Vantage

Q1 2024

Vantage Cloud Cost Report

How Businesses are Allocating Cloud Spend in Q1 of 2024

Fresh off the quarterly earnings for Microsoft, Google, and Amazon, we are releasing the Q1 2024 Cloud Cost Report, an analysis of cloud usage based on anonymized Vantage customer usage. Vantage is a cloud cost visibility and optimization platform, with a unique view into industry trends, thanks to tens of thousands of connected infrastructure accounts across 14 cloud providers.

This report uses anonymized, real-world data to quantify how cloud spending is shifting and changing across the tech industry. To discuss this report in more detail, join our growing Slack Community of over 1,000 engineering leaders, FinOps professionals, and CFOs. View past reports here.

Top Services by Spend Across Clouds

The big three cloud providers, AWS, Google Cloud, and Azure, despite being competitors, share many similarities in their top services. Compute instances consistently lead the pack, followed by either relational databases or object storage. This is unsurprising given that these services serve as the foundational elements for most workloads.

This quarter, there was a growing emphasis on logging and monitoring services. As applications and infrastructure grow in complexity, monitoring and logging tools play an important part in maintaining visibility, troubleshooting issues, and ensuring reliable operations. However, these tools often come at a premium, underscoring the importance of a strategic approach to monitoring.

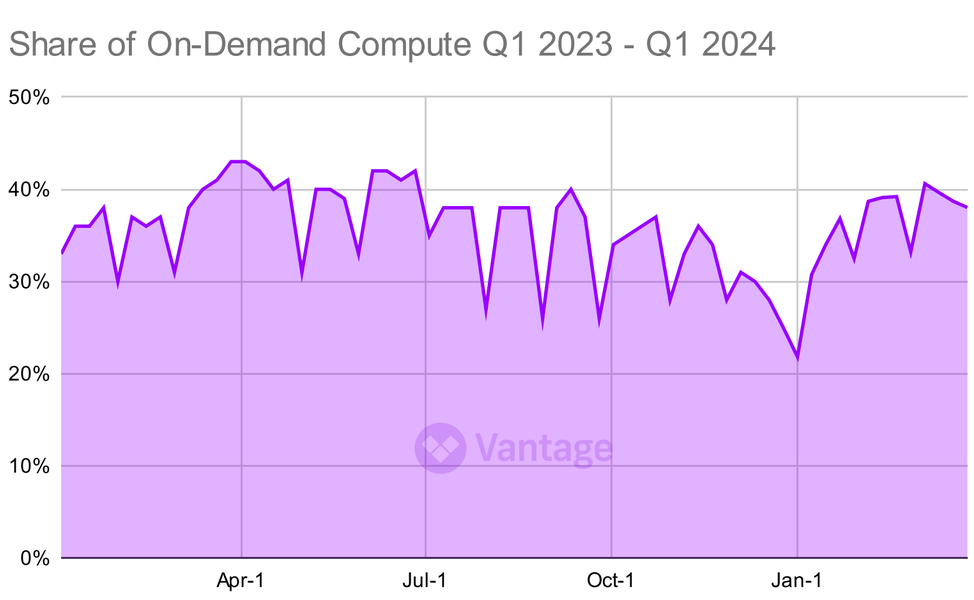
Top 10 Services by Spend on AWS, Google Cloud, and Azure Q1 2024

Ranking	AWS		Google Cloud		Azure	
1	EC2	37%	Compute Engine	36%	Virtual Machines	31%
2	RDS	11%	Cloud Storage	13%	Storage	16%
3	S3	5%	BigQuery	8%	Azure SQL	6%
4	EBS	4%	Dataflow	7%	Log Analytics	5%
5	DynamoDB	4%	Bigtable	6%	Azure App Service	4%
6	OpenSearch	3%	Cloud Logging	4%	SQL Managed Instance	4%
7	CloudFront	3%	Pub/Sub	4%	Microsoft Defender for Cloud	4%
8	CloudWatch	2%	Cloud SQL	3%	Azure PostgreSQL	3%
9	NAT Gateways	2%	Networking	2%	Bandwidth	2%
10	VPC	2%	Spanner	2%	Virtual Network	2%

An Increase in On-Demand Spend

After a steady decline in On-Demand spend proportionate to commitment-based discounts (i.e., Reserved Instances and Spot Instances), this quarter saw a slight rise in EC2 On-Demand. One reason is the increased use of GPU-based instances for ML workloads.

