

Exploring Self-Optimized Technologies in Cloud Data Warehouse

A SCALABLE AI WHITE PAPER

SCALABLE
AI

TABLE OF CONTENT

| | |
|--|----|
| Introduction..... | 01 |
| Self-Optimization: Hack Your Way to Success..... | 02 |
| From Chaos to Control: How Self-Optimization Solves Business Problems..... | 03 |
| Cloud Storage with Self-Optimization Helps Businesses in Many Ways..... | 04 |
| Demystifying the Process of Data Query Optimization..... | 05 |
| Optimizing Data Queries for Maximum Business Benefits..... | 06 |
| Query Optimization: Your Weapon Against Common Business Challenges..... | 08 |
| Automated Resource Allocation..... | 10 |
| Challenges and their mitigation..... | 11 |
| Realizing the Potential of Self-Awareness on Data Security..... | 14 |
| Final Thoughts..... | 15 |

INTRODUCTION

Data warehouse is an integral source of information for companies for years and will continue to play a pivotal role in providing insights for decision-making. However, they are undergoing massive change in terms of technology and adoption. New players are driving the market with innovative solutions that best suit the customers in terms of scalability, performance, and accessibility with less cost compared to on-premise Data warehouse solutions.

After the successful deployment of the warehouse, when the users can utilize the data, the focus shifts to the maintenance and growth part. The data warehouse team becomes responsible for providing the users with the required data, views, and processing capabilities.

However, as the scale and the size of the data warehouse increase the complexity in handling and managing will also increase exponentially. Companies must constantly monitor the connections between the data sources, architecture maintenance, updates, and patches, manage security challenges, access management, and more.

Cloud-based data warehouse solutions solve these issues to a great extent since companies can shift the responsibility of managing the resources to cloud providers and focus on their core business.

For instance, AWS Redshift and Google Cloud Platform (GCP) are highly scalable cloud-based data warehouse solutions with managed service offerings. In addition, players such as Snowflakes offer an innovative cloud solution by understanding customer needs and challenges. However, an increasing number of customers are looking for solutions that can provide collaborative features without compromising on security. As their demand increase, we need to investigate the technology frameworks that can self-adapt and self-optimize based on the workload and resource requirements.



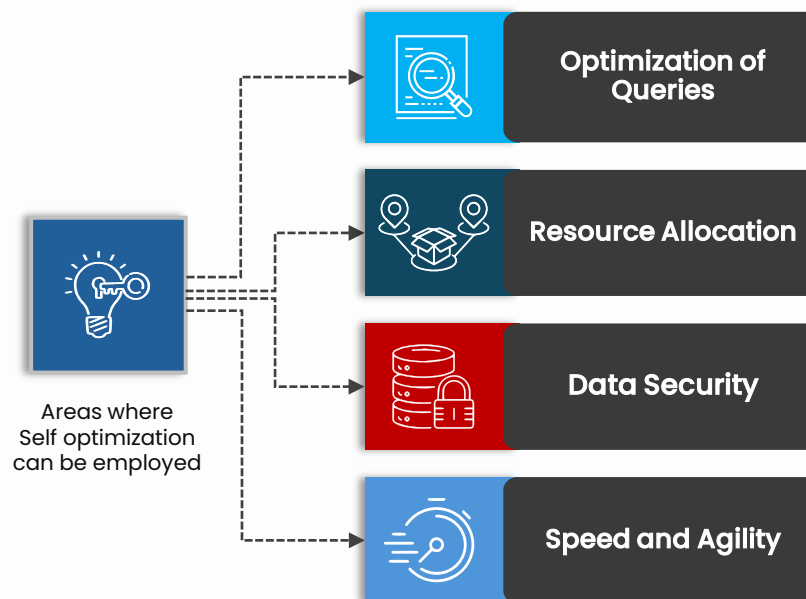
SELF-OPTIMIZATION:

Hack Your Way to Success

Self-optimization is the key to successful implementation and management of Data Warehouse solutions, given the dynamic nature of the requirements based on business needs. Cloud-based Data warehouse solutions are designed to utilize the self-optimization features in terms of mapping virtual to physical resources and managing data loads to applications based on usage patterns. As technology evolves to meet new data demands, it also provides new areas of opportunity for business growth and operational efficiency.

