Frontiers in Computer Science and Artificial Intelligence
ISSN: 2978-8048
DOI: 10.32996/fcsai
Journal Homepage: www.al-kindipublisher.com/index.php/fcsai

FCSAI
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Designing High-Performance OLAP Cubes for Advanced Analytical Decision-Making

JAGADEESWAR ALAMPALLY
Software Development Manager, USA
Corresponding Author: JAGADEESWAR ALAMPALLY, **E-mail**: jagadeeswar.alampally81@gmail.com

| ABSTRACT

The growing size of enterprise data requires analytical systems that can facilitate effective decision-making. Online Analytical Processing (OLAP) cubes have the advantage of allowing multidimensional analysis but have challenges in scalability and latency with handling heterogeneous data. The framework to be proposed in this paper is based on developing high-performance OLAP cubes, combining dimensional modeling optimization, scalable structuring, and contemporary storage approaches. However, the strategy enhances the responsiveness and decision capability of analytics by matching cube architecture and performance evaluation metrics. The article offers insights into the balance between scalability, usability and computational efficiency in the modern analytics setting.

| KEYWORDS

OLAP Cubes, Data Warehousing, Decision Support Systems, High-performance Analytics, Multidimensional Modeling, Big Data Processing

| ARTICLE INFORMATION

ACCEPTED: 01 May 2022 **PUBLISHED:** 25 June 2022 **DOI:** 10.32996/fcsai.2022.1.1.4x

I. Introduction

Increased organizational data has raised the pressure on analytical infrastructures that can transform massive amounts of data into insight that can be acted upon. Online Analytical Processing (OLAP) facilitates this objective by means of multidimensional aggregation, slicing and hierarchical analysis in decision-support systems [9]. OLAP allows the planning, monitoring, and predictive assessment of an enterprise by allowing interaction with aggregated data structures that are already aggregated.

Scalability and performance issues such as storage overheads and query latency are however introduced by increasing dimensionality and data heterogeneity. The industrial applications demonstrate the necessity of cube optimization to enable the complex analytical tasks to process huge volumes of operation data [3]. Enterprise analytics solutions also reflect dependency on OLAP architectures that are combined with novel visualization and cloud intelligence solutions to enable decisions [10].

Designing of cubes has a great impact on responsiveness of the system and analytical capacity. Direct effects on computational efficiency and insight generation are schema construction, aggregate strategies and hierarchy management. The available literature usually considers modeling and performance optimization individually and provides little combined advice on how to design high-performance OLAP systems.

The present paper will provide a unified system of building high-performance OLAP cubes to support advanced decision-making. It incorporates dimensional modelling, scalable structuring approaches, and performance considerations into one design approach. Part II is a review of related work, part III is the methodology, part IV provides the definition of evaluation principles, part V speaks of implications and part VI concludes the study.

II. Background and Related Work

A. OLAP Architecture and Cube Models

The architectures of Online Analytical Processing are based on multidimensional models which organize data into hypercubes of facts and dimensions so that it can be aggregated and explored hierarchically to support decision analytics [9]. The traditional schemas like star and snowflake are flexible in querying yet have scalability problems with the rise in the number of dimensions.

The cube models that are based on graphs are extensions of the relational models, as they do take into account the relationships between events along with their depending situations, making the analysis of process-oriented data sets more expressive [1]. Moreover, the hierarchical representation methods also increase the efficiency of roll-up and drill-down of the multi-level data in terms of analytics [6].

B. Data Warehouse Combination.

The underlying warehouse design has a great impact on OLAP performance. The optimised schema building and ETL pipelines are shown to enhance the speed of analytical responsiveness and decision-support ability in healthcare data settings by domain-specific implementations [4].

The introduction of heterogenous sources of big data in agricultural warehouses also demonstrates the significance of structural consistency in generating cubes and providing predictive analytics [11], [12]. The architecture of the warehouse also helps in enhancing the results of the modeling in crop yield analytics using efficient multidimensional aggregation [15].

C. Analytical Systems of High Performance.

Scalability enhancement of the OLAP processing is more dependent on the architectural optimization. The hybrid memory cube systems are able to minimise latency by implementing optimal data placement schemes in high dimensional workloads [5]. Large-scale analytics on environmental data can be performed on distributed models by the use of parallel processing frameworks [13], and fast query execution through the use of in-memory virtual modelling platforms through reduced disk I/O dependency [14].

