RESEARCH ARTICLE

An Empirical Analysis of Server Utilization and Optimization Strategies in a Proxmox VE Environment

iD Levent Emmungil[1*]

[1]Ostim Technical University; 0000-0003-2220-5333

Corresponding author:
Levent Emmungil, Ostim Technical University
E-mail address: emmungil@gmail.com

ABSTRACT

Server underutilization in enterprise and institutional data centers represents a significant source of financial waste and environmental impact. While proprietary virtualization platforms are extensively documented in literature, open-source alternatives like Proxmox Virtual Environment (VE) require further empirical analysis to validate their role in cost-effective IT management. This paper presents a detailed case study investigating the resource utilization of two heterogeneous Proxmox VE servers deployed in a corporate data center. Through a multi-resolution analysis (yearly, monthly, weekly, daily, hourly) of telemetry data – including CPU, memory, swap, load average, and network statistics – we diagnose critical inefficiencies: severe memory over-commitment on one server and systemic underutilization on the other. These patterns highlight common pitfalls such as imbalanced resource allocation, inadequate workload distribution, and insufficient VM right-sizing. Based on our findings, we propose a set of actionable optimization strategies, including cross-server workload migration, dynamic resource allocation, and energy-aware consolidation. Our results, derived from a real-world deployment in one of the biggest technopolis in Türkiye, demonstrate that a data-driven approach to managing Proxmox VE infrastructures can substantially reduce capital and operational expenditures, mitigating the rising costs of server hardware and operations.

Keywords: Server Virtualization, Resource Optimization, Proxmox VE, Data-Driven Management, Cost Efficiency

1. Introduction

Institutional IT departments – spanning universities, government facilities, and corporate entities – face the persistent challenge of scaling computational resources to meet growing demand amidst constrained budgets. A seminal study by Barroso and Hölzle (2009) revealed that large-scale data centers frequently operate at a remarkably low 10–30% utilization, resulting in substantial wasted capital expenditure on idle hardware and operational expenditure on power and cooling. Virtualization has emerged as the primary technological solution to this problem, enabling higher consolidation ratios and improved resource elasticity.

While commercial platforms like VMware vSphere and Microsoft Hyper-V dominate both industry and academic discourse before, open-source alternatives have gained significant traction among cost-conscious organizations. Especially after new VMware licensing strategy the popularity of Proxmox VE increased drastically. Proxmox Virtual Environment (VE) is a prominent open-source platform that uniquely integrates full Kernel-based Virtual Machine (KVM) virtualization with lightweight Linux Containers (LXC) under a unified management interface. Despite its growing adoption, there remains a gap in empirical academic research focused on optimization strategies specific to Proxmox VE deployments.

Merely deploying a virtualization platform is insufficient; realizing its full potential requires a deliberate, metrics-driven strategy. Enterprises often collect extensive telemetry but fail to translate this data into actionable insights for optimization. This study aims to bridge that gap by conducting a granular analysis of two production Proxmox servers. We demonstrate how systematic evaluation of built-in usage statistics can uncover hidden inefficiencies and inform effective remediation strategies, ultimately leading to reduced costs and improved operational sustainability.

2. Method

To conduct this empirical analysis, we employed a structured methodology for data collection and evaluation.

2.1. Case Study Subjects

The study focuses on two production servers in an active data center of a technopolis in Türkiye:

Server1 (Has more than one active Virtual Server): A high-capacity server with two physical 48 logical cores and 62.5 GiB of RAM.

Server2 (Has a single active Virtual Server): A mid-range server with sing physical 16 logical cores and 31 GiB of RAM.

2.2. Data Collection

Resource utilization data was collected directly from the integrated Proxmox VE management dashboard over a continuous one year period. Data was aggregated and analyzed at multiple temporal resolutions to identify both long-term trends and short-term patterns:

Yearly and Monthly Trends: For identifying baseline utilization and long-term patterns.

Weekly and Daily Trends: For understanding workload cycles (e.g., weekday vs. weekend).

