



# **Amazon Web Services**

## **Data Engineering Immersion Day**

---

Lab 3. Consuming data with Athena and Quicksight

***August 2020***

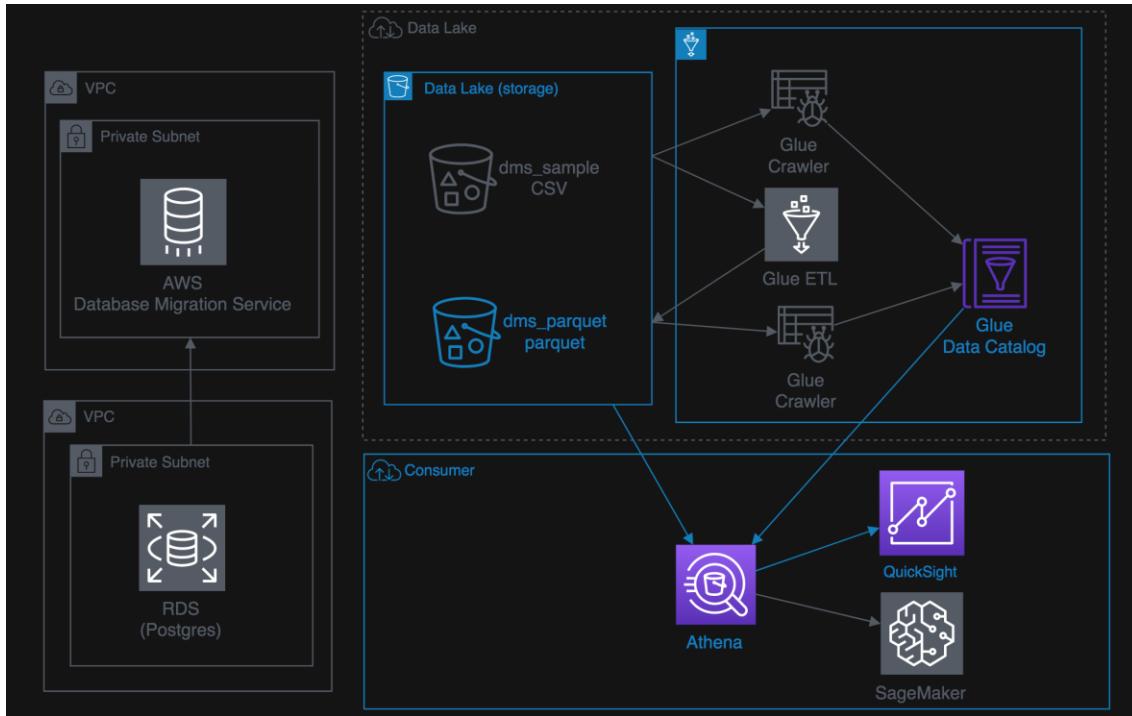
## Table of Contents

<i>Introduction</i> .....	2
Prerequisites .....	2
Getting Started .....	2
<i>Get Started Using the Lab Environment</i> .....	3
<i>Query Data with Amazon Athena</i> .....	6
<i>Build an Amazon QuickSight Dashboard</i> .....	14
Set up QuickSight.....	14
Create QuickSight Charts.....	21
Create QuickSight Parameters.....	24
Create a QuickSight Filter .....	27
Add Calculated Fields .....	29
<i>Amazon QuickSight ML-Insights (Optional)</i> .....	32
<i>Athena Workgroups to Control Query Access and Costs (Optional)</i> .....	34
Workflow setup to separate workloads .....	34
Explore the features of workgroups.....	37
Managing Query Usage and Cost .....	43
Cost Allocation Tags .....	50

## Lab 3. Consuming data with Athena and Quicksight

### Introduction

This lab introduces you to AWS Glue, Amazon Athena, and Amazon QuickSight. AWS Glue is a fully managed data catalog and ETL service; Amazon Athena queries data; and Amazon QuickSight provides visualization of the data you import.



### Prerequisites

The DMS Lab and Glue ETL lab is a prerequisite for this lab.

### Getting Started

In this lab, you will complete the following tasks:

1. [Query data and create a view with Amazon Athena](#)
2. [Athena Workgroups to Control Query Access and Costs](#)
3. [Build a dashboard with Amazon QuickSight](#)

The Lab is also available - <https://aws-dataengineering-day.workshop.aws/>

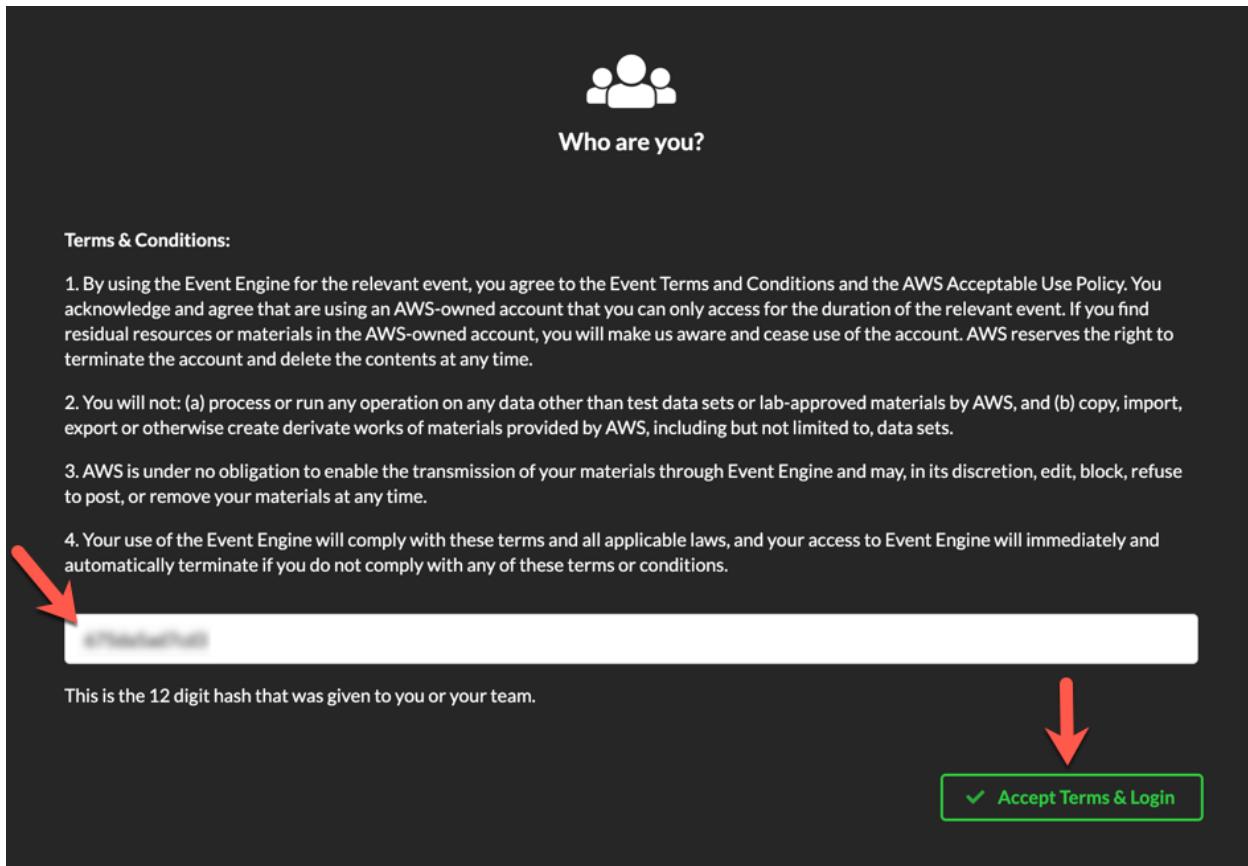
### Get Started Using the Lab Environment

Please skip this section if you are running the lab on your own AWS account.

Today, you are attending a formal event and you will have been sent your access details beforehand. If in the future you might want to perform these labs in your own AWS environment by yourself, you can follow instructions on GitHub - <https://github.com/aws-samples/data-engineering-for-aws-immersion-day>.

A 12-character access code (or ‘hash’) is the access code that grants you permission to use a dedicated AWS account for the purposes of this workshop.

1. Go to <https://dashboard.eventengine.run/>, enter the access code and click Proceed:



2. On the Team Dashboard web page you will see a set of parameters that you will need during the labs. Best to save them to a text file locally, alternatively you can always go to this page to review them. Replace the parameters with the corresponding values from here where indicated in subsequent labs:

## Lab 3. Consuming data with Athena and Quicksight

Because you're at a formal event, some AWS resources have been pre-deployed for your convenience, for example:

The screenshot shows a web interface titled "Environment Setup". At the top right is a "Readme" button. Below it, under the heading "Outputs:", there is a list of AWS resources with their corresponding ARNs:

- S3 Bucket name: mod-3fccddd609114925-dmslabs3bucket-1ngcgzzcnd15u
- BusinessAnalystUser: mod-3fccddd609114925-BusinessAnalystUser-MB0XFZLQLOXX
- DMSLabRoleS3 ARN: arn:aws:iam::377243295828:role/mod-3fccddd609114925-DMSLabRoleS3-O2VT1RSN43SG
- Glue Lab Role: mod-3fccddd609114925-GlueLabRole-YLTJA13WW6WT
- S3BucketWorkgroupA: mod-3fccddd609114925-s3bucketworkgroupa-tbon3m1mkunh
- S3BucketWorkgroupB: mod-3fccddd609114925-s3bucketworkgroupb-18ygl8nfp8ead
- WorkgroupManagerUser: mod-3fccddd609114925-WorkgroupManagerUser-5IVE0UQNIBG4

3. On the Team Dashboard, please click AWS Console to log into the AWS Management Console:

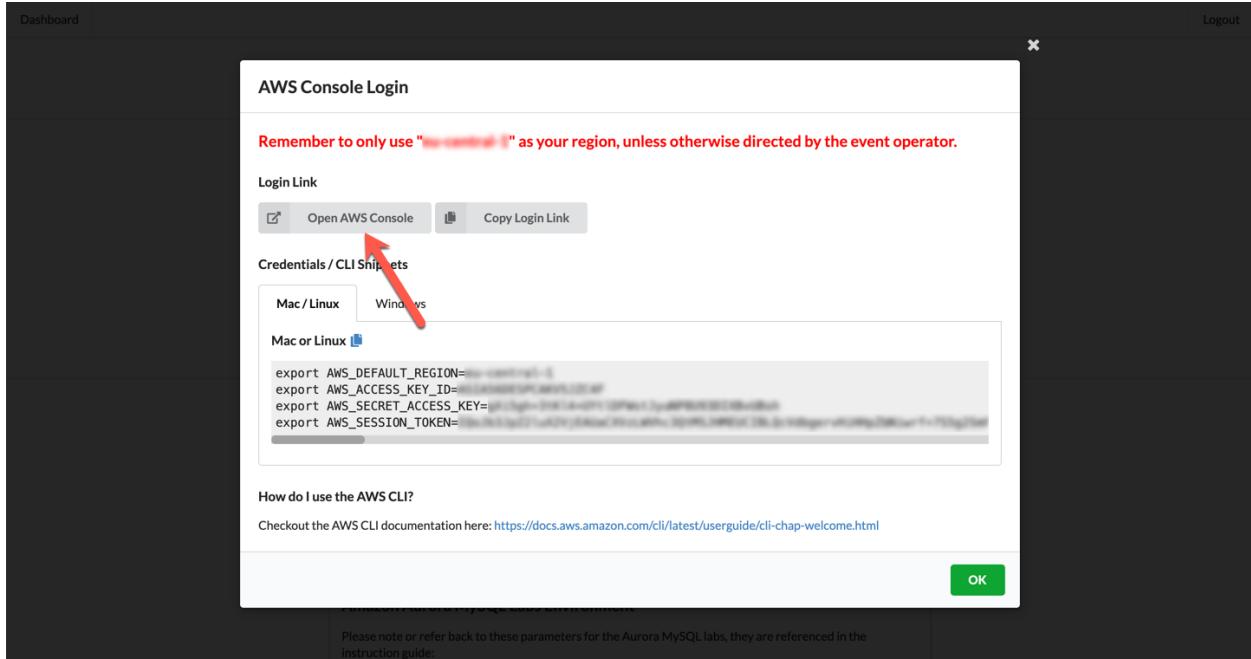
The screenshot shows a "Team Dashboard" interface. At the top center is a "Event" icon with a downward arrow pointing to a section below. This section contains two buttons: "AWS Console" and "SSH Key". Below these buttons, the event details are listed:

Event: Data Engineering Immersion Day - Test  
Team Name:

Event ID: d2302d4ae9ff4ea2857846b74f7de7e2  
Team ID: 1c2f7ad7ec044b0b8276f917c5983133

4. Click Open Console. For the purposes of this workshop, you will not need to use command line and API access credentials:

## Lab 3. Consuming data with Athena and Quicksight



Once you have completed these steps, you can continue with the rest of this lab.

## Lab 3. Consuming data with Athena and Quicksight

### Query Data with Amazon Athena

1. In the AWS services console, search for **Athena**.

The screenshot shows the AWS services console. A search bar at the top contains the text "athena". Below the search bar, the "Athena" service is listed with the subtext "Query Data in S3 using SQL". Other services like AWS Glue, IAM, S3, and Lambda are also visible in the navigation pane.

2. If you are using Athena first time, click on “Get Started” button in introduction screen.

The screenshot shows the "Amazon Athena" landing page. It features a central icon and text: "Amazon Athena is a fast, cost-effective, interactive query service that makes it easy to analyze petabytes of data in S3 with no data warehouses or clusters to manage." Below this are three main buttons: "Get Started", "Getting started guide", "Select a data set", "Create a table", and "Query data". There is also some descriptive text about data formats and supported SQL.

3. Setup the S3 buckets to store the query results. For this, Navigate to S3 console, click on bucket created as part of student lab ( e.g: <dmslab-student-dmslabs3bucket-xg1hdyq60ibs>). Inside bucket, click on “create folder”. Name folder as **athenaquery**. Click on **Save**.

The screenshot shows the AWS S3 console for the bucket "dmslab-student-dmslabs3bucket-xg1hdyq60ibs". The "Properties" tab is selected. A search bar at the top has "athenaquery" typed into it. Below the search bar are buttons for "Upload", "+ Create folder", "Download", and "Actions". The location "US East (N. Virginia)" is shown. The main area displays a table with one item, "athenaquery", which was just created. At the bottom, there is a note about encryption settings and three radio buttons: "None (Use bucket settings)" (selected), "AES-256", and "AWS-KMS". The "Save" and "Cancel" buttons are at the bottom right.

## Lab 3. Consuming data with Athena and Quicksight

4. Note down the path of S3 folder created above and save it.

The screenshot shows the Amazon S3 console interface. The path is 'Amazon S3 > dmslab-student-dmslabs3bucket-xg1hdq60ibs > athenaquery'. The 'Overview' tab is selected. A search bar contains the placeholder 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are buttons for 'Upload', 'Create folder', 'Download', and 'Actions'. On the right, it says 'US East (N. Virginia)' with a refresh icon. A message at the bottom states 'There are no objects under this path.'

5. In Athena Console, click on **Settings**.

The screenshot shows the Athena Query Editor. The top navigation bar includes 'Athena', 'Query Editor' (which is selected), 'Saved Queries', 'History', 'Data sources', 'Workgroup : primary', 'Settings' (which is highlighted), 'Tutorial', 'Help', and 'What's new'. The main area has tabs for 'Data source' (set to 'AwsDataCatalog') and 'Database' (set to 'ticketdata'). On the left, there is a sidebar titled 'Tables (24)' with a list of tables: bookmark\_parquet\_purchase\_history (Partitioned), cdc\_sporting\_event\_ticket, cdc\_ticket\_purchase\_hist, mib\_data, name\_data, rnf\_data, rnf\_stadium\_data, and parquet\_person. The main panel shows a query editor with a code completion dropdown containing various SQL examples. Buttons for 'Run query', 'Save as', and 'Create' are at the bottom, along with a note 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'. A 'Format query' and 'Clear' button are also present.

6. Copy paste the folder path as shown below:

The screenshot shows the 'Settings' dialog box. It has a title 'Settings' with a close button 'X'. Below it, a note says 'Settings apply by default to all new queries. [Learn more](#)'. The 'Workgroup: primary' section is highlighted. The 'Query result location' field is set to 's3://dmslab-student-dmslabs3bucket-xg1hdq60ibs/athenaquery/' with an info icon. Below it, an example is given: 'Example: s3://query-results-bucket/folder/'. There are two sections at the bottom: 'Encrypt query results' (unchecked) and 'Autocomplete' (unchecked). At the bottom right are 'Cancel' and 'Save' buttons.

7. In the Query Editor, select your newly created database e.g., "**ticketdata**".
8. Click the table named "**parquet\_sporting\_event\_ticket**" to inspect the fields.

**Note:** The type for fields id, sporting\_event\_id and ticketholder\_id should be (double).

## Lab 3. Consuming data with Athena and Quicksight

Next, we will query across tables `parquet_sporting_event`, `parquet_sport_team`, and `parquet_sport location`.

9. Copy the following SQL syntax into the New Query 1 tab and click **Run Query**.

```
SELECT
    e.id AS event_id,
    e.sport_type_name AS sport,
    e.start_date_time AS event_date_time,
    h.name AS home_team,
    a.name AS away_team,
    l.name AS location,
    l.city
FROM parquet_sporting_event e,
     parquet_sport_team h,
     parquet_sport_team a,
     parquet_sport_location l
WHERE
    e.home_team_id = h.id
    AND e.away_team_id = a.id
    AND e.location_id = l.id;
```

The results appear beneath the query window.

