# Data Quality Issues and Metadata Repository of Data Warehouse

**Ashish Kumari[1], Navdeep Bohra[2], Pardeep Sangwan[3], Dinesh Sheoran[4], Sandeep Kumar[5]**

*[1]Department of IT, Maharaja Surajmal Institute of Technology, New Delhi, India,*
*[2]Department of CSE, Maharaja Surajmal Institute of Technology, New Delhi, India,*
*[3]Department of ECE, Maharaja Surajmal Institute of Technology, New Delhi, India,*
*[4]Department of ECE, Maharaja Surajmal Institute of Technology, New Delhi, India,*
*[5]Department of CSE, Maharaja Surajmal Institute of Technology, New Delhi, India*

**Abstract:** We can characterize Data warehouses in the form of subject-arranged, coordinated, time-fluctuating, non-unstable assortments of information that can be utilized in a hierarchical manner. These days, data warehouses turned into a significant methodology to incorporate heterogeneous data sources for the associations, and are used to empower OLAP (On-Line Analytical Processing). Regrettably, not the capacity interaction and the collection appear to be totally dependable. Because of the foremost job of a Data warehouse (DW) in generating methodology choices, quality of data warehouse is the main thing for associations. Thus, we have to utilize strategies, designs, methods, apparatuses etc. to assist us in planning and keeping up with excellent DWs. Over the most recent couple of years, there have been a few ways to deal with plan DWs from the conceptual, logical and physical perspectives. This paper presents an overall working of metadata repository for data warehouse metadata. The repository design is madusesast two metadata layers and three methods or perspectives. We conclude these three perspectives models and two layers to measure the issue of DW quality and understandability.

**Index Terms:** Data Warehouse (DW), Metamodel, Conceptual, Logical and Physical Model, Quality Metrics

## 1. INTRODUCTION

A Data Warehouse (DW) is an assortment of innovations pointed toward empowering the information labourer (leader, administrator, expert and so on) to make better and quicker decisions. In the present era, it is the information that assumes a pivotal part in the progress or failure of any kind of framework. Data warehouses contain big repositories for collection of data or information. Data in data warehouses are based on the organization's decision, thus assuring the data warehouse's quality is crucial [1][2]. An effective data warehouse framework gives significant data which can examine the previous trends, relate with the present working and anticipate real modern patterns. It is the nature of the data warehouse that chooses the achievement or disappointment of frameworks [3]. Data warehouses have to deliver high-quality data as well as high-quality services. Coherency, newness, exactness, openness, accessibility and execution all these are quality components needed by the end clients of the DW.

The association of paper is as per the following: Section 2 portrays the previous research in DW with a basic architecture. Section 3 provides an outline of the fundamental two instantiation layers utilized in DW for the aggregation or accumulation, integration and customization for the quality information. Segment 4 relation of Quality with DW. Segment 5 metamodel Framework with quality metamodel. Section 6 Data Quality and tools used for DW. Section 7 Quality Metamodel with GQM and Quality Metrics.

## 2. ARCHITECTURE

Officially, the architecture of DW may be described as layers of eventuate sees in the stacked manner that one upon another. A data warehouse architecture depicts various levels of data or information, where information from one layer is derived from information from a lower layer, or where one layer's output becomes the input for the next lower layer.

**Operational Databases**: Data sources from the lower layer of the model, additionally called operational databases. They might comprise structured or organized keep inaccessible database frameworks and inheritance frameworks, or empirical data stored in records.

**Global Data Warehouse**: The global (or fundamental or principal) Data Warehouse is described in the architecture's core layer. The global data warehouse maintains an accurate information about the output calculated using the aggregation, integration and evaluation of point-by-point values from many data sources. An operational Data Store is a type of data store that stores unreliable, moderate data and is used to combine data from many sources (ODS). This operational store also acts as a temporary storage area for the modified information and also for clean out the dirty information, to ensuring that the main storage centre contains only clean and error-free data.

**Client Warehouses**: Furthermore, local or client warehouses, the next tier of architecture, which includes profoundly accumulated information, straightforwardly obtained from the worldwide distribution center or warehouses. There are different sorts of local warehouses, like the information stores or the OLAP data sets, which might utilize social data set frameworks or explicit multidimensional information structures.

Each and every component, data and process of a data warehouse are basically ought to be followed and directed from a metadata store.

**Metadata Repository**: Repository is a common word for the dataset that has been separated for both the purpose of information tracking and reporting. It is used as a guide or hints for the designer as well as administrator of a data warehouse. To be sure, the data warehouse is an extremely intricate framework, the quantities of recorded data are enormous, and the processes for extracting, transforming, purifying, storing, and aggregating it are many, volatile, and dependent on time. The metadata repository is utilized a roadmap, providing the history including all changing data on its design and components as well as a trace of all plan decisions. The fundamental design of a data warehouse is depicted in Figure 1.

