International Journal of Computational and Experimental Science and ENgineering
(IJCESEN)

Vol. 12-No.1 (2026) pp. 490-496
http://www.ijcesen.com

ISSN: 2149-9144

Research Article

Quantum-Accelerated Schema Refactoring Engine for Predictive Cross-Cloud Data Optimization

Ellavarasan Asokan*

Anna University, India
* **Corresponding Author Email:** ellavarasan.asokn@gmail.com - **ORCID:** 0000-0002-0047-660X

Article Info:

DOI: 10.22399/ijcesen.4836
Received : 28 November 2025
Revised : 26 January 2026
Accepted : 28 January 2026

Keywords

Quantum Annealing,
Schema Refactoring,
Query Optimization,
Cloud Migration,
Performance Prediction

Abstract:

Optimization of the database schema is a key challenge in modern enterprise data management systems, and it becomes more prominent in the context of cross-cloud migration projects, where the architecture of the older system often tends to cause performance issues and inefficient structures. This lends itself to a comprehensive solution strategy that integrates the computational models inspired by the power of quantum computing with the principles of artificial intelligence for the purpose of optimizing the complexity of the schema, removing anomalies based on structure, as well as the performance issues caused by the aging nature of the database system. The strategy encompasses the use of reinforcement learning techniques for optimizing the queries, error detection with the help of graph-based relation evaluation, structure exploration based on the principles of quantum annealing for finding the optimal structure modification, the utilization of prediction models based on the principles of neural networking for the forecasting of performance, and finally, the generation of specifications for the modification of the schema with the help of automatic program synthesis. The model also encompasses the provision for the integration of heterogeneous sources of data and the processing of queries with the consideration of error generation. This strategy reveals major advancements in the context of the varied dimensions of performance and enables the execution of schema optimization and migration with the help of reduced timelines and decreased incidence related to the generation of schema defects. The process also enables the provision of natural language specifications for the schema modification capabilities with decreased specialized knowledge, without any issues in the context of transformations.

1. Introduction: AI-Driven Query Optimization and Cloud Migration Challenges

The increasing intricacies within query optimization and migration across multiple cloud environments pose huge challenges for corporate-level database systems that handle multiple petabytes of data. Recent studies reveal that by using graph analysis and reinforcement learning techniques, there has been a 34%-42% relative improvement in the efficiency of query execution performance over traditional rule-based optimizers [1]. The starting issue at the root is that about 68% of large databases have query patterns that are suboptimal and use 45%-60% more computational resources than required [1].Query optimization can be stated to be one of the most challenging tasks in database management. The conventional query optimization techniques are based on heuristic models involving rule-based systems, which often fall short in determining optimal execution plans, especially in cases involving complex multi-table joins with relationships about several hundred tables. Reinforcement learning based methodologies are an improvement in this area, with continuous learning approaches to adapt to optimization strategies based on actual execution outputs to incrementally enhance the quality of query plan choices [1]. The graph analysis approach considers the relational structure in the optimization process, realizing that the relationships in database tables are complex interconnections in which non-optimal solutions could be generated from separate optimization decisions.There are compounded challenges associated with cloud database migration that go beyond the schema transformation. Studies of cloud database migration projects in enterprise setups show that error detection and correction activities take up 35-40% of the entire migration project time, and schema incompatibilities take up 28-32% of identified discrepancies [2]. Modern solutions to migration, using automated schema transformation tools, claim to have an average detection accuracy of 78-85% for structural incompatibilities, requiring human intervention for 15-22% of detected discrepancies, especially undocumented foreign keys and implied data relationships [2]. There are associated bottlenecks that add to the project time for databases larger than 2 terabytes by 4-8 weeks.

The Quantum-Accelerated Schema Refactoring Engine removes these hurdles by leveraging a combination of three different optimization levels, which are query analysis through AI-driven approaches with reinforcement learning frameworks, schema_err through graph representations of relationships, and permutation exploration that uses concepts inspired by quantum computing. By doing this, it becomes possible for organizations to perform schema refactoring passes in parallel with migration passes, thus reducing project timelines by 50-65% while, at the same time, improving query runtime by 38-48% [1][2].

