# Athena Query Cheat Sheet

This cheat sheet provides AWS Athena queries to assist during the training labs.

**Contact us**
**info@invictus-ir.com**

## > Querying Basics

### Querying a table

To query data, you first need to specify the table name you want to explore.

```
SELECT *
FROM ctf1_cloudtrail_logs
```

Querying the table without any filters or operations will return all rows and columns from the **ctf1_cloudtrail_logs** table.

### Using LIMIT to limit the number of rows

The **LIMIT** clause is used to restrict the number of rows returned in the query result. This is useful when you only want to see a subset of the data.

```
SELECT *
FROM ctf1_cloudtrail_logs
LIMIT 10
```

This query returns the first 10 rows from the specified table.

### Selecting specific columns

The **SELECT** statement is used to specify the columns you want to include in your results. This is useful when you only need to display or analyze certain columns from a table.

```
SELECT eventtime, eventSource, sourceipaddress
FROM ctf1_cloudtrail_logs
```

In this query, the SELECT statement retrieves and displays only the columns eventtime, eventSource, and sourceipaddress from the ctf1_cloudtrail_logs table.

## > Filtering Data

### Where clause

Use the **WHERE** clause to filter data based on a condition. This allows you to focus on specific rows that match the criteria.

```
SELECT eventtime, eventSource, eventName, sourceipaddress
FROM ctf1_cloudtrail_logs
WHERE eventSource = 'iam.amazonaws.com'
```

This query will return all rows where the eventSource is iam.amazonaws.com.

### LIKE operator

The **LIKE** operator in Athena is used for searching partial matches in a specific column. It allows you to find the occurrence of a value or pattern within the data in a column.

```
SELECT eventtime, eventSource, eventName, sourceipaddress
FROM ctf1_cloudtrail_logs
WHERE sourceipaddress LIKE '%192.168%'
```

This query filters rows where the sourceipaddress column contains the substring "192.168". The % symbols act as wildcards, meaning that it will match anything before or after 192.168.

### Using subfields

To extract values from subfields in AWS Athena, you can directly query a nested or structured field. This allows you to work with specific properties that are part of a more complex field.

```
SELECT useridentity.arn
FROM ctf1_cloudtrail_logs
```

In this query, useridentity.arn is used to extract the arn value from the useridentity field in the ctf1_cloudtrail_logs table.

### Ordering results by eventtime

The **ORDER BY** clause is used to sort the results of a query by a specific column, either in ascending or descending order. This is useful when you need to analyze data over time or prioritize certain records.

```
SELECT eventtime, eventSource, eventName, sourceipaddress
FROM ctf1_cloudtrail_logs
ORDER BY eventtime ASC
```

This query retrieves data from the ctf1_cloudtrail_logs table and sorts the results by the eventtime field in ascending order (**ASC**). To sort in descending order, use **DESC** instead.

## > Counting

### Counting rows with COUNT

The **COUNT** function is used to return the total number of rows that match a specific condition.

```
SELECT COUNT(*)
FROM ctf1_cloudtrail_logs
WHERE eventSource = 'guardduty.amazonaws.com'
```

This query counts the total number of rows in the ctf1_cloudtrail_logs table where the eventSource is 'guardduty.amazonaws.com'.

### Summarize data

You can use aggregate functions like **SUM**, **AVG**, **MIN**, **MAX**, and **COUNT** to summarize data. This is helpful when you want to calculate metrics like totals, averages, or other statistics for your dataset.

```
SELECT eventSource, COUNT(*) AS total_events
FROM ctf1_cloudtrail_logs GROUP BY eventSource
```

In this query, the **COUNT(*)** function is used to count the total number of events for each eventSource.

## > Important CloudTrail fields

| Column name | Description |
|---|---|
| useridentity | Contains details about the identity that made the request, including the type (e.g., IAM user, root) and ARN. |
| eventtime | The time when the event occurred. |
| eventname | The specific action that was performed, like StartInstances or CreateUser. |
| sourceipaddress | The IP address from where the request was made. |
| useragent | Information about the requester's environment, such as browser or tool used. |
| eventSource | The AWS service where the event originated, such as ec2.amazonaws.com or guardduty.amazonaws.com. |

INVICTUS INCIDENT RESPONSE