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS QuickSight interface. At the top, there are tabs for 'New query 1', 'New query 2', 'New query 3', and 'New query 4'. Below the tabs is a code editor window containing an Athena query:

```

SELECT
    e.id AS event_id,
    e.sport_type_name AS sport,
    e.event_date_time AS event_date_time,
    h.name AS home_team,
    a.name AS away_team,
    l.name AS location,
    c.name AS city
FROM parquet_sport_event e,
    parquet_sport_team h,
    parquet_sport_team a,
    parquet_sport_location l
WHERE
    e.home_team_id = h.id
    AND e.away_team_id = a.id
    AND e.location_id = l.id
    AND e.event_date_time > '2020-09-01 00:00:00'
    AND e.event_date_time < '2020-09-30 23:59:59'

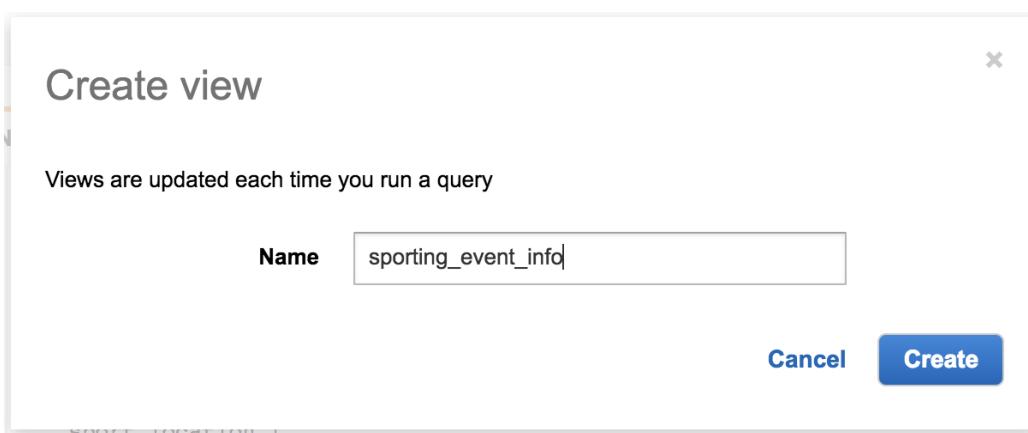
```

Below the code editor are buttons for 'Run query', 'Save as', 'Create', and 'Format query'. A status message indicates '(Run time: 2.76 seconds, Data scanned: 12.67 KB)'. A note says 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'. A tooltip for 'Create view from query' is shown over the 'Create' button.

The main area displays the query results in a table format. The columns are labeled: event\_id, sport, event\_date\_time, home\_team, away\_team, location, and city. The data includes various sports events from different teams and locations across North America.

10. As shown above Click **Create** and then select **Create view from query**

11. Name the view "sporting\_event\_info" and click **Create**.



Your new view is created

The screenshot shows the AWS Athena Query Editor. At the top, there are tabs for 'Athena', 'Query Editor' (which is selected), 'Saved Queries', 'History', 'Data sources', and 'Workgroup : primary'. The 'Query Editor' tab has a sub-menu with 'Data source', 'Database', 'Tables (24)', 'Views (1)', and 'Create table' and 'Create view' buttons.

In the 'Views (1)' section, there is a single entry: 'sporting\_event\_info'. Clicking on it reveals its schema:

```

event_id (bigint)
sport (string)
event_date_time (timestamp)
home_team (string)
away_team (string)
location (string)
city (string)

```

To the right of the schema, there is a 'Create view' button with a dropdown menu icon.

On the right side of the screen, there is another 'Query Editor' window titled 'New query 3'. It contains the following SQL syntax:

```

CREATE OR REPLACE VIEW "sporting_event_info" AS
SELECT
    e.id AS event_id,
    e.sport_type_name AS sport,
    e.event_date_time AS event_date_time,
    h.name AS home_team,
    a.name AS away_team,
    l.name AS location,
    c.name AS city
FROM parquet_sporting_event e,
    parquet_sport_team h,
    parquet_sport_team a,
    parquet_sport_location l
WHERE
    e.home_team_id = h.id
    AND e.away_team_id = a.id
    AND e.location_id = l.id

```

Below the query are buttons for 'Run query', 'Save as', 'Create', and a note '(Run time: 0.71 seconds, Data scanned: 0 KB)'. A note at the bottom says 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'.

12. Copy the following SQL syntax into the New Query 2 tab and click on **Save Query**.

### Lab 3. Consuming data with Athena and Quicksight

```
SELECT t.id AS ticket_id,
       e.event_id,
       e.sport,
       e.event_date_time,
       e.home_team,
       e.away_team,
       e.location,
       e.city,
       t.seat_level,
       t.seat_section,
       t.seat_row,
       t.seat,
       t.ticket_price,
       p.full_name AS ticketholder
  FROM sporting_event_info e,
       parquet_sporting_event_ticket t,
       parquet_person p
 WHERE
       t.sporting_event_id = e.event_id
  AND t.ticketholder_id = p.id
```

The screenshot shows the AWS Athena console interface. At the top, there's a tab labeled "sporting\_event\_i..." and a button "New query 2" with a delete icon. Below the tabs is a large text area containing the SQL query with line numbers 1 through 20. At the bottom of the text area are three buttons: "Run query" (highlighted in blue), "Save as", and "Create".

```
1 SELECT t.id AS ticket_id,
2       e.event_id,
3       e.sport,
4       e.event_date_time,
5       e.home_team,
6       e.away_team,
7       e.location,
8       e.city,
9       t.seat_level,
10      t.seat_section,
11      t.seat_row,
12      t.seat,
13      t.ticket_price,
14      p.full_name AS ticketholder
15  FROM sporting_event_info e,
16      parquet_sporting_event_ticket t,
17      parquet_person p
18 WHERE
19      t.sporting_event_id = e.event_id
20      AND t.ticketholder_id = p.id
```

Click on **Save as** button Give this query a name:  
**create\_view\_sporting\_event\_ticket\_info** and some description and then, click on **Save**.

### Lab 3. Consuming data with Athena and Quicksight

Choose a name

Name  Use 1 - 128 characters

Description  Use upto 1024 characters

Back to the query editor, you will see the query name changed. Now, click on **Run Query**.

**sporting\_event\_i...** **create\_view\_spor... ✖** +

```
1 SELECT t.id AS ticket_id,
2     e.event_id,
3     e.sport,
4     e.event_date_time,
5     e.home_team,
6     e.away_team,
7     e.location,
8     e.city,
9     t.seat_level,
10    t.seat_section,
11    t.seat_row,
12    t.seat,
13    t.ticket_price,
14    p.full_name AS ticketholder
15 FROM sporting_event_info e,
16     parquet_sporting_event_ticket t,
17     parquet_person p
18 WHERE
19     t.sporting_event_id = e.event_id
20     AND t.ticketholder_id = p.id
```

**Run query** **Save as** **Create ▾**

The results appear beneath the query window.

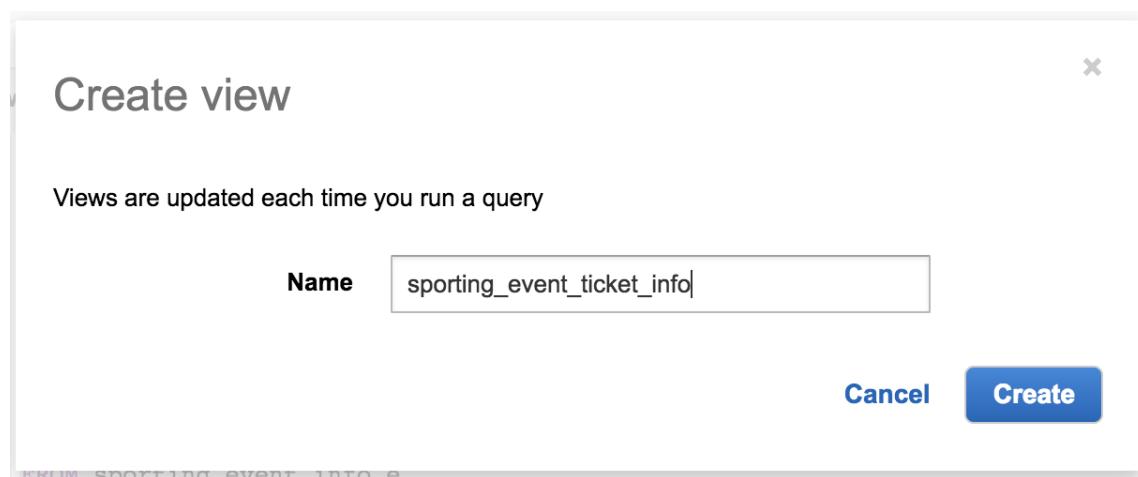
## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Query Editor interface. A query has been run, and the results are displayed in a table. The columns are: ticket\_id, event\_id, sport, event\_date\_time, home\_team, away\_team, location, city, seat\_level, seat\_section, seat\_row, seat, ticket\_price, ticketholder. The results show 10 rows of data from a football game between Arizona Cardinals and Kansas City Chiefs at the University of Phoenix Stadium in Glendale, Arizona, on November 18, 2020.

	ticket_id	event_id	sport	event_date_time	home_team	away_team	location	city	seat_level	seat_section	seat_row	seat	ticket_price	ticketholder
1	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	B	2	48.00	Hoses Sumer
2	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	B	3	48.00	Hoses Sumer
3	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	A	3	48.00	Hoses Sumer
4	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	A	4	48.00	Hoses Sumer
5	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	A	1	48.00	Hoses Sumer
6	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	17	A	2	48.00	Hoses Sumer
7	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	16	B	2	48.00	Hoses Sumer
8	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	16	B	1	48.00	Hoses Sumer
9	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	16	B	3	48.00	Hoses Sumer
10	7605981.0	10821	football	2020-11-18 18:00:00.000	Arizona Cardinals	Kansas City Chiefs	University of Phoenix Stadium	Gilbert, Arizona	3	16	B	4	48.00	Hoses Sumer

13. As shown above Click **Create view from query**.

14. Name the view "sporting\_event\_ticket\_info" and click **Create**.



15. Copy the following SQL syntax into the New Query 3 tab.

```
SELECT
    sport,
    count(distinct location) as locations,
    count(distinct event_id) as events,
    count(*) as tickets,
    avg(ticket_price) as avg_ticket_price
FROM sporting_event_ticket_info
GROUP BY 1
ORDER BY 1;
```

Click on **Save Query** and give this query name: **analytics\_sporting\_event\_ticket\_info** and some description and then, click on **Save**.

## Lab 3. Consuming data with Athena and Quicksight

Choose a name

Name: analytics\_sporting\_event\_ticket\_info  
Use 1 - 128 characters

Description: to analyze the view: sporting\_event\_ticket\_info  
Use upto 1024 characters

Cancel Save

The name of the New Query 3 will be changed to one assigned in previous step. Click on **Run Query**.

```
1 SELECT
2   sport,
3   count(distinct location) as locations,
4   count(distinct event_id) as events,
5   count(*) as tickets,
6   avg(ticket_price) as avg_ticket_price
7 FROM sporting_event_ticket_info
8 GROUP BY 1
9 ORDER BY 1;
10
```

Run query Save as Create ▾

You query returns two results in approximately five seconds. The query scans 25 MB of data, which prior to converting to parquet, would have been 1.59GB of CSV files.

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Query Editor interface. At the top, there are tabs for 'Athena', 'Query Editor' (which is selected), 'Saved Queries', 'History', and 'AWS Glue Data Catalog'. The 'Workgroup' is set to 'primary'. Below the tabs, there are three tabs for 'New query 1', 'New query 2', and 'New query 3', with 'New query 3' currently active. The query editor contains the following SQL code:

```
1 SELECT
2   sport,
3   count(distinct location) as locations,
4   count(distinct event_id) as events,
5   count(*) as tickets,
6   avg(ticket_price) as avg_ticket_price
7   FROM sporting_event_ticket_info
8   GROUP BY 1
9   ORDER BY 1;
```

Below the code, there are buttons for 'Run query', 'Save as', and 'Create'. A status message indicates '(Run time: 6.21 seconds, Data scanned: 25.97 MB)'. A note says 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'. The results section shows a table with the following data:

	sport	locations	events	tickets	avg_ticket_price
1	baseball	30	294	958680	53.89345581425812
2	football	25	113	810304	57.40977502271104

The purpose of saving the queries is to have clear distinction between the results of the queries running on one view. Otherwise, your query results will be saved under “Unsaved” folder within the S3 bucket location provided to Athena to store query results. Please navigate to S3 bucket to observe these changes, as shown below:

The screenshot shows the AWS S3 console. The path is 'Amazon S3 > dmslab-student-dmslabs3bucket-xg1hdyq60lbs > athenaquery'. The 'Overview' tab is selected. A search bar at the top has the placeholder 'Type a prefix and press Enter to search. Press ESC to clear.' Below the search bar are buttons for 'Upload', '+ Create folder', 'Download', and 'Actions'. The location 'US East (N. Virginia)' is shown. The main area displays a table of objects:

Name	Last modified	Size	Storage class
analytics_sporting_event_ticket_info	--	--	--
create_view_sporting_event_ticket_info	--	--	--

A note at the bottom right says 'Viewing 1 to 2'.

## Build an Amazon QuickSight Dashboard

### Set up QuickSight

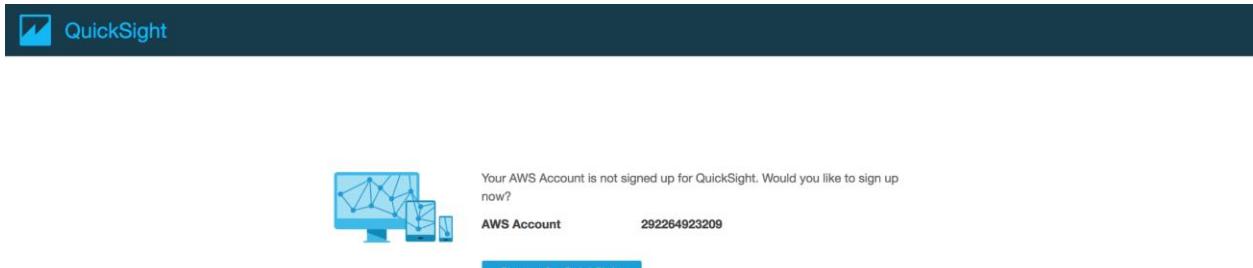
1. In the AWS services console, search for **QuickSight**.

The screenshot shows the AWS services search results. The search bar at the top contains 'QuickSight'. Below the search bar, 'QuickSight' is listed with the description 'Fast, easy to use business analytics'. Other services listed include 'Athena', 'S3', 'AWS Glue', 'IAM', and 'CloudWatch'. At the bottom, there is a link 'All services'.

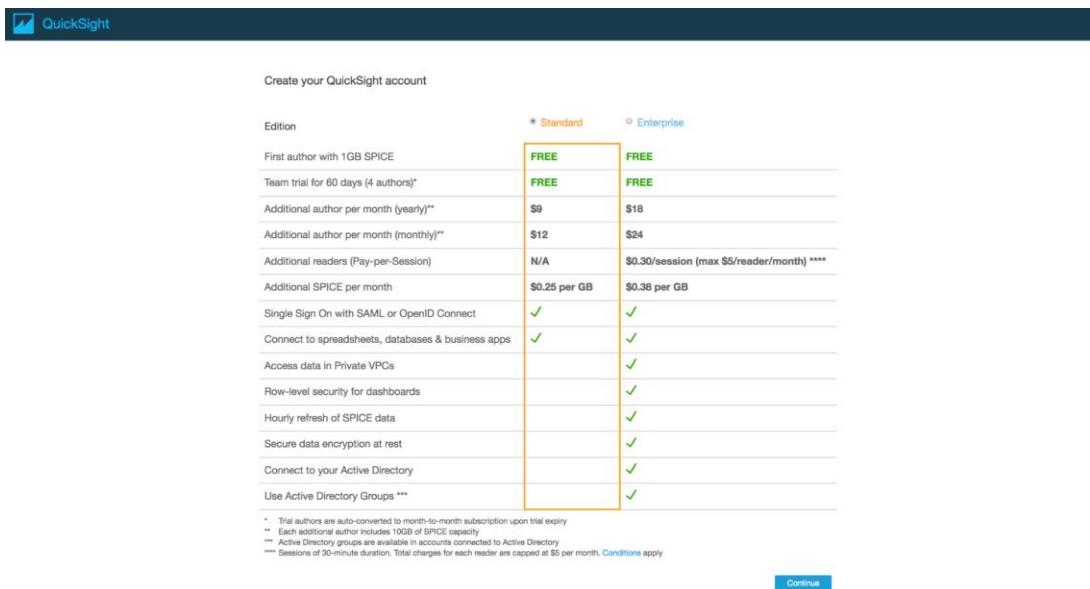
## Lab 3. Consuming data with Athena and Quicksight

If this is the first time you have used QuickSight, you are prompted to create an account.

2. Click **Sign up for QuickSight**.



3. For account type, choose **Standard**. If you plan to complete the Bonus Exercise, please choose **Enterprise** Version
4. Click **Continue**.



5. On the Create your QuickSight account page, fill out your name and email address.
6. Select region and the check boxes to enable auto discovery, Amazon Athena, and Amazon S3.
7. Click **Choose S3 buckets** and select your DMS bucket (e.g., "dmslab-student-").
8. Click **Finish**.

## Lab 3. Consuming data with Athena and Quicksight

Create your QuickSight account

Edition **Standard**

QuickSight account name **Glue-Lab-George**

Notification email address **julbrigh+dataenglab@amazon.com**

QuickSight capacity region **US East (N. Virginia)**

Enable autodiscovery of data and users in your Amazon Redshift, Amazon RDS and AWS IAM services.

Amazon Athena  
Enables QuickSight access to Amazon Athena databases

Please ensure the right Amazon S3 buckets are also enabled for QuickSight.

Amazon S3 (1 bucket)  
Enables QuickSight to auto-discover your Amazon S3 buckets

Amazon S3 Storage Analytics  
Enables QuickSight to visualize your S3 Storage Analytics data

Amazon IoT Analytics  
Enable QuickSight to visualize your IoT Analytics data

**Choose S3 buckets**

**Finish**

9. On the **QuickSight landing page**, on the **top right corner**, click on “**Manage QuickSight**”.

Search for analyses, data sets, and dashboards

New analysis

All analyses All dashboards Favorites Tutorial videos

Last updated

Manage QuickSight

Community Send feedback What's new

10. Choose “Security and Permissions” and under “QuickSight access to AWS Services, click on “Add or Remove” button.

Manage users

Your subscriptions

SPICE capacity

Account settings

**Security & permissions**

Manage VPC connections

Mobile settings

Domains and Embedding

QuickSight access to AWS services

Amazon Redshift Amazon RDS IAM Amazon S3 Amazon Athena Amazon S3 Storage Analytics

AWS IoT Analytics

Add or remove

## Lab 3. Consuming data with Athena and Quicksight

11. If you will observe there is an unchecked box against S3 buckets for “dmslab-student-“, please check the box.



**QuickSight access to AWS services**

QuickSight can connect to the selected AWS products & services below for all users & groups:

	Amazon Redshift Enables QuickSight to auto-discover clusters	<input checked="" type="checkbox"/>
	Amazon RDS Enables QuickSight to auto-discover instances	<input checked="" type="checkbox"/>
	IAM Enables you to invite IAM users from this AWS Account to access QuickSight	<input checked="" type="checkbox"/>
	Amazon S3 Enables QuickSight to auto-discover your Amazon S3 buckets <a href="#">Details</a>	<input type="checkbox"/>
	Amazon Athena Enables QuickSight access to Amazon Athena databases	<input checked="" type="checkbox"/>
	Amazon S3 Storage Analytics Enables QuickSight to visualize your S3 Storage Analytics data	<input checked="" type="checkbox"/>
	AWS IoT Analytics Enables QuickSight to visualize your IoT Analytics data	<input checked="" type="checkbox"/>
	Amazon SageMaker	<input type="checkbox"/>

12. Select Amazon S3. Select the S3 bucket created as part of student lab (for e.g: dmslab-student-dmslabs3bucket-xg1hdyq60ibs) which will have all the folders for your source data.

