



Amazon Web Services

Data Engineering Immersion Day

Lab 3. Consuming data with Athena and Quicksight

March 2020

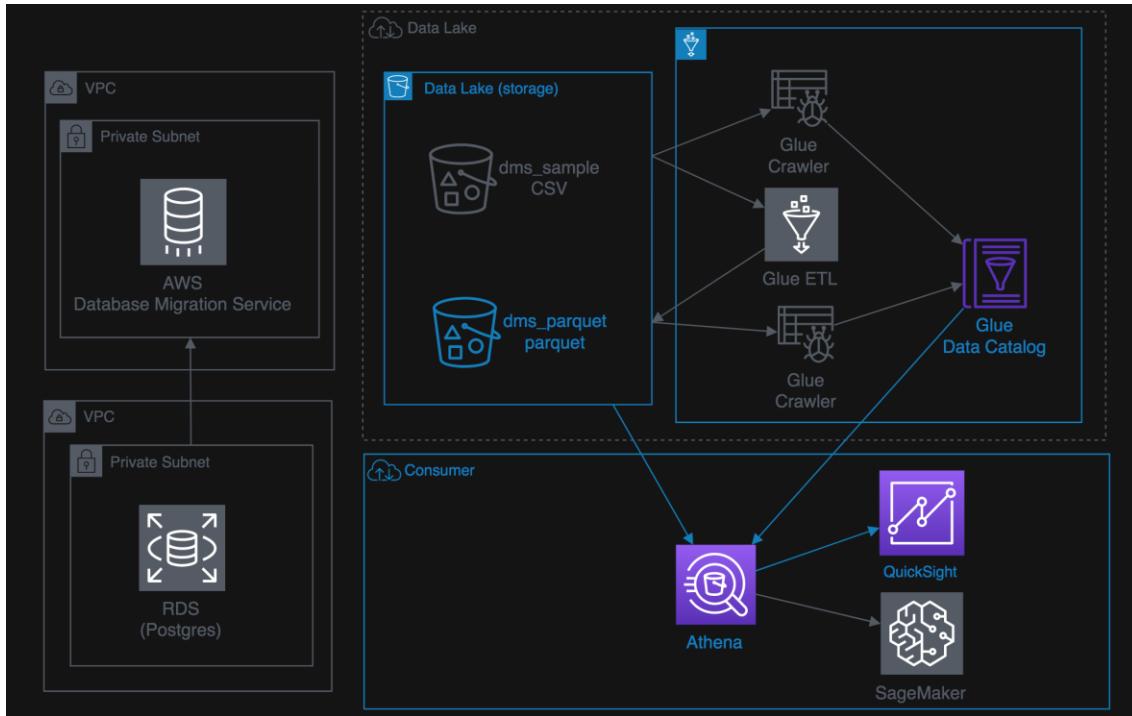
Table of Contents

<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/>

Lab 3. Consuming data with Athena and Quicksight

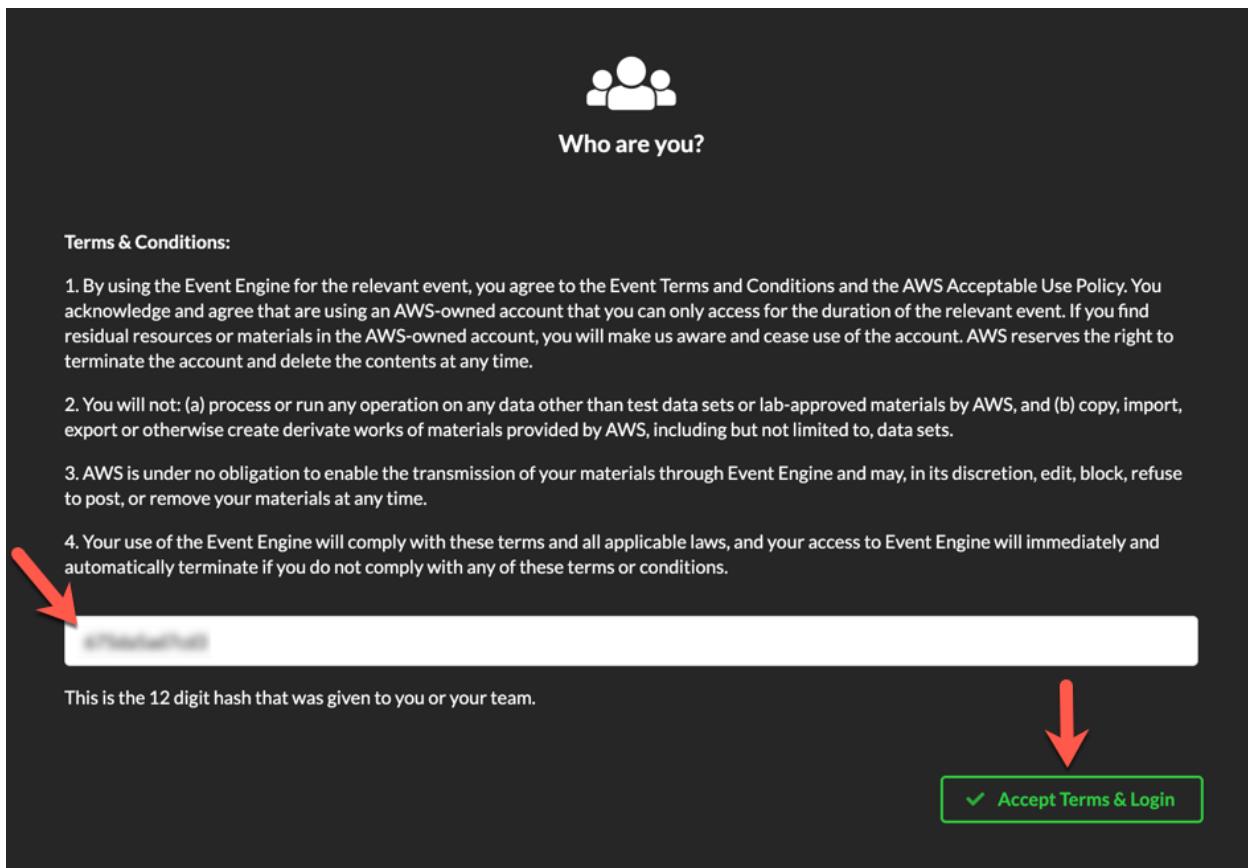
Get Started Using the Lab Environment

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-based interface for managing AWS resources. At the top, there's a navigation bar with a gear icon and the word "Modules". Below it, a section titled "Environment Setup" contains a "Readme" link. The main content area is titled "Outputs:" and lists several AWS resources with their corresponding ARNs:

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

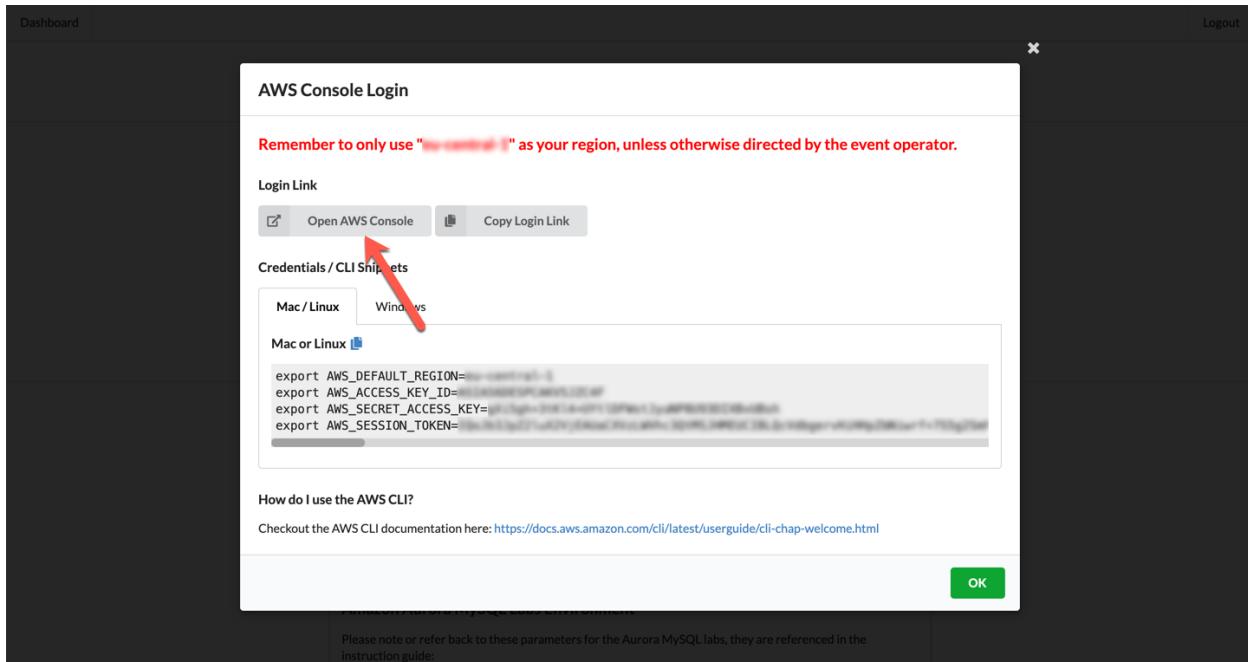
3. On the Team Dashboard, please click AWS Console to log into the AWS Management Console:

The screenshot shows the "Team Dashboard" interface. At the top, there's a "Event" section with a cloud icon and a downward arrow. Below it, there are two buttons: "AWS Console" and "SSH Key". The main content area displays information about an event:

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

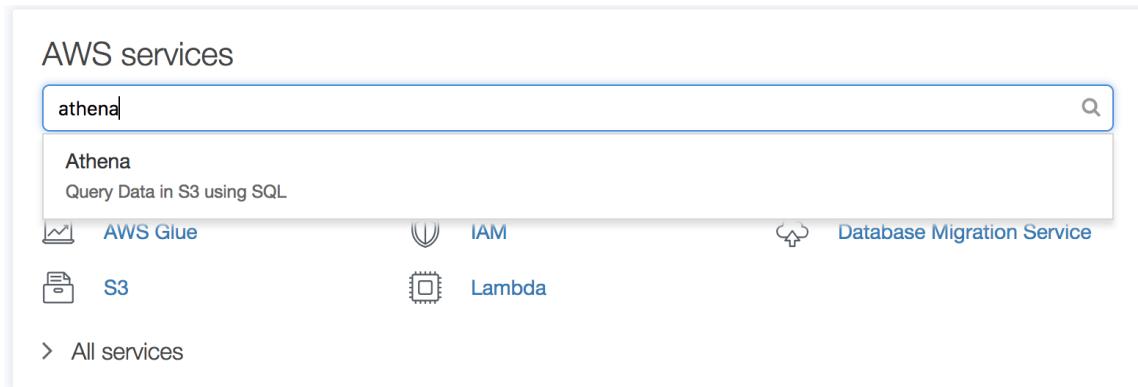


Once you have completed these steps, you can continue with the rest of this lab.

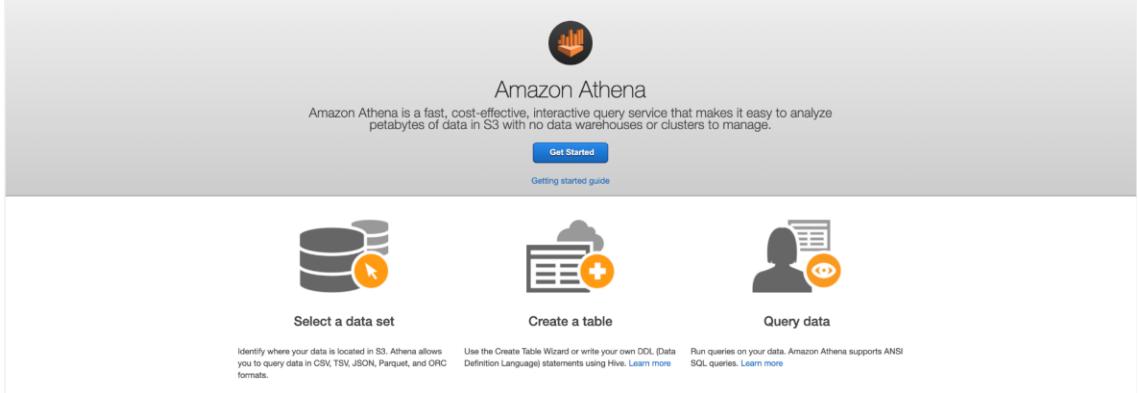
Lab 3. Consuming data with Athena and Quicksight

Query Data with Amazon Athena

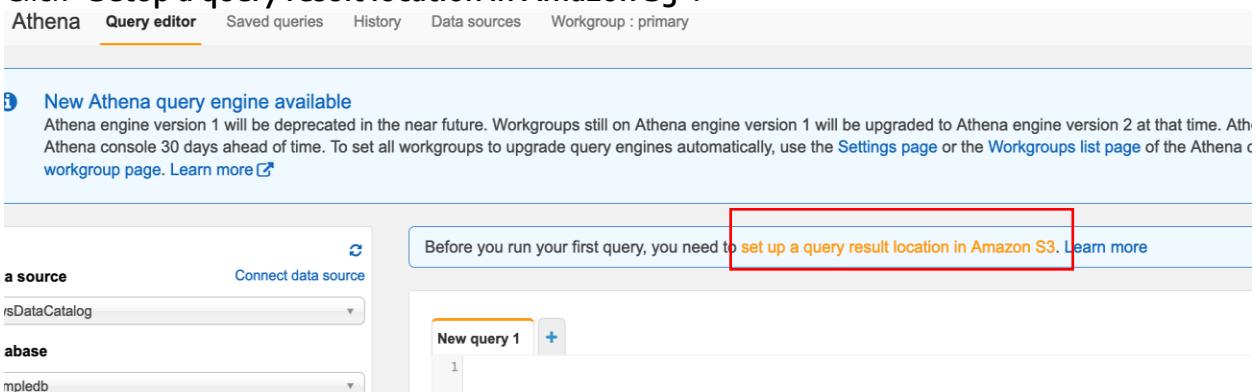
1. In the AWS services console, search for **Athena**.