Today we can observe how automation and machine learning play a great role in creating a self-optimization system that can automate and make a decision on a variety of functions that are manual in nature. This can be extended to Data Warehouse systems which enable them to tune themselves based on the requirements. This capability is accelerating the speed at which data warehouses deliver value to businesses.

Cloud-based Data warehouse systems can employ and exploit the self-optimization features for a better user experience.



FROM CHAOS TO CONTROL

HOW SELF-OPTIMIZATION SOLVES BUSINESS PROBLEMS

- Managing exponential data growth

Businesses often struggle to keep pace with the ever-increasing volume of data. Self-optimization helps by automatically scaling storage and optimizing resource allocation.

- Slow query response times

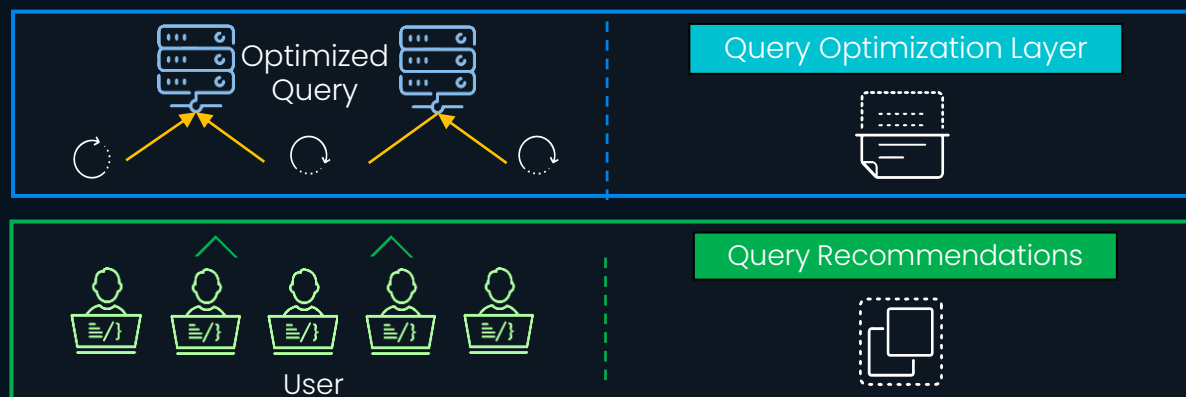
Slow data retrieval can hinder analysis and decision-making. Self-optimization through data organization and skipping techniques improves query response times.

- Inefficient resource utilization

Businesses might unknowingly pay for unused storage capacity. Self-optimization identifies and eliminates this waste, leading to cost savings.

- IT staff workload

Managing a large data warehouse can be a burden on IT staff. Self-optimization automates tasks, freeing them for more strategic work.



Overall, self-optimization in cloud data storage empowers businesses to manage their data more effectively, reduce costs, improve performance, and ultimately gain a competitive edge through data-driven decision-making.

CLOUD STORAGE WITH **SELF-OPTIMIZATION** HELPS BUSINESSES IN MANY WAYS

Reduced Costs

- Eliminates waste

Self-optimization identifies and removes underutilized storage resources. You only pay for what you use, preventing unnecessary spending on cloud storage.

- Automatic scaling

During low usage periods, self-optimization can automatically scale down storage needs, further reducing costs.

Enhanced Efficiency

- Reduced manual work

Mundane tasks like data organization and storage management become automated, freeing up IT staff for more strategic initiatives.

- Reduced errors

Automating routine tasks minimizes the risk of human error in configuration and data management.

Improved Performance

- Faster access

Data organization and skipping irrelevant sections during retrieval (thanks to self-partitioning) lead to quicker retrieval times for queries.

- Optimized resource allocation

By analyzing workloads and allocating resources efficiently, self-optimization ensures smooth operations and faster data processing.

Better Decision-Making

- Faster data analysis

Improved performance through self-optimization allows for quicker data analysis, leading to more timely insights for better decision-making.

- Focus on strategy

With IT staff freed from routine tasks, they can focus on developing data governance strategies and ensuring data security.

