



# **Amazon Web Services**

## **Data Engineering Immersion Day**

---

Lab 3. Consuming data with Athena and Quicksight

***July 2021***

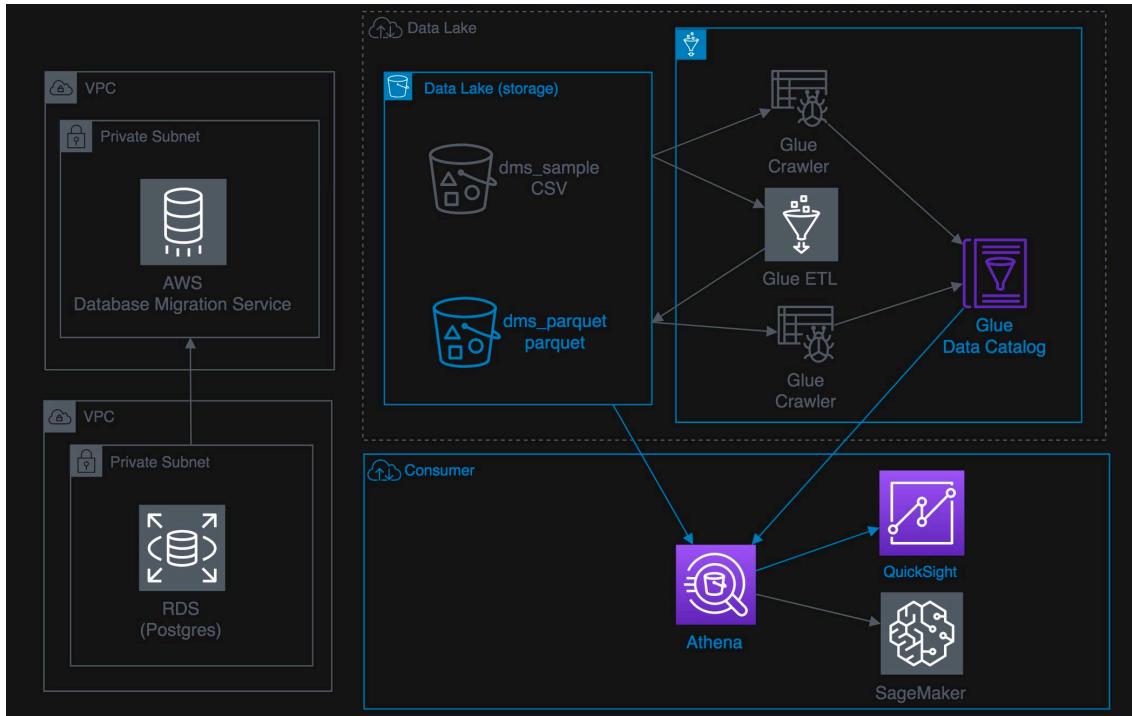
## Table of Contents

<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 .....	18
Create QuickSight Parameters .....	21
Create a QuickSight Filter.....	24
Add Calculated Fields.....	26
<i>Amazon QuickSight ML-Insights (Optional)</i> .....	29
<i>(Optional)Athena Workgroups to Control Query Access and Costs</i> .....	31
Workflow setup to separate workloads .....	31
Explore the features of workgroups .....	34
Managing Query Usage and Cost .....	41
Cost Allocation Tags.....	48

## 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/>

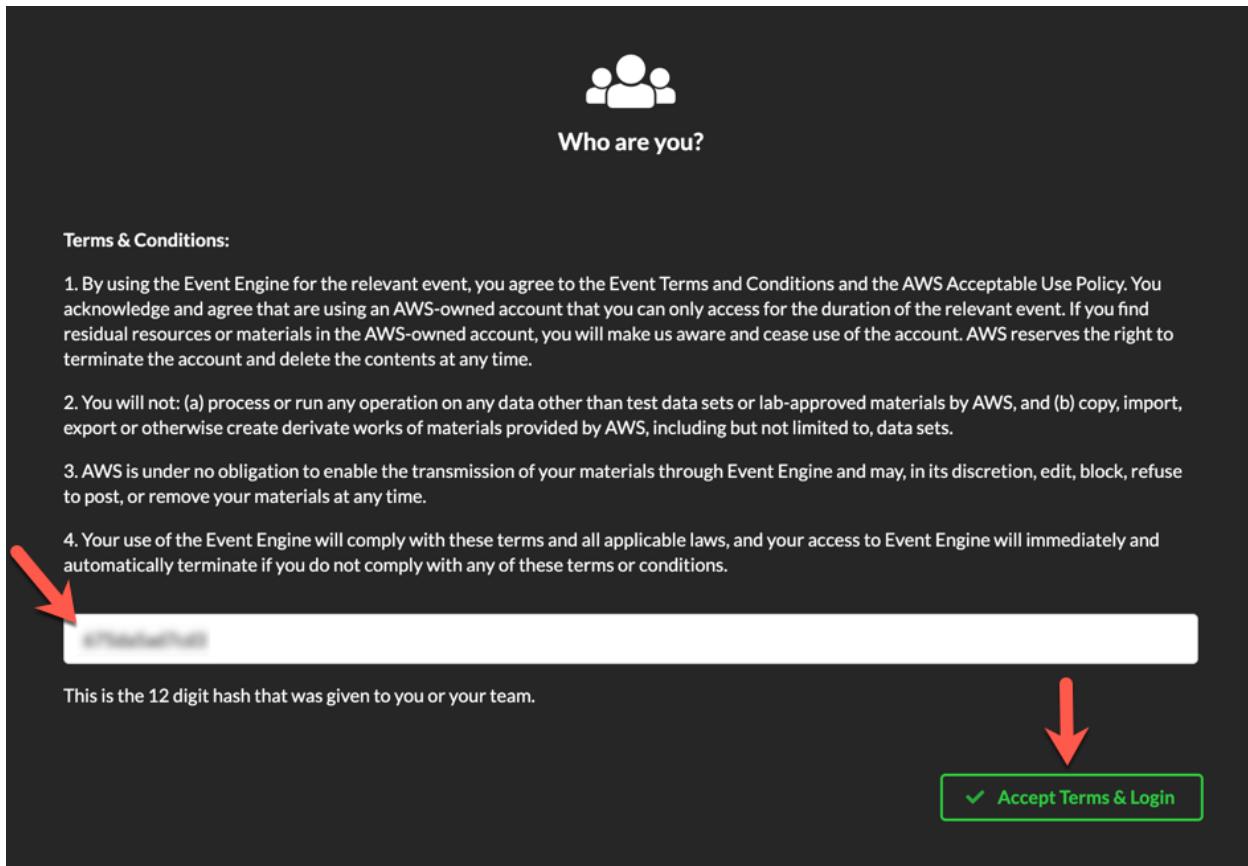
### Get Started Using the Lab Environment

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 "Modules" interface with a "Environment Setup" tab selected. The "Outputs:" section lists several AWS resources with their 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]

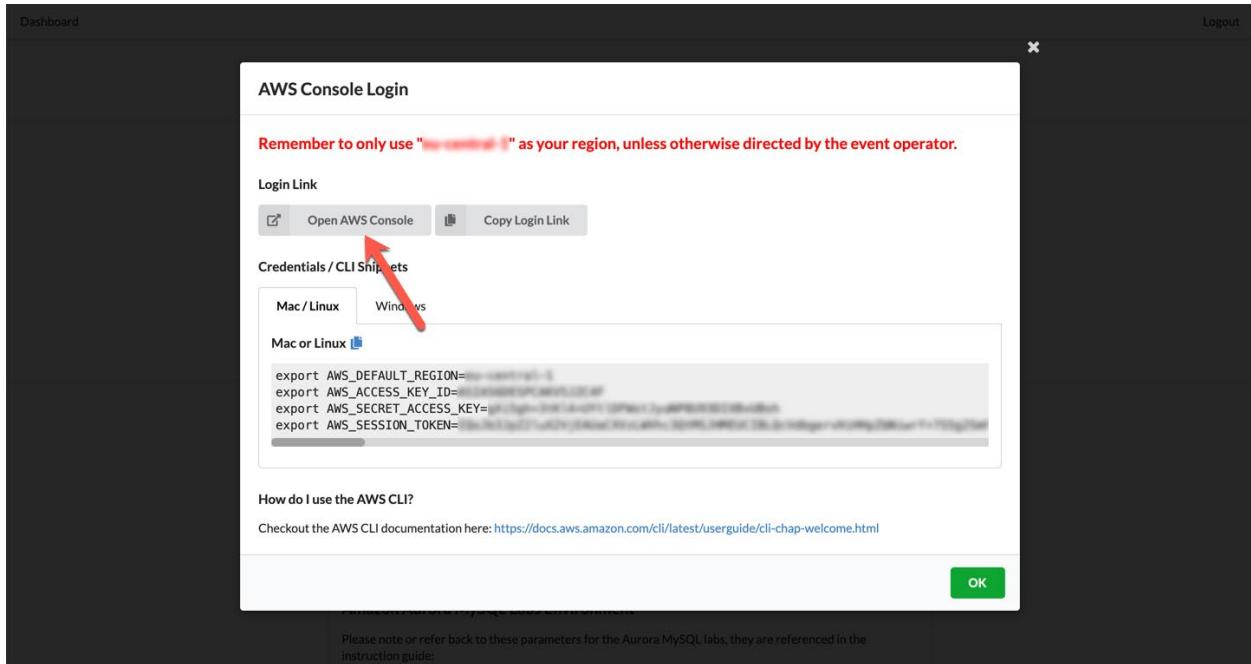
3. On the Team Dashboard, please click AWS Console to log into the AWS Management Console:

The screenshot shows the "Team Dashboard" with an "Event" section. An arrow points from the "Event" section down to two buttons: "AWS Console" and "SSH Key". Below the buttons, 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

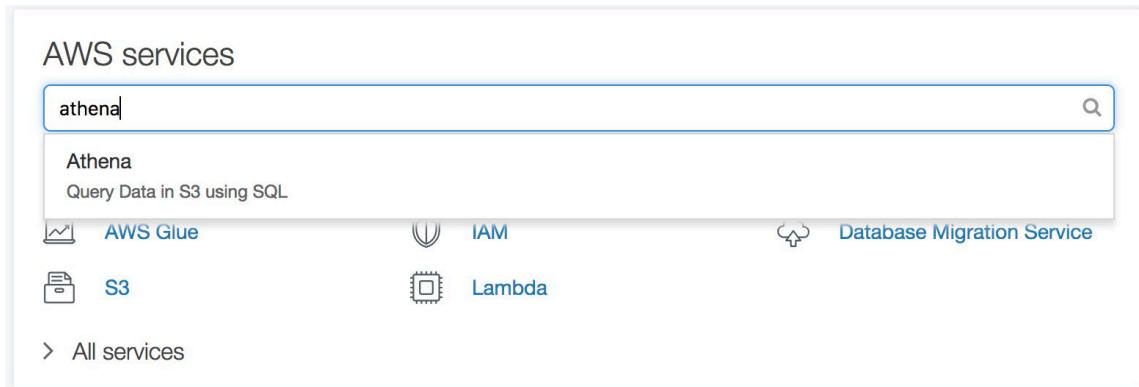


Once you have completed these steps, you can continue with the rest of this lab.

