FUDMA Journal of Sciences (FJS)
ISSN online: 2616-1370
ISSN print: 2645 - 2944
Vol. 10 No. 1, January, 2026, pp 280 – 290
DOI: https://doi.org/10.33003/fjs-2026-1001-4424

A SYSTEMATIC MAPPING REVIEW OF DATA WAREHOUSE SOLUTIONS IN DATA MANAGEMENT

***Mpi Chimekwa, Bilkisu Muhammad-Bello, Ibrahim Anka Salihu and Joshua Abah**

Department of Software Engineering, Nile University of Nigeria, Nigeria.

*Corresponding authors' email: 246370029@nileuniversity.edu.ng

ABSTRACT

Organizations heavily rely on data for operations, analysis and data-driven business support decisions for productivity, profitability and growth. Numerous data warehouse solutions exist with varying scope, capabilities and comparative strengths which most industry players are not aware of. Therefore, this paper is concerned about systematically mapping the research on Data Warehouses for Data Management by identifying and characterizing the publication landscape, the gaps, research focus and methods. The papers were selected using filtering and inclusion. Study design and time frames were used to get papers from academic databases. The papers were screened using the titles, review of abstracts, which were followed by full-text review for relevant papers. Evaluation of papers were undertaken by answering Research Questions (RQs), summarizing and presenting the results from the primary studies in tables and charts for ease of analysis. The research findings using 30 reviewed papers showed Lakehouse and Hybrid are the most used Data Warehouse architectural styles followed by Cloud Natives. Data Base Management System is still the most prevalent technology for Data Warehouse storage. Scalability and Throughput are the most reported factors in Data Warehouse cost optimization. In conclusion, security, maintainability, reliability and data quality are the most identified attributes affecting Data Warehouse quality.

Keywords: Big Data, Data, Data Warehouse, Data Management, Database Management System

INTRODUCTION

Big Data or generally "Data" is called the new "Oil" if processed into useful information to support actionable data-driven decisions for organizations. At the heart of this powerful insights derivable from data is "Data Analytics" which is popular in business practices today. Many companies are attempting to incorporate data-driven decision-making because of the explosive rise in big data analysis (Kim et al., 2025). The demand for data-analytics will likely grow as technology increasingly pervades society. To meet these growing demands in Data Management (DM), requires a Data Warehouse (DW) which serves as the repository for a wide array of sensitive or strategic data (Fugkeaw et al., 2024). These DWs are implemented on Snowflake, Google BigQuery, Databricks, Lakehouse, Azure Synapse and Amazon Redshift which are scalable pay-as-you-go cloud solutions for handling huge dataset. (Zagan & Danubianu, 2023) in their work proposed Data Lake (DL) in cloud architecture using Azure Data Lake Storage Gen2 (ADLS Gen2) for Extract–Load– Transform (ELT) pipelines. The result enables large volume of data to be stored in their raw forms, afterwards transformed and analyzed without the need for structured schemas. These solutions are affordable, easily accessible and flexible.

Considering the growing interest in Data Warehouse for Data Management, this systematic mapping study attempts to bridge a gap by extensively analyzing relevant studies published between 2020 and 2025. The duration choice is to enable this research focus on most recent advances in Data Warehouse technologies and solutions by industry experts. The aim is to understand and analyze Data Warehouse architectures, integration, trends, and contribution of Data Warehouses in Data-Driven Decision Making or better still "Data Management" and demographics of the primary studies. The remainder of this paper is organized as follows: Section II provides works related to reviews of Data Warehouse in Data-Driven Decision making. Section III will focus on the research method, Section IV will discuss the results and their implications while section V will contain the conclusions and recommendation for future study respectively.

Related Work

This section aims to summarize the key points of various papers published in the domain of Data-Warehouse. The main goal is to ascertain their contributions to the research topic and impact in advancing Data-Warehouse as the preferred solution in Data-Management (Data-Driven Decision Making.)

(Rique et al., 2023) worked on a case study of the information needs and the perceptions of data-driven decision-making of practitioners from one software development organization. By interviewing 19 software leadership personnel, (Rique et al., 2023) argued that their result contributed to helping organizations adopt data-driven solutions in their analytical initiatives. (Colombari & Neirotti, 2024) used a mixed research method to test if production workers' involvement and frontline managers' competency are associated with the use of Data-Driven Decision-Making (DDM). By building a regression model, triangulated with qualitative interview data, the researchers' result suggested that firms with skilled frontline managers are more likely to adopt DDM. (Sunwoo et al., 2025) reported result of utilizing 1,407 experimental cases from 48 companies on the Hackle online experimentation platform. Their work compared decision-making outcomes experiment designers and Generative AI, analyzing agreement rates and identifying patterns across companies.

Interestingly. Cloud Data Warehouses (CDW) have been widely investigated for their reliability, elasticity and scalability in handling huge datasets. Their unique architecture makes them suitable for real-time data warehouse and big data analytics integration. Proposing a Privacy-Preserving Access Control with Fast Online Analytical Processing (OLAP) Query and Efficient Revocation for CDW (PPAC-CDW) scheme (Fugkeaw et al., 2024) reported that the scheme has fast query performance based on blockchain with optimized ciphertext retrieval.

