Q2 2024

Vantage Cloud Cost Report

How Businesses are Allocating Cloud Spend in Q2 of 2024

Fresh off the quarterly earnings for Microsoft, Google, and Amazon, we are releasing the Q2 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

Quarterly reports show market leader AWS with consistent growth. In contrast, Azure and Google Cloud are expanding rapidly and are competing for more market share. This expansion can be attributed to several factors, including generous credits, easy integrations, and of course, the increasing impact of AI.

Notably, for the first time, Vertex AI—Google Cloud's platform for experimentation and prototyping with Gemini models—has entered the top 10 services by spend. Despite this shift, the core offerings: compute, object storage, and relational databases, continue to remain the primary drivers of costs across providers.

Top 10 Services by Spend on AWS, Google Cloud, and Azure Q2 2024

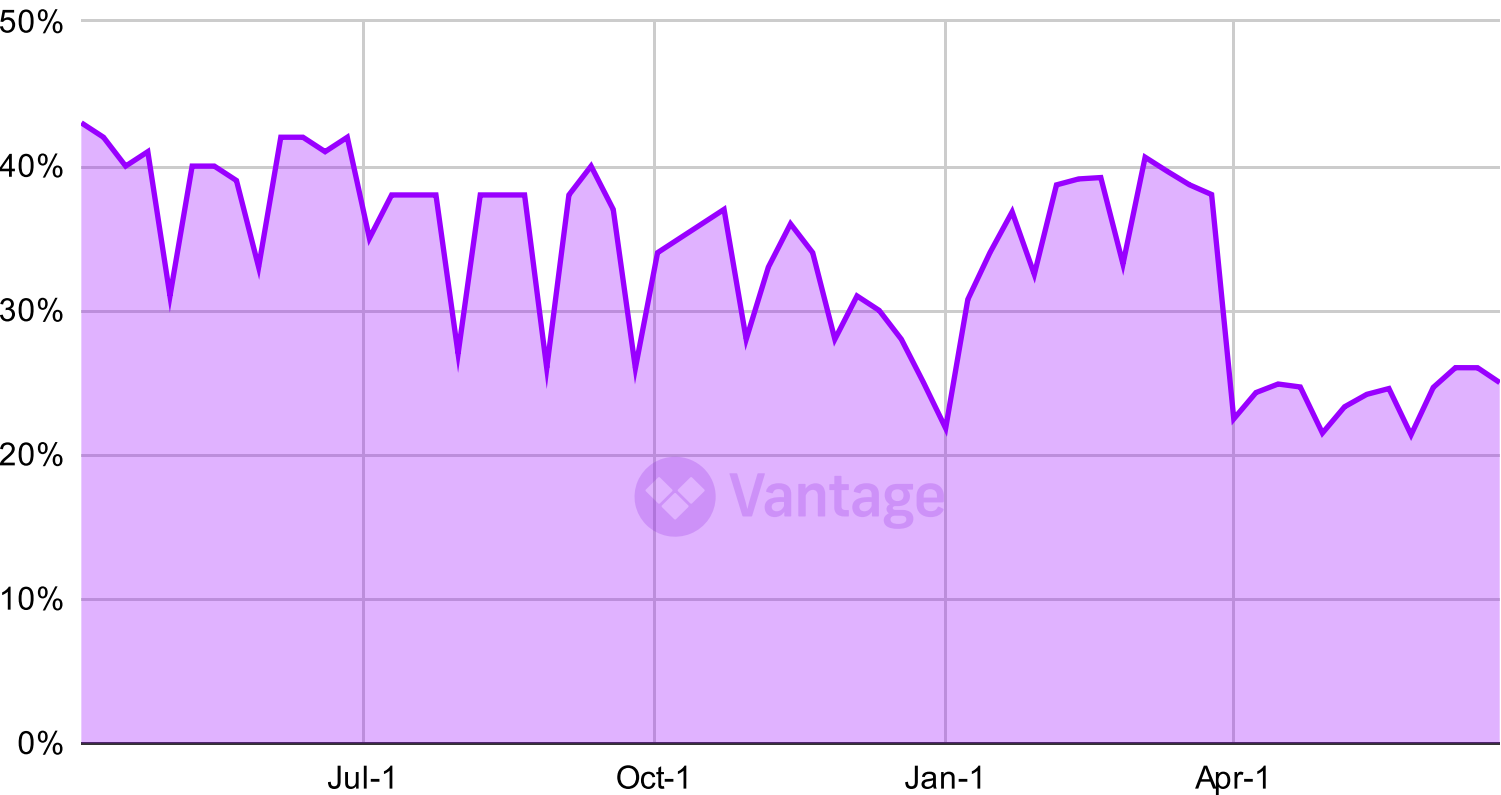| Ranking	AWS		Google Cloud		Azure	
1	EC2	39%	Compute Engine	23%	Virtual Machines	20%
2	RDS	12%	Cloud Storage	13%	Storage	19%
3	S3	6%	Bigtable	9%	Azure SQL	7%
4	CloudFront	4%	Dataflow	9%	SQL Managed Instance	5%
5	EBS	4%	BigQuery	8%	Log Analytics	5%
6	DynamoDB	3%	Cloud Logging	5%	Microsoft Defender for Cloud	4%
7	OpenSearch	3%	Cloud Pub/Sub	5%	Azure App Service	4%
8	CloudWatch	2%	Cloud SQL	3%	Bandwidth	2%
9	NAT Gateways	2%	Cloud Spanner	2%	Application Gateway	2%
10	ElastiCache	2%	Vertex AI	1%	Virtual Network	2%