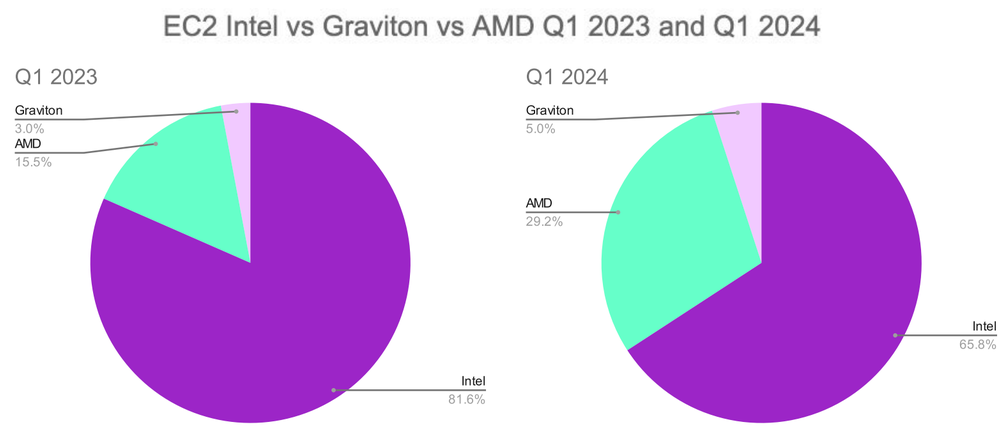
AI workloads are often only required for the short term or are experimental, so companies opt to use On-Demand rather than commit to Reserved or Spot Instances. This leads to less predictable patterns in On-Demand spend and is just one example of the growing impact AI is having on cloud spending.

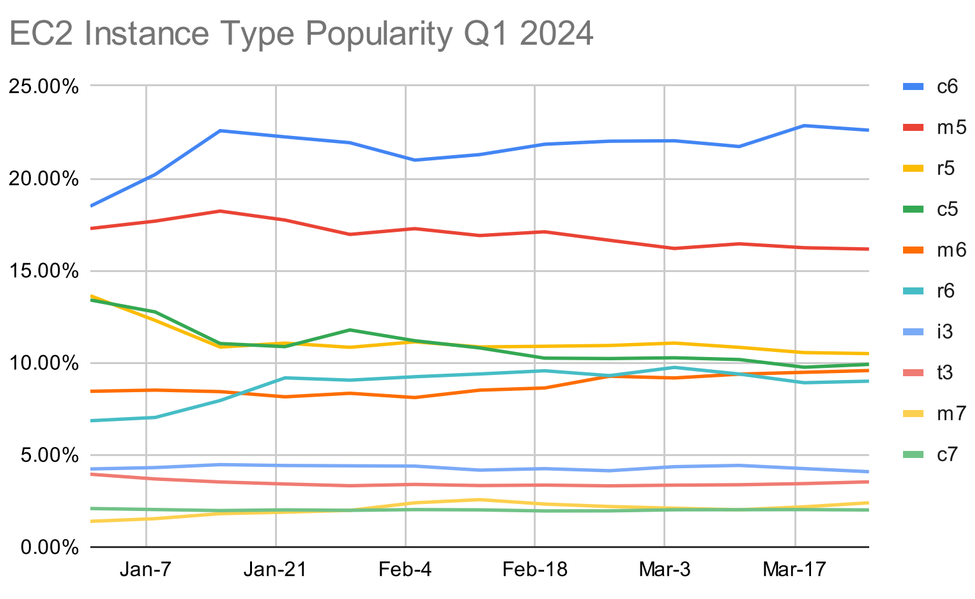


The Rise of Graviton and AMD

There is more ability to choose between processors than ever before since just about every new generation of EC2 instance family has instances with AMD and Graviton processors. As such, the percentage share of AMD and Graviton costs for all EC2 instances compared to Intel has nearly doubled in the past year.

