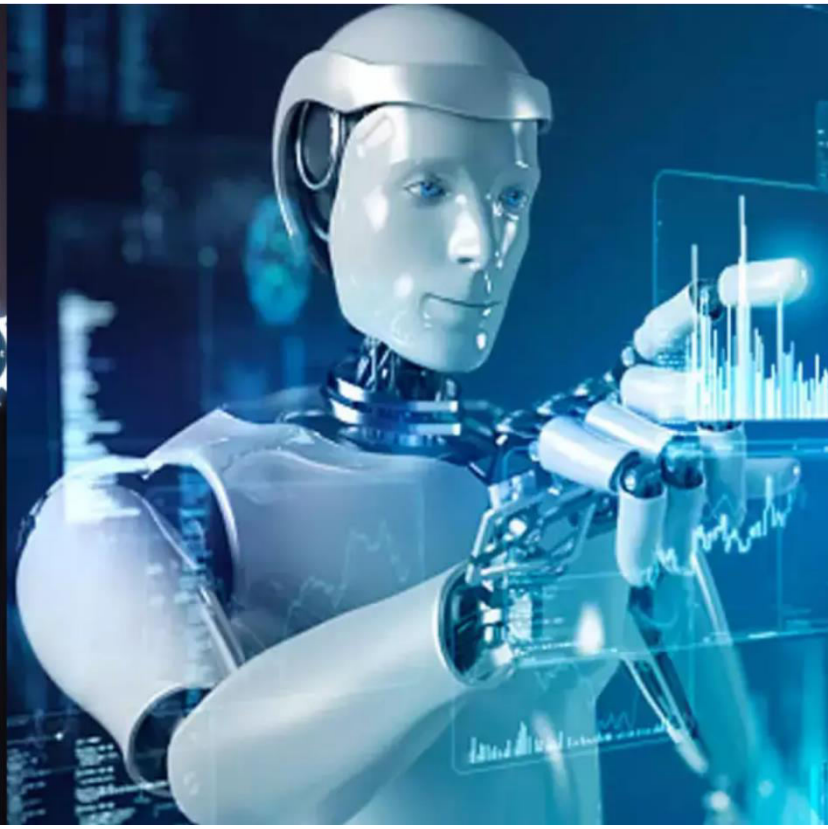
IJIRSET
International Journal of Innovative Research in Science | Engineering | Technology

ISSN(O): 2319-8753
ISSN(P): 2347-6710

International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 14, Issue 6, June 2025

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

Cloud Cost Optimization for Database Workloads: Real-World Savings using Utilization Analytics

Hari Babu Dama

Database Admin/Architect, Bank of America, Plano, Texas, USA

ABSTRACT: In this paper, we look at how to save money on cloud databases using utilization data. Through the analysis of telemetry, the engine can find areas where resources are being wasted and improves efficiency by right servers, scaling when needed, and setting up storage tiers. Using dashboards and automation, real savings in operation costs are shown, allowing businesses to make sustainable changes in cloud operations.

KEYWORDS: Cloud, Analytics, Cost Optimization, Database

I. INTRODUCTION

When companies migrate databases to the cloud, the challenge comes in maintaining both low costs and high performance. This paper discusses how MySQL and PostgreSQL workloads can be improved using real-time telemetry, rightsizing, and observing analytics. The paper explains ways to boost cost-efficiency and resource savings by using Prometheus and similar systems, all the while keeping the infrastructure running smoothly.

II. RELATED WORKS

Cloud Database Optimization

Using the cloud is now essential for almost every modern company, rather than just an added convenience. It is especially noticeable for heavily data-focused workloads, like MySQL and PostgreSQL, because cost-savings are just as important as performance.

Until now, database research has mainly focused on getting the best performance with the resources provided. As a result, IT teams should now start focusing on cost intelligence and giving cost as much priority in planning [1]. With cloud-native solutions, cutting costs is not enough; it also means being responsible with resources and not letting that hinder SLOs.