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

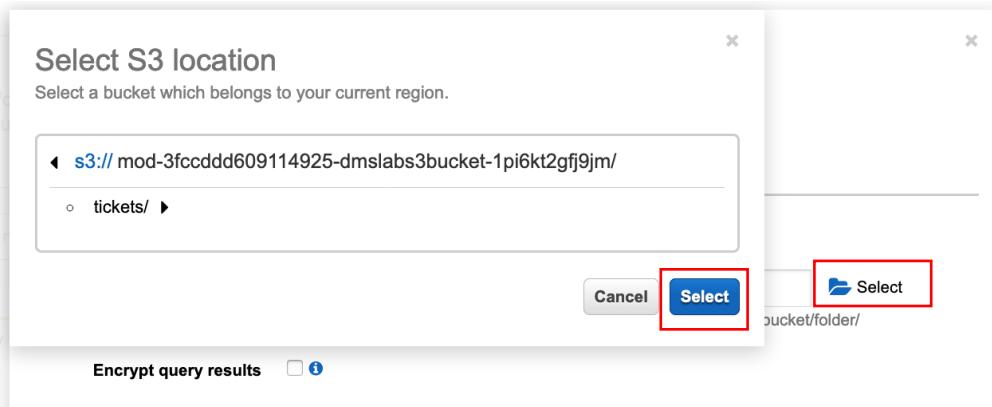


3. Click “**Setup a query result location in Amazon S3**”.

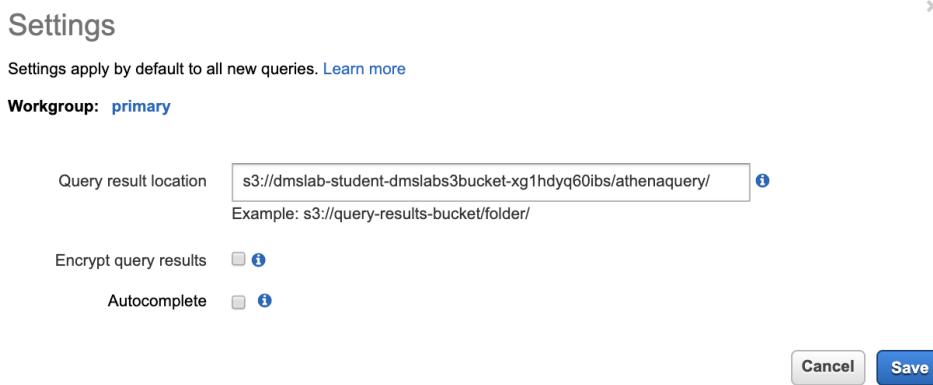


4. It navigates to a **Settings** page. Click on “**Select**” folder icon, choose the **dmslabs3bucket** (e.g: <dmslab-student-dmslabs3bucket-xg1hdyq6oibs>). then click on “**Select**” button.

Lab 3. Consuming data with Athena and Quicksight



- Append **athenaquery/** at the end of the S3 location. Click on **Save**.



- In the **Query Editor**, select your newly created database e.g., "**ticketdata**".
 - 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**).

Athena **Query Editor** Saved Queries

Catalog Glue Database ticketdata

Tables (22) Create table

- ▶ mib_data
- ▶ name_data
- ▶ nfl_data
- ▶ nfl_stadium_data
- ▶ parquet_person
- ▶ parquet_sport_location
- ▶ parquet_sport_team
- ▶ parquet_sporting_event
- ▶ parquet_sporting_event_ticket
 - ▶ id (double)
 - ▶ sporting_event_id (double)
 - ▶ sport_location_id (string)
 - ▶ seat_level (bigint)
 - ▶ seat_section (bigint)
 - ▶ seat_row (string)
 - ▶ seat (bigint)
 - ▶ ticketholder_id (double)
 - ▶ ticket_price (double)
- ▶ person
- ▶ player

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

Lab 3. Consuming data with Athena and Quicksight

8. 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.

The screenshot shows the AWS Athena console interface. At the top, there is a tab labeled "New query 1". Below it is the SQL query code. At the bottom of the query editor, there are several buttons: "Run query" (highlighted in blue), "Save as", "Create", and others. To the right of the buttons, it says "(Run time: 1.41 seconds, Data scanned: 12.38 KB)". Further down, there are links for "Format query" and "Clear". Below the editor, a message says "Use Ctrl + Enter to run query, Ctrl + Space to autocomplete". On the far right, it shows "Athena engine version 1" and "Release versions".

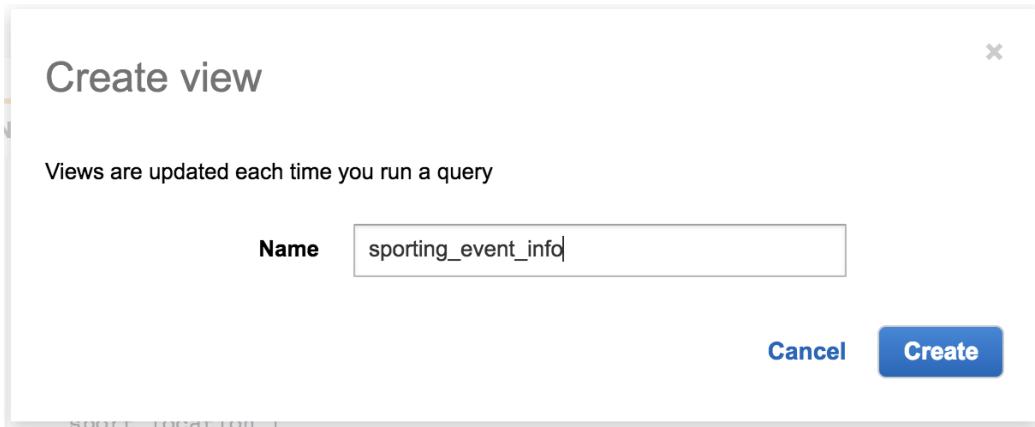
The screenshot shows the results of the query in a table format. The table has columns: event_id, sport, event_date_time, home_team, away_team, location, and city. There are 6 rows of data. The first row is highlighted in light blue.

▲	event_id ▼	sport ▼	event_date_time ▼	home_team ▼	away_team ▼	location ▼	city ▼
1	1	baseball	2019-04-07 00:00:00.000	New York Mets	Detroit Tigers	Citi Field	Queens New York
2	11	baseball	2019-04-14 00:00:00.000	New York Mets	Atlanta Braves	Citi Field	Queens New York
3	21	baseball	2019-04-21 00:00:00.000	New York Mets	Minnesota Twins	Citi Field	Queens New York
4	31	baseball	2019-04-28 00:00:00.000	New York Mets	Los Angeles Dodgers	Citi Field	Queens New York
5	41	baseball	2019-05-05 00:00:00.000	New York Mets	Kansas City Royals	Citi Field	Queens New York
6	51	baseball	2019-05-12 00:00:00.000	New York Mets	Colorado Rockies	Citi Field	Queens New York

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

10. Name the view **sporting_event_info** and click **Create**.

Lab 3. Consuming data with Athena and Quicksight



Your new view is created

Athena Query Editor Saved Queries History Data sources Workgroup : primary

Data source Connect data source

Database

Tables (24)

Views (1)

sporting_event_info

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

New query 1 New query 2 New query 3 +

```
1 CREATE OR REPLACE VIEW "sporting_event_info" AS
2 SELECT
3   e.id AS event_id,
4   e.sport_type_name AS sport,
5   e.start_date_time AS event_date_time,
6   h.name AS home_team,
7   a.name AS away_team,
8   l.name AS location,
9   l.city
10  FROM parquet_sporting_event e,
11    parquet_sport_team h,
12    parquet_sport_team a,
13    parquet_sport_location l
14  WHERE
15    e.home_team_id = h.id
16    AND e.away_team_id = a.id
17    AND e.location_id = l.id
```

Run query Save as Create (Run time: 0.71 seconds, Data scanned: 0 KB)

Use Ctrl + Enter to run query, Ctrl + Space to autocomplete

11. Copy the following SQL syntax into the **New Query 3** tab.

```
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,
```

Lab 3. Consuming data with Athena and Quicksight

```
parquet_person p
WHERE
    t.sporting_event_id = e.event_id
    AND t.ticketholder_id = p.id
```

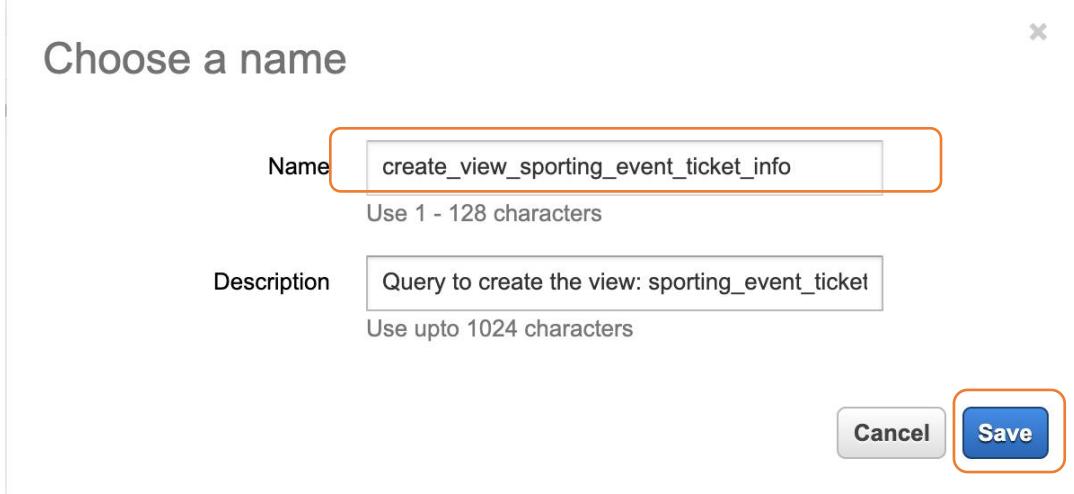
The screenshot shows the AWS Athena Query Editor interface. At the top, there are three tabs: "New query 1", "New query 2", and "New query 3" (which is selected, indicated by an orange border). Below the tabs is a code editor containing the provided SQL query. The code includes column aliases like "ticket_id" and "full_name", and joins between tables "sporting_event_info", "parquet_sporting_event_ticket", and "parquet_person". A WHERE clause filters the results based on event and ticketholder IDs. The code editor has line numbers from 1 to 20. Below the code editor are three buttons: "Run query" (blue), "Save as" (gray), and "Create" (gray). To the right of the "Run query" button is the text "(Run time: 21.04 seconds, Data scanned: 139.22 MB)". At the bottom of the editor, a note says "Use Ctrl + Enter to run query, Ctrl + Space to autocomplete".

```
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 (Run time: 21.04 seconds, Data scanned: 139.22 MB)

Use Ctrl + Enter to run query, Ctrl + Space to autocomplete

12. Click on **Save as** button Give this query a name: `create_view_sporting_event_ticket_info` and some description and then, click on **Save**.



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

Lab 3. Consuming data with Athena and Quicksight

```
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  
21
```

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

The results appear beneath the query window.