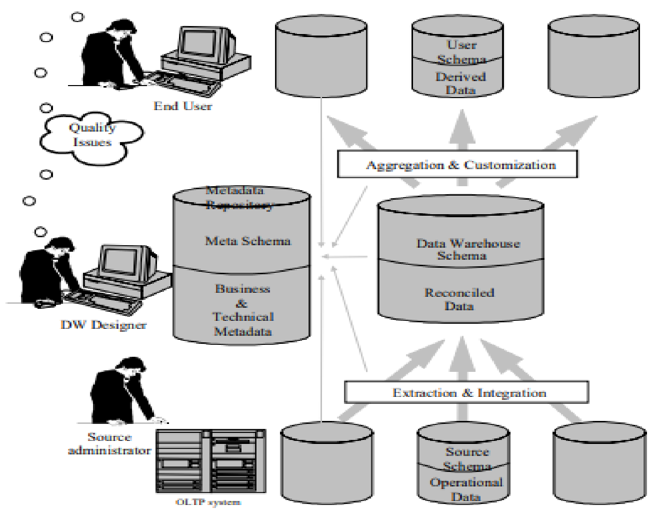



Fig. 1. Architecture of DW

Data Repository contains following things:

1. Data warehouse design, place of data, hierarchy structure and view of data.
2. Monitoring, linking (previous data to currently updated data) and activation of data.
3. Algorithm for the data combining and summed up, that can provide final conclusion.
4. Mapping of currently working data with data warehouse that have all the information about required data.
5. Improve the performance of warehouse.

DW have demonstrated their worth to fill in as repositories for incorporated, homogenized and clean data. We can say that they don't serve just as data storage for responding to complex inquiries rapidly yet in addition as moderate stages in the handling of data inside the data arrangement of a venture, where the data turns out to be more exact and valuable. Consequently, at last, the ultimate quality of the data that is delivered to the end client is determined by the information preparation chain, which is located on the front side of an OLAP application.

## 3. LITERATURE

Notably, a definitive objective of associations, states and organizations, when they amass and store data is the capacity to handle it later and exploit DW.

Key choices for storing the data in data warehouses are based on the organization's decision, thus assuring the data warehouse's quality is crucial. One way to confirm the quality of designs, which will further use in their design model, possible with the help of process nature such as all three models (conceptual, logical, and physical models) [1].

Set up a causal connection between the interior quality ascribes dictated by underlying properties like complexity, size, coupling, and so forth and the outer quality credits like maintainability, analyzability, and so on. According to this relationship, specialists have measured the nature of conceptual schemas based on this connection by offering measurements or metrics based on their structural features [4]-[8].

OLAP is used for dealing with user queries and analyzing the information on outer level as well as having an effective problem-solving system and for managing the metadata warehouse in online mode to understand online schemas [9]-[13].

Information quality can be estimated by different boundaries. Past work gives various groupings of the data quality measurements [14]. Most of the time referenced measurements are timeliness, exactness, completeness, and consistency. All these measurements managed using the empirical model of DW. It provides many mathematical calculations, modelling and simulation techniques to check the boundary and develop more systems [17].

Basically, quality of data is strongly based on the way of representation or reengineering of data as well as quality of DW. Companies believes on the DW for storing the important information that can manage it in proper way, keep that secure and make the decisions easy [2][18]-[20].

Information such as the amount of unique or absent values in such a field, data types of characteristics, or occurrence trends with related recurrence are all examples of what may be learned by relevant to the development [21].

Data warehouse developing in both the educational sector and industry. Both the sectors have different requirements and priorities. Industry gives preference to the quantity of data that produce statistical patterns from the huge data warehouse and ready to compromise with the quality of data, whereas the academic sector focus on the quality of data to understand the analytics [22]. Data quality factor is very important factor that manage the data warehouse. To evaluate the quality factor, Goal-Question –Matrix is used. That contains some question related to the concern data and get responses from the user [23][2].

Speculative parallelization [24]-[25], a technique that enables processors to continue executing subsequent instructions without waiting for the previous instruction to complete, can improve the performance and efficiency of parallel computing systems, especially in data-intensive applications. It can help improve the efficiency and scalability of data warehouses, making them better suited for handling large amounts of data, which is relevant to the data quality issues and metadata repository of data warehouses.

Blockchain is a distributed ledger technology that has the potential to revolutionize the way data is stored and shared. Its inherent security features, such as immutability and decentralization, make it a promising technology for improving the data quality and security of data warehouses. This technology can provide a tamper-proof and transparent record of data changes and transactions, which can improve the accuracy and integrity of data stored in data warehouses. Furthermore, the use of blockchain technology in data warehousing can help address the issue of data privacy by providing a secure and decentralized way of storing and sharing data [26]-[28].

Using a variety of methods, data profiling is defined as the analysis of a dataset to acquire metadata. Thus, understanding a dataset is a necessary step before performing either data quality measurement or monitoring. In general, the use of neural networks [29]-[30] and machine learning methods [34]-[35] can help improve the accuracy and completeness of data stored in data warehouses by providing advanced pattern recognition and data analysis capabilities.