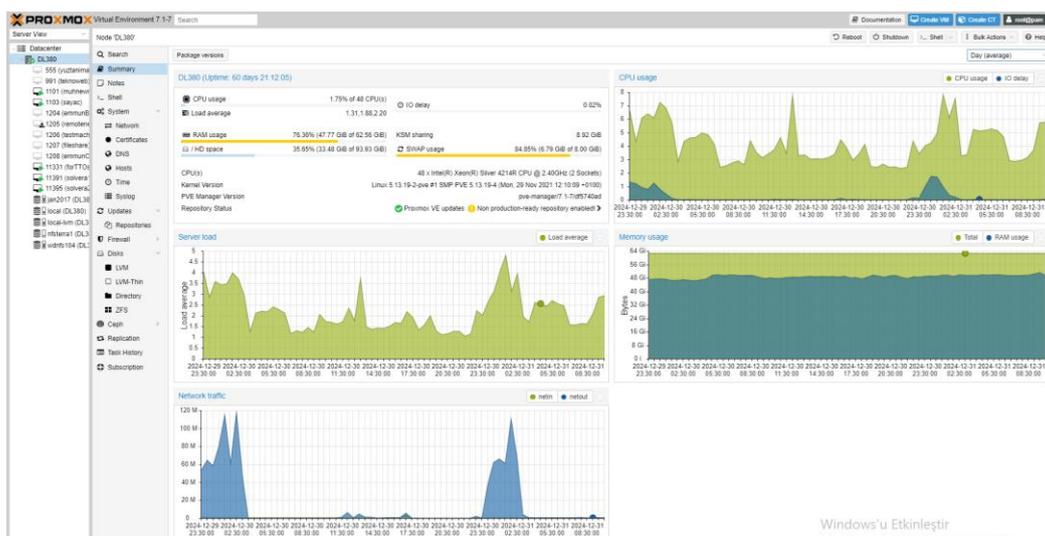Hourly Trends: For pinpointing peak usage and transient bottlenecks.
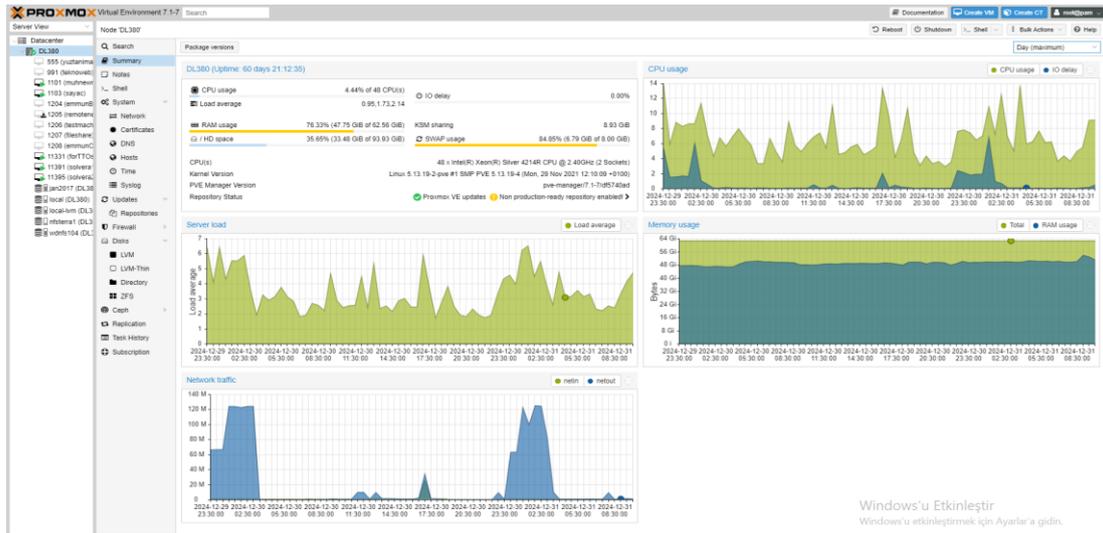