<input checked="" type="checkbox"/>	dmslab-student-dmslabs3bucket-xg1hdyq60ibs	<input type="checkbox"/>
-------------------------------------	--	--------------------------

13. Then, click on Finish.



14. You will observe that now there is a check mark in the checkbox for Amazon S3. This confirms that QuickSight has required permissions. Then, click on “Update”.

## Lab 3. Consuming data with Athena and Quicksight

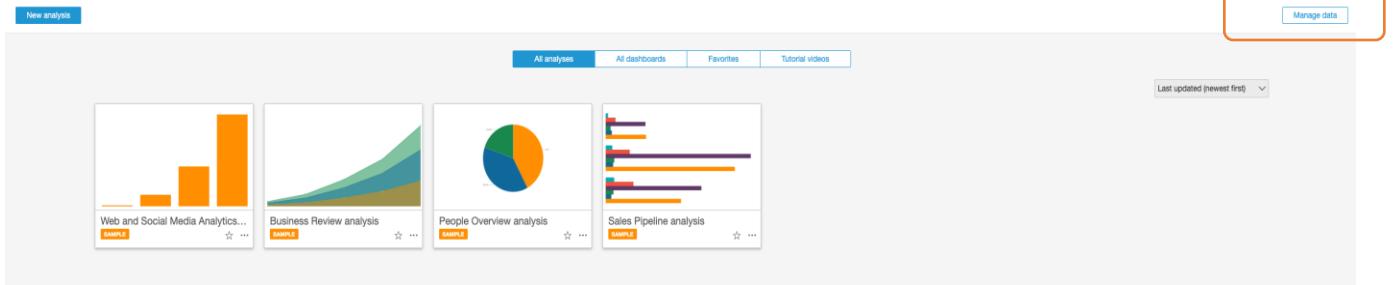
**QuickSight access to AWS services**

QuickSight can connect to the selected AWS products & services below for all users & groups:

	Amazon Redshift Enables QuickSight to auto-discover clusters	<input checked="" type="checkbox"/>
	Amazon RDS Enables QuickSight to auto-discover instances	<input checked="" type="checkbox"/>
	IAM Enables you to invite IAM users from this AWS Account to access QuickSight	<input checked="" type="checkbox"/>
	Amazon S3 Enables QuickSight to auto-discover your Amazon S3 buckets <a href="#">Details</a>	<input checked="" type="checkbox"/>
	Amazon Athena Enables QuickSight access to Amazon Athena databases	<input checked="" type="checkbox"/>
	Amazon S3 Storage Analytics Enables QuickSight to visualize your S3 Storage Analytics data	<input checked="" type="checkbox"/>
	AWS IoT Analytics Enables QuickSight to visualize your IoT Analytics data	<input checked="" type="checkbox"/>
	Amazon SageMaker Enables QuickSight to infer fields from custom ML models	<input type="checkbox"/>

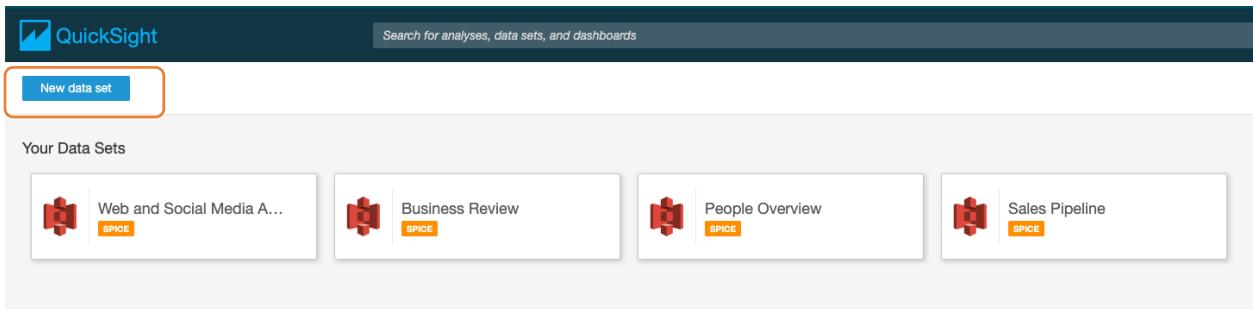
**Cancel** **Update**

15. Navigate to QuickSight landing page by clicking on the QuickSight logo on the top left.  
On the top right corner, click **Manage Data**.



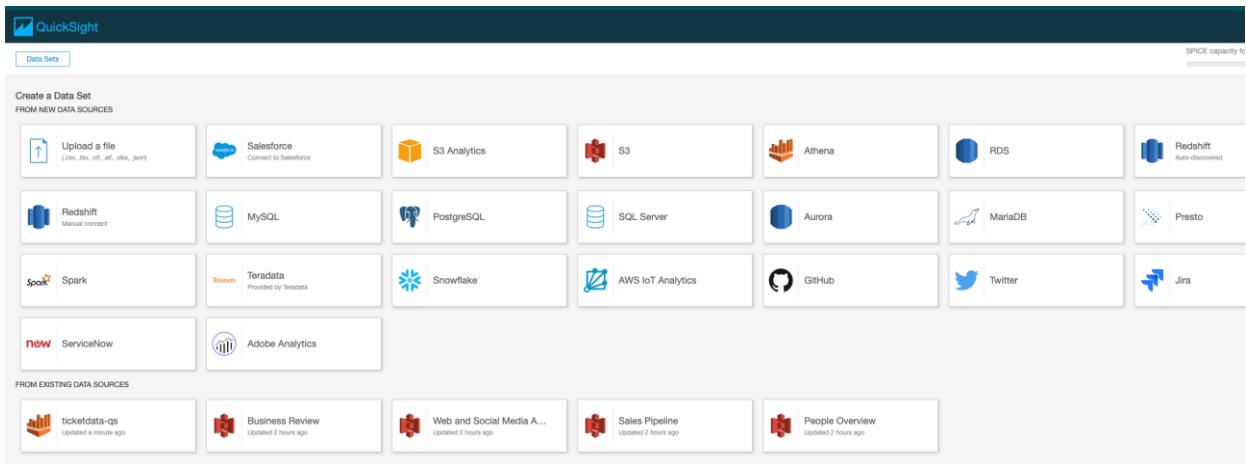
16. Click **New Data Set**.

## Lab 3. Consuming data with Athena and Quicksight



The screenshot shows the QuickSight interface with the title bar "QuickSight" and a search bar "Search for analyses, data sets, and dashboards". Below the search bar, there is a button labeled "New data set" which is highlighted with an orange border. The main area is titled "Your Data Sets" and displays four data sets: "Web and Social Media A..." (SPICE), "Business Review" (SPICE), "People Overview" (SPICE), and "Sales Pipeline" (SPICE). Each data set is represented by a red cube icon and its name followed by a small orange "SPICE" badge.

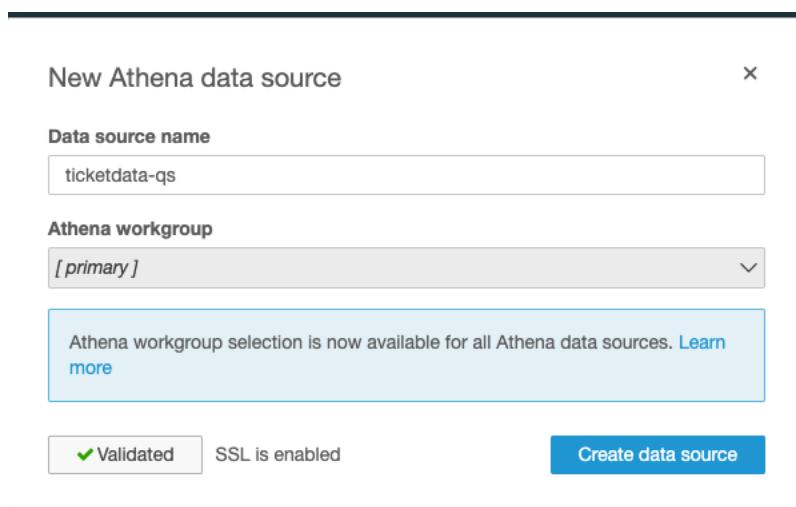
17. On the Create a Data Set page, select **Athena** as the data source.



The screenshot shows the "Create a Data Set" page in the QuickSight console. The top navigation bar has a "Data Sets" tab selected. The main area is titled "Create a Data Set" and "FROM NEW DATA SOURCES". It lists various data sources in a grid: "Upload a file" (CSV, JSON, XML, AVRO, Parquet), "Salesforce" (Connect to Salesforce), "S3 Analytics", "S3", "Athena" (highlighted with a blue selection bar), "RDS", "Redshift" (Auto-discovered), "Redshift" (Manual connect), "MySQL", "PostgreSQL", "SQL Server", "Aurora", "MariaDB", "Presto", "Spark", "Teradata" (Provided by Redshift), "Snowflake", "AWS IoT Analytics", "GitHub", "Twitter", "Jira", "ServiceNow", and "Adobe Analytics". Below this section, there is another titled "FROM EXISTING DATA SOURCES" which lists five existing data sets: "ticketdata-qS" (Updated a minute ago), "Business Review" (Updated 2 hours ago), "Web and Social Media A..." (Updated 2 hours ago), "Sales Pipeline" (Updated 2 hours ago), and "People Overview" (Updated 2 hours ago).

18. For Data source name, type “ticketdata-qS” and click **Validate connection**.

19. Click **Create data source**.

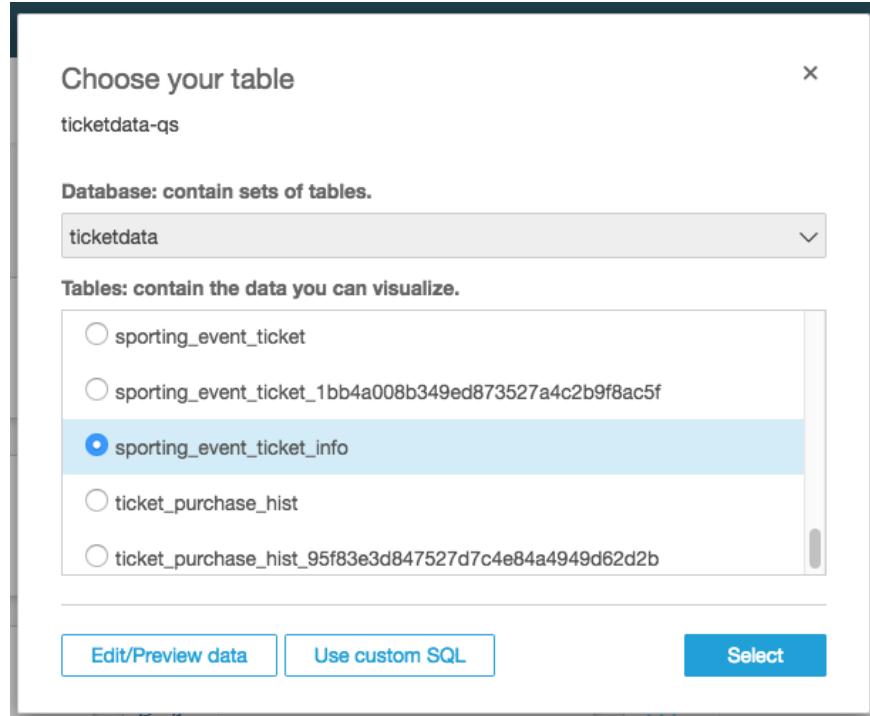


The screenshot shows the "New Athena data source" dialog box. It has a header "New Athena data source" and a close button "x". The "Data source name" field contains "ticketdata-qS". The "Athena workgroup" dropdown is set to "[ primary ]". A message box at the bottom left says "Athena workgroup selection is now available for all Athena data sources. [Learn more](#)". At the bottom right, there are two buttons: a "Validated" button with a green checkmark and an "SSL is enabled" button, and a "Create data source" button in a blue box.

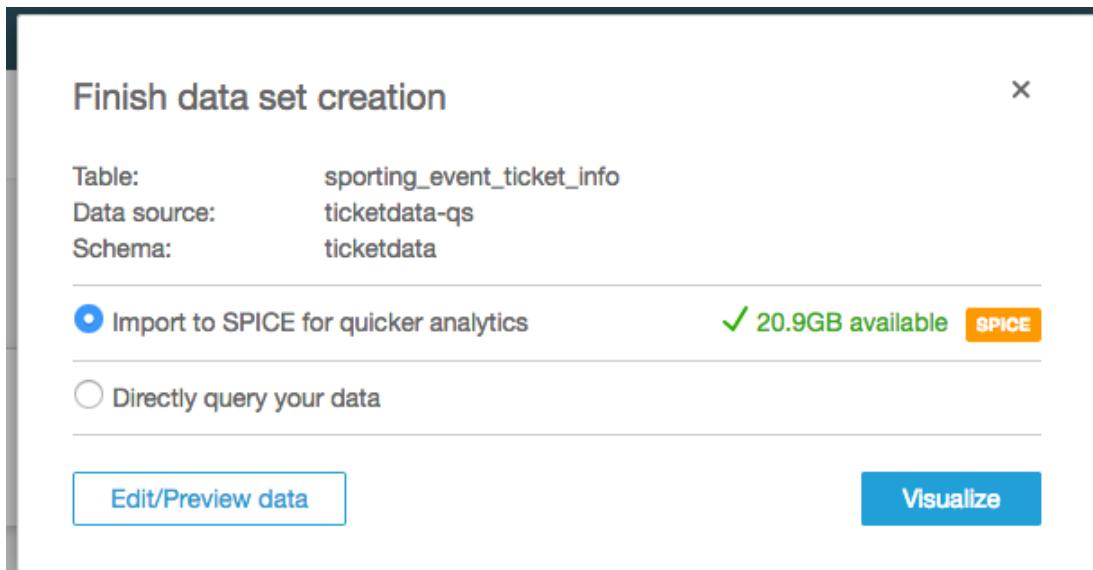
20. In the Database drop-down list, select the database name you created in the AWS Glue lab.

21. Choose the "sporting\_event\_ticket\_info" table and click **Select**.

### Lab 3. Consuming data with Athena and Quicksight

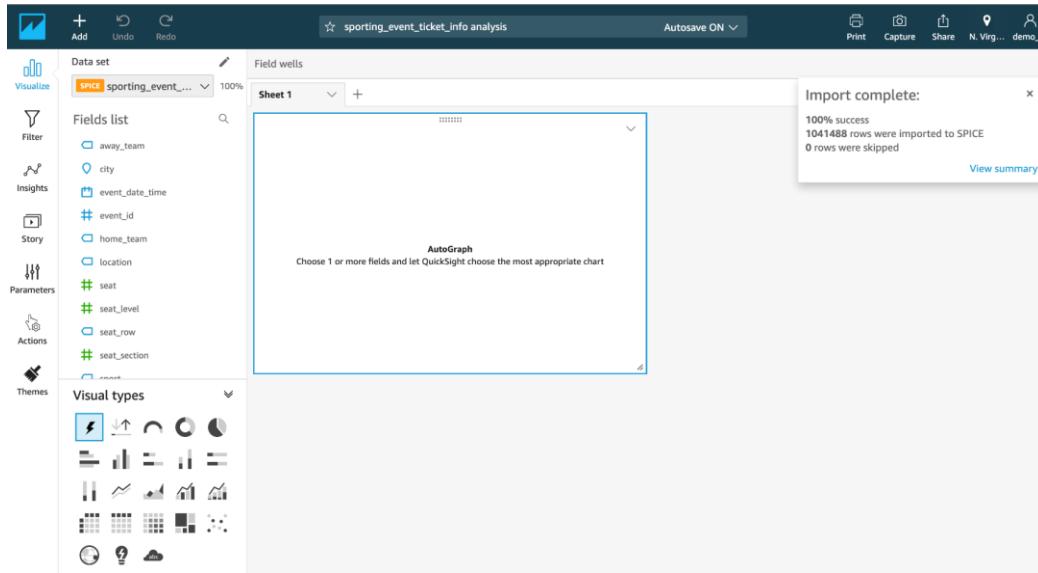


22. To finish data set creation, choose the option **Import to SPICE for quicker analytics** and click **Visualize**.



You will now be taken to the QuickSight Visualize interface where you can start building your dashboard.

## Lab 3. Consuming data with Athena and Quicksight

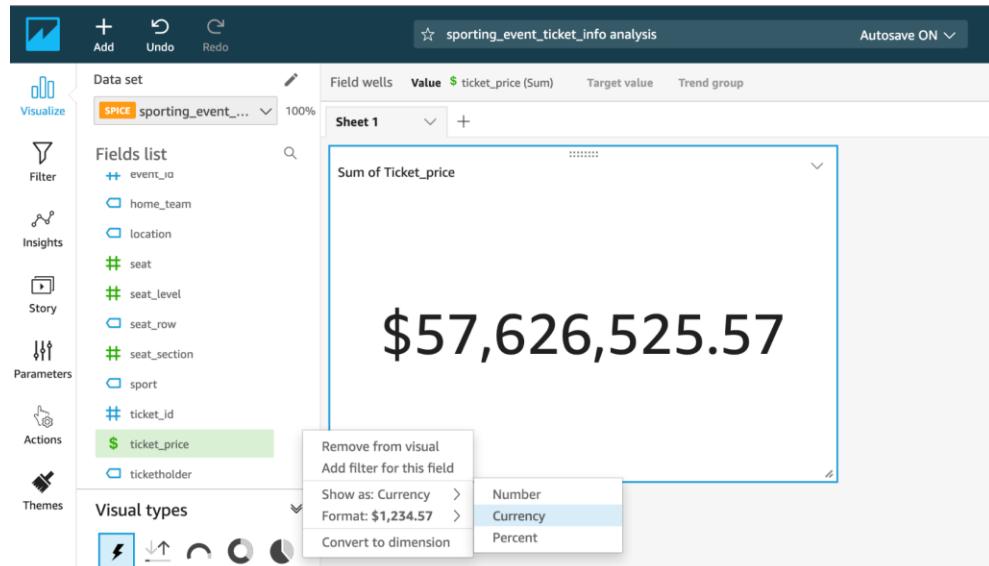


**Note:** The SPICE dataset will take a few minutes to be built, but you can continue to create some charts on the underlying data.