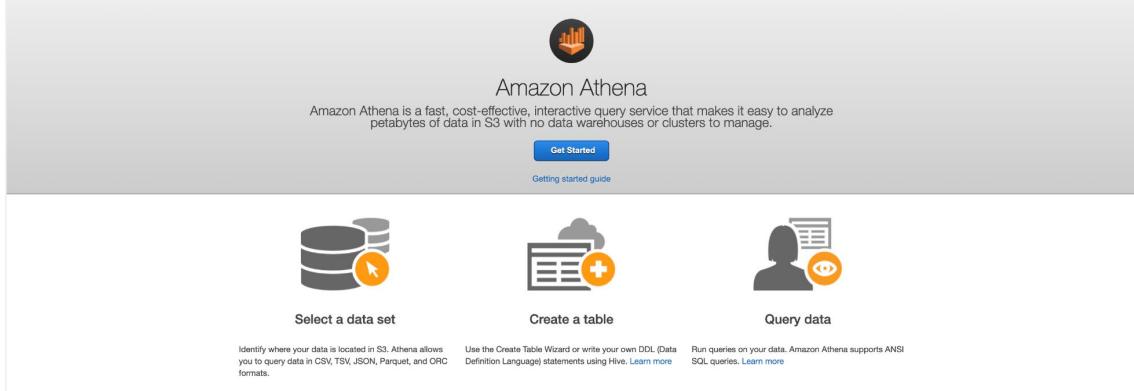
## Lab 3. Consuming data with Athena and Quicksight

### Query Data with Amazon Athena

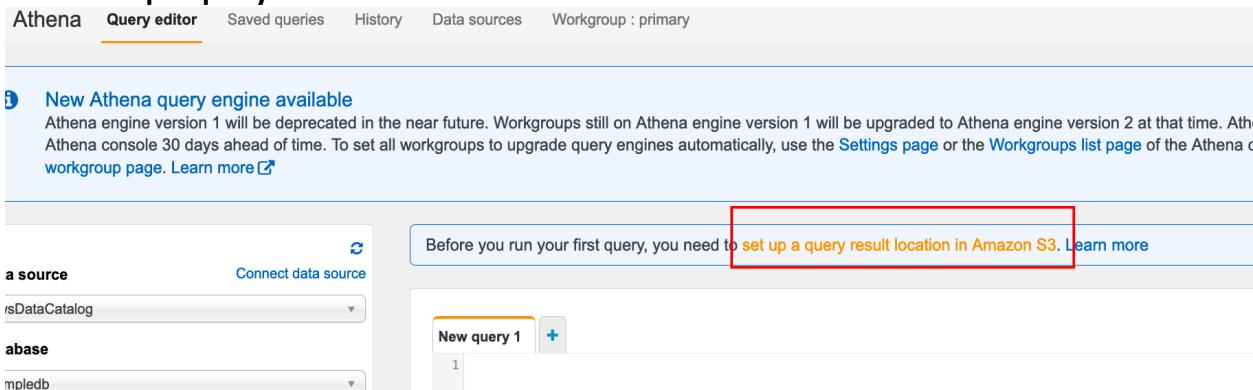
1. In the AWS services console, search for **Athena**.



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

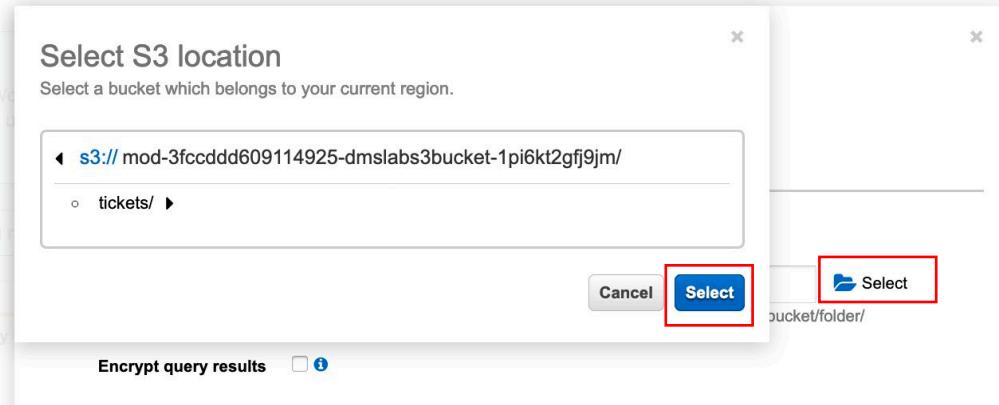


3. Click “**Setup a query result location in Amazon S3**”.

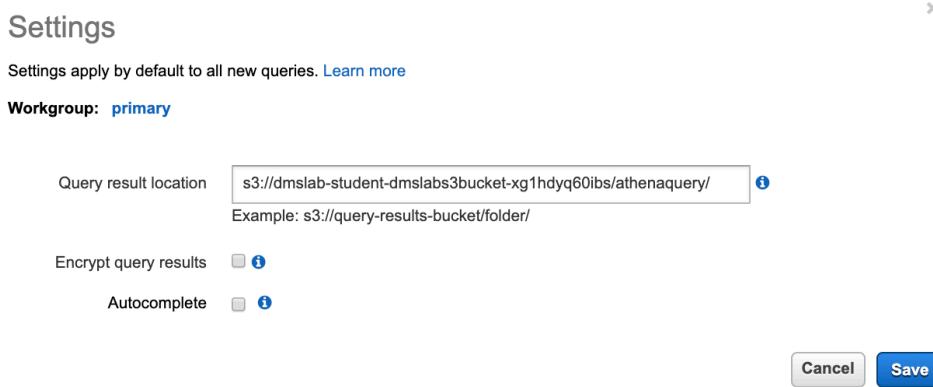


4. It navigates to a **Settings** page. Click on “**Select**” folder icon, choose the **dmslabs3bucket** (e.g: <dmslab-student-dmslabs3bucket-xg1hdq60ibs>). then click on “**Select**” button.

## Lab 3. Consuming data with Athena and Quicksight



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



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

7. 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)**.

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.

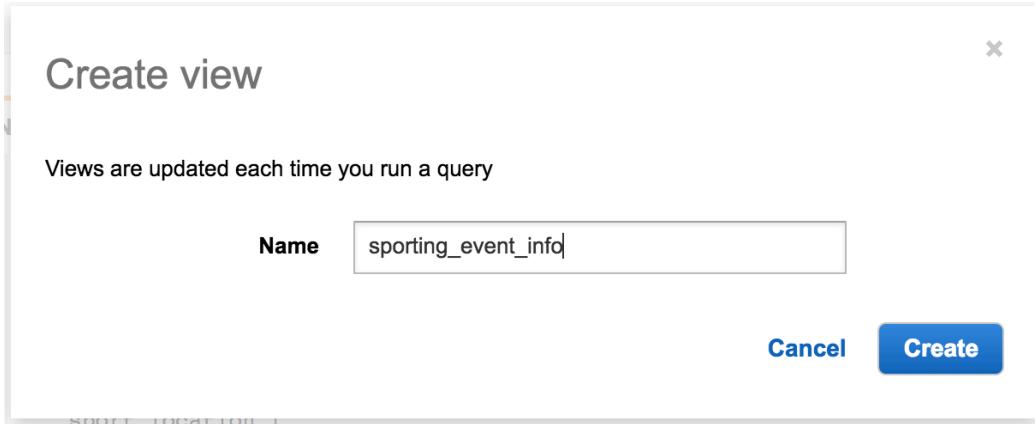
The screenshot shows the AWS QuickSight interface. At the top, there's a 'New query 1' tab and a '+' button. Below it is the SQL code for the query. At the bottom of the editor, there are buttons for 'Run query', 'Save as', 'Create', and 'Format query'. It also shows run time and data scanned information. To the right, there are buttons for 'Athena engine version 1' and 'Release versions'. Below the editor is a 'Results' section with a table showing six rows of data from the query.

	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

```
CREATE OR REPLACE VIEW "sporting_event_info" AS
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
```

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
```

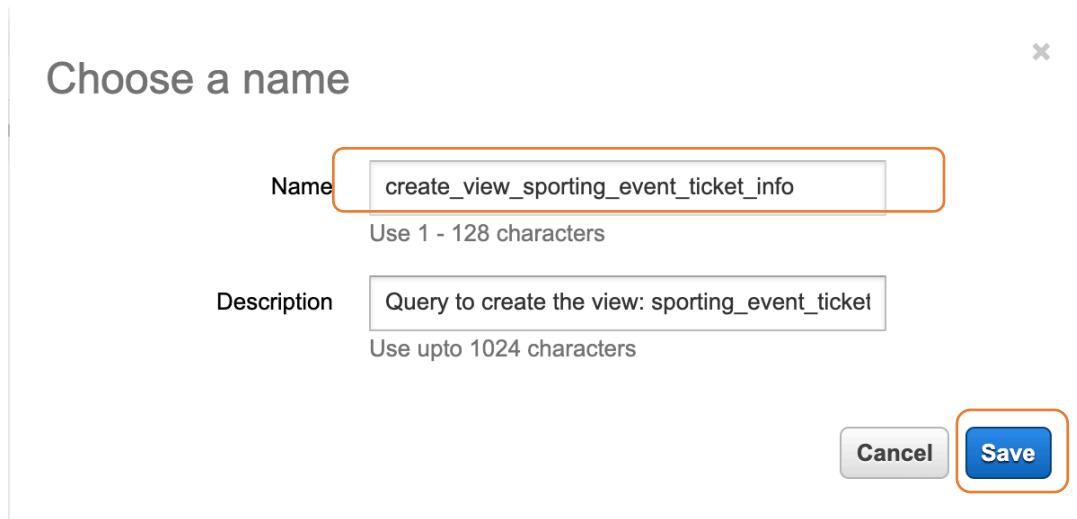
The screenshot shows the AWS Athena Query Editor interface. At the top, there are three tabs: "New query 1", "New query 2", and "New query 3" (which is currently selected and highlighted with an orange border). Below the tabs is a code editor containing the provided SQL query. The code is numbered from 1 to 21. The query selects various columns from three tables: sporting\_event\_info, parquet\_sporting\_event\_ticket, and parquet\_person, joining them on event\_id and ticketholder\_id respectively. A WHERE clause filters the results to only include rows where the sporting\_event\_id matches the event\_id and the ticketholder\_id matches the id in the parquet\_person table. At the bottom of the editor, there are three buttons: "Run query" (blue), "Save as" (grey), and "Create" (grey dropdown). To the right of these buttons, it says "(Run time: 21.04 seconds, Data scanned: 139.22 MB)". Below 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
21
```

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

```
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.

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
```

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

Use Ctrl + Enter to run current query    ...    Athena engine ver

Create table from query  
Create view from query

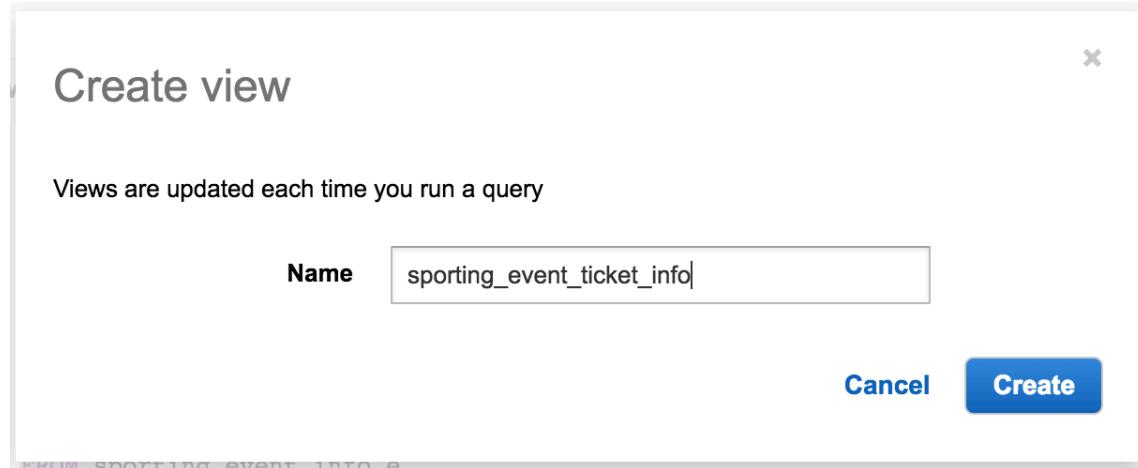
Results

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**.