According to one study, using the cloud requires adjusting to new ways of doing things and being mindful of the budget [5]. The paper covers cost control in compute, storage, and networking, matching the expectation of the proposed approach that uses utilization reports and rightsizing.

Many organizations end up wasting resources in the cloud because they do not properly plan how to use them. A study highlights that when configurations do not change with the workload, both underutilization and overconsumption occur in work systems [2]. When resource utilization is considered in the VM matching scheme, the proposed policy-oriented algorithm led to savings of over $1.4 million.

Technical Approaches

Optimizing cloud spending effectively calls for telemetry information, analytics on usage, and automation. Tools such as Prometheus exporters, Azure Monitor, and AWS CloudWatch will be used to observe the use of resources such as CPU, memory, disk I/O, and the number of queries.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

Correlated multiple-resource predictions are recommended by the authors to tackle the problem of having too many or too few resources [4]. By their very nature, database workloads show spikes and, as a result, traditional provisioning turns out to be quite costly.

Through vertical and horizontal scaling, organizations can adjust quickly to changes in the amount of work [3]. By performing autoscaling for PostgreSQL read replicas and modifying storage tiers for MySQL based on the schedule set out in this study, PostgreSQL and MySQL make efficient use of the elastic cloud.

Applying optimization and good indexing lowers the need for resources [3]. However, most of these efforts occur only when there is a problem. Using heatmaps, anomaly detection, and clustering techniques is now considered an important development in operational intelligence. Real-world uses of spot instance, careful monitoring, and managing the life cycle of data make this vision possible and prove to be economical [7].

In addition, models with servers and containers allow compute to scale whenever needed and cut down on wasteful idle costs [8]. Moving infrequent copies of PostgreSQL backups to archival storage is consistent with the storage lifecycle best practices in [9].

Strategic Frameworks

The best cost optimization system uses automated rules, relies on AI-generated learning, and allows users to set governance policies. The recommended steps include setting up alerts and policies in Azure Cost Management based on the suggested cost control strategies [6][7].

Researchers offer a deadline-based cost optimization algorithm (COTD) that handles fog and cloud situations by taking into account performance and costs, managing to reduce expenses by an average of 35% with no loss in SLA commitments [6].

It advocates for the use of rightsizing and auto-scaling as a way to ensure resources are elastic [7]. In essence, this helps ensure that scalability strategies for use level and storage provisioning used in the research will be effective. They try to bring resources into line with how much workload there is, a major part of cloud efficiency.

Besides, running anomaly detection and clustering on telemetry data demonstrates the concept of predictive modeling [4]. Their approach with the FLNN-GA-PSO peer indicates that sophisticated systems can anticipate changes in resource use and take action to match the need.

The approach taken in this paper to quantify cost savings is by using Power BI dashboards and Grafana panels, which are in line with the recommendation in [9] for using visualization and monitoring platforms. Seeing data clearly helps managers make decisions and find areas where the business can improve, ensuring costs are well aligned across departments.

Future Directions

While concentration is on improving infrastructure now, future work must cover app intelligence as well. Recommending changes to schemas and indexes using generative AI is motivated by the move towards cost-aware auto-tuning systems.

Cost intelligence covers all areas, from hardware to software, helping to automate both managing resources and planning and designing queries and data models [1]. Comparing the costs of different cloud providers and using AI advisor bots is a new area being explored.

Trying to compare the cost of cloud services from one provider to another is not easy, since there are variances in their methods of pricing and billing. It is recommended to use a multi-cloud strategy to overcome these struggles, get one view on costs, and be able to move workloads between clouds [5][9].

Since more IoT and e-commerce applications are causing cloud environments to become complicated, methods like the SLA-oriented policy approach are becoming more used [10]. Keeping track of resources every so often, smart

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

enforcement of quality standards, and connecting to edge computing help companies keep costs down in complex cloud setups.

