

Practical - 6

Name: Sakshi Deshmukh

PRN: 202301040191

Roll no. 154

AWS Athena

product_data

product_id	product_name	category	price	stock_quantity
101	Classic Le	Electronics	1200	50
102	Smart-Hor	Electronics	150	120
103	4K Ultra-H	Electronics	350	75
104	Gaming M	Electronics	80	300
105	Wireless E	Electronics	199	250
201	The Great	Books	15	500
202	A Tale of	Books	12	450
203	Learning F	Books	45	150
204	A Brief Hi	Books	22	200
205	Dune	Books	18	300
301	Men's Rui	Clothing	90	400
302	Women's	Clothing	55	600

electronics_data

product_id	product_name	category	price	stock_quantity
101	Classic Le	Electronics	1200	50
102	Smart-Hor	Electronics	150	120
103	4K Ultra-H	Electronics	350	75
104	Gaming M	Electronics	80	300
105	Wireless E	Electronics	199	250

Setting up S3 bucket

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

Create bucket

Create bucket

Info

Buckets are containers for data stored in S3.

General configuration

AWS Region

Europe (Stockholm) eu-north-1

Bucket type

Info

☒ General purpose

Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

☐ Directory

Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name

Info

sakshi-athena-practical

Bucket names must be 3 to 63 characters and unique within the global namespace. Bucket names must also begin and end with a letter or number. Valid characters are a-z, 0-9, periods (.), and hyphens (-). [Learn More](#)

Copy settings from existing bucket - optional

Only the bucket settings in the following configuration are copied.

Choose bucket

Format: s3://bucket/prefix

Object Ownership

Info

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

Successfully created bucket "sakshi-athena-practical"

To upload files and folders, or to configure additional bucket settings, choose [View details](#).

General purpose buckets

All AWS Regions

Directory buckets

General purpose buckets (1)

Info

Buckets are containers for data stored in S3.

Find buckets by name

Name

▲

AWS Region

▼

Creation date

▼

☐ sakshi-athena-practical

Europe (Stockholm) eu-north-1

October 28, 2025, 15:35:48 (UTC+05:30)

Account snapshot

Info

Updated daily

Storage Lens provides visibility into storage usage and activity trends.

External access summary - new

Info

Updated daily

External access findings help you identify bucket permissions that allow public access or access from other AWS accounts.

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates.

Privacy

Terms

Cookie preferences

Bucket for storing results

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

Create bucket

Create bucket [Info](#)

Buckets are containers for data stored in S3.

General configuration

AWS Region
Europe (Stockholm) eu-north-1

Bucket type [Info](#)

☒ **General purpose**
Recommended for most use cases and access patterns. General purpose buckets are the original S3 bucket type. They allow a mix of storage classes that redundantly store objects across multiple Availability Zones.

☐ **Directory**
Recommended for low-latency use cases. These buckets use only the S3 Express One Zone storage class, which provides faster processing of data within a single Availability Zone.

Bucket name [Info](#)

sakshi-athena-practical-results

Bucket names must be 3 to 63 characters and unique within the global namespace. Bucket names must also begin and end with a letter or number. Valid characters are a-z, 0-9, periods (.), and hyphens (-). [Learn More](#)

Copy settings from existing bucket - optional
Only the bucket settings in the following configuration are copied.

[Choose bucket](#)

Format: s3://bucket/prefix

Object Ownership [Info](#)

Control ownership of objects written to this bucket from other AWS accounts and the use of access control lists (ACLs). Object ownership determines who can specify access to objects.

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

Successfully created bucket "sakshi-athena-practical-results"

To upload files and folders, or to configure additional bucket settings, choose [View details](#).

General purpose buckets

All AWS Regions

Directory buckets

General purpose buckets (2) [Info](#)

Buckets are containers for data stored in S3.

[Find buckets by name](#)

Name	AWS Region	Creation date
sakshi-athena-practical	Europe (Stockholm) eu-north-1	October 28, 2025, 15:35:48 (UTC+05:30)
sakshi-athena-practical-results	Europe (Stockholm) eu-north-1	October 28, 2025, 15:54:28 (UTC+05:30)

[Copy ARN](#)

[Empty](#)

[Delete](#)

[Create bucket](#)

Account snapshot [Info](#)

Updated daily

[View dashboard](#)

Storage Lens provides visibility into storage usage and activity trends.