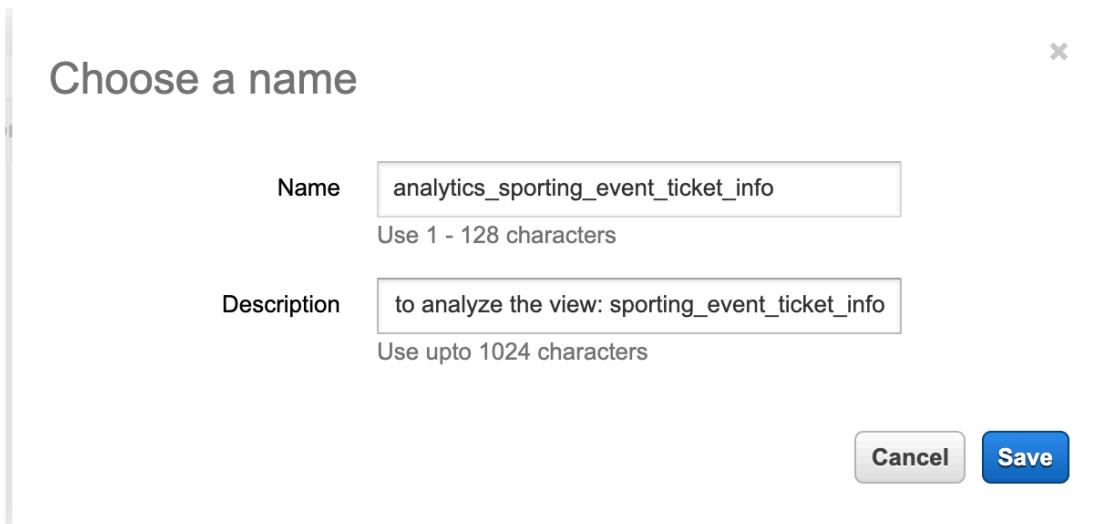
## Lab 3. Consuming data with Athena and Quicksight



15. Copy the following SQL syntax into the New Query 4 tab.

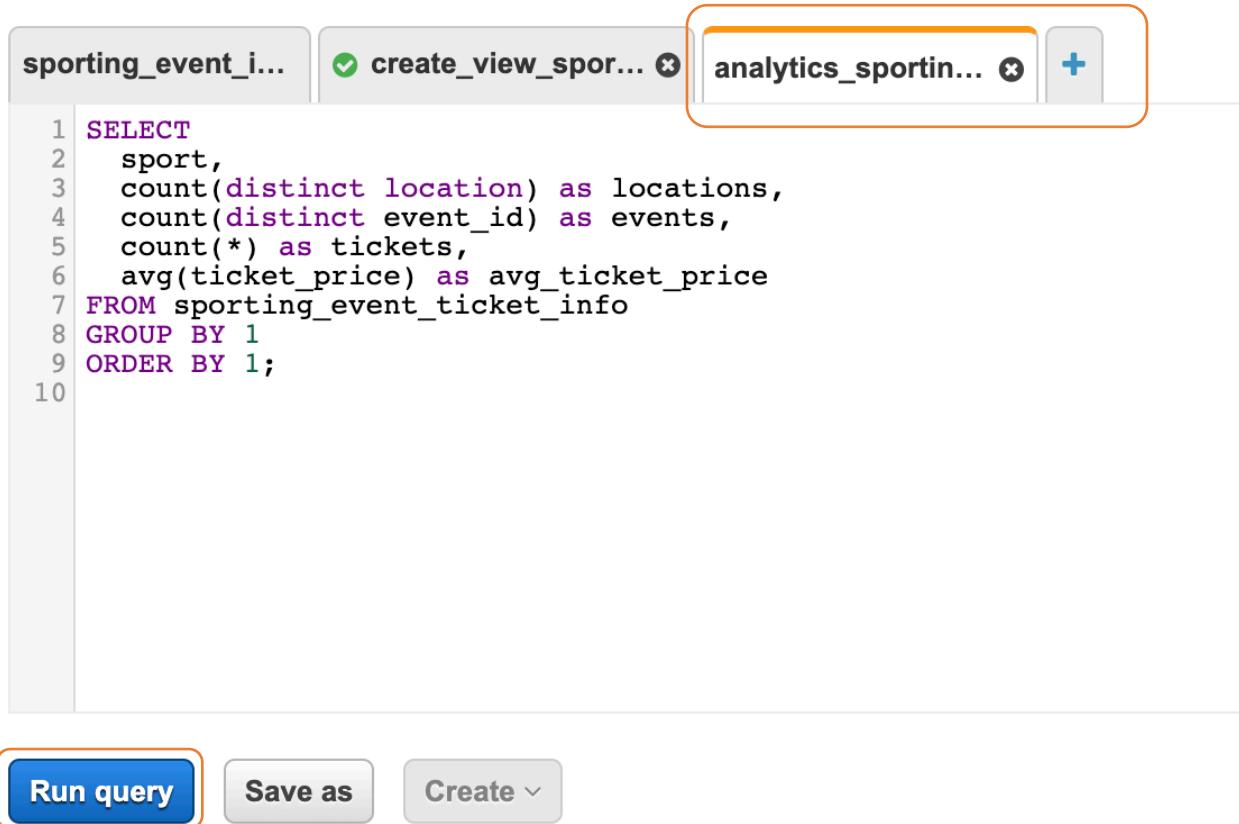
```
SELECT
  sport,
  count(distinct location) as locations,
  count(distinct event_id) as events,
  count(*) as tickets,
  avg(ticket_price) as avg_ticket_price
FROM sporting_event_ticket_info
GROUP BY 1
ORDER BY 1;
```

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**.

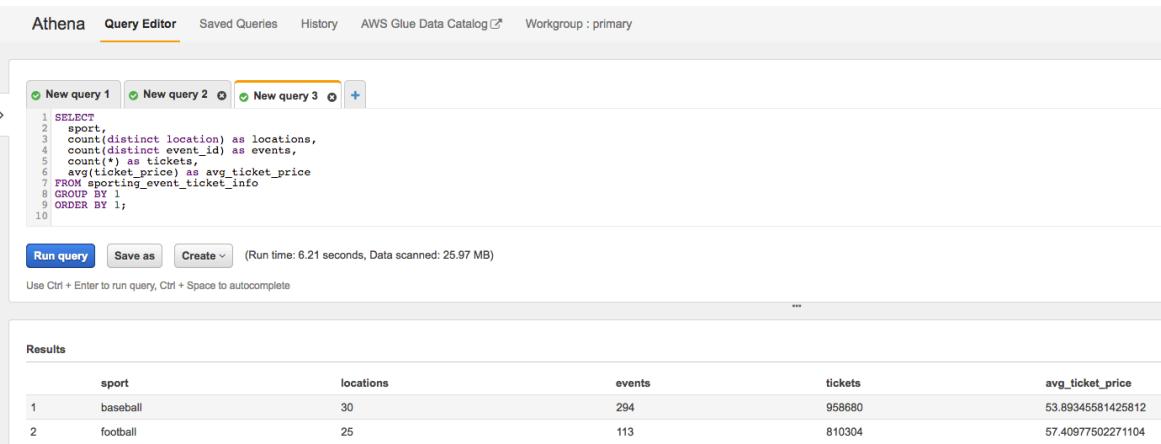
## Lab 3. Consuming data with Athena and Quicksight



```
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;
```

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.



Athena    Query Editor    Saved Queries    History    AWS Glue Data Catalog    Workgroup : primary

```
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;
```

Run query    Save as    Create ▾    (Run time: 6.21 seconds, Data scanned: 25.97 MB)

Results

	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:

## Lab 3. Consuming data with Athena and Quicksight

Amazon S3 > dmslab-student-dmslabs3bucket-xg1hdq60ibs > athenaquery

dmslab-student-dmslabs3bucket-xg1hdq60ibs

Overview

Q Type a prefix and press Enter to search. Press ESC to clear.

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

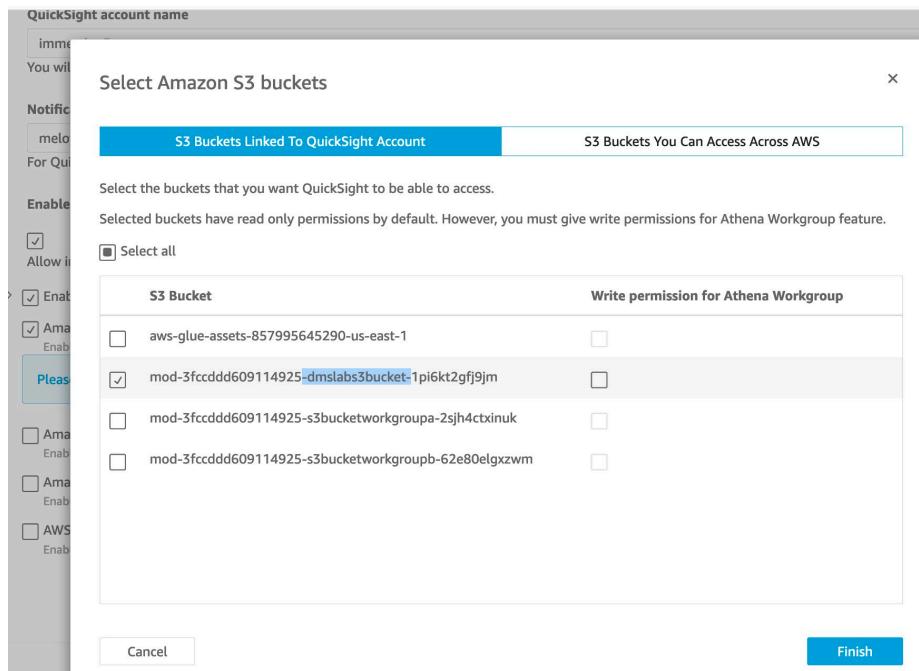
3. For account type, choose the default **Enterprise** Version.
4. Click **Continue**.

## Lab 3. Consuming data with Athena and Quicksight

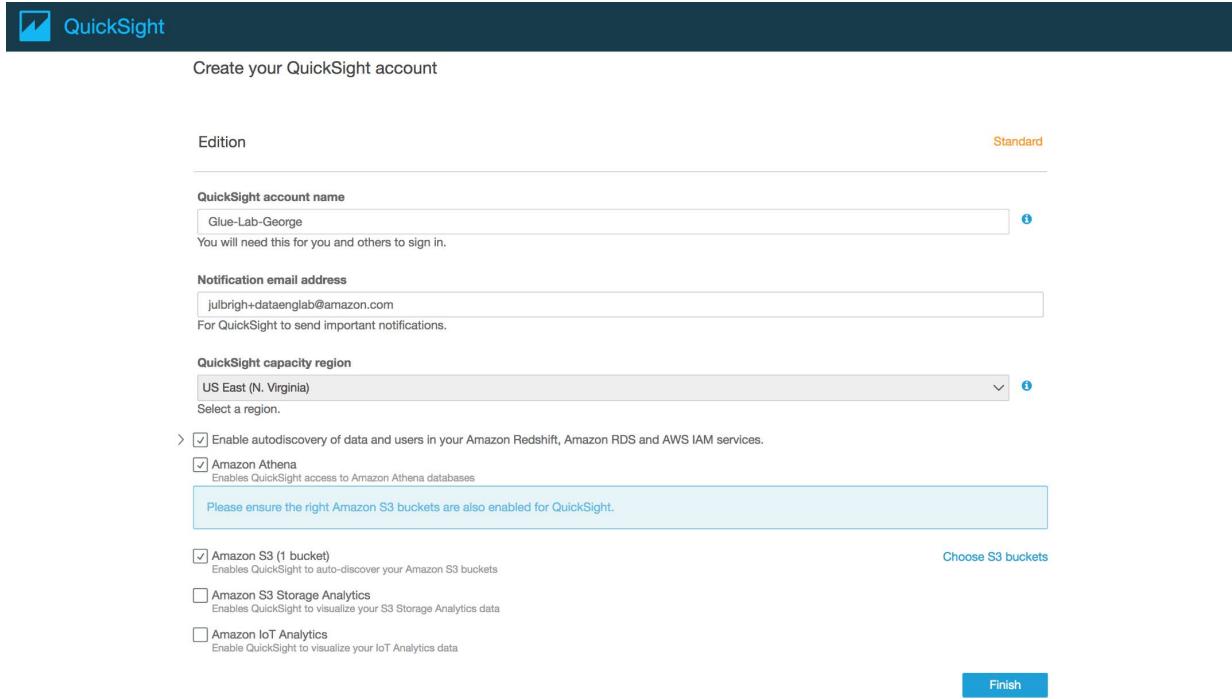


Create your QuickSight account

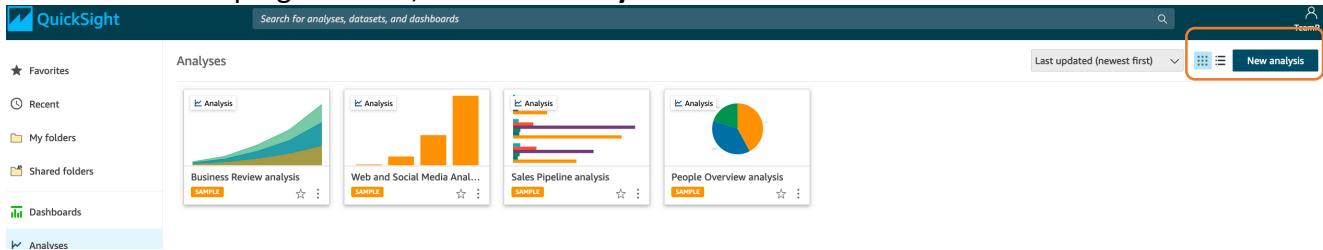
5. On the Create your QuickSight account page, for QuickSight account name give a unique name (e.g., quicksight-lab-<initials>-<randomstring>) and email address.
6. Choose the appropriate AWS region based on where you are running this workshop on and the check boxes to enable auto discovery, Amazon Athena, and Amazon S3.
7. Select your DMS bucket (e.g., "xxx-dmslabs3bucket-xxx"), Click **Finish**.