Automated machine learning [36] is another method that helps address the data quality issues in data warehouses by automating the data cleaning and preprocessing steps. The use of AutoML helps improve the accuracy and reliability of data analysis, allowing organizations to make more informed decisions based on their data.

Data warehouses and big data [37]-[38] are related in that both are used to store and manage large volumes of data for analytical purposes.

## 4. QUALITY AND DATA WAREHOUSES

The issue of information quality is on a very basic level entwined in “how” the clients utilize the information in the framework since the clients are definitive adjudicators of the nature of the information delivered for them: if no one utilizes the information, no one will at any point take care to work on its quality.

Then the main question, at that point, is how to sort out the plan, organization and advancement of the DW so that all the unique, and some of the time-restricted, quality necessities of the clients can be at the same time fulfilled. Another point is the way to focus on these necessities to fulfil them in terms of their importance. This problem is usually illustrated by the DW's real strategy where the issue is to track down a bunch of emerged sees that streamline client demands reaction time and the worldwide data warehouse upkeep cost simultaneously.

This problem is usually illustrated by the DW's real strategy, in which the problem is to track down several newly discovered views that streamline client demand response time while also lowering the global data warehouse maintenance cost.

Some of the main terms are Interpretability, accessibility, availability, timeliness, usefulness, believability, etc. used for the management of quality.

## 5. METAMODEL FRAMEWORK

Our idea depends on the understanding that the metadata store of the information distribution centre can fill in as the cockpit for the administration of value. We should give a cognizant system that catches the DW from three distinct perspectives, specifically the design or architecture, process and quality perspectives, to make the store utilitarian or functional.

Why to make an architectural model? An Architecture model relates to a mapping construction of such a meta-information base that regulates the typically dispersed as well as heterogeneous arrangement of DW parts and in this way is the fundamental beginning stage for plan and functional improvement. In light of existing approaches and research, the models were created in order to pass on a powerful, statistically specified, and technically understandable conceptual displaying language. The metadata diagram's expressiveness and administration are critical for DW quality.

Why to design a process model? The static representation of the design sections of the DW, such as data warehouse operations, can be enhanced by developing the dynamic aspects of the warehouse. Giving information distribution centres a process model effectively captures the framework's behaviour as well as the intricate stakeholders' interdependencies.

Why to use a quality model? DW centres may also act as a bridge between subjective customer requirements on data quality and analytical detection and assessments of information gaps. Because the DW centre is a multi-layered, group-connected structure, it may act like a "data cleaner" for something like the data presented to the client. By storing the nature of a DW information and processes in a metadata repository is giving, in this way, the DW partners with adequate ability to assess, comprehend and perhaps respond against the construction and data of the warehouse.

The structure necessitates metadata classification in two implementation layers and three views or perspectives. The metadata layer contains its true meta-data for a given DW, whereas the metamodel layer contains the metadata repository's blueprint. The conceptual, logical, and physical perspectives are very famous by nature, in the field of data set and data frameworks.

### 5.1. STRUCTURE OF QUALITY METAMODEL

The design cycle of the DW is empowered through the investigation of main three perspectives concerning the data warehouse i.e., design, processor cycles, and quality. The architecture defines the data warehouse's static components, while the process captures a dynamic aspect of an information centre's environment. At last, quality is a proportion of the satisfaction of the assumptions for the elaborate stakeholders in such a climate. These three metamodels regard a coherent framework and fit effortlessly with one another. Metadata keep away in a repository, where it very well may be gotten to from each part of the DW.

A storage unit or repository is used to store Metadata, from where the metadata can be accessed by each and every component that presents in the data warehouse.

The repository could be an actual area or a virtual data set that pulls metadata from a few sources. Metadata can include a variety of items, like how to get explicit information or more data about it. The need for a three-point-of-view metadata schema in a Data Warehouse is contested [23]:

1. The conceptual model perspective is created around the undertaking model of the association.

2. The logically correct perspective covering all the representations of the outline model of the Information Centre. It handles “what” type of queries on the process structure.

3. The physical point of view is addressing the capacity and properties of execution of the Data Warehouse parts. Physical model handles “how” queries for the processing of model.

Every one of these points of view, and the relationship between them, are symmetrically connected to the three established layers of DW, specifically sources, information centre (DW), and customers. Figure 2 shows the generated metamodel.