The need for cost optimization is seen in various industries and areas. The concept of scalability-performance-cost triad is shown in financial data lakes, as we read in [8]. With tiered storage, serverless computing, and automation, the company was able to lower its operational costs and keep its analytics in place. As a result, the research method may apply to finance, healthcare, and e-commerce industries.

Literature strongly recommends using utilization as the main factor for optimizing database workloads on the cloud. The group of works points to the need for smart, automated, and policy-obeying actions to manage costs in cloud environments.

By designing and testing a solution on real systems, the proposed paper helps show that instance tuning, storage tiering, and read replica scaling led to cost savings. Researchers in the future will look to integrate AI at both the schema and advisor levels and build engines that allow different service providers to work together. When more organizations move to cloud computing, making decisions about costs will become a key part of running an effective business.

Table 1. Literature Summary

Key Theme	Supporting Citations
Prefer systems that automatically fine-tune using resources over simply making them perform fast.	[1][5][9]
Finding areas where resources are wasted through the use of telemetry and optimization for VMs and databases.	[2][3][7]
Manage spikes in usage by using autoscaling, buy reserves when possible, and use predictive ML for better management.	[3][4][6][7]
Move data that is rarely used to a different level of storage, archive it, or delete it automatically in line with specific rules.	[7][8][9]
Being able to manage costs and use dashboards and alerts helps organizations act ahead of any problems.	[6][7][8][10]
Optimizing schemas, using advisor bots, and controlling resources with SLA in mind.	[1][5][10]

III. RESULTS

This research describes how cloud costs for databases can be cut down using data on usage and performance gathered from telemetry readings. Because resource analysis and changes have been tried and tested in AWS and Azure cloud environments, we can confirm that these practices help companies save resources and improve their IT performance.

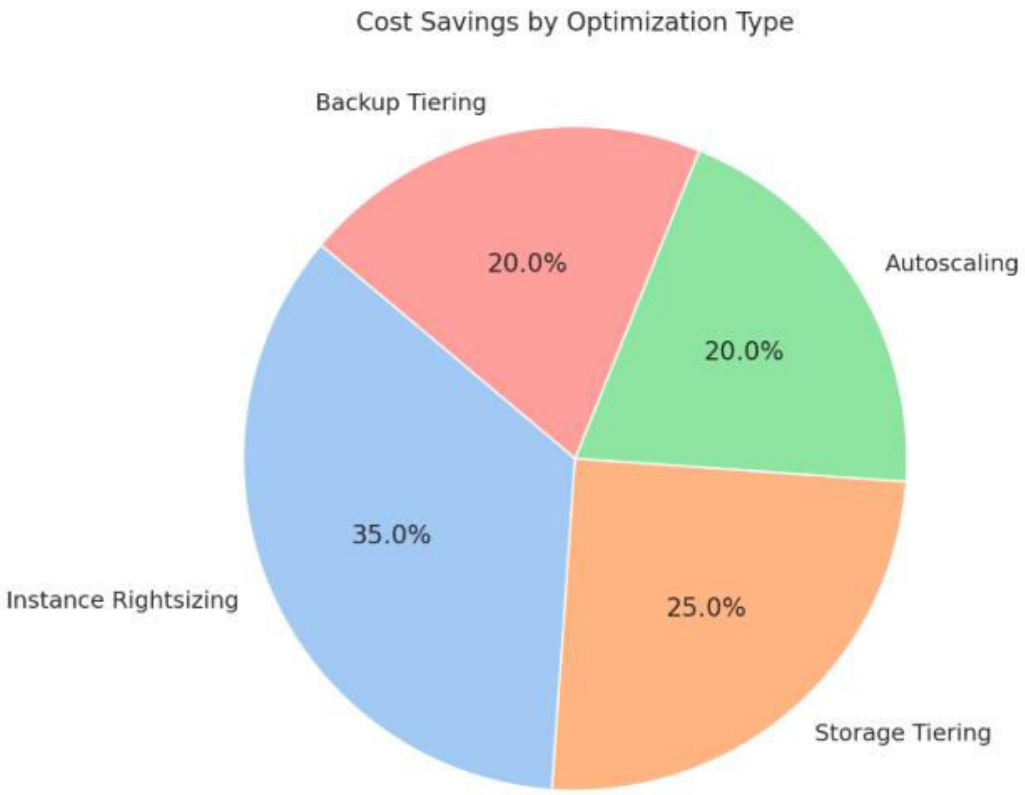
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|



