agile BI development products, Datawarehouse appliances, Big data analytics, In-memory data, I workspaces and dashboards, and collaborative sharing of BI content are examples of new Datawarehouse technologies. Mobile BI includes data federation, as well as mash-ups and complex event processing.

The following hardware and software help to consolidate data from various sources in a Datawarehouse for reporting and analytics:

1. Datawarehouse Appliances and Big Data technologies like Hadoop and Map Reduce
2. CDC (Change Data Capture)
3. Compression
4. Data Reduplication
5. Data Replication
6. EAI (Enterprise Application Integration)
7. ETL (Extract, Transform, Load)

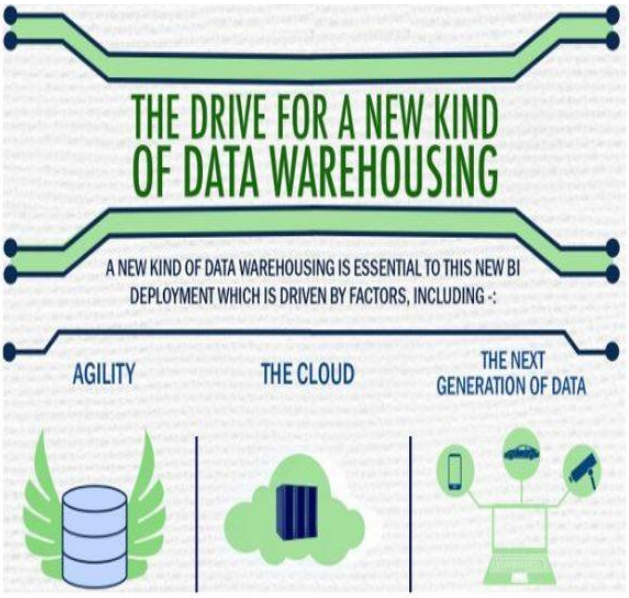

Figure 1: The New Kind of Warehousing[3].

II. DISCUSSION

A. Components Of Datawarehouse

Figure 2 above illustrates the main components of Datawarehouse. It includes- database, architecture, acquisition, sourcing, metadata, access tools, data marts, and information delivery system.

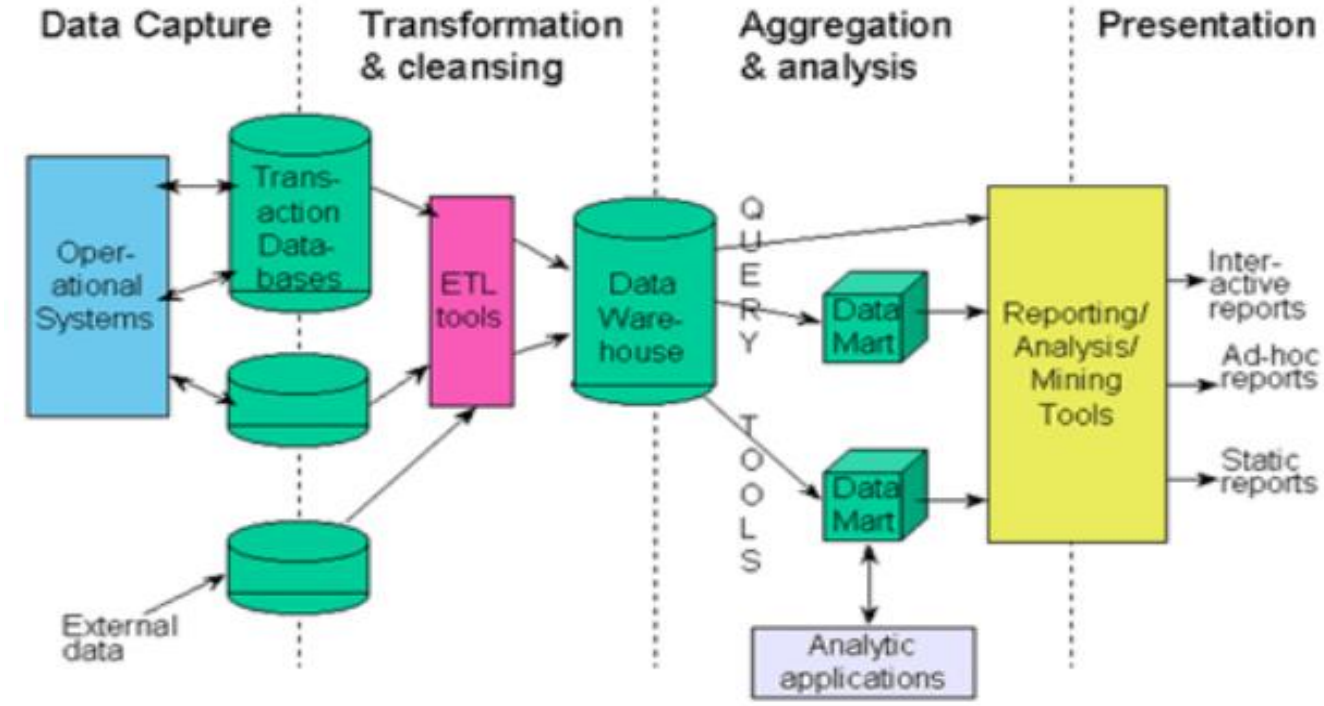

Figure 2: Components of Datawarehouse [4]

