

Media sharing service architecture in AWS

Media sharing is one of the hottest markets on the Internet. Customers have a staggering appetite for placing photos and videos on social networking sites, and for sharing their media in custom online photo albums.

The growing popularity of media sharing means scaling problems for site owners, who face ever-increasing storage and bandwidth requirements and increased go-to-market pressure to deliver faster than the competition.

Since most businesses today have limited manpower, budget, and data center space, AWS offers a unique set of opportunities to compete and scale without having to invest in hardware, staff, or additional data center space. Utilizing AWS is not an all or nothing proposition. Depending on the project, different services can be used independently.

This diagram shows an example of a highly available, durable, and cost-effective media sharing and processing platform.

Gliffy Macro Error

You do not have permission to view this diagram.

#	Description
1	Sharing content first involves uploading media files to the online service. In this configuration, an Elastic Load Balancer distributes incoming network traffic to upload servers, a dynamic fleet of Amazon Elastic Compute Cloud (Amazon EC2) instances. Amazon CloudWatch monitors these servers and an Auto Scaling group manages them, automatically scaling EC2 capacity up or down based on load. In this example, a separate endpoint to receive media uploads was created in order to off-load this task from the website's servers.
2	Original uploaded files are stored in Amazon Simple Storage Service (Amazon S3), a highly available and durable storage service.
3	To submit a new file to be processed, upload web servers push a message into an Amazon Simple Queue Service (Amazon SQS) queue. This queue acts as a communication pipeline between the file reception and file processing components.
4	The processing pipeline is a dedicated group of Amazon EC2 instances used to execute any kind of post-processing task on the uploaded media files (video transcoding, image resizing, etc.). To automatically adjust the needed capacity, Auto Scaling manages this group. You can use Spot Instances to dynamically extend the capacity of the group and to significantly reduce the cost of file processing.
5	Once processing is completed, Amazon S3 stores the output files. Original files can be stored with high durability. Processed files could use reduced redundancy.
6	Media-related data can be put in a relational database like Amazon Relational Database Service (Amazon RDS) or in a key-value store like Amazon SimpleDB.
7	A third fleet of EC2 instances is dedicated to host the website front-end of the media sharing service.
8	Media files are distributed from Amazon S3 to the end user via Amazon CloudFront, a content delivery network. Amazon CloudFront offers low-latency delivery through a worldwide network of edge locations.