What is a CDN (Content Delivery Network)?

What is a CDN?

A content delivery network (CDN) is a network of interconnected servers that speeds up webpage loading for data-heavy applications. CDN can stand for content delivery network or content distribution network. When a user visits a website, data from that website's server has to travel across the internet to reach the user's computer. If the user is located far from that server, it will take a long time to load a large file, such as a video or website image. Instead, the website content is stored on CDN servers geographically closer to the users and reaches their computers much faster.

Why is a CDN important?

The primary purpose of a content delivery network (CDN) is to reduce latency, or reduce the delay in communication created by a network's design. Because of the global and complex nature of the internet, communication traffic between websites (servers) and their users (clients) has to move over large physical distances. The communication is also two-way, with requests going from the client to the server and responses coming back.

A CDN improves efficiency by introducing intermediary servers between the client and the website server. These CDN servers manage some of the client-server communications. They decrease web traffic to the web server, reduce bandwidth consumption, and improve the user experience of your applications.

What are the benefits of CDNs?

Content delivery networks (CDNs) provide many benefits that improve website performance and support core network infrastructure. For example, a CDN can do the following tasks:

Reduce page load time

Website traffic can decrease if your page load times are too slow. A CDN can reduce bounce rates and increase the time users spend on your site.

Reduce bandwidth costs

Bandwidth costs are a significant expense because every incoming website request consumes network bandwidth. Through caching and other optimizations, CDNs can reduce the amount of data an origin server must provide, reducing the costs of hosting for website owners.

Increase content availability

Sandpiper EX2016
Google v. Sandpiper
IPR2025-00826

Too many visitors at one time or network hardware failures can cause a website to crash. CDN services can handle more web traffic and reduce the load on web servers. Also, if one or more CDN servers go offline, other operational servers can replace them to ensure uninterrupted service.

Improve website security

Distributed denial-of-service (DDoS) attacks attempt to take down applications by sending large amounts of fake traffic to the website. CDNs can handle such traffic spikes by distributing the load between several intermediary servers, reducing the impact on the origin server.

What is the history of CDN technology?

Content delivery network (CDN) technology emerged in the late 1990s with the focus on faster content delivery over the internet:

First generation

First-generation CDN services focused on networking principles of intelligent network traffic management and data centers for replication.

Second generation

Second-generation CDNs arose in response to the rise of audio and video streaming services, especially video on demand and news on demand. The technology also evolved to solve new challenges in content delivery on mobile devices. Companies used cloud computing techniques and peer-to-peer networks to accelerate content delivery.

Third generation

Third-generation CDNs are still evolving. AWS is driving innovation as one of the leading CDN service providers in the world. With most web services centralized in the cloud, the focus is now on edge computing—managing bandwidth consumption using smart devices that communicate intelligently. Autonomous and self-managed edge networks might be the next step in CDN technology.

What internet content can a CDN deliver?

A content delivery network (CDN) can deliver two types of content: static content and dynamic content.

Static content

Static content is website data that does not change from user to user. Website header images, logos, and font styles remain the same across all users, and the business does not change them frequently. Static data does not need to be modified, processed, or generated and is ideal for storage on a CDN.

Dynamic content

Dynamic content such as social media news feeds, weather reports, login status, and chat messages vary among website users. This data changes based on the user's location, login time, or user preferences, and the website must generate the data for every user and every user interaction.

How does a CDN work?

Content delivery networks (CDNs) work by establishing a point of presence (POP) or a group of CDN edge servers at multiple geographical locations. This geographically distributed network works on the principles of caching, dynamic acceleration, and edge logic computations.

Caching

Caching is the process of storing multiple copies of the same data for faster data access. In computing, the principle of caching applies to all types of memory and storage management. In CDN technology, the term refers to the process of storing static website content on multiple servers in the network. Caching in CDN works as follows:

1. A geographically remote website visitor makes the first request for static web content from your site.

2. The request reaches your web application server or origin server. The origin server sends the response to the remote visitor. At the same time, it also sends a copy of the response to the CDN POP geographically closest to that visitor.
3. The CDN POP server stores the copy as a cached file.
4. The next time this visitor, or any other visitor in that location, makes the same request, the caching server, not the origin server, sends the response.

