International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com OPEN ACCESS E-ISSN : 3049-2424
Volume 3 Issue 4 , April 2026
DOI: https://doi.org/10.71366/ijwos03042689514

ETL Process in Data Migration

Pardeep Singh[1], Taruna Gulati[2]
[1] *Head of Department (CSE),Guru Tegh Bahadur 4th Centenary Engineering College, G-8 Area, Rajouri Garden, New Delhi-110064, India*
[2]*Associate Professor, Guru Tegh Bahadur 4th Centenary Engineering College,G-8 Area, Rajouri Garden, New Delhi-110064, India.*

Article Info

Article History:

Published:30 April 2026

Publication Issue:
Volume 3, Issue 4
April-2026

Page Number:
354-360

Corresponding Author:
Pardeep Singh

Abstract:

Data Migration is the process of transferring data between storage types, formats, or computer systems. It is a key consideration for any system implementation, upgrade, or consolidation. It has become a priority in many industries, spawned by a variety of business needs. The ETL process is one of the most important processes of data warehouse building. It involves all sorts of environments, involves multiple technical means to extract data from the data sources of various operation types and load them to the target: data warehouse. This process actually integrates data from different data sources and data of different operation types, and transforms the nonstandard data into standard ones to some extent. The systems and methods may involve analyzing a source language construct of the source data integration platform to determine a logical syntax.

Keywords: Conceptual Model, Data cleaning , Data Mart, Data Migration, Data Quality, Data Warehouses, Decision Making, ETL, Metadata, OLAP, OLTP

1.INTRODUCTION

The Extract-Transform-Load (ETL) system is the foundation of any data warehouse [1]. ETL is used to migrate data from one database to another, to form data marts and data warehouses and also to convert databases from one format or type to another. Data Warehouses (DW) are complex systems employed to integrate the organization's data from several distributed, heterogeneous sources An ETL system consists of three consecutive functional steps: extraction, transformation, and loading [2]:

A. **Extraction:** The ETL Extraction step is responsible for extracting data from the source systems. Each data source has its distinct set of characteristics that need to be managed in order to effectively extract data for the ETL process. The process needs to effectively integrate systems that have different platforms, such as different database management systems, different operating systems, and different communications protocols.

B. **Transformation**: The second step in any ETL scenario is data transformation. The transformation step tends to make some cleaning and con-forming on the incoming data to gain accurate data which is correct, complete, consistent and unambiguous. This process includes data cleaning, transformation, and integration. It defines the granularity of fact tables, the dimension tables, DW schema (star or snowflake), derived facts, slowly changing fact tables and dimension tables . All transformation rules and the resulting schemas are described in the metadata repository.

C. Loading data to the target multidimensional structure is the final ETL step. In this step, extracted and transformed data is written into the dimensional structures actually.

From the 3 processes, we can see metadata plays an important role in ETL, whose mismanagement can lead to the ineffectiveness of ETL processes directly. ETL processes often fails through its triviality and fallibility. The architecture of ETL is shown as Fig. 1[3]. The phases of extract, transform and load were executed in one single process. Under the framework of conventional ETL, the ETL process is defined: for different data source, develop and compile program or script; retrieval records from database; after extract, exchange the data according to users' requirement; load the data to target data warehouse; and process the records piece by piece until the end of source database. The framework of ETL is simple and would be easily implemented under the conventional architecture, but the weakness is obvious: The efficiency and reliability of load is lame which makes the overall scenario weak and difficult [4]

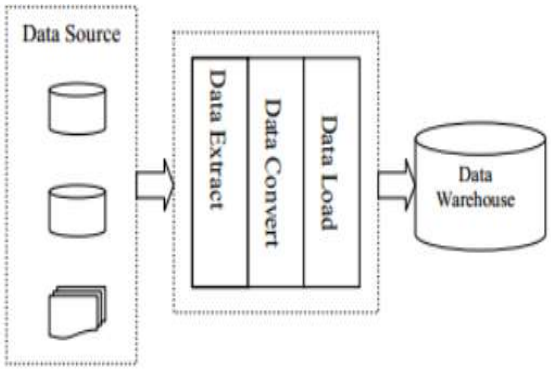

Fig 1 : Architecture of ETL [3]