Figure 1: Server1 Daily Average
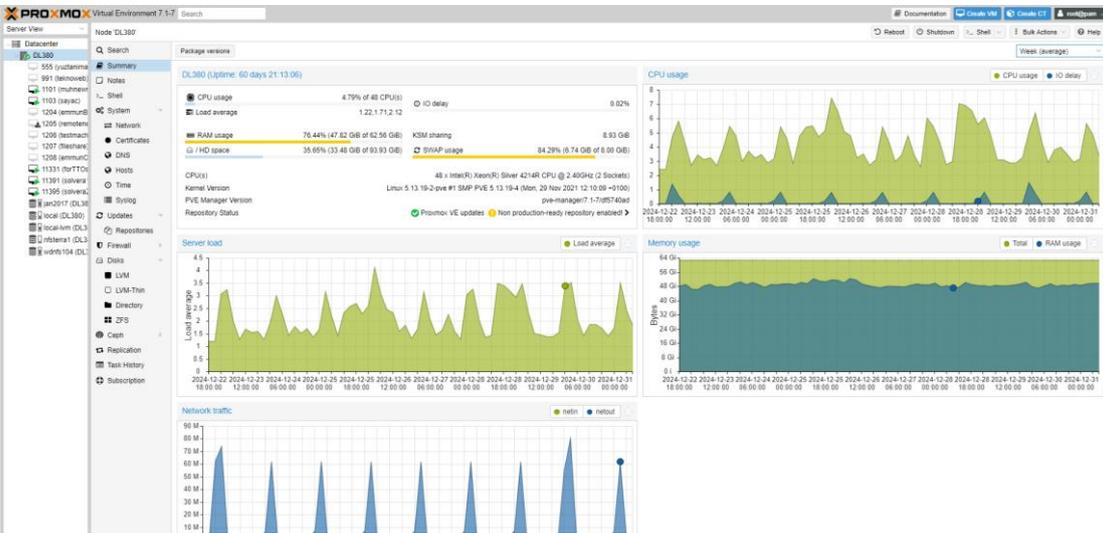
Figure 2: Server1 Daily Maximum