(Sreepathy et al., 2024) utilized First Unified Data Ingestion Connectors (UDIC), Adaptive Data Variety Transformation (ADVT) and Intelligent Metadata Management (IMM) reported how the void for unified data ingestion interface challenge was addressed using an integrated framework for data ingestion in data lake. The researchers addressed the challenges posed by heterogeneous data sources, formats and metadata management. (Zagan & Danubianu, 2023) reported results from Data Lake (DL) architecture in cloud using Azure Data Storage Gen2 (ADLS Gen2) for extract-load-transform (ETL) pipeline which allows large volume of data to be stored in their raw form. More affordable, easily accessible web server access log data ingestion, storage and transformation over the newest technologies is the main contribution of this study. Similarly, the systematic mapping by (Ataei & Litchfield, 2022) focused on state of big data reference architectures and how Big Data systems development can be better developed. By assessing 22 Big Data reference architecture from academic practices, this study described the common components of reference architecture, their limitations and how Big Data system development can be tackled using effective artefact can tackle complex. (Shaheen et al., 2021) reported the result from performance tuning investigation of large-scale data repositories using the proposed Optimized GAN-based Deep Learning (OGDL) performance prediction model. From results of series of experiments, this study observed that deep learning models performed better than classical machine learning model and about 6-8% increase in accuracy recorded.

(Oreščanin et al., 2024) proposed a new metadata model that supports handling of personal identifiable information in a data lake. The study considered the current Data Lake architecture, metadata models and proposed enhancements to improve the collection, discovery, storage, processing, and removal of personal identifiable information. Similarly, (Hai et al., 2023) reported survey reviews of the development, architectures and system data lakes. The study classified the existing approaches and systems based on their provided functions for data lakes. By this study, (Hai et al., 2023) successfully provided useful technical reference for designing, implementing and deploying data lakes.

Researchers have proposed different methodologies for designing a Data Warehouse solution, however, Bill Inmon's and Ralph Kimball's approaches still stands out. Depending on the design and requirement, researchers or developers can choose which best fits their scenarios.

The approach to use for building a Data Warehouse for Data Management greatly depends on the business objectives, business model, overall requirements, budget and team's expertise and other dependencies respectively.

Comparing both approaches shows that Inmon's top-down approach involves data extracted from external data sources, transformed and loaded into a central Data Warehouse before it is divided into different data marts or business units. While the Kimball's bottom-up approach loads clean data directly into the data marts which are later loaded into a central Data Warehouse.

In this study, the results of a systematic mapping on Data Warehouse in Data Management from **2020** to **2025** were reviewed. The goal is to collect and analyse a set of relevant research papers that study Data Warehouse in Data Management. This will enable cataloguing available evidences with the aim to identify research gaps which would be the basis for future studies.

MATERIALS AND METHODS

Systematic Mapping research method was adopted in this study. It aims to present and categorize research publications within a particular time frame and their availability on a specific topic or research development. It also explores existing literature, the frequency of publication, publication venue. (Sofian et al., 2022) outlined the process steps which entails defining research questions, searching for relevant papers, screening the papers, keywording the abstracts, extracting the data and mapping them as shown in Figure 1.

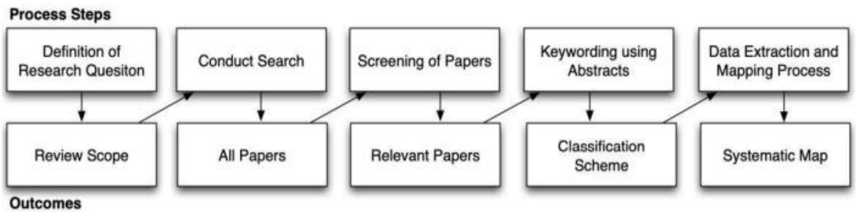

Figure 1: Systematic Mapping Process (Sofian et al., 2022)

Each process step has an outcome and the complete process is the systematic map as explained in (Sofian et al., 2022) in Table 1 as follows:

Table 1: Outcome of Systematic Mapping Process Steps

S/No.	Process Step	Outcome
A	Definition of Research Questions (Research Scope)	The primary goal of a systematic mapping study is to provide an overview of a research area and identify the quantity and type of research and results available within this area.
B	Conduct Search for Primary Studies (All Papers)	Primary studies were identified using search strings on scientific databases or browsing manually through relevant conference proceedings or journal publications.
C	Screening of Papers for Inclusion and Exclusion (Relevant Papers)	Inclusion and exclusion criteria were used to exclude studies that were not relevant for answering the research questions.
D	Keywording of Abstracts (Classification Scheme)	Keywording is a way to reduce the time needed to develop the classification scheme and ensure that the method considers the existing studies.
E	Data Extraction and Mapping of Studies (Systematic Map)	Once the classification scheme was in place, the relevant articles were sorted into the scheme, i.e., the actual data extraction took place.

Research Questions
The primary Research Question (RQ) of this systematic mapping study was: 'What architectural styles are used for data warehouses (e.g., Inmon, Kimball/Dimensional, Data Vault, Lakehouse-hybrid, cloud-native MPP)?' This primary question was divided into seven RQs. Table 2 lists the formulated RQs along with the rationale behind each RQ.

Table 2: Formulated Research Questions and Rationale

RQ No	Research Question	Motivation
RQ1	What architectural styles are used for data warehouses?	To identify the Data Warehouse architectural styles whether Inmon, Kimball, Data Vault, Lakehouse-hybrid or Cloud Native have been applied.
RQ2	Which enabling technologies are most prevalent, and how has this evolved over time?	To identify which technology be it Database Management System, visualization tools, streaming, ETL/ELT tools etc. that are more in use and how they have evolved of time.
RQ3	What is the evidence on performance and cost?	To understand how latency, concurrency, scalability and throughput affect performance and cost.
RQ4	How do solutions address quality attributes?	To know how reliability, security, cost, governance, data quality and maintainability affect quality attributes.
RQ5	What trade-offs exist between on premises and cloud DW solutions?	To understand the key trade-offs of using on-premises as against cloud Database solutions.

Data Sources
Six electronic databases were considered as primary data sources for potentially relevant studies. Google Scholar was not included in the list due to its low-precision results and the overlapping of results from other data sources. The electronic databases used in the search process are listed in Table 3.