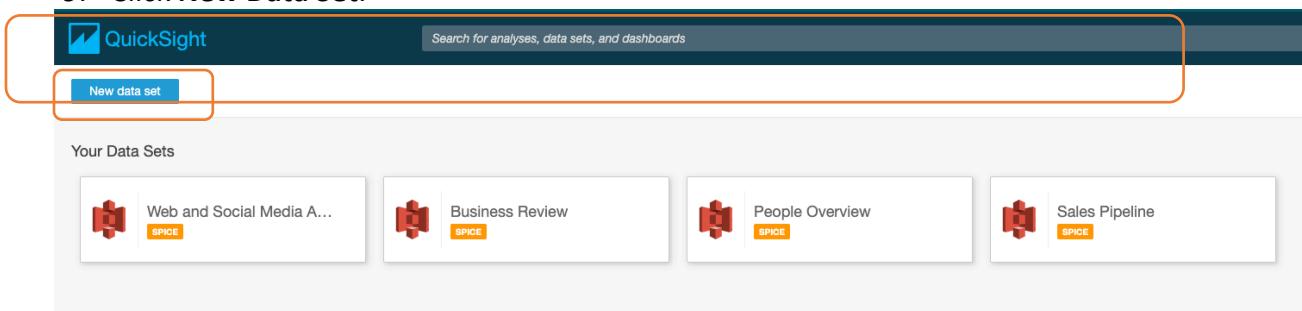
## Lab 3. Consuming data with Athena and Quicksight



8. On the top right corner, click **New analysis**.



9. Click **New Data Set**.



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

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

12. 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

13. In the Database drop-down list, select the database **ticketdata**.

14. 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

15. 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**

Finish data set creation

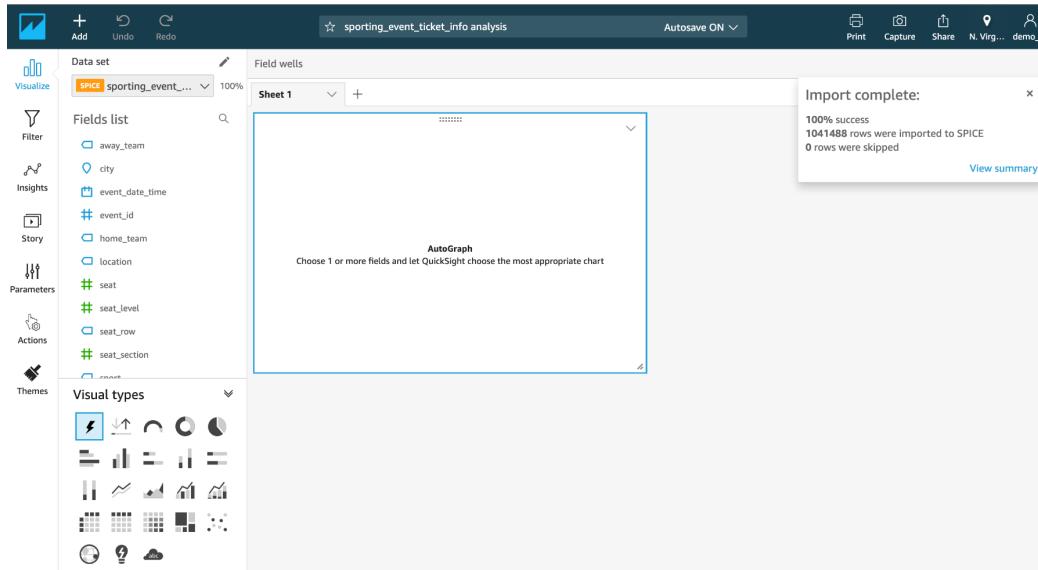
Table: sporting\_event\_ticket\_info  
Data source: ticketdata-qs  
Schema: ticketdata

Import to SPICE for quicker analytics    ✓ 20.9GB available    SPICE  
 Directly query your data

Edit/Preview data    Visualize

## 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.



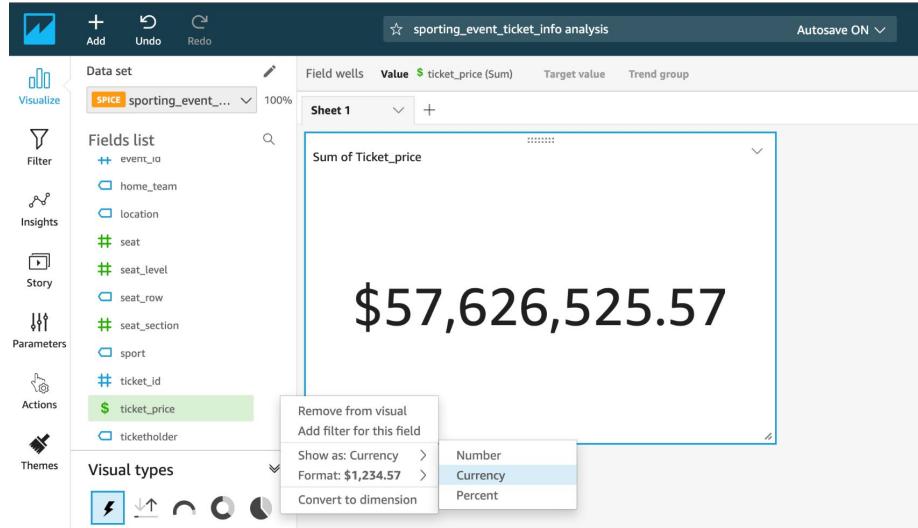
**Note:** The SPICE dataset will take a few minutes to be built, but you can continue to create some charts on the underlying data.

### 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

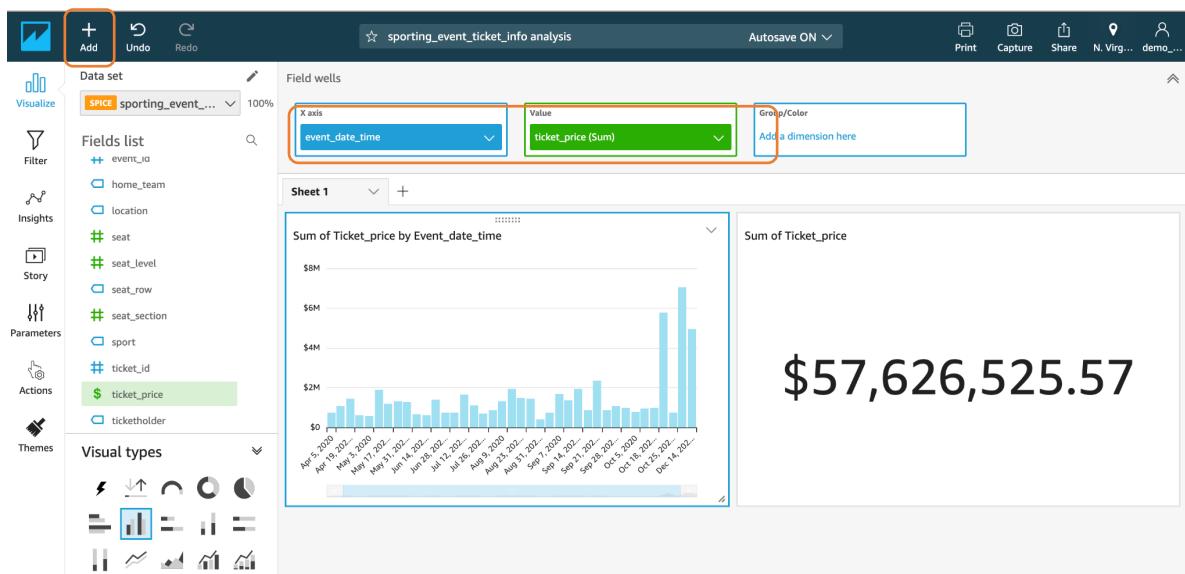


3. You can **add visual** by clicking **Add button** at top left corner of screen.

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

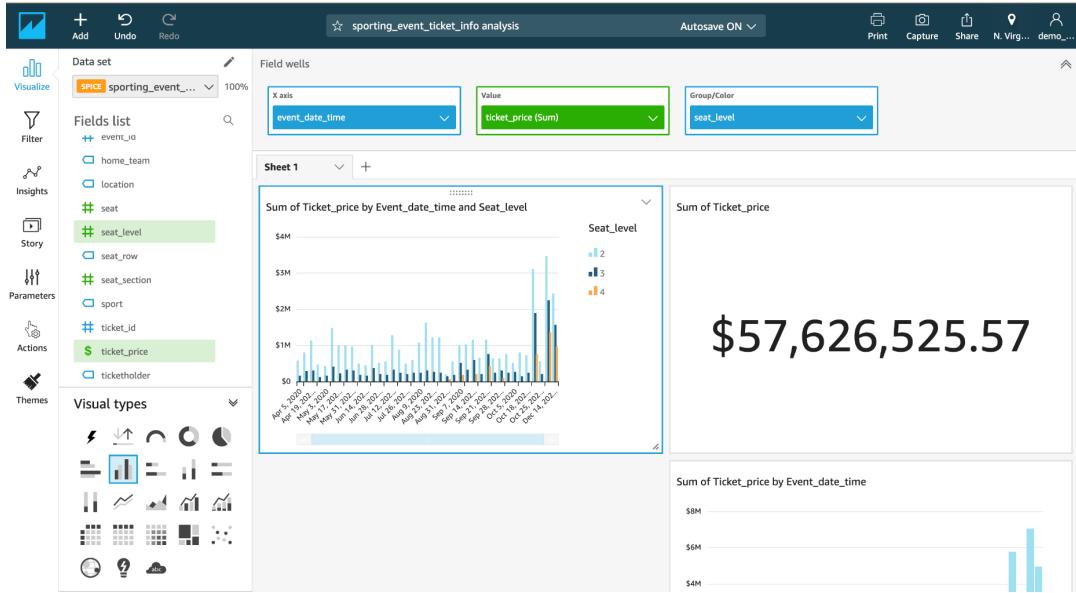
This layout requires a value for the X-axis. In Fields list, select the "**event\_date\_time**" field and you should see the visualization update.