A Dramatic Decrease in EC2 On-Demand Spend

After an uptick in On-Demand spend last quarter, EC2 On-Demand spend has significantly decreased in favor of Reserved Instances and Savings Plans.
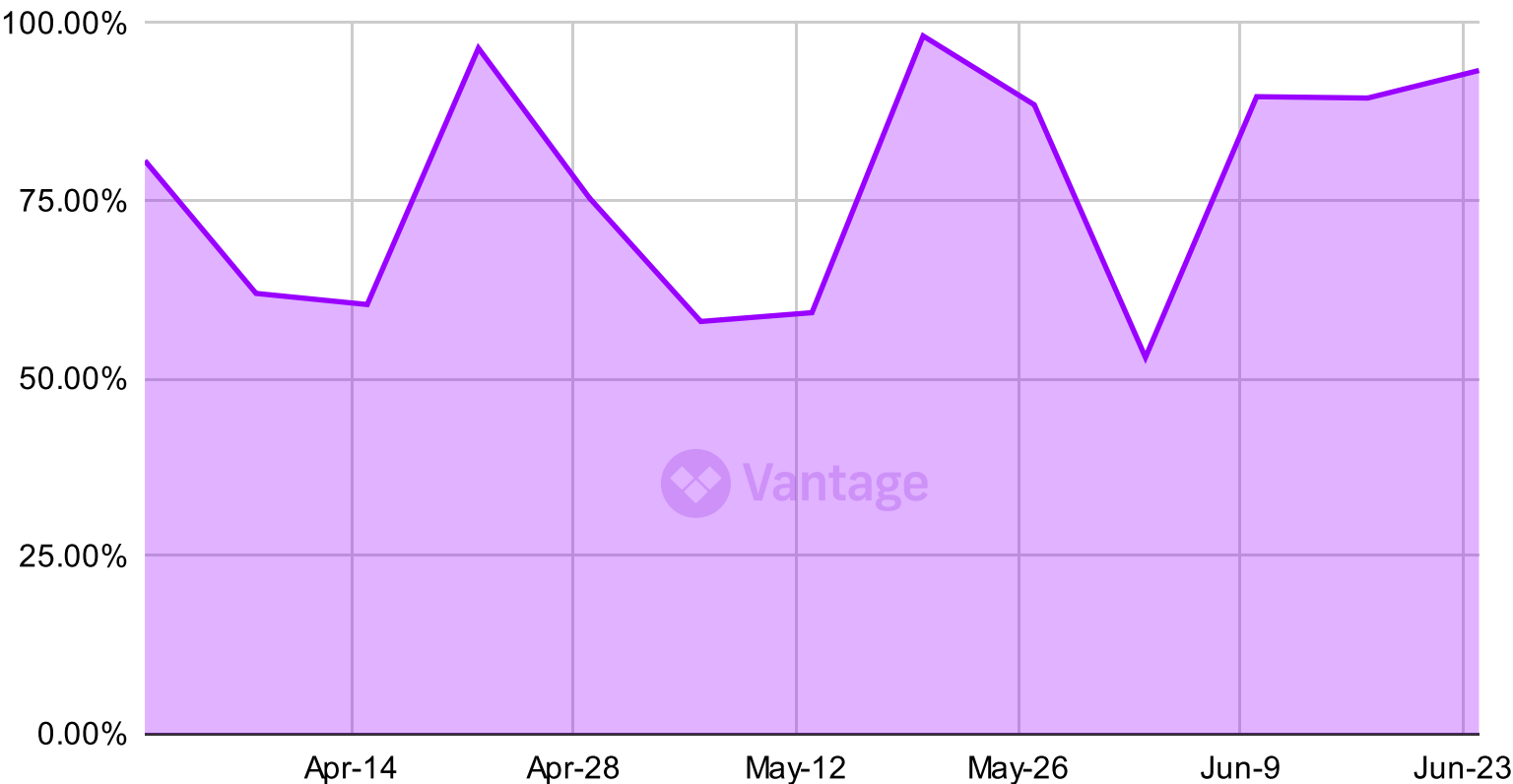
This shift could be due to the growing adoption of Savings Plans, especially after recent changes in the Reserved Instances marketplace caused some users to temporarily switch from Reserved Instances to On-Demand. Organizations are increasingly opting for Savings Plans for flexibility and savings.

