Adobe Experience Cloud

Adobe

WHITEPAPER

Adobe Experience Manager as a Cloud Service

Performance and scalability

Table of Contents

Adobe Experience Manager as a Cloud Service

Performance and scalability.

Adapt to ever-changing market expectations and stay ahead of the competition with Adobe Experience Manager offered as a cloud-native service that is always current, scalable, and globally available.

Brands are under tremendous pressure to adapt to ever-evolving consumer expectations and to deliver uninterrupted experiences across devices and at scale. Auto-scaling in Cloud Service can accommodate variable traffic patterns, predictable or unpredictable, for both the end user who is engaging with the content and marketers who are creating the experience. Cloud Service constantly monitors the need for capacity based on traffic and automatically scales using horizontal and vertical auto-scaling based on specific requirements for the memory or processing capacity needed to maintain high performance. Cloud Service also includes an out-of-the-box, Adobe-managed content delivery network (CDN) to help deliver the best digital experience to your customers, regardless of their geographical location. This paper describes how Experience Manager as a Cloud Service maintains high performance at scale.

Author and publish performance.

In a traditional dedicated hosted environment, the resources available to the system are static and limited, requiring infrastructure planning for compute capacity and memory to ensure sufficient capacity. However, predicting traffic is challenging and can put the consumer experience at risk, which leads many brands to overestimate capacity, hence increasing total costs.

Experience Manager as a Cloud Service is built with a fully cloud-native architecture that uses lightweight container-based virtualization for the author and publish services that can adapt to variable workloads much faster. Thus, multiple author and publish nodes can run concurrently, ensuring high performance not just for end consumers but also for authors creating those experiences.

The author tier in Cloud Service is set up as a cluster by default, and it can scale horizontally or vertically based on CPU utilization. Once the system detects a high utilization, a scaling event is triggered automatically to adjust the number of nodes to bring the utilization to its target level. In addition, author nodes are designed for high availability with complete separation between the code and content. As illustrated in Figure 1, these nodes are connected to the Content Repository Service that stores the structured content in a distinct store, while all binaries are separately stored in a blob store. This allows for product updates to happen without any downtime or interruption to content creators. While new nodes are getting updated with latest features, existing nodes on previous version keep running in the background. As soon as the new nodes are ready, they get connected to the content repository, and start receiving serving requests. At this point, old nodes are retired. This redundancy helps ensure that the system is always up to date with no interruption to service.

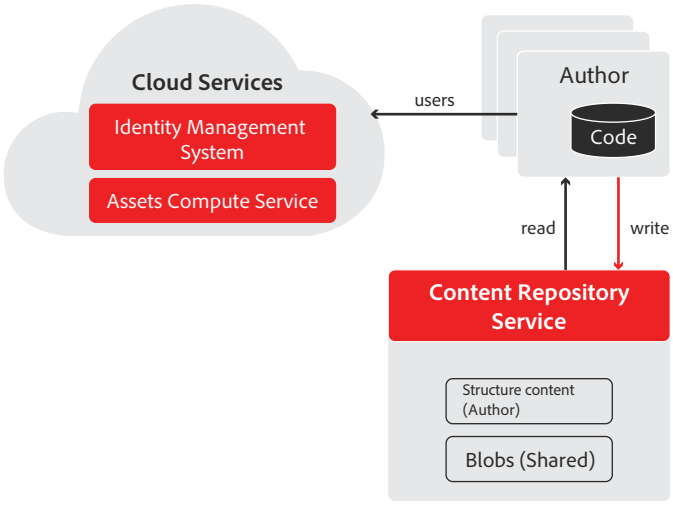

Figure 1. Author cluster

Given the critical nature of publish nodes, the publish tier uses a farm approach in which each publish node is equipped with its own disk for content storage to make it self-sufficient (Figure 2). They are designed to scale not only based on the CPU utilization target but also on the incoming request rate. As soon as an increase in capacity is triggered, new nodes are added to distribute the workload. To ensure no lag in scaling, these nodes are fired up with the most recent content state on the master publish, allowing the system to scale very quickly. In addition, a web server in front of every publisher node takes care of additional caching, rewriting, and blocking of requests.