Table 3: List of Electronic Databases

Database name	Link
IEEE Xplore	https://ieeexplore.ieee.org/Xplore/home.jsp
Science Direct	https://www.sciencedirect.com/
Google Scholar	https://scholar.google.com/
Web of Science Core Collection	https://clarivate.com/
ACM Digital Library	https://dl.acm.org/
Scopus	https://www.elsevier.com/

Search Terms
The search term is very important for getting the best result from the relevant studies. As suggested by (Sofian et al., 2022) the PICO viewpoint which stands for Population, Intervention, Comparison and Outcome was used:

Population: Primary studies in Data Warehouse?
Intervention: Data Warehouse techniques?
Comparisons: Techniques, Contribution, Dataset, Performance metrics, and Data Warehouse phases.
Outcome: Classification of the types and combination of Data Warehouse techniques applied in Data Management.
Using the PICO structure, a generic search string was constructed to ensure consistency of the search across multiple databases:

Core concept string
("data warehouse" OR "data warehousing" OR "enterprise data warehouse" OR "EDW" OR "dimensional model" OR "data vault" OR "lake house" OR "modern data warehouse") AND (architecture OR design OR implementation OR performance OR scalability OR "quality attribute*" OR security OR governance OR ETL OR ELT OR streaming OR real-time OR MPP OR columnar) `

Technology-focused augmentations (combine with core)
("massively parallel" OR MPP) OR (columnar OR "column store") OR (Hadoop OR Hive OR Spark OR Presto OR Trino) OR (Snowflake OR Redshift OR BigQuery OR Synapse) OR (Oracle Exadata OR Teradata) OR (Parquet OR ORC OR Delta))

Inclusion and Exclusion Criteria
Different criteria were used to include and exclude the studies from the data sources in this systematic study as shown in Table 4 and Table 5 below:

Table 4: Inclusion Criteria

Inclusion Criteria	
IC1	Articles published between 2020 and 2025
IC2	Inclusion of the most recent article in the case of multiple study retrieval
IC3	Articles that are peer-reviewed
IC4	Articles providing the Data Warehouse phases and Data Management techniques

Table 5: Exclusion Criteria

Exclusion Criteria	
EC1	Articles that do not meet the inclusion criteria
EC2	Articles not complete or available as just abstract
EC3	Study language not English
EC4	Articles providing the Data Warehouse phases and Data Management techniques
EC5	Articles providing unclear results

Data Warehouse Process

The authors introduced different Data Warehouse processes architectures. According to (*Data Warehouse Architecture*, 2025) the common architectures are Single tier (in which the data warehouse is designed to be a central repository for all data as the platform for analysis and querying), the Two tier architecture (where source systems is separated from the data warehouse which results to having two layers) and the Three tier (which comprises of the data sources and data storage, the data access methods and data ingestion or extractor layer). This study will use the data warehouse architecture key components and processes provided by (*Data Warehouse Architecture*, 2025) as shown in Figure 2 below:

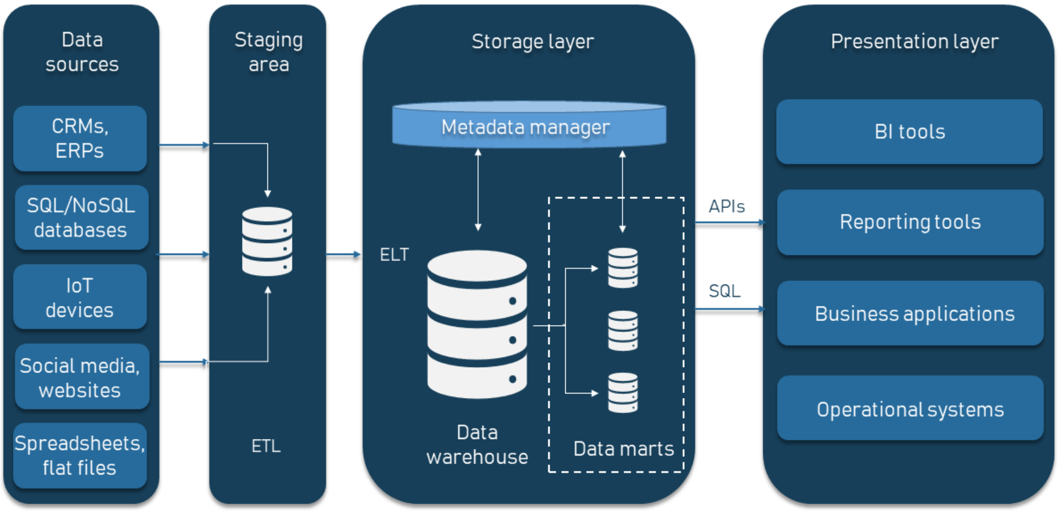

Figure 2: Data Warehouse Architecture (*Data Warehouse Architecture*, 2025)

These processes and components includes:

Data Source Layer

This is where the raw data is collected from different sources depending on the business requirement and design. This may include but not limited to CRM systems, ERP systems and other 3rd party data providers through APIs etc.

Staging/Transformation Layer

This is an intermediate area where data is cleaned, integrated and transformed. This includes basically the ETL/ELT (Extract, Transform, Load/Extract, Load, and Transform) which is the core process of moving/processing data before storage in the warehouse.

Storage/Data Serving Layer

The storage/Data Serving Layer is the central data repository for storing the transformed, integrated and historical data for the purpose of easy querying and analytics.

Presentation Layer

This serves as the top tier component for connecting tools and applications which enable end-users to seamlessly interact with the data for reporting, analysis and data mining.

RESULTS AND DISCUSSION

In this section, all the Research Questions (RQs) are answered by analysing, summarizing and presenting the results extracted from the collection of primary studies. These results cover overviews, tables and graphs for ease of understanding.

Rq1: What Architectural Styles Are Used for Data Warehouses?