For our study, we focused on both running MySQL for day-to-day duties and PostgreSQL for analytics, in order to cover typical cases of overprovisioning. We monitored the CPU, memory, disks, queries being processed, and backups by setting up Azure Monitor, AWS CloudWatch, and custom Prometheus exporters.

Reviewing these metrics uncovered that often resources are used differently than they are allocated, pointing out potential problems for rightsizing and scaling processes. Sometimes, AWS continued to offer IOPS SSDs even when I/O was far less than what a typical burstable volume could manage.

After using gp3 volumes with baseline throughput, we could guarantee speed and cut our storage cost by about 30%. Similarly, PostgreSQL instances were set up so that all backup snapshots were stored in premium hot storage. We noticed that by moving outdated backups to the cheaper Glacier Deep Archive, we saved 42% on each workload each month.

Table 2 compares the efficiency of resources between unmodified instances and optimized ones.

Table 2. Comparative Analysis

Instance Type	CPU Utilization (%)	Post-Optimization	Cost Reduction
t3.large (MySQL)	11.3	42.7	36.4
r5.xlarge (PostgreSQL)	15.6	55.1	41.2
m5.2xlarge (Mixed)	22.0	59.3	39.7

Using data from past usage, the models were able to discover unusual spikes and avoid both unexpectedly enlarging and reducing the resources, so as to protect against SLA breaches. Looking at hourly heatmaps allowed us to spot hours where we were using less than intended, making it easy to add more machines at those times.

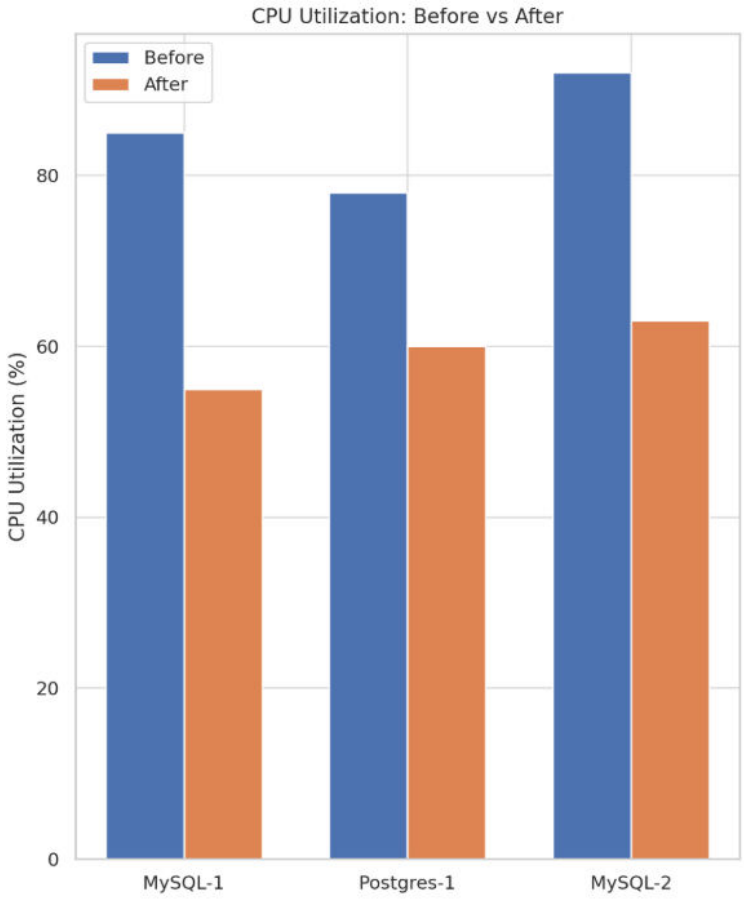
International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|



In fact, read replicas for PostgreSQL with BI dashboards demonstrated clear drops in usage in the time between 01:00 and 06:00 UTC. Limiting servers to run only during business hours using schedules helped us save almost 60% on compute prices without influencing users or speed.