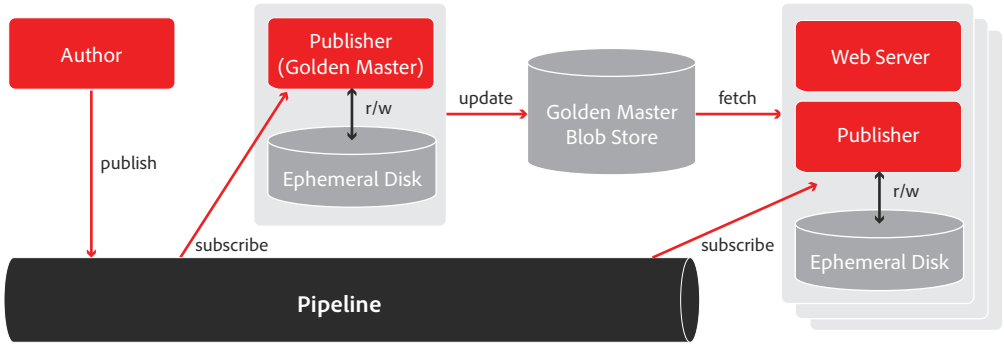

Figure 2. Publish farm

The auto-scaling capabilities of Cloud Service are not limited to serving traffic. Cloud Service also constantly monitors storage usage and automatically triggers a scaling event up or down to accommodate the required storage capabilities. Like the author and publish tiers, the storage tier is redundant by design. This means that adjustments can be made to the storage tier without impacting the functional or nonfunctional behavior of storage operations. This helps ensure that Cloud Service always delivers content storage capacities at cloud-level scale.

Service-based architecture.

Performance for brands involves the ability not only to deliver faster experiences at scale, but also to produce experiences at a high velocity. Creatives and marketers need to be able to move large volumes of content through its lifecycle efficiently. The modern, service-based architecture of Cloud Service enables brands to achieve this with the Asset Compute Service. The following sections detail how the Asset Compute Service optimizes author efficiency for the entire asset lifecycle, starting from uploading and ingesting 100% pixel-quality assets to processing and finally downloading them.

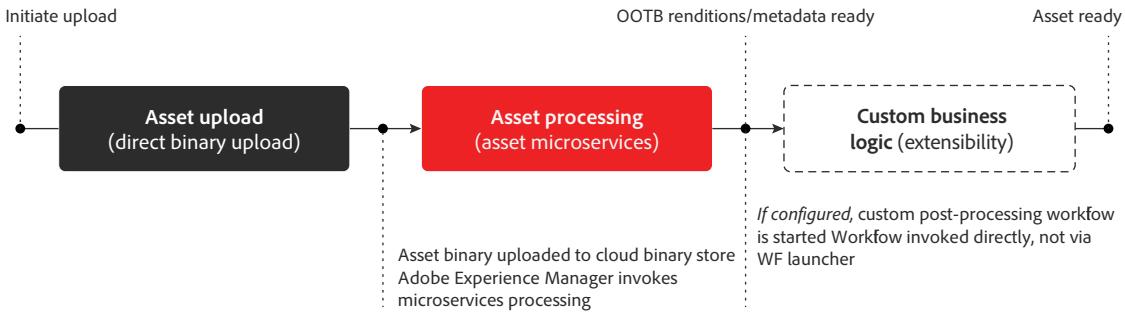

Figure 3. Scalable asset processing and ingestion

Direct binary upload for superior ingestion performance.

Cloud Service leverages direct binary access to significantly improve content ingestion performance. Traditionally, this operation has required significant processing power, depending on the volume and size of assets, and has been limited by server-side bandwidth. With Cloud Service, this limitation is overcome by enabling users to upload binaries directly and in partitions to the cloud storage container. As illustrated in Figure 4, when an authorized user initiates an asset upload request, Experience Manager is notified to trigger optimization of partitioning, followed by allocation of a private location in the binary cloud storage container. This partitioning of high-quality and large-size content into smaller and manageable chunks, instead of handling one big file, is key to significantly speeding up the ingestion process. Once all the parts of the asset have been uploaded to the storage container, Experience Manager is notified of its completion, allowing it to register the new asset in its repository.

This improvement in performance is not limited to local authors. Cloud Service also includes an out-of-the-box CDN that is leveraged to manage asset transfer from the point of presence (PoP) closest to the user. Given the private nature of the CDN connection, it is dedicated entirely to the upload process and not limited by the bandwidth of a public internet connection. This helps ensure high upload performance, even for cross-region asset transfers.