Established on the current analysis of the 30 selected primary study papers, five architectural technologies used in the various studies were identified. In total, 2 (PS5, PS7) expressly used Inmon, 2 (PS4, PS5) used Kimball (2), 1 (PS1) used Data Vault, 10 (PS1, PS6, PS7, PS10, PS15, PS16, PS17, PS20, PS22, PS23) used Lakehouse-Hybrid and 9 (PS2, PS8, PS11, PS15, PS20, PS21, PS24, PS28 and PS30) used Cloud Native architectures respectively.

Table 6: Architectural Style Usage in Data Warehouse

PS	Architectural style used for data warehouses				
	Inmon	Kimball	Data Vault	Lakehouse-Hybrid	Cloud Native
PS1	0	0	1	1	0
PS2	0	0	0	0	1
PS3	0	0	0	0	0
PS4	0	1	0	0	0
PS5	1	1	0	0	0
PS6	0	0	0	1	0
PS7	1	0	0	1	0
PS8	0	0	0	0	1
PS9	0	0	0	0	0
PS10	0	0	0	1	0
PS11	0	0	0	0	1
PS12	0	0	0	0	0
PS13	0	0	0	0	0
PS14	0	0	0	0	0
PS15	0	0	0	1	1
PS16	0	0	0	1	0
PS17	0	0	0	1	0
PS18	0	0	0	0	0
PS19	0	0	0	0	0
PS20	0	0	0	1	1
PS21	0	0	0	0	1
PS22	0	0	0	1	0
PS23	0	0	0	1	0
PS24	0	0	0	0	1
PS25	0	0	0	0	0
PS26	0	0	0	0	0
PS27	0	0	0	0	0
PS28	0	0	0	0	1
PS29	0	0	0	0	0
PS30	0	0	0	0	1
Total	**2**	**2**	**1**	**10**	**9**

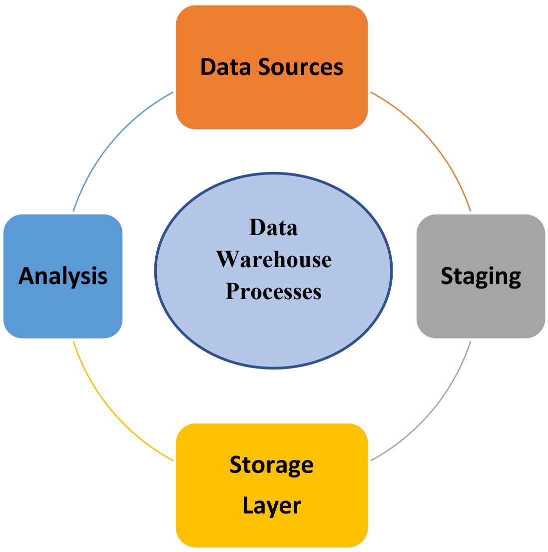

Figure 3: Data Warehouses Processes

Using the above Figure 3 Data Warehouse Processes, the current analysis shows that the Lakehouse was the most used architecture as shown in Figure 4. This was followed by Cloud Natives which are gaining increasing attention too as Cloud Computing solutions become more affordable and easily accessible.

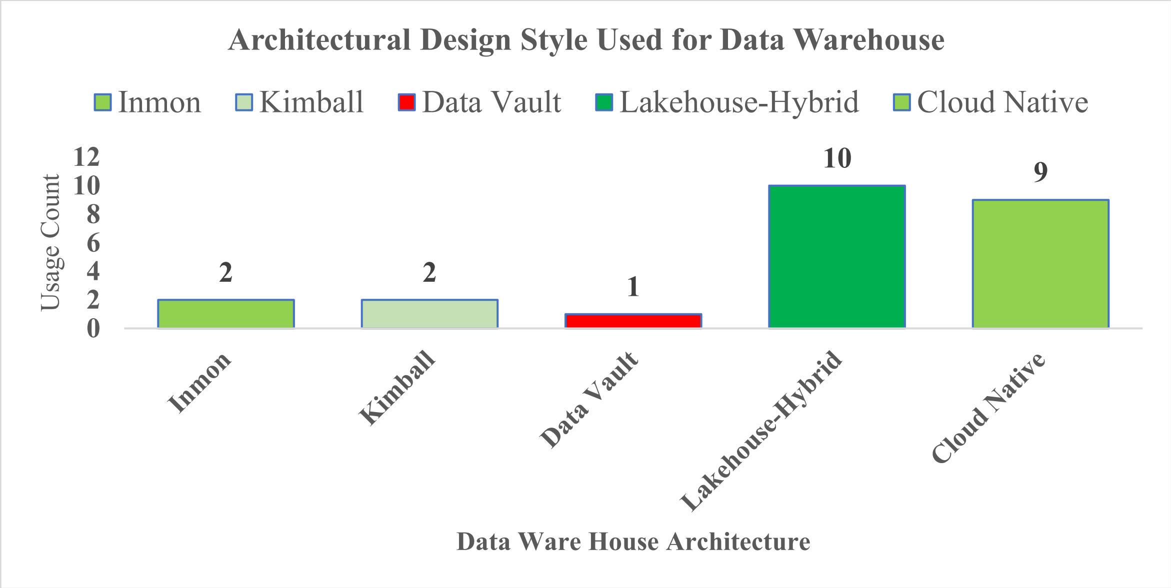

Figure 4: Visualization of the Architectural Style Usage for Data Warehouses

RQ2: Which Enabling Technologies Are Most Prevalent, And How Has Evolved Over Time?