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



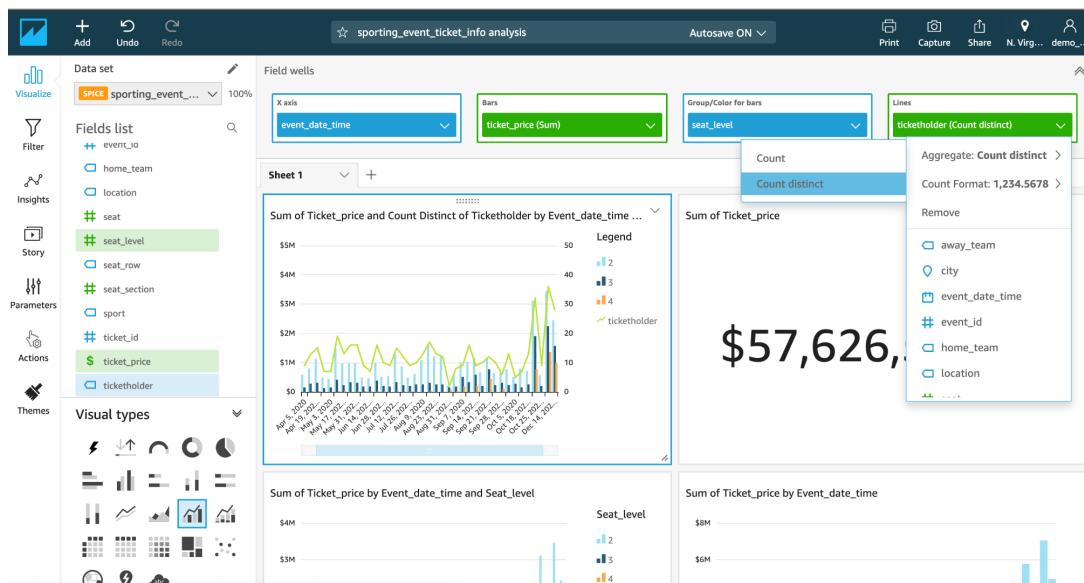
4. You can drag and move other visuals to adjust space in dashboard. In the Fields list, click and drag the **seat\_level** field to the **Group/Color** box. You can also use the slider below the x axis to fit all of the data.

## Lab 3. Consuming data with Athena and Quicksight



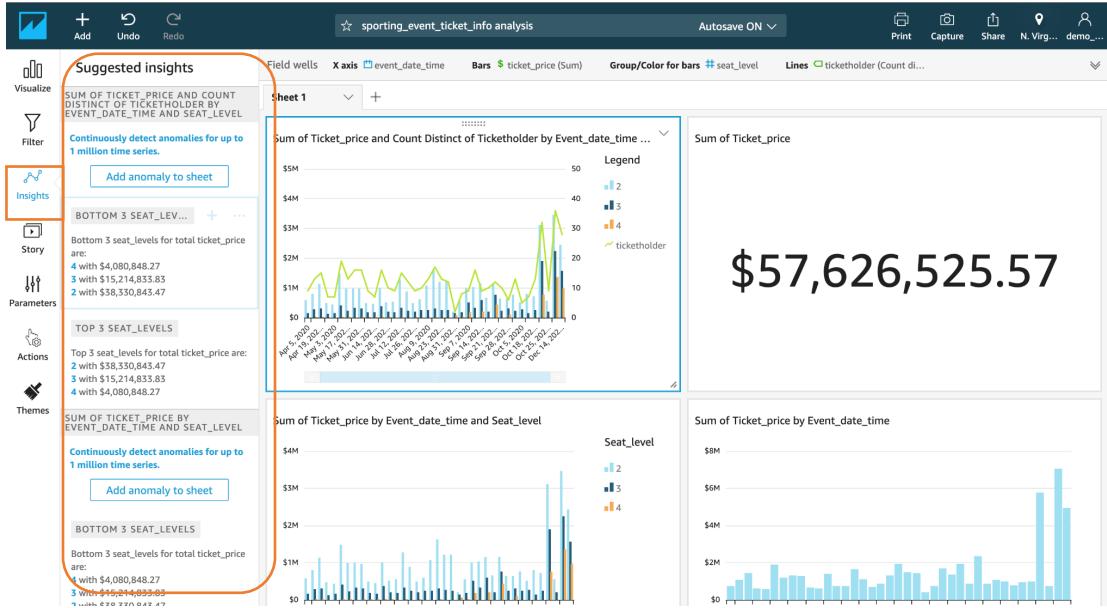
Let's build on this one step further by changing the chart type:

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

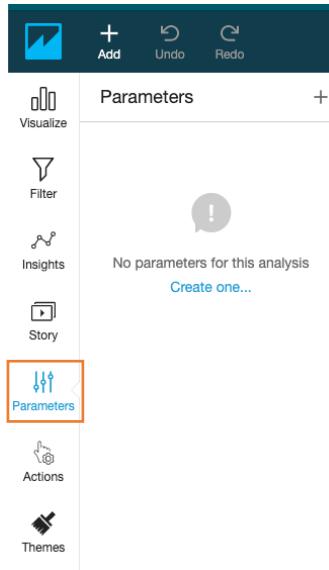


Feel free to experiment with other chart types and different fields to get a sense of the data.

## 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.
  - a. For Name, type **EventFrom**.
  - b. For Data type, choose **Datetime**.
  - c. For Time granularity, set **Hour**.

## Lab 3. Consuming data with Athena and Quicksight

- d. For Default value, select the value from calendar as start date available in your graph for **event\_date\_time**. For example, **2021-01-01 00:00**.
- e. Click **Create**, and then **close** the Parameter Added dialog box.

Create new parameter ×

Use parameters to dynamically control values in your fields, filters, and sheet.

Name

Data type (Not alterable after creation)

Time granularity

Default date

Dynamic default [Set a dynamic default](#)

3. Create another parameter with the following attributes:
  - a. **Name: EventTo**
  - b. **Data type: Datetime**
  - c. For Time granularity, set **Hour**.
  - d. For Default value, select the value from calendar as end date available in your graph for **event\_date\_time**. For example, **2022-01-01 00:00**
  - e. Click **Create**

Create new parameter ×

Use parameters to dynamically control values in your fields, filters, and sheet.

Name

Data type (Not alterable after creation)

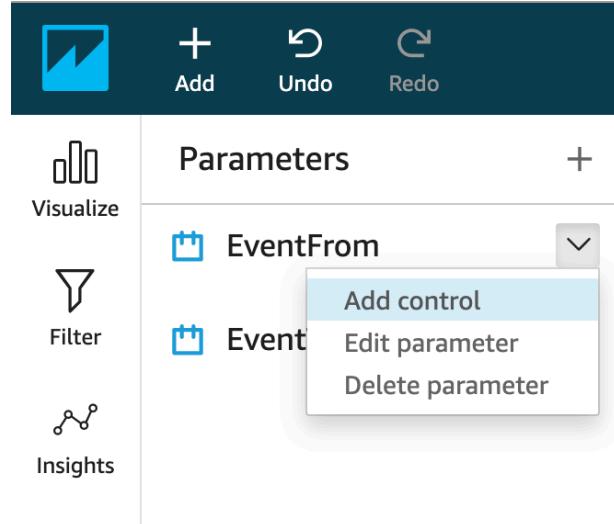
Time granularity

Default date

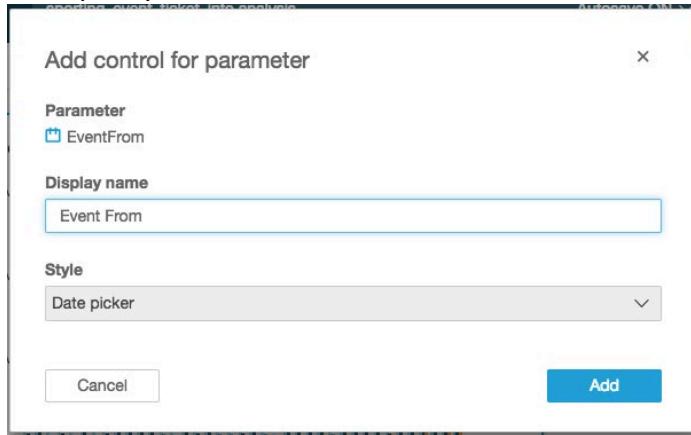
Dynamic default [Set a dynamic default](#)

## Lab 3. Consuming data with Athena and Quicksight

4. 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**.

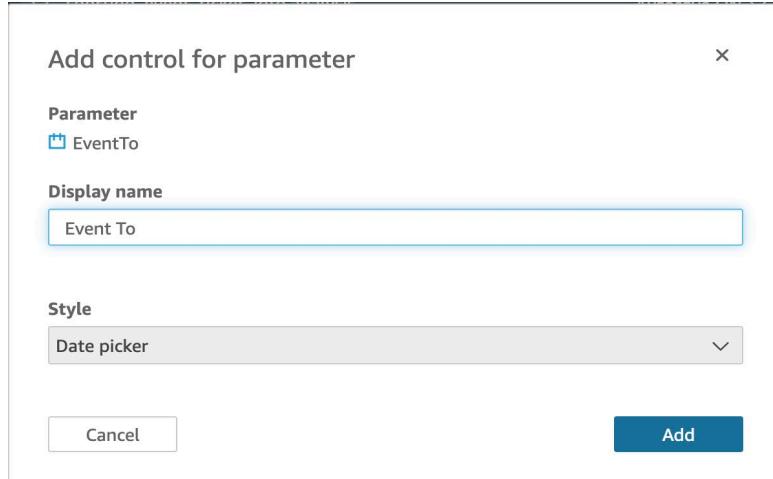


5. For Display name, specify **Event From** and click **Add**.

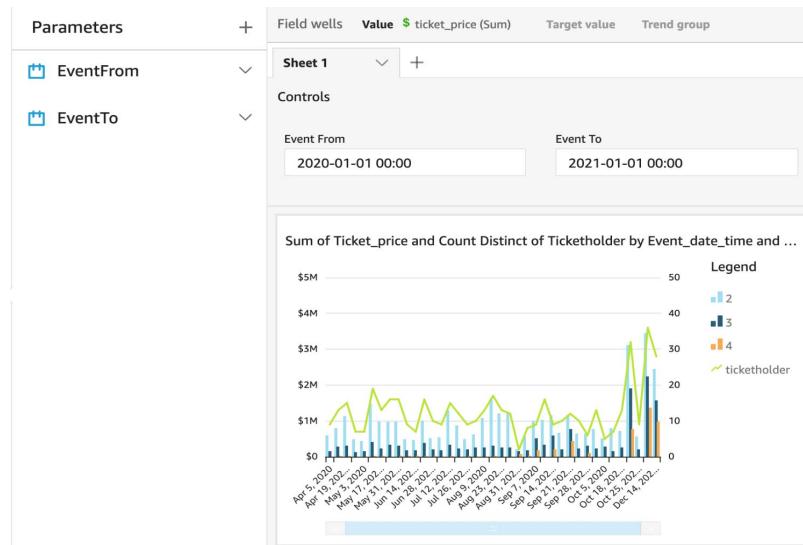


6. Repeat the process to add a control for **EventTo** with display name **Event To**

## Lab 3. Consuming data with Athena and Quicksight



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

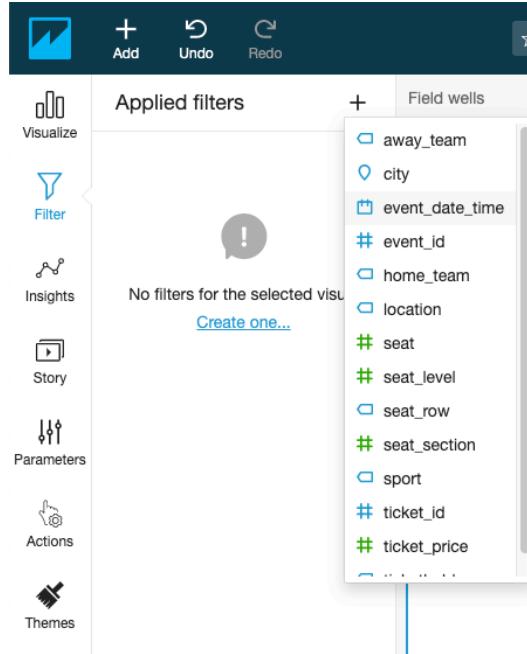


### Create a QuickSight Filter

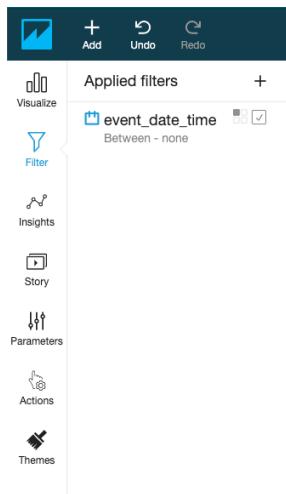
To complete the process, we will wire up a filter to these controls for all visuals.