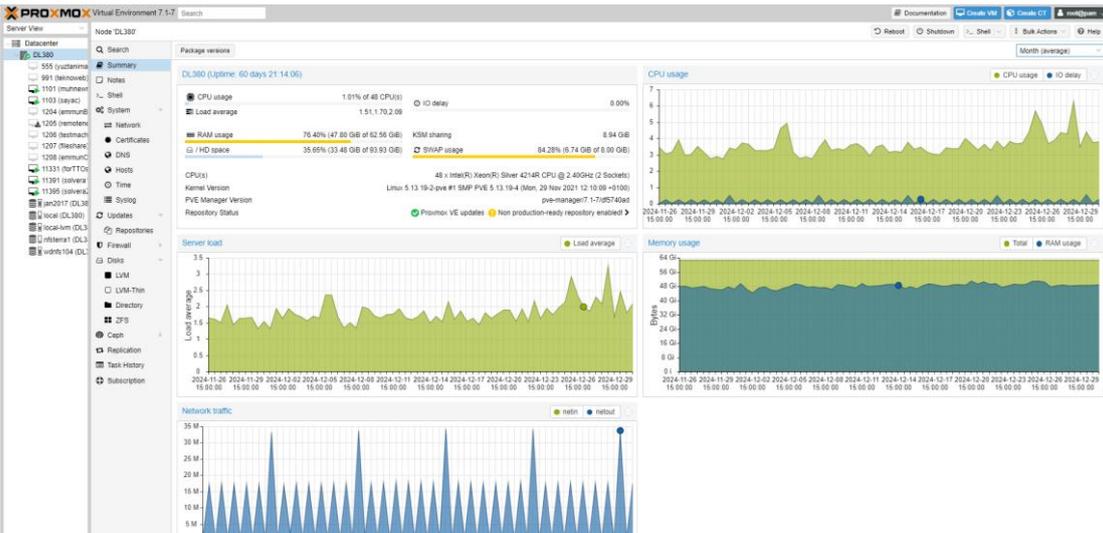**Figure 3**: Server1 Weekly Average

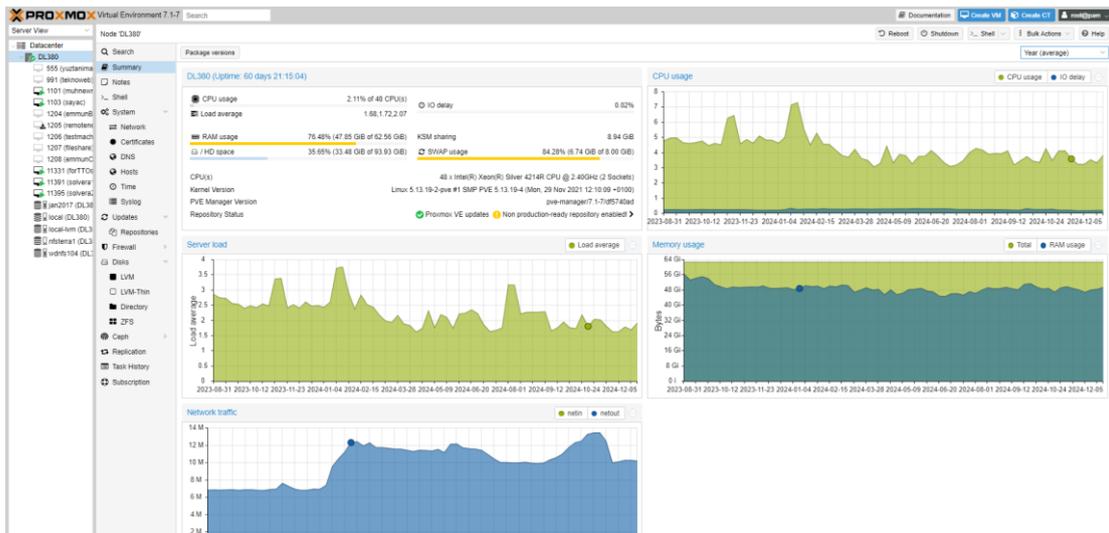**Figure 4**: Server1 Monthly Average

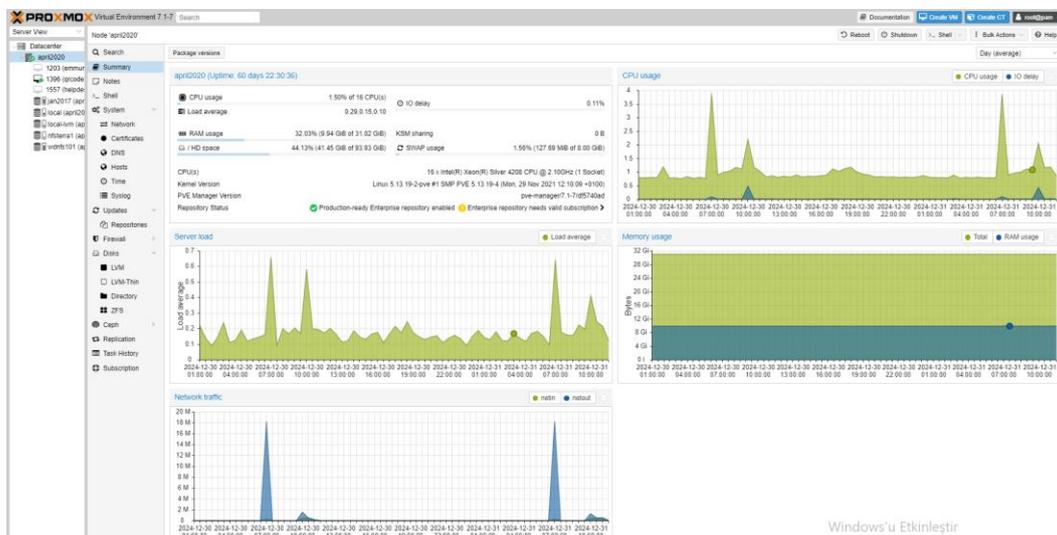**Figure 5**: Server1 Yearly Average