```
New query 1 New query 2 + 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  
21
```

Run query **Save as** **Create** (Run time: 20.65 seconds, Data scanned: 139.22 MB)

Use Ctrl + Enter to run current query. Use Shift + Enter to run previous query. Use Esc to cancel a query. Use F5 to refresh the results.

Athena engine ver

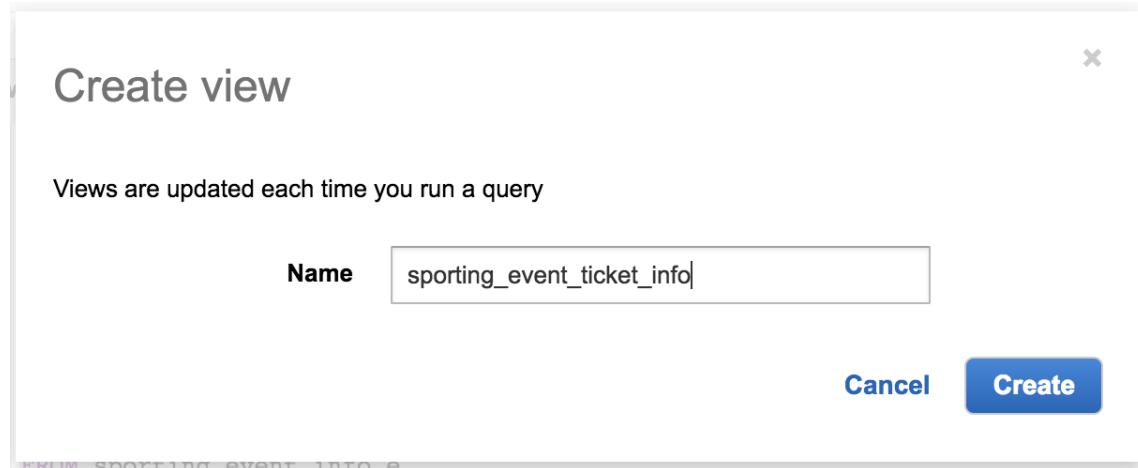
Create table from query **Create view from query**

Results

	ticket_id	event_id	sport	event_date_time	home_team	away_team	location	city	seat_level
1	241271.0	8771	football	2019-09-23 12:00:00.000	New England Patriots	Buffalo Bills	Gillette Stadium	Foxborough, Massachusetts	3
2	247911.0	8771	football	2019-09-23 12:00:00.000	New England Patriots	Buffalo Bills	Gillette Stadium	Foxborough, Massachusetts	3
3	247901.0	8771	football	2019-09-23 12:00:00.000	New England Patriots	Buffalo Bills	Gillette Stadium	Foxborough, Massachusetts	3

13. As shown above, click **Create view from query**.
14. Name the view "sporting_event_ticket_info" and click **Create**.

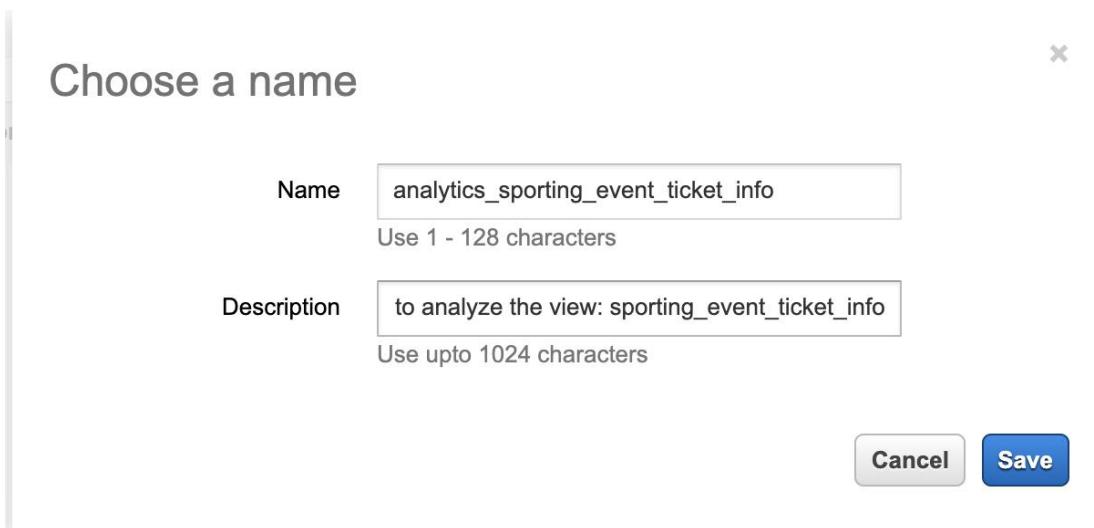
Lab 3. Consuming data with Athena and Quicksight



15. Copy the following SQL syntax into the New Query 4 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 as** and give this query name: **analytics_sporting_event_ticket_info** and some description and then, click on **Save**.



The name of the New Query 4 will be changed to one assigned in previous step. Click on **Run Query**.

Lab 3. Consuming data with Athena and Quicksight

```
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

Your 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.

sport	locations	events	tickets	avg_ticket_price
baseball	30	294	958680	53.89345581425812
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:

Lab 3. Consuming data with Athena and Quicksight

Amazon S3 > dmslab-student-dmslabs3bucket-xg1hdyq60ibs > athenaquery

dmslab-student-dmslabs3bucket-xg1hdyq60ibs

Overview

Upload + Create folder Download Actions

US East (N. Virginia)

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

Viewing 1 to 2

Build an Amazon QuickSight Dashboard

Set up QuickSight

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

AWS services

QuickSight

QuickSight
Fast, easy to use business analytics

Athena S3 AWS Glue

IAM CloudWatch

All services

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

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

Your AWS Account is not signed up for QuickSight. Would you like to sign up now?

AWS Account 292264923209

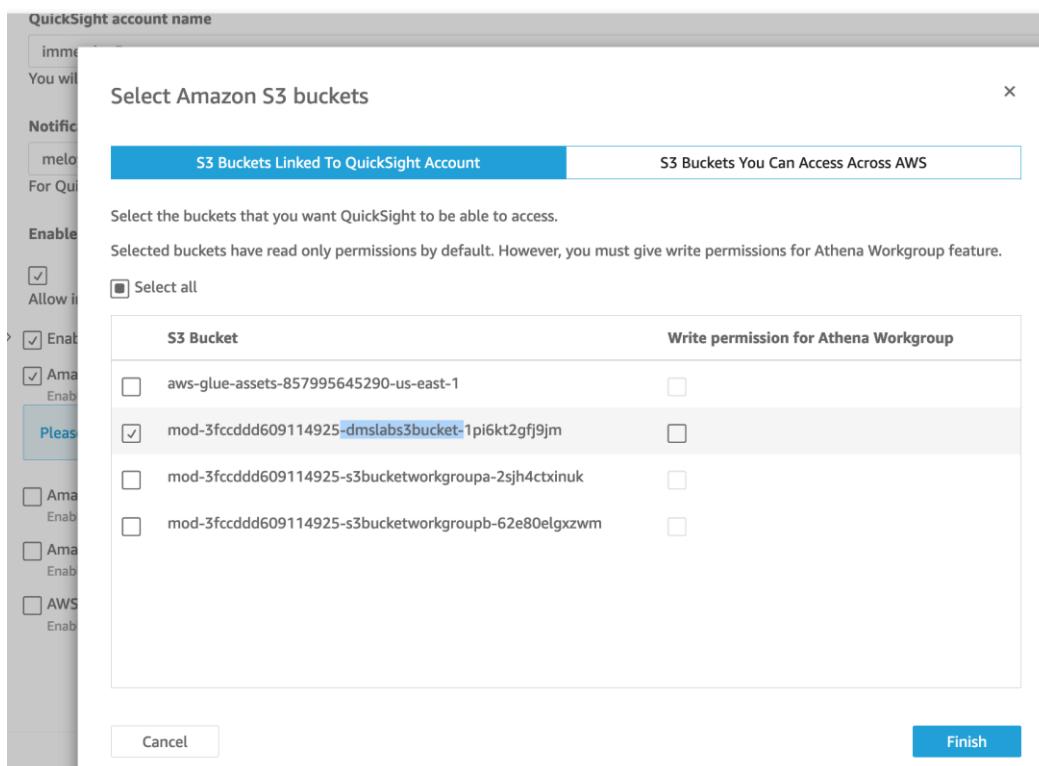
Sign up for QuickSight

3. For account type, choose the default **Enterprise** Version.
4. Click **Continue**.

Lab 3. Consuming data with Athena and Quicksight



5. On the Create your QuickSight account page, fill out your name and email address.
6. Keep the default region “**US East (N. Virginia)**” and the check boxes to enable auto discovery, Amazon Athena, and Amazon S3.
7. Select your DMS bucket (e.g., “xxx-dmslabs3bucket-xxx”), 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 julbright+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

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

Amazon S3 (1 bucket)

Amazon S3 Storage Analytics

Amazon IoT Analytics

Choose S3 buckets

Finish

- On the **QuickSight landing page**, on the **top right corner**, click on "**Manage QuickSight**".

New analysis

All analyses All dashboards Favorites Tutorial videos

All analyses Last updated

Manage QuickSight

Community Send feedback What's new

- 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

Security & permissions

QuickSight can control access to AWS resources for the entire account in addition to individual users and groups

QuickSight access to AWS services

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

By configuring access to AWS services, QuickSight can access the data in those services. Access by users and groups can be controlled through the options below.

Add or remove

Lab 3. Consuming data with Athena and Quicksight

10. If you will observe there is an unchecked box against S3 buckets for "xxx-dmslabs3bucket-", please **check the box**.



11. Select the **dmslabs3bucket** (e.g: xxx-dmslabs3bucket-xxx), which will have all the folders for your source data.

Select Amazon S3 buckets ×

S3 Buckets Linked To QuickSight Account S3 Buckets You Can Access Across AWS

Select the buckets that you want QuickSight to be able to access.
Selected buckets have read only permissions by default. However, you must give write permissions for Athena Workgroup feature.

Select all

S3 Bucket	Write permission for Athena Workgroup
<input type="checkbox"/> aws-glue-assets-857995645290-us-east-1	<input type="checkbox"/>
<input checked="" type="checkbox"/> mod-3fccddd609114925-dmslabs3bucket-1pi6kt2gfj9jm	<input type="checkbox"/>
<input type="checkbox"/> mod-3fccddd609114925-s3bucketnetworkgroupa-2sjh4ctxinuk	<input type="checkbox"/>
<input type="checkbox"/> mod-3fccddd609114925-s3bucketnetworkgroupb-62e80elgxzwm	<input type="checkbox"/>

Lab 3. Consuming data with Athena and Quicksight

12. Then, click on **Finish**.

13. 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**".

QuickSight access to AWS services

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

	Amazon Redshift	<input checked="" type="checkbox"/>
	Enables QuickSight to auto-discover clusters	
	Amazon RDS	<input checked="" type="checkbox"/>
	Enables QuickSight to auto-discover instances	
	IAM	<input checked="" type="checkbox"/>
	Enables you to invite IAM users from this AWS Account to access QuickSight	
	Amazon S3	<input checked="" type="checkbox"/>
	Enables QuickSight to auto-discover your Amazon S3 buckets Hide	
Select S3 buckets 1 buckets selected		
	Amazon Athena	<input checked="" type="checkbox"/>
	Enables QuickSight access to Amazon Athena databases	
	Amazon S3 Storage Analytics	<input type="checkbox"/>
	Enables QuickSight to visualize your S3 Storage Analytics data	
	AWS IoT Analytics	<input type="checkbox"/>
	Enables QuickSight to visualize your IoT Analytics data	
	Amazon Elasticsearch Service	<input type="checkbox"/>
	Enable QuickSight access to your Amazon Elasticsearch Service domains	
	Amazon SageMaker	<input type="checkbox"/>
	Enables QuickSight to infer fields from custom ML models	
	Amazon Timestream	<input type="checkbox"/>
	Enable QuickSight access to your Amazon Timestream databases	

[Cancel](#) Update

14. Navigate to QuickSight landing page by clicking on the **QuickSight logo** on the top left. On the top right corner, click **New analysis**.

Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS QuickSight interface. On the left, there's a sidebar with navigation links: Favorites, Recent, My folders, Shared folders, Dashboards, and Analyses (which is selected). The main area is titled 'Analyses' and displays four sample analyses: 'Business Review analysis', 'Web and Social Media Anal...', 'Sales Pipeline analysis', and 'People Overview analysis'. Each analysis card includes a 'SAMPLE' button and three dots for more options. In the top right corner, there are buttons for 'Last updated (newest first)', a search bar, and a user icon. A blue box highlights the 'New analysis' button.

15. Click New Data Set.

The screenshot shows the 'Your Data Sets' section of the AWS QuickSight console. It lists four existing data sets: 'Web and Social Media A...' (SPICE), 'Business Review' (SPICE), 'People Overview' (SPICE), and 'Sales Pipeline'. Each data set is represented by a red puzzle piece icon and a 'SPICE' badge. In the top left corner, there is a blue 'New data set' button, which is highlighted with a blue box.

16. On the Create a Dataset page, select Athena as the data source.

The screenshot shows the 'Create a Data Set' page. At the top, there are tabs for 'Data Sets' and 'SPICE capacity (0)'. Below, there are two sections: 'FROM NEW DATA SOURCES' and 'FROM EXISTING DATA SOURCES'. The 'FROM NEW DATA SOURCES' section contains various data source icons: 'Upload a file (.csv, .tsv, .ctf, .elv, .json)', 'Salesforce', 'S3 Analytics', 'S3', 'Athena', 'RDS', 'Redshift (Auto-discovered)', 'Redshift (Manual connect)', 'MySQL', 'PostgreSQL', 'SQL Server', 'Aurora', 'MariaDB', 'Presto', 'Spark', 'Teradata (Provided by Teradata)', 'Snowflake', 'AWS IoT Analytics', 'GitHub', 'Twitter', 'Jira', 'ServiceNow', and 'Adobe Analytics'. The 'Athena' icon is highlighted with a blue box. The 'FROM EXISTING DATA SOURCES' section shows four data sources: 'ticketdata-qs' (updated a minute ago), 'Business Review' (updated 2 hours ago), 'Web and Social Media A...' (updated 2 hours ago), and 'Sales Pipeline' (updated 2 hours ago).

17. For Data source name, type **ticketdata-qs**, then click **Validate connection**.

18. Click **Create data source**.

Lab 3. Consuming data with Athena and Quicksight

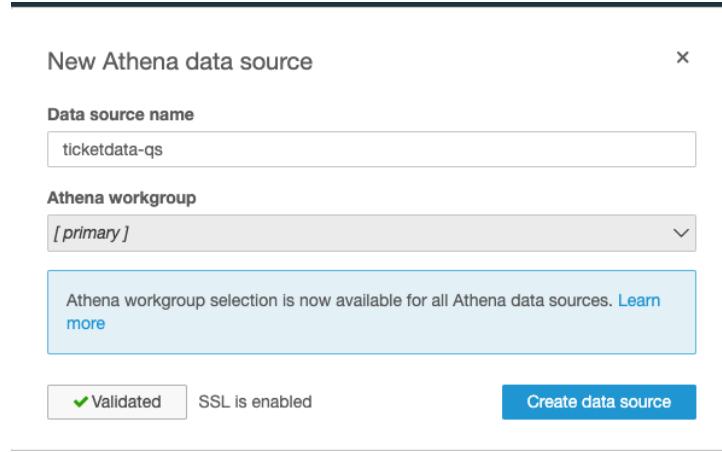
New Athena data source ×

Data source name
ticketdata-qs

Athena workgroup
[primary]

Athena workgroup selection is now available for all Athena data sources. [Learn more](#)

✓ Validated SSL is enabled Create data source



19. In the Database drop-down list, select the database **ticketdata**.
20. Choose the "**sporting_event_ticket_info**" table and click **Select**.