Additionally, recent quarters have seen a rise in experimentation with GPU instances. For these cases, organizations were hesitant to commit to one- or three-year plans, opting instead for On-Demand. However, this quarters decline in On-Demand indicates the experimentation phase is stabilizing. Organizations may have a more defined understanding of their GPU needs and are either including them in their long-term planning or are moving away from the experiments altogether.

Share of EC2 On-Demand Compute Q2 2023 - Q2 2024

Azure Virtual Machine Users Could Commit and Save

Azure Virtual Machines show a significantly higher On-Demand spend compared to AWS EC2, indicating a preference for flexibility over commitments to Reserved Instances or savings plan for compute.
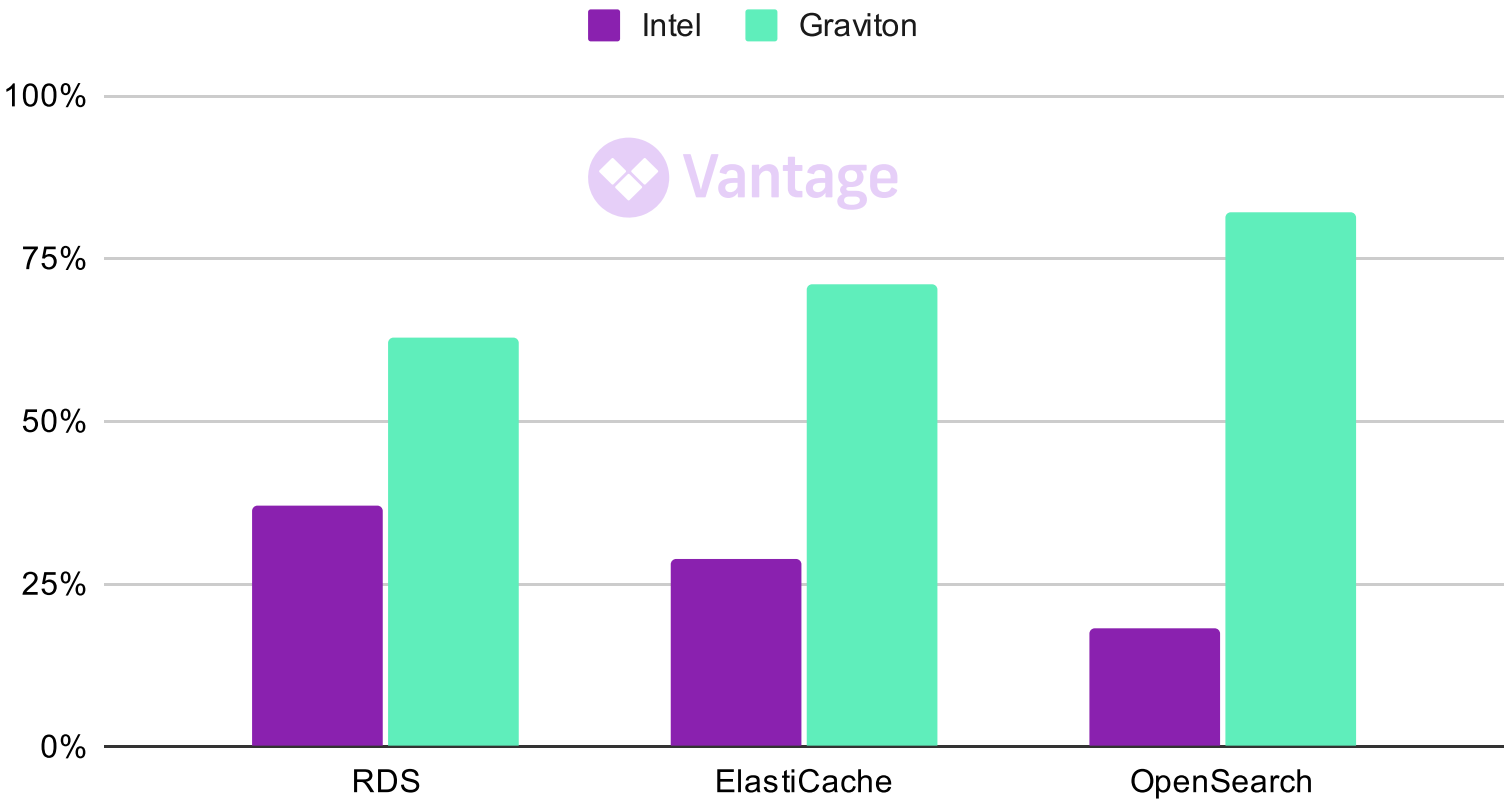
However, for cases where it is feasible to switch to Reserved Instances or savings plan for compute there are significant cost savings (up to 72% for Reserved Instances and 65% for savings plan for compute, according to Azure). Ideal use cases include steady-state workloads or long-term projects.