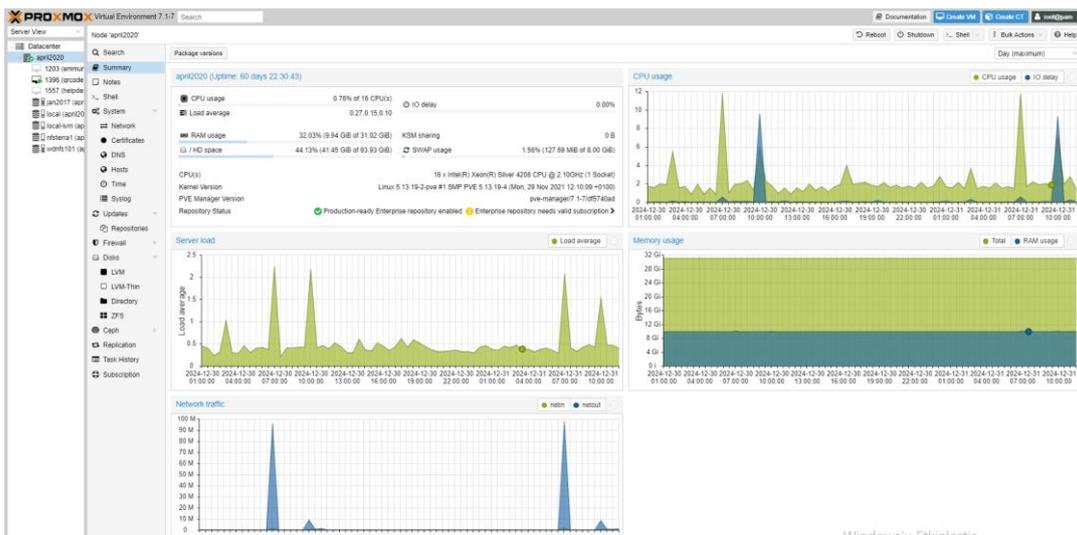**Figure 6**: Server2 Daily Average

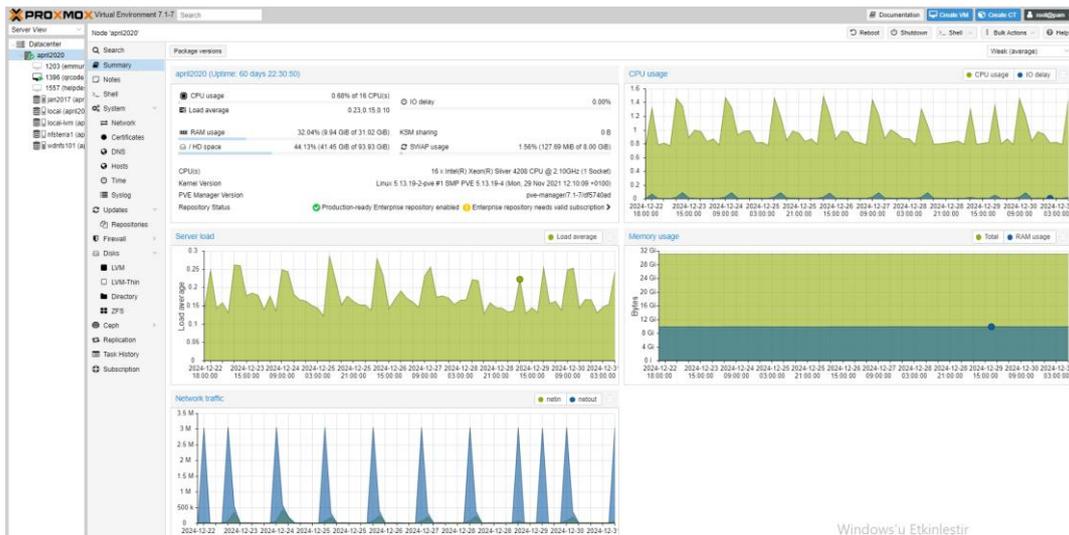
Figure 7: Server2 Daily Maximum
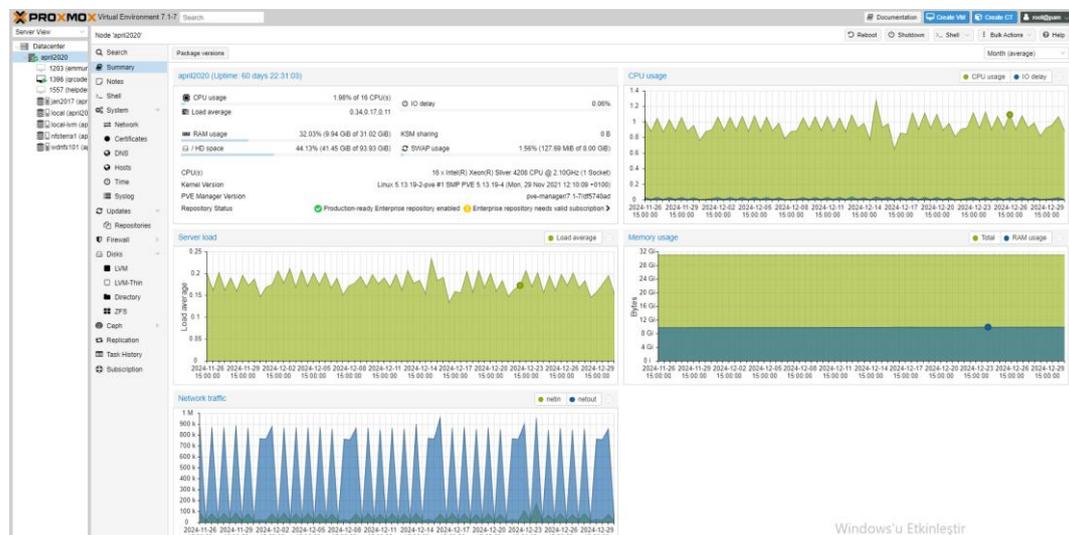
Figure 8: Server 2 Weekly Average