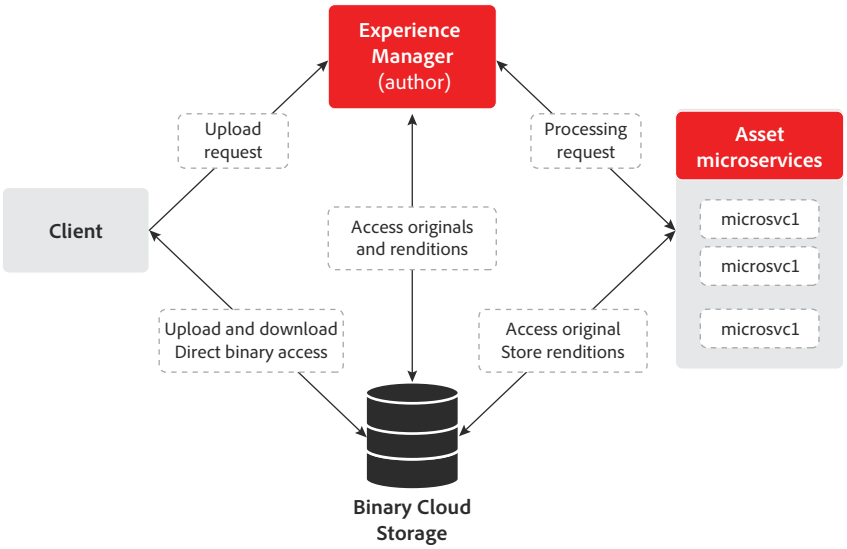

Figure 4. Direct binary upload for superior ingestion performance

Microservice for faster asset processing.

Once assets are ingested, they need to be rendered and processed before they can be used widely. Rendering and processing generally entail a series of back-end operations such as extraction of rich XMP metadata, usability of text to power search, creation of thumbnails, Sensei-powered smart tagging, Photoshop imaging engine and web renditions to provide previews of the assets, custom image renditions, and limited video transcoding. Depending on the volume and quality of assets, this operation can slow down an author with highly demanding processing requests. Traditionally, it has required IT teams to plan for extra compute capacity depending on predictable seasonal demands, such as preparation by marketing teams for product launches or holiday season sales. This not only results in overestimated capacity but also leaves assets vulnerable in unpredictable times. Asset processing in Cloud Service uses a microservice architecture to carry out this demanding operation outside the Experience Manager environment, saving resources and capacity planning.

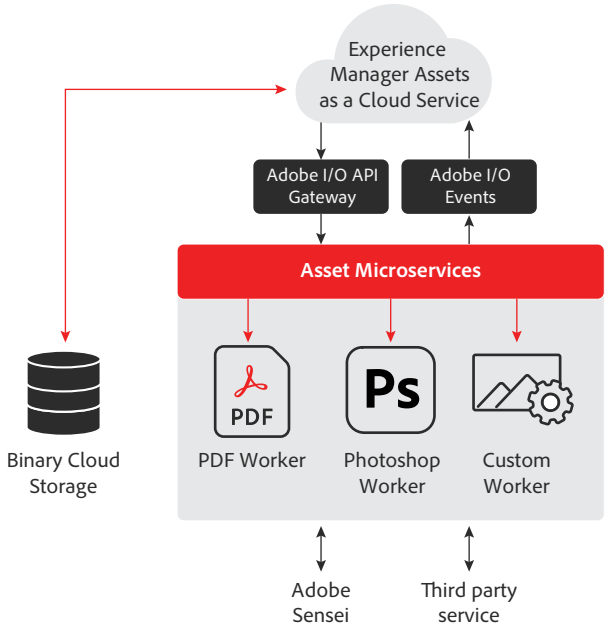

Figure 5. Asset microservice for faster rendering and processing

The Asset Compute Service uses the serverless service provided by the Adobe I/O Runtime. As illustrated in Figure 5, it is architected as a collection of workers, which are stateless, independent of the Experience Manager environment, and much faster to run compared with traditional virtualization methods. As soon as processing of an asset is requested, it triggers the dispatch of one or more microservices, depending on the processing profile configuration. Since the compute capacity is independent of Experience Manager and is dedicated to the specific operation of accessing binary directly from cloud storage and processing the asset, it is a much faster process. In addition, capacity is increased or reduced automatically based on the volume of processing requests being received. Given the stateless nature of the workers, there is no requirement to load the previous state while firing up additional capacity, resulting in quick instantiation of the workers. In addition, the Asset Compute Service also provides the flexibility to develop and deploy custom workers that can be used to read and manipulate binary data of assets stored in Experience Manager, most commonly, to create asset renditions.

Summary

Experience Manager as a Cloud Service makes it easy to scale as your business grows. The modern, modular, and elastic architecture of Cloud Service improves performance for the end consumer as well as marketing teams with no worry about infrastructure. In addition, all the systems are pro-actively monitored for all critical metrics to ensure best service health of the systems so that your teams can focus on growing without worrying about scale and performance.

For more information.

www.adobe.com/go/aem

Adobe