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

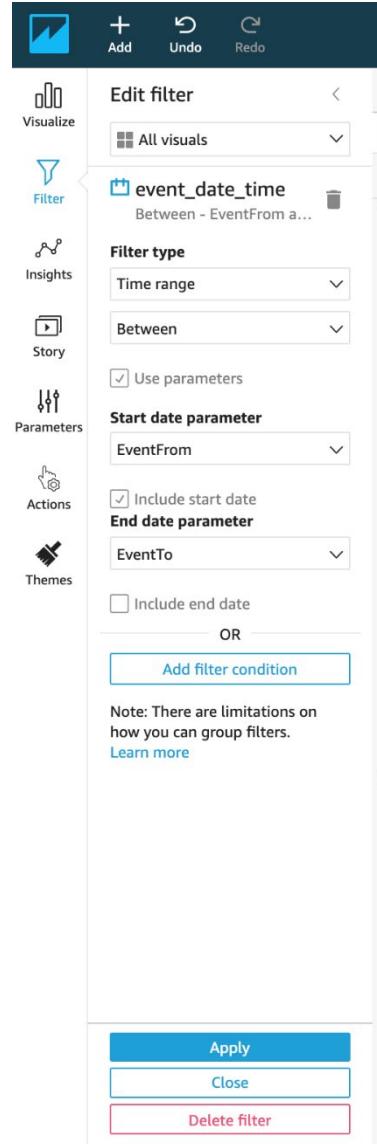


3. Click this filter to edit the properties.



4. 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

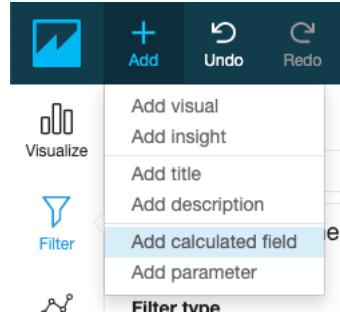


## Add Calculated Fields

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**.

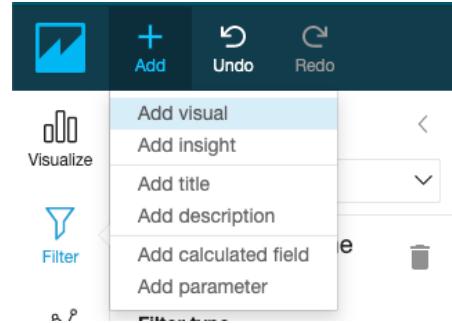
A screenshot of the 'Edit calculated field' dialog box. The title bar says 'Edit calculated field'. The main area shows the formula `1 extract('WD',{event_date_time})`. On the right side, there are sections for 'Fields', 'Parameters', and 'Functions'. The 'Functions' section has a search bar and a list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

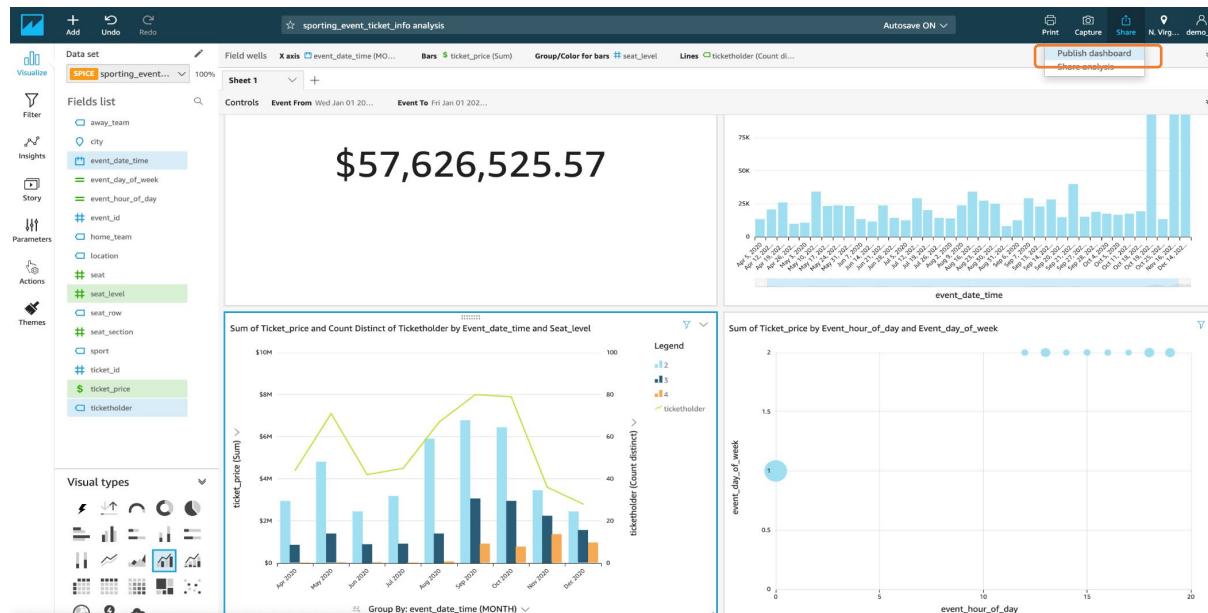
## Lab 3. Consuming data with Athena and Quicksight



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

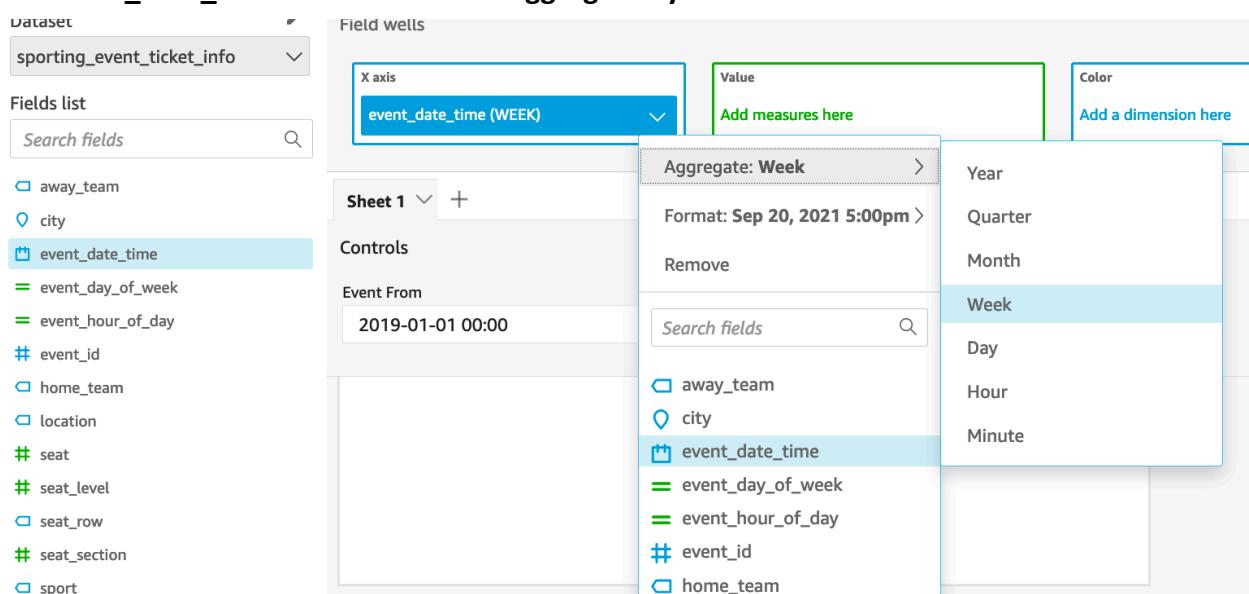
- Narrative
- Anomaly Detection
- Forecasting

ML-Insights are only available to enterprise version of QuickSight. You will need to upgrade to Enterprise Edition before you start with the task (*if you haven't selected Enterprise Edition in the beginning of this lab*). 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>

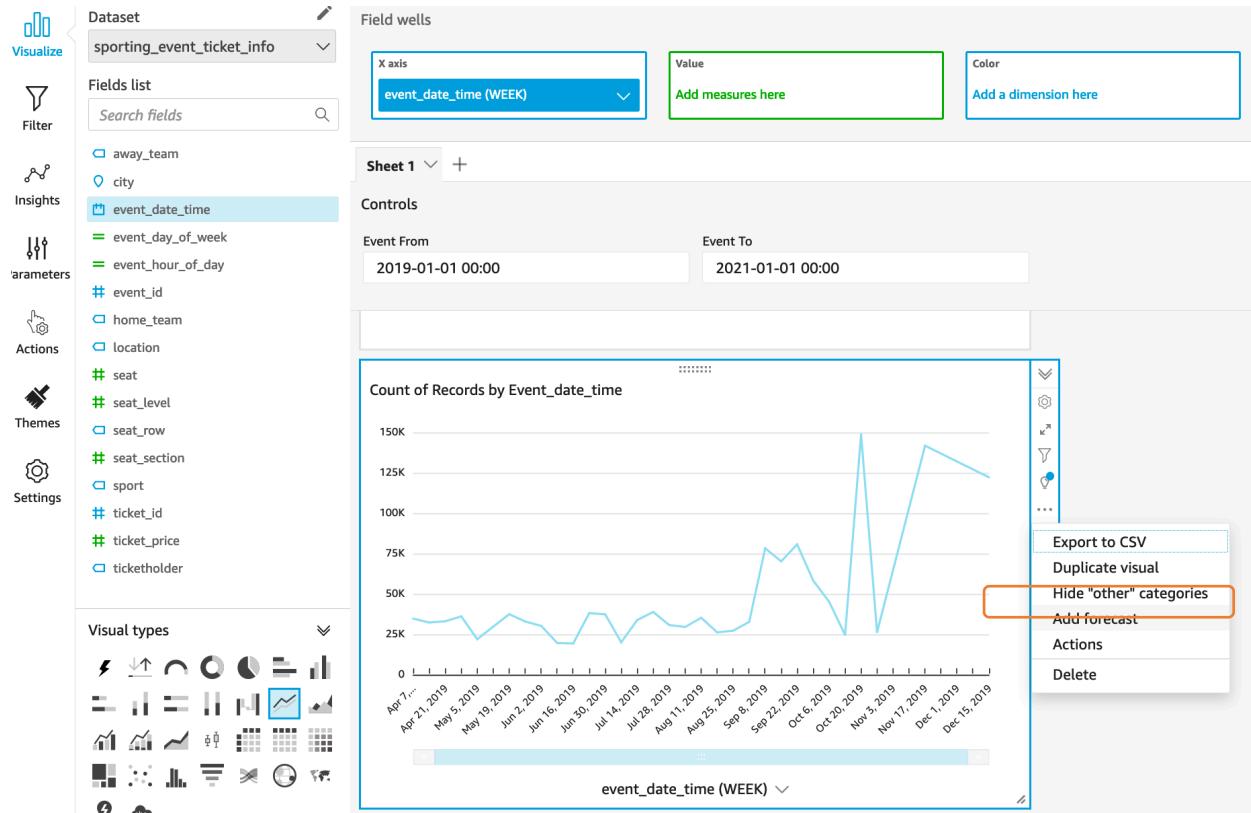
Let's see how we can add a bit of forecasting in our dashboard. Forecasting works with timeseries, which is better represented with a line graph. Let's first create a line graph.