### Create QuickSight Charts

In this section we will take you through some of the different chart types.

1. In the Fields list, click the "**ticket\_price**" column to populate the chart.
2. Click the **expand icon** in corner of "ticket\_price" field and select format as **currency** to show numbers in dollar amount.



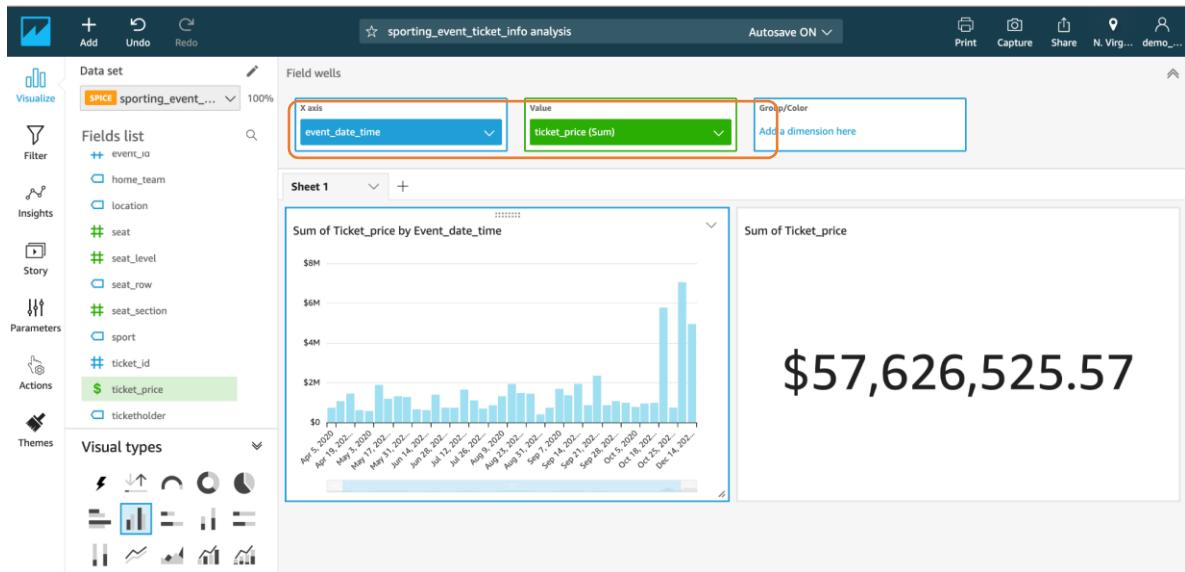
## Lab 3. Consuming data with Athena and Quicksight

3. You can add new visual and keep building your dashboard by clicking **Add button** at top left corner of screen.

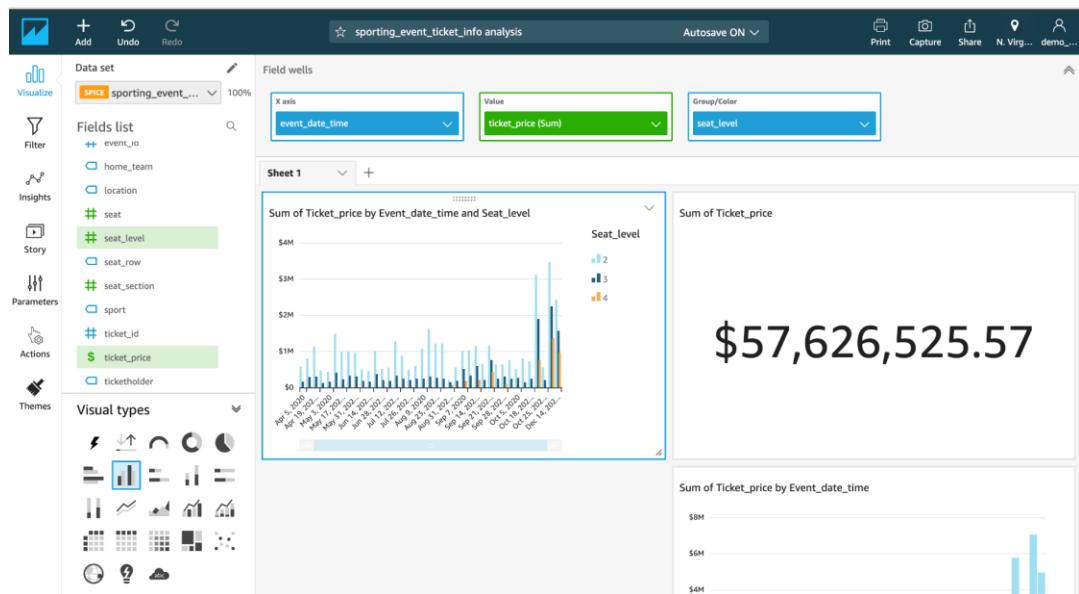
In the **Visual types** area, choose the **Vertical bar chart** icon.

This layout requires a value for the X-axis. In Fields list, select the "**event\_date\_time**" field and you should see the visualization update.

For Y-axis, select "**ticket\_price**" from the Field list.



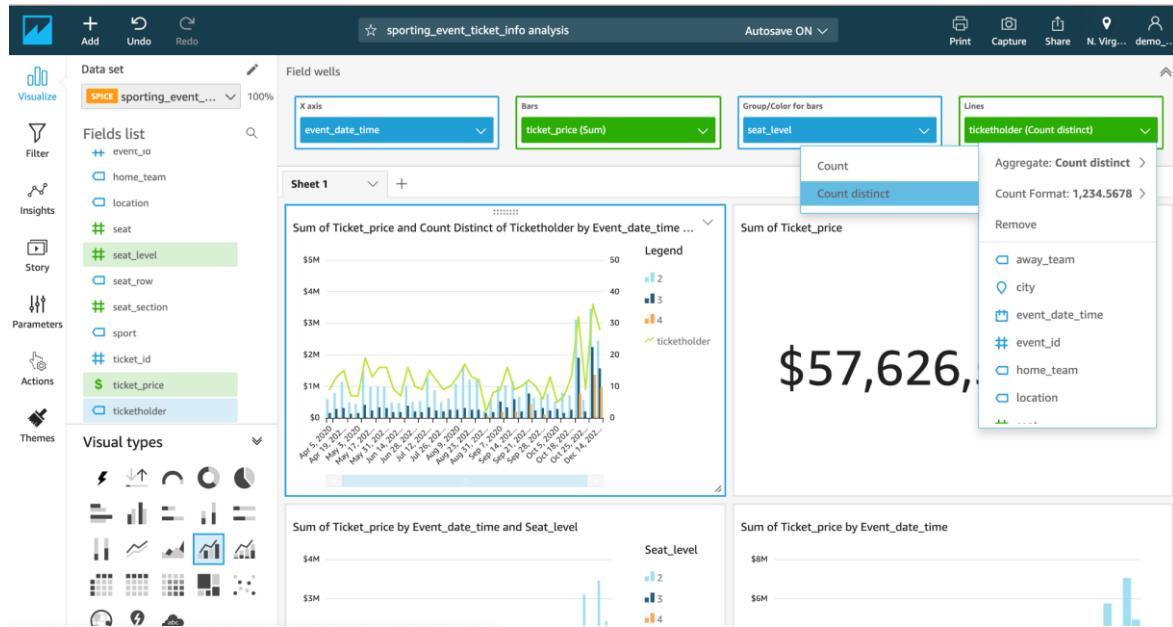
4. Add new Visual and you can drag and move other visuals to adjust space in dashboard. In the Fields list, click and drag the **seat\_level** field to the **Group/Color box** in the **Field wells pane**. You can also use the slider below the x axis to fit all of the data.



## Lab 3. Consuming data with Athena and Quicksight

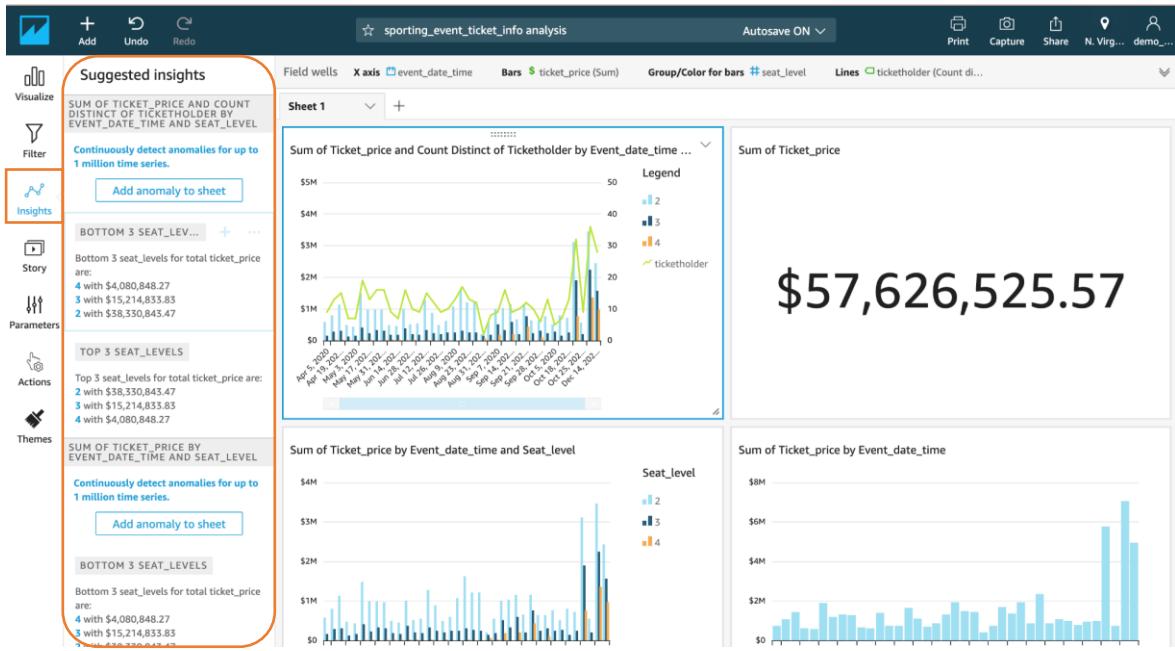
Let's build on this one step further by changing the chart type to "Clustered bar combo chart" and adding in the **ticketholder** field for the **Lines**.

5. In the Visual types area, choose the Clustered bar combo chart icon.
6. In the Fields list, click and drag the **ticketholder** field to the Lines box in the Field wells pane.
7. In the Field wells pane, click the Lines box and choose **Count Distinct** for Aggregate. You can then see the y-axis update on the right-hand side.



8. Click on **insight** icon on the left tabs section and explore insight information in simple English.

## Lab 3. Consuming data with Athena and Quicksight

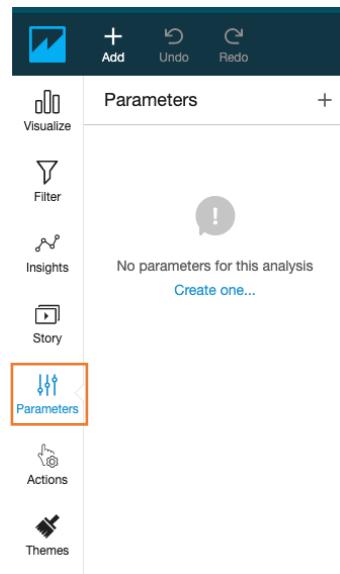


Feel free to experiment with other chart types and different fields to get a sense of the data.

## Create QuickSight Parameters

In the next section we are going to create some parameters with controls for the dashboard, then assign these to a filter for all the visuals.

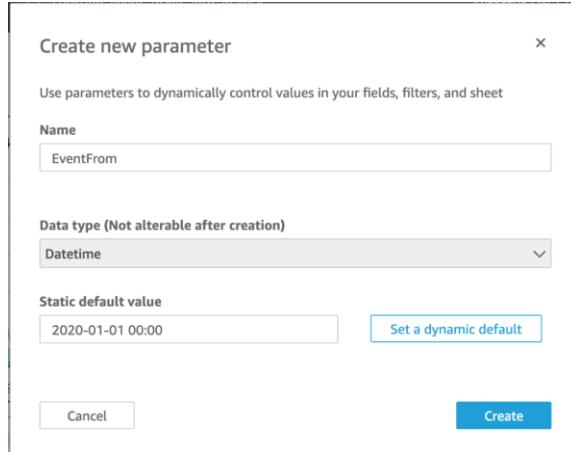
1. In the left navigation menu, select **Parameters**.



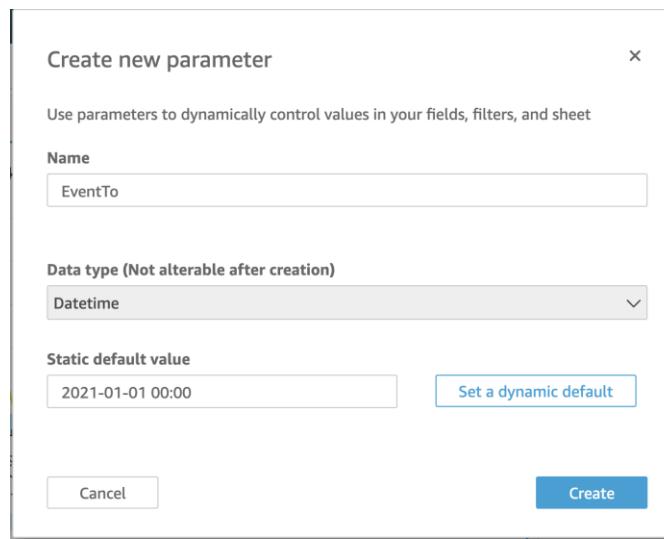
2. Click **Create one** to create a new parameter with a Name.
3. For Name, type **EventFrom**.
4. For Data type, choose **Datetime**.

## Lab 3. Consuming data with Athena and Quicksight

5. For Default value, select the value from calendar as start date available in your graph for **event\_date\_time**. For example, **2020-01-01 00:00**.
6. Click **Create**, and then close the Parameter Added dialog box.

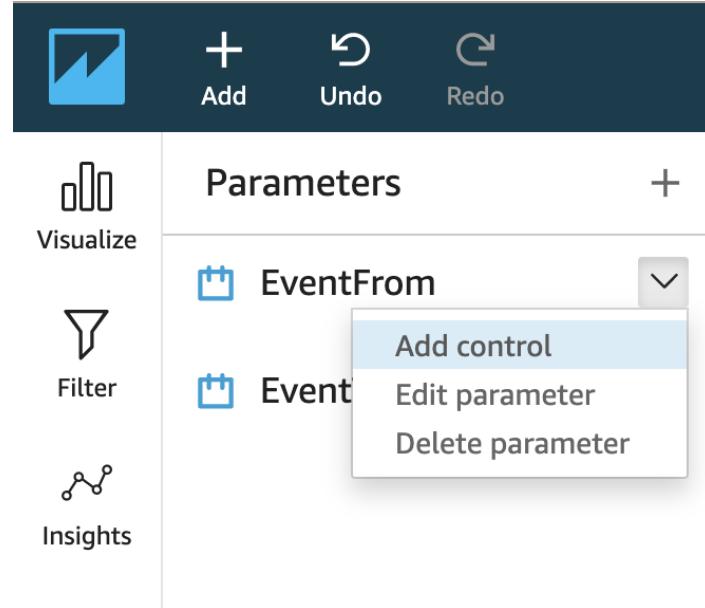


7. Create another parameter with the following attributes:
  - a. **Name: EventTo**
  - b. **Data type: Datetime**
  - c. For Default value, select the value from calendar as end date available in your graph for **event\_date\_time**. For example, **2021-01-01 00:00**
  - d. Click **Create**



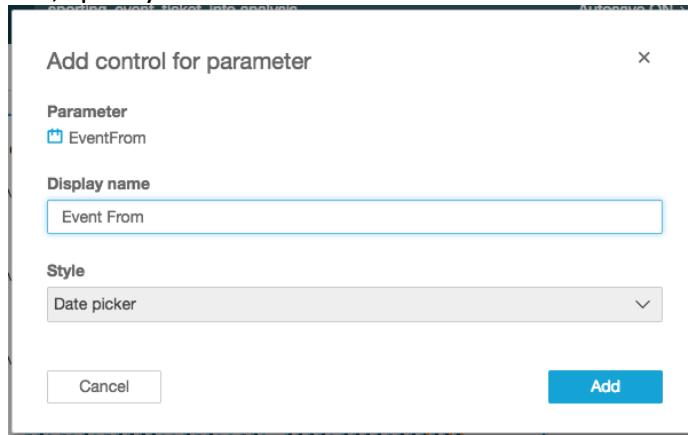
8. In next window, you can select any option to perform any operation with the parameter. Alternatively, you can click the drop-down menu for the **EventFrom** parameter and choose **Add control**.