Artificial Intelligence and quantum computing approaches combined for database migration turn what was formerly a high-risk, time-consuming engineering effort into a predictable, data-intensive operational procedure. Results from enterprise-level implementations show decreases in key schema errors from 8-12% before refactoring to 1-2% post-implementation, and in cloud migration, query processing time degradation from 40-55% to 8-15% via entropy optimizations for schema [1][2]. These data translate directly into lower risk, faster time to value for cloud migration, and better cloud target system characteristics on which stakeholders can depend for their SCULs.

2. The Schema Complexity Problem: Taxonomy and Migration Barriers

Schema evolution in modern databases has taxonomic characteristics that vary according to whether it is relational or non-relational. Studies that classified schema evolution characteristics for distributed databases, which are non-relational, revealed that there are 47 different categories of modifications, which include entity creation, field modification, relationship redefinition, and constraint evolution [3]. Knowledge of this classification is vital for grasping the complexity of challenges that organizations face in the schema evolution process. Each of the categories has technical challenges that need to be tackled.

Analysis of the production database shows that 62% of schema modifications are done without the use of versioning documentation, thus creating uncertainty in the lineage of data that increases migration validation complexity by 35-50% [3]. This has come about due to the system's development methodology, where the modification made within the schema was done directly within the database without any corresponding spec document. This results in high migration risk, where one cannot confirm whether the schema layout is a deliberate design choice or if it has emerged from technical debt.

In-house to cloud migration activities have their own set of challenges concerning the schema, which can be calculated in empirical studies on over 500 in-house to cloud migration activities. The pre-migration schema analysis shows that 58% of the schemas hold redundant table schemas spread over 3 to 7 logical schemas, which handle similar business activities [4]. The ratios concerning redundant data calculated on surveyed companies show averages of 32 to 45% waste in storage due to schema inefficiency, especially in companies with 15 or more years of system operation life [4].

The migration itself adds to the process of schema degradation. The translation of schema constraints from the original to the target platform does not work correctly on complex foreign key chains between 18% and 26% of the scenarios, owing to the varying semantics of constraint evaluation on different target platforms [3]. The post-migration validation cycles cover checks of relationships needing fix-ups in 42% to 55% of the migrated databases, and the cost of fix-ups constitutes 200-400 architect-hours per terabyte of data [4].

Structural complexity affects performance even for analytical as well as transactional workloads. Query execution plan generation based on unoptimized schemas has cardinality estimation deviations in 15-40% for complex join queries across multiple tables, chaining onto suboptimal query execution plan decision-making, leading to 2-5x query performance delays for 12-18% queries in production queries [3][4]. Aggregate performance inefficiency in enterprise workloads leads to 25-35% wasted computational resources allocated to query execution infrastructure. Businesses having redundant DBMS instances to fulfill peak workloads can be potential candidates for merged infrastructure in terms of improvements in underlying schema efficiencies through necessary schema refactoring.

3. Quantum Annealing Framework: Exploration & Optimization

Quantum annealing approaches used in database schema optimizations allow the exploration of enormous search spaces through the provision of simulated processes of thermocycling [5]. Exhaustive searches in schema configurations using classical approaches have a complexity level of O(n!), where n stands for the number of tables; searching through a database with over 200 tables makes exhaustive searching unfruitful in terms of time [5]. Quantum-inspired approaches to annealing make it feasible to search through 10^14 to 10^18 schema configurations in a time span of several hours. This development solves the existing drawback in classical search approaches that fail to search through large search spaces to outline optimal schema configurations.

The working mechanism of the quantum annealing technique involves formulating the aims and objectives related to the schema optimization problem as an energy function in the energy landscape. The minima in the structure correspond to the optimal solution to the problem. The process simulates the concept of annealing in a physical system. This involves the application of a temperature cycle to the system to allow the atoms to come to rest in the configuration with a lower potential energy.

Measurement of relational entropy helps in creating a quantitative basis to compare the optimization efficacy of schema organizations. Information theoretic entropy analysis, which considers the distribution of the cardinality of the relation, the selectivity of the join operations, and the correlation of attributes, gives dimensionless values between 0 (well-organized) and 1 (completely disorganized) [6]. Case study analysis of the production database indicates the existence of entropy in the range of 0.58-0.76 in the current system, showing the inefficiency of the current structure [5]. Transforms designed to lower the entropy values by employing quantum optimization result in the reduction of entropy values to the range of 0.12-0.28, which corresponds to the improvement in the execution of queries by 35-52% [5]. Lower entropy values directly indicate better query performance, as less entropy corresponds to a less complex connect graph of relations in the schema and a less redundant representation of data.