2.Literature Survey

Data warehouse is a system that gives access to the information sources which helps in decision making process. Starting from the requirement analysis phase, we find that a goal oriented approach to requirement analysis states that for decision making, requirements are needed that must be integrated from the heterogeneous sources. It works from the requirement to the conceptual design of the data warehouse. Based on the demand and mixed driven approaches, a perspective view of the data warehouse decision making is proposed [5]. Further E. Soler [6] faces on the requirement analysis for the DWs which includes the security requirements of the DWs. This uses the MDA approach for the requirements of the data warehouse. The information and the QoS requirements are discussed in this approach which follows the conceptual and the logical design phases. Requirement analysis and the conceptual design are the primary steps for the DW design process [7]. It presents a comprehensive approach for the development of data warehouse. Munawar describes a methodology for requirement analysis to the designing of DW in context with the appropriate approach to reduce the risk of failure [8]. Now a framework for the requirement analysis in support of DW is proposed by ArihamSarkar [9]. It works upto the conceptual design phase with the construction of MD model. It deals with the approaches of refinement with high level designing components. Enterprise data warehouse implementation can add significant value to the enterprise through increased strategic advantage levels if executed properly yet development projects in this area remain very challenging and involve high levels of risk. Adelman et al [10] aver that risks in enterprise data warehouse development projects remains more than that experienced in other projects. Projects in the information technology field are notorious for poor performance compared to their counterparts in other industries. McBride [11] found that one-third of all money spent on software is used to repair botched projects and that billions are spent each year reworking software projects that do not match/fit requirements.

Enterprise data warehouse development projects have shown to follow the same trend with an embarrassingly low success rate across diverse industries.

3. Overview Of Modelling Approaches

The development of the data warehouse involves various approaches. These approaches help to reduce the building complexity of the DW.

					Security Issues			Parameters		
S.No	**Approach**	**Technology Used**	**Concept Used**	**SpecifiedLevel**	**Confiden-Tiality**	**Integrity**	**Autho-Rity**	**Trace-Bility**	**Port-ability**	**Guidelines Supported**
1.	MDA (Soler E., 2008)	i* modeling framework	i* models	Requirement level	Yes	Yes	Yes	No	No	No
2.	MDA (Sergio Lujan Mora, 2005)	Extension of UML	UML 2.0	Conceptual	No	No	No	No	Yes	Yes
3.	ExtendingUML (Mario Plattini, 2006)	Extension of UML 2.0	UML Metamodel	Conceptual And logical	No	Yes	No	No	Yes	Yes
4.	Pragmatic Approach (Torsten Priebe, 2001)	ADAPTed UML Notation	UML	Conceptual	Yes	No	No	Yes	No	No
5.	Relational Metamodel (E.Soler, 2006)	Extension of UML metamodel	CWM	Logical	Yes	No	No	Yes	No	Yes

6.	MDA (E. Soler, 2007)	MDA and QVT framework	ACA	Logical	No	No	Yes	Yes	Yes	No
7.	MDA (C. Blanco, 2008)	UML Extension	ACA	All abstraction level	Yes	Yes	Yes	No	No	Yes

Table 1 :Comparison between various Modelling approach

MDA approach deals with security requirements of the data warehouse at requirement analysis level. This approach includes the information which is needed to define the security at the same level. The security issues like confidentiality, integrity and authority are all discussed here define the requirement at the conceptual and the logical level. UML profile for MD modeling considers main properties of modeling at conceptual level with the specified security constraint using MDA approach [12]. A UML profile is portable with the productivity of the conceptual level to the logical level. This profile needs the primitive security as no other issues are considered at the desired level of the development of the data warehouse. Mario pattinni uses the extension of UML 2.0 for defining the integrity at the conceptual level to the logical level of integration [13]. This profile is portable and supported the guideline as well. A pragmatic approach uses the OLAP systems for mapping the OLAP security [14]. This used the conceptual modeling with the ADAPted UML notation and works on the confidentiality conflict. This approach is highly traceable. E. Soler [15] works on the extension of UML metamodel and continues his research up to the logical level. This approach also promotes the security conflicts with supported guidelines. CWM concept is used for the modeling of this approach as to reduce the conflicts. The MDA approach using the QVT transformations at logical level is presented in [16]. The access control and audit (ACA) model works on the authorization of the conducted designing phase. Another approach of MDA using the UML extension with the access control and audit models defining the security rules at all levels of abstraction [17] and also that deals with the security issues. The above table 1 describes the comparison of the modeling approaches.

