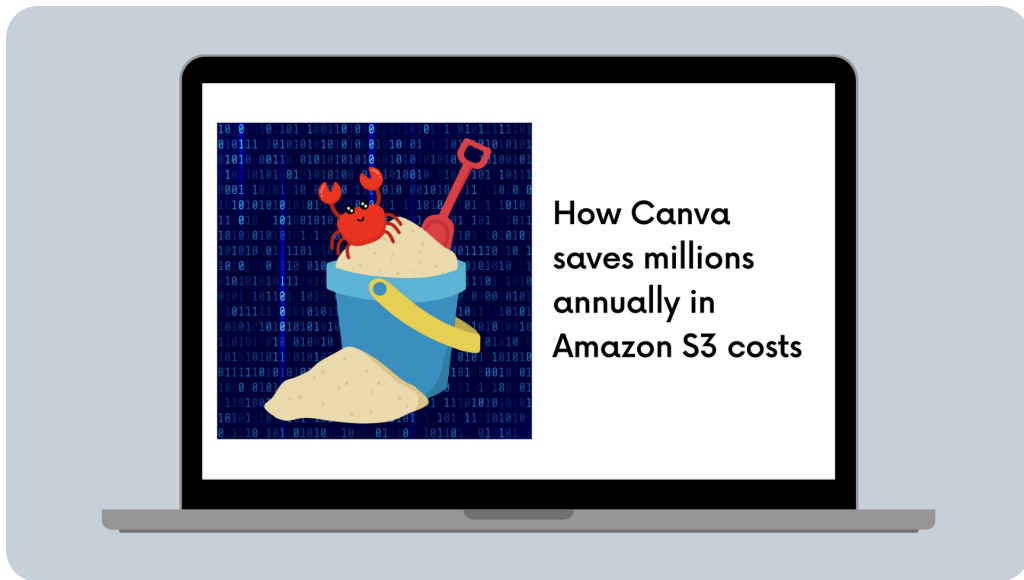


Amazon S3

# How Canva saves millions annually in Amazon S3 costs

Understanding our data and usage patterns was the real key.



---

**Josh Smith**

May 4

---



[Canva](#) is an online design tool that empowers users worldwide to design, edit and publish anything they can dream up. Canva runs most of its production workloads on AWS, leveraging several core services, including [Amazon S3](#), [Amazon ECS](#), [Amazon RDS](#), and [Amazon DynamoDB](#).





## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

Accept all

Only essential

videos from our curated library of over [75 million stock photos and graphics](#). However, this number is relatively small compared to the total amount of content we store. This is because Canva allows every brand to create and publish their own templates, media, and other content, leading to an explosion of user-generated content that we need to store somewhere!



A brand-specific template for a Canva presentation

Amazon S3 provides a reliable, extremely durable, low-cost storage solution. In particular, Amazon S3 provides various storage classes designed for different access patterns and workloads. Understanding the differences between these S3 storage classes is crucial to operating cost-effectively at scale. Across Canva's production environments, we store

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

requirements of your workloads, as outlined in the following diagram.