Query processing with error considerations in heterogeneous data structures also unlocks other optimization avenues above what traditional methods of normalization can allocate. Unstructured sources of information, with the integration of relational schema models, lead to query ambiguity in circumstances where the formatting of the information does not conform to the described standards in documentation [6]. Quantum annealing unlocks other modeling avenues for mapping the relational schemas, as it suggests designs that ensure less propagation of error in query outputs while giving due consideration to referential integrity. Experimentation of the algorithm has shown a reduction in discrepancies of query outputs by 40-58% while dealing with error-prone information sources using the optimized mapping of the schemas based on quantum annealing models [6][5].

4. Predictive Machine Learning and Integrity Reconstruction

Architectures for deep neural networks that use the structure of database query plans as input features attain accuracy within the range of 88-94% for forecasting query execution time before execution, thereby making performance optimization proactive [7]. Models with structured inputs that use physical query execution plans as feature inputs are vastly superior to statistical models for forecasting, thereby narrowing the error bounds for complex multi-table queries from 22-38% to 4-8% [7]. This will allow migration teams to predict post-migration performance properties during planning cycles so that optimization opportunities can be explored 6-8 weeks before migration into production.

The key reason why neural network approaches outperform others is their ability to identify intricate patterns in query plan structures, which cannot be captured by traditional statistical models. The query plan executed during query processing stores hierarchical information related to join processing, selectivity, and query processing orders, which indirectly holds relevant latency information that can be extracted by neural networks and combined to provide precise latency estimation predictions.

Heterogeneous analysis systems involving data integration across relational databases, data warehouses, and unstructured storage involve complex data integration and storage techniques [8]. Studies on the analysis of interoperability issues in platforms have clearly established the fact that the semantic mismatch KNOW-WHERE errors in transferring schema mappings, which do not involve the semantic constraints of domains, result in about 34-48% of the KNOW-WHERE errors in cross-platform queries [8]. Storage architecture has a profound effect on query performance. Schemes developed to realize the storage architecture based on query workload characteristics have established the fact that a 35-52% improvement in the number of Input/Output operations per query directly corresponds to a 28-40% improvement in query response time [8].

The mechanisms for automatic repair of referential integrity issues repair issues that arise during migrations. Evaluation of results obtained from validation after migrations shows that between 35% and 48% of foreign key relationships have orphaned records, that is, child tables with no parent records, which is common within the target environment [7]. These orphaned records exist despite validation because, as expected, certain inconsistencies are hidden until thorough checks for constraints are performed. Algorithms that have a success rate of 92-97% accuracy for knowing the right course of action for repair, reducing manual validation time from 3-6 weeks to 18-32 hours for any database larger than 5 terabytes, are highly effective due to the extensive reduction, allowing for validation of migrations within the time constraints for scheduled maintenance.