4. VARIOUS OLAP TECHNOLOGIES AND THEIR IMPACT ON DECISION MAKING

The OLAP concept was introduced in 1993 by Codd. OLAP is an approach to quickly answer multidimensional analytical queries [21]. In OLAP, a dimension is a sequence of analyzed parameter values. An important goal of multidimensional modeling is to use dimensions to provide as much context as possible for facts [22]. Combinations of dimension values define a cube's cell. A cube stores the result of different calculations and aggregations. There are three variants of OLAP: MOLAP, ROLAP, Hybrid OLAP (HOLAP). We compare these approaches in table 2.Multidimensional OLAP is appropriate for decision making. It offers a number of advantages, including automatic aggregation, visual querying and good query performance due to the use of preaggregation [18]. Besides, MOLAP may be a good solution for the situations in which small to medium sized DBs are the norm and application software speed is critical [19]. If the user wants to change dimensions, the whole deployment process needs to be redone [20].

Table 2: Comparison between various OLAP technologies

	MOLAP	ROLAP	HOLAP
Data Storage	Multidimensional Database	Relational Database	Uses MOLAP to store higher level summary data and ROLAP to store detailed data.
Result sets	Stores in MOLAP cube	Stores no result sets	Stores result sets but not all
Capacity	Requires significant capacity	Requires the least storage capacity	Requires the average storage capacity
Performance	Fastest	Slowest	Average
Dimensions	Minimum	Maximum	Average
Vulnerability	Poor storage utilization	Database design recommended by ER diagram are inappropriate for Decision Support systems	Better in comparison.
Advantages	Fast query performances	No limitations in data volume	Fast access at all levels of aggregation.
Disadvantages	Data redundancy	Slow performance	As slow as ROLAP when we try to access leaf level data

5.CONCLUSION AND FUTURE WORK

In this paper, we make the observation that data warehouse creation is an important task which is increasingly becoming a bottleneck, preventing the rapid deployment of data warehouses [23]. While a number of techniques and software exist for storing the data warehouse, and performing analyses on it, there is a marked lack of tools for the warehouse creation task. Furthermore, it has been observed that doing this on an ad hoc basis has proven to be labor intensive, error prone, and generally frustrating[24]. However, not all problems relevant to warehouse creation have been solved, and a number of research issues remain. The principal goal of this paper has been to identify the common issues in data integration and data warehouse creation[25]. We hope this will lead the developers of warehouse creation tools to examine, and where appropriate incorporate, the techniques developed for data integration, and the researchers in both data integration and data warehousing community to address the open research issues in this important area. In our future work, we intend to study the security problems related to OLAP operations likeaccess control, authorizations, inferences related to complex dynamic queries etc. Further identification of malicious attempts and vulnerabilities will be focused to prevent unauthorized access of data.

References

[1] Ralph Kimball and Joe Caserta, The Data Warehouse ETL Toolkit: Practical Techniques for Extracting,Cleaning, Conforming, and Delivering Data, John Wiley & Sons, Inc., 2004.

[2] Anand, N., & Kumar, M. (2013, June). An overview on data quality issues at data staging etl. In *Proceedings of the International Conference on Advances in Computer Science and Application, Lucknow, India* (pp. 21-22).

[3]. N. Anand and M. Kumar, "Modeling and optimization of extraction-transformation-loading (ETL) processes in data warehouse: An overview," *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Tiruchengode, India, 2013, pp. 1-5 https://doi.org/10.1109/ICCCNT.2013.6726592

[4]. Anand, N., Kumar, N., Dash, J., & Patra, D. (2015). Using ETL for Optimizing Business Intelligence Success in Multiple Investment Combinations. International Journal of Applied Engineering Research (IJAER), 10(6), 5037–5043. https://doi.org/10.5281/zenodo.18378109

[5]. Anand, N. (2014). ETL and its impact on Business Intelligence. *International Journal of Scientific and Research Publications*, *4*(2), 1.

[6]Anand, N. (2012). Application of ETL tools in business intelligence. *International Journal of Scientific and Research Publications*, *2*(11), 1-4.

[7]MatteoGolfarelli: From user requirements to conceptual design in Data warehouse design- A survey. In: bias.csn.unibo.it (2008).

[8] Munawar, Salim, N., Ibrahim, R.: Towards data warehouse quality through integrated requirement analysis. In: ICACSIS, IEEE, international conferences, pp. 259-264(2011).