B. Datawarehouse Design Approach:

Datawarehouse design methods are a crucial part of the datawarehousing process. The proper Datawarehouse design may help you save a lot of time and money on your project. When building a Datawarehouse, two distinct Datawarehouse Design Approaches are often used, depending on the needs. There are four main methods to constructing a DW milieu. The following architypes are often mentioned to as:

1. Bottom-Up Methodology (BUM)
2. Top-Down Methodology (TDM)
3. Federated Methodology (FM)
4. Hybrid Methodology (HM)

a. Bottom-Up Methodology (BUM)

Kim R. is a well-known author who specializes in data warehousing [5]. Dimension modeling, or the Kimball technique, is his Datawarehouse design approach. The bottom-up technique is used in this methodology. The aim of a bottom-up strategy is to provide business value as soon as possible by establishing dimensional data marts. These data marts, unlike the top-down method, include all of the data, both atomic and summary. The Datawarehouse bus architecture is used to link data marts together. An aspect is pooled across facts in two or even more datamarts in the bus architecture. These are referred to as conforming dimensions. Data marts are used to combine these conformed dimensions, and thus a Datawarehouse is created.

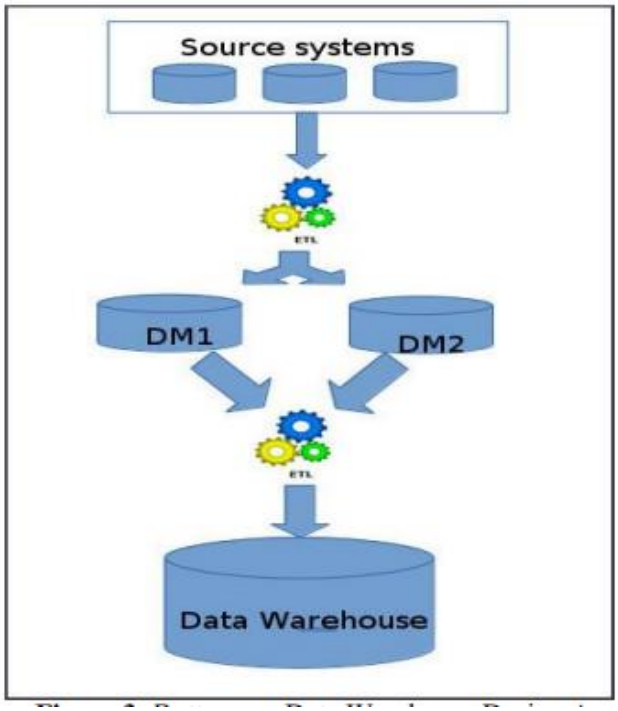

Figure 3: Bottom-up Datawarehouse Design Approach [6]

b. The Benefits and Drawbacks of a Bottom-Up Methodology

1. A chief advantage of a BUM strategy is that it focuses on utilizing dimensional, star schema models to create user-friendly, flexible data structures. It also provides immediate benefit since it does not need the installation of a large infrastructure.
2. Because datamarts include both detailed and nuclear data, users don't need to "dig through" from one structure to another to get detailed or transactional information.
3. This model includes a set of consistent data marts that can be supplied rapidly.
4. A bottom-up approach has the disadvantage of requiring companies to impose the usage of typical magnitudes and evidences in order to guarantee integration and provide a single version of the truth.

c. Top-Down Methodology (TDM)

Bill Inmon is often regarded as the "creator of data warehousing" because of the creator's top-down style [5]. It is a component of a larger system of business intelligence (BI). In case one uses a top-down strategy, they'll have to look at worldwide commercial requirements, strategize the way to build a Datawarehouse, design it, and put it all together. Datawarehouse contains nuclear or operational data collected from a single or multi-source systems and merged into a regularized corporate data model. The information is then aggregated, dimensioned, and transmitted to one or more "dependent" data marts. Since the user get all of the data from a central Datawarehouse, these data marts are "dependent."