## Lab 3. Consuming data with Athena and Quicksight



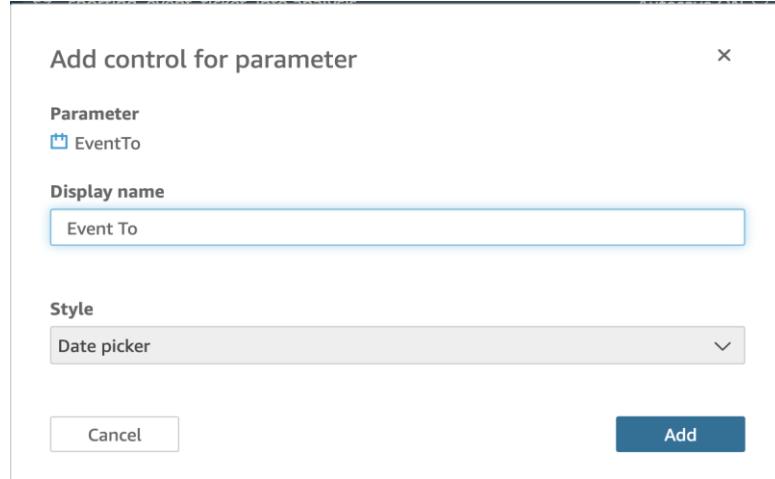
9. Click on **Control**

10. For Display name, specify **Event From** and click **Add**.

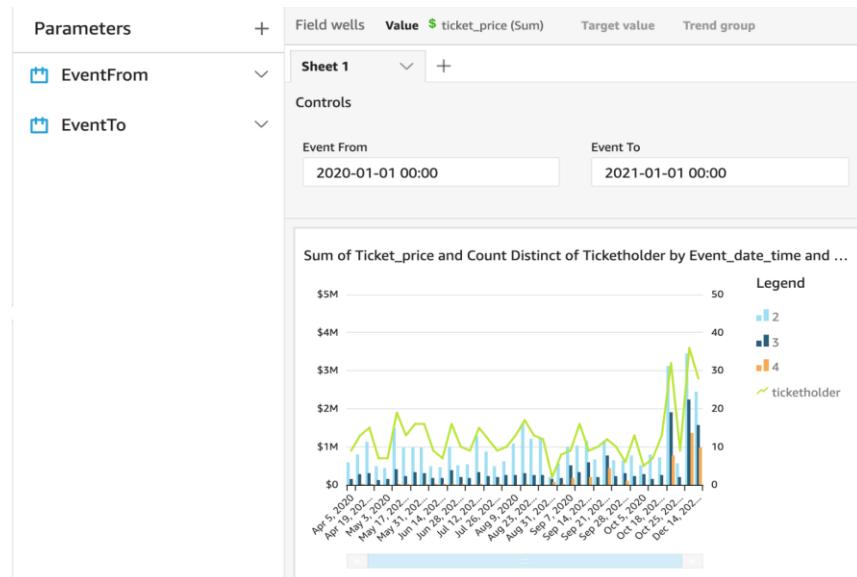


11. Repeat the process to add a control for **EventTo** with display name **Event To**.

## Lab 3. Consuming data with Athena and Quicksight



You should now be able to see and expand the Controls section above the chart.

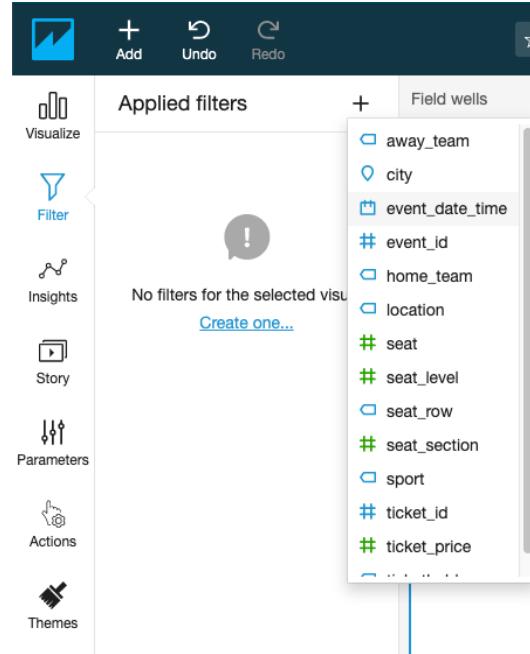


### Create a QuickSight Filter

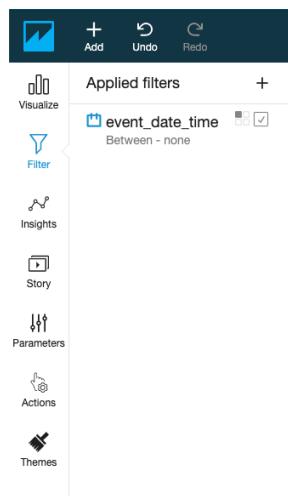
To complete the process, we will wire up a filter to these controls for all visuals.

1. In the left navigation menu, choose **Filter**.
2. Click the plus icon (+) to add a filter for the field "**event\_date\_time**".

## Lab 3. Consuming data with Athena and Quicksight

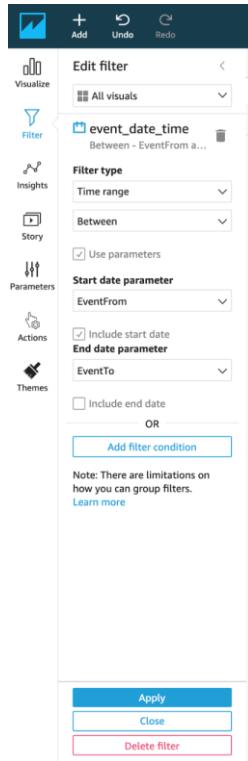


3. Click this filter to edit the properties.



4. Choose to make this filter apply to **All visuals**.
5. For Filter type, choose **Time range** and **Between**.
6. Select option **Use Parameter**.
7. For Start date parameter, choose **EventFrom**.
8. For End date parameter, choose **EventTo**.
9. Click **Apply**.

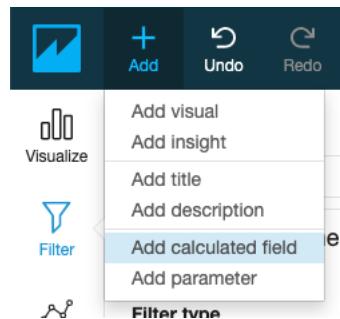
## Lab 3. Consuming data with Athena and Quicksight



## Add Calculated Fields

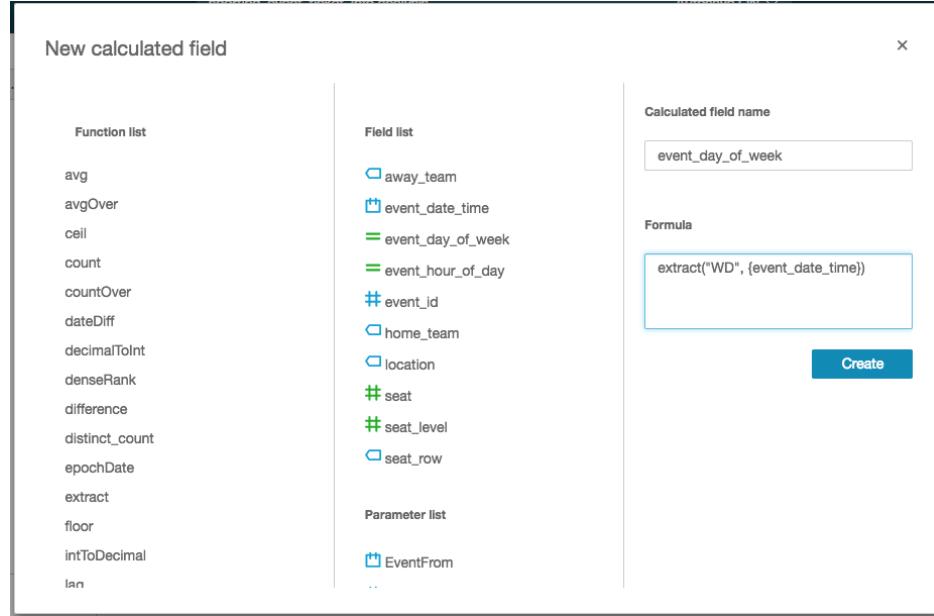
In the next section, you will learn, how to add calculated fields for "day of week" and "hour of day" to your dataset and a new scatter plot for these two dependent variables.

1. Click the Add button on the top left and select **Add a calculated field**.



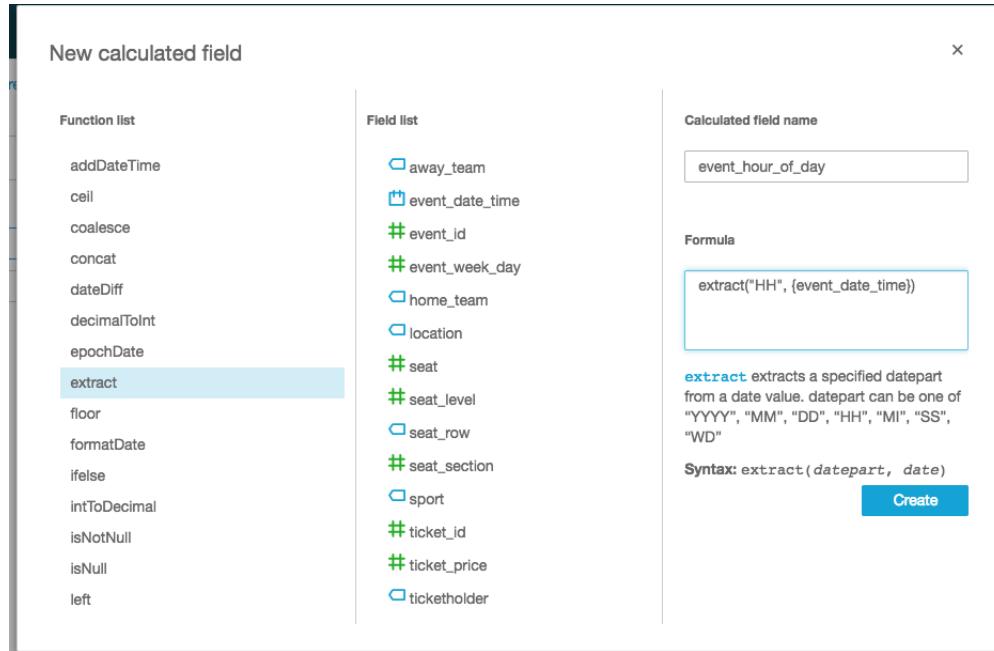
2. For **Calculated field name** type "**event\_day\_of\_week**".
3. For **Formula**, type `extract("WD",{event_date_time})`  
Note: extract returns a specified portion of a date value. Requesting a time-related portion of a date that doesn't contain time information returns 0. WD: This returns the day of the week as an integer, with Sunday as 1.
4. Click **Create**.

## Lab 3. Consuming data with Athena and Quicksight



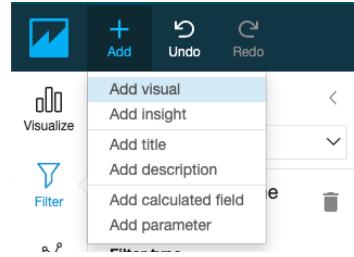
5. Add another calculated field with the following attributes:
  - a. Calculated field name: "**event\_hour\_of\_day**"
  - b. Formula: **extract("HH",{event\_date\_time})**

Note: HH: This returns the hour portion of the date.

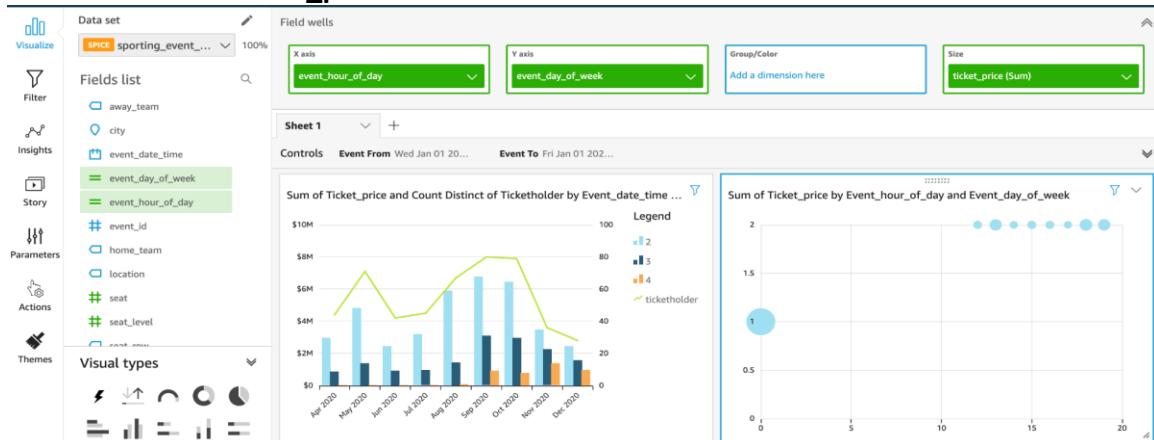


6. Click Add button in the top left and choose **Add visual**.

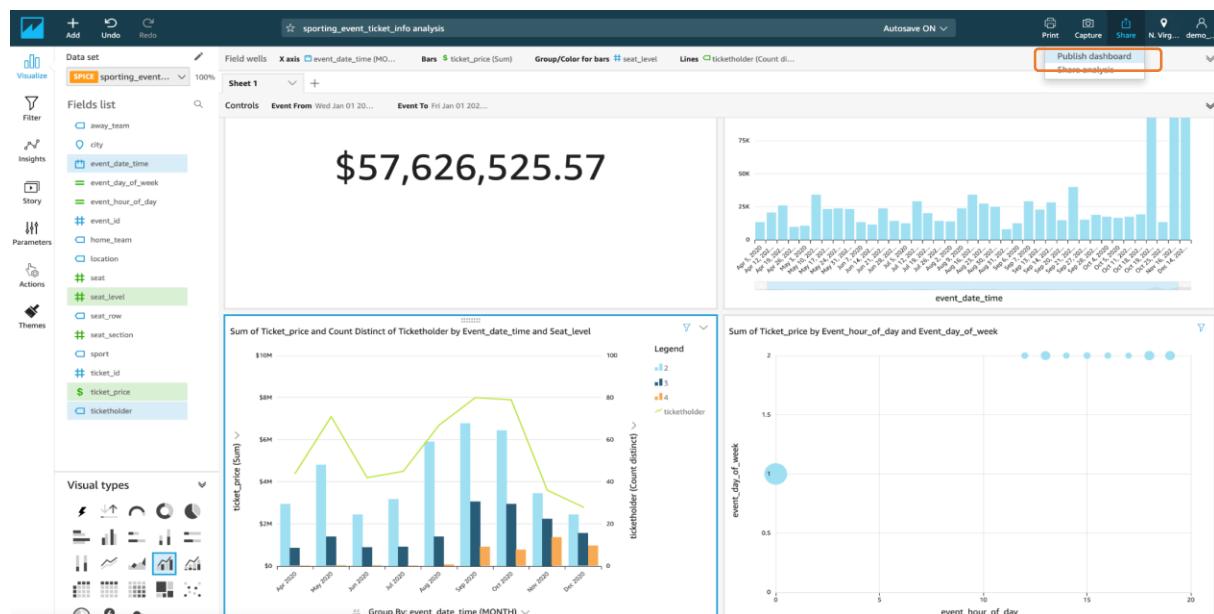
## Lab 3. Consuming data with Athena and Quicksight



7. For field type, select the **scatter plot**.
8. In the Fields list, select and drag the following attributes to the Field wells pane to set the graph attributes:
  - a. **X-axis: "event\_hour\_of\_day"**
  - b. **Y-axis: "event\_day\_of\_week"**
  - c. **Size: "ticket\_price"**



Since now you have completed your dashboard then you can **publish** it by clicking on top right corner of screen.



## Lab 3. Consuming data with Athena and Quicksight

A *dashboard* is a read-only snapshot of an analysis that you can share with other Amazon QuickSight users for reporting purposes. In Dashboard other users can still play with visuals and data but that will not modify dataset.

You can share an analysis with one or more other users with whom you want to collaborate on creating visuals. Analysis provide other uses to write and modify data set.

### Amazon QuickSight ML-Insights (Optional)

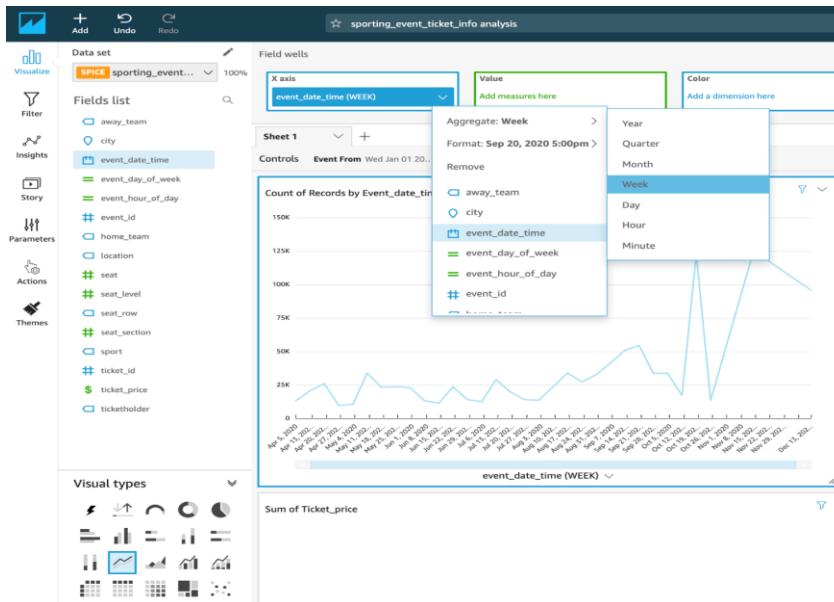
With Amazon QuickSight, you can add Machine Learning capabilities to your visuals, easily, with one click action. There are 3 types of Machine Learning Insights

- Narrative
- Anomaly Detection
- Forecasting

ML-Insights is only available to enterprise version of QuickSight. You will need to upgrade to Enterprise Edition before you start with the task. To upgrade your Amazon QuickSight Subscription from Standard Edition to Enterprise Edition please follow this guide  
<https://docs.aws.amazon.com/quicksight/latest/user/upgrading-subscription.html>

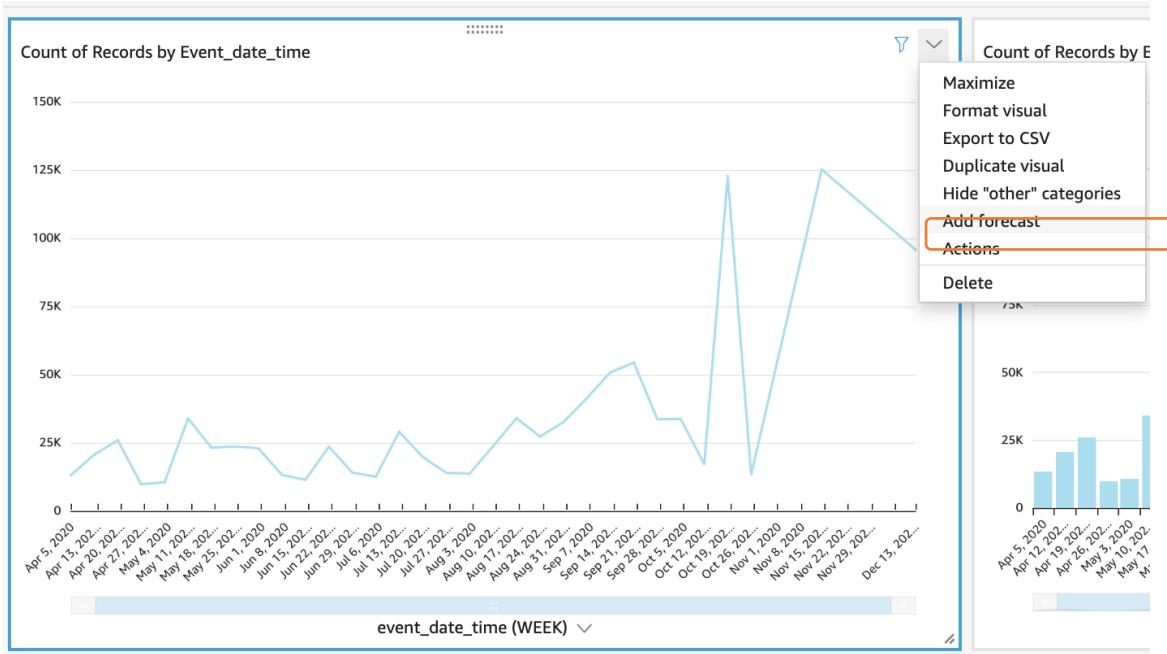
Let's see how we can add a bit of forecasting in our dashboard. Forecasting works with timeseries, which is better represented with a line graph. Let's first create a line graph.