[9]AribanSarkar: Data warehouse requirement analysis framework business-object based approach. In: IJACSA,

[10] Adelman, S. Moss, L., 2004, Data Warehouse Risks, Sid Adelman & Associates, Sherman Oaks, CA, http://www.sidadelman.com/3_data_warehouse_risks.htm [Accessed on 04 April 2009]vol.3, No.1, 2012

[11]McBride, S., 2004, Poor project management leads to high failure rate, Available from: http://www.itworld.com/041015poor [Accessed on 10 Sep 2009]

[12]Trojillo, J., Soler, E.: A UML 2.0 profile to define security requirements for data Warehouses. In: Computer Science direct, Elsevier, pp. 969-983(2009).

[13] Emilio Soler , Rodolfo Villarroel, Juan Trujillo" Representing security and audit rules in Data warehouse at the logical levels by using the common warehouse meta model" Proceedings of the First International Conference on Availability, Reliability and Security (ARES'O6)0- 7695-2567-9106 $20.00 2006 IEEE

[14] Nayak, D., Anand, N., Prusty, T., & Das, S. Transfer learning for corn leaf disease detection: experimental comparison of MobileNetV2 and Efficient NetBO. In Connecting Intelligence (pp. 408-412). CRC Press. https://doi.org/10.1201/9781003773504-69

[15] Sharma, V., Kumari, P., Das, S., Anand, N., & Kundu, R. Energy efficient IoT enabled smart irrigation system leveraging LoRa technology. In Connecting Intelligence (pp. 413-418). CRC Press. https://doi.org/10.1201/9781003773504-70

[16] S. Agarwal, A. P. Singh and N. Anand, "Evaluation performance study of Firefly algorithm, particle swarm optimization and artificial bee colony algorithm for non-linear mathematical optimization functions," *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Tiruchengode, India, 2013, pp. 1-8. https://doi.org/10.1109/ICCCNT.2013.6726474

[17] Anand, N., Singh, K.J. (2024). A Comprehensive Study of DDoS Attack on Internet of Things Network. In: Swain, B.P., Dixit, U.S. (eds) Recent Advances in Electrical and Electronic Engineering. ICSTE 2023. Lecture Notes in Electrical Engineering, vol 1071. Springer, Singapore. https://doi.org/10.1007/978-981-99-4713-3_56

[18]Anand, N., Singh, K.J. (2023). An Overview on Security and Privacy Concerns in IoT-Based Smart Environments. In: Rao, U.P., Alazab, M., Gohil, B.N., Chelliah, P.R. (eds) Security, Privacy and Data Analytics. ISPDA 2022. Lecture Notes in Electrical Engineering, vol 1049. Springer, Singapore. https://doi.org/10.1007/978-981-99-3569-7_21

[19]Kumar, M., Lal, I.B., Ranjan, R., Kumar, N., Kumar, N., Anand, N. (2025). Optimizing Modulation Schemes for 5G Efficiency Networks. In: Ghonge, M.M., Liu, H., Khan, M., Tran, T.A. (eds) Advances in Emerging Technologies and Computing Innovations. ICETCI 2025. Sustainable Artificial Intelligence-Powered Applications. Springer, Cham. https://doi.org/10.1007/978-3-031-92854-3_39

[20]Kumar, N., Asmita, Kaushik, S., Soni, R.S., Verma, R., Anand, N. (2026). Experimental Results Region-Based Convolutional Neural Network Algorithm for Deep Face Detection. In: Udgata, S.K., Mohapatra, D., Sethi, S., Rana, M.E. (eds) Intelligent Systems. ICMIB 2025. Lecture Notes in Networks and Systems, vol 1624. Springer, Cham. https://doi.org/10.1007/978-3-032-05117-2_26

[21] Anand, N. (2014). The Conceptual Modeling of ETL Processes. Journal of Global Research in Computer Science, 5(1), 16-18.

[22] Anand, N., & Sharma, P. (2014). Data Warehouse Security through Conceptual Models.

[23] Anand, N. (2012). An Overview on Physical Implementation of Secure ETL Workflow. Journal of Global Research in Computer Science, 3(10), 43-45.

[24] Nitin Anand, Vatsala Sharma, Pardeep Singh (2025); ETL and Data Warehousing: Architecture, Vulnerabilities, and Security Mechanisms; International Journal of Scientific and Research Publications (IJSRP) 15(10) (ISSN: 2250-3153). DOI:http://dx.doi.org/10.29322/IJSRP.15.10.2025.p16612

[25] Anand, P. S. N. (2014). Framework for The Integrated And Validated Model of Data Warehouse. American Journal of Engineering Research (AJER), e-ISSN, 2320-0847.