Case Study

UnifyCloud

Slashing Cloud Infrastructure Costs by 40% with AI-Driven Cloud Optimization

THE OVERVIEW

A leading technology organization was experiencing growing pains – increasing IT infrastructure expenses in their Azure environment – and they needed better tools to control utilization and cost. By implementing UnifyCloud's Cloud Optimization, a SaaS-based cloud and AI optimization tool, the company shifted from tedious manual oversight to automated, data-driven resource management.

THE CHALLENGE

The company was experiencing significant growth, with new employees, projects and resources coming online. With increased development work and extensive use of AI, the organization faced increasing compute costs, which accounted for most of their cloud spend due to the high price of these services. Key issues included:

- **Poor Resource Management**: No structured approach to optimize costs for VMs, App Services, and SQL Databases
    - **Over-Provisioned Resources**: Virtual Machines (VMs) were often created in higher tiers than required for their actual workload.
    - **Orphaned Resources**: Lack of visibility into idle or unused resources led to unnecessary costs.
    - **Reserved Instance Visibility**: No process in place to take advantage of cost-saving RI-based opportunities.
- **Token Consumption**: Limited visibility and insight into token usage meant minimal ability to control AI usage costs.
- **Idle Infrastructure**: Development environments were running 24/7, despite typically only being needed during business hours.
- **Infrastructure and Inventory Drift**: Inadequate monitoring of resource creation, deletion, regional changes, and SKU configuration drift made it difficult to know what and where resources were being used and what they were used for.
- **Shadow IT**: Unapproved and under-utilized production VMs with heavy cores and RAM were being created without oversight, leading to unexpected cost spikes.

With so much going on, constant manual oversight to address these issues was difficult and time-consuming – a losing proposition. They needed to find a way to better manage their IT environment and eliminate unnecessary expense and do this cost-effectively without requiring additional dedicated resources.

CloudAtlas Optimization Case Study

UnifyCloud

THE SOLUTION

The company deployed CloudAtlas Cloud Optimization to better monitor and manage their IT infrastructure. With an easy 15-minute setup, they were able to get 24/7 monitoring across their tenant with automated management tools to keep costs in check. The tool leverages an agentic approach to provide three levels of optimization:

- **Utilization Analytics:** Cloud Optimization monitors CPU and memory consumption over a rolling 30-day period. If utilization is identified at less than a set minimum, 40% in this case, the system provides workload reallocation or downsizing recommendations.
- **Automated Scheduling:** Using Microsoft Graph APIs, the company implemented auto-shutdown and restart schedules to eliminate costs incurred during idle hours. Machines can be excluded and the timing can be configured regionally to ensure accessibility.
- **Resource Approvals:** The system provides a single-click review capability. New resources are flagged for human administrator oversight where they can provide consent or denial directly in the dashboard.
- **Agentic Rightsizing:** An AI agent monitors resources for utilization, configuration, and other changes and automatically adjusts or reconfigures VMs, App Services and SQL Databases to a more cost-effective configuration when needed.
- **AI Assistant Support:** Using natural language processing, users can query an AI assistant to better assess proper resource configuration to avoid incorrect sizing and setup. They can also ask questions about any resource and get instant answers.

THE RESULTS

The implementation of AI Powered Cloud Optimization delivered significant financial and operational improvements without adding new resource requirements:

- **Reduced Operational Expense:** The organization achieved an overall 40% reduction in cloud costs thanks to automated scheduling, rightsizing and better Reserved Instance utilization.
- **Inventory Drift and Shadow IT Savings:** Monitoring for orphaned resources and shadow IT enabled an instant $7,085 in monthly savings in a single tenant. Another right-sizing recommendation for a single VM identified immediate savings of $3,555 monthly.
- **Proactive Management:** Real-time expense visibility with hourly trend analysis and intelligent anomaly detection enabled the organization to improve budgeting accuracy. They can quickly identify irregular expense, respond to provisioning or usage changes, and take decisive action to eliminate surprises and hit budget targets within 5%.
- **Token Control**: Token tracking across teams and projects identifies cost drivers, applies predictive analytics, and monitors for anomalies to proactively control consumption, and enforces policy-driven governance to prevent overages and align AI usage with budget constraints.

By replacing manual resource management with UnifyCloud's agentic, data-driven automation, the organization transformed its cloud environment into a model of efficiency, reducing total infrastructure spend by 40% while materially improving budget accuracy and predictability. With rapid, low-friction deployment, the platform established 24/7 intelligent monitoring and autonomous optimization, immediately uncovering savings opportunities and enforcing proactive cost control through dynamic rightsizing, real-time anomaly detection, and automated scheduling that eliminated surprises and enhanced financial discipline.