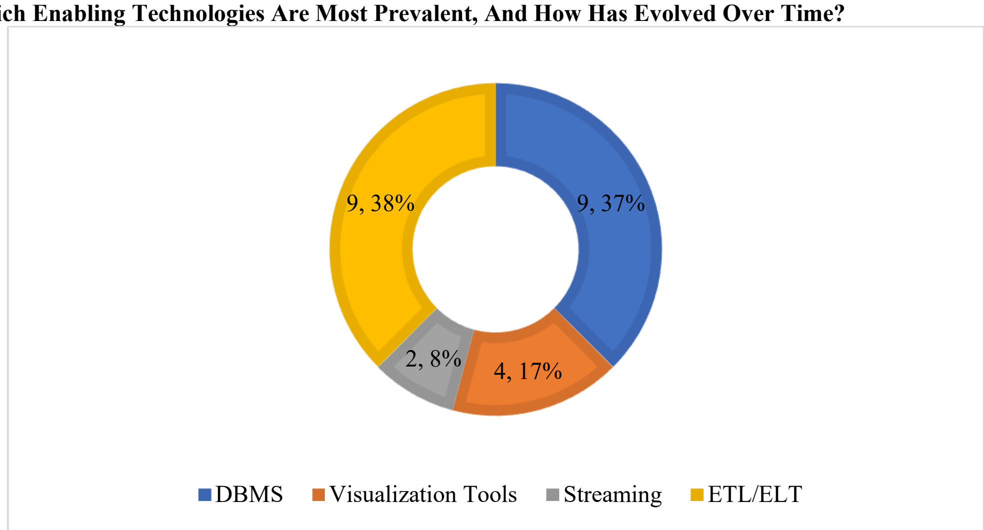

Figure 5: Percentage Spread of Enabling Technologies most prevalent

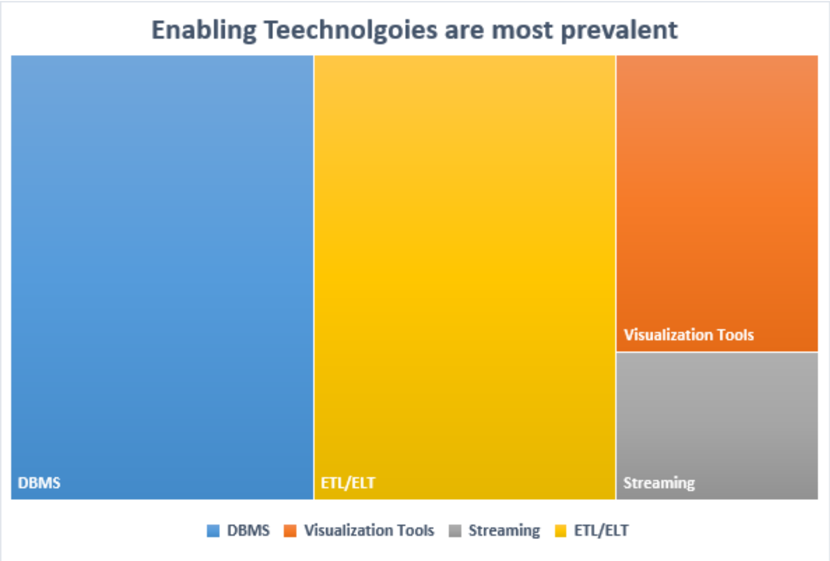

Figure 6: Percentage Spread of Enabling Technologies Are Most Prevalent

Figure 5 and 6 show the spread of the enabling technologies and most prevalent. The analysis shows that ETL/ELT and DBMS are the most enabling technologies for Data Warehouse solutions.

Table 7: Enabling Technologies Spread and Prevalence

PS	Enabling Technologies Spread and Prevalence			
	DBMS	**Visualization Tools**	**Streaming**	**ETL/ELT**
PS1	0	1	0	1
PS2	1	0	0	0
PS3	1	0	1	1
PS4	1	0	0	1
PS5	1	0	0	0
PS6	0	1	0	0
PS7	1	0	0	1
PS8	1	0	0	0
PS9	1	0	0	1
PS10	0	0	0	0
PS11	0	0	0	1
PS12	0	0	0	0
PS13	0	0	0	1
PS14	0	0	0	0
PS15	0	1	1	0
PS16	1	0	0	0
PS17	0	0	0	0
PS18	0	1	0	1
PS19	1	0	0	0
PS20	0	0	0	0
PS21	0	0	0	1
PS22	0	0	0	0
PS23	0	0	0	0
PS24	0	0	0	0
PS25	0	0	0	0
PS26	0	0	0	0
PS27	0	0	0	0
PS28	0	0	0	0
PS29	0	0	0	0
PS30	0	0	0	0
Total	**9**	**4**	**2**	**9**

RQ3: What Is the Evidence On Performance and Cost?

Table 8: Distribution of Papers Based on Reports on Performance and Cost

	Latency	**Concurrency**	**Scalability**	**Throughput**
Papers	PS3, PS13, PS15, PS20, PS24, PS29	PS20, PS24, PS26, PS29	PS1, PS2, PS7, PS11, PS13, PS20, PS24, PS29	PS10, PS11, PS13, PS20, PS24, PS24, PS26, PS29
Total	**6**	**4**	**8**	**7**

The systematic review of the 30 Papers in this study showed that a total of 25 researchers through their respective works discussed and showed evidences on how them researches contributed to improving performance and minimizing cost in Data Ware House solutions. Table 8 shows that 6 Papers (PS3, PS13, PS15, PS20, PS24, PS29) discussed how their solutions reduced Latency; 4 papers (PS20, PS24, PS26, PS29) outlined how their solution improved Concurrency; 8 Papers (PS1, PS2, PS7, PS11, PS13, PS20, PS24, PS29) enumerated how their studies contributed to enhancing scalability while 7 Papers (PS10, PS11, PS13, PS20, PS24, PS24, PS26, PS29) mentioned how Throughput was enhanced in their respective Data Warehouse solutions.