Share of Azure VM On-Demand Compute Q2 2024

Graviton is Taking Over

Last quarter, we discussed the growing distribution of Graviton and AMD compared to Intel for EC2 instances. In particular, we've been seeing a much more even distribution of the three processor types in newer generation instances.

This quarter, we queried Amazon services with only Intel and Graviton processor options. The data showed Graviton had surpassed Intel in RDS, ElastiCache, and OpenSearch. This highlights AWS's continued investment for ARM-based architecture and companies willingness to adopt it for performance and cost benefits.

AWS Intel vs Graviton Q2 2024

Datadog Log Management Q2 2024

- flex stored logs 1.6%
- flex compute logs 3.2%
- logs indexed 30 day 6.5%
- logs indexed 7 day 7.5%
- logs indexed 3 day 11.0%
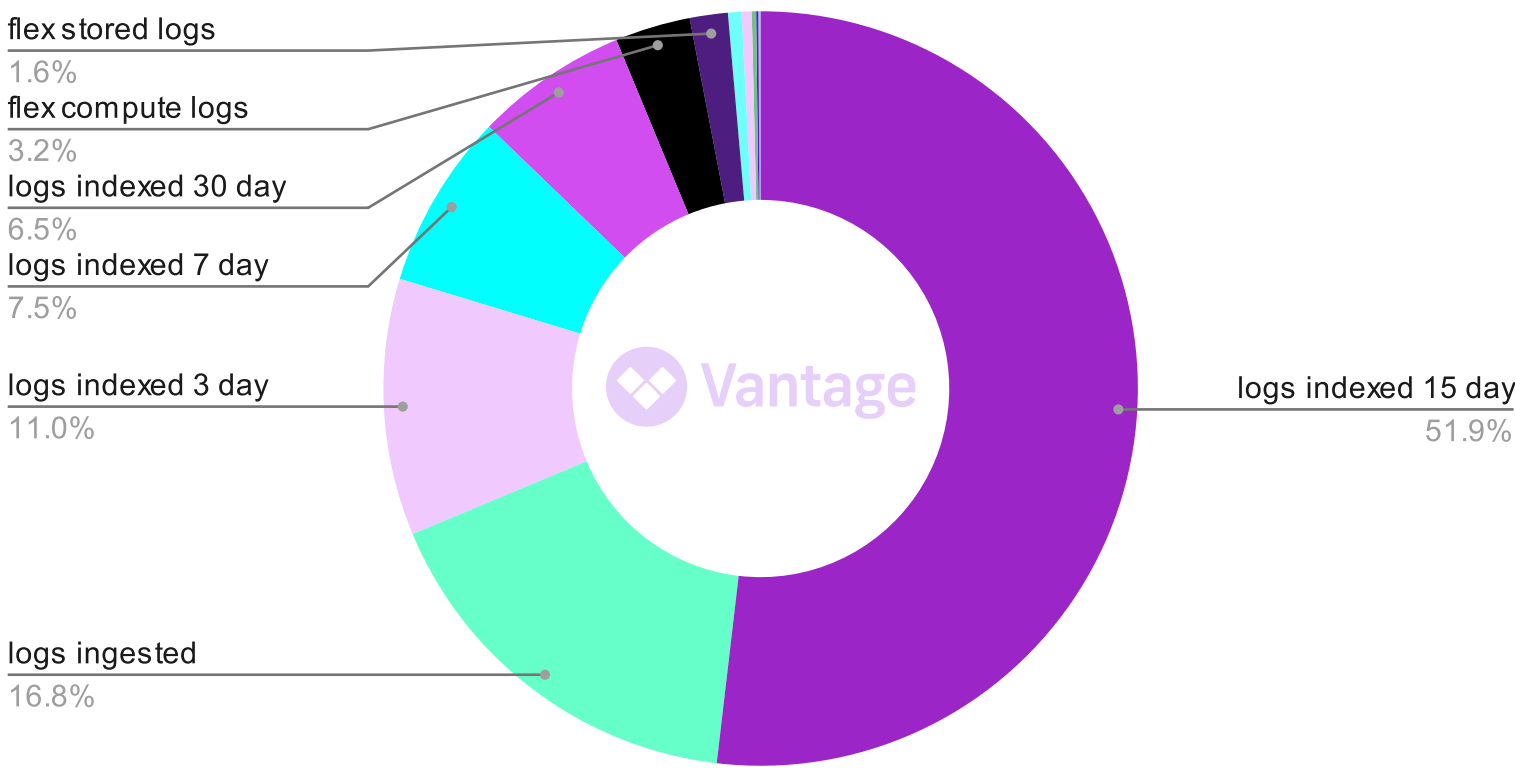- logs ingested 16.8%
- logs indexed 15 day 51.9%

Room for Optimization With Datadog Log Management

Since we started the Cloud Cost Report in 2022, Log Management has remained the top spend category for Datadog. Within that category, log indexing makes up a majority of spend. In particular, logs indexed for 15 days.