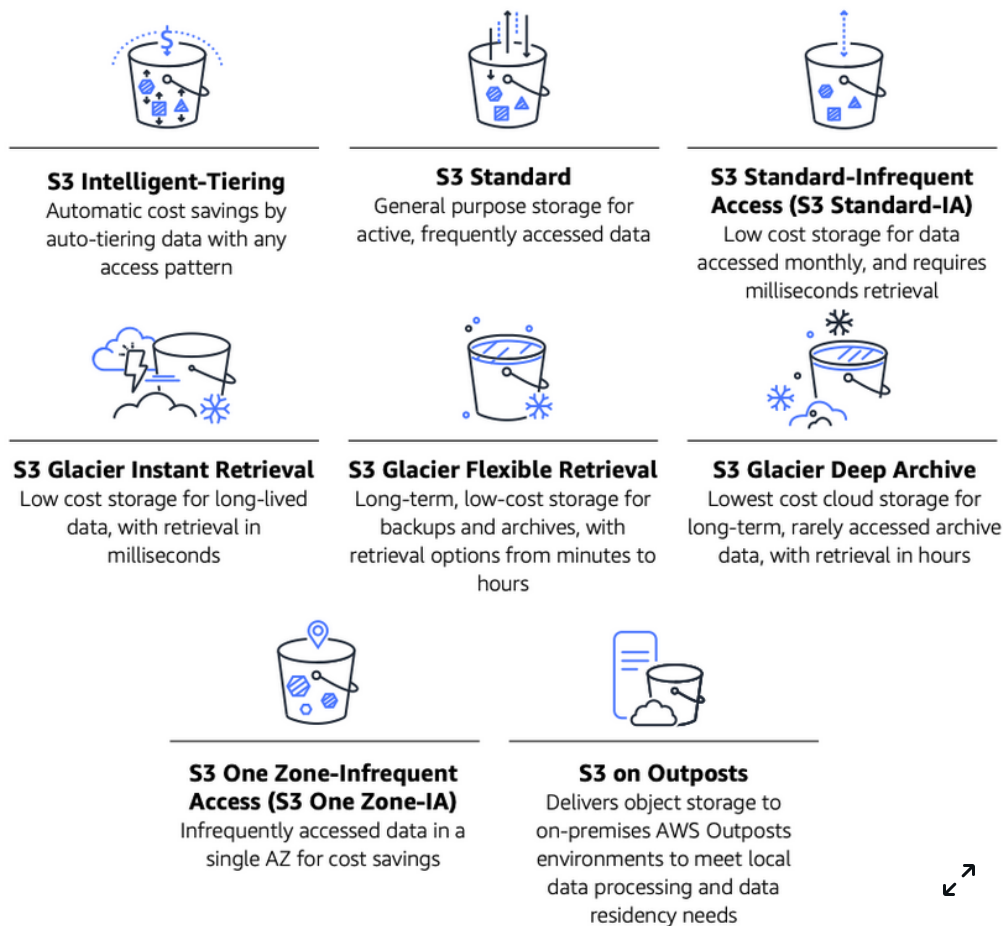


Figure 1: Amazon S3 storage classes

For Canva's common libraries of templates, stock photos, and graphics, the S3 Standard storage class suits us best. Many users frequently access this content daily, so keeping it in a storage class optimized for frequently

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

complete picture of our diverse user base's access patterns, have a read of our recent blog post [“Understanding a Diverse User Base with Frequency Segmentation at Scale”](#).

For some specific use cases, like log archival and backups, we also use S3 Glacier Flexible Retrieval. S3 Glacier Flexible Retrieval is ideal for data that can be retrieved within minutes to hours. This requirement is acceptable for log archives and backups, which may only need to be accessed on rare occasions, but generally isn't suited for frequent access to user-generated content as we have with templates, stock photos, and graphics.

When AWS launched S3 Glacier Instant Retrieval in November 2021, it offered the best of both worlds for infrequently accessed data, low-cost archive storage, and fast retrieval in milliseconds. So naturally, this prompted us to ask ourselves how much we could save by migrating our infrequently accessed data to this S3 Glacier IR storage class. And which buckets should we migrate?

### Understanding our Data

While we knew our user-generated content tended to be accessed soon after its creation, we weren't entirely sure of the breakdown. We'd previously used [S3 lifecycle policies](#) that would transition data from S3 Standard into S3 Standard-IA after 30 days but hadn't analyzed that saving.

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

after the first 15 days, in line with the expected access patterns we mentioned previously.



Figure 2: Amazon S3 Standard storage access patterns over time

This next graph shows the same information, but for data stored in S3 Standard-IA. The bytes stored seem a bit strange, but looking closely at the Days axis, see that the time periods get longer (first 15 days, then 30, then 180, then 365). If you flattened this graph out to have equally spaced time periods, it would look very similar to Canva's user growth over time. Importantly though, the rate of retrieval looks to be reasonably flat across all time periods.

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

The final graph in our analysis shows the total amount of data accessed across the bucket in each storage class. In this particular bucket, roughly 10% of the total data was stored in S3 Standard, while 90% was stored in S3 Standard-IA. The graph shows that 60-70% of all accessed data came from S3 Standard, while 30-40% came from S3 Standard-IA. This aligns very well with S3's pricing model for the storage classes. S3 Standard has a higher storage cost per GB, but very low data access costs, while S3 Standard-IA (and S3 Glacier Instant Retrieval) has a lower storage cost per GB but higher data access costs.