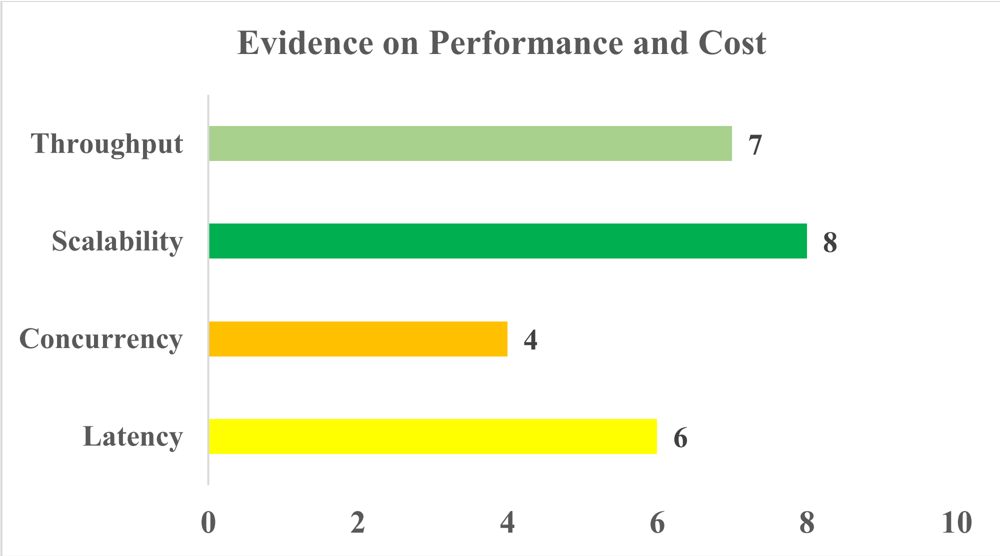

Figure 7: Spread of Papers On Their Evidence On Performance and Cost.

Rq4: How Do Solutions Address Quality Attributes?

Table 9: Spread of Papers Across Data Warehouse Quality Attributes

PS	Quality Attributes					
	Maintainability	Agility	Data Quality	Security	Governance	Reliability
PS1	0	0	0	0	0	0
PS2	0	0	0	0	0	0
PS3	0	0	0	0	0	0
PS4	1	0	0	0	0	0
PS5	0	0	0	0	0	0
PS6	0	0	0	0	0	0
PS7	0	0	0	1	0	1
PS8	0	0	0	0	0	1
PS9	0	0	0	0	0	0
PS10	0	0	0	0	0	0
PS11	0	0	0	1	0	0
PS12	0	0	0	0	0	0
PS13	1	0	0	0	0	0
PS14	0	0	0	0	0	0
PS15	0	0	1	1	1	0
PS16	0	0	0	0	0	0
PS17	0	1	0	0	0	0
PS18	0	0	0	0	0	0
PS19	0	0	0	0	0	0
PS20	1	0	0	0	0	0
PS21	0	0	0	0	0	0
PS22	0	0	0	0	0	0
PS23	0	0	0	0	0	0
PS24	0	0	0	0	0	0
PS25	0	1	0	0	0	0
PS26	0	0	0	0	0	0
PS27	0	0	0	0	0	0
PS28	0	0	0	0	0	0
PS29	0	0	0	1	0	0
PS30	0	0	0	0	0	0
Total	**3**	**2**	**1**	**4**	**1**	**2**

Table 9 above is the summary of data collected from the primary study, where the 30 selected papers were Systematically Reviewed and categorized into different quality attributes such as Maintainability, Agility, Data Quality, Security, Governance and Reliability. The primary data was further processed as shown in the Table 10.

Table 10: Summary of Papers Across Data Warehouse Quality Attributes

Quality Attributes	Primary Studies	Number of Papers
Maintainability	PS4, PS13, PS20	3
Agility	PS17, PS25	2
Data Quality	PS15	1
Security	PS7, PS11, PS15, PS29	4
Governance	PS15	1
Reliability	PS7, PS8	2
	Total	**13**

From the above, 3 papers (PS4, PS13, and PS20) reported how maintainability affected the quality of Data Warehouse, Similarly, 2 papers (PS17, PS25) reported how Agility affected their research result, and 1 paper (PS15) also showed how Security contributed to the quality of their Data Warehouse, Governance and Reliability Quality Attributes were reported by 1 Paper (PS15) and 2 Papers (PS7 and PS8) respectively. The Figure below further simplifies the results

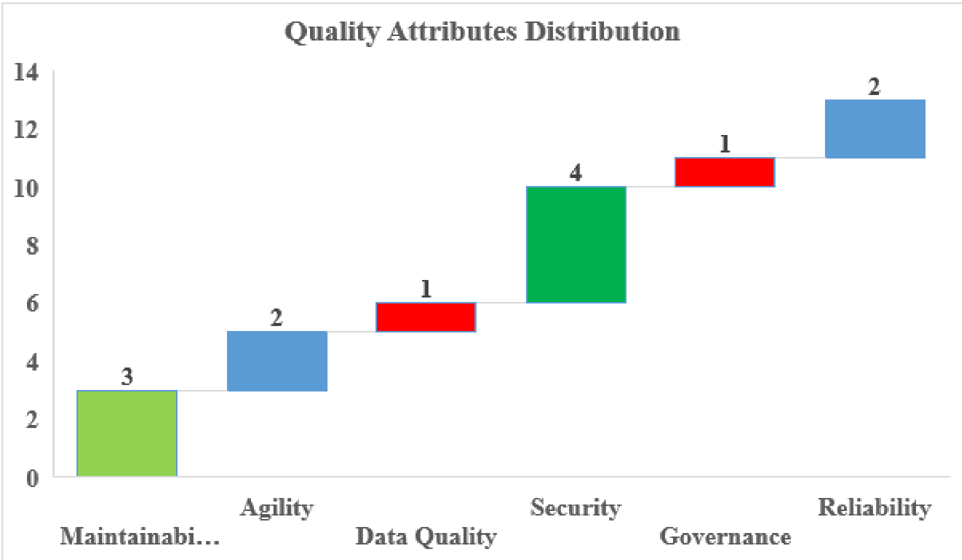

Figure 8: Chart of Primary Studies Across Data Warehouse Quality Attributes

Rq5: What Trade-Offs Exist Between on Premises and Cloud DW Solutions?