Choose your table ×

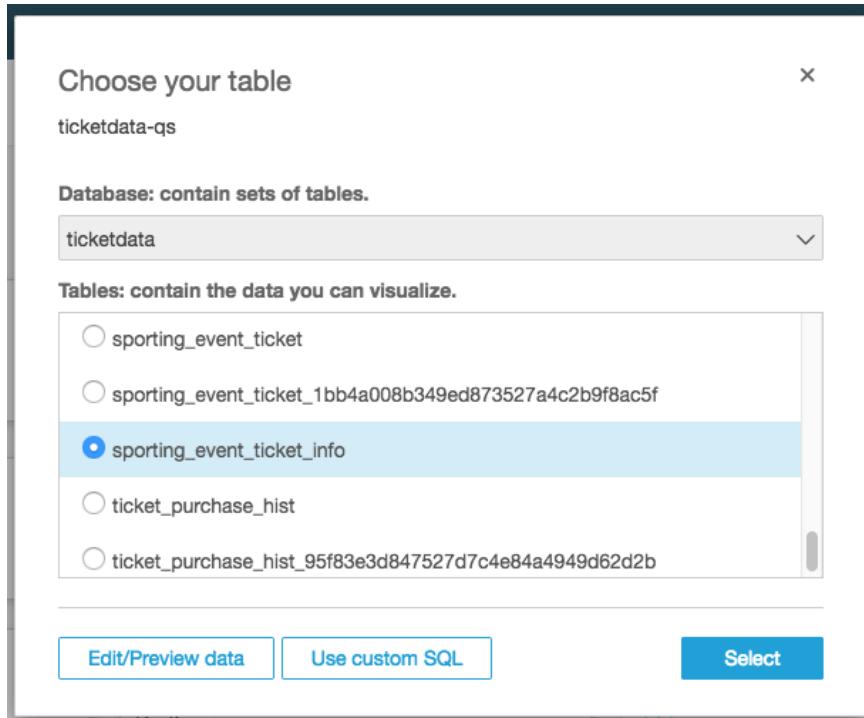
ticketdata-qs

Database: contain sets of tables.
ticketdata

Tables: contain the data you can visualize.

sporting_event_ticket
 sporting_event_ticket_1bb4a008b349ed873527a4c2b9f8ac5f
 sporting_event_ticket_info
 ticket_purchase_hist
 ticket_purchase_hist_95f83e3d847527d7c4e84a4949d62d2b

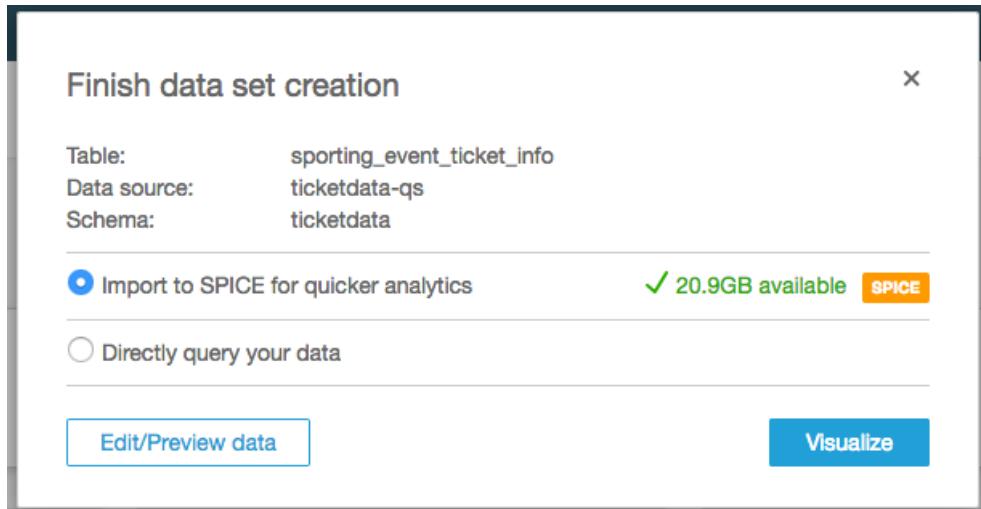
Edit/Preview data Use custom SQL Select



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

If your SPICE has **0 bytes available**, choose the second choice **Directly query your data**

Lab 3. Consuming data with Athena and Quicksight



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

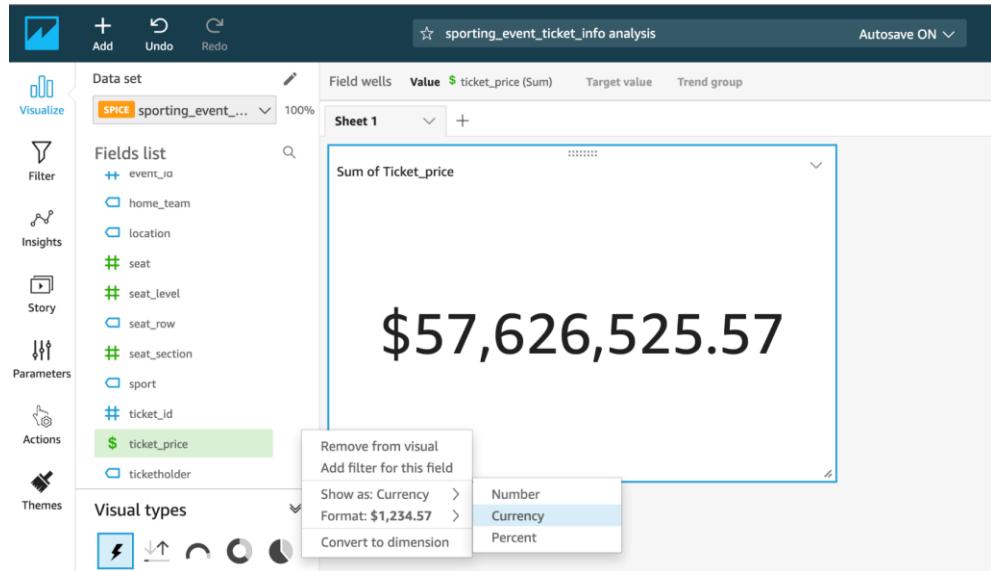
Note: The SPICE dataset will take a few minutes to be built, but you can continue to create some charts on the underlying data.

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 **Show as Currency** to show the number in dollar value.

Lab 3. Consuming data with Athena and Quicksight

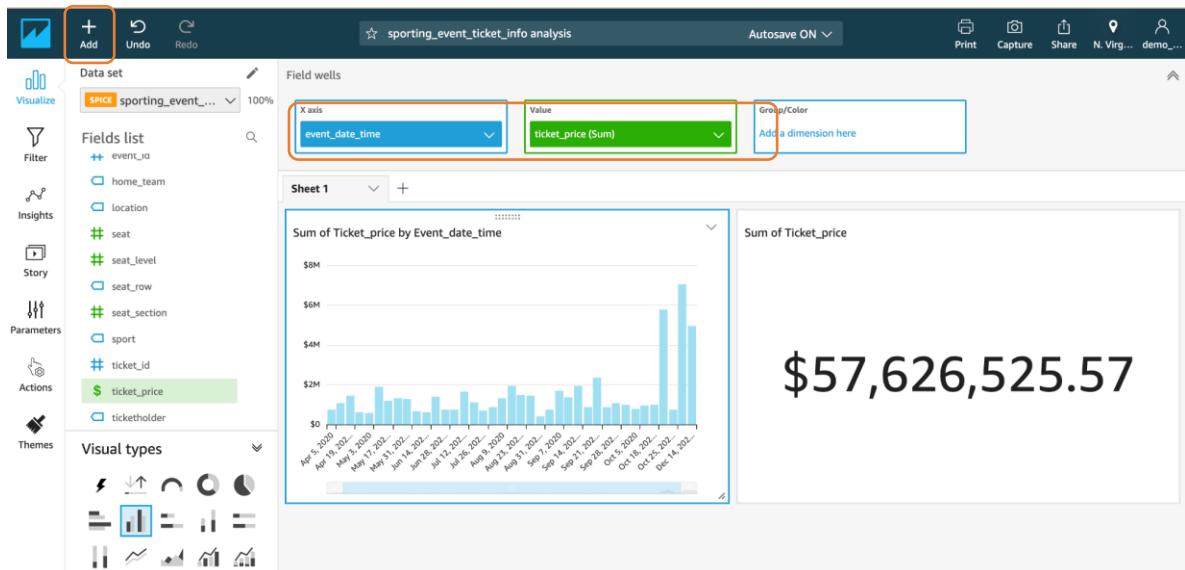


3. You can **add visual** by clicking **Add button** at top left corner of screen.

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

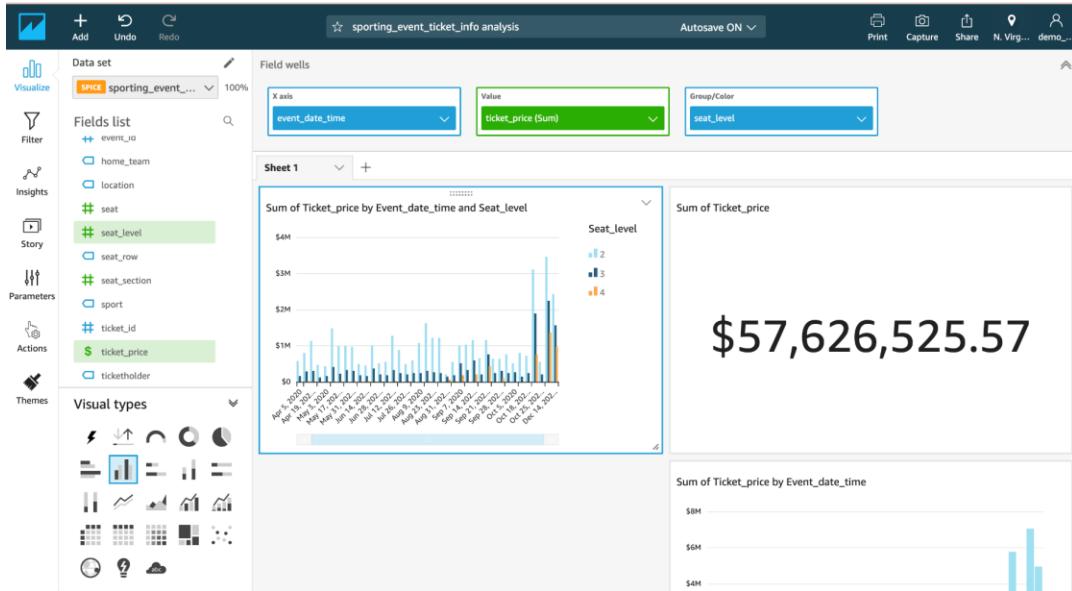
This layout requires a value for the X-axis. In Fields list, select the "**event_date_time**" field and you should see the visualization update.

For Value Y-axis, select "**ticket_price**" from the Field list.



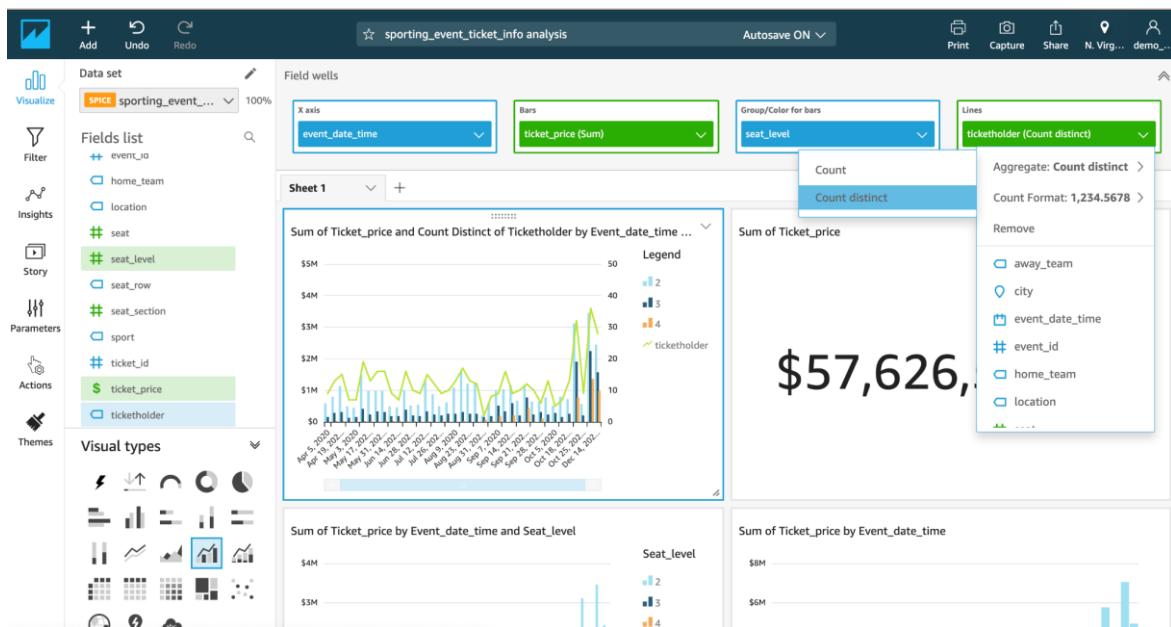
4. You can drag and move other visuals to adjust space in dashboard. In the Fields list, click and drag the **seat_level** field to the **Group/Color** box. You can also use the slider below the x axis to fit all of the data.