Dynamic acceleration

Dynamic acceleration is the reduction in server response time for dynamic web content requests because of an intermediary CDN server between the web applications and the client. Caching doesn't work well with dynamic web content because the content can change with every user request. CDN servers have to reconnect with the origin server for every dynamic request, but they accelerate the process by optimizing the connection between themselves and the origin servers.

If the client sends a dynamic request directly to the web server over the internet, the request might get lost or delayed due to network latency. Time might also be spent opening and closing the connection for security verification. On the other hand, if the nearby CDN server forwards the request to the origin server, they would already have an ongoing, trusted connection established. For example, the following features could further optimize the connection between them:

- Intelligent routing algorithms
- Geographic proximity to the origin
- The ability to process the client request, which reduces its size

Edge logic computations

You can program the CDN edge server to perform logical computations that simplify communication between the client and server. For example, this server can do the following:

- Inspect user requests and modify caching behavior.
- Validate and handle incorrect user requests.
- Modify or optimize content before responding.

Distribution of application logic between the web servers and the network edge helps developers offload origin servers' compute requirements and improve website performance.

What is a CDN used for?

A content delivery network (CDN) improves normal website functions and increases customer satisfaction. The following are some example use cases.

High-speed content delivery

By combining static and dynamic internet content delivery, you can use CDNs to provide your customers with a global, high-performing, whole-site experience. For example, Reuters is the world's largest news wholesaler to top channels such as BBC, CNN, the New York Times, and the Washington Post. The news media challenge for Reuters is to deliver news content promptly to customers around the globe. Reuters uses Amazon's CDN service, Amazon CloudFront, with Amazon Simple Storage Service (Amazon S3) to minimize dependence on satellite link communication and create a cheaper, highly available, and secure globally distributed network platform.

Real-time streaming

CDNs help reliably and cost-effectively deliver rich and high-quality media files. Companies streaming video and audio use CDNs to overcome three challenges: reduce bandwidth costs, increase scale, and decrease delivery time. For example, Hulu is an online video streaming platform owned by the Walt Disney Company. It uses Amazon CloudFront to consistently stream more than 20 GBps of data to its growing customer base.

Multi-user scaling

CDNs help support a large number of concurrent users. Website resources can manage only a limited number of client connections at a time. CDNs can rapidly scale this number by taking some of the load from the application server. For instance, King is a gaming company that builds socially connected, cross-platform games that can be played anytime, anywhere, and from any device. King has over 350 million players at any time, a play 10.6 billion games a day on the platform.