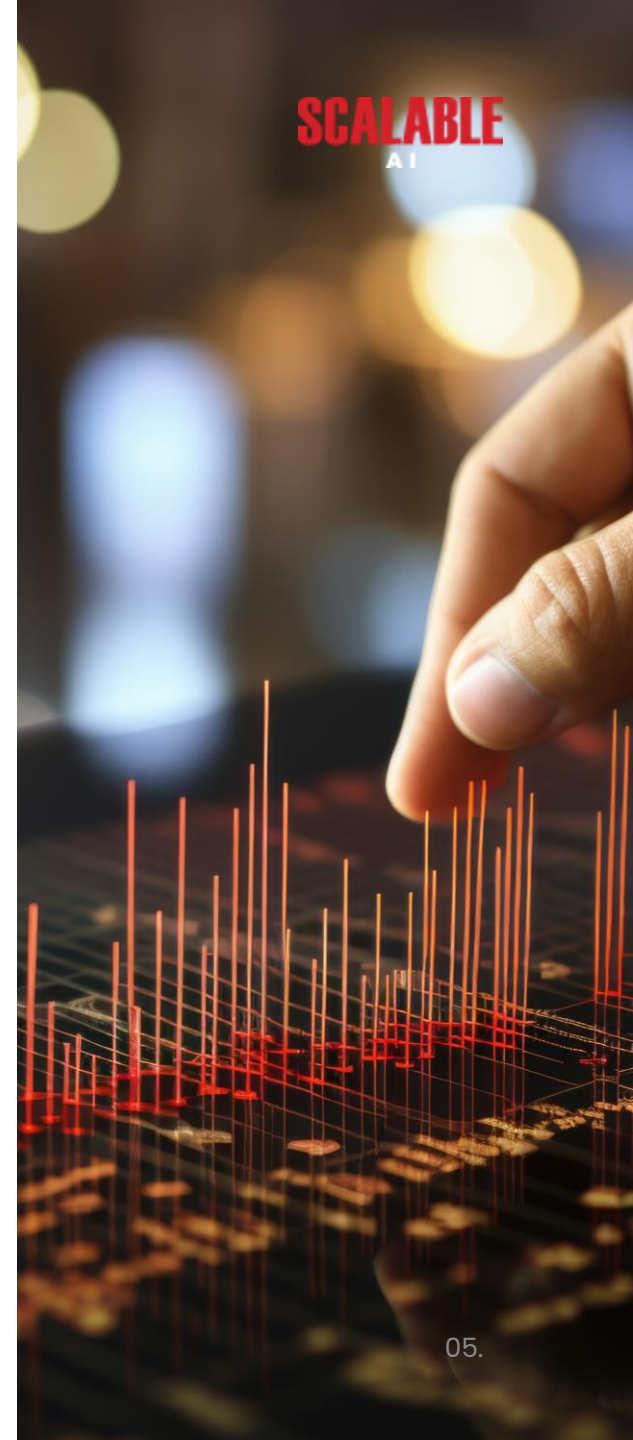
DEMYSTIFYING THE PROCESS OF DATA QUERY **OPTIMIZATION**

Being the central hub of data for an organization, a data warehouse handles millions of queries from users for extracting data for business purposes. Some of these queries follow best practices whereas most will be ad-hoc, which can cause an increased load to the system. An intelligent cloud data warehouse system can analyse these query patterns and optimize them or provide query recommendations to the users that they can use for extraction. This can significantly improve the performance of the Data Warehouse system.

Advanced Machine learning algorithms combined with automation features, analyze the historical pattern of the user queries and identify the best possible recommendations to the user to efficiently extract the data for their business need.

For example, performing a join on multiple tables is intensive and slow. In rare scenarios, poorly written queries based on joins can lead to a crash of the system. To avoid these situations, there should be a mechanism in place that can identify these issues and provide a better query recommendation to the user. ML-based optimization techniques solve this purpose by streamlining the joins between the tables to extract the data, thus reducing the load and increasing the speed of execution.

Many cloud players have started to implement ML-based query optimization techniques to predict and recommend queries to users based on their activities. These solutions are incorporated into the cloud and are made available to the users based on the subscriptions.



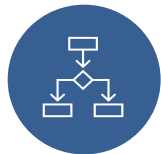
OPTIMIZING DATA QUERIES FOR MAXIMUM BUSINESS BENEFITS

There are several advantages to optimizing data searches, including faster and more effective data access and analysis.



Faster Results and Improved User Experience

Sluggish queries can slow down applications and reports, frustrating users. Optimized queries deliver results quicker, enhancing user experience and productivity.