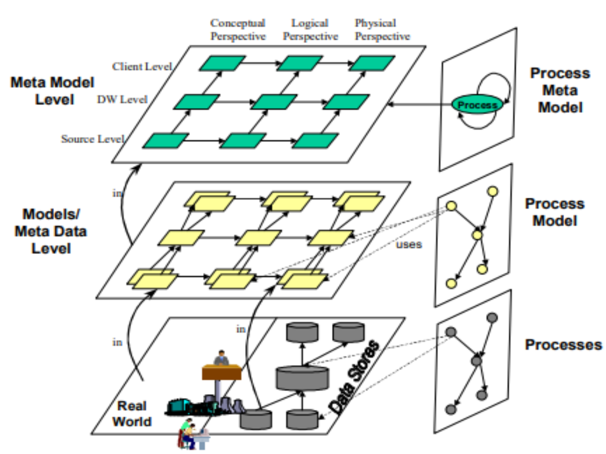
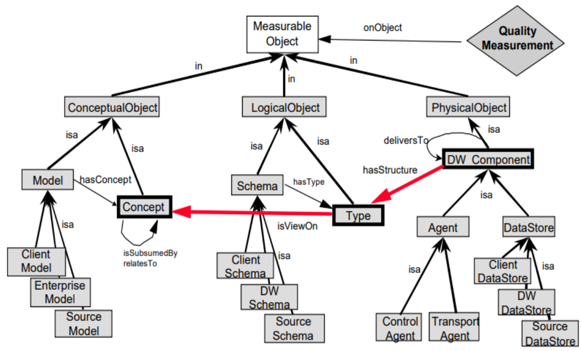



Fig. 2.[6] The Metadata Framework

A high-level item is a Measurable Object. It groups objects in any perspective i.e. physical, logical, or conceptual, and at any level of DW, customer, or source. Inside every perspective, we found and understood the modules it offers (for example customer model) and the sorts of data found inside these modules. Horizontal level used in the data structures are mainly constructed with three parts of subclasses attached with system, the design and the Data Store in the horizontal levels, each for a different point of view. The two horizontal connections that are, is View On and has Structure set up the connection as to how the flat connections in Figure 3 are deciphered: all types of relational or design that is multidimensional are characterized as the logical perspectives on the concepts inside the conceptual perspectives. Then again, the parts of the actual point of physical perspective get a pattern according to the legitimate viewpoint as their diagram. And Quality Measurements focuses on the object's material, means to get some valuable stuff from the processing. Each object can have an associated set of materialized views




called Quality Measurements.

Fig. 3. Structure of the Meta Model Repository [29]

Every item in Figure 3 is a meta-class: logical schemas, conceptual models and DW pieces are all addressed in the meta-data set as occurrences of each. In reality we can see all these three levels in Figure 2 and can easily differentiate them. The metamodel (for example, on top layer in Figure 2) provides documentation to the DW for nonexclusive aspects such as construction and customers, as well as from a commercial standpoint. That metamodel is started with the DW's metadata (the inner layer of Figure 2), for example: relational schemas detailed information or a visual representation for conceptual model. This is addressed by the most basic layer in Figure 2. The most basic level in Figure 2 addresses the current reality, where actual cycles and information exist: at this level, metadata are instantiated with information, such as the data sets of a correlation or even the objects of this current reality addressed first by conceptual model components.

Moreover, companies have countless heterogeneous information origins and very hard to explain a general image for which type of information is accessible from the main origin as well as for monitoring the interrelations of the information origins. Subsequently, we can see the design of DW through two ways that are: first, contains the conceptual point of view and second, a perfect detachment between areas or defining the levels for data origins, DW and at last level for customers. The conceptual venture model is the internal object of the designed metamodel portrayed in Figure 3, which gives a conceptual description of the information in the undertaking. As far as this business model is concerned, most of the available data inside the sources and indicated by the DW's clients is conveyed.

## 6. DATA QUALITY

Quality is referred to as necessary and should not only include features of the internal data itself but also user data testing [22]. Analysis of data, design part of data products and developments for data manufacturing systems are three major issues that generates data quality issues as well as data quality dimensions.

Data quality policy is the general expectation and bearing of an association for issues concerning the nature of information items. Information quality management is the administration work that decides and carries out the information quality policy. A data quality framework envelops the hierarchical design, obligations, strategies, process, and required data for carrying out information quality management.

The strategy manages information quality prerequisites assortment and definition. The parameters quality is at first treated as client requirements (subjective) and afterwards transferred in the technical solutions.

Basically, the most well-known instances of filthy information are: Format contrasts, Information concealed in freestyle text, Violate the integrity rules, Missing numbers, Schema contrasts. The embedded information can be assembled in the accompanying classifications: Data Transformation, Data Fetching, Data Cleaning, Data Transition, Data Quality Analysis with Scrubbing etc. These classification phases compromise with the quality of data. At every phase data is changed by all the previous layers. Considering the precautions taken, there is still a portion of data which is not maximized the quality of data in Data Warehouse.

**Data Quality issues can generate at some places:**

i. Data procedures not handled properly;

ii. Cannot follow a specific method for processes and procedures;

iii. Inability of data entry, as well as maintenance in DW;

iv. Process may have lost during the system change;

v. New data which is yet not correlated;

vi. Perform of data handling is very low;

vii. Outsider data cannot easily have arranged with any organization DW.