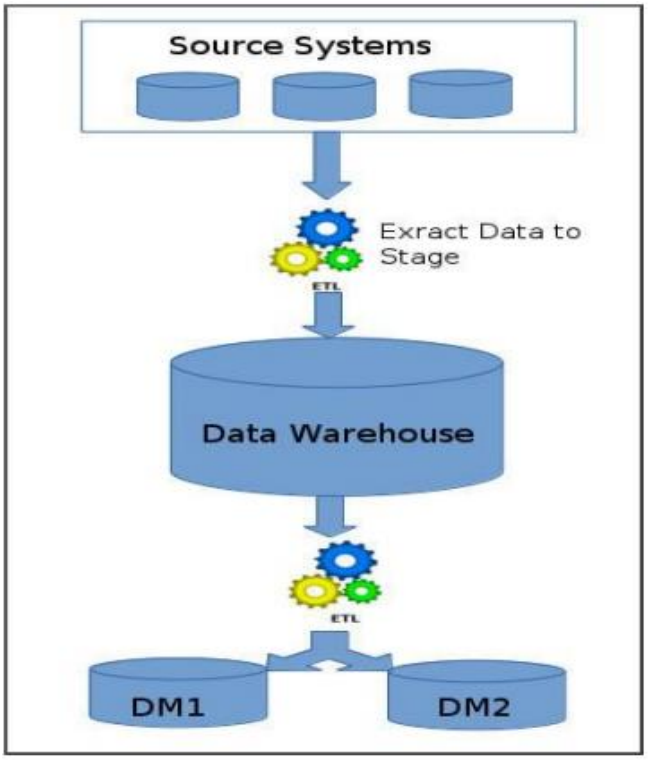

Figure 4: Top-down Datawarehouse Design Approach [7]

d. The Benefits and Drawbacks of a Top-Down Approach

1. It supports downstream analytic data structures with an integrated, flexible design. To begin with, this implies that the Datawarehouse serves as a starting point for all data marts, ensuring consistency and uniformity in order for businesses to reach a single version of the truth.
2. Second, the warehouse's atomic data allows businesses to repurpose data in a variety of ways to suit new and unanticipated business requirements. A Datawarehouse, for example, may be used to generate rich data sets for statisticians to provide operational reports, as well as to provisions operational data stores (ODS) and investigative usages.
3. The major disadvantage of a top-down strategy is that it may take longer and cost more to implement than other techniques, particularly in the beginning.

C. Federated Methodology (FM)

The federated method often mistaken with the hybrid strategy described previously, as well as top-down data warehousing systems such as "hub-and-spoke" [8]. Though, the FM as described by a data warehouse scientist Doug Hackney, its most outspoken proponent. The federated method, according to Hackney, is "architecture of architectures." A FM, Hackney admits, will not ever win prizes for beauty or be put out on pristine white boards as an "ideal solution."

a. The Benefits and Drawbacks of a Federated Approach

1. The strategy simply encourages companies to communicate the "highest value" metrics, dimensions, and measurements as widely as feasible. Several data marts, datawarehouses, or analytic apps are available.
2. The federated method has a significant flaw in that it is not described.
3. Another issue is that, deprived of a particular architype in mind, a FM may continue to decentralize and fragment logical assets, making it more difficult to provide an enterprise perspective in the end.

D. Hybrid Methodology (HM)

The HM method attempts to combine the greatest aspects of both the top-down and bottom-up techniques. It tries to take advantage of the "bottom-up" method's speed and user-orientation while sacrificing the integration imposed by a Datawarehouse in a "top down" approach [9]. To store and maintain the enterprise and local models in the data marts, as well as synchronize the differences between them, the hybrid method depends on an extraction, transformation, and load (ETL) tool.

a. The Benefits and Drawbacks of a Hybrid Approach

1. It blends fast development methods with a foundation for corporate architecture.
2. It iterates on an enterprise data model, building a heavyweight infrastructure only when it's ready.

3. Backfilling a Datawarehouse may be a time-consuming operation that provides little apparent benefit and consequently may not ever be financed.
4. Rely profoundly on an ETL system to bring into line metadata across local and enterprise versions, create aggregates, load detail set of information, and manage the move to a DW architecture.
5. Local groups may find it too simple to deviate irreversibly from the corporate data model using the hybrid approach.
+

E. Different Extraction Methods in Datawarehouse

Extraction is the initial stage in the ETL process. The method of digging out data from a base system for usage in a Datawarehouse environment is known as extraction. Datawarehouse extraction techniques are divided into two categories: logical extraction and physical extraction.