Better Decision Making

Faster access to data allows for more timely analysis and decision-making. Businesses can react quicker to market trends, identify operational inefficiencies, or capitalize on new opportunities.



Cost Savings

Optimized queries use fewer computing resources, reducing the cost of running database servers or cloud services. This can lead to significant savings, especially for businesses dealing with large datasets.



Increased Scalability

As businesses grow, their data volume also increases. Optimized queries can handle larger datasets efficiently, allowing businesses to scale their data operations without performance degradation.



Improved Resource Utilization

Optimized queries minimize the workload on database servers, freeing up resources for other tasks. This can improve overall system performance and stability.



Reduced Risk of Errors

Poorly written queries can lead to inaccurate results. Optimization techniques help identify and fix errors in queries, ensuring data integrity and reliable insights.

Optimizing data queries helps businesses unlock the true value of their data by making it faster, more accessible, and more reliable. This translates to better decision-making, improved efficiency, and ultimately, a competitive advantage.



QUERY OPTIMIZATION

YOUR WEAPON AGAINST COMMON BUSINESS CHALLENGES

- **Slow Reporting and Analytics**

Imagine waiting hours for a sales report or a key marketing metric. Inefficient queries lead to sluggish dashboards and reports, delaying insights and hindering informed actions. Optimization ensures these critical tools deliver results quickly.

- **Inefficient Resource Allocation**

Unoptimized queries can overload database servers, impacting other applications and processes. This can lead to system slowdowns and bottlenecks. Optimization frees up resources, allowing for smoother overall system operation.

- **Data-Driven Decision Delays**

When crucial data takes forever to retrieve, it hinders timely decision-making. Businesses might miss out on key opportunities or struggle to react swiftly to market changes. Optimization ensures data is readily available for informed and rapid action.

- **Data Inaccuracy and Misconceptions**

Inaccurate queries can lead to misleading data and faulty conclusions. Businesses might base decisions on incorrect information, potentially harming their strategies. Optimization helps identify and rectify errors in queries, ensuring data accuracy and reliable insights.

- **Wasted IT Costs**

Businesses pay for resources used by their databases. Poorly written queries can consume excessive resources, driving up costs unnecessarily. Optimization helps businesses get the most out of their existing infrastructure, minimizing IT expenses.

- **Limited Scalability Issues**

As businesses grow, their data volume grows as well. Unoptimized queries might struggle to handle this increasing workload, leading to performance degradation. Optimization allows businesses to scale their data operations efficiently, handling larger datasets without compromising speed or accuracy.

“

By tackling these problems, data query optimization empowers businesses to leverage their data effectively. It translates to faster insights, improved operational efficiency, cost savings, and ultimately, better decision-making for a competitive edge.

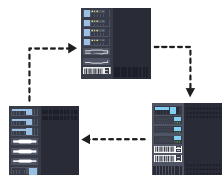
AUTOMATED RESOURCE ALLOCATION

Cloud-based Data warehouse systems offer a flexible and efficient approach to automate resource allocation based on workload demands. This ensures a high availability of data sources to the consuming applications downstream. Cloud computing shifts computing from local dedicated resources to distributed, virtual, elastic, multi-tenant resources. This paradigm provides end-users with On Demand access to computing, storage, and software services.

Quality of service (QoS) requirements is an important parameter when allocating resources to the end user. The resource allocation should support quick response time and throughput of a variety of downstream applications. Cloud providers offer a menu of server types with different configurations of CPU capacity, memory, network capacity, disk I/O performance, and disk storage size. An important objective of cloud platform providers is minimizing operational costs while guaranteeing the best resource allocation that meets their client QoS requirements.

As the demand increases the cloud-based Data warehouse solution should be able to support the client's requirements. Autoscaling is one of the solutions where the cloud platforms can provide necessary resources to the demand, but this is not a viable solution as it may not be cost-effective. The solution should focus on self-optimization when allocating resources thus enabling a better resource utilization that can provide better cost benefits as part of the QoS requirement.

Like query optimization, we can enable machine learning-based algorithms for resource optimizations. This provides an effective way to analyze the QoS requirement and efficiently optimizes cloud resource allocation. A self-adapting data warehouse will also optimize workloads, whereas legacy data warehouses could not easily optimize all queries and response times. A self-adapting data warehouse can also streamline and organize data with automatic adjustment of computing power so that it provides the best performance for each user and maximizes the ability to handle concurrent load execution.