Lab 3. Consuming data with Athena and Quicksight



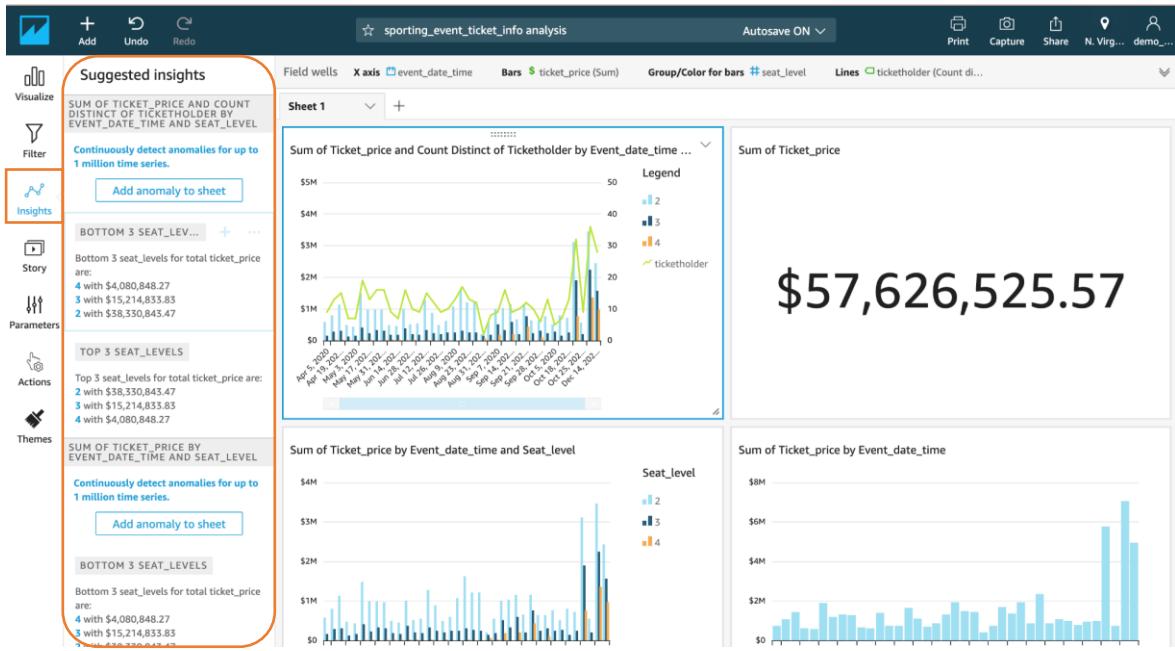
Let's build on this one step further by changing the chart type:

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.
7. In the **Lines** box, click the dropdown box and choose **Aggregate: 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

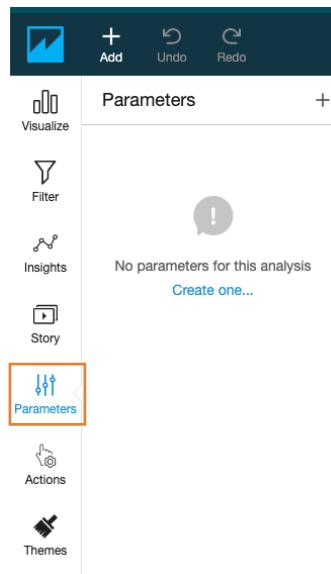


Feel free to experiment with other chart types and different fields to get a sense of the data.

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.

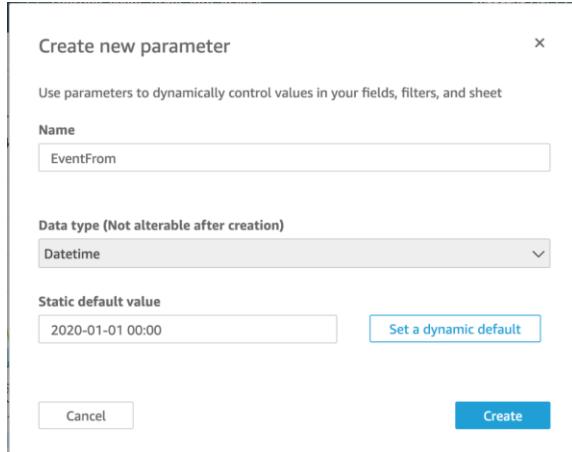
1. In the left navigation menu, select **Parameters**.



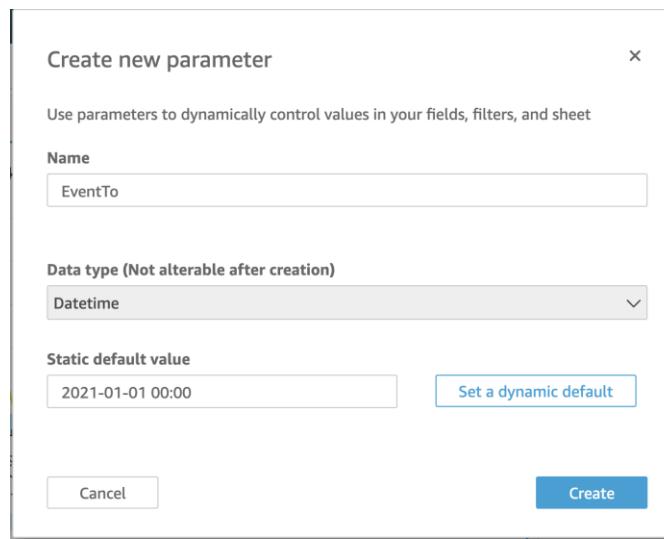
2. Click **Create one** to create a new parameter with a Name.
3. For Name, type **EventFrom**.
4. For Data type, choose **Datetime**.

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, **2019-01-01 00:00**.
6. Click **Create**, and then **close** the Parameter Added dialog box.

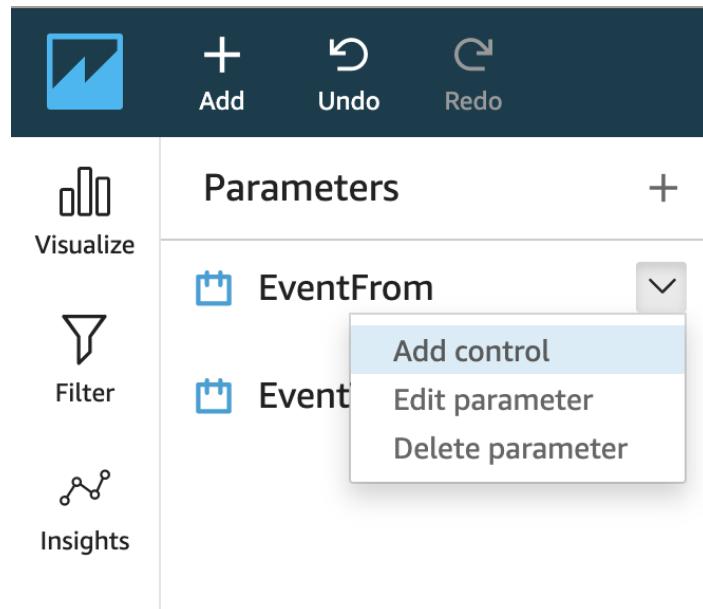


7. Create another parameter with the following attributes:
 - a. **Name:** EventTo
 - b. **Data type:** Datetime
 - c. For Default value, select the value from calendar as end date available in your graph for **event_date_time**. For example, **2021-01-01 00:00**
 - d. Click **Create**



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

The dialog box has a title 'Add control for parameter' and a close button 'x'. It contains the following fields:

- Parameter:** EventFrom
- Display name:** Event From
- Style:** Date picker

At the bottom are 'Cancel' and 'Add' buttons.

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

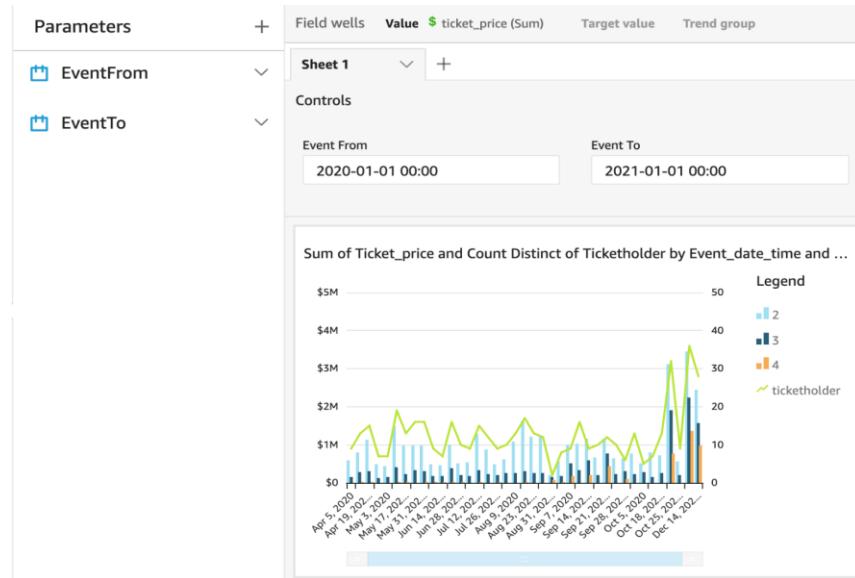
The dialog box has a title 'Add control for parameter' and a close button 'x'. It contains the following fields:

- Parameter:** EventTo
- Display name:** Event To
- Style:** Date picker

At the bottom are 'Cancel' and 'Add' buttons.

Lab 3. Consuming data with Athena and Quicksight

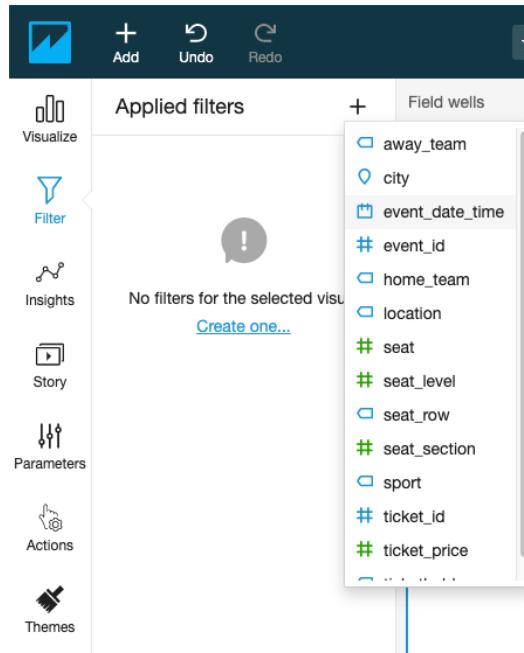
You should now be able to see and expand the Controls section above the chart.



Create a QuickSight Filter

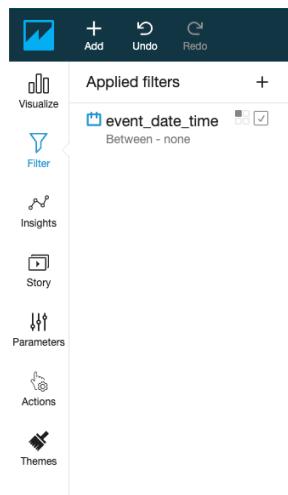
To complete the process, we will wire up a filter to these controls for all visuals.

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



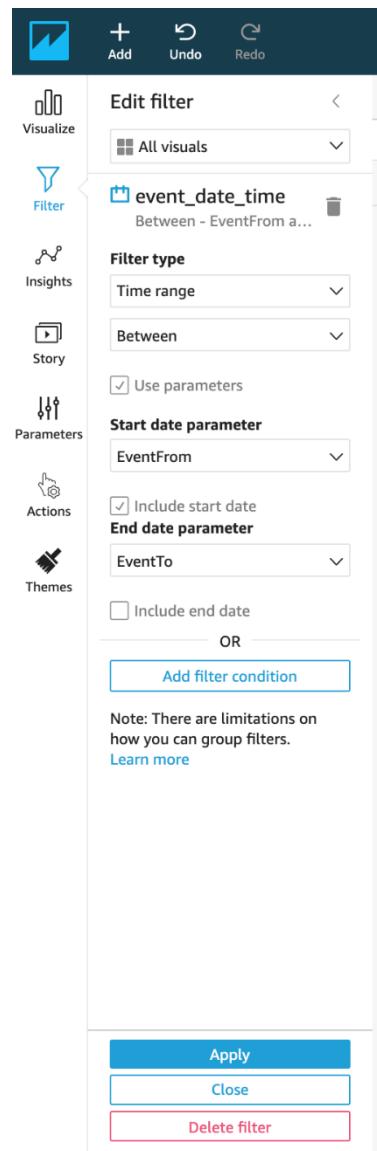
3. Click this filter to **edit** the properties.

Lab 3. Consuming data with Athena and Quicksight



4. For Filter type, choose **Date & Time range** and **Between**.
5. Select option **Use Parameter**, click **Yes** to apply to all visual.
6. For **Start date parameter**, choose **EventFrom**.
7. For **End date parameter**, choose **EventTo**.
8. Click **Apply**.

Lab 3. Consuming data with Athena and Quicksight

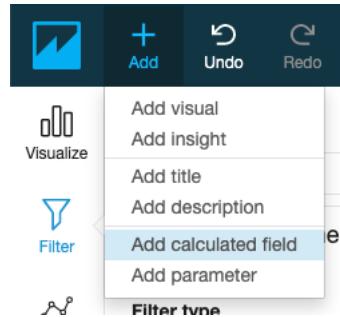


Add Calculated Fields

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**.

Lab 3. Consuming data with Athena and Quicksight



2. Give it a name **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 **Save**.

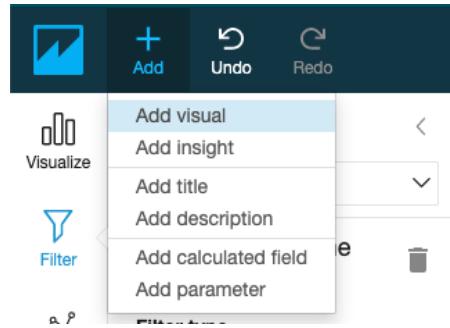
A screenshot of the 'Edit calculated field' dialog box. The title bar says 'Edit calculated field'. The main area shows the field name 'event_day_of_week' and the formula '1 extract('WD',{event_date_time})'. On the right side, there are sections for Fields, Parameters, and Functions. The 'Functions' section is expanded, showing a search bar and a list of functions including 'All', 'abs', 'addDateTime', 'avg', 'avgIf', 'avgOver', and 'ceil'. The 'Save' button at the top right is highlighted with a red box.