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

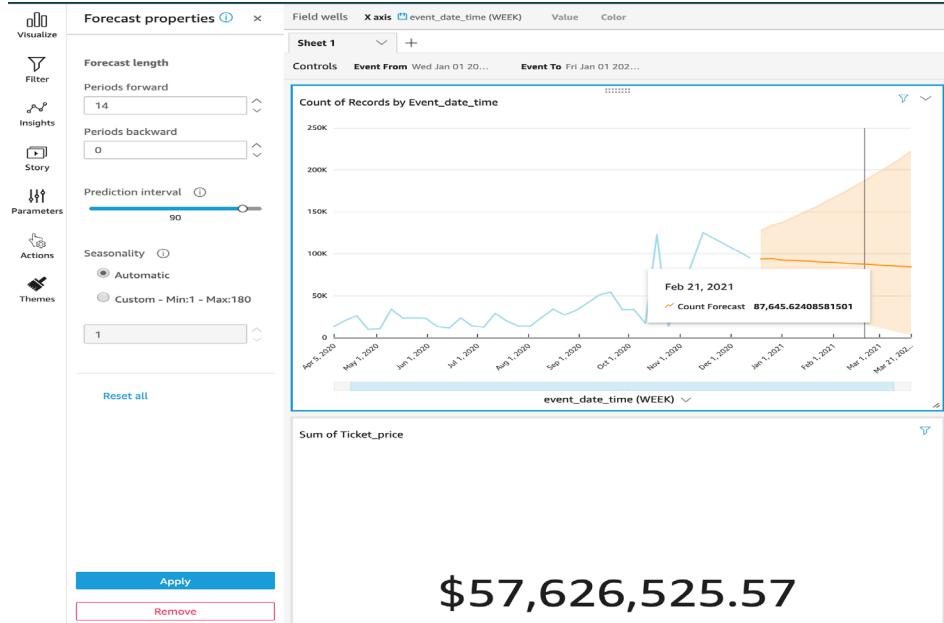


## Lab 3. Consuming data with Athena and Quicksight

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



## 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.

## Lab 3. Consuming data with Athena and Quicksight

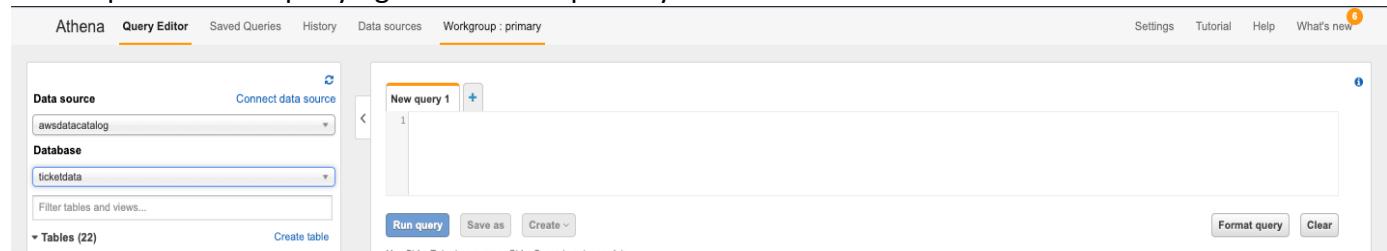
The resources have been already created before starting the lab. You can go to the [CloudFormation](#) console, choose the oldest stack. Navigate to the “**Resources**” to understand the different resources created by the template. Navigate to **outputs** section to see the results of resources created with description.

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-ldtj44qkwygle	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**”

## Lab 3. Consuming data with Athena and Quicksight

The screenshot shows the AWS Athena Workgroups page. At the top, there are tabs for Athena, Query Editor, Saved Queries, History, Data sources, and Workgroup (which is currently selected). Below the tabs, there's a section titled 'Workgroups' with a note about using workgroups to separate users, teams, applications, or workloads. A 'Create workgroup' button is available. The main table lists one workgroup:

Name	Description	Creation time	Workgroup status
primary		2019/10/28 13:01:55 UTC-4	Enabled

At the bottom right of the table, there are navigation links: 'Beginning of List', 'Previous Page', and 'Next Page'.

3. Provide the following:

- a. Workgroup **Name**: “**workgroupA**”
- b. **Description**: (optional):
  - 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**”

## 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**.

[Explore the features of workgroups](#)

## Lab 3. Consuming data with Athena and Quicksight

- 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-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	-

- Navigate to [IAM dashboard](#) and copy the Sign-in URL for IAM users in this account as shown below

The screenshot shows the AWS IAM dashboard. On the left, there's a sidebar with 'Identity and Access Management (IAM)' and a 'Dashboard' link. The main area is titled 'IAM dashboard' and contains the message 'Sign-in URL for IAM users in this account'. Below this is a redacted URL: [https://\[REDACTED\].signin.aws.amazon.com/console](https://[REDACTED].signin.aws.amazon.com/console). There are 'Edit' and 'Customize' buttons next to the URL. A red arrow points from the text 'copy the Sign-in URL for IAM users in this account' in the instructions above to the URL in the screenshot.

- Next, Open the copied link in a different browser or in private window, select **IAM user** and login with following credential:
  - AccountID:** This should be pre-filled
  - IAM User name:** <value copied for **BusinessAnalystUser**>
  - Password:** **Master123!**
  - Make sure the region is the same as where you are running the whole of this workshop
- 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 top navigation bar includes services like Services, Resource Groups, Athena, S3, AWS Glue, and RDS. The main area has tabs for Athena, Query Editor, Saved Queries, History, Data sources, and Workgroup : workgroupA. The left sidebar shows a Data source (AwsDataCatalog) and a Database (ticketdata). Under ticketdata, there are Tables (3) and Views (1). The 'sporting\_event\_info' view is highlighted with a red box. The main panel displays a query in New query 1: `SELECT * FROM "ticketdata"."sporting_event_info" limit 10;`. Buttons for Run query, Save as, and Create are present. A note at the bottom says 'Use Ctrl + Enter to run query, Ctrl + Space to autocomplete'. The results section 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 in the top right is now set to 'workgroupA', as indicated by a red box around the text. The rest of the interface, including the database selection and the running query, remains 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 (which is highlighted with a red box). Below is a table listing workgroups:

	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

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, 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;'. Below the tabs 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 setup as before, with the 'Run query' button visible.

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 'Settings' dialog box is open. It shows the 'Workgroup' dropdown set to 'workgroupA'. The 'Query result location' field is populated with 's3://dmslab-student-s3bucketworkgroupa-ltj44qkwyle/'. 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 right 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;
```

Below the editor, the results are shown in a table:

	event_id	sport	event_date_time	home_team	away_team	location	city
1	4491	baseball	2020-07-19 00:00:00.000	Seattle Mariners	New York Mets	Safeco Field	Seattle Washington
2	4581	baseball	2020-09-20 00:00:00.000	Seattle Mariners	Boston Red Sox	Safeco Field	Seattle Washington
3	4191	baseball	2020-07-05 00:00:00.000	Houston Astros	Texas Rangers	Minute Maid Park	Houston Texas
4	4451	baseball	2020-06-14 00:00:00.000	Seattle Mariners	Philadelphia Phillies	Safeco Field	Seattle Washington
5	4481	baseball	2020-07-05 00:00:00.000	Seattle Mariners	Toronto Blue Jays	Safeco Field	Seattle Washington
6	4611	baseball	2020-10-11 00:00:00.000	Seattle Mariners	Los Angeles Angels	Safeco Field	Seattle Washington
7	5661	baseball	2020-07-26 00:00:00.000	San Diego Padres	Kansas City Royals	Petco Park	San Diego California
8	5591	baseball	2020-05-31 00:00:00.000	San Diego Padres	Texas Rangers	Petco Park	San Diego California
9	9151	football	2020-09-21 19:00:00.000	San Diego Chargers	Denver Broncos	Qualcomm Stadium	San Diego, California
10	4071	baseball	2020-04-12 00:00:00.000	Houston Astros	Tampa Bay Rays	Minute Maid Park	Houston Texas

- 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](#)

	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 interface. A red box highlights the error message in the results section:

Your query has the following error(s):