External access summary - new [Info](#)

Updated daily

External access findings help you identify bucket permissions that allow public access or access from other AWS accounts.

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Creating folder for non-partitioned data

The screenshot shows the 'Create folder' page in the AWS Management Console. The breadcrumb navigation is 'Amazon S3 > Buckets > sakshi-athena-practical > Create folder'. The page title is 'Create folder' with an 'Info' link. A note states: 'Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)'. A warning box indicates: 'Your bucket policy might block folder creation. If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.' The 'Folder' section has a 'Folder name' input field containing 'non-partitioned-data' followed by a slash. A note below says: 'Folder names can't contain "/>

Upload data into folder

The screenshot shows the 'Upload' page in the AWS Management Console. The breadcrumb navigation is 'Amazon S3 > Buckets > sakshi-athena-practical > Upload'. The page title is 'Upload' with an 'Info' link. A note states: 'Add the files and folders you want to upload to S3. To upload a file larger than 160GB, use the AWS CLI, AWS SDKs or Amazon S3 REST API. [Learn more](#)'. A dashed box contains the text: 'Drag and drop files and folders you want to upload here, or choose [Add files](#) or [Add folder](#)'. Below this, a section titled 'Files and folders (1 total, 632.0 B)' shows a table of files to be uploaded. The table has columns for 'Name', 'Folder', 'Type', and 'Size'. One file is listed: 'product_data.csv' with a folder of '-' and a size of '632.0 B'. Buttons for 'Remove', 'Add files', and 'Add folder' are at the top right. The 'Destination' section shows the destination as 's3://sakshi-athena-practical' and includes a 'Destination details' link.

Name	Folder	Type	Size
product_data.csv	-	text/csv	632.0 B

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Upload succeeded

For more information, see the [Files and folders](#) table.

Close

Upload: status

After you navigate away from this page, the following information is no longer available.

Summary

Destination	Succeeded	Failed
s3://sakshi-athena-practical	1 file, 632.0 B (100.00%)	0 files, 0 B (0%)

Files and folders

Configuration

Files and folders (1 total, 632.0 B)

Find by name

< 1 >

Name	Folder	Type	Size	Status	Error
product_data.csv	-	text/csv	632.0 B	Succeeded	-

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

sakshi-athena-practical

non-partitioned-data/

non-partitioned-data/

Copy S3 URI

Objects

Properties

Objects (1)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

Find objects by prefix

< 1 >

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	product_data.csv	csv	October 28, 2025, 15:41:06 (UTC+05:30)	632.0 B	Standard

CloudShell

Feedback

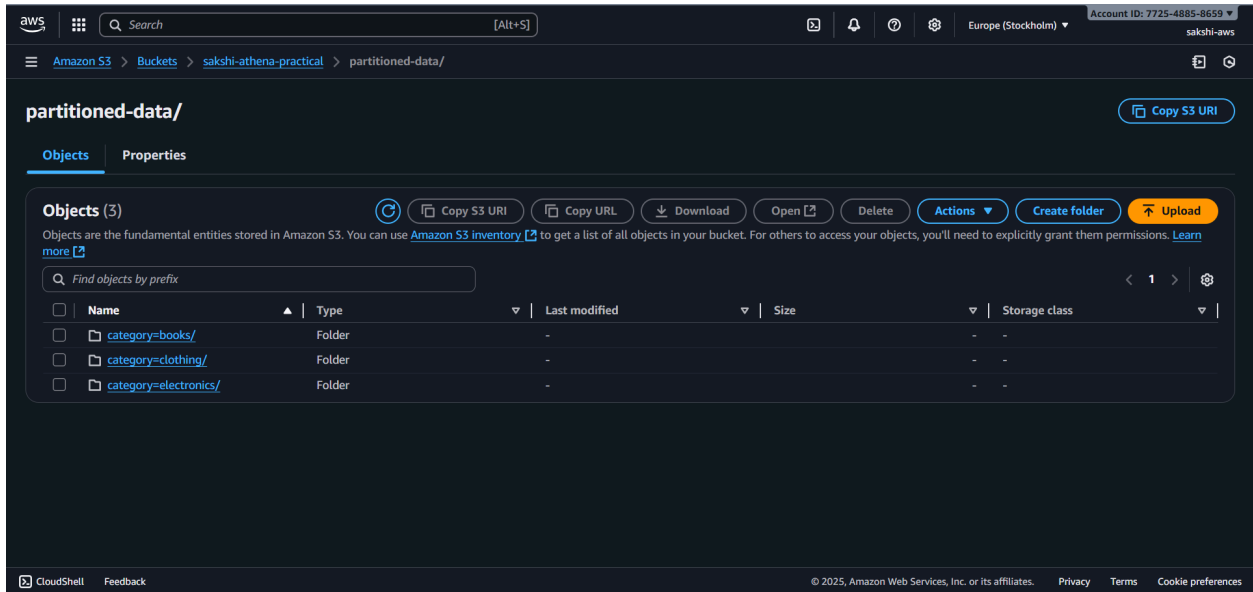
© 2025, Amazon Web Services, Inc. or its affiliates. [Privacy](#) [Terms](#) [Cookie preferences](#)