15 days is a common retention period as companies often need to search and analyze log data for a couple of weeks. However, if a company is looking to optimize its Datadog costs and finds that it is not utilizing the data for the entire 15-day period, reducing the retention duration is a straightforward way to reduce costs.
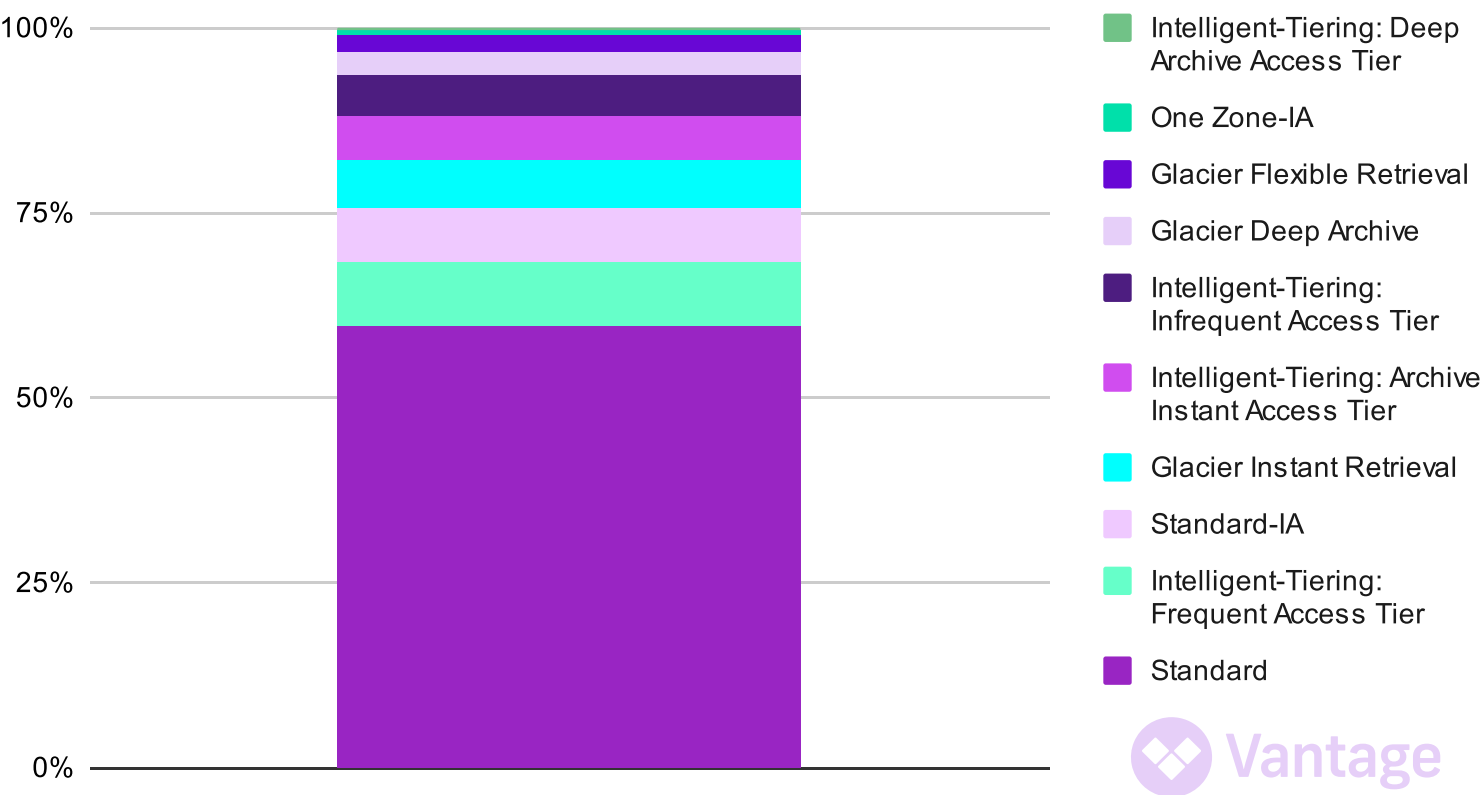
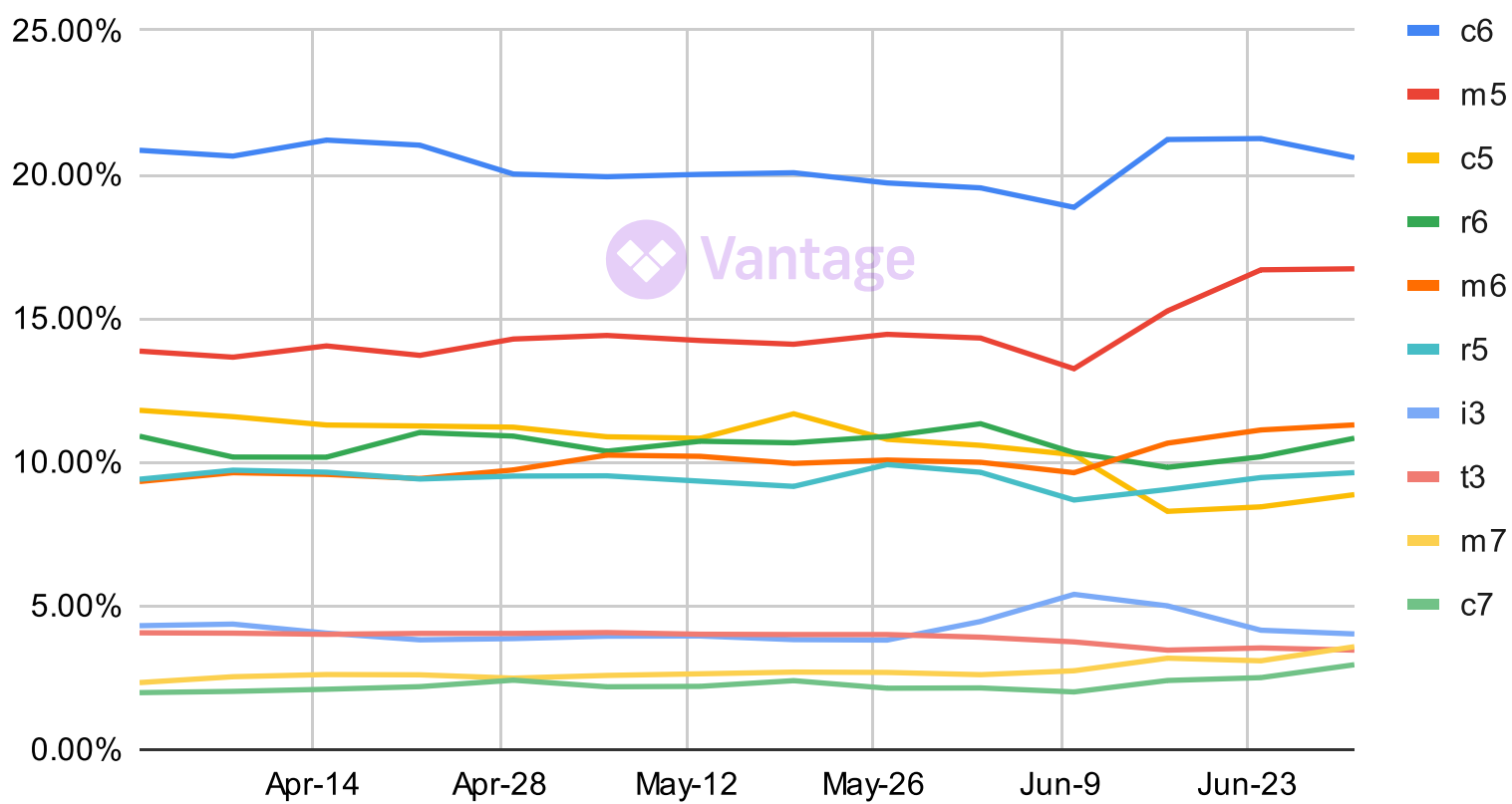# S3 Storage Classes Still a Majority Standard