1. Click **add Visual at top left corner of screen**, and select **Line Chart** and add the **event\_date\_time** as the **x-axis** and **aggregate by week**. As shown in below screenshot

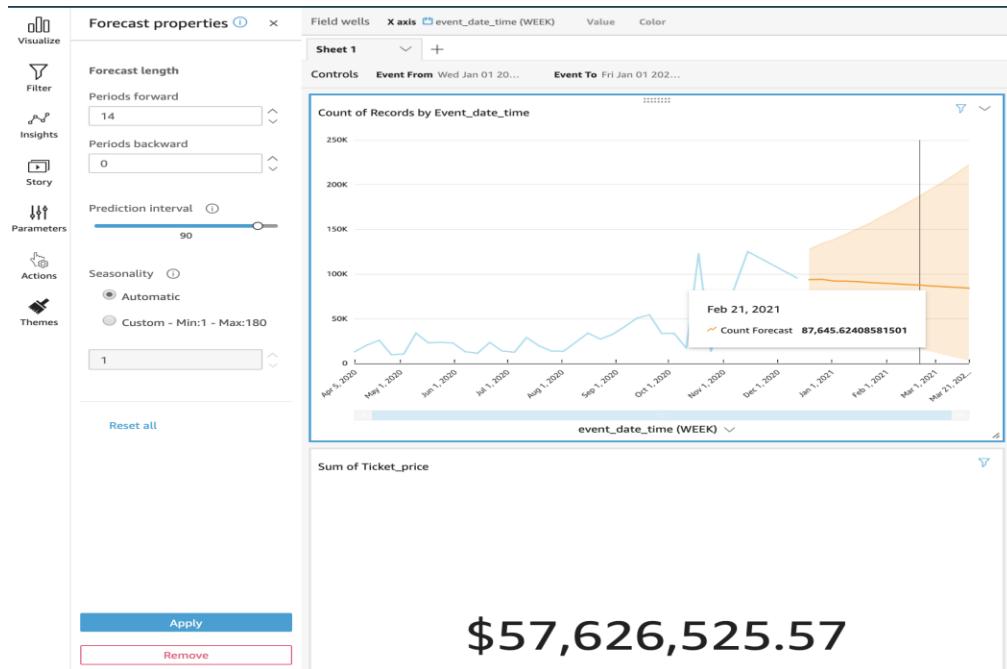


2. Add forecasting to the visual. To do that, click on the drop-down arrow on the top right corner of the visual, and then click **Add forecast**.

## Lab 3. Consuming data with Athena and Quicksight



The visual will add forecast, you can hover over and explore forecasted data as shown below. Feel free to explore with the properties of the forecast algorithm.



Congratulations!! You have successfully completed this lab, Continue to Next section if you want to dive deep into Athena query access and cost

### (Optional)Athena Workgroups to Control Query Access and Costs

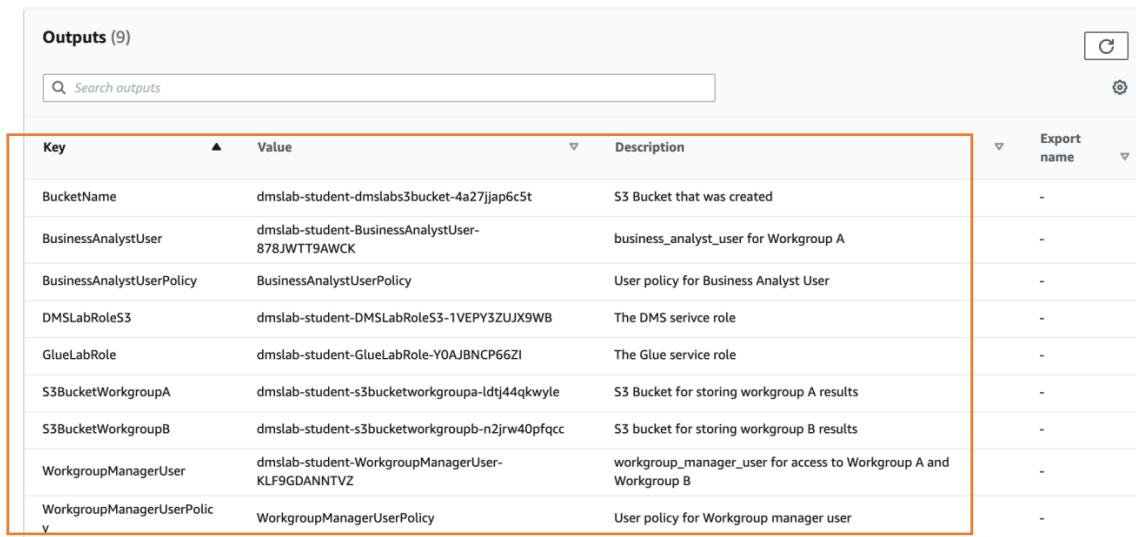
Use workgroups to separate users, teams, applications, or workloads, to set limits on amount of data each query or the entire workgroup can process, and to track costs. Because workgroups act as resources, you can use resource-level identity-based policies to control access to a specific workgroup. You can also view query-related metrics in Amazon CloudWatch, control costs by configuring limits on the amount of data scanned, create thresholds, and trigger actions, such as Amazon SNS, when these thresholds are breached.

#### Workflow setup to separate workloads

For this lab, we will create two workgroups: “workgroupA” and “workgroupB”. Before creating the workgroups, you need to have users, appropriate IAM policies to assigned to each user and S3 buckets to store the query results. This has been created using Cloud Formation template for your convenience. It is recommended to go through the template for better understanding of pre-requisites. We will have two users: “business\_analyst\_user” and “workgroup\_manager\_user” created in IAM with different policies:

- The **business\_analyst\_user** will have access to **workgroupA** and query **sporting\_event\_info** table.
- The **workgroup\_manager\_user** will have access to both workgroups **workgroupA** and **workgroupB** for management purposes.

The resources have been already created as part of the DMS lab. You can click on the **dmslab\_student** stacks on [CloudFormation](#) and navigate to the “**Resources**” to understand the different resources created by the template. Navigate to **outputs** section to see the results of resources created with description.



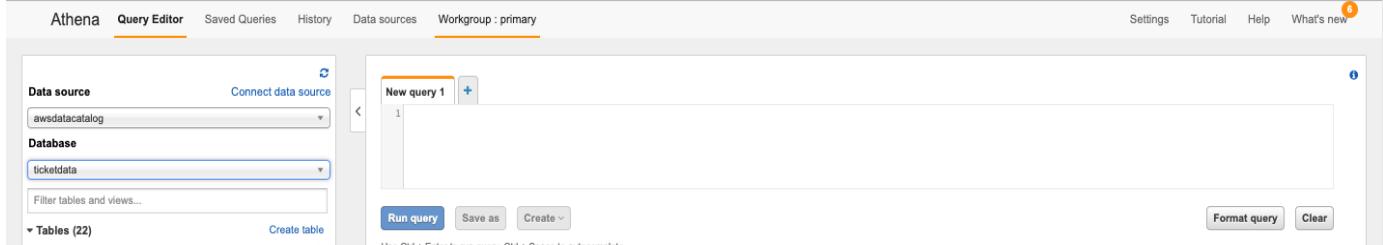
Outputs (9)			
Key	Value	Description	Export name
BucketName	dmslab-student-dmslabs3bucket-4a27jjap6c5t	S3 Bucket that was created	-
BusinessAnalystUser	dmslab-student-BusinessAnalystUser-878JWTT9AWCK	business_analyst_user for Workgroup A	-
BusinessAnalystUserPolicy	BusinessAnalystUserPolicy	User policy for Business Analyst User	-
DMSLabRoleS3	dmslab-student-DMSLabRoleS3-1VEPY3ZUJX9WB	The DMS service role	-
GlueLabRole	dmslab-student-GlueLabRole-Y0AJBNCP66ZI	The Glue service role	-
S3BucketWorkgroupA	dmslab-student-s3bucketworkgroupa-ldtj44qkwyle	S3 Bucket for storing workgroup A results	-
S3BucketWorkgroupB	dmslab-student-s3bucketworkgroupb-n2jrw40pfqcc	S3 bucket for storing workgroup B results	-
WorkgroupManagerUser	dmslab-student-WorkgroupManagerUser-KLF9GDANNNTVZ	workgroup_manager_user for access to Workgroup A and Workgroup B	-
WorkgroupManagerUserPolicy	WorkgroupManagerUserPolicy	User policy for Workgroup manager user	-

## Lab 3. Consuming data with Athena and Quicksight

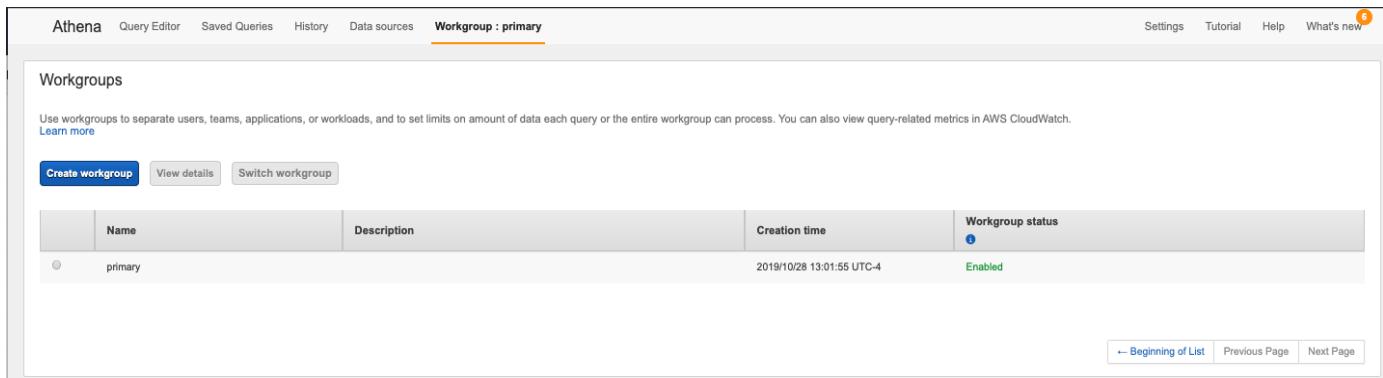
We will utilize the values from the outputs wherever required in the following steps.

### Now we will create workgroups.

1. Navigate to **Athena Console** and click on “**Workgroup: primary**”. The default workgroup provided for querying in Athena is “primary”.



2. Click on “**Create workgroup**”



3. Provide the following:

- a. Workgroup **Name**: “workgroupA”
- b. **Description**: (optional):
  - “workgroupA for BusinessAnalystUser”
  - “workgroupB for workgroup manager user”
- c. **Query result location**: Provide the query location, You can find S3 bucket name from Cloudformation output tab of student lab
  - For workgroupA, the s3 path would look something like: “s3:// dmslab-student-s3bucketworkgroupa-ldtj44qkwyle/”.
  - For workgroupB, provide S3 path as: “s3:// dmslab-student-s3bucketworkgroupb-n2jrw40pfqcc/”.
- d. For “**Encrypt query results**”, leave as default i.e. unchecked. You can check this if you want your query results to be encrypted.
- e. Check the checkbox for “**Metrics: Publish query metrics to AWS CloudWatch**”

## Lab 3. Consuming data with Athena and Quicksight

Workgroup name\* workgroupA

Description workgroupA for BusinessAnalystUser  
Use up to 1024 characters.

Query result location and Encryption

Query result location s3://dmslab-student-s3bucketworkgroupa-ldtj44 Select Enter a path to an S3 bucket or prefix.

Encrypt query results  Encrypt results stored in S3

Metrics

Metrics  Publish query metrics to AWS CloudWatch ⓘ

Settings

Override client-side settings  ⓘ

Requester pays S3 buckets  Enable queries on requester pays buckets in Amazon S3 ⓘ

Tags

A tag is a label that you assign to an Athena workgroup resource. It consists of a key and a value. Use tags to categorize workgroups by purpose, owner, or environment. You can also use tag specific values for a tag key. Use [best practices](#) and create a consistent set of tags. Do not use duplicate tag keys the same workgroup. [Learn more](#)

Key	name	Value (Optional)	workgroupA	x
Use 1 - 128 characters. (A-Z,a-z,0-9,_,-,;,:/+=,-,@)		Use up to 256 characters. (A-Z,a-z,0-9,_,-,;,:/+=,-,@)		

4. Provide the following:

- a. Optionally, you can click on **Override client-side settings**. This will override the client-side settings and keep the defaults for query execution and storing results.
- b. **Tag** your workgroup to analyze later with CloudWatch or perform any analytics on query execution and results.

- i. For **workgroupA**: provide **key:"name"**, **value:"workgroupA"**
  - ii. For **workgroupB**: Provide **key:"name"**, **value:"workgroupB"**

- c. For “**Requester Pays S3 buckets**”, keep as **default**. This is Optional.

Choose **Enable queries on Requester Pays buckets in Amazon S3** if workgroup users will run queries on data stored in Amazon S3 buckets that are configured as Requester Pays. The account of the user running the query is charged for applicable data access and data transfer fees associated with the query.

5. Click on **create workgroup**

6. Follow the above procedure to create **workgroupB**.

## Lab 3. Consuming data with Athena and Quicksight

### Explore the features of workgroups

- From the **Outputs** tab of DMS student lab, Note down user name **BusinessAnalystUser** and bucket name **S3BucketWorkgroupA** and save it.

Outputs (9)			
Key	Value	Description	Export name
BucketName	dmslab-student-dmslabs3bucket-4a27jap6c5t	S3 Bucket that was created	-
BusinessAnalystUser	dmslab-student-BusinessAnalystUser-878JWTT9AWCK	business_analyst_user for Workgroup A	-
BusinessAnalystUserPolicy	BusinessAnalystUserPolicy	User policy for Business Analyst User	-
DMSLabRoleS3	dmslab-student-DMSLabRoleS3-1VEPY3ZUJX9WB	The DMS service role	-
GlueLabRole	dmslab-student-GlueLabRole-YOAJBNCP66ZI	The Glue service role	-
S3BucketWorkgroupA	dmslab-student-s3bucketworkgroupa-ldtj44qkwyle	S3 Bucket for storing workgroup A results	-
S3BucketWorkgroupB	dmslab-student-s3bucketworkgroupb-n2jrw40pfqcc	S3 bucket for storing workgroup B results	-
WorkgroupManagerUser	dmslab-student-WorkgroupManagerUser-KLF9GDANNTVZ	workgroup_manager_user for access to Workgroup A and Workgroup B	-
WorkgroupManagerUserPolicy	WorkgroupManagerUserPolicy	User policy for Workgroup manager user	-

- Note down 12 digit AWS account id . Follow steps here to find out account id - <https://www.apn-portal.com/knowledgebase/articles/FAQ/Where-Can-I-Find-My-AWS-Account-ID>
- Next, Open AWS console log-in different browser, select **IAM user** and login with following credential:
  - AccountID:** <your-account-name-or-alias>
  - IAM User name:** <value copied for **BusinessAnalystUser**>
  - Password:** **Master123!**
- From new **BusinessAnalystUser** user, Navigate to Athena Console . You will notice that you can see your workgroup designated as “workgroupA” and you can also view table: **sporting\_event\_info** as shown below:

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Query Editor interface. The 'Workgroup' dropdown at the top is set to 'workgroupA'. In the left sidebar, under 'Views (1)', the 'sporting\_event\_info' view is highlighted with a red box. The main area displays a successful query execution:

```
1: SELECT * FROM "ticketdata"."sporting_event_info" limit 10;
```

Buttons for 'Run query', 'Save as', and 'Create' are visible below the query text.

If your workgroup is other than **workgroupA**, click on Workgroup:

This screenshot is identical to the one above, but the 'Workgroup' dropdown at the top is now set to 'workgroupB'. The 'sporting\_event\_info' view in the 'Views (1)' section is highlighted with a red box.

Select **workgroupA** from the workgroup list and then click on **Switch Workgroup**.

### Workgroups

Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch.  
[Learn more](#)

The screenshot shows the AWS Workgroups console. The 'Switch workgroup' button is highlighted with a red box. The table lists the following workgroups:

	Name	Description	Creation time	Workgroup status
●	workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
●	workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
●	primary		2020/03/13 19:35:43 UTC-4	Enabled

5. If you see that your bucket is not setup with Athena to store the query results, as shown below, then proceed to setup the bucket.

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Query Editor interface. On the left, there's a sidebar with 'Data source' set to 'awsdatacatalog' and 'Database' set to 'ticketdata'. Under 'Tables (3)', there are three entries: 'parquet\_sport\_location', 'parquet\_sport\_team', and 'parquet\_sporting\_event'. On the right, the main area has a message box stating: 'Before you run your first query, you need to set up a query result location in Amazon S3'. Below this message, there are two tabs: 'New query 1' and 'New query 2'. The 'New query 1' tab contains the SQL command: '1 SELECT \* FROM "ticketdata"."sporting\_event\_info" limit 10;'. At the bottom of the editor are buttons for 'Run query', 'Save as', and 'Create'.