Creating folder for partitioned data

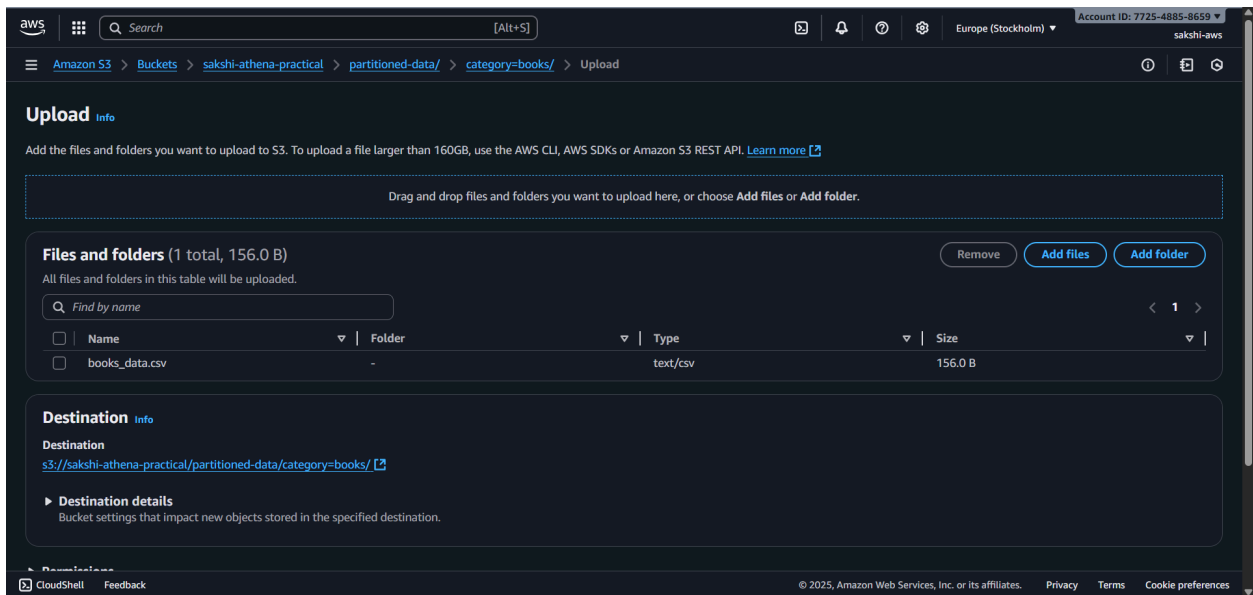
The screenshot shows the AWS Management Console interface for creating a folder in the 'sakshi-athena-practical' bucket. The breadcrumb navigation is 'Amazon S3 > Buckets > sakshi-athena-practical > Create folder'. The page title is 'Create folder' with an 'Info' link. A note states: 'Use folders to group objects in buckets. When you create a folder, S3 creates an object using the name that you specify followed by a slash (/). This object then appears as folder on the console. [Learn more](#)'. A warning box indicates: 'Your bucket policy might block folder creation. If your bucket policy prevents uploading objects without specific tags, metadata, or access control list (ACL) grantees, you will not be able to create a folder using this configuration. Instead, you can use the [upload configuration](#) to upload an empty folder and specify the appropriate settings.' The 'Folder' section has a 'Folder name' input field containing 'partitioned-data'. Below it, a note says 'Folder names can't contain "/>