**Tools for Data Quality**

Every warehouse system requires ETL tool for implement the data quality. That provide the usability, assurance and functioning of data, usability in data quality factor. ETL stands for Extraction, Transformation and Loading [30], shown in figure 4. This is used to transfer the data from one side to another side. Main work of this ETL is integration of data and data warehousing for business purpose.

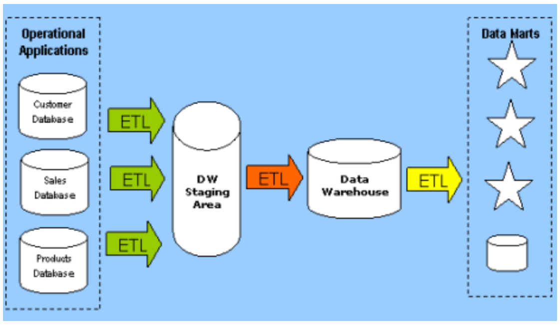



Fig. 4. Data Warehouse Environment

There are many quality tools some of Quality tools can be separated by following points:

Cleansing Tools: Cleansing process is used to clarify the data as well as remove anomalies form the data. Also provide the parsing, changing capability and match the information with standard criteria.

Predicting Tools: Data styles, verification and matching data with already existing data rules for the company.

Data Uploading and Downloading Tools: to convert the data from one system to another or can say provide the portability. Data Migration, Data Quality Analysis tools etc. are many more features.

The following factors were used to determine whether or not a given tool met the requirements for consideration [41]:

- re-employed for superior performance in the past
- found throughout the investigation
- discovered by chance.

Some of Data quality tools are:

Statistical Analysis System (SAS): It combines all the quality of data in seamless way with the existing data and new entered data. Data integration and profiling of data handled with this tool. All the statistical calculation performed with SAS, like removing redundancies, semantics correctness, remove errors, analyze all type of stored data; and create a reliable data store.

ORACLE 10G: It is a DW builder tool, that provide huge information management, provide online support and final report management.

SAP: It provides data warehousing, acquiring of information, scheduling of architecture, calculate data on internet, intellectual calculation of data with analysis, using already existing data to generate a new structure.

TIBCO: It provides fast decision making and real-time data process faster. TIBCO used for correlation between large volume of data as well as for the pattern identification.

These tools improve the information correctness when you want to generate information on source side, to fetch the data before the final storage and change in post during storage in DW. Somehow today's analysis tools are also failed in integration process. This problem is still existing in the data quality area.

The Data Management Association defines data quality management as assessing, enhancing, and guaranteeing data accuracy. Numerous DQ strategies, such as Total Data Quality Management (TDQM), A Methodology for Information Quality Evaluation (AIMQ) [41][42], Hybrid Information Quality Management (HIQM) [42], Comprehensive Data Quality (CDQ) [42], Data Quality Assessment (DQA) [42], DQ assessment methods [41]-[42], etc. have been proposed over the years. All data go through data profiling, measurement, cleaning, and monitoring. In the presence of management associations, management, and new tools, all are facing data quality issues.

Quality issues generate in each area, such as geolocation records [43], shopping records, online payment [43], electronic health records [45], etc.

## 7. QUALITY METAMODEL

In the wake of introducing the metadata framework and the design metamodel, we continue to talk about the quality of metamodel. Metamodel methods are generally founded on human expertise participation and measurable models for ad-hoc issue settling. The quality metamodel's proper prerequisites can be summed up as follows: (1) metadata structure's orders that we have taken on ought to be regarded. Accordingly, each of the three viewpoints and instantiation layers ought to be precisely present at the quality metamodel. It would increment the accessibility of the methodology as well as re-ease of use of the all-accomplished arrangements. (2) Moreover, the violation of value ought to be done using a repository, empowering subsequently the concentrated assortment, storage and questioning of all important meta-data in a predictable manner. (3) Ultimately, the metadata store should permit the violation of the elaborate quality elements, using well developed computerized procedures and calculations or algorithms.

### *7.1. GOAL-QUESTION-METRIC (GQM)*

The solution for this issue expands on the broadly utilized Goal-Question-Metric (GQM) way to deal with programming quality administration, figure 5. Some queries posed by the clients are normally not answered straightforwardly but display with metrics using either the item or process being referred to; explicit procedures and algorithms are then performed to determine the appropriate response of an inquiry from the matrices.

A group of inquiries is utilized to describe how the appraisal of a particular objective will be calculated that is dependent on any portraying design. Questions attempt to show the main motive of the Metrix (such as process, asset, and item) as for a chosen issue of quality to its quality decision from the chosen perspective.