### 6. Setup the S3 bucket for storing the query results. Click on **Settings**.

This screenshot shows the same Athena Query Editor interface as above, but with a different tab selected. The top navigation bar now includes 'Settings' (which is highlighted with an orange border), 'Tutorial', 'Help', and 'What's new'. The main area of the editor remains the same, displaying the 'New query 1' tab with the SQL command and the 'Run query' button.

Provide the S3 bucket location for workgroupA, copied and saved from the Output tab of cloud formation template, as shown below. Then, click on **Save**.

The screenshot shows the 'Settings' dialog box. It has a heading 'Settings' and a sub-instruction 'Settings apply by default to all new queries. [Learn more](#)'. Below this, it says 'Workgroup: [workgroupA](#)'. There are two main configuration sections: 'Query result location' (set to 's3://dmslab-student-s3bucketworkgroupa-ltj44qkwyle/') and 'Encrypt query results' (with an unchecked checkbox). At the bottom are 'Cancel' and 'Save' buttons.

### 7. Back to Athena Query Editor, click on the three dots against **sporting\_event\_info** view and then click on **Preview**. You will be able to see query results. This shows that you as "business\_analyst\_user" has access to query the view **sporting\_event\_info** and store the query results in S3 bucket designated for workgroupA.

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Query Editor interface. A query is being run against the 'ticketdata' database, specifically targeting the 'sporting\_event\_info' view. The results are displayed in a table format, showing 10 rows of data from various baseball games across different locations and teams. The entire results section is highlighted with a red box.

event_id	sport	event_date_time	home_team	away_team	location	city
1	baseball	2020-07-19 00:00:00.000	Seattle Mariners	New York Mets	Safeco Field	Seattle Washington
2	baseball	2020-09-20 00:00:00.000	Seattle Mariners	Boston Red Sox	Safeco Field	Seattle Washington
3	baseball	2020-07-05 00:00:00.000	Houston Astros	Texas Rangers	Minute Maid Park	Houston Texas
4	baseball	2020-06-14 00:00:00.000	Seattle Mariners	Philadelphia Phillies	Safeco Field	Seattle Washington
5	baseball	2020-07-05 00:00:00.000	Seattle Mariners	Toronto Blue Jays	Safeco Field	Seattle Washington
6	baseball	2020-10-11 00:00:00.000	Seattle Mariners	Los Angeles Angels	Safeco Field	Seattle Washington
7	baseball	2020-07-26 00:00:00.000	San Diego Padres	Kansas City Royals	Petco Park	San Diego California
8	baseball	2020-05-31 00:00:00.000	San Diego Padres	Texas Rangers	Petco Park	San Diego California
9	football	2020-09-21 19:00:00.000	San Diego Chargers	Denver Broncos	Qualcomm Stadium	San Diego, California
10	baseball	2020-04-12 00:00:00.000	Houston Astros	Tampa Bay Rays	Minute Maid Park	Houston Texas

- Logged in as “business\_analyst\_user”, click on **workgroup** and try switching to other workgroups which this user does not have access to. Select **workgroupB** and then click on **switch workgroup**.

The screenshot shows the AWS Workgroups console. It lists three workgroups: 'workgroupB', 'workgroupA', and 'primary'. The 'workgroupB' row is highlighted with a red box. Below the table, there is a 'Switch workgroup' button, which is also highlighted with a red box.

Name	Description	Creation time	Workgroup status
workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
primary		2020/03/13 19:35:43 UTC-4	Enabled

- If you try running the query, you will get the error “Access Denied” as shown below:

The screenshot shows the AWS Athena Query Editor. A query is being run against the 'ticketdata' database, specifically targeting the 'sporting\_event\_info' view. The results section is highlighted with a red box. An error message is displayed in the top right corner: "User: arn:aws:iam::665953140268:user/lakeformation-BusinessAnalystUser-7H32Wd4CW86 is not authorized to perform: athena:StartQueryExecution on resource: arn:aws:athena:us-east-1:665953140268:workgroup/workgroupB (Service: AmazonAthena; Status Code: 400; Error Code: AccessDeniedException; Request ID: 40b3397b-f49b-4d1c-b44c-cd0aa47e977)".

This means that we have achieved the user segregation for different workgroups as defined by the IAM policy and attached to that user. Any query executed and its results within a particular workgroup will be isolated to that workgroup.

## Lab 3. Consuming data with Athena and Quicksight

10. To view the query results, navigate to “workgroup”, select the workgroup and click on “View Details”.

Workgroups

Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)

[Create workgroup](#) [View details](#) [Switch workgroup](#)

	Name	Description	Creation time	Workgroup status
<input type="radio"/>	workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
<input checked="" type="radio"/>	workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
<input type="radio"/>	primary		2020/03/13 19:35:43 UTC-4	Enabled

11. You will be able to see the details, as shown below. Navigate to S3 bucket by clicking on the link and see the query results stored inside the “Unsaved” folder within the **workgroupA** bucket.

Workgroup: workgroupA

[Edit workgroup](#) [Delete workgroup](#) [Disable workgroup](#) [Enable workgroup](#)

[Overview](#) [Metrics](#) [Data usage controls](#) [Tags](#)

To grant access to the workgroup, create an IAM policy and attach it to a user, group or role. [Learn more](#)

Description	Not defined
Query result location	<a href="s3://dmslab-student-s3bucket/workgroupa-ldtj44qkyle/">s3://dmslab-student-s3bucket/workgroupa-ldtj44qkyle/</a> 
Amazon CloudWatch Metrics	Enabled
Encrypt query results	Not defined
Workgroup status	Enabled
Workgroup ARN	arn:aws:athena:us-east-1:678691952726:workgroup/workgroupA 
Bytes scanned cut off per query	Not defined
Override client-side settings	Disabled
Queries with requester pays buckets	Disabled

12. Now, login as workgroup\_manager\_user.

- a. Account ID or Alias: <you-account-id-or-alias>
- b. IAM User Name: <Copy the IAM User Name from cloud formation outputs tab> (for e.g: in this lab: dmslab-student-WorkgroupManagerUser-KLF9GDANNTVZ)
- c. Password: Master123!

This user has access to workgroupA and workgroupB for management purposes.

Switch the workgroups to workgroupA, workgroupB and primary and you will not be able to access the primary workgroup because this user **does not have access to “primary” workgroup**.

## Lab 3. Consuming data with Athena and Quicksight

Workgroups

Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)

Create workgroup View details Switch workgroup

Name	Description	Creation time	Workgroup status
workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
primary		2020/03/13 19:35:43 UTC-4	Enabled

AWS Services Resource Groups Athena S3 AWS Glue RDS Workgroup : workgroup

Workgroup: primary

Edit workgroup Delete workgroup Disable workgroup Enable workgroup

Overview Metrics Data usage controls Tags

To grant access to the workgroup, [create an IAM policy](#) and attach it to a user, group or role. [Learn more](#)

User: arn:aws:iam::665953140268:user/lakeformation-WorkgroupManagerUser-1WHLHD4RLAS0 is not authorized to perform: athena:GetWorkGroup on resource: arn:aws:athena:us-east-1:665953140268:workgroup/primary (Service: AmazonAthena; Status Code: 400; Error Code: AccessDeniedException; Request ID: 07164bd-a04-4571-9aa7-f72507e020d)

Also note that this user does not have access to any tables or cannot run any queries. This is how we can isolate the responsibilities of different users within different workgroups by defining policies and attaching that to the user.

AWS Services Resource Groups Athena S3 AWS Glue RDS Workgroup : workgroup

Query Editor Saved Queries History Data sources Workgroup : workgroup

Data source Connect data source

AwsDataCatalog

Database Choose a database... Filter tables and views... No databases or tables found.

Before you run your first query, you need to [set up a query result location in Amazon S3](#). Learn more

New query 1 New query 2 +

1: SELECT \* FROM "ticketdata"."sporting\_event\_info" limit 10;

Run query Save as Create Use Ctrl + Enter to run query, Ctrl + Space to autocomplete Format query Clear

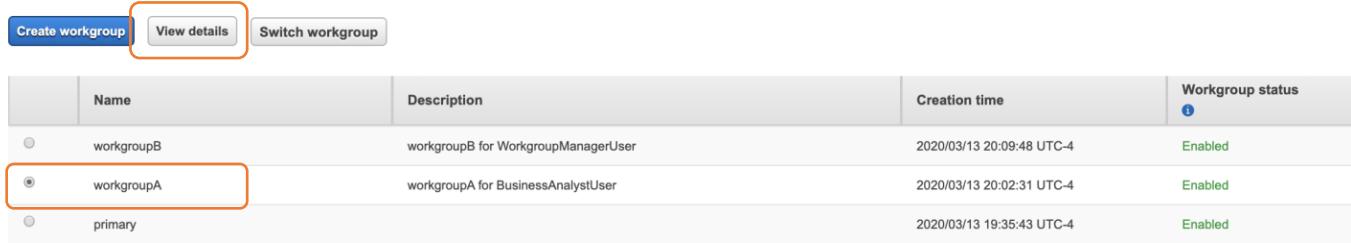
At any point of time, **you can edit, delete and disable your workgroups** as shown:

Select the workgroup and click on “**View Details**”.

## Lab 3. Consuming data with Athena and Quicksight

### Workgroups

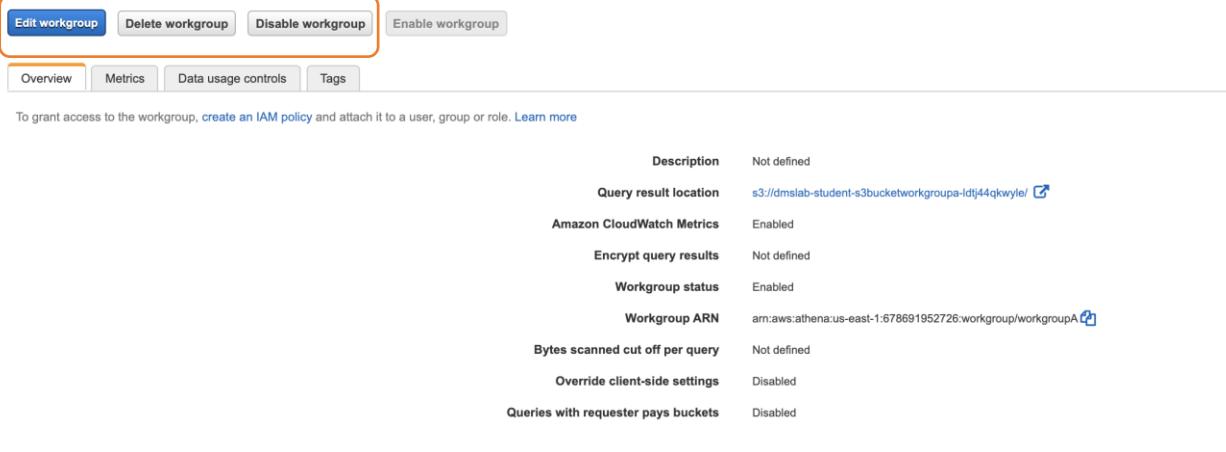
Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)



	Name	Description	Creation time	Workgroup status
●	workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
●	workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
●	primary		2020/03/13 19:35:43 UTC-4	Enabled

Click on “Edit Workgroup” to make changes, “Delete workgroup” to delete the entire workgroup and “Disable workgroup” to disable the workgroup and disable any queries to be run within that workgroup.

### Workgroup: workgroupA



	Description	Not defined
Query result location	s3://dmslab-student-s3bucketworkgroupa-ltj44qkwyel/ 	
Amazon CloudWatch Metrics	Enabled	
Encrypt query results	Not defined	
Workgroup status	Enabled	
Workgroup ARN	arn:aws:athena:us-east-1:678691952726:workgroup/workgroupA 	
Bytes scanned cut off per query	Not defined	
Override client-side settings	Disabled	
Queries with requester pays buckets	Disabled	

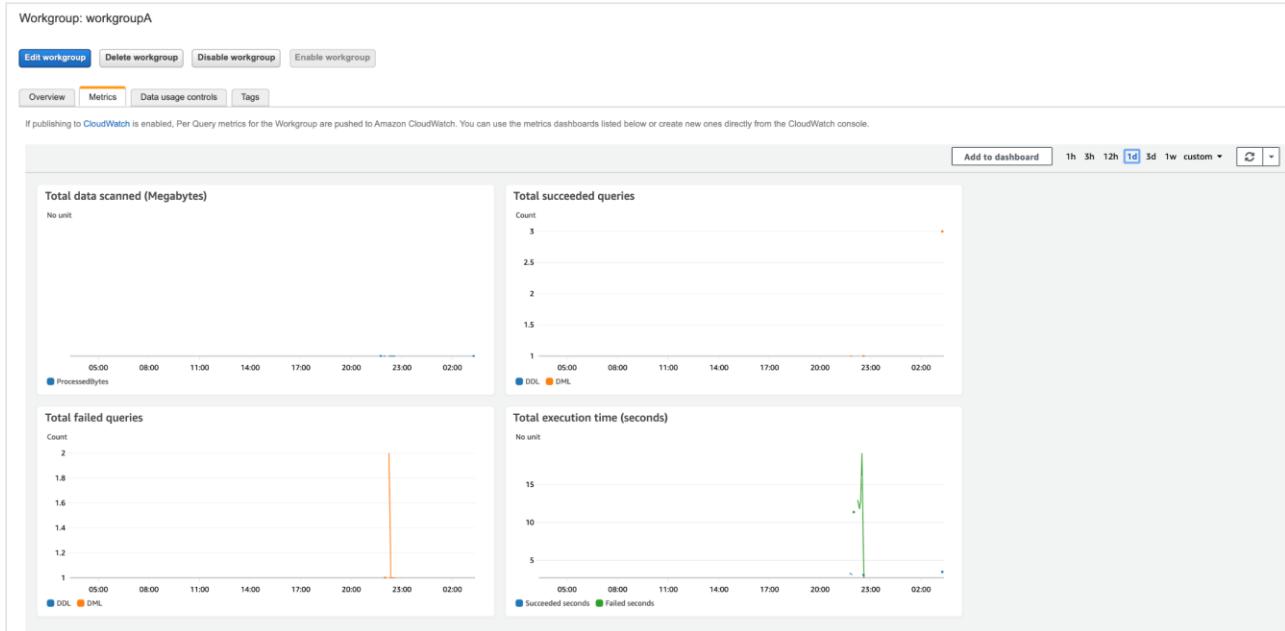
**Please Note:** For lab purpose, we are attaching policies directly to users. For Best practices, we recommend creating separate groups in IAM for different workgroups and then attaching policies for different workgroups to their respective groups in IAM.

## Managing Query Usage and Cost

**\*\*Please Note** that the following section of this lab is carried out under **admin** account and not the **BusinessAnalystUser** and **WorkgroupManagerUser**, so please login to your account with admin credentials\*\*

Once you **enable the CloudWatch metrics** for your workgroups, you will be able to see **Metrics**, by selecting the desired workgroup and click on **Metrics** as shown:

## Lab 3. Consuming data with Athena and Quicksight



Choose the **metrics interval** that Athena should use to fetch the query metrics from CloudWatch, or choose the **refresh** icon to refresh the displayed metrics.

1h 3h 12h 1d 3d 1w custom ▾



Let's setup data usage controls which means setting up the threshold for the amount of data scanned. There are two types of data usage controls: **per-query** and **per-workgroup**.

**Per-query data usage control** will check the total amount of data scanned by per query within the workgroup and if the amount exceeds the threshold, the query will be cancelled automatically. Let's setup **per-query data usage for “primary workgroup”**.

1. From Athena console, click on **Workgroup** and select **primary**. Click on **View Details**

Workgroups				
Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. <a href="#">Learn more</a>				
<a href="#">Create workgroup</a>		<a href="#">View details</a>	<a href="#">Switch workgroup</a>	
Name	Description	Creation time	Workgroup status	
workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled	
workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled	
primary		2020/03/13 19:35:43 UTC-4	Enabled	

2. Click on **Data usage controls**. In **Per query data usage control**, the default minimum limit is 10MB per query. We will select the default value- 10MB. Also, note the default

## Lab 3. Consuming data with Athena and Quicksight

**"Action"** for per query data usage control. If the query exceeds the limit, it will be cancelled.

3. Click **Update**
4. The per-query threshold has been set.

The screenshot shows the AWS Workgroup console for the 'primary' workgroup. The 'Data usage controls' tab is selected. A data limit of 10 Megabytes MB is set, and the action is set to cancel the query if it exceeds the limit. The 'Update' button is visible at the bottom.

5. Navigate to query editor on Athena console. Run the following query:  
**SELECT \* FROM "ticketdata"."cdc\_sporting\_event\_ticket"**
6. This query scans 200 MB of data, but since we have set the threshold as 10MB, this query execution will be cancelled, as shown:

The screenshot shows the AWS Athena Query Editor. A query for 'SELECT \* FROM "ticketdata"."cdc\_sporting\_event\_ticket"' is running. A red warning box indicates 'Query cancelled! : Bytes scanned limit was exceeded'. The status bar shows the run time and data scanned.

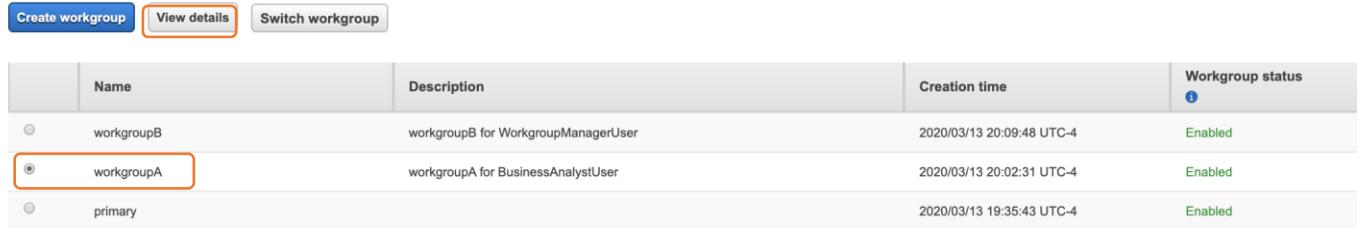
For **per-workgroup data usage control**, you can configure the maximum amount of data scanned by all queries in the workgroup during a specific period. This is useful when you have few analytics reports to run, where you probably have a good idea of how long the process should take and the total amount of data that queries scan during this time. You only have a few reports to run, so you can expect them to run in a few minutes, only scanning a few megabytes of data.