Cloud Data Warehouse
Service Optimization



Automated Resource Pooling
and infrastructure Optimization



End User Optimization

CHALLENGES

While automated resource allocation in cloud data warehouses (CDWs) offers significant benefits, it's not without its challenges. Here's a closer look at some of the hurdles to consider:

Limited Scope

- Complex Tasks Remain Manual

Automation excels at handling routine tasks like scaling compute power. However, complex tasks like schema design, data governance, and setting optimization goals still require human expertise.

Accuracy and Control

- Machine Learning Shortcomings

Machine learning algorithms used for automated allocation are still under development. They might not always perfectly predict workload demands, potentially leading to resource over- or under-provisioning in certain situations.

- Limited Visibility

Understanding how the automation algorithms make decisions can be challenging. Ideally, CDWs should provide some level of explainable AI to give users insights into the allocation process.

Configurability Challenges

- Finding the Right Balance

Balancing automation with human oversight can be tricky. Over-reliance on automation can lead to inefficiencies if the system isn't configured correctly for your specific needs. Conversely, under-utilizing automation negates its benefits.



Security Considerations

- Accidental Overspending

Automated scaling based on unexpected workload surges could lead to exceeding budget limitations if not properly monitored and controlled.

- Data Security Risks

Scaling up resources might involve adding new nodes or instances. It's crucial to ensure these new resources adhere to the same security protocols as the existing infrastructure.

MITIGATING THE CHALLENGES

- **Clear Goals & Monitoring**

Clearly define your goals for automation and continuously monitor its effectiveness. Regularly review resource allocation decisions made by the system to identify any patterns or areas for improvement.

- **Human Expertise & Oversight**

Don't let automation replace human oversight entirely. Involve IT staff in setting up and monitoring the automation system to ensure it aligns with your business needs and security protocols.

- **Start Small & Scale Gradually**

Begin with automating simpler tasks and gradually introduce more complex automation features as you gain confidence in the system's capabilities.



By understanding these challenges and implementing proper safeguards, businesses can leverage automated resource allocation effectively and reap the numerous benefits it offers for managing cloud data warehouses. Remember, automation is a powerful tool, but it should be used strategically with a healthy dose of human oversight.

REALIZING THE POTENTIAL OF SELF-AWARENESS ON DATA SECURITY

Self-awareness of data security isn't quite about a data warehouse itself becoming sentient, but rather about building a system that incorporates features to consciously address security threats and data breaches. Numerous security breaches are reported across industries, where critical data is being compromised. To avoid these situations, a data warehouse system should be able to self-analyze in terms of data security threats and provide early warning to the necessary stakeholders or enable defense mechanisms. Every day a cloud data warehouse will handle millions of transactions which provides a huge resource base for itself to identify genuine and fraudulent transactions.

We can enable the cloud solutions to learn every day to detect and profile the transactions based on the origin, query request, transaction type, and more. When a critical parameter is breached, an automated response should be in place to analyze the type of transaction request in real time to identify the threat. Customer can avail of the services as part of their subscriptions and can enable an On-premises monitoring solution that can work in sync with the Cloud data warehouse to identify the threats in transactions.

FINAL THOUGHTS

As cloud technology evolves to handle various customer requirements, self-optimization is a key factor that can enable a self-sustaining data warehouse ecosystem. Customers should focus on the solutions that can provide this important feature as a part of their overall cloud data warehouse architecture.

We at Scalable Systems are aware of this critical requirement and have developed a center of excellence in this practice. We work closely with our vendor partner and provide the necessary inputs to align the solution that can be self-optimized and self-sustain in the long run. An organization's Cloud Data warehouse strategy should focus on a long-term roadmap to ensure business growth and success. In the age of automation and machine learning, we need to fully capitalize on the opportunities these technologies bring to the table and implement a solution that sustains itself with minimal manual intervention.



About Scalable AI

We deliver actionable insights and predictive analytics that organizations can use to identify opportunities, manage risks, achieve operational excellence, and to gain an innovative edge.

www.scalableai.com

About Scalable Systems

Scalable Systems is a Data, Analytics & Digital Transformation Company focused on vertical specific innovative solutions. By providing next generation technology solutions & services, we help organizations to identify risks & opportunities, achieve sales & operational excellence to gain an innovative edge.

www.scalable-systems.com