

# Web Server Log Analyzer

The **Web Server Log Analyzer** project is designed to process and analyze raw server log data stored in **Amazon S3**. Web servers generate log files that contain valuable information about user activity, error messages, IP addresses, request paths, timestamps, and more. These logs, when analyzed properly, can provide insights into website performance, security issues, user behavior, and server health.

- Apache or Nginx web logs.
- Glue Transformations:
- Parse log lines into structured fields
- Extract metrics (top IPs, most visited URLs, error rate)
- Output (S3): Structured, queryable logs or analytics-ready data.

Apache or Nginx web logs:-

**Apache or Nginx web server logs** using a cloud-native architecture built on **Amazon Web Services (AWS)**. Both Apache and Nginx generate structured log files in formats such as the **Common Log Format (CLF)** or **Combined Log Format**, which contain valuable information about incoming HTTP requests—such as IP addresses, timestamps, URLs, request methods, status codes, user-agents, and referrers. These logs are typically generated on EC2 instances or web servers running in the cloud, and are then uploaded periodically to an **Amazon S3** bucket for centralized storage.

The screenshot shows the AWS Management Console interface for creating a new S3 bucket. The top navigation bar includes the AWS logo, a search bar, and user information. The breadcrumb trail indicates the path: Amazon S3 > Buckets > Create bucket. The main heading is 'Create bucket' with an 'Info' link. Below this, a note states 'Buckets are containers for data stored in S3.' The 'General configuration' section is active, showing the 'AWS Region' as 'US East (N. Virginia) us-east-1'. Under 'Bucket type', the 'General purpose' option is selected, with a description: 'Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.' The 'Directory' option is also visible, described as 'Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.' The 'Bucket name' field contains 'apache-server'. A note at the bottom explains that bucket names must be 3 to 63 characters long, unique globally, and follow specific character rules. The footer includes 'CloudShell', 'Feedback', and copyright information for Amazon Web Services, Inc. (2025), along with links for 'Privacy', 'Terms', and 'Cookie preferences'.

aws [Search] [Alt+S] United States (N. Virginia) omvishwambharpatil

EC2

Amazon S3 > Buckets > Create bucket

## Create bucket [Info](#)

Buckets are containers for data stored in S3.

### General configuration

**AWS Region**  
US East (N. Virginia) us-east-1

**Bucket type** [Info](#)

☒ **General purpose**  
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

☐ **Directory**  
Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

**Bucket name** [Info](#)  
apache-server

Bucket names must be 3 to 63 characters and unique within the global namespace. Bucket names must also begin and end with a letter or number. Valid characters are a-z, 0-9, periods (.), and hyphens (-). [Learn More](#)

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

[Alt+S]

United States (N. Virginia)

omvishwambharpatil

☰

EC2 > Instances

Instances (1) [Info](#)

Connect

Instance state ▾

Actions ▾

Launch instances ▾

All states ▾

<

1

>

<input type="checkbox"/>	Name	Instance ID	Instance state	Instance type	Status check	Alarm status	Availability Zone	Public IP
<input type="checkbox"/>	apache-webse...	i-0d9c2263767bd25f8	Running	t2.micro	Initializing	<a href="#">View alarms +</a>	us-east-1c	ec2-3-93-

Select an instance

```
#
~\ ##### Amazon Linux 2023
~~\#####
~~\###|
~~\#/ https://aws.amazon.com/linux/amazon-linux-2023
~~V~' '~>
~~~~
~~~.
~/m/'
```

```
last login: Tue Jun 10 06:29:50 2025 from 18.206.107.29
ec2-user@ip-172-31-84-114 ~]$ sudo yum install httpd
Last metadata expiration check: 0:38:22 ago on Tue Jun 10 06:33:47 2025.
Package httpd-2.4.62-1.amzn2023.x86_64 is already installed.
Dependencies resolved.
Nothing to do.
Complete!
ec2-user@ip-172-31-84-114 ~]$ rpm -qa | grep httpd
httpd-tools-2.4.62-1.amzn2023.x86_64
```

```
aws [Search] [Alt+S] United States (N. Virginia)
EC2

ec2-user@ip-172-31-84-114 ~]$ rpm -qa | grep httpd
httpd-tools-2.4.62-1.amzn2023.x86_64
httpd filesystem-2.4.62-1.amzn2023.noarch
httpd-core-2.4.62-1.amzn2023.x86_64
generic-logos-httpd-18.0.0-12.amzn2023.0.3.noarch
httpd-2.4.62-1.amzn2023.x86_64
ec2-user@ip-172-31-84-114 ~]$ sudo systemctl status httpd
httpd.service - The Apache HTTP Server
Loaded: loaded (/usr/lib/systemd/system/httpd.service; disabled; preset: disabled)
Active: active (running) since Tue 2025-06-10 06:56:49 UTC; 17min ago
Docs: man:httpd.service(8)
Main PID: 27175 (httpd)
Status: "Total requests: 0; Idle/Busy workers 100/0; Requests/sec: 0; Bytes served/sec: 0 B/sec"
Tasks: 230 (limit: 1111)
Memory: 16.0M
CPU: 749ms
CGroup: /system.slice/httpd.service
└─27175 /usr/sbin/httpd -DFOREGROUND
└─27184 /usr/sbin/httpd -DFOREGROUND
```