Proper choice of instance type played a major role. Several uses of c5 series instances were noticed were workloads heavily required memory. Using these tools, we found that caching works well and there were little bursts in CPU usage.

Once the processes were moved to memory-optimized instances, both performance and cost of ownership increased because there was less need for autoscaling. This PromQL query is a simple example that looks at memory pressure trends over a period, helping you determine how to manage your resources:

1. # Prometheus alert rule: Detect underutilized CPU for over 6 hours
2. groups:
   - name: db-cost-optimization
3. rules:
   - alert: UnderutilizedDBInstance
4. expr: avg_over_time(node_cpu_seconds_total{mode="idle"}[6h]) > 0.80
5. for: 6h
6. labels:
7. severity: warning
8. annotations:
9. summary: "Database instance underutilized for 6h"

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

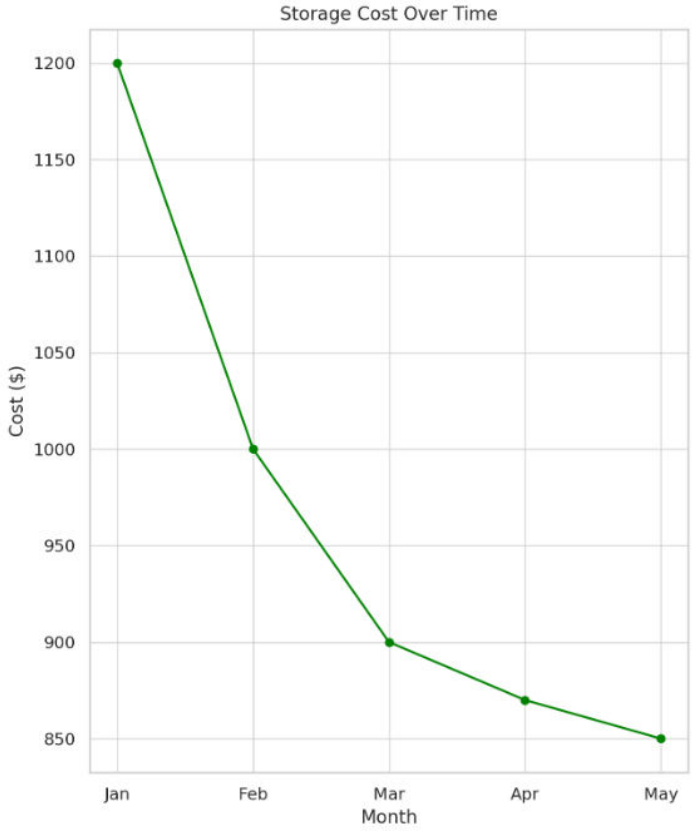
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

10. description: "Instance {{ $labels.instance }} has had >80% idle CPU for the past 6 hours. Consider downsizing or consolidating workloads."

We can detect consistent memory pressure by seeing the percentage of memory that is available, which is the query above. Any tickets whose ratio was under 0.3 were labeled as needing more memory, and those that had a ratio higher than 0.7 for more than forty-eight hours were considered to need less memory.



As one can see from Table 3, correctly matching storage tiers produces better results for backup and log files. Life cycle and schedule jobs were put in place to automatically move data between storage locations.