Table 11: Technology Trade-Offs

PS	Technology Trade-Offs			
	Visualization	**Cloud**	**Other**	**Database**
PS1	Power BI			
PS2		Azure Cloud	Parquet	HDFS
PS3				MySQL Database
PS4				OLAP Database
PS5				NoSQL
PS6		Data Lake		MySQL
PS7				Central Database
PS8				
PS9				NoSQL
PS10				
PS11				
PS12				
PS13				
PS14			OLTP	MySQL
PS15				
PS16		AI Cloud Platform		Hive Data Warehouse
PS17				
PS18				
PS19				PostgreSQL
PS20				Azure Data Lake
PS21			AWS IoT Core	

PS22				Data Lake
PS23			OLAP cubes	
PS24			Edge Computing, IoT	
PS25				
PS26			OLAP, Data Cube	
PS27				
PS28				Document Database
PS29				
PS30		Amazon S3	Apache Kafka	
Total	**1**	**4**	**7**	**13**

Table 11 represents the distribution of the Primary Studies (PSs) Technologies and the Trade-Offs across Technologies used for Visualization, Cloud, Database and Other Technologies. The result shows that Power BI was used by PS1 for visualization, Cloud Technologies used includes Microsoft Azure, Data Lake, AI Cloud Platforms and Amazon S3.
Similarly, different 13 papers (PS2, PS3, PS4, PS5, PS6, PS7, PS8, PS14, PS16, PS19, PS20, PS22 and PS28) used HDFS, MySQL Database, OLAP Database, MySQL, Central Database, Hive Data Warehouse, PostgreSQL, Azure Data Lake and Document NoSQL Databases for data storage purposes.
Interestingly, 7 Papers (PS1, PS14, PS21, OS23, PS26 andPS30) applied different technologies Parquet, OLTP, AWS IoT Core, OLAP cubes, Edge Computing and Apache Kafka were used to support implementation of the different Data Warehouse solutions as shown in Figure 9.

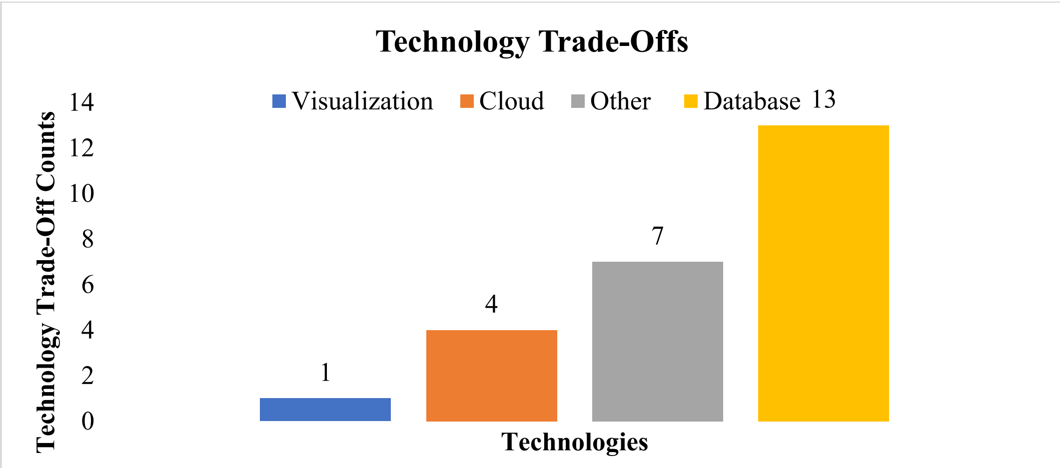

Figure 9: Technology Trade-Offs for Data Warehouses

Threat to Validity

Construct Validity

The validity of the results from this study is subject to the papers collected on Data Warehouse for this study, which is a function of the inclusion, search term and inclusion terms. This study was based on papers downloaded using Data Warehouse, Data Lake, Data Management and Data Warehouse Applications. This led to having expanded papers, thereby increasing the scope of the study. In order to improve the construct validity in any future study on this topic, future studies may include variant search term around warehouse, possibly including Data Warehouse in different industries and Data Warehouse and Machine Learning.

Internal Validity

In order to reduce the internal validity threat, this study went further to consider papers that were not just 100% Data Warehouse based. The study also considered papers with Data Warehouse even when they were Big Data Papers.

Conclusive Validity

This study focused on identifying and categorizing Architectural Styles which support Data Warehouse Technology. To ensure replicability of this study, the researcher may consider further review of papers which extensively applied the Inmon or Kimball's Architectural Style

External Validity

External validity threat to this study affects the generalization of this study. Possibly, future studies may consider up to 100 papers in order to conclude the extent to which the result can be generalized.

CONCLUSION

This work systematically reviewed papers on Data Warehouse and how different researchers applied different technologies across the architectural design. The 30 papers selected as Primary Studies (PSs) show the spread of Data Warehouse technologies in solving challenging industrial and business problems. This study was successful because the research questions enabled narrowing down to understand how different authors apply the diverse Architectural Process in their designs. This study shows that Data Warehouse technologies are used to process and transform data, store data and utilize the data for presentation which supports data-driven decision making. Achieving the aforementioned is via different architectural styles. In course of this study, it was observed that Data Vaults, Data Lake, Cloud Natives, and Visualization Tools are some of the technologies enabling Data Warehouse solutions. From the study, it was noted that effectively designed and implemented Data Warehouse solutions impact on Performance and Cost through improving the Data Warehouse Latency, Concurrency, Scalability and Throughput. *More so, through an extensive systematic review of the papers, this study is concluding that Agility, Data Quality, Security, Governance, Reliability and Maintainability of a properly designed Data Warehouse affects its Quality*. Additionally, this research revealed that different tools and technologies such as Power BI, MySQL, NoSQL, PostgreSQL, Azure Data, and Amazon S3 are some technological trade-offs which enable successful Data Warehouse solutions. Further research on this topic would focus on investigating researches which are on Data Warehouses based on the search terms, but did not contribute to the mapping results.