# Nginx web logs

Nginx web logs provide critical visibility into web server activity and user interactions. Nginx generates logs in formats such as the **Access Log** and **Error Log**, typically following the **Combined Log Format**

Nginx web logs are continuously pushed to **Amazon S3**, acting as a scalable and durable log repository. This enables decoupling of log generation and analysis processes.

### Launch an instance [Info](#)

Amazon EC2 allows you to create virtual machines, or instances, that run on the AWS Cloud. Quickly get started by following the simple steps below.

#### Name and tags [Info](#)

Name

[Add additional tags](#)

#### ▼ Application and OS Images (Amazon Machine Image) [Info](#)

An AMI is a template that contains the software configuration (operating system, application server, and applications) required to launch your instance. Search or Browse for AMIs if you don't see what you are looking for below

#### ▼ Summary

Number of instances [Info](#)

Software Image (AMI)  
Amazon Linux 2023 AMI 2023.7.2...[read more](#)  
ami-02457590d33d576c3

Virtual server type (instance type)  
t2.micro

[Cancel](#) [Launch instance](#) [Preview code](#)

CloudShell Feedback © 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

EC2

Package	Architecture	Version	Repository	Size
Installing:				
nginx	x86_64	1:1.26.3-1.amzn2023.0.1	amazonlinux	33 k
Installing dependencies:				
generic-logos-httpd	noarch	18.0.0-12.amzn2023.0.3	amazonlinux	19 k
gperftools-libs	x86_64	2.9.1-1.amzn2023.0.3	amazonlinux	308 k
libunwind	x86_64	1.4.0-5.amzn2023.0.2	amazonlinux	66 k
nginx-core	x86_64	1:1.26.3-1.amzn2023.0.1	amazonlinux	670 k
nginx-filesystem	noarch	1:1.26.3-1.amzn2023.0.1	amazonlinux	9.6 k
nginx-mimetypes	noarch	2.1.49-3.amzn2023.0.3	amazonlinux	21 k

Transaction Summary

Install 7 Packages

Total download size: 1.1 M

Installed size: 3.6 M

i-06fe8aece6d35a262 (nginx-server)

PublicIPs: [98.82.2.43](#) PrivateIPs: 172.31.92.28

# Glue Transformations:

AWS Glue is a fully managed extract, transform, and load (ETL) service that enables automatic data discovery, cleaning, enrichment, and preparation for analytics. After web logs are ingested into Amazon

The image shows two screenshots of the AWS IAM console. The top screenshot is the IAM dashboard, and the bottom screenshot is the 'Create role' wizard.

**IAM dashboard:**

- Security recommendations:** 1 recommendation. **Root user has MFA** (green checkmark). Having multi-factor authentication (MFA) for the root user improves security for this account. **Deactivate or delete access keys for root user** (red warning icon). Deactivate or delete the access keys for the root user. Instead, use access keys attached to an IAM user to improve security. [Manage access keys](#)
- IAM resources:**

User groups	Users	Roles	Policies	Identity providers
1	2	154	91	0
- What's new:** Updates for features in IAM. [View all](#)
  - Right-size permissions for more roles in your account using IAM Access Analyzer to generate 50 fine-grained IAM policies per day. 9 months ago
  - Amazon S3 Object Ownership can now disable access control lists to simplify access management for data in S3. 9 months ago
  - Amazon Redshift simplifies the use of other AWS services by introducing the default IAM role. 9 months ago
- AWS Account:** Account ID: 052674914236, Account Alias: 052674914236, [Create](#), Sign-in URL: <https://052674914236.signin.aws.amazon.com/console>
- Quick Links:** [My security credentials](#) (Manage your access keys, multi-factor authentication (MFA) and other credentials).
- Tools:** [Policy simulator](#) (The simulator evaluates the policies that you choose and determines the effective permissions for each of the actions that you specify), [Web identity federation playground](#) (Authenticate yourself to any of the supported web identity providers, see the requests and