1. Login as **Admin** to the account. On Athena console, click on **Workgroup** and Select **workgroupA**. Click on **View Details**.

## Lab 3. Consuming data with Athena and Quicksight

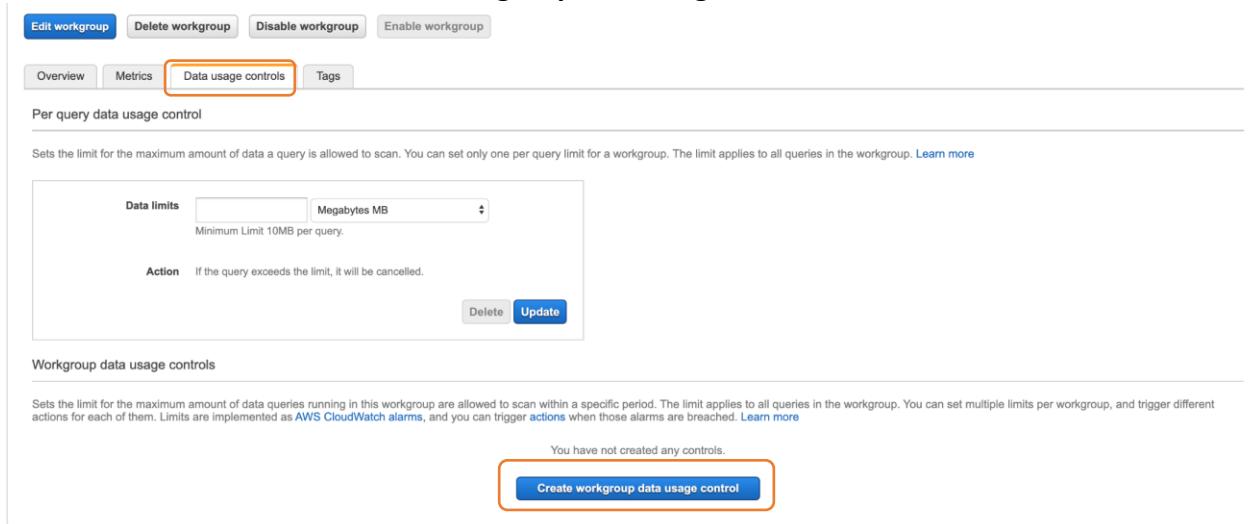
### Workgroups

Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)



	Name	Description	Creation time	Workgroup status
●	workgroupB	workgroupB for WorkgroupManagerUser	2020/03/13 20:09:48 UTC-4	Enabled
●	workgroupA	workgroupA for BusinessAnalystUser	2020/03/13 20:02:31 UTC-4	Enabled
●	primary		2020/03/13 19:35:43 UTC-4	Enabled

2. Click on **Data usage Controls** and scroll down to section **Workgroup data usage controls**. Click on **Create workgroup data usage control**



Per query data usage control

Sets the limit for the maximum amount of data a query is allowed to scan. You can set only one per query limit for a workgroup. The limit applies to all queries in the workgroup. [Learn more](#)

Data limits	<input type="text"/>	Megabytes MB
-------------	----------------------	--------------

Minimum Limit 10MB per query.

Action If the query exceeds the limit, it will be cancelled.

Delete Update

Workgroup data usage controls

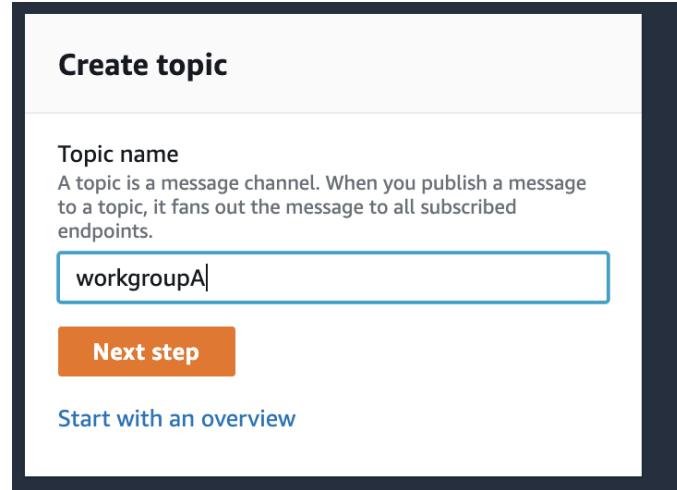
Sets the limit for the maximum amount of data queries running in this workgroup are allowed to scan within a specific period. The limit applies to all queries in the workgroup. You can set multiple limits per workgroup, and trigger different actions for each of them. Limits are implemented as [AWS CloudWatch alarms](#), and you can trigger [actions](#) when those alarms are breached. [Learn more](#)

You have not created any controls.

Create workgroup data usage control

3. The select query on **sporting\_event\_info** returns more than 10KB of data. For this lab, we have only this table to query from. So, let's set the threshold accordingly.
  - a. Set **Data Limits** to **10 KBs**
  - b. Set **Time period** to **1 minute**
  - c. Set **Action** as "**Send a notification to**". Here, click on **Create SNS Topic**.
    - i. This will take you to **SNS Console**. Provide **Topic Name** as **workgroupA**.

## Lab 3. Consuming data with Athena and Quicksight



- ii. Click on **Next Step**
- iii. Click on **Create Subscription**. We will subscribe to this topic with **email address**. Whenever the threshold is breached, we will get an email notification to the email address which is our subscriber.

The screenshot shows the Amazon SNS 'Topics' page. On the left is a sidebar with 'Amazon SNS' and links for 'Dashboard', 'Topics' (which is highlighted in orange), 'Subscriptions', 'Mobile' (with 'Push notifications' and 'Text messaging (SMS)'), and 'AWS Lambda'. The main content area shows a topic named 'workgroupA'. The 'Details' section shows the 'Name' as 'workgroupA', 'Display name' as '-', 'ARN' as 'arn:aws:sns:us-east-1:665953140268:workgroupA', and 'Topic owner' as '665953140268'. Below this are tabs for 'Subscriptions', 'Access policy', 'Delivery retry policy (HTTP/S)', 'Delivery status logging', 'Encryption', and 'Tags'. The 'Subscriptions' tab is selected, showing '(0)' and buttons for 'Edit', 'Delete', 'Request confirmation', 'Confirm subscription', and an orange 'Create subscription' button.

- iv. In **Create Subscription**, select **Protocol** as **Email**. In **Endpoint**, Provide **email address**, then click on **Create subscription**.

## Lab 3. Consuming data with Athena and Quicksight

Amazon SNS > Subscriptions > Create subscription

### Create subscription

**Details**

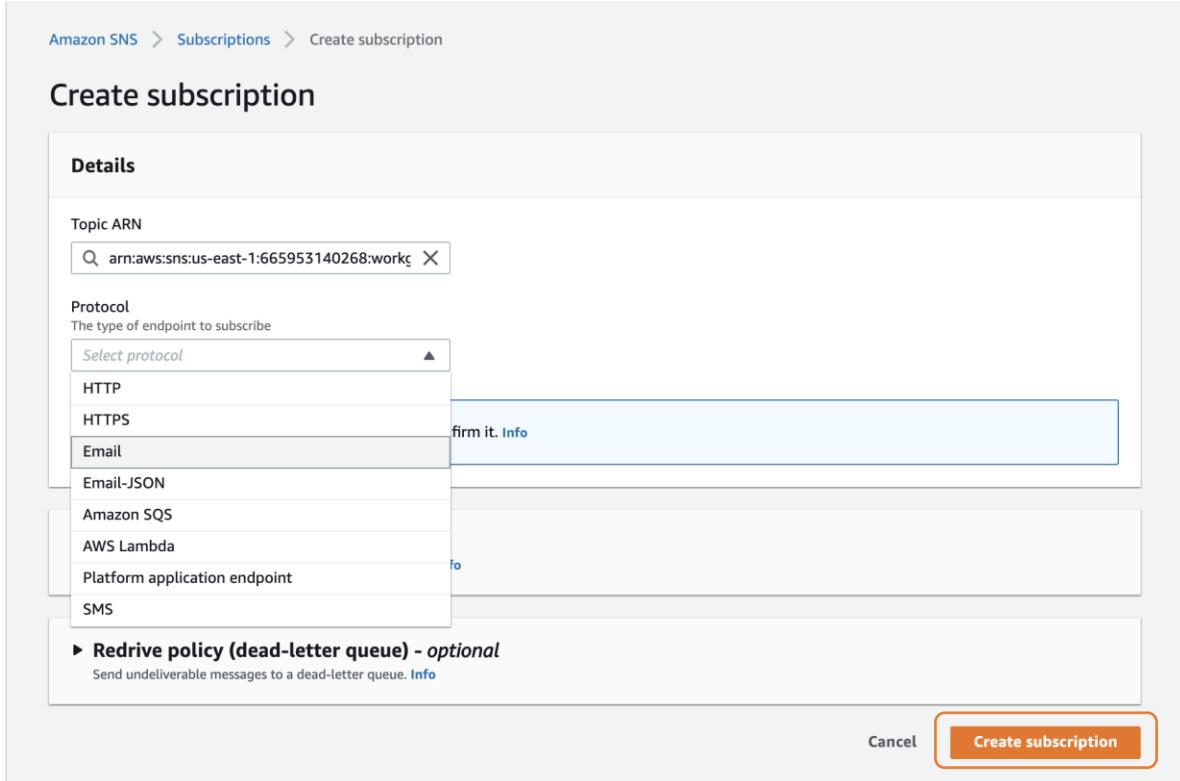
Topic ARN  
arn:aws:sns:us-east-1:665953140268:workgroupA

Protocol  
The type of endpoint to subscribe

Select protocol ▾  
HTTP  
HTTPS  
Email  
Email-JSON  
Amazon SQS  
AWS Lambda  
Platform application endpoint  
SMS

► Redrive policy (dead-letter queue) - optional  
Send undeliverable messages to a dead-letter queue. [Info](#)

Cancel **Create subscription**



- v. Verify your email for subscription to be validated.
- vi. Back to WorkgroupA workgroup data usage control, for **Action**, select **workgroupA** for the **SNS topic**. Click on **Create**.

**Create workgroup data usage control**

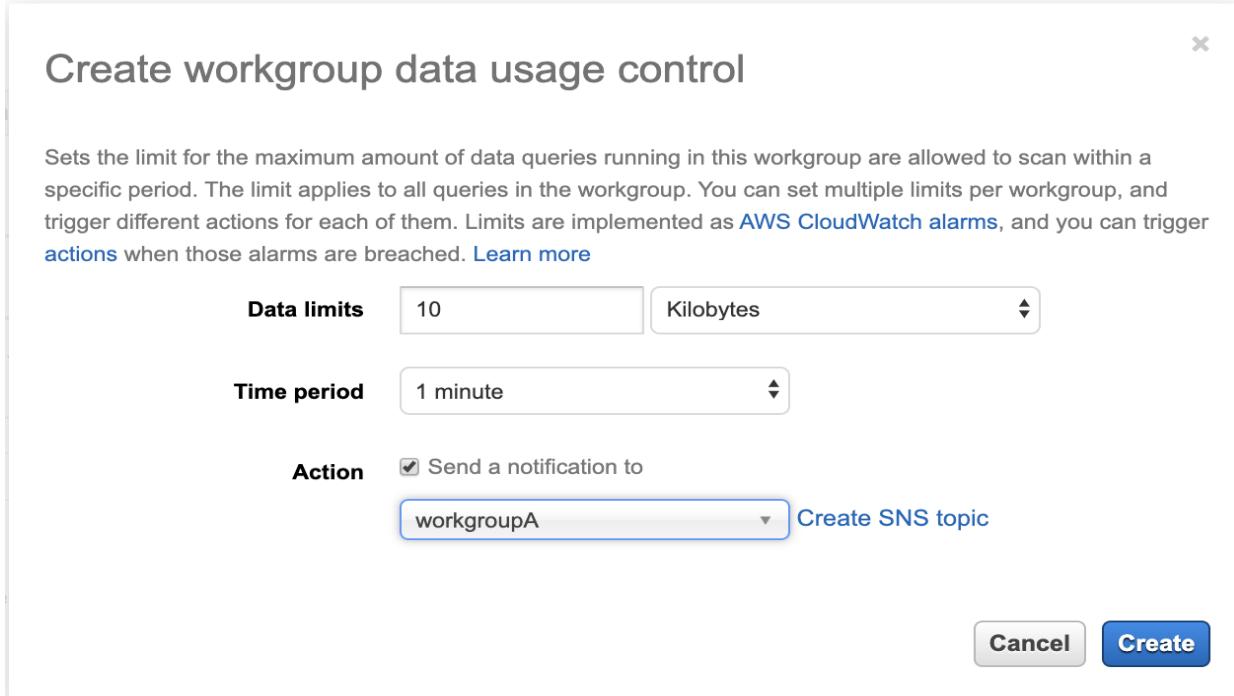
Sets the limit for the maximum amount of data queries running in this workgroup are allowed to scan within a specific period. The limit applies to all queries in the workgroup. You can set multiple limits per workgroup, and trigger different actions for each of them. Limits are implemented as [AWS CloudWatch alarms](#), and you can trigger [actions](#) when those alarms are breached. [Learn more](#)

**Data limits** 10 Kilobytes

**Time period** 1 minute

**Action**  Send a notification to  
workgroupA [Create SNS topic](#)

Cancel **Create**



## Lab 3. Consuming data with Athena and Quicksight

vii. Once created, this control will be listed like this:

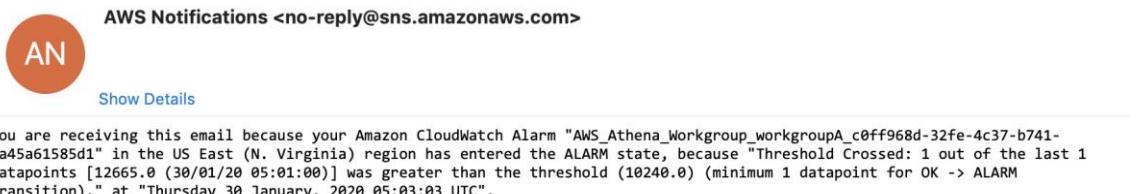
The screenshot shows the 'Data usage controls' section of the AWS Workgroup console. It includes fields for setting data limits (Megabytes MB) and defining actions if the limit is exceeded. A table lists existing data usage controls, showing one entry for 10.24 KB with a time period of 1 minute and an action of sending a notification to a specific SNS topic.

4. Back to **Athena Query Editor**, run the following query, by logging in as **Business Analyst User** to the console and selecting **Workgroup: workgroupA**:

```
SELECT * FROM "ticketdata"."sporting_event_info";
```

5. You will receive an **email notification from AWS Notifications** stating that workgroup data usage threshold has been breached, which will look something like this:

ALARM: "AWS\_Athena\_Workgroup\_workgroupA\_c0ff968d-32fe-4c37-b741-fa45a61..." in US East (N. Virgi...



6. You can also check **CloudWatch Alarms** and get more details on CloudWatch console:

The CloudWatch Alarms console displays a table of alarms categorized by AWS service. An alarm for 'AWS/Athena' is shown in the 'ALARM' state. To the right, a 'Recent alarms' chart shows a step increase in 'ProcessedBytes' at 05:03 UTC on January 30, 2020, reaching a value of 10.8k, which triggered the alarm.

## Lab 3. Consuming data with Athena and Quicksight

7. Alternatively, you can have AWS Lambda as the subscriber endpoint, so as soon as the threshold is breached, SNS will call the lambda function, which in turn will disable the workgroup and preventing from executing further queries within that workgroup. Feel free to explore multiple subscriber endpoints.

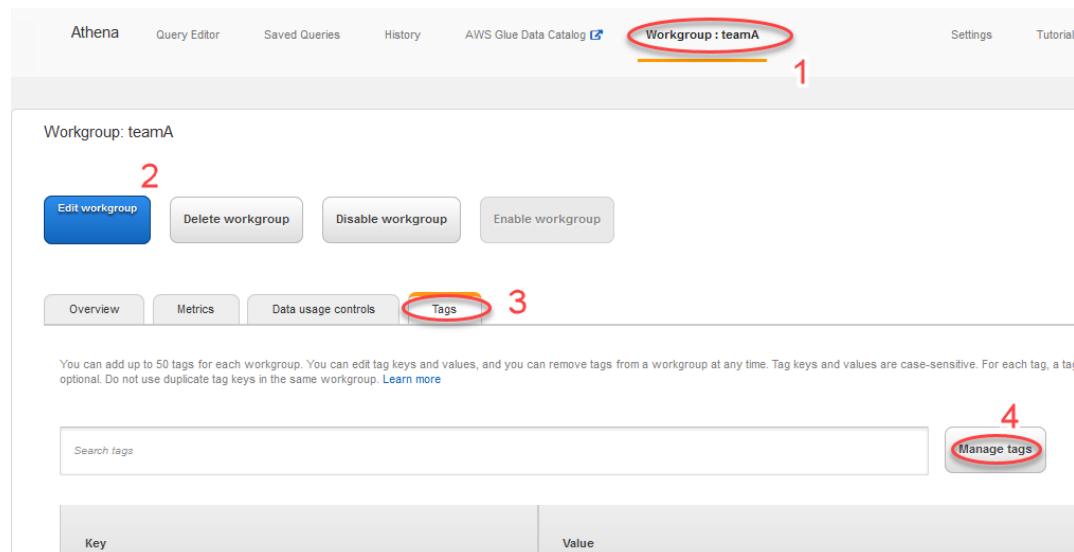
### Cost Allocation Tags

When you created two workgroups: **workgroupA** and **workgroupB**, you also created **name as tags**. These tags can be utilized in Billing and Cost Management console to determine the usage per workgroup.

For example, you can create a set of tags for workgroups in your account that helps you track workgroup owners, or identify workgroups by their purpose. You can **view tags for a workgroup in “View Details” page** for the workgroup under consideration.

You can add tags later after you have created workgroup. To create tags:

1. Open the Athena console at <https://console.aws.amazon.com/athena/>, choose the **Workgroups** tab, and select the workgroup.
2. Choose **View details** or **Edit workgroup**.
3. Choose the **Tags** tab.
4. On the **Tags** tab, choose **Manage tags**, and then specify the key and value for each tag.
5. When you are done, choose **Save**.



For more details on best practices: <https://docs.aws.amazon.com/athena/latest/ug/tags-console>