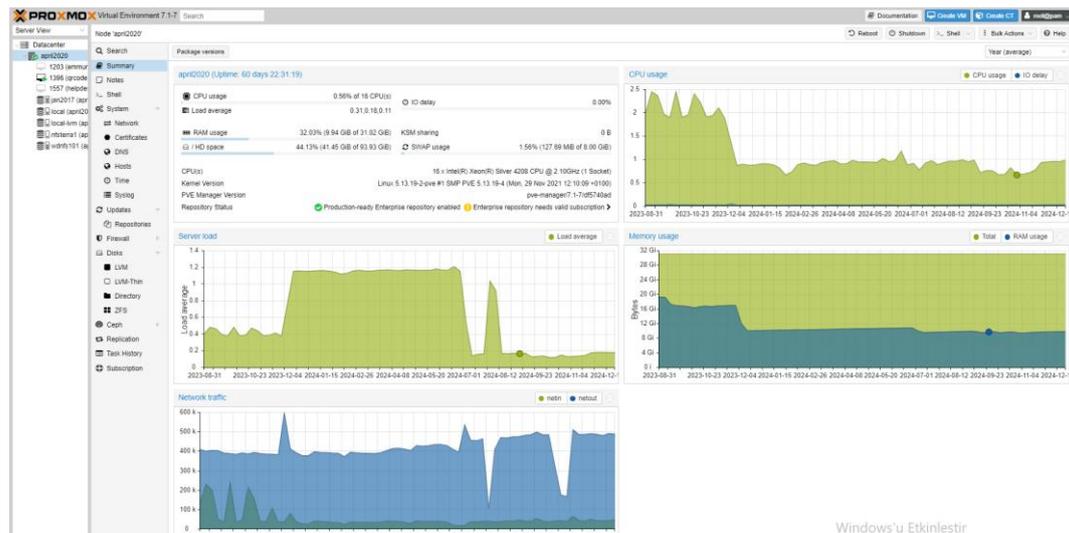**Figure 9**: Server2 Monthly Average

Figure 10: Server2 Yearly Average

2.3. Metrics Evaluated

The analysis focused on five key performance metrics:

- **CPU Utilization:** Percentage of total CPU capacity consumed.
- **Memory Utilization:** Allocated RAM versus actual usage, including swap activity.
- **Swap Usage:** Indicator of memory pressure and over-commitment.
- **Load Average:** System load over 1, 5, and 15-minute intervals.
- **Network Throughput:** Inbound and outbound data traffic.

2.4. Analytical Approach

Each server was evaluated individually to characterize its utilization profile. A comparative analysis was then performed to identify imbalances and opportunities for workload redistribution.

3. Findings

The telemetry data revealed starkly different and suboptimal utilization profiles for the two servers.

3.1. Individual Server Analysis

Server1: Exhibited critically low CPU utilization, averaging 2–4%. In contrast, memory utilization was consistently high at approximately 76% of its 62.5 GiB capacity. This memory pressure was further evidenced by significant swap usage (~85%), indicating severe over-commitment. This profile identifies Server1 as memory-bound, where memory capacity is the limiting factor for further workload consolidation.

Server2: Demonstrated systemic underutilization across all metrics. CPU utilization averaged a minimal 0.5–1.5%, while memory usage remained at a low 32% of its 31 GiB capacity. Swap usage was negligible. This server is classified as largely idle, representing wasted resources.

3.2. Comparative Analysis

These servers illustrates a classic case of inefficient resource management within a cluster. The workloads are poorly distributed: Server1 is oversubscribed on memory while its CPU lies mostly idle, whereas Server2 is underutilized on all fronts. This imbalance suggests an absence of effective cross-server load balancing and a common practice of static, over-provisioned virtual machine allocation without subsequent right-sizing.

Figure 11: Comparative Utilization Metrics provides a snapshot comparison of the key metrics between the two servers, highlighting the stark contrast in their resource profiles.
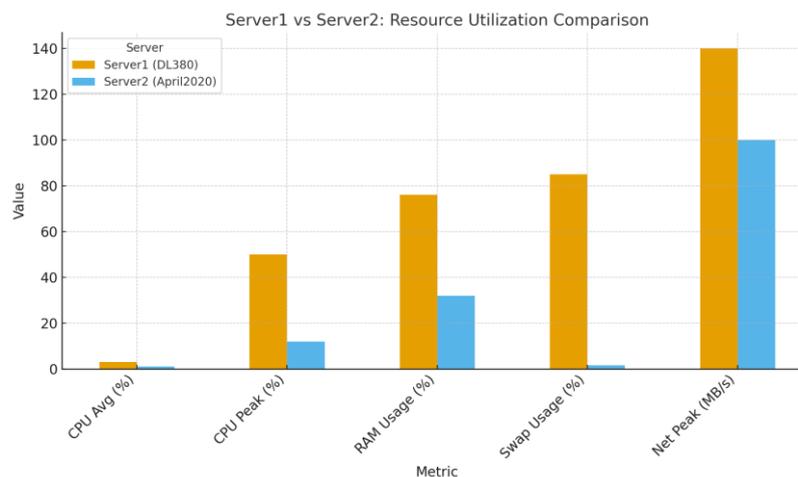