Figure 4: Relative Amazon S3 storage class access patterns over time

This kind of data access was typical for many of the buckets we used to store user-generated content. So

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

reason why it's important to maximize storage cost visibility and have a good understanding of your data footprint.

Interestingly, this cost to transition is based on the number of objects we're moving. At the same time, the potential savings from S3 Glacier Instant Retrieval largely comes from the total amount of data in the storage class.

Furthermore, the cost to transition all the data is a one-time charge, while the savings from the cheaper storage class are ongoing.

This meant we could calculate the approximate time it would take to move from S3 Standard to S3 Glacier Instant Retrieval to break even based on the average object size in the bucket. For example, if you need to transition a small number of very large objects, the payoff happens quickly. Alternatively, transitioning a large number of small objects can take months before breaking even. We did the same analysis for S3 Standard-IA because we already had large amounts of data in the storage class for cost savings.

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

Figure 5: Amazon S3 Standard to Amazon S3 Glacier Instant Retrieval breakeven (months) and Amazon S3 Standard-IA to Amazon S3 Glacier Instant Retrieval breakeven (months)

The graph's x-axis begins on the left at 128 KB because that's the minimum billable size of objects in S3 Standard-IA and S3 Glacier Instant Retrieval (even if your objects are smaller than 128 KB, you're still billed for 128 KB).

Therefore the time to break even when moving from S3 Standard IA to S3 Glacier Instant Retrieval is constant for objects below 128 KB because all objects are billed as being 128 KB. You can see this in the following graph.



Figure 6: Amazon S3 Standard to Amazon S3 Glacier Instant Retrieval breakeven (months) and Amazon S3 Standard-IA to Amazon S3 Glacier Instant Retrieval breakeven (months) vs. average object size (KB)



## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

buckets with larger average object sizes. When we performed this analysis, we decided to target buckets (with the appropriate usage patterns from the previous section) with an average object size of 400 KB or more first because these buckets would show a positive return on investment within 6 months or less.

### Conclusion

There's very little to write about regarding the migration effort, which was an engineer's dream! We applied a lifecycle policy to each of the buckets, and quickly migrated nearly 80 billion objects in approximately two days. At the time of writing this blog, roughly 130 petabytes of Canva's total 230 petabytes of data in S3 resides in S3 Glacier Instant Retrieval. We have found it extremely valuable as a tool for reducing costs on our infrequently accessed user data. It offers the best of both worlds from S3 Standard-IA and S3 Glacier Flexible Retrieval.

Canva saves roughly \$300,000 per month (\$3.6 million annually) thanks to these changes, and given the ever-growing amount of user-generated data we store, these savings continue to grow over time. However, it's very important to remember that these savings required us to first understand the access patterns for our data, as well as a one-off spend of over \$1.6 million to transition

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

## More from Canva Engineering

### UI TESTING

#### Speeding up UI tests with CodeceptJS and Playwright

Lessons learned from migrating 1000+ browser-functional tests from Chimp, WebdriverIO, and Saucelabs to CodeceptJS and Playwright.

BS

**Benjamin Sejas,  
Srini Ramasamy**  
Mar 3, 2022

### BACKEND ENGINEERING

#### From Zero to 50 Million Uploads per Day: Scaling Media at Canva

The evolution of media persistence during hypergrowth at Canva

**Jacky Chen,  
Robert Sharp**  
Nov 29, 2022

### DEVOPS

#### Functional completeness and local hermeticity

Maintaining a functionally complete yet airtight local development environment is something we consider to be of critical importance.

**David Hearnden**  
Mar 25, 2015

---

## We care about your privacy

By clicking "Accept All", you agree to the storing of cookies on your device to give you the most optimal experience using our website. We may use cookies to enhance performance, analyze site usage, and personalize your experience.

---

© 2023 All Rights Reserved. Canva®

[Privacy policy](#) [Terms](#)