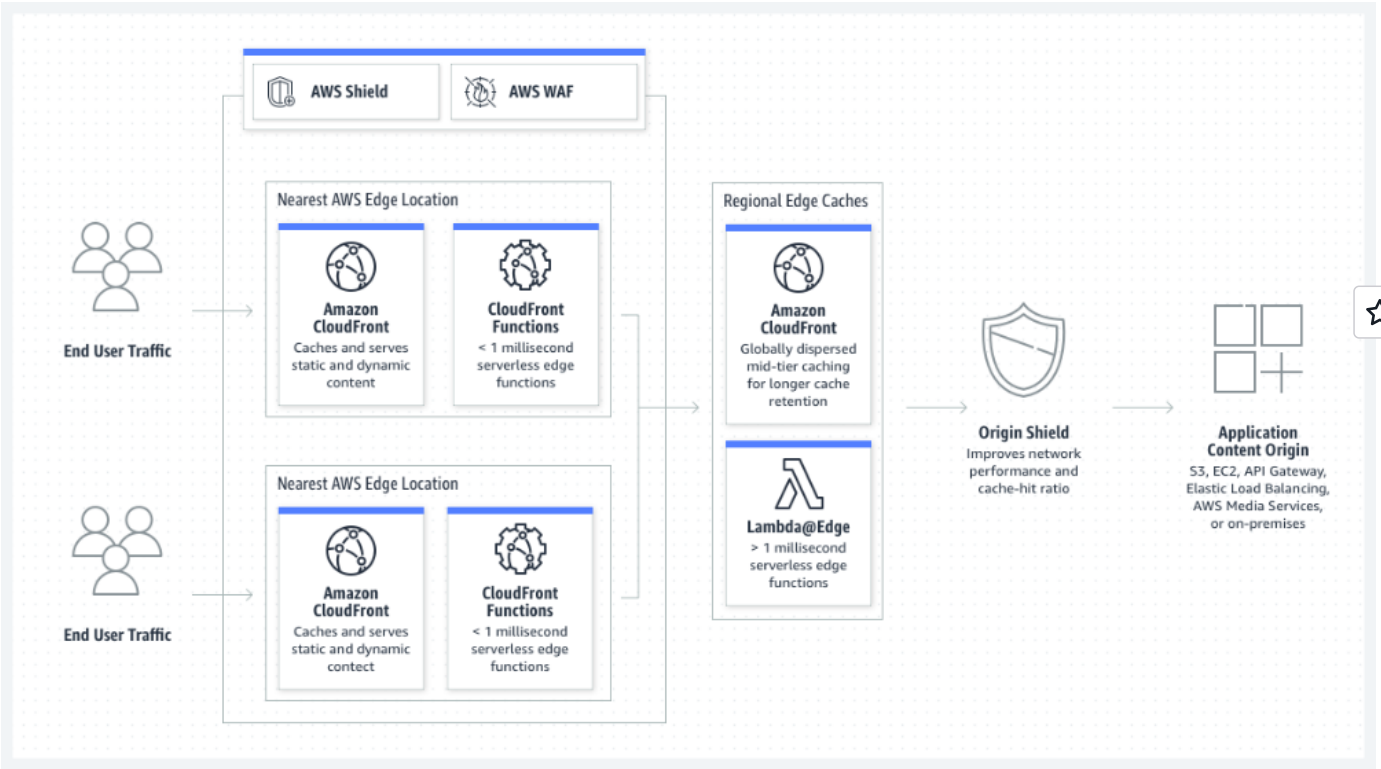
King's game applications record users' game data on central data centers, allowing them to play on different devices without losing progress. The data centers aim to give users a consistent experience, even if users join the game on old machines with limited bandwidth.

King uses Amazon CloudFront to deliver hundreds of terabytes of content daily, with spikes to half a petabyte or more when it launches a new game or initiates a large-scale marketing program.

What is Amazon CloudFront?

Amazon CloudFront is a content delivery network (CDN) service built for high performance, security, and developer convenience. You can use Amazon CloudFront to do these tasks:

- Deliver data through 450+ globally dispersed points of presence (POPs) with automated network mapping and intelligent routing.
- Improve security with traffic encryption and access controls, and use AWS Shield Standard to defend against distributed denial-of-service (DDoS) attacks at no additional charge.
- Customize the code you run at the AWS network edge using serverless compute features to balance cost, performance, and security.
- Scale automatically to deliver software, game patches, and IoT updates with high transfer rates.



Start building on AWS CloudFront with 50 GB of free data transfer out for 12 months. Create a free AWS account today.

Next steps on AWS

Check out additional product-related resources

→

Sign up for a free account

Instant get access to the AWS Free Tier.

→

Start building in the console

Get started building in the AWS management console.

→

Learn

What Is AWS?

What Is Cloud Computing?

What Is Generative AI?

Cloud Computing Concepts Hub

AWS Cloud Security

What's New

Blogs

Press Releases

Resources

Getting Started

Training

AWS Trust Center

AWS Solutions Library

Architecture Center

Product and Technical FAQs

Analyst Reports

AWS Partners

Developers

Builder Center

SDKs & Tools

.NET on AWS

Python on AWS

Java on AWS

PHP on AWS

JavaScript on AWS

Help

Contact Us

File a Support Ticket

AWS re:Post

Knowledge Center

AWS Support Overview

Get Expert Help

AWS Accessibility

Legal

Amazon is an Equal Opportunity Employer: Minority / Women / Disability / Veteran / Gender Identity / Sexual Orientation / Age.

Back to top

Privacy Site terms Your Privacy Choices Cookie Preferences

© 2025, Amazon Web Services, Inc. or its affiliates. All rights reserved.