Metrics is an information group, related to each question to address it quantitatively. The information can be the type:

1. Objective, which defines the program's size, precious working time on an undertaking, and different types of variants related to a record).
2. Subjective which defines the level of client fulfilment, the intelligibility of a text)

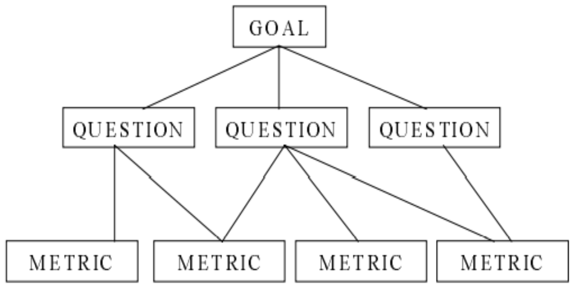



Fig. 5. Goal, Question, Metric

Processing steps of GQM:

i) Define the set of rules and goals of the project for a group or company that satisfy the end users with target and better performance.

ii) Generate questions based on goal criteria for checking the goal achievement.

iii) Measurements metric that collect answers from derived questions.

iv) To invent a mechanism that can collect data and verify that.

Goal coordinates are: target achieved on time, objective fulfillers, a group or company views and the purpose (that can improve the target goal).

Question coordinates about: processing speed, performance, comparisons between processes, manager view point, improvement in performance.

### *7.2. QUALITY METRIC*

A portrayal of the activity performed, made to accomplish a bunch of objective or subjective information which answer an inquiry quantitatively [46][48].

Data model quality based on Figure 1:

**Goals of conceptual quality**: The stakeholders straight communicate the quality objectives, in their high-level, subjective jargon, including abstract necessities.

**Measurements of physical quality**: The estimations are performed by the explicit algorithms or apparatuses, against the explicit DW objects, at explicit time.

**Questions of Logical quality:** The questioner part is the intermediate point between the conceptual level and physical level. That concrete the client requirements and the substantial, physical estimations and activities.

Five principle stages go from establishment of aims and concepts through metric implementation, validation and retiring i.e. recognizable proof, Creation, Metrics definition, Theoretical approval and Empirical approval.

### *7.3. QUALITY FACTOR*

Electron Various process steps, which runs during the execution of an operation, can differentiate or analyze the output by using various mathematical formulas and analyzing techniques of quality metrics. The accuracy of such stored information is factored into the evaluation model. As a result, these are:

1. Architecture and Governance
2. Development and Assessment of Software
3. Data Storage
4. Use of data
5. Quality level of data

These above factors have some operations on each level of DW. The Quality model not define only for the DW and stakeholders but is also used to define the data directory. Architecture and Governance include the design and data evaluation factors used to observe other quality factors. Development of software is a difficult and time taking process and the Assessment of software testing the design according to the data quality factors as well as keeping track for all operations. Data is stored in DW using a batch system and also take care about new updates. Use of data defines the uses and query on the data, accessibility, access and utility of data. Finally, the Quality of data provides the level of data in the field of Correctness, Authenticity, stability etc.

## 8. PROBLEMS WITH QUALITY OF SERVICES

Supporting all data quality requirements for DW users demands a data quality system of measurement which is intended for match their demands to the attributes. Several past research represent that all types or stakeholder requirements are different for DW [46]-[49]. History depending on the machine code which is automated by nature, that's why it makes a difference between user and the quality of data. User cannot assume and see the transparency about the data warehouse and integration of data is still not up to mark also. This problem is still existing in the data quality area.

## 9. CONCLUSION AND FUTURE SCOPE

In this paper, we introduced the procedure in the data repository that describes the DW metadata of the overall system. The structure needs grouping of metadata at some place, like two instantiation layers and three perspectives, and how they are associated with one another. The metamodel layer make the comprised outline for the metadata repository and for the metadata layer makes real metadata for specific information distribution center. We connected this structure to an obvious methodology for the design of the DW. Then, at that point, a quality metamodel, that talks about broadly acknowledged Goal-Question-Metric methodology for use of quality administration in data frameworks. It is very vast in itself. The deliberate literature survey has shown that impressive examination exists in the information quality evaluation field, information quality administration field, and observing field. Meeting the information quality necessities of the data warehouse stakeholders requires an information quality estimation framework that is intended to adjust their prerequisites to the information quality measurements. A large number of the current information quality estimation apparatuses are planned as automated or computerized frameworks, with predetermined boundaries, leaving the interface with the stakeholders. Because of these limits, the issue of information quality persists.

## References


[1] Manuel Serrano, Juan Trujillo, Coral Calero, Mario Piattini "Metrics for data warehouse conceptual models understandability" Information and Software Technology 49 (2007) 851-870.

[2] Calero C, Piattini M, Pascual C, Serrano M.A. "Towards Data warehouse quality metrics", In 3rd International workshop on design and Management of Data warehouses (DMDW 2001), Interlaken, Switzerland;2001.

[3] Naveen Dahiya, Vishal Bhatnagar, Manjeet Singh "A Fuzzy Based Matrix Methodology for Evaluation and Ranking of Data Warehouse Conceptual Models Metrics" The International Arab Journal of Information Technology, Vol. 15, No. 2, March 2018.