```
User: arn:aws:iam::665953140268:lakeformation-BusinessAnalystUser-7H52lWD4CWS6 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)
```

Below the error message, 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-s3bucket/workgroupa-ldtj44qkyle/ <a href="#">Edit</a>
Amazon CloudWatch Metrics	Enabled
Encrypt query results	Not defined
Workgroup status	Enabled
Workgroup ARN	arn:aws:athena:us-east-1:678691952726:workgroup/workgroupA <a href="#">Edit</a>
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::656953140268:user/lakeformation-WorkgroupManagerUser is not authorized to perform: athena:GetWorkGroup on resource: arn:aws:athena:us-east-1:656953140268:workgroup:primary (Service: AmazonAthena; Status Code: 400; Error Code: AccessDeniedException; Request ID: 07164bd-bd-a0-4571-9aa7-f72507e620a0)

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

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

No databases or tables found.

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

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

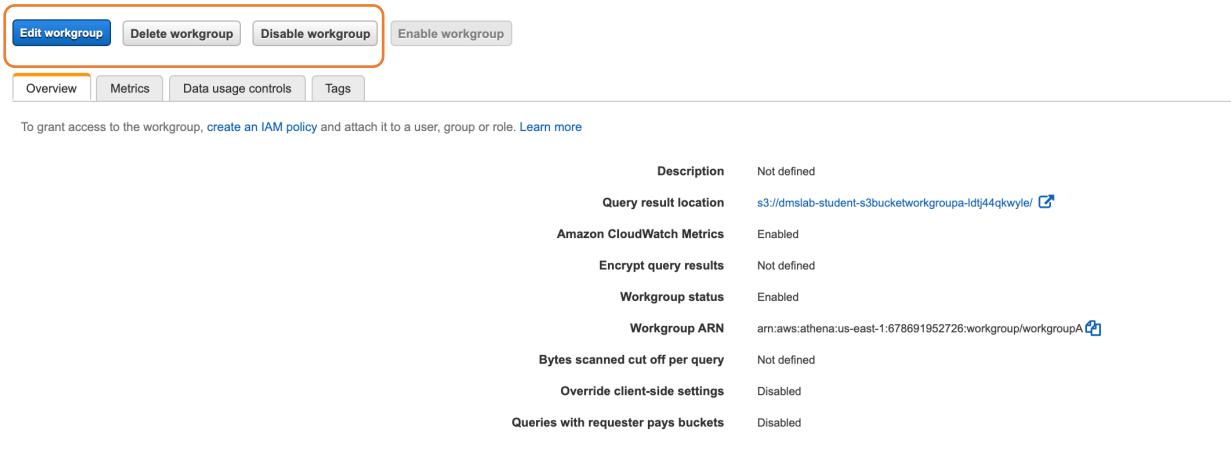
Use workgroups to separate users, teams, applications, or workloads, and to set limits on amount of data each query or the entire workgroup can process. You can also view query-related metrics in AWS CloudWatch. [Learn more](#)



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

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-ldj44qkwyle/ 
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

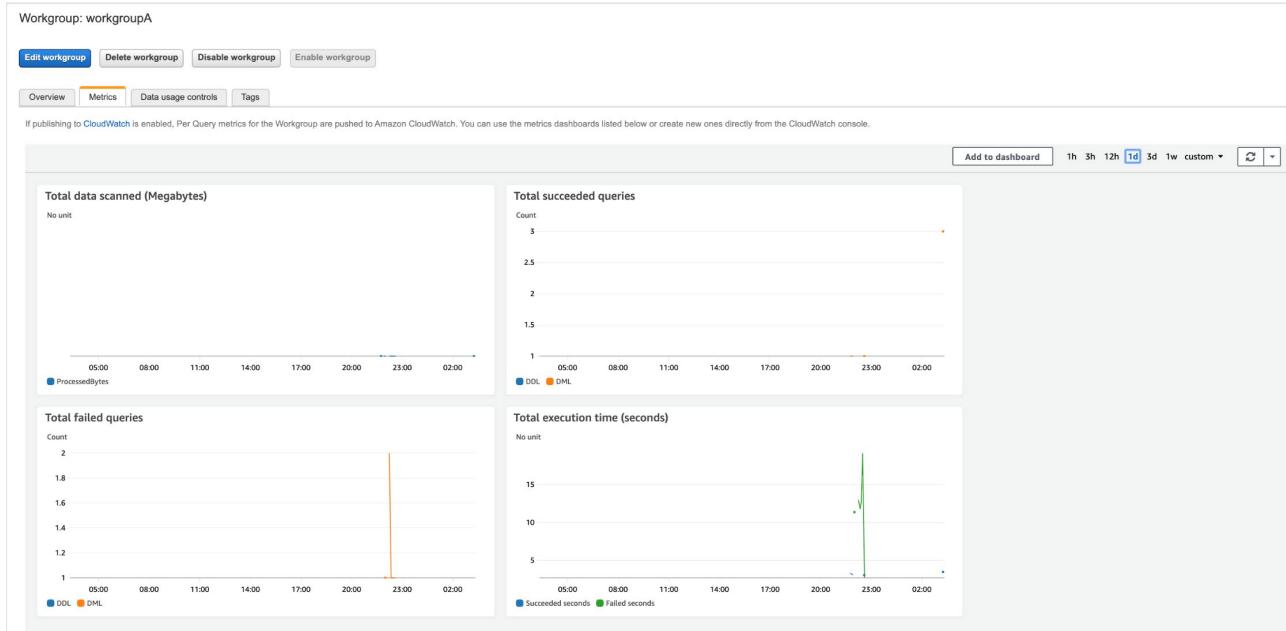
**Please Note:** For lab purpose, we are attaching policies directly to users. For Best practices, we recommend creating separate groups in IAM for different workgroups and then attaching policies for different workgroups to their respective groups in IAM.

## 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 **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.

The screenshot shows the AWS Workgroup console for the 'primary' workgroup. The 'Data usage controls' tab is selected. A configuration box is open, showing 'Data limits' set to 10 Megabytes MB. Below it, an 'Action' section states: 'If the query exceeds the limit, it will be cancelled.' There are 'Delete' and 'Update' buttons at the bottom of the box.

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. A message box displays: 'Query cancelled! : Bytes scanned limit was exceeded'. The main interface shows a query window with the following code:  
New query 1  
1 SELECT \* FROM "ticketdata"."sporting\_event\_ticket"  
Run query Save as Create (Run time: 3.33 seconds, Data scanned: 10 MB)

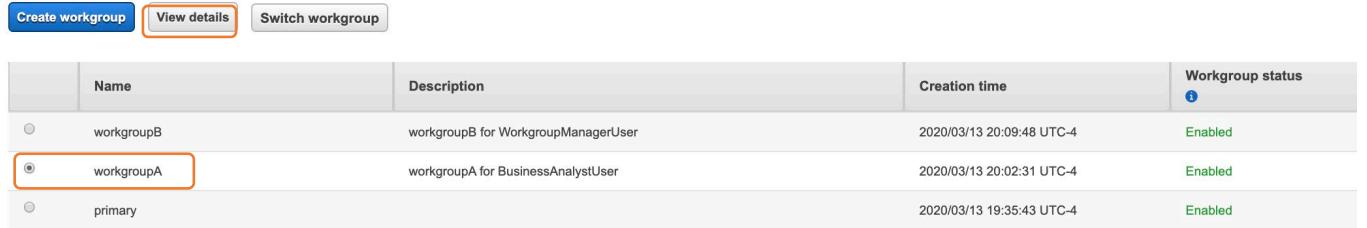
For **per-workgroup data usage control**, you can configure the maximum amount of data scanned by all queries in the workgroup during a specific period. This is useful when you have few analytics reports to run, where you probably have a good idea of how long the process should take and the total amount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

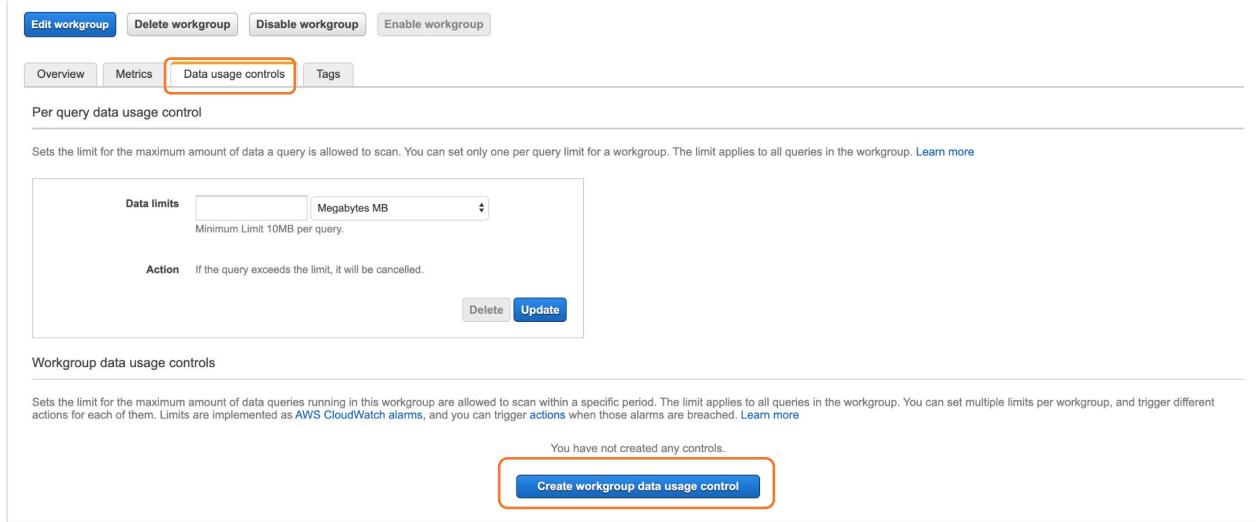
### Workgroups

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  Megabytes MB

Minimum Limit 10MB per query.

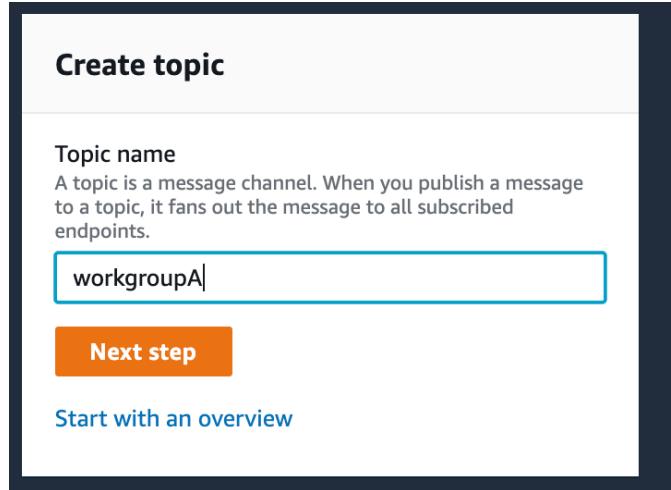
Action If the query exceeds the limit, it will be cancelled.

Workgroup data usage controls

You have not created any controls.

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**, then **Create Topic**.
- iii. Note down the topic **ARN number**. Looks like  
**arn:aws:sns:<region>:<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 'Topics' page. On the left is a sidebar with 'Amazon SNS' and links for 'Dashboard', 'Topics' (which is selected), 'Subscriptions', 'Mobile' (with 'Push notifications' and 'Text messaging (SMS)'), and 'AWS Lambda'. The main area shows a topic named 'workgroupA'. The 'Details' tab is selected, showing the 'Name' (workgroupA), 'ARN' (arn:aws:sns:us-east-1:665953140268:workgroupA), 'Display name' (empty), and 'Topic owner' (665953140268). Below the details 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  Kilobytes

Time period

Action  Send a notification to

Cancel **Create**

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

## Lab 3. Consuming data with Athena and Quicksight

Workgroup: workgroupA

The screenshot shows the AWS Workgroup configuration page for 'workgroupA'. At the top, there are buttons for 'Edit workgroup' (highlighted in blue), 'Delete workgroup', 'Disable workgroup', and 'Enable workgroup'. Below these are tabs for 'Overview', 'Metrics', 'Data usage controls' (which is selected and highlighted in orange), and 'Tags'. Under 'Per query data usage control', it says '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.' A 'Learn more' link is provided. A form allows setting a 'Data limits' value (e.g., 10 MB) and selecting an 'Action' if exceeded (e.g., 'If the query exceeds the limit, it will be cancelled'). Buttons for 'Delete' and 'Update' are at the bottom. Below this, under 'Workgroup data usage controls', it says '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.' A 'Learn more' link is provided. A table lists a single data usage control entry: 'Data limits' (10 KB), 'Time period' (1 minute), 'Action' (Send notification to topic : arn:aws:sns:us-east-1: [redacted]), and 'workgroupA' (under 'Created by').

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

### AWS Notification - Subscription Confirmation

The screenshot shows an AWS SNS subscription confirmation email from 'AWS Notifications <no-reply@sns.amazonaws.com>' to a recipient (represented by a redacted email address). The email was sent 'Today at 5:58'. It contains a message: 'You have chosen to subscribe to the topic: arn:aws:sns:us-east-1: [redacted] workgroupA'. Below this is a blue button with the text 'To confirm this subscription, click or visit the link below (If this was in error no action is necessary): [Confirm subscription](#)'. At the bottom, there is a note: '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";
```

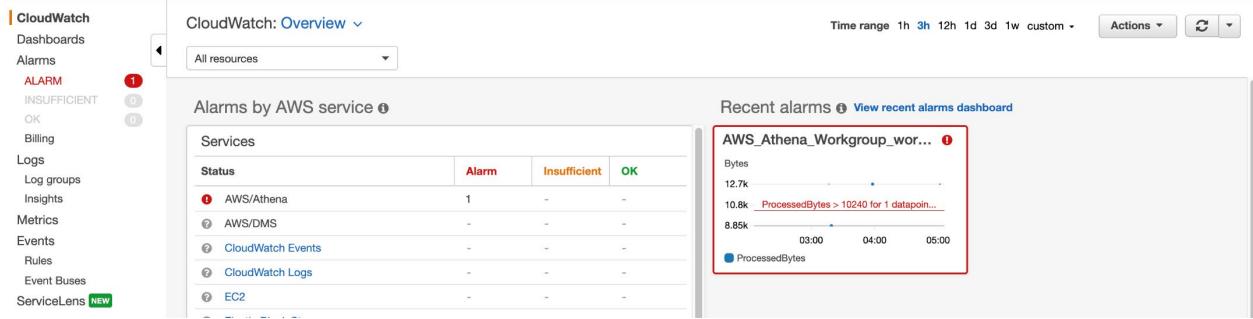
6. You will receive an **email notification from AWS Notifications** stating that workgroup data usage threshold has been breached, which will look something like this:

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

The screenshot shows an AWS CloudWatch alarm notification email from 'AWS Notifications <no-reply@sns.amazonaws.com>' to a recipient (represented by a redacted email address). The email is from 'AN'. It contains a message: 'You are receiving this email because your Amazon CloudWatch Alarm "AWS\_Athena\_Workgroup\_workgroupA\_c0ff968d-32fe-4c37-b741-fa45a61..." 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".'

## Lab 3. Consuming data with Athena and Quicksight

7. You can also check [CloudWatch Alarms](#) and get more details on CloudWatch console:



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 tabs for 'Athena', 'Query Editor', 'Saved Queries', 'History', 'AWS Glue Data Catalog', 'Workgroup : teamA' (which is circled in red), 'Settings', and 'Tutorial'. Below this, the 'Workgroup: teamA' section is displayed. It includes buttons for 'Edit workgroup' (circled in red with number 2), 'Delete workgroup', 'Disable workgroup', and 'Enable workgroup'. A navigation bar at the bottom has tabs for 'Overview', 'Metrics', 'Data usage controls', and 'Tags' (which is circled in red with number 3). A note below the tabs 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. Below it is a table with columns 'Key' and 'Value'.

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