That's not even taking into consideration Intel processors having both the majority of instances and the highest-priced instances. For newer generations, such as the recent generation of the m family, there is a more distributed split. See our take.





Most Popular EC2 Instance Types

By looking through past Cloud Cost Reports, we can observe trends and adoption patterns. The c, m, and r families have continuously occupied a majority of the board and provide insight into the compute instance upgrade cycle. In Q2 of 2023, c6 surpassed c5. Then, in the final week of last quarter, c6 surpassed m5 for the first time.

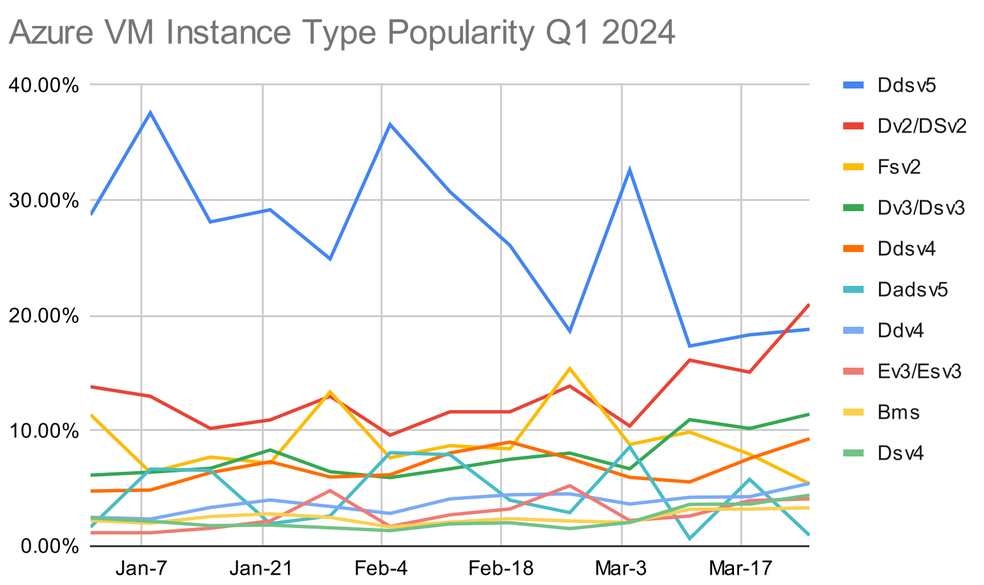
This quarter, m5 continues to trend downward, as m6 trends slightly upward, indicating an upgrade to m6 instances. Similarly, r6 appears to be bridging the gap towards r5.

Finally, c7 and m7 instances have made it to the board, indicating the adoption of the latest generation of compute and general-purpose instances is already underway.

Most Popular Azure Virtual Machines

For the first time, we look at the popularity of Azure Virtual Machines (VMs). There is a higher degree of fluctuation in instance type usage compared to EC2, in part due to a larger portion of Azure users using On-Demand.

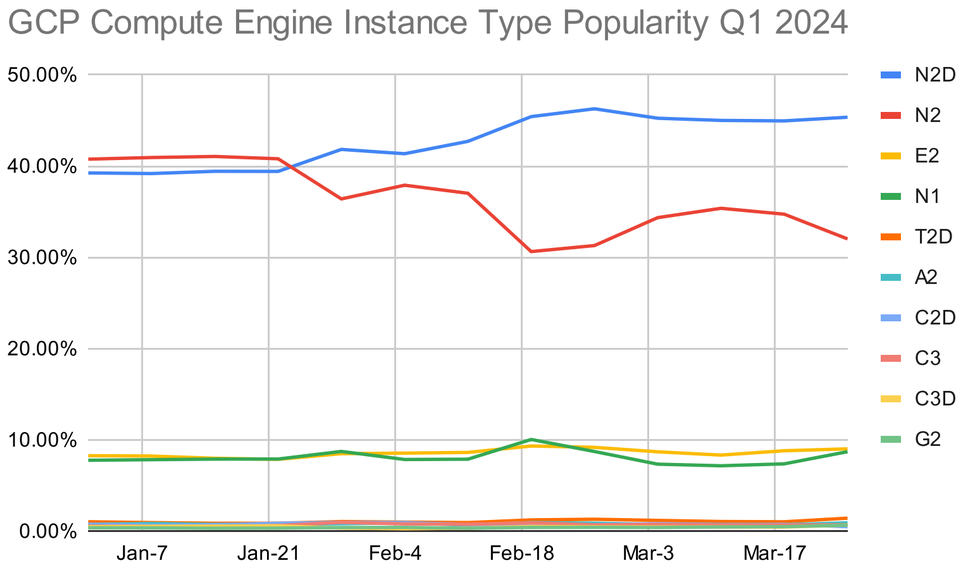
The general-purpose D-series instances occupy the top two spots and most of the board, highlighting its versatility and widespread adoption for a variety of workloads. Toward the end of the quarter, there was a decline in Ddsv5, accompanied by a growth in the adoption of most other instance types. This trend suggests that customers are moving toward a more distributed mix of instance types, tailoring their choices to meet the specific demands of their workloads, as they continue to optimize and upgrade their Azure infrastructure.

