aws re/start

AWS Service Integration with Amazon Athena

The module will now explore AWS Service Integration with Amazon Athena.

What you will learn

At the core of the lesson

You will learn how to:

- Explain how to use Amazon Athena to query data from other AWS services

Topics:

- Amazon Athena

Key terms:

Amazon Athena

aws re/start

This module includes a technical discussion of AWS service integration with Amazon Athena. Amazon Athena is an interactive query service that makes it easy to analyze data in Amazon Simple Storage Service (Amazon S3) by using standard structured query language (SQL).

At the end of this module, you will be able to:
- Explain how to use Amazon Athena to query data from other AWS services

Amazon Athena

Amazon Athena characteristics:

- Handles large scale datasets with ease
- Serverless—Avoids need to extract, transform, and load (ETL)
- Pay only for the queries that you run
- Athena automatically runs queries in parallel so most results come back in seconds.

To use Amazon Athena, point to your data in Amazon Simple Storage Service (Amazon S3), define the schema, and start querying by using standard SQL. Most results are delivered within seconds. With Athena, you do not need complex ETL jobs to prepare your data for analysis. Athena makes it easy for anyone with SQL skills to quickly analyze large-scale datasets.

Athena works with various standard data formats, including comma-separated values (CSV), JavaScript Object Notation (JSON), Optimized Row Columnar (ORC), Apache Avro, and Apache Parquet. Athena is ideal for quick, ad hoc querying. However, it can also handle complex analysis, including large joins and arrays. Athena uses Amazon S3 as its underlying data store, which makes your data highly available and durable.

Athena is serverless, so there is no infrastructure to manage, and you pay only for the queries that you run. You can quickly query your data without needing to set up and manage any servers or data warehouses. Athena enables you to query all your data in Amazon S3 without needing to set up complex processes to extract, transform, and load ETL) the data.

Also, Athena offers fast, interactive query performance. Athena automatically runs queries in parallel, so most results come back within seconds.

Getting started with Amazon Athena

1. **Create** an Amazon Simple Storage Service (Amazon S3) bucket and load data into it
   - Alternatively, have an existing bucket with data already in it

2. **Define** the schema (table definition) that describes the data structure

3. **Start** querying data by using structured query language (SQL)

Data

SQL query

User   Query results   Amazon Athena   S3 bucket
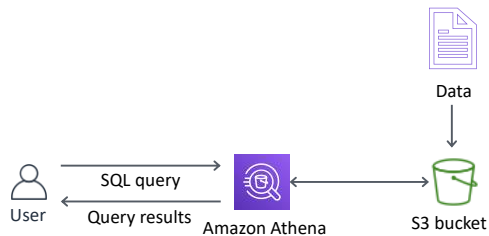
aws re/start

To use Athena, point to your data in Amazon S3, define the schema, and start querying by using the built-in query editor.

1. Get started by browsing to the **Amazon Athena** service in the AWS Management Console, and then choose **Get Started**. The Athena Query Editor opens. It lists databases and tables in the left column. You can create a database by running a command such as `CREATE DATABASE mydatabase`. Then, you create a table. The table defines the *schema* of the data. Athena table definitions look similar to the table definitions for relational databases. To see an example table definition, refer to Getting Started in the Amazon Athena documentation.

   The table definition ends with a `LOCATION` statement, which points to the S3 bucket where your data resides (or where it will reside).

2. After you define the table, you can start running standard SQL queries by using the *Athena Query Editor*. For example, you could use the following query:
   `SELECT * FROM tableName WHERE columnName='value';`

3. Results will display in the Athena Query Editor. However, you can also download

the results as comma-separated values (CSV) files. Also, you can create a client to access Athena and run SQL queries programmatically. For an example of such a client (written in Java), refer to the Code Samples page of the Amazon Athena documentation.

AWS service integrations with Amazon Athena

Amazon Athena works with other AWS services

Athena makes it easier to query logs from services, such as:
- AWS CloudTrail
- Application Load Balancer logs
- Amazon Virtual Private Cloud (Amazon VPC) Flow Logs

aws re/start

You can query data from other AWS services in Athena, including services that monitor and secure on your account (which is why Athena is covered in this module).

For example, using Athena with *CloudTrail logs* can enhance your analysis of AWS service activity. You can automatically create tables for querying CloudTrail logs directly from the CloudTrail console, and use those tables to run queries in Athena. You can then use queries to identify trends and further isolate activity by attribute, such as source IP address or user.

Querying *Application Load Balancer logs* with Athena enables you to see the source of traffic, latency, and bytes that are transferred to and from Elastic Load Balancing instances and backend applications.

Amazon Virtual Private Cloud (Amazon VPC) flow logs capture information about the IP traffic that goes to and from network interfaces in a VPC. Query the logs in Athena to investigate network traffic patterns, and to identify threats and risks across your Amazon VPC network.

For the full list of AWS services that integrate with Amazon Athena, refer to AWS

[Service Integrations with Athena](#).

Key takeaways

You can use Amazon Athena to query logs, comma-separated values (CSV) files, and other data files. Some examples are:

- **AWS CloudTrail** logs to enhance your analysis of AWS service activity.

- **Elastic Load Balancing logs** to see the source of traffic, latency, and bytes that are transferred to and from ELB instances and backend applications.

- **Amazon VPC Flow Logs** to investigate network traffic patterns, and to identify threats and risks across your Amazon VPC network.

aws re/start

Some key takeaways from this section include:

- Use Athena with **CloudTrail logs** to enhance your analysis of AWS service activity.

- Use Athena to query **Application Load Balancer logs**. This query enables you to see the source of traffic, latency, and bytes that are transferred to and from ELB instances and backend applications.

- Use Athena with **Amazon Virtual Private Cloud Flow Logs** to capture information about the IP traffic that goes to and from network interfaces (NIC) in a VPC. Query the logs in Athena to investigate network traffic patterns, and to identify threats and risks across your Amazon VPC network.