5. Add another calculated field with the following attributes:
 - a. Calculated field name: **event_hour_of_day**
 - b. Formula: `extract('HH', {event_date_time})`

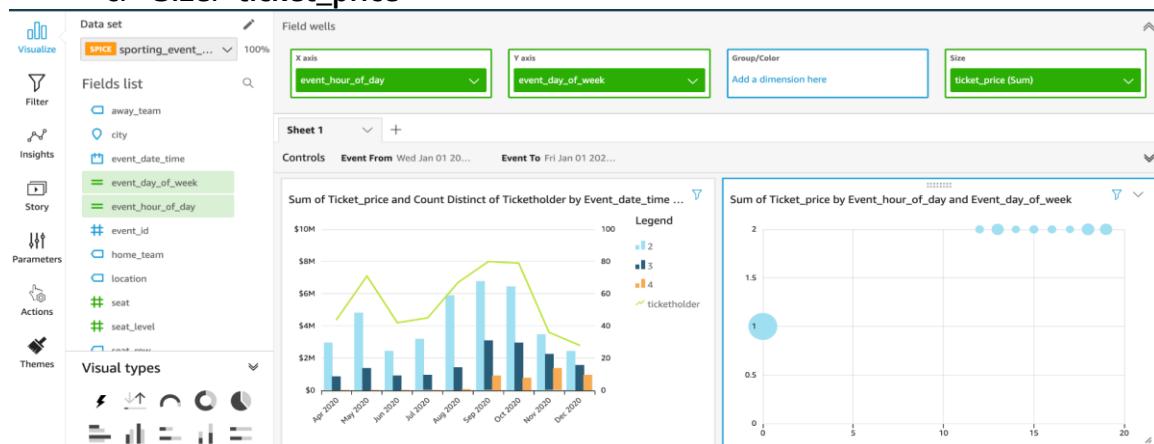
Note: HH: This returns the hour portion of the date.

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

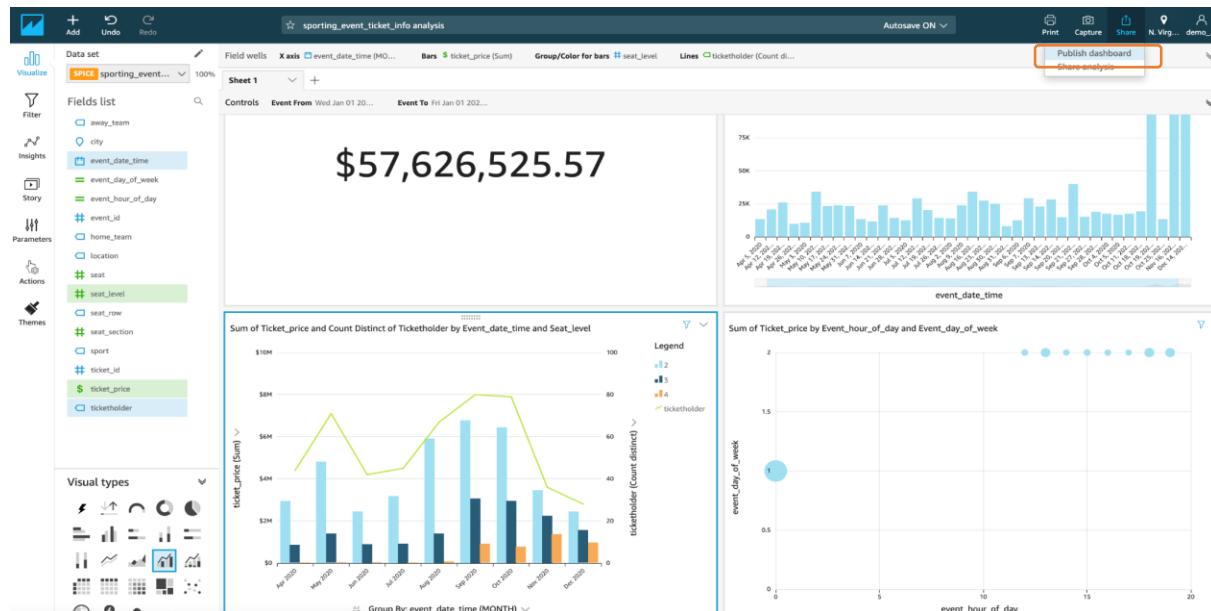
Lab 3. Consuming data with Athena and Quicksight



7. For field type, select the **scatter plot**.
8. In the Fields list, click the following attributes 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 the **Share** menu on the 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 provides 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>

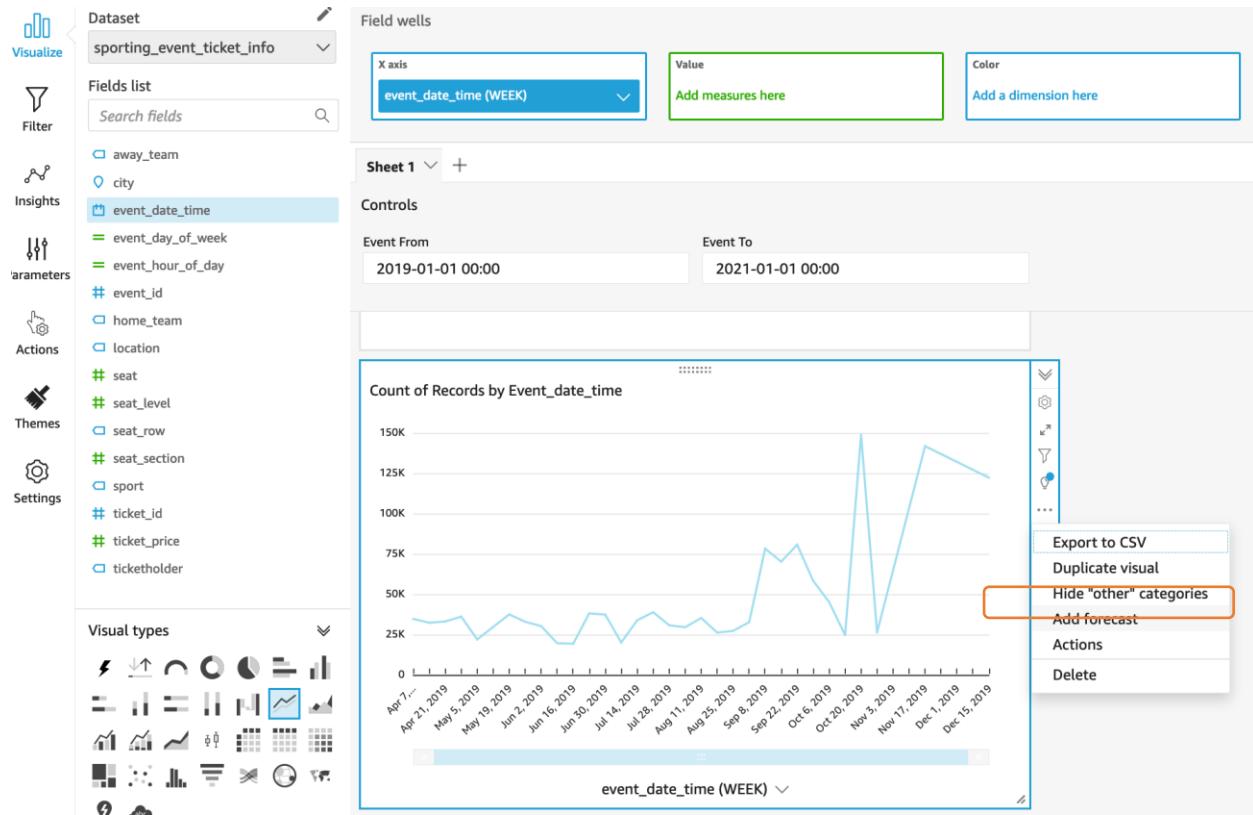
Let's see how we can add a bit of forecasting in our dashboard. Forecasting works with timeseries, which is better represented with a line graph. Let's first create a line graph.

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

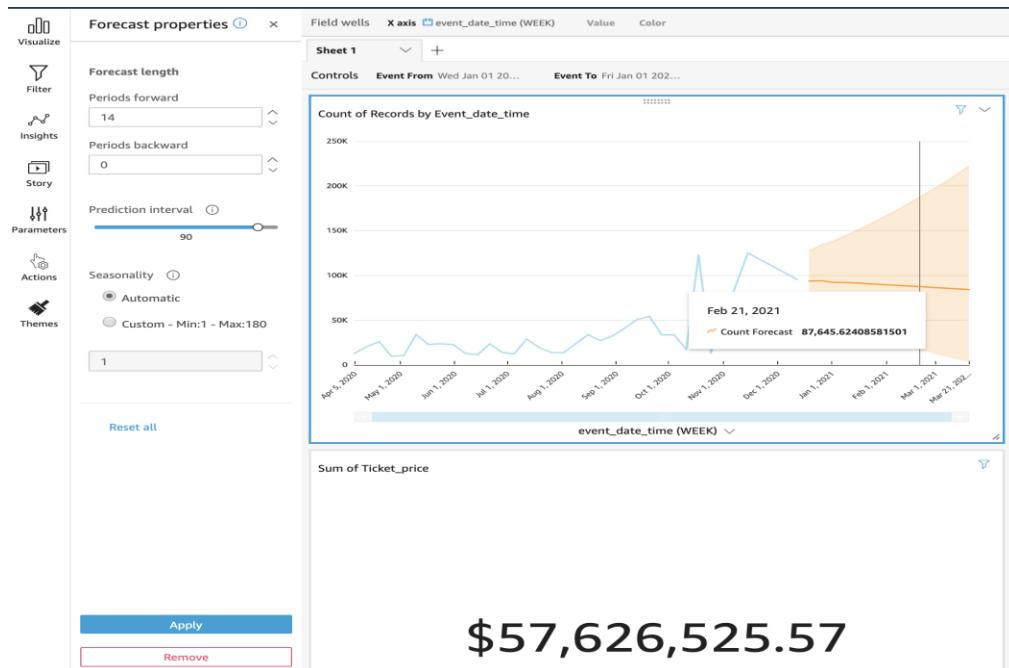
The screenshot shows the QuickSight visual editor interface. On the left, there's a 'Dataset' dropdown set to 'sporting_event_ticket_info' and a 'Fields list' sidebar with various event-related fields like away_team, city, event_date_time, etc. In the center, under 'Field wells', the 'X axis' is set to 'event_date_time (WEEK)' and the 'Aggregate' dropdown is set to 'Week'. The 'Y axis' well has a placeholder 'Add measures here'. On the right, there's a 'Color' section with 'Add a dimension here' and a dropdown menu for time aggregation levels: Year, Quarter, Month, Week (which is selected), Day, Hour, and Minute. Below the dropdown is a large empty area where the chart would be displayed.

Lab 3. Consuming data with Athena and Quicksight

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



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.



Lab 3. Consuming data with Athena and Quicksight

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 before starting the lab. You can go to the [CloudFormation](#) console, choose the oldest stack. 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.

Lab 3. Consuming data with Athena and Quicksight

Outputs (9)

Search outputs

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-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	-

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):

Lab 3. Consuming data with Athena and Quicksight

- i. "workgroupA for BusinessAnalystUser"
- ii. "workgroupB for workgroup manager user"
- iii. **Query result location:** Provide the query location. You can find S3 bucket name from **Cloudformation output** tab with the key name "**S3BucketWorkgroupA**" & "**S3BucketWorkgroupB**".
- iv. For workgroupA, the s3 path would look something like: "s3://xxx-s3bucketworkgroupa-xxx/".
- v. For workgroupB, provide S3 path as: "s3://xxx-s3bucketworkgroupb-xxx/".
- c. For "**Encrypt query results**", leave as default i.e. unchecked. You can check this if you want your query results to be encrypted.
- d. Check the checkbox for "**Metrics: Publish query metrics to AWS CloudWatch**"

Workgroup name* workgroupA

Description workgroupA for BusinessAnalystUser
Use up to 1024 characters.

Query result location and Encryption

Query result location s3://dmslab-student-s3bucketworkgroupa-ldtj4/
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 [i](#)

Settings

Override client-side settings [i](#)

Requester pays S3 buckets Enable queries on requester pays buckets in Amazon S3 [i](#)

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.

Lab 3. Consuming data with Athena and Quicksight

- 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**.

Explore the features of workgroups

1. From the **Outputs** tab of **CloudFormation** console, note down user name **BusinessAnalystUser** and bucket name **S3BucketWorkgroupA** and save it.

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-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-KLF9GDANNNTVZ	workgroup_manager_user for access to Workgroup A and Workgroup B	-
WorkgroupManagerUserPolicy	WorkgroupManagerUserPolicy	User policy for Workgroup manager user	-

2. 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>
3. Next, Open [AWS console log-in](#) different browser, select **IAM user** and login with following credential:
 - a. **AccountID**: <your-account-name-or-alias>
 - b. **IAM User name**: <value copied for **BusinessAnalystUser**>
 - c. Password: **Master123!**

Lab 3. Consuming data with Athena and Quicksight

- d. Make sure the region is: **US East (N. Virginia)**
4. 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:

The screenshot shows the AWS Athena Query Editor interface. The top navigation bar includes 'Services', 'Resource Groups', 'Athena', 'S3', 'AWS Glue', 'RDS', and 'Support'. The 'Athena' tab is selected. The main area has 'Query Editor' tabs: 'Query Editor' (selected), 'Saved Queries', 'History', 'Data sources', and 'Workgroup : workgroupA'. On the left, the 'Data source' dropdown is set to 'AwsDataCatalog' and the 'Database' dropdown is set to 'ticketdata'. Under 'Tables (3)', there are three entries: 'parquet_sport_location', 'parquet_sport_team', and 'parquet_sporting_event'. Under 'Views (1)', there is one entry: 'sporting_event_info', which is highlighted with an orange border. In the center, 'New query 1' contains the SQL query: 'SELECT * FROM "ticketdata"."sporting_event_info" limit 10;'. Below the query are buttons for 'Run query', 'Save as', and 'Create'. A note says 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'. On the right, there is a 'Results' section which is currently empty.