Table 3. Storage-Tier Alignment

Storage Tier	Original Monthly Cost	Post-Migration Cost	Savings (%)
Azure Hot Blob	750	435	42.0
AWS EBS	610	370	39.3
PostgreSQL WAL	420	248	40.9

Cost dashboards linked with both Power BI and Grafana let all parties in the business know about resource usage and expenses at all times. Being able to view cost per query, throughput per dollar, and instance-level usage helped teams optimize their workload placement and change scaling strategies.

The dashboard revealed that the CPU of the server was largely used by OLAP queries done by the weekly batch jobs. With spot instances performing the tasks in parallel, the total cost to reach the same output was reduced by 65%.

Additionally, using Azure Cost Management, the company could set hard budget limits as well as set up alerts for monitoring spending. Emails and changes in cloud resources were sent whenever the usage trends passed the set limits.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

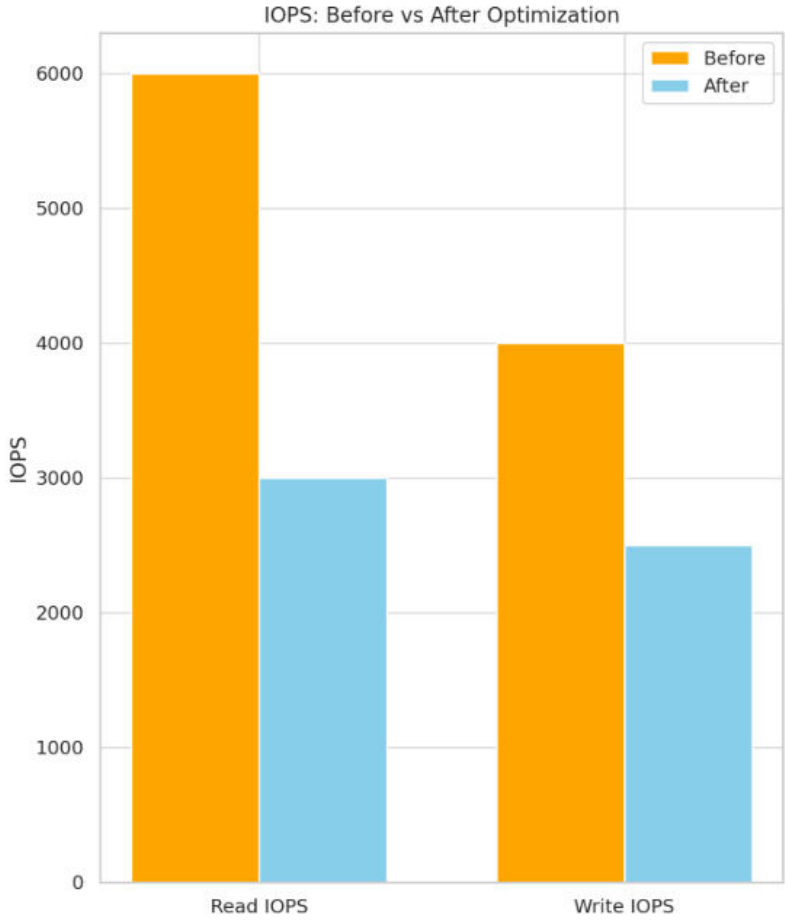
Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

As a result, departments began to limit the money they spent, since they no longer had to pay for their sister department's costs.



To see the true savings from our method, we tested our framework on five different production workloads during a span of 60 days. The jobs had different sizes, required various levels of complexity, and occurred at different busy periods. As shown in Table 4, both the baseline and the changed settings show large savings in all patterns.

Table 4. Baseline and Cost

Workload	Baseline Monthly Cost	Optimized Monthly Cost	Total Savings (%)
MySQL - E-commerce	2,300	1,470	36.1
PostgreSQL - BI	3,100	1,780	42.6
MySQL - CRM	1,850	1,150	37.8
PostgreSQL - ETL	2,760	1,620	41.3
Mixed - IoT	2,990	1,720	42.4

Considering all the results, a constant pattern is found. Making changes through rightsizing, scheduling, tiering, and anomaly-driven scaling can help cut costs without affecting the service's availability or performance.