These trends suggest that the approaches toward the design of cubes and the computational performance engineering have been converged in the context of contemporary analytics system.

Table 1: Comparative Review of OLAP Design Strategies

Approach	Core Technique	Performance Advantage	Limitation	Source
Multidimensional Schema OLAP	Star/snowflake modeling	Flexible aggregation	Dimensional scalability issues	[9]
Graph-Based OLAP	Relationship-centric structuring	Improved contextual queries	Higher complexity	[1]
Hierarchical Modeling	Aggregation trees	Efficient navigation	Maintenance overhead	[6]
Domain Warehousing	Schema/ETL optimization	Decision relevance	Limited portability	[4], [11]
Hybrid Memory Cubes	Data placement optimization	Reduced latency	Hardware reliance	[5]
In-Memory Virtual Models	Runtime modeling	Faster execution	Resource intensive	[14]

III. Designing of High-Performance OLAP Cubes Methodology.

A. Optimization of Dimensional Modelling.

Cube performance is achieved effectively by refining dimensional schema. Converting facts into dimensions into which they belong minimizes the redundancy and facilitates the flexibility of analysis in the course of aggregation [8]. Formatted modeling plans enhance the consistency of schema and cube hierarchies of warehouses which can be quickly executed and navigated.

Semi-automatic design methods extend the method of cube construction by controlling the restructuring of the schema with the help of an analytical pattern recognition. They are scalable modeling decision-making methods that do not compromise business logic or analytical requirements [8]. It is also known that hypercube interfaces work on a relational environment and exhibit better interaction of the user with the multidimensional environment by allowing intuitive querying via abstraction layers [2].

B. Cube Data Organization Scalability.

With a higher data dimensionality, cube structuring should give more priority on storage efficiency and query responsiveness. Cube compression High-dimension compression methods reduce memory overheads and still support aggregation, are better than in high-dimensional analytical settings [5]. These plans work well especially when the cube complexity is beyond the traditional storage limits.

Hypercube-based query models also assist in scaling since, multidimensional relationships are arranged into pliable structures which streamline analytic traversal and contextual connection among datasets [2]. The graph-oriented processing mechanisms also facilitate a scalable analysis of the process execution data to facilitate effective management of the relational dependencies in cube operations [1]. Taken together, these structuring methods take care of restrictions of dimensional explosion and heterogeneity of data.

D. Retail Distribution and Supply Chain Management.

The computation infrastructure and storage infrastructure is found to be a strong dictator of cube efficiency. The deployment of hybrid memory allows to place the data within memory levels in an optimal way, cutting latency and enhancing the throughput of high-volume analytical processes [5]. These architectures are appropriate in the setting where a high rate of aggregation must be performed with large volumes of data.

Parallel processing also enhances the responsiveness of cubes by allocating the computing tasks to scalable computing infrastructure. Analytical pipelines made distributed with the help of cluster based models of execution prove to be effective in the handling of large environmental and industrial data that allow real time decision support capabilities [13]. All these strategies reinforce cube performance during heavy workloads of analytical work.

Figure 1: OLAP Cube Design Workflow

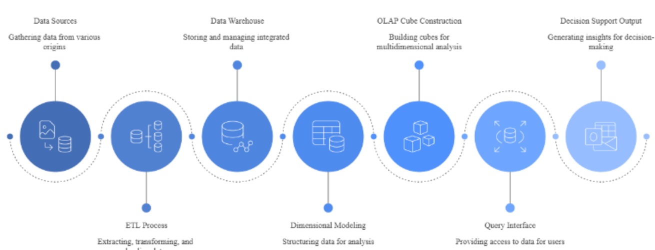

Table 2: Cube Design Optimization Techniques

Design Component	Optimization Method	Expected Impact
Schema Modeling	Fact–dimension transformation	Reduced redundancy and improved navigation
Cube Structuring	High-dimensional compression	Lower memory overhead

Query Interface	Hypercube abstraction models	Enhanced usability
Storage Layer	Hybrid memory deployment	Reduced latency
Processing Layer	Parallel computation	Improved throughput

IV. Performance Evaluation Framework.

A. Analytical Workload Issues.

The assessment of the efficacy of OLAP cube architectures should be done through organized analysis of analytical loads which affect computational demand and responsiveness. The complexity of a query, degree of aggregation and dimensional traversal are of important consideration in processing requirements in interaction with multidimensional structures [9]. Operations of roll-up and drill-down over hierarchical dimensions require extra computational overhead and especially when there are nesting levels of categorical data within datasets. Complex hierarchical representation and navigation are then very important in ensuring consistent performance in the execution when exploring in a multidimensional way [6].