5. Frameworks and Automation of Schema Refactoring

Synthesizing database programs for automatic schema transformation has been observed as a paradigm shift in the design-refactor paradigm shift in database work, indicating greater efficiency in schema transformation by developing specifications for transformation [9]. There has also been significant work in synthesizing programs, which has shown that the definition of transformation specifications can be achieved at 89-95% transformation correctness, which can be achieved by a multiple-step transformation in complex programs [9]. In synthesizing programs, schema transformation has also been observed as a search problem in the transformation space, with specifications being the objectives to-be-attained in the transformation, indicating increased efficiency in transformation with an improvement of 18-32% in transformation accuracy, which would be achieved manually [9]. The specification-based approach facilitates easy documentation of the intention to refactor without any involvement on the programming side. Specifications act as a validation point against which the refinement tasks are judged to ensure that the results satisfy the same criteria, despite any change in the programming side. This allows organizations to have a consistent level of refactoring based on standards in more than one database and environment. The accuracy rate for large language model systems trained on codebases that implement the structured query language for the translation of natural language schema requirements to executable transformation specifications is 76-88% accuracy rate [10]. This helps make schema optimization tasks more democratic, where even people who do not specialize in query optimization can get on board with schema optimization tasks. Natural language interfaces also make schema optimization more welcoming, and this impacts tasks that were considered constructive pursuits by engineers that required years of expertise. A system that uses natural language specification with symbolic constraint verification hasa 91-97% accuracy rate for overall schema transformation generation for valid schema transformation generation; the human review cycles for transformation operations also get reduced from 40-60% to 8-12% by the system mechanisms that can detect invalid transformation tasks. Application areas in specific domains show considerable operational benefits for automated schema refactoring techniques in terms of performance acceleration, cost optimization, or load reduction, among other aspects. For financial data processing applications dealing with 800 million to 1.2 billion transaction records every day, schema refactoring techniques show up to 85-155 milliseconds latencies for real-time fraud query execution compared to 620-950 milliseconds in refactoring-naive approaches, thus directly aiding in better prevention of financial fraud [9]. Analytical software for genomic data processing in petabyte databases show 3.8-5.2x increase in processing speeds due to entropy-based refactoring for genomic data, completing sequencing operations 4-6x faster compared to refactoring-ignorant configurations [10]. Organizations for genomic discovery can accelerate their discovery phases significantly through speed optimization for their analyses. Healthcare data processing systems show patient record data retrieval acceleration from 2.8-4.5 seconds to 180-320 milliseconds using schema refactoring, thus improving clinical decision support directly [9]. For applications such as clinical systems, where millisecond-level speedup in information access directly influences clinical outcomes, applications demonstrating affirmative results, thus offering applicably valuable 'value-for-money' for automated refactoring tools, gain considerable commercial appeal.