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 of the interface is now set to 'workgroupA', indicated by a red box. The rest of the interface and the query results are the same.

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. At the top, there are buttons for 'Create workgroup', 'View details', and 'Switch workgroup', with 'Switch workgroup' highlighted by a red box. Below is a table of 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)', 'parquet_sport_location' and 'parquet_sport_team' are listed. On the right, a main panel 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, a query editor window is open with 'New query 1' selected. The query code is: `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 the 'Settings' tab highlighted in the top navigation bar. The main area displays the same query editor with the same code and buttons.

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 'Workgroup' dropdown set to 'workgroupA'. Under 'Query result location', the value 's3://dmslab-student-s3bucketworkgroupa-ltj44qkwyle/' is entered. Below it, there's an example placeholder 'Example: s3://query-results-bucket/folder/'. There are also sections for 'Encrypt query results' and 'Autocomplete'. 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. In the top navigation bar, 'Athena' is selected. The main area displays a query in the editor:

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

The results pane shows a table with 10 rows of data from the 'sporting_event_info' view. The columns are: event_id, sport, event_date_time, home_team, away_team, location, and city. The data includes various baseball games and a football game.

- 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**.

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 page. The 'Switch workgroup' button is highlighted with an orange box. The table lists three 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

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

The screenshot shows the AWS Athena Query Editor with an error message in a red box:

Your query has the following error(s):
User: arn:aws:iam::665953140268:lakeformation-BusinessAnalystUser-7H32WD4CWS6 is not authorized to perform: athena:StartQueryExecution on resource: arn:aws:athena:us-east-1:665953140268:workgroupB (Service: AmazonAthena; Status Code: 400; Error Code: AccessDeniedException; Request ID: 40b3397b-f49b-4d1c-b44c-dcaaf47e977)

The query editor shows the same query as before:

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

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 **workgroupA** 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	s3://dmslab-student-s3bucketworkgroupa-ldtj44qkyle/ 
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

Athena Services Resource Groups Workgroup : workgroup

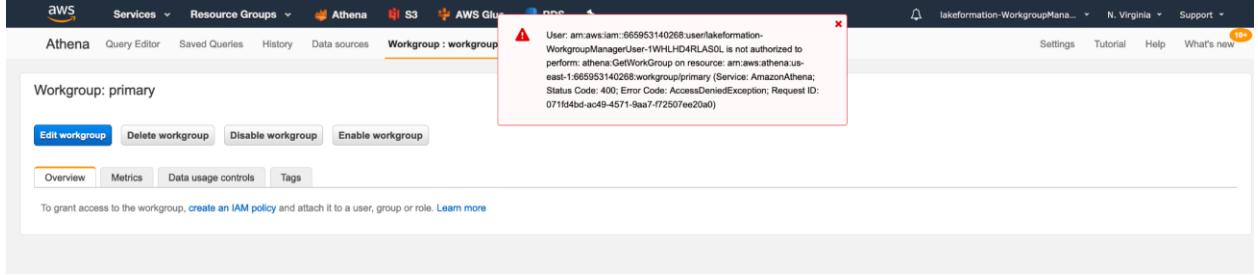
Workgroup: primary

[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.

Athena Services Resource Groups Workgroup : workgroup

New query 1 New query 2 +

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

Run query Save as Create

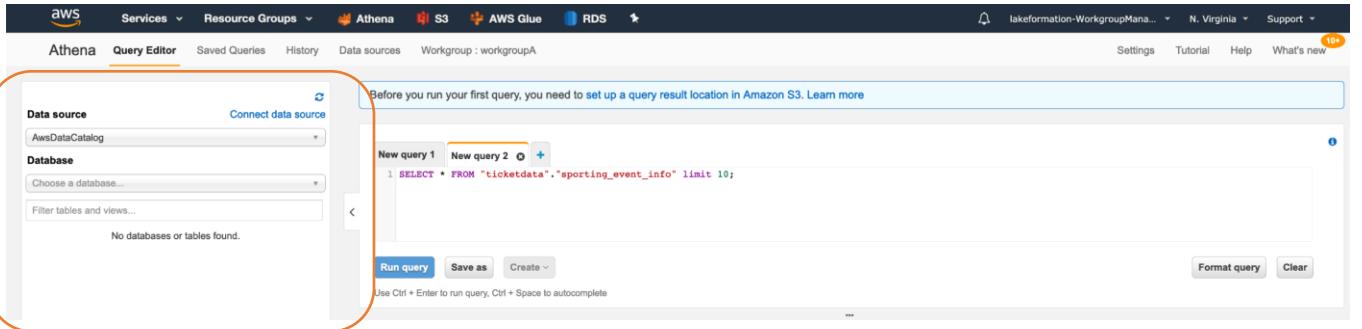
Before you run your first query, you need to [set up a query result location in Amazon S3](#). Learn more

Data source Connect data source

AwsDataCatalog

Database Choose a database...

No databases or tables found.



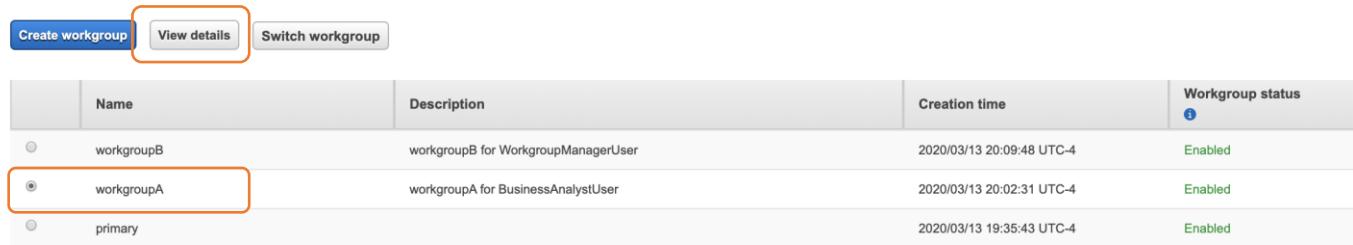
At any point of time, **you can edit, delete and disable your workgroups** as shown:

Select the workgroup and click on "**View Details**".

Lab 3. Consuming data with Athena and Quicksight

Workgroups

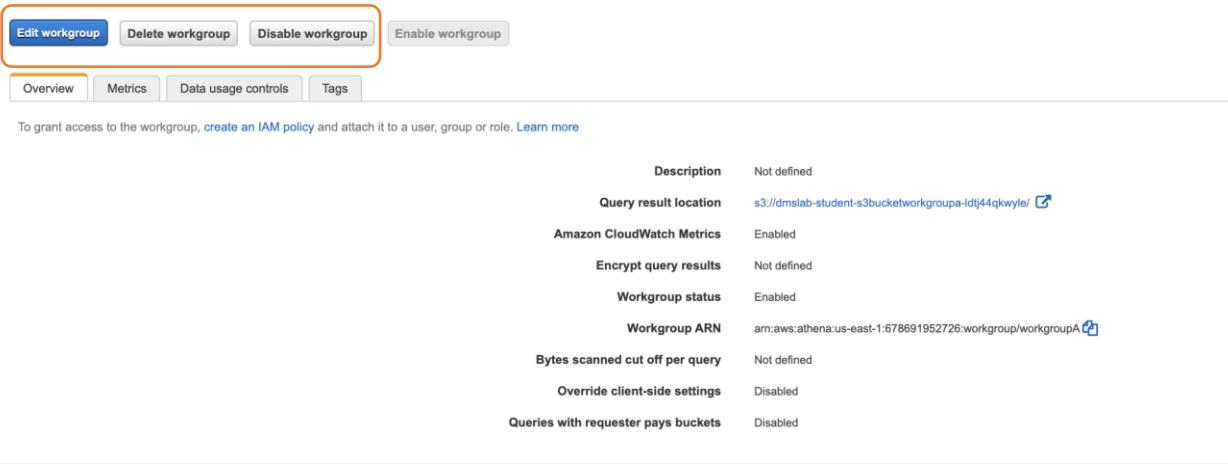
Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)



	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	

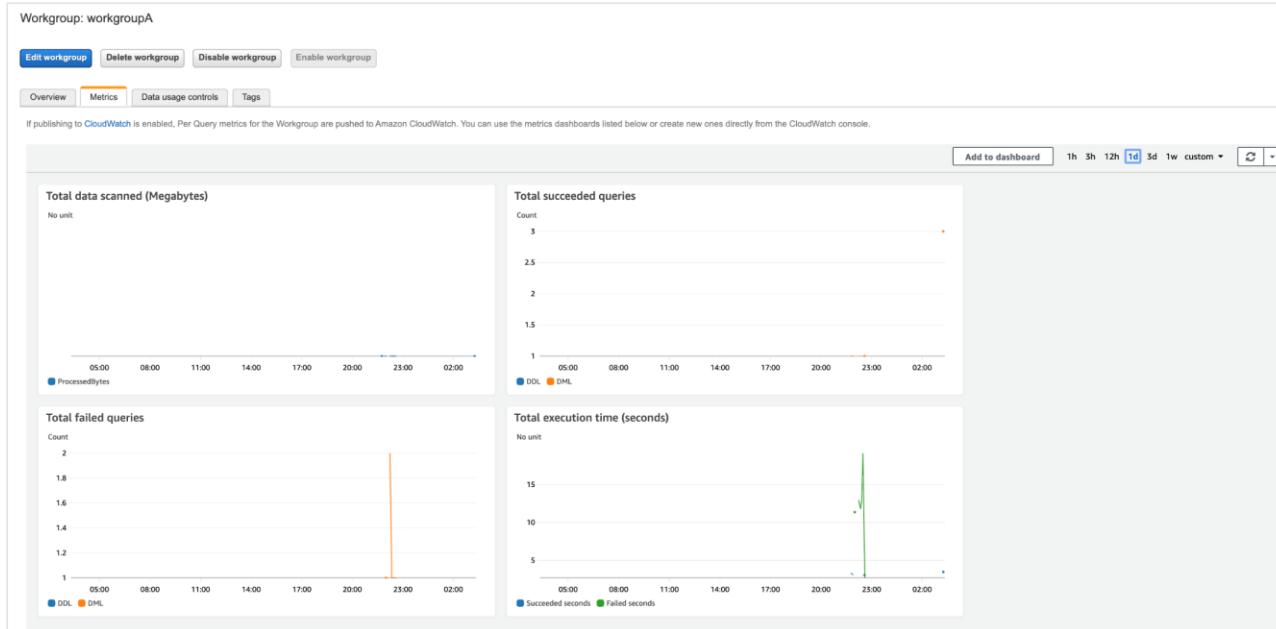
Please Note: For lab purpose, we are attaching policies directly to users. For Best practices, we recommend creating separate groups in IAM for different workgroups and then attaching policies for different workgroups to their respective groups in IAM.

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.



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**

The screenshot shows the AWS Athena Workgroups page. The 'primary' workgroup is selected and highlighted with an orange border. The 'View details' button is also highlighted with an orange border.

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 **10 MB** 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.

Workgroup: primary

Edit workgroup Delete workgroup Disable workgroup Enable workgroup

Overview Metrics Data usage controls Tags

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 Megabytes MB Minimum Limit 10MB per query.