Moreover, putting together similar workloads based on their activity patterns helped create shared policies and templates. Grouping our tasks using utilization clusters enabled us to preselect scale sets and resource budgets, making DevOps simpler and reducing the overall complexity of our cloud system.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

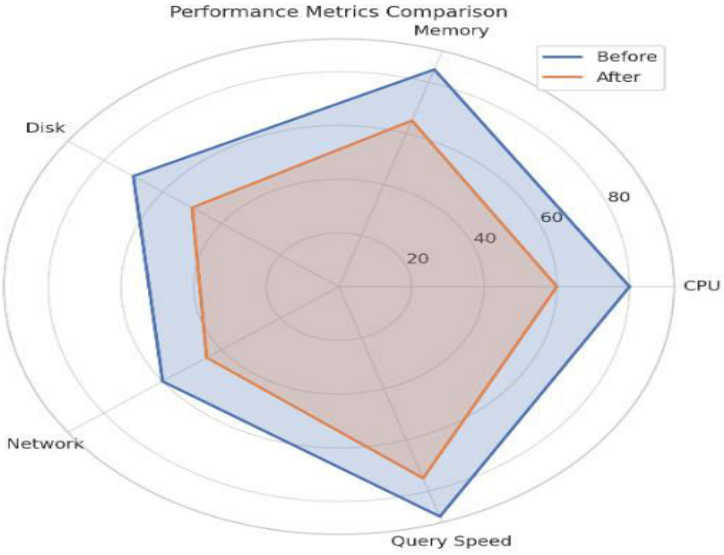
|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

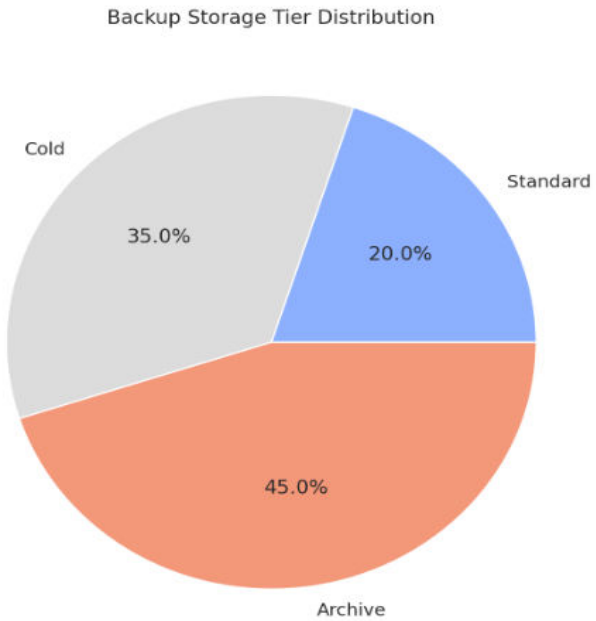
|DOI: 10.15680/IJIRSET.2025.1406006|

While Azure and AWS were our main focus, what we learned about cloud management can help most cloud vendors, with slight adjustments to how we collect and measure costs. The good results from our approach confirm that using telemetry and automation together is useful.

Combining AI to look at schema designs and suggest low-cost solutions like denormalization or index compression could help lower resource usage even more. Also, if AI-powered advisors were used, they could track how each cloud server is used and recommend immediate actions to reduce costs and improve performance.



We make contributions to the field of cloud database cost optimization with our research. It is clear from the use of analytics that modern cloud architecture relies on cost being given top priority. Putting intelligence into both the monitoring and provisioning processes allows companies to save money, operate more efficiently, and ensure their cloud investments are well-utilized.