**Create role wizard:**

- Step 1: Select trusted entity**
- Trusted entity type:**
  - ☒ **AWS service**: Allow AWS services like EC2, Lambda, or others to perform actions in this account.
  - ☐ **AWS account**: Allow entities in other AWS accounts belonging to you or a 3rd party to perform actions in this account.
  - ☐ **Web identity**: Allows users federated by the specified external web identity provider to assume this role to perform actions in this account.
  - ☐ **SAML 2.0 federation**: Allow users federated with SAML 2.0 from a corporate directory to perform actions in this account.
  - ☐ **Custom trust policy**: Create a custom trust policy to enable others to perform actions in this account.
- Use case:** Allow an AWS service like EC2, Lambda, or others to perform actions in this account.
  - Common use cases:**
    - ☐ **EC2**: Allows EC2 instances to call AWS services on your behalf.
    - ☐ **Lambda**: Allows Lambda functions to call AWS services on your behalf.
  - Use cases for other AWS services:** [Choose a service to view use case](#)

aws

Services

Search for services, features, blogs, docs, and more

[Option+S]

Global

we've redesigned the IAM roles experience to make it easier to use. [Let us know what you think.](#)

IAM > Roles > Create role

Step 1  
Select trusted entity

Step 2  
Add permissions

Step 3  
Name, review, and create

Name, review, and create

Role details

Role name

Enter a meaningful name to identify this role.

aws-glue-service-role-demo

Maximum 64 characters. Use alphanumeric and '+', '@', '-' characters.

Description

Add a short explanation for this role.

Allows Glue to call AWS services on your behalf.

Maximum 1000 characters. Use alphanumeric and '+', '@', '-' characters.

Step 1: Select trusted entities

1- {

2-   "Version": "2012-10-17",

3-   "Statement": [

4-     {

5-       "Effect": "Allow",

6-       "Principal": {

7-          "Service": "glue.amazonaws.com"

8-       },

9-       "Action": "sts:AssumeRole"

10-   }

aws

Services

Search

Global

Identity and Access Management (IAM)

Search IAM

Dashboard

Access management

User groups

Users

**Roles**

Policies

Identity providers

Account settings

Access reports

Access analyzer

Archive rules

Analizers

Settings

Credential report

Organization activity

Service control policies (SCPs)

Related consoles

[IAM Identity Center](#) **New**

New! Securely access AWS services from your data center with IAM Roles Anywhere. [Learn more](#)

Role aws-glue-service-role-demo created

[View role](#)

IAM > Roles

Roles (155) [Info](#)

An IAM role is an identity you can create that has specific permissions with credentials that are valid for short durations. Roles can be assumed by entities that you trust.

Search

< 1 2 3 4 5 6 7 8 > ⚙

<input type="checkbox"/>	Role name	Trusted entities	Last activity
<input type="checkbox"/>	A2ISageMaker-ExecutionRole-20211024T130530	AWS Service: sagemaker	250 days ago
<input type="checkbox"/>	A2ISageMaker-ExecutionRole-20211219T140706	AWS Service: sagemaker	251 days ago
<input type="checkbox"/>	A2ISageMaker-ExecutionRole-20211219T145014	AWS Service: sagemaker	251 days ago
<input type="checkbox"/>	A2ISageMaker-ExecutionRole-20211220T194051	AWS Service: sagemaker	51 days ago
<input type="checkbox"/>	accessTokenAuth-role-y59kizz6	AWS Service: lambda	80 days ago
<input type="checkbox"/>	AmazonAppStreamServiceAccess	AWS Service: appstream	1 hour ago
<input type="checkbox"/>	AmazonComprehendServiceRole-compre-train	AWS Service: comprehend	-
<input type="checkbox"/>	AmazonComprehendServiceRole-comprehend-custom	AWS Service: comprehend	-
<input type="checkbox"/>	AmazonComprehendServiceRole-entity-recognizer	AWS Service: comprehend	-
<input type="checkbox"/>	AmazonSSMRoleForAutomationAssumeQuickSetup	AWS Service: ssm	-

aws

Services

Search for services, features, blogs, docs, and more

[Option+S]

Global

Query with S3 Select [Info](#)

Use Amazon S3 Select to retrieve a subset of data from an object using standard SQL queries. Pricing is based on the size of the input, query results, and data transferred. [Learn more](#) or [see Amazon S3 pricing](#)

Input settings

Path

s3://aws-glue-demo-tutorial-series/data-store/annual\_reports/csv\_reports/annual-enterprise-survey-2021-financial-year-provisional-csv.csv

Size

6.3 MB (6617174.0 B)

Format

☒ CSV

☐ JSON

☐ Apache Parquet

CSV delimiter

☒ Comma

☐ Tab

☐ Custom

☐ Exclude the first line of CSV data

Enable this setting if CSV contains a header row.