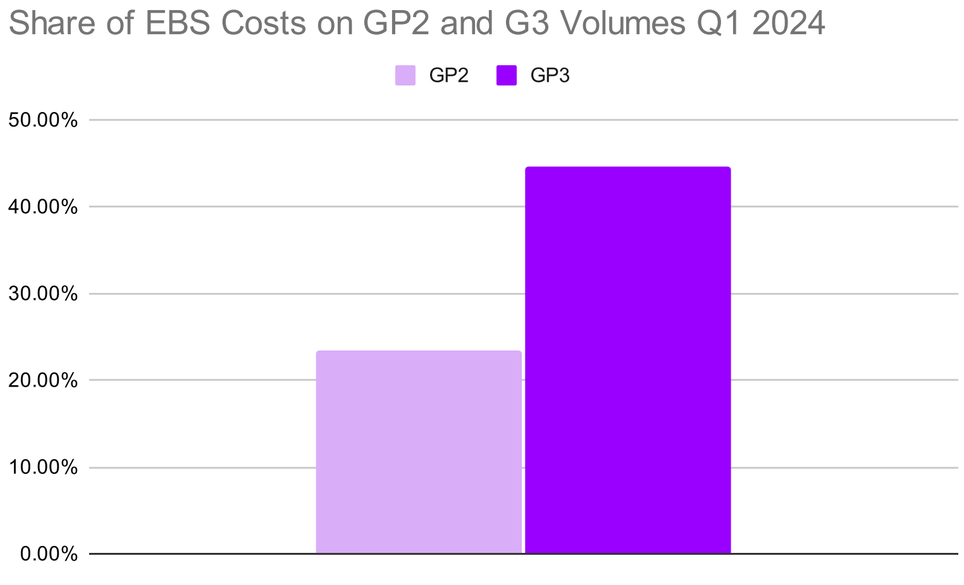


Most Popular GCP Compute Engine VMs

Also for the first time, we look at the popularity of GCP Compute Engine VMs. Compared to EC2, there is a much more dramatic difference between the leading instance types and the rest, with about 80% of the spend belonging to N2D and N2.

N2D and N2 belong to the general-purpose machine family and have the best price-performance for many workloads. One of the main differences between them is that N2 has Intel processors, while N2D has AMD. The Tau T2A, which is not yet on the board but is growing in adoption, was the first instance type to use Arm-based processors. As Arm-based processors are gaining more traction, we expect to see an even distribution in the future.





The Growth of GP3

In the very first Cloud Cost Report (Q3 2022), we saw EBS customers still slow to upgrade to GP3 for General Purpose SSD, even though it is more cost-effective and has higher I/O performance. Finally, GP3 volumes have surpassed GP2.

This further highlights how customers are increasing priority in the optimization of their cloud performance and costs.

Rising GCP AI Adoption

Google has been investing heavily in its AI offerings as more and more companies are using and implementing it. Last quarter saw a plethora of AI announcements from Google, including the rebranding of Bard to Gemini and the release of several Gemini models.

As their offerings continue to expand into a complete unified platform, we can expect to see even more adoption.