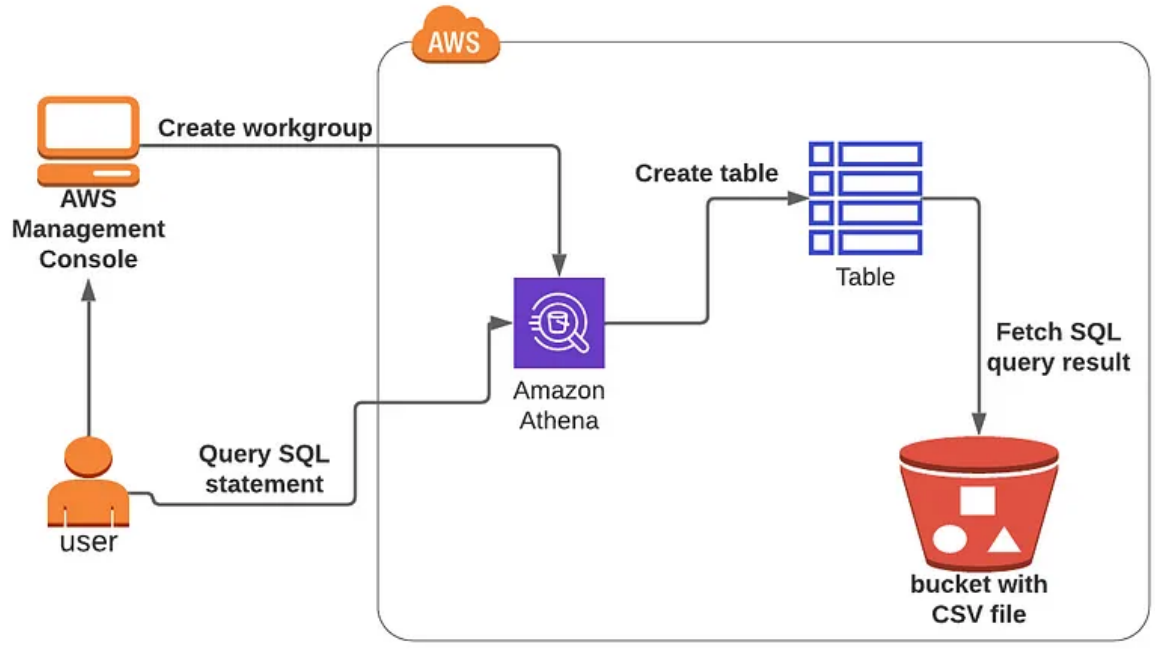
Athena

Named after the Greek goddess of wisdom, Amazon Athena is a serverless, interactive query service that allows you to analyze data stored in different locations and formats as a single dataset using standard SQL, mostly sourcing the data from Amazon S3 and supports a wide variety of file formats such as CSV, JSON, ORC and Apache Paraquet right out of the box thereby eliminating the need for complex ETL processes and infrastructure setup, enabling the user to start querying their data immediately.

Though not necessarily low latency in nature, Athena usually returns the results of queries in a matter of seconds, making it rather quick to use which alongside its relative cost-effectiveness has led to its widespread integration and use with business intelligence tools such as Microsoft PowerBI and Amazon Quicksight (A service we will learn more about later). In fact, Amazon even provides its own Microsoft Power BI Connector for Amazon Athena which can be used to analyze data present in Amazon Athena from the Microsoft PowerBI desktop application directly (at least on Windows machines).

Also worth mentioning is that workloads that utilize Apache Spark can also be run using Amazon Athena making it extremely easy to run analytics applications without the need to provision or manage resources and infrastructure, due to the serverless nature of Amazon Athena. In fact, since Athena is a serverless offering, when using it to run queries on the desired S3 buckets, the user usually only has to pay for the queries they run, and nothing else, making it a cost-effective solution for ad-hoc data analysis and reporting.



Amazon AthenaSQL Architecture (Credit: Whizlabs)

Amazon Athena works through something called workgroups, where users can control and manage their different query statements as well as the resources required to run said queries. Once the workgroup is created, the user can input SQL queries into Athena through the AWS Management Console or Command Line Interface (CLI) as shown in the diagram above. Athena, a query service, interacts with data stored in Amazon S3, running SQL statements directly on the files without requiring data movement.

Next, a table is created in Athena which serves as a metadata layer over the data in the S3 bucket. The actual data, in this case, is stored as CSV files in S3, and Athena reads the table structure to know how to query it. Finally, after the query is executed, Athena fetches the results from the CSV file in the S3 bucket and presents them to the user in the form of SQL query results, which can be retrieved through the console/ CLI.

Similar architecture and methods are used when running Apache Spark applications using Amazon Athena.

TLDR;

Amazon Athena can be used to run queries on-demand on an S3 bucket with support for a wide variety of file formats such as CSV, JSON, Apache Paraquet, etc. The queries run by Amazon Athena usually return results within a matter of seconds and the service only charges for the queries and not the associated resources required to run them.

The service also supports running analytics applications that utilize Apache Spark workloads.