Compression

☒ None

☐ GZIP

☐ BZIP2

Output settings

# Parse log lines into structured fields:-

parse raw log lines into structured fields. Web logs from servers like Nginx or Apache are typically stored as unstructured text, where each line represents a request in a specific log format such as the Common Log Format (CLF) or the Combined Log Format. These lines contain information like the client IP address, request timestamp, HTTP method, requested URL, HTTP status code, user-agent, and referrer—all separated by spaces or enclosed in quotes.

The screenshot displays the AWS CloudWatch Logs Insights console. The left sidebar shows navigation options like 'dWatch', 'ites and recents', 'boards', 'ns', 'roups', 'Insights', 'cs', 'y traces', 'ts', 'cation monitoring', 'hts', 'igs', and 'g Started'. The main area is titled 'CloudWatch > Logs Insights'. It features a search bar, a query editor with the following query:

```
1 fields @timestamp, @message
2 | sort @timestamp desc
3 | limit 20
```

Buttons for 'Run query', 'Save', and 'History' are present. Below the query editor, it states 'Queries are allowed to run for up to 15 minutes.' The 'Logs' tab is selected, showing 'No results' and a prompt to 'Run a query to see related events'. The 'Visualization' tab is also visible, showing a histogram of log records. The histogram displays a distribution of log records over time, with a peak around 02:40. Below the histogram, a table shows the first 8 results of the query:

#	@timestamp	@message
1	2022-08-31T15:37:29.3...	10.0.0.69 - - [31/Aug/2022:20:37:29 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
2	2022-08-31T15:37:28.4...	10.0.0.160 - - [31/Aug/2022:20:37:28 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
3	2022-08-31T15:36:59.3...	10.0.0.69 - - [31/Aug/2022:20:36:59 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
4	2022-08-31T15:36:59.3...	10.0.0.69 - - [31/Aug/2022:20:36:59 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
5	2022-08-31T15:36:58.4...	10.0.0.160 - - [31/Aug/2022:20:36:58 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
6	2022-08-31T15:36:58.4...	10.0.0.160 - - [31/Aug/2022:20:36:58 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
7	2022-08-31T15:36:29.3...	10.0.0.69 - - [31/Aug/2022:20:36:29 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"
8	2022-08-31T15:36:29.3...	10.0.0.69 - - [31/Aug/2022:20:36:29 +0000] "GET / HTTP/1.1" 200 615 "-" "ELB-HealthChecker/2.0" "-"

## Extract metrics (top IPs, most visited URLs, error rate)

After parsing the raw Nginx or Apache log data into structured fields, the next crucial step is to extract key metrics that provide insights into server usage and performance. Among the most

valuable metrics are the top client IP addresses, which help identify the most active users or potential malicious actors (e.g., in cases of denial-of-service



**Logs**  
**1,313** Events/day  
[View logs](#)

**Metrics**  
**108** Points/day  
[View metrics](#)

**Resources**  
**126** Updates/day  
[View resources](#)

#### Metrics [API polling] (87)

Selected AWS service metrics and data points collected in the last hour for each service.

AWS services		Metrics selected	Points/Hour ↓
	AWS/EC2	23 of 34 metrics	108
	AWS/RDS	64 of 128 metrics	0

### Edit AWS/EC2 metrics

#### Select metrics

[All](#)[Selected](#)[Unselected](#)

Metrics (23/34)

[Select all](#)[Unselect all](#)[Reset to defaults](#)

- ☒ CPUCreditBalance
- ☒ CPUCreditUsage
- ☒ CPUSurplusCreditBalance
- ☒ CPUSurplusCreditsCharged
- ☒ CPUUtilization
- ☒ DedicatedHostCPUUtilization

#### Tags

CloudWatch Metrics can be filtered by tags. A tag is a key-value pair that you have applied to the resources of this AWS service.

[+ Add tag](#)

#### Dimensions

[Cancel](#)[Apply settings](#)