Moreover, the variability of workload in various situations of the decision-support should be taken into account when determining the effectiveness of cube design. Interactive analytical sessions are typically characterized by queries of unknown patterns and the cube forms need to be able to support dynamic aggregation paths. Introducing workload-conscious structuring measures are used to keep the resources use balanced and minimizes the bottlenecks in the high-frequency analytical interactions.

B. Decision-Support Metrics.

The technical measurements also need to be consistent with decision-support goals as performance evaluation to make the output of the analysis reliable and actionable. As descriptors of system responsiveness, latency and throughput are two key metrics of how effectively analytical queries can be answered with increase in load [5]. Scalability is the ability of the system to perform in a stable manner in any case of a dimensionality change, an increase in the amount of data, or a similar increase in simultaneous access, it is an important metric when making decisions on large scale deployments.

In addition to computational measures, analytical reliability has an impact on the decision outcomes. The integration of multi-dimensional analytics with visualization and intelligence layers on the platform show how the performance indicators can be converted into better interpretation and strategic planning capabilities [10]. The assessment of cube systems based on the measurement of the metrics combined with technical and decision-making will guarantee the correspondence between the optimization of the infrastructure and the usefulness of the analytic value.

C. Application Domain Validation.

Practical applications give a confirmation to performance assessment models under a variety of environments. Farm analytics with warehouse-based cube architectures show improved sensitivity to handle heterogeneous datasets in making predictive decisions [11], [12]. The same can be said about the healthcare decision environment where structured multidimensional storage and retrieval mechanisms are useful to promote clinical analysis processes and enhance information accessibility [4].

The role of multidimensional analytics in decision monitoring of operational performance and improvement of process-level decision strategies is further seen in industrial optimization studies [7]. The total confirmation of these domain applications on the ability of performance evaluation based on the computation metrics as well as the contextual validation is that it provides a holistic view of OLAP cube effectiveness.

Figure 2: Performance Evaluation Model

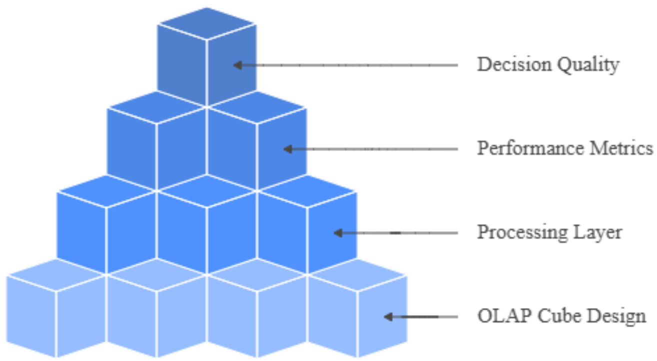

Table 3: Performance Metrics Mapping

Metric	Measurement Focus	Decision Impact
Latency	Query response time	Faster insight availability
Throughput	Queries processed per interval	Improved workload handling
Scalability	Stability under dimensional growth	Sustained system reliability
Analytical Accuracy	Insight consistency	Better decision confidence

V. Discussion

High-performance OLAP cube design has to be a balance between the complexity of structure and responsiveness to analysis. Richness of dimensions increases the power of the exploratory, but with a cost of computational overhead that will likely decrease query throughput. Compression, hybrid deployment of memory, parallel processing methods are shown to give quantifiable scalability and latency control enhancements but can take up more architectural complexity and resource dependency [5], [13]. Making an effective balance thus depends on balancing cube structure with anticipated workload properties as opposed to making dimensional detail as fine and gross as possible.

The approach to be used in the study is the combination of modeling refinement, scalable structuring and infrastructure-aware optimization in one framework. The previous studies have indicated that schema transformation and automated dimensional designing techniques can enhance the consistency of the aggregations and efficiency of navigating in analytics [8]. Cube systems can be used together with hypercube interaction models and graph-based processing strategies to enable flexible querying of heterogeneous datasets as well as be useful in decision-support tasks [1], [2]. This integration brings out the importance of both conceptual modeling and computational execution by having a coordinated design practices.