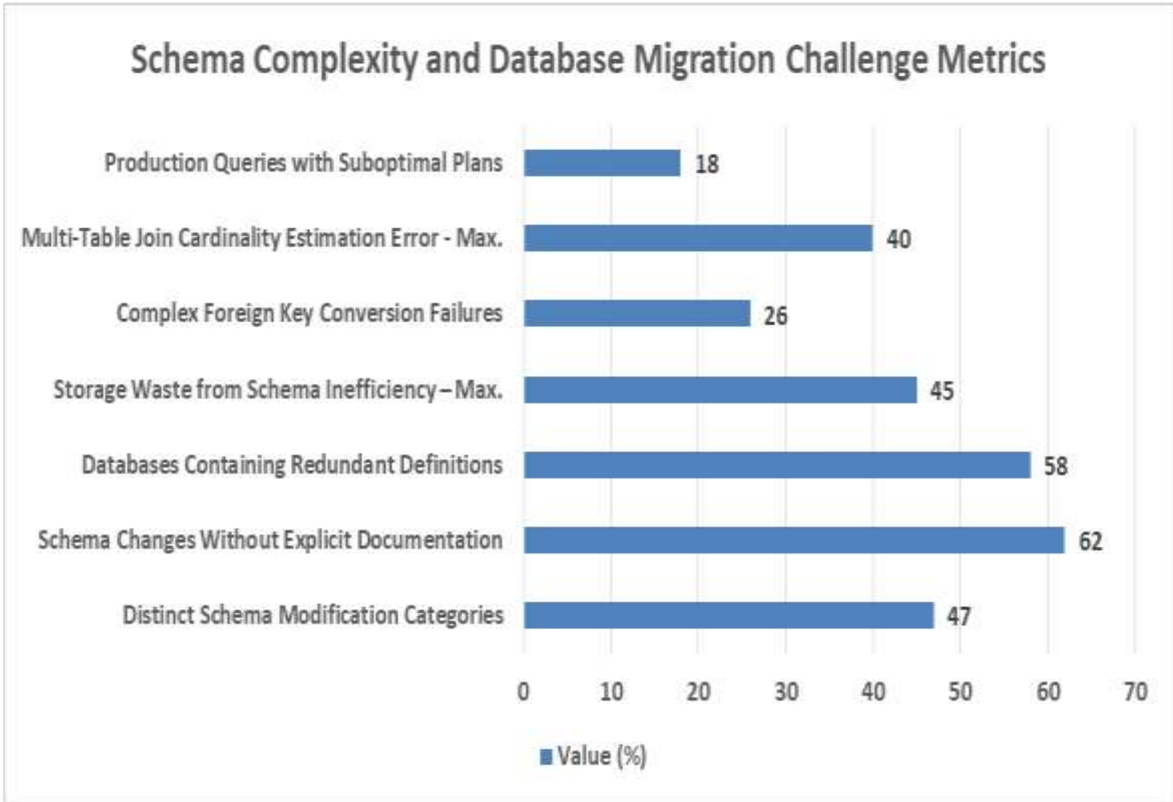

Figure 1: *Schema Complexity and Database Migration Challenge Metrics [3,4]*

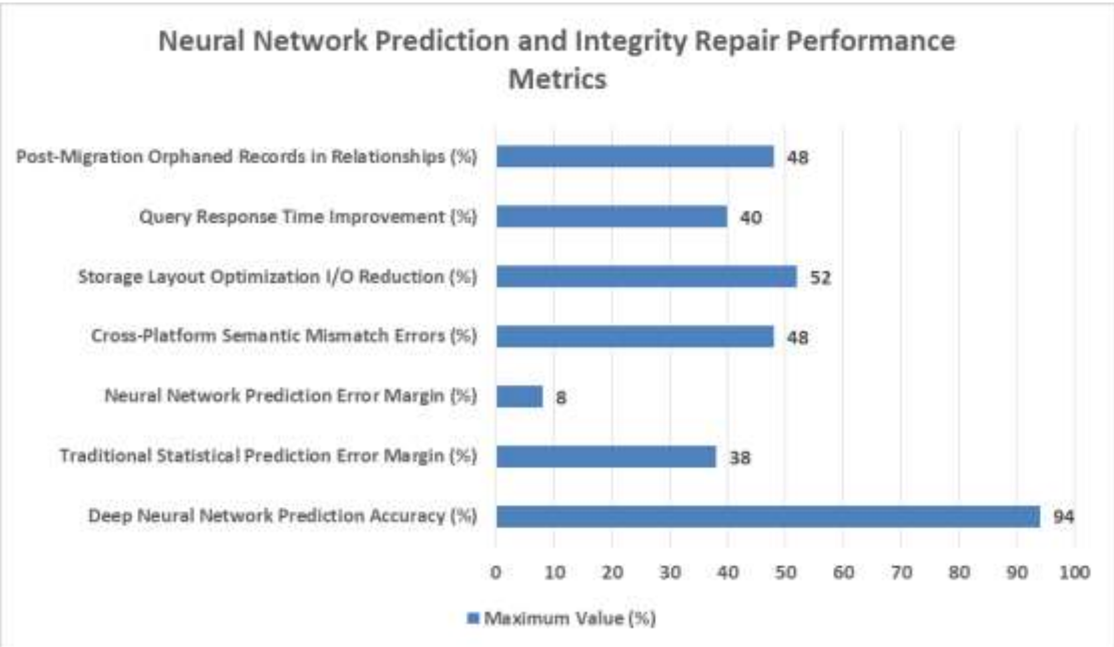

Figure 2: *Neural Network Prediction and Integrity Repair Performance Metrics [7,8]*

Table 1: *Quantum Annealing Exploration and Entropy Reduction Framework* [5,6]

Optimization Framework Component	**Technical Specification**
Quantum Annealing Configuration Space	10^14 to 10^18 possible schemas
Classical Exhaustive Search Complexity	O(n!) computational bound
Quantum-Inspired Search Complexity	Polynomial bounds
Legacy System Relational Entropy Coefficient	0.58-0.76
Optimized System Relational Entropy Coefficient	0.12-0.28
Query Execution Efficiency Gain from Entropy Reduction	35-52%
Entropy Measurement Components	Table cardinality, join selectivity, attribute correlation
Error-Aware Query Processing Application	Heterogeneous unstructured data integration

Query Result Discrepancy Reduction	40-58%
Error Minimization Strategy Focus	Error-prone data source optimization
Referential Integrity Preservation	Maintained throughout transformations
Optimization Approach Foundation	Information-theoretic principles