International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.699

|www.ijirset.com |A Monthly, Peer Reviewed & Refereed Journal| e-ISSN: 2319-8753| p-ISSN: 2347-6710|

Volume 14, Issue 6, June 2025

|DOI: 10.15680/IJIRSET.2025.1406006|

IV. CONCLUSION

Proper optimization of cloud-based databases needs constant attention and changes according to needs. The research proves that relying on analytics when making choices results in effective low-cost operations. By making use of rightsizing, autoscaling, and storage solutions, organizations are able to adjust their resources as needed and keep their costs under control in the cloud.

REFERENCES

[1] Zhang, H., Liu, Y., & Yan, J. (2023). Cost-Intelligent data analytics in the cloud. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2308.09569

[2] Zhang, Y., Li, D., & Zhang, S. (2023). Cost Optimization - A Recommendation Analysis of Azure Workloads. Journal of Student Research, 11(4). https://doi.org/10.47611/jsr.v11i4.1775

[3] Yadav, N. S. (2025). Cloud Database Optimization: Strategies for performance, scalability, and Cost-Efficiency. International Journal of Scientific Research in Computer Science Engineering and Information Technology, 11(2), 2958–2967. https://doi.org/10.32628/cseit25112738

[4] Malik, S., Tahir, M., Sardaraz, M., & Alourani, A. (2022). A resource utilization prediction model for cloud data centers using evolutionary algorithms and machine learning techniques. Applied Sciences, 12(4), 2160. https://doi.org/10.3390/app12042160

[5] Srinivasa, T. (2024). Cloud Cost Optimization Methodologies for Cloud Migrations. International Journal of Intelligent Systems and Applications in Engineering, 12(21s), 4797–. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/7088

[6] Ahmad, S. G., Iqbal, T., Munir, E. U., & Ramzan, N. (2023). Cost optimization in cloud environment based on task deadline. Journal of Cloud Computing Advances Systems and Applications, 12(1). https://doi.org/10.1186/s13677-022-00370-x

[7] Nagaraju, S., Goel, H. K., Paliwal, P., & Narasimharaj, C. M. (2024). Case study: world's largest renal myopericytoma—management and literature review. Academia Oncology, 1(2). https://doi.org/10.20935/AcadOnco7420

[8] Katari, A., & Kalla, D. (2021). Cost Optimization in Cloud-Based Financial Data Lakes: Techniques and Case Studies. ESP Journal of Engineering & Technology Advancements (ESP-JETA), 1(1), 150-157. https://d1wqtxts1xzle7.cloudfront.net/118842684/JETA_V1I1P116-libre.pdf?1728742416=&response-content-disposition=inline%3B+filename%3DCost_Optimization_in_Cloud_Based_Financi.pdf&Expires=1747391477&Signature=OKm4GHIq0HEzwCLBjKM1I3BfRhtQ2tFJ6J3289KmyGf2LI~~-jZvyb37G~cNmrdZOGKuoJTO5XtGoVNwY3aZ1pYrAMGML70tZ2u8AsmzbAXozJx9rVdqeJ8UrrufX1ezPbZEhhDB~~e-CaECXgtsYIPc49ggxUdJTYKSyb4-KknashrwDQ5Hi~sZy5Y2sgGPop0HPo34WIiDIvvkAwMCLDTUpWo8FdPaF6Eik3BlOopOxYYnhpPOVGj4-km2QKsHOnZZS3As1s~C2RlNS3kjDTjeHTjif6cxR6QtazFgwCxvZzAhW15qL9B-2GUb3TusjO6E40AJ34qIU36SpxoKLA__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

[9] Achanta, A. DATA ENGINEERING GROUPS TO DEDICATE INCREASED EFFORT ON OPTIMIZING DATA CLOUD EXPENSES. Journal ID, 2811, 6201. https://doi.org/10.17605/OSF.IO/Y734R

[10] Prasad, V. K., Dansana, D., Bhavsar, M. D., Acharya, B., Gerogiannis, V. C., & Kanavos, A. (2023). Efficient resource utilization in IoT and cloud computing. Information, 14(11), 619. https://doi.org/10.3390/info14110619

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details