Even industrial and domain specific implementations justify the usefulness of optimized cube engineering even more. In the examples of multidimensional analytics deployed in enterprise and sector-specific warehouses, structured aggregation increases the interpretability and decision alignment across the operational environments [4], [11], [12]. There is, however, a limitation in dealing with very high dimensional data and dynamic evolving schemas, showing that adaptive modeling and intelligent automation is still necessary in future cube design studies.

VI. Conclusion

In this paper, a structure of high-performance OLAP cubes design was provided to facilitate analytical decision-making. The study overcame the performance constraints of high-dimensional analytical environments by incorporating the refinement of dimensional modeling, scalable structuring, and infrastructure-conscious optimization. Sharing of schema design, storage strategies and distributed computation enhances responsiveness and reliability in the multidimensional systems.

Domain implementation evidence has supported the idea that the optimized level of warehouse integration and hierarchical modeling have efficiency in aggregation, and the scalable architecture has efficiency in terms of latency and throughput stability [4], [5], [11]. These conclusions underscore the importance of integrated cube engineering, as opposed to those methods of optimization that are separated.

New studies in the future ought to be done on adaptive restructuring, automated schema evolution and smart workload-aware optimization. Other potential solutions include conversational analytical interfaces and machine-assisted tuning to make evolving analytics ecosystems more accessible and relevant to a decision maker [2].

References

[1] S. M. R. Beheshti, B. Benatallah, and H. R. Motahari-Nezhad, "Scalable graph-based OLAP analytics over process execution data," *Distributed and Parallel Databases*, vol. 34, no. 3, pp. 379–423, 2016.

[2] F. H. Hazboun, M. Owda, and A. Y. Owda, "A natural language interface to relational databases using an online analytic processing hypercube," *AI*, vol. 2, no. 4, pp. 720–737, 2021.

[3] L. Kuwertz et al., "Leveraging online analytical processing cubes in the world's largest offshore and land drilling record system," in *Proc. Offshore Technology Conf.*, 2019, p. D021S026R003.

[4] A. Hamoud, H. Adday, T. Obaid, and R. Hameed, "Design and implementing cancer data warehouse to support clinical decisions," *Int. J. Scientific & Engineering Research*, vol. 7, no. 2, pp. 1271–1285, 2016.

[5] R. R. Silva, C. M. Hirata, and J. de Castro Lima, "Big high-dimension data cube designs for hybrid memory systems," *Knowledge and Information Systems*, vol. 62, no. 12, pp. 4717–4746, 2020.

[6] G. Tsilinikos, *Analysis and Processing of Hierarchical Data*, 2018.

[7] K. Karapetyan, "Realization of big data analytics tool for optimization processes within the Finnish engineering company," 2016.

[8] L. Sautot, S. Bimonte, and L. Journaux, "A semi-automatic design methodology for (big) data warehouse transforming facts into dimensions," *IEEE Trans. Knowledge and Data Engineering*, vol. 33, no. 1, pp. 28–42, 2019.

[9] K. Dhanasree and C. Shobabindu, "A survey on OLAP," in *2016 IEEE Int. Conf. Computational Intelligence and Computing Research (ICCIC)*, 2016, pp. 1–9.

[10] V. Gole and S. Shiralkar, *Empower Decision Makers with SAP Analytics Cloud*. Berkeley, CA, USA: Apress, 2020.

[11] V. M. Ngo, N. A. Le-Khac, and M. T. Kechadi, "Designing and implementing data warehouse for agricultural big data," in *Int. Conf. Big Data*. Cham, Switzerland: Springer, 2019, pp. 1–17.

[12] V. M. Ngo, N. A. Le-Khac, and M. T. Kechadi, "Data warehouse and decision support on integrated crop big data," *Int. J. Business Process Integration and Management*, vol. 10, no. 1, pp. 17–28, 2020.

[13] A. Sadiq, A. El Fazziki, J. Ouarzazi, and M. Sadgal, "Air quality analysis based on MapReduce and K-means: A decision making system," *Int. J. Advance Soft Comput. Appl.*, vol. 9, no. 2, 2017.

[14] A. Pattanayak, "High performance analytics with SAP HANA virtual models," *Journal of Computer and Communications*, vol. 5, no. 7, pp. 1–7, 2017.

[15] V. M. Ngo, N. A. Le-Khac, and M. T. Kechadi, "An efficient data warehouse for crop yield prediction," *arXiv preprint arXiv:1807.00035*, 2018.