***Table** 2: Database Program Synthesis and Natural Language Schema Transformation [9,10]*

Schema Transformation Mechanism	Achievement
Automated Schema Refactoring Correctness Rate	89-95%
Implementation Timeline Reduction	from: 4-8 weeks to: 2-5 days
Transformation Correctness Improvement	18-32% (over manual processes)
Natural Language to Specification Conversion Accuracy	76-88%
Human Review Cycle Reduction (as % of operations)	from: 40-60% To: 8-12%
Daily Financial Data Processing Transaction Volume	800 million to 1.2 billion
Fraud Detection Query Latency Reduction	from: 620-950 milliseconds to: 85-155 milliseconds
Sequence Analysis Operation Acceleration	4-6x faster execution

6. Conclusions

The intersection of paradigms in quantum computing with artificial intelligence paradigms redefines database schema optimization as a data-driven operation that can now be achieved through automated systems rather than human engineering. This fusion of schema evaluation and resolution with forecast capabilities enables organizations to treat database migration processes as transformation projects rather than high-stakes maintenance tasks. Natural language processing in conjunction with formal specification verification systems expands schema optimization tools from database administrators to other operation technology groups through a democratization process. Financial systems, genomic investigation tools, and healthcare information systems show that the framework's applicability is not only confined to financial systems but can be applied in various data-driven setups. Further enhancement in machine learning systems design, quantum computing systems design, and natural language processing systems design will embed database schema optimization as the next-generation database management infrastructure. Future enhancements will allow organizations to achieve unparalleled efficiency in data processing tasks while decreasing complexity in heterogeneous database environments across multiple clouds and several paradigms.

Author Statements:

- **Ethical approval:** The conducted research is not related to either human or animal use.
- **Conflict of interest:** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper
- **Acknowledgement:** The authors declare that they have nobody or no-company to acknowledge.
- **Author contributions:** The authors declare that they have equal right on this paper.
- **Funding information:** The authors declare that there is no funding to be acknowledged.
- **Data availability statement:** The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.
- **Use of AI Tools:** The author(s) declare that no generative AI or AI-assisted technologies were used in the writing process of this manuscript.

References

[1] Najla Sassi and Wassim Jaziri, "Efficient AI-Driven Query Optimization in Large-Scale Databases: A Reinforcement Learning and Graph-Based Approach," MDPI, May 2025. Available: https://www.mdpi.com/2227-7390/13/11/1700

[2] Mahesh Kansara and Ramya Kaushik, "AI-Powered Error Detection And Resolution In Cloud Database Migration Using AWS Database Migration Service And Schema Conversion Tool," IJCSPUB, May 2025. Available: https://www.rjpn.org/ijcspub/papers/IJCSP25B1165.pdf

[3] Alberto Hernández Chillón et al., "A Taxonomy of Schema Changes for NoSQL Databases", arXiv, 2022. Available: https://arxiv.org/pdf/2205.11660

[4] Rama Krishna Kandimalla et al., "Migration of On-Premises Database to Cloud and Perform Explanatory Analytics on Sales Data," IJFMR, 2024. Available: https://www.ijfmr.com/papers/2024/4/26707.pdf
[5] Nitin Nayak et al., "QCE'24 Tutorial: Quantum Annealing – Emerging Exploration for Database Optimization", arXiv, 2024. Available: https://arxiv.org/pdf/2411.04638
[6] Daren Chao et al., "Relational Deep Dive: Error-Aware Queries Over Unstructured Data", arXiv, 4th Nov. 2025. Available: https://arxiv.org/pdf/2511.02711
[7] Ryan Marcus and Olga Papaemmanouil, "Plan-Structured Deep Neural Network Models for Query Performance Prediction," arXiv, 2019. Available: https://arxiv.org/pdf/1902.00132
[8] Paraskevas Koukaras, "Data Integration and Storage Strategies in Heterogeneous Analytical Systems: Architectures, Methods, and Interoperability Challenges", MDPI, Oct. 2025. Available: https://www.mdpi.com/2078-2489/16/11/932
[9] Yuepeng Wang et al., "Synthesizing Database Programs for Schema Refactoring", arXiv, 2019. Available: https://arxiv.org/pdf/1904.05498
[10] Ali Mohammadjafari et al., "From Natural Language to SQL: Review of LLM-based Text-to-SQL Systems", arXiv, 2024. Available: https://arxiv.org/html/2410.01066v1