RECOMMENDATION FOR FURTHER STUDY

The result from this subject may not be conclusive and generalizable because 30 papers may not be sufficient to make informed decision. For instance the use of Architectural Usage in Data Warehouse showed that Inmon architecture was used in just 2 papers. Similarly 2 papers used Kimball architecture, Lakehouse-Hybrid was used in 10 papers and 9 Cloud Native was used by 9 papers. This result is completely inconclusive due to the size of papers. More so, the results may not be generalizable because the Power BI Visualization Tool was only used in 1 paper, in spite of being one of the most successful commercial Business Intelligence (BI) tools. More so, out of the 30 papers, only 4 papers used Cloud Solutions like Azure Cloud, Data Lake, AI Cloud and Amazon in their Data Warehouse solutions. In addition, the Primary Study size of 30 is not a good representative to generalize the Data Warehouse Quality Attribute. This is because, out of 30 Papers used in the Systematic Review, only 3 related Maintainability, just 2 papers reported Agility, 1 discussed Data Quality, 4 papers discussed security, and 1 paper discussed Governance while only 2 papers mentioned Reliability. From the aforementioned, t is obvious that future study is required where at least 100 to 200 papers could be systematically reviewed to get a better representation of How Data Warehouses enhance Data-Driven Decisions in Data Management processes. Finally, future study should include search from more publication venues because this study used mostly papers from IEEE Explorer, Science Direct and Google Scholar.

REFERENCES

Ataei, P., & Litchfield, A. (2022). The State of Big Data Reference Architectures: A Systematic Literature Review. *IEEE Access*, *10*, 113789–113807. https://doi.org/10.1109/ACCESS.2022.3217557

Colombari, R., & Neirotti, P. (2024). Leveraging Frontline Employees' Knowledge for Operational Data-Driven Decision-Making: A Multilevel Perspective. *IEEE Transactions on Engineering Management*, *71*, 13840–13851. https://doi.org/10.1109/TEM.2023.3291272

Data Warehouse Architecture. (2025, May 14). Databricks. https://www.databricks.com/discover/data-warehouse-architecture

Fugkeaw, S., Hak, L., & Theeramunkong, T. (2024). Achieving Secure, Verifiable, and Efficient Boolean Keyword Searchable Encryption for Cloud Data Warehouse. *IEEE Access*, *12*, 49848–49864. https://doi.org/10.1109/ACCESS.2024.3383320

Hai, R., Koutras, C., Quix, C., & Jarke, M. (2023). Data Lakes: A Survey of Functions and Systems. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12571–12590. https://doi.org/10.1109/TKDE.2023.3270101

Inmon, W. H. (n.d.). *Building the Data Warehouse*.

Kim, J., Park, G., Woo, M., & Geum, Y. (2025). How Big Data Has Changed Technology Roadmapping: A Review on Data-Driven Roadmapping. *IEEE Access*, *13*, 8297–8309. https://doi.org/10.1109/ACCESS.2025.3526173

Oreščanin, D., Hlupić, T., & Vrdoljak, B. (2024). Managing Personal Identifiable Information in Data Lakes. *IEEE Access*, *12*, 32164–32180. https://doi.org/10.1109/ACCESS.2024.3365042

Rique, T., Perkusich, M., Dantas, E., Albuquerque, D., Gorgônio, K., Almeida, H., & Perkusich, A. (2023). On Adopting Software Analytics for Managerial Decision-Making: A Practitioner's Perspective. *IEEE Access*, *11*, 73145–73163. https://doi.org/10.1109/ACCESS.2023.3294823

Shaheen, N., Raza, B., Shahid, A. R., & Malik, A. K. (2021). Autonomic Workload Performance Modeling for Large-Scale Databases and Data Warehouses Through Deep Belief Network with Data Augmentation Using Conditional Generative Adversarial Networks. *IEEE Access*, *9*, 97603–97620. https://doi.org/10.1109/ACCESS.2021.3096039

Sofian, H., Yunus, N. A. M., & Ahmad, R. (2022). Systematic Mapping: Artificial Intelligence Techniques in Software Engineering. *IEEE Access*, *10*, 51021–51040. https://doi.org/10.1109/ACCESS.2022.3174115

Sreepathy, H. V., Dinesh Rao, B., Kumar Jaysubramanian, M., & Deepak Rao, B. (2024). Data Ingestions as a Service (DIaaS): A Unified Interface for Heterogeneous Data Ingestion, Transformation, and Metadata Management for Data Lake. *IEEE Access*, *12*, 156131–156145. https://doi.org/10.1109/ACCESS.2024.3479736

Sunwoo, C., Kwon, H., Min Kim, J., Lim, H.-H., Kim, Y., Hwang, D., & Kim, J. (2025). Automated Business Decision-Making Using Generative AI in Online A/B Testing: Comparative Analysis with Human Decision-Making. *IEEE Access*, *13*, 124530–124542. https://doi.org/10.1109/ACCESS.2025.3588480

Zagan, E., & Danubianu, M. (2023). Data Lake Architecture for Storing and Transforming Web Server Access Log Files. *IEEE Access*, *11*, 40916–40929. https://doi.org/10.1109/ACCESS.2023.3270368