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

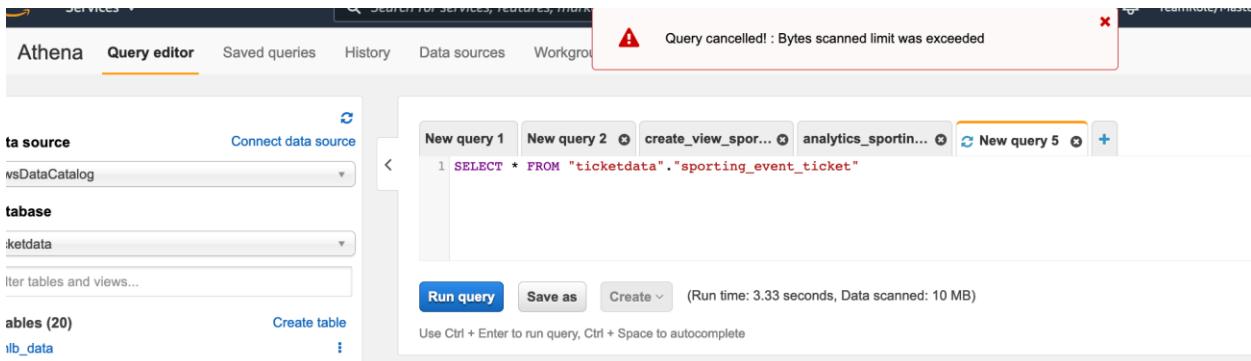
Delete Update



5. Navigate to query editor on Athena console. Run the following query:

```
SELECT * FROM "ticketdata"."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 interface. On the left, there are dropdown menus for 'Data source' (set to 'awsDataCatalog'), 'Database' (set to 'ticketdata'), and 'Tables (20)' (showing 'lib_data'). The main area contains a query editor with the following content:

```
1 SELECT * FROM "ticketdata"."sporting_event_ticket"
```

Below the query, status information is displayed: "Run time: 3.33 seconds, Data scanned: 10 MB". A red alert message at the top right states: "Query cancelled! : Bytes scanned limit was exceeded".

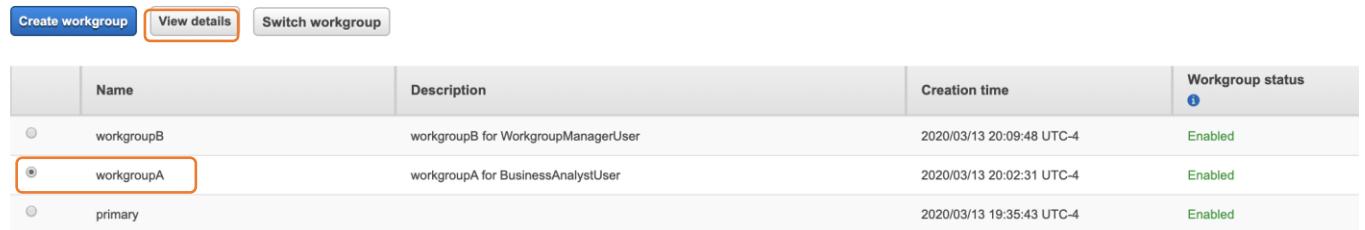
For **per-workgroup data usage control**, you can configure the maximum amount of data scanned by all queries in the workgroup during a specific period. This is useful when you have few analytics reports to run, where you probably have a good idea of how long the process should take and the total amount 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

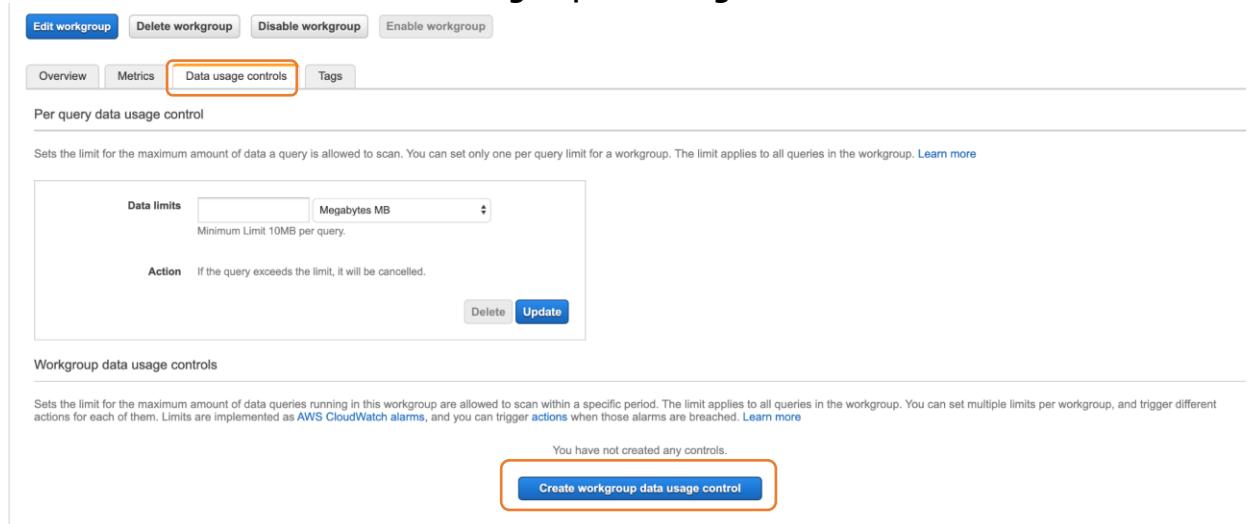
Workgroups

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

- 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 Megabytes MB

Minimum Limit 10MB per query.

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

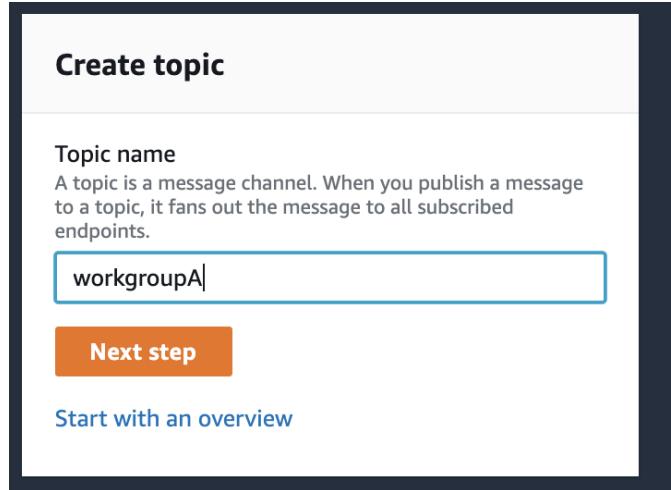
Workgroup data usage controls

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.

- 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.
 - Set **Data Limits** to **10 KBs**
 - Set **Time period** to **1 minute**
 - Set **Action** as "**Send a notification to**". Here, click on **Create SNS Topic**.
 - 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**, then **Create Topic**.
- iii. Note down the topic **ARN number**. Looks like **arn:aws:sns:us-east-1:<accountID>:workgroupA**
- iv. 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' interface. On the left, there's a sidebar with 'Dashboard', 'Topics' (which is selected), and 'Subscriptions'. Under 'Topics', there's a 'Mobile' section with 'Push notifications' and 'Text messaging (SMS)'. The main area shows a topic named 'workgroupA'. The 'Details' tab is selected, showing the following information:
Name: workgroupA
ARN: arn:aws:sns:us-east-1:665953140268:workgroupA
Display name: -
Topic owner: 665953140268
Below the details, there 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 'Create subscription'.

- v. 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

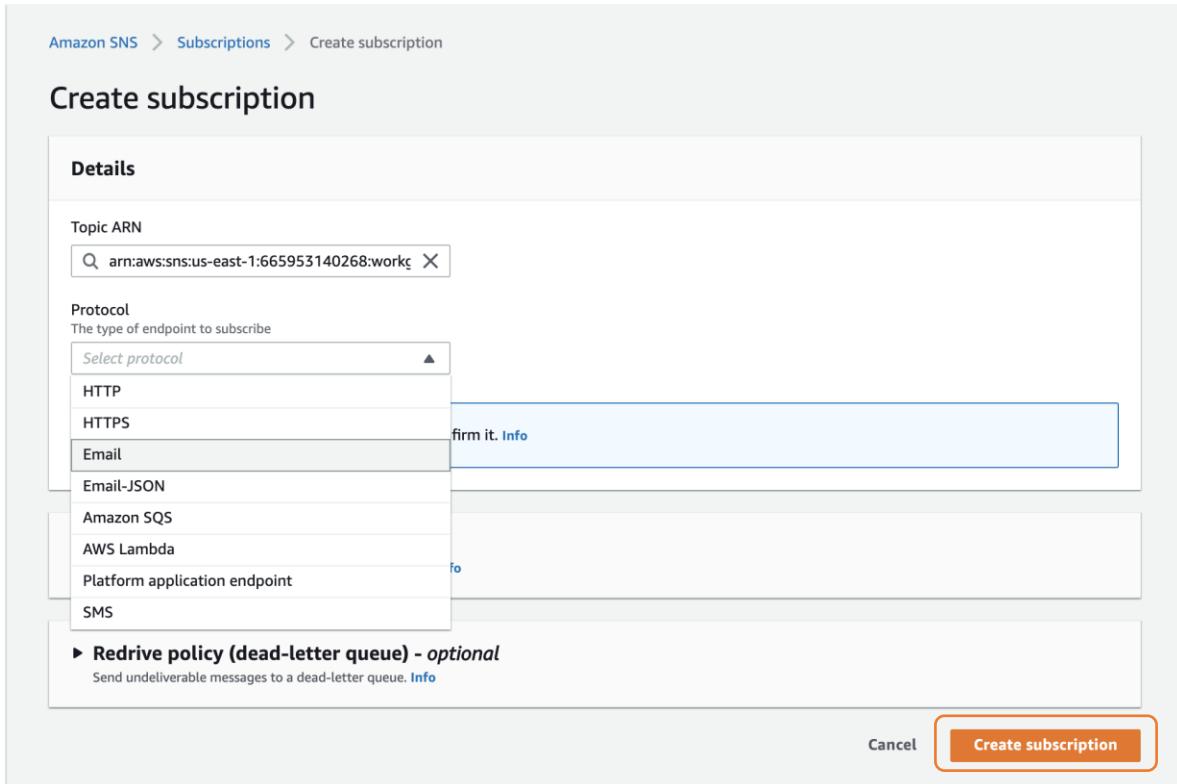
Topic ARN
arn:aws:sns:us-east-1:665953140268:workgroup

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**



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

Create workgroup data usage control

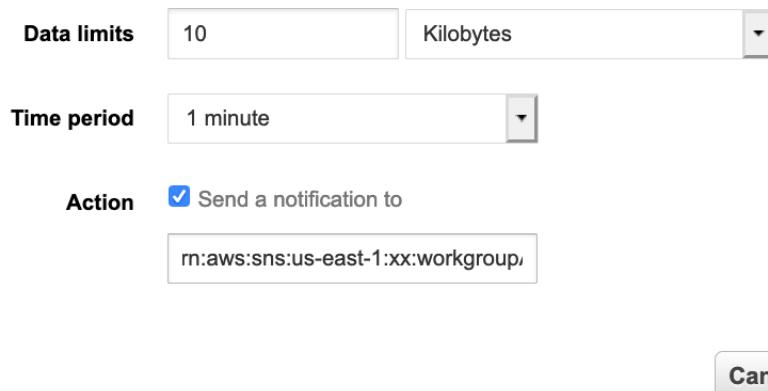
Sets the limit for the maximum amount of data queries running in this workgroup are allowed to scan within a specific period. The limit applies to all queries in the workgroup. You can set multiple limits per workgroup, and trigger different actions for each of them. Limits are implemented as [AWS CloudWatch 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
arn:aws:sns:us-east-1:xx:workgroup/

Cancel **Create**



- viii. Once created, this control will be listed like this:

Lab 3. Consuming data with Athena and Quicksight

Workgroup: workgroupA

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

Overview Metrics Data usage controls Tags

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 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](#)

[Create](#) [Delete](#)

	Data limits	Time period	Action
<input type="radio"/>	10 KB	1 minute	Send notification to topic : arn:aws:sns:us-east-1: workgroupA

4. Go to your email box, click **confirm subscription** in the first notification email:

AWS Notification - Subscription Confirmation

 AWS Notifications <no-reply@sns.amazonaws.com> Today at 5:58

You have chosen to subscribe to the topic:
arn:aws:sns:us-east-1: [REDACTED] workgroupA

To confirm this subscription, click or visit the link below (If this was in error no action is necessary):
[Confirm subscription](#)

Please do not reply directly to this email. If you wish to remove yourself from receiving all future SNS subscription confirmation requests please send an email to [sns-opt-out](#)

5. 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";
```

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

Lab 3. Consuming data with Athena and Quicksight

ALARM: "AWS_Athena_Workgroup_workgroupA_c0ff968d-32fe-4c37-b741-fa45a61..." in US East (N. Virgi...

AWS Notifications <no-reply@sns.amazonaws.com>

AN

Show Details

You are receiving this email because your Amazon CloudWatch Alarm "AWS_Athena_Workgroup_workgroupA_c0ff968d-32fe-4c37-b741-fa45a61585d1" in the US East (N. Virginia) region has entered the ALARM state, because "Threshold Crossed: 1 out of the last 1 datapoints [12665.0 (30/01/20 05:01:00)] was greater than the threshold (10240.0) (minimum 1 datapoint for OK -> ALARM transition)." at "Thursday 30 January, 2020 05:03:03 UTC".

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

The screenshot shows the CloudWatch Alarms Overview page. On the left, a sidebar lists various CloudWatch services: Dashboards, Alarms (with 1 new), INSUFFICIENT, OK, Billing, Logs, Log groups, Insights, Metrics, Events, Rules, Event Buses, and ServiceLens. The main area displays 'Alarms by AWS service' with a table. The table has columns: Services, Status, Alarm, Insufficient, and OK. It lists six services: AWS/Athena (Status: 1, Alarm: 1, Insufficient: -, OK: -), AWS/DMS (Status: -, Alarm: -, Insufficient: -, OK: -), CloudWatch Events (Status: -, Alarm: -, Insufficient: -, OK: -), CloudWatch Logs (Status: -, Alarm: -, Insufficient: -, OK: -), and EC2 (Status: -, Alarm: -, Insufficient: -, OK: -). To the right, a panel titled 'Recent alarms' shows a chart for 'AWS_Athena_Workgroup_wor...'. The chart tracks 'Bytes' over time from 03:00 to 05:00. The data points are: 12.7k (blue dot), 10.8k (red dot, labeled 'ProcessedBytes > 10240 for 1 datapoin...'), and 8.85k (blue dot). A red box highlights the 10.8k point.

8. 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**.

Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Workgroup configuration interface. At the top, there are navigation links: Athena, Query Editor, Saved Queries, History, AWS Glue Data Catalog, Workgroup : teamA (circled in red), Settings, and Tutorial. Below this, the workgroup name 'teamA' is displayed. A row of buttons includes 'Edit workgroup' (circled in red with number 2), 'Delete workgroup', 'Disable workgroup', and 'Enable workgroup'. A horizontal menu bar contains 'Overview', 'Metrics', 'Data usage controls', and 'Tags' (circled in red with number 3). A note below states: 'You can add up to 50 tags for each workgroup. You can edit tag keys and values, and you can remove tags from a workgroup at any time. Tag keys and values are case-sensitive. For each tag, a tag optional. Do not use duplicate tag keys in the same workgroup.' A 'Learn more' link is provided. On the right, a 'Manage tags' button is circled in red with number 4. A search bar labeled 'Search tags' is present. A table header with columns 'Key' and 'Value' is shown.

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