Figure 11: Comperative Utilization

Figure 12: Resource Utilization Trends (30 days) illustrates the time-series data for each metric, confirming the consistency of these patterns over the observation period. Server1's memory pressure and swap usage are persistent, while Server2's resources remain abundantly available. Network throughput is more volatile on Server1, correlating with its higher workload activity.
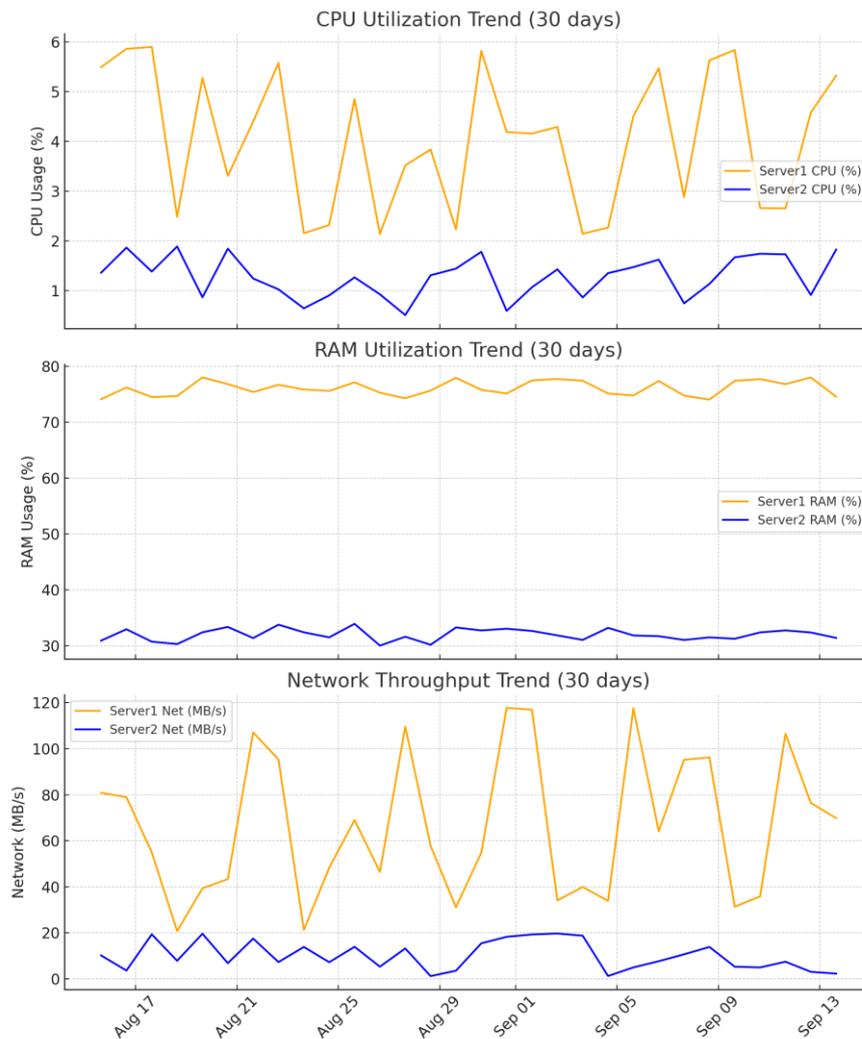

Figure 12: Resource Utilization (30 days)

4. Discussion and Conclusions

The results underscore a critical misalignment between resource provisioning and actual workload requirements, a common issue in enterprise environments that lack a continuous optimization cycle. The high swap usage on Server1 is a primary concern, as it significantly degrades performance and indicates that VMs are contending for insufficient physical memory. Conversely, the underutilization of Server2 represents a direct financial loss.