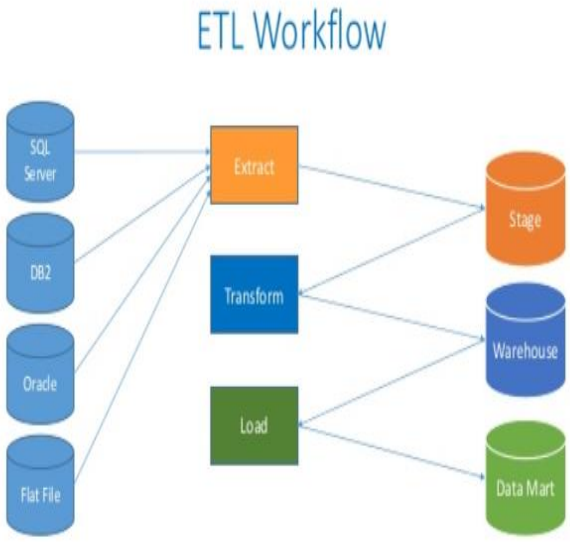

Figure 5: Extraction, Transform, Load (ETL) Workflow [10]

F. Characteristics of Datawarehouse:

There are three distinct features of Datawarehouses

a. Integrated

The harvesting and modification of information is coherent irrespective of the original input. A DW is a centralized collection of knowledge that might be studied to help make better judgments. Flow of data together into DW on a constant schedule via operational platforms, RDBMS, and certain other sort of sources. Data is accessed by industry experts, computer scientists, database administrators (DBAs), and policymakers via business intelligence (BI) software, Postgresql interface, and related advanced analytics.

b. Non-Volatile

The information in a Datawarehouse is not updated in real time. It is kept up to date by uploading data on a regular basis, shielding it from the effects of temporary change.

G. Datawarehouse Benefits

a. Facilitate Improved Choice Considerations

Corporate choice considerations are not at all the need to hedge their bets or make critical business choices based on incomplete or inadequate facts.

b. Access to Data Quickly and Easily

Users may nearly immediately access a vast amount of data from a variety of sources. It implies you won't have to waste time manually extracting data from different sources.

c. Consistent Data of High Quality

Datawarehouse collects data from a variety of sources and converts it into a standardized format that can be utilized throughout your company.

H. Datawarehouse Applications

A datawarehouse assists company leaders in organizing, analyzing, and using their data to make decisions. A Datawarehouse server is the single component of an enterprise management plan-execute-assess “closed loop” feedback system. The following are some of the areas where datawarehouse is extensively utilized.

1. Analytical Processing
2. Banking services
3. Consumer goals
4. Controlled manufacturing
5. Data Mining
6. Decision Making
7. Financial services
8. Information Processing
9. Mailing Box Applications
10. Real Life
11. Retail sectors
12. Statistical Analysis
13. Various Industries

III. CONCLUSION

The Datawarehouse is a database that is maintained distinct from the operating database of the company. A Datawarehouse offers us with generalized and consolidated data in a multidimensional perspective. A Datawarehouse also gives us with Online Analytical Processing (OLAP) capabilities in addition to generalized and consolidated data. These technologies enable us to conduct interactive and effective multidimensional data analysis. Data generalization and data mining are the outcomes of the study. Datawarehouse Overview, Datawarehouse Design Approach, Different Datawarehouse Extraction Methods, Datawarehouse Characteristics and Benefits, Datawarehouse Components, Datawarehouse Types, and Datawarehouse Applications are all covered here.

REFERENCES

[1] Verma H. Data-warehousing on Cloud Computing. Int J Adv Res Comput Eng Technol. 2013;

[2] Gupta P. DATAWAREHOUSING AND ETL PROCESSES: An Explanatory Research. Int J Eng Dev Res. 2016;

[3] What Is the Future of Data Warehousing? [Internet]. 2016 [cited 2018 Sep 10]. Available from: https://www.thedigitaltransformationpeople.com/channels/enabling-technologies/what-is-the-future-of-data-warehousing/

[4] Chandra P, Gupta MK. Comprehensive survey on data warehousing research. Int J Inf Technol. 2018;

[5] Interview – Bill Inmon, Father of Data Warehouse [Internet]. 2016 [cited 2018 Sep 10]. Available from: https://analyticsindiamag.com/interview-bill-inmon-father-of-data-warehouse/

[6] S V, Srinath M, Kumar AC, A.S N. Data Warehousing Architecture and Pre-Processing. IJARCCE. 2017;

[7] Jaroli P, Masson P. Data Warehousing and OLAP Technology (Data warehousing). Int J Eng Trends Technol. 2017;

[8] Khnaisser C, Lavoie L, Diab H, Éthier J-F. Data Warehouse Design Methods Review for the Healthcare Domain. East Eur Conf Adv Databases Inf Syst ADBIS 2015 New Trends Databases Inf Syst. 2015;

[9] S.Kulkarni P, W. Bakal J. Hybrid Approaches for Data Cleaning in Data Warehouse. Int J Comput Appl. 2014;

[10] Kimball R, Caserta J. The Data Warehouse ETL Toolkit. The effects of brief mindfulness intervention on acute pain experience: An examination of individual difference. 2015.