International Journal of Core Engineering & Management
Volume-8, Issue-02, 2025 ISSN No: 2348-9510

COST OPTIMIZATION STRATEGIES IN CLOUD DATA WAREHOUSING: A COMPARATIVE STUDY OF AWS REDSHIFT AND SNOWFLAKE

Ujjawal Nayak
Experian Information Solutions, Inc.
Costa Mesa, CA, USA

Abstract

Optimizing costs in cloud data warehousing is crucial as enterprises seek scalable, efficient, and economically sustainable solutions. AWS Redshift and Snowflake are among the leading platforms used for cloud data warehousing, each offering unique strengths. This paper explores strategies to achieve cost optimization in AWS Redshift and Snowflake, including data ingestion efficiency, storage optimization, query performance tuning, resource management, and effective monitoring.

Index Terms—AWS Redshift, Snowflake, Cost Optimization, Cloud Data Warehousing, Performance Tuning, Resource Management.

I. INTRODUCTION

Cloud data warehousing solutions such as AWS Redshift and Snowflake are widely adopted due to their scalability, flexibility, and analytical power. However, managing operational costs effectively remains a critical challenge. Efficient cost management in these platforms involves comprehensive strategies that optimize resource usage without compromising performance [1], [2].

II. STRATEGIES FOR AWS REDSHIFT COST OPTIMIZATION

A. Data Compression and Distribution Keys

Selecting appropriate compression encodings and distribution keys reduces storage costs and improves query performance [3].

B. Efficient Data Ingestion

Using COPY commands for bulk loading and AWS Glue for data transformation optimizes ingestion workflows [4].

C. Query Optimization

Regularly analyzing query plans and performance through Redshift's Query Analyzer and recommending automated improvements ensures resources are efficiently utilized [5].

D. Auto Scaling and Cluster Management

Utilizing Redshift's Elastic Resize and concurrency scaling capabilities allows dynamic scaling of clusters according to workload demands [6], [7].

E. Data Archiving and Lifecycle Management

Leveraging Amazon S3 for data archiving and Redshift Spectrum for querying archived data significantly reduces storage expenses [8].

Source Systems (RDBMS, NoSQL, Files) --> AWS Glue (ETL Processing) --> Amazon S3 (Staging) --> Redshift COPY Command --> AWS Redshift

Fig 1: AWS Redshift Data Load Flow

III. STRATEGIES FOR SNOWFLAKE COST OPTIMIZATION

A. Warehouse Size and Auto-Suspend

Selecting appropriate warehouse sizes and configuring auto-suspend periods can dramatically reduce idle compute costs [9].

B. Efficient Data Storage Management

Leveraging Snowflake's automatic compression and micro-partitioning reduces storage costs [10].

C. Optimized Query Design

Creating generic views and using query profiling to optimize queries minimizes resource usage [11].

D. Resource Monitoring and Management

Continuous monitoring of resource utilization through Snowflake's built-in Resource Monitors allows proactive adjustments [12].

E. ELT Pipeline Optimization

Utilizing efficient Extract-Load-Transform (ELT) approaches reduces processing overhead [13].

Source Systems (RDBMS, NoSQL, Files) --> Cloud Storage (AWS S3, Azure Blob Storage) --> Snowpipe/External Stages --> Snowflake Tables

Fig 2: Snowflake Data Load Flow

IV. LIMITATIONS/CHALLENGES

Challenges include complexity in migrating legacy ETL processes, ensuring compliance with data security standards, managing costs during deployment and scaling, and potential performance degradation [14].

V. FUTURE SCOPE

Future work includes AI-driven resource prediction, automated query analysis tools, enhanced data governance frameworks, and real-time analytics [15].

VI. COMPARATIVE ANALYSIS AND RECOMMENDATIONS

Table I: Feature/Strategy Comparison

Feature/Strategy	**AWS Redshift**	**Snowflake**
Data Compression	Manual selection needed	Automatic compression
Scaling	Elastic Resize, Concurrency Scaling	Instant auto-scaling
Query Optimization	Manual tuning, Query Analyzer	Automatic query profiling
Data Ingestion	COPY command, AWS Glue	Snowpipe, External Stages
Storage Management	Amazon S3 Archiving	Automatic micro-partitioning
Resource Monitoring	AWS CloudWatch integration	Built-in Resource Monitors
Cost Management	Reserved Instances, Predictable cost	Pay-per-use pricing
ETL/ELT Pipeline	ETL-based traditional approach	ELT-based modern approach

AWS Redshift suits predictable workloads with extensive AWS integrations, while Snowflake offers flexibility and scalability. Continuous evaluation of workloads and operational needs is recommended [16].

VII. CONCLUSION

Cost optimization in AWS Redshift and Snowflake is achievable through strategic resource management, optimized data practices, and monitoring, maximizing ROI.

REFERENCES

1. AWS Redshift Documentation, 2024. https://aws.amazon.com/redshift/
2. Snowflake Documentation, 2024. https://docs.snowflake.com
3. AWS Redshift Compression Guide, 2024. https://docs.aws.amazon.com/redshift/latest/dg/c_Compression_encodings.html
4. AWS Glue Documentation, 2024. https://docs.aws.amazon.com/glue/latest/dg/what-is-glue.html
5. AWS Redshift Query Analyzer Guide, 2024. https://aws.amazon.com/redshift/features/query-analyzer/
6. AWS Elastic Resize Documentation, 2023. https://docs.aws.amazon.com/redshift/latest/mgmt/elastic-resize.html
7. AWS Concurrency Scaling, 2023. https://docs.aws.amazon.com/redshift/latest/dg/concurrency-scaling.html
8. AWS Redshift Spectrum Guide, 2023. https://aws.amazon.com/redshift/features/spectrum/
9. Snowflake Warehouse Management, 2024. https://docs.snowflake.com/en/user-guide/warehouses-overview.html
10. Snowflake Storage Optimization, 2024. https://docs.snowflake.com/en/user-guide/tables-storage.html
11. Snowflake Query Profiling Guide, 2024. https://docs.snowflake.com/en/user-guide/ui-query-profile.html
12. Snowflake Resource Monitoring, 2024. https://docs.snowflake.com/en/user-guide/resource-monitors.html
13. ELT Pipeline Optimization, Snowflake, 2024. https://docs.snowflake.com/en/user-guide/data-pipelines.html
14. Cloud Data Warehousing Challenges, IEEE, 2023. https://ieeexplore.ieee.org
15. AI and Analytics in Cloud, IEEE, 2024. https://ieeexplore.ieee.org
16. Cloud Data Warehousing Comparative Study, IEEE, 2023. https://ieeexplore.ieee.org
17. S. Ahmad and R. Patel, "Optimizing Snowflake Costs: Techniques for Query Optimization and Warehouse Scaling," Int. J. Cloud Comput., 2022.
18. L. Schmitz, "Maximizing Efficiency in Snowflake: Best Practices for Reducing Storage and Compute Costs," Cloud Data Strategies J., 2021.
19. A. Kapoor, "Reducing Snowflake Costs: Auto-Suspend, Resource Monitoring, and Query Optimization," Data Eng. J., 2023.