This imbalance is not a limitation of the Proxmox VE platform itself but rather a consequence of operational practices. It reveals a need for: (1) better initial and ongoing VM right-sizing, (2) implementation of live migration for dynamic load balancing, and (3) adoption of memory optimization techniques native to modern hypervisors. The findings align with existing literature on resource management in cloud environments (Liu et.al. 2021) (Ajankar et.al., 2011), but provide a specific context for Proxmox VE deployments.

This case study demonstrates the potent value of telemetry data inherent in Proxmox VE for diagnosing and rectifying server utilization inefficiencies. The analysis of two production servers revealed a typical scenario of one server being memory-bound and another being profoundly underutilized – a situation that leads to unnecessary hardware costs, inflated energy bills, and potential performance degradation.

By adopting a data-driven approach to resource management, institutions can transform their Proxmox VE deployments from a static infrastructure into a dynamic, efficient, and cost-effective cloud environment. The recommended strategies – workload migration, right-sizing, and energy-aware scheduling – provide a clear pathway to achieving higher performance-per-watt, reducing total cost of ownership (TCO), and promoting more sustainable IT operations. Future work will involve implementing these recommendations and quantifying the resulting improvements in efficiency and cost savings.

5. Recommendations

Based on our analysis, we propose the following actionable strategies to optimize the studied environment and similar Proxmox VE deployments:

- Cross-Server Workload Migration: Migrate memory-intensive VMs from Server1 to Server2 to immediately alleviate memory pressure and balance utilization across the cluster.

- VM Right-Sizing: Conduct an audit of all VMs to reduce overstated vCPU and RAM allocations, especially for consistently underutilized guests. This frees up resources for better consolidation.

- Advanced Memory Management: On Server1, enable memory ballooning and configure Transparent Huge Pages (THP) to improve memory management efficiency. If workloads justify it, a physical RAM upgrade should be considered.

- Energy-Aware Consolidation: Strategically consolidate workloads onto one server during periods of low demand (e.g., nights, weekends), allowing the other server to be powered down or placed in a low-power state to reduce energy consumption Beloglazov et.al. 2012).

- Implement Quality of Service (QoS): Monitor for "noisy neighbor" VMs and use Proxmox's resource limits (CPU units, I/O bandwidth) to ensure predictable performance for all workloads.

Acknowledgment

References

Ajankar, S., Mohta, A., & Sane, S. (2011). *Optimization in virtualization*. https://doi.org/10.1007/978-81-8489-989-4_16

Barroso, L. A., & Hölzle, U. (2009). The case for energy-proportional computing. *IEEE Computer*, 40(12), 33–37.

Beloglazov, A., Abawajy, J., & Buyya, R. (2012). Energy-aware resource allocation heuristics for efficient management of data centers for cloud computing. *Future Generation Computer Systems, 28*(5), 755–768. https://doi.org/10.1016/j.future.2011.04.017

Clark, C., Fraser, K., Hand, S., Hansen, J., Jul, E., Limpach, C., Pratt, I., & Warfield, A. (2005). *Live migration of virtual machines*. In *Proceedings of the 2nd Symposium on Networked Systems Design & Implementation* (NSDI).

Cost optimization in dedicated servers in 2025. (2025). *UMA Technology*. https://umatechnology.org/cost-optimization-in-dedicated-servers-in-2025/

Toxigon Infinite. (2025). *How to optimize server performance in 2024*. https://toxigon.com/how-to-optimize-server-performance-in-2024

Hsu, C. H., Chen, Y., Wang, T., & Wu, C. (2011). Energy-aware task consolidation technique for cloud computing. In *2011 IEEE 3rd International Conference on Cloud Computing Technology and Science* (pp. 115–121). IEEE. https://doi.org/10.1109/CloudCom.2011.25

Peng, J., Chen, J., Kong, S., Liu, D., & Qiu, M. (2016). Resource optimization strategy for CPU intensive applications in cloud computing environment. In *2016 IEEE 3rd International Conference on Cyber Security and Cloud Computing (CSCloud)* (pp. 124–128). IEEE. https://doi.org/10.1109/CSCloud.2016.29

Proxmox Server Solutions GmbH. (2025). *Proxmox VE documentation*. https://pve.proxmox.com/pve-docs/

Singh, H., Bhasin, A., Kaveri, P. R., & Chavan, V. (2020). Cloud resource management: Comparative analysis and research issues. *International Journal of Scientific & Technology Research, 9*(6), 96–113.

Verma, A., Ahuja, P., & Neogi, A. (2008). pMapper: Power and migration cost aware application placement in virtualized systems. In *Proceedings of the 9th ACM/IFIP/USENIX International Conference on Middleware*.