Creating folder for each category

This screenshot shows the same 'Create folder' page in the AWS Management Console, but with the 'Folder name' input field set to 'category=books'. The breadcrumb navigation is 'Amazon S3 > Buckets > sakshi-athena-practical > partitioned-data/ > Create folder'. The rest of the page content, including the warning box and the 'Server-side encryption' section, remains identical to the previous screenshot.



Uploading files in each folder



aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

sakshi-athena-practical

partitioned-data/

category=books/

category=books/

Copy S3 URI

Objects

Properties

Objects (1)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	books_data.csv	csv	October 28, 2025, 15:49:56 (UTC+05:30)	156.0 B	Standard

CloudShell

Feedback

© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences

aws

Search

[Alt+S]

Europe (Stockholm)

Account ID: 7725-4885-8659

sakshi-aws

Amazon S3

Buckets

sakshi-athena-practical

partitioned-data/

category=clothing/

category=clothing/

Copy S3 URI

Objects

Properties

Objects (1)

Copy S3 URI

Copy URL

Download

Open

Delete

Actions

Create folder

Upload

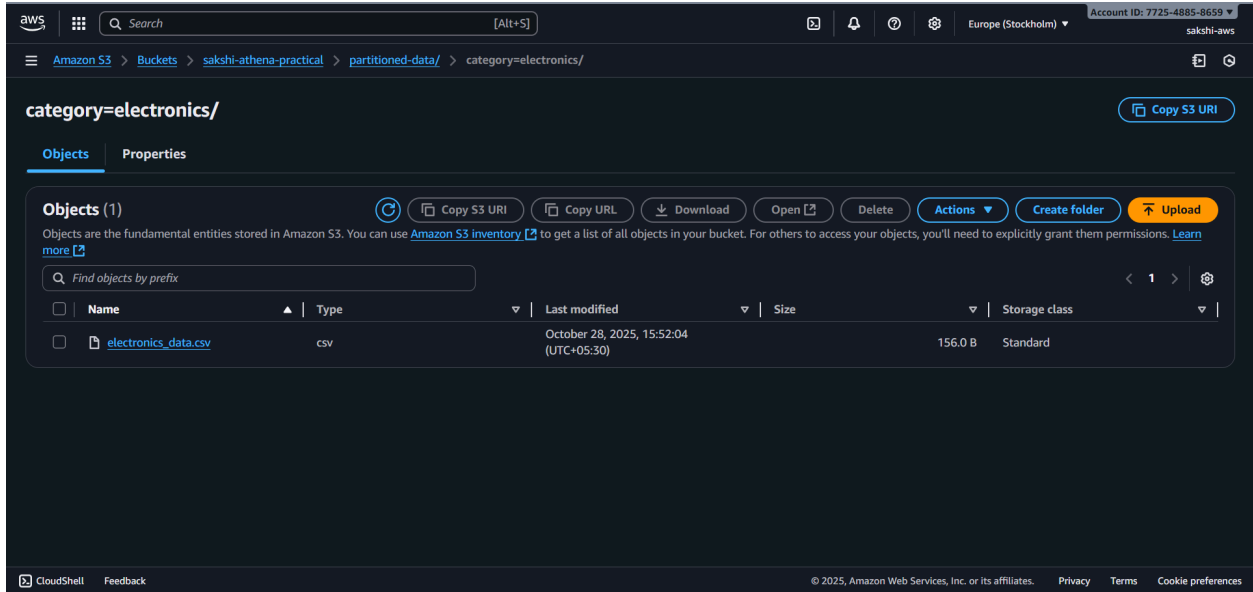
Find objects by prefix

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	clothing_data.csv	csv	October 28, 2025, 15:51:17 (UTC+05:30)	135.0 B	Standard

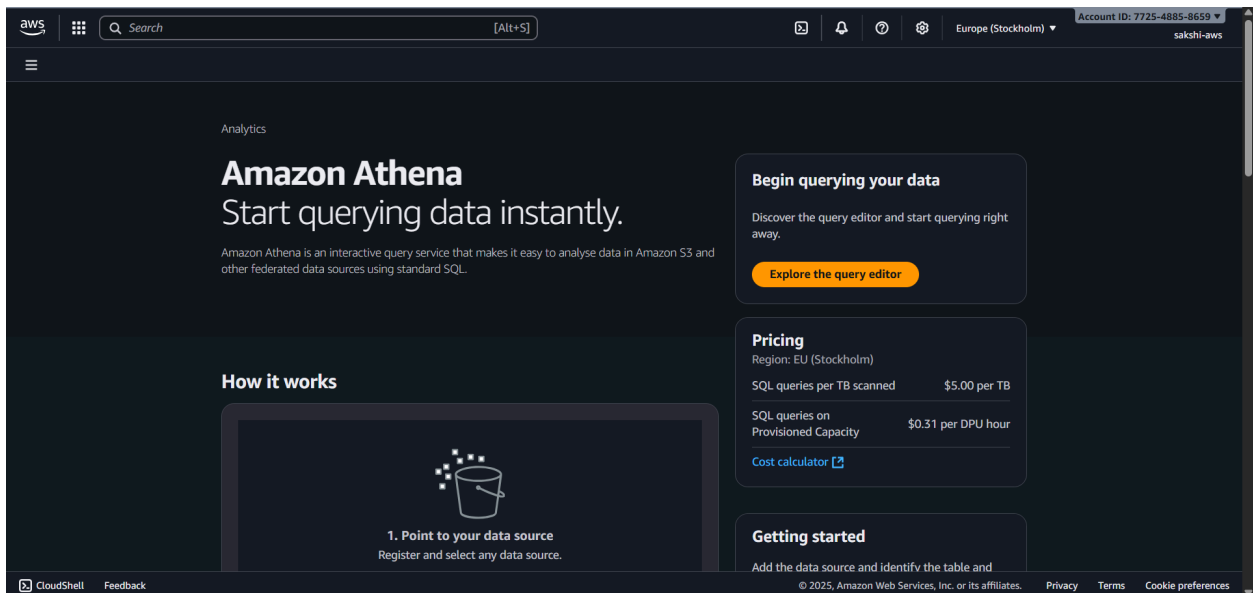
CloudShell

Feedback

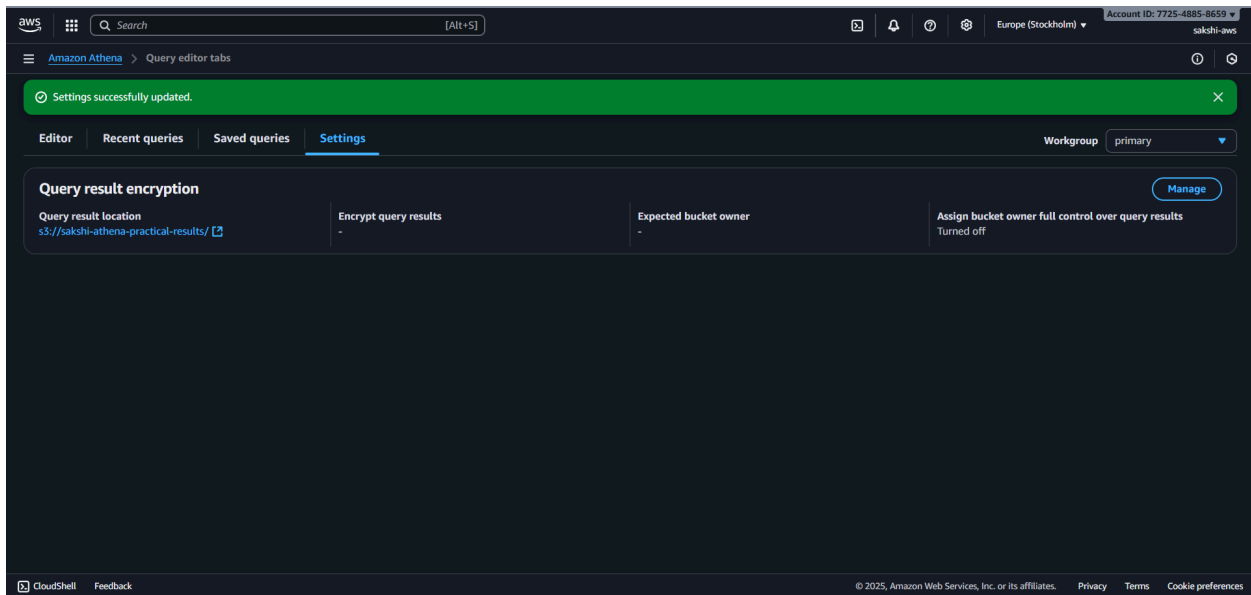
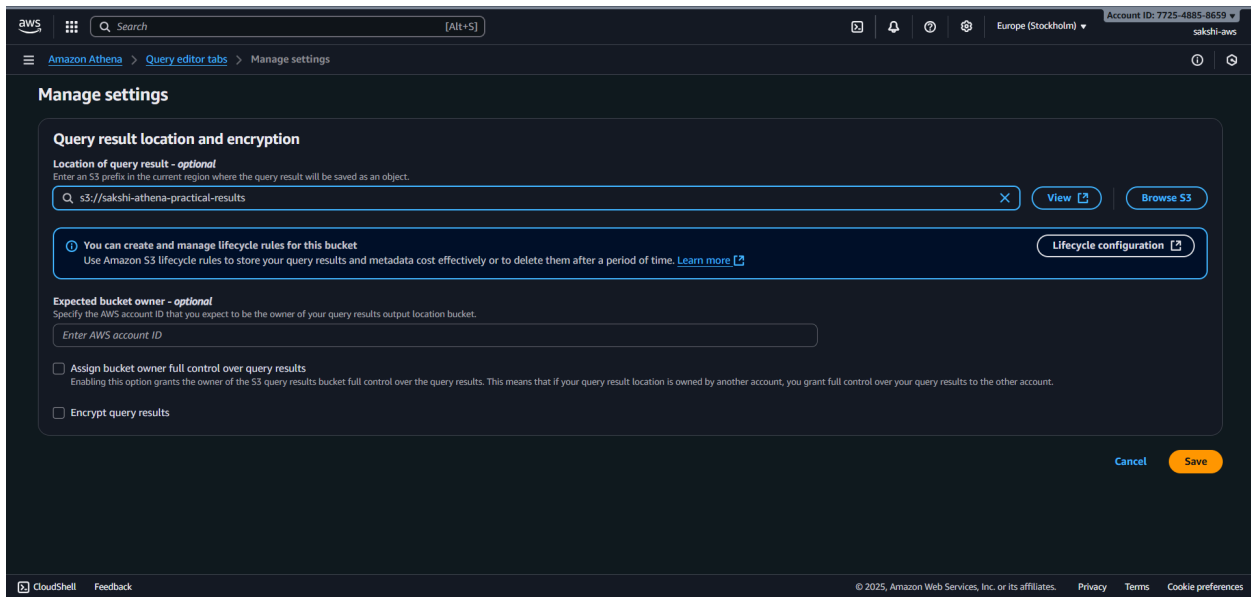
© 2025, Amazon Web Services, Inc. or its affiliates. Privacy Terms Cookie preferences



Launching Athena



Setting query result location



Amazon Athena Query Editor interface showing the execution of a SQL query.

Header: AWS logo, Search bar, [Alt+S], Account ID: 7725-4885-8659, sakshi-aws, Europe (Stockholm).

Navigation: Amazon Athena > Query editor tabs. Editor, Recent queries, Saved queries, Settings. Workgroup: primary.

Data Panel:

- Data source: AwsDataCatalog
- Catalogue: None
- Database: my_csv_database
- Tables and views: Create, Filter tables and views
- Tables (2): partitionedproduct (Partitioned), producttest
- Views (0)

Query Editor:

```
1 CREATE DATABASE IF NOT EXISTS pectest;
```

Query