[4] Gargi Aggarwal, Sangeeta Sabharwal, Sushama Nagpal "Theoretical and Empirical Validation of Coupling Metrics for Object-Oriented Data Warehouse Design" Arab J Sci Eng DOI 10.1007/s13369-017-2692-y, 24 July 2017.

[5] Bailey V. , Rawal B.S. (2022) Challenges of the System and Network Administration. Data Science Insights Magazine, Insights2Techinfo, Volume 1, pp. 1-6. 2022.

[6] Sharma A., Chhabra A., (2022) Big Data: The Future of Information Management, Data Science Insights Magazine, Insights2Techinfo, Volume 1, pp. 21-24. 2022.

[7] Mishra A, (2022) Analysis of the Development of Big data and AI-Based Technologies for the Cloud Computing Environment, Data Science Insights Magazine, Insights2Techinfo, Volume 2, pp. 9-12. 2022.
[8] Suraj Juddoo , “Overview of data quality challenges in the context of Big Data”, 2015 International Conference on Computing, Communication and Security (ICCCS).
[9] M. Jarke, M.A.Jeusfeld, C. Quix, P. Vassiliadis “Architecture and quality in data warehouses: An extended repository approach. Information Systems” 24(3), pp. 229-253, 1999. (a previous version appeared in Proc. 10th Conference of Advanced Information Systems Engineering (CAiSE ’98), Pisa, Italy, 1998)
[10] Shanmugam R, Rafsanjani M K, (2022) Time Series Database (TSD): A Approach for Time Series Data Management, Data Science Insights Magazine, Insights2Techinfo, Volume 2, pp. 17-20. 2022.
[11] Gupta, B. B., Yamaguchi, S., & Agrawal, D. P. (2018). Advances in security and privacy of multimedia big data in mobile and cloud computing. Multimedia Tools and Applications, 77, 9203-9208.
[12] Adil At-taibe; Badreddine El Mohajir “From ER models to multidimensional models: The application of Moody and Kortink technique to a University information system”, Third IEEE International Colloquium in Information Science and Technolology (CIST),2014.
[13] Harco Leslie Hendric Spits Wamars, R. Randriatoamanana, “Datawarehouser: A data warehouse artist who have ability to understand data warehouse schema pictures” IEEE Conference (TENCON), 2016.
[14] C Batini, M Scannapieco “ Data and information quality” - Cham, Switzerland: Springer International, 2016.
[15] Xu, Z., He et al. (2021). Certificateless public auditing scheme with data privacy and dynamics in group user model of cloud-assisted medical wsns. IEEE Journal of Biomedical and Health Informatics.
[16] Gupta, B. B., & Quamara, M. (2020). Internet of Things Security: Principles, Applications, Attacks, and Countermeasures. CRC Press.
[17] Li Cai, Yangyong ZhuThe,” Challenges of Data Quality and Data Quality Assessment in the Big Data Era”, 2016.
[18] Cipriano Forza “Survey Research in Operations management: A Process-Based Perspective”, International Journal of Operations & Production Management 22(2):152-194, February 2002.
[19] A. Abello´, J. Samos, F. Saltor, A framework for the classification and description of multidimensional data models, in: 12th International Conference on Database and Expert Systems Applications (DEXA’ 01), Springer-Verlag, Munich (Germany), 2001.
[20] R Jindal, S Taneja, “Comparative study of data warehouse design approaches: a survey”, International Journal of Database Management Systems ( IJDMS ) Vol.4, No.1, February 2012.
[21] Ehrlinger, L., & Wöß, W. (2022). A survey of data quality measurement and monitoring tools. Frontiers in Big Data, 28.
[22] A. Abello´, J. Samos, F. Saltor, YAM2 (Yet Another Multidimensional Model): An Extension of UML, in: International Database Engineering and Applications Symposium (IDEAS 2002), IEEE Computer Society, Edmonton (Canada), 2002, pp. 172-181.
[23] An Overview of Data Warehousing and OLAP Technology, Sweta Singh, P. Malhan, Published 2014.
[24] Kumar, S., Singh, S. K., Aggarwal, N., Gupta, B. B., Alhalabi, W., & Band, S. S. (2022). An efficient hardware supported and parallelization architecture for intelligent systems to overcome speculative overheads. International Journal of Intelligent Systems, 37(12), 11764-11790.
[25] Kumar, S., Singh, S. K., Aggarwal, N., & Aggarwal, K. (2021). Evaluation of automatic parallelization algorithms to minimize speculative parallelism overheads: An experiment. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1517-1528.
[26] Singh, S. K., Sharma, S. K., Singla, D., & Gill, S. S. (2022). Evolving Requirements and Application of SDN and IoT in the Context of Industry 4.0, Blockchain and Artificial Intelligence. Software Defined Networks: Architecture and Applications, 427-496.
[27] Singla, D., Singh, S. K., Dubey, H., & Kumar, T. (2021, December). Evolving requirements of smart healthcare in cloud computing and MIoT. In International Conference on Smart Systems and Advanced Computing (Syscom-2021) (pp. 102-109).
[28] Sharma, S., & Singh, S. K. (2022). IoT and its uses in Security Surveillance, Insights2Techinfo, pp.1.
[29] Kaur, P., Singh, S. K., Singh, I., & Kumar, S. (2021, December). Exploring Convolutional Neural Network in Computer Vision-based Image Classification. In International Conference on Smart Systems and Advanced Computing (Syscom-2021).
[30] Singh, I., Singh, S. K., Kumar, S., & Aggarwal, K. (2022, July). Dropout-VGG based convolutional neural network for traffic sign categorization. In Congress on Intelligent Systems: Proceedings of CIS 2021, Volume 1 (pp. 247-261). Singapore: Springer Nature Singapore.
[31] Quamara, M., Gupta, B. B., & Yamaguchi, S. (2019, October). MQTT-driven remote temperature monitoring system for IoT-based smart homes. In 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE) (pp. 968-970). IEEE.
[32] Gupta, B. B., Gupta, S., & Chaudhary, P. (2017). Enhancing the browser-side context-aware sanitization of suspicious HTML5 code for halting the DOM-based XSS vulnerabilities in cloud. International Journal of Cloud Applications and Computing (IJCAC), 7(1), 1-31.
[33] Peñalvo, F. J. G., Sharma, A., Chhabra, A., Singh, S. K., Kumar, S., Arya, V., & Gaurav, A. (2022). Mobile cloud computing and sustainable development: Opportunities, challenges, and future directions. International Journal of Cloud Applications and Computing (IJCAC), 12(1), 1-20.
[34] Peñalvo, F. J. G., Maan, T., Singh, S. K., Kumar, S., Arya, V., Chui, K. T., & Singh, G. P. (2022). Sustainable Stock Market Prediction Framework Using Machine Learning Models. International Journal of Software Science and Computational Intelligence (IJSSCI), 14(1), 1-15.
[35] Singh, I., Singh, S. K., Singh, R., & Kumar, S. (2022, May). Efficient loop unrolling factor prediction algorithm using machine learning models. In 2022 3rd International Conference for Emerging Technology (INCET) (pp. 1-8). IEEE.
[36] Mengi, G., Singh, S. K., Kumar, S., Mahto, D., & Sharma, A. (2023, February). Automated Machine Learning (AutoML): The Future of Computational Intelligence. In International Conference on Cyber Security, Privacy and Networking (ICSPN 2022) (pp. 309-317). Cham: Springer International Publishing
[37] A. Sharma et al., "Fuzzy Based Clustering of Consumers' Big Data in Industrial Applications," 2023 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2023, pp. 01-03, doi:

10.1109/ICCE56470.2023.10043451.
[38] Singh, A., Singh, S. K., & Mittal, A. (2022). A Review on Dataset Acquisition Techniques in Gesture Recognition from Indian Sign Language. Advances in Data Computing, Communication and Security: Proceedings of I3CS2021, 305-313.
[39] Vaishali A. Kherdekar, Pravin S. Metkewar "A Technical Comprehensive Survey of ETL Tools", International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 4 (2016) pp 2557-2559
[40] Dr. Suman Mann and Meena Siwach, "Data Model Quality Metrics of Data Warehouse: A Survey", International Conference on Innovative Computing and Communication (ICICC 2020).
[41] Zong, W., & Wu, "The Challenge of Data Quality in the Big Data Age" Journal of Xi'an Jiaotong University Social Sciences 335, 2013, pp 38-43
[42] Matthias Jarke, Manfred A Jeusfeld, Christoph Quix, Panos Vassiliadis, "Architecture and quality in data warehouses: An extended repository approach", information system, 1999.
[43] An overview of data warehousing and OLAP technology, S. Chaudhuri, U. Dayal Published 1997.
[44] Pickering, D., & Blaszczynski, A. (2021). Paid online convenience samples in gambling studies: Questionable data quality. International Gambling Studies, 21(3), 516-536.
[45] Cichy, C., & Rass, S. (2019). An overview of data quality frameworks. IEEE Access, 7, 24634-24648.
[46] Bähr, S., Haas, G. C., Keusch, F., Kreuter, F., & Trappmann, M. (2022). Missing data and other measurement quality issues in mobile geolocation sensor data. Social Science Computer Review, 40(1), 212-235.
[47] Jain, A. K., & et al (2018). Two-level authentication approach to protect from phishing attacks in real time. Journal of Ambient Intelligence and Humanized Computing, 9, 1783-1796.
[48] Cappiello, C., Francalanci, C., & Pernici, B. (2004) Data quality assessment from user 's perspective. Procedures of the 2004 International Workshop on Information Quality in Information Systems, New York: ACM, pp 78-73.
[49] Cook, L. A., Sachs, J., & Weiskopf, N. G. (2022). The quality of social determinants data in the electronic health record: a systematic review. Journal of the American Medical Informatics Association, 29(1), 187-196.