S3 is the third largest category of AWS spend and almost 70% of that spend is from storage costs. Optimizing storage is a huge opportunity for savings with minimal engineering effort required. Yet, if we look back to the Q4 2022 report, we can see that customers have not made significant changes in their storage class optimization strategies.

AWS offers a variety of storage classes tailored to different use cases. A substantial portion of S3 costs is spent on the Standard storage class, which although is ideal for general-purpose, frequently accessed data, is one of the most expensive classes. Other classes (e.g., Intelligent-Tiering) offer drastic cost-savings for unpredictable, changing, archive, or infrequent access use cases.

S3 Storage Classes Q2 2024

Legend:
- Intelligent-Tiering: Deep Archive Access Tier
- One Zone-IA
- Glacier Flexible Retrieval
- Glacier Deep Archive
- Intelligent-Tiering: Infrequent Access Tier
- Intelligent-Tiering: Archive Instant Access Tier
- Glacier Instant Retrieval
- Standard-IA
- Intelligent-Tiering: Frequent Access Tier
- Standard

EC2 Instance Type Popularity Q2 2024

Legend: c6, m5, c5, r6, m6, r5, i3, t3, m7, c7

AWS EC2 Newer Generation Instance Type Popularity

This quarter reinforced the ongoing trend of customers migrating to newer generation instance types for cost and performance improvements. 6th generation c, r, and m instances have, or almost have, surpassed their 5th generation family in terms of usage. Some 7th generation instances are continuing to gain momentum.

Specifically, in the Compute Optimized c instance family, the c6 instance type continues to lead the pack as c5 usage steadily declines. While c6 maintains a strong lead, c7 is gaining some traction, as well. Similarly, for the memory optimized r instance family, r6 has surpassed r5, showcasing the shift towards more efficient instances. The general purpose m family is also adopting newer generations, with m6 closing the gap to m5 and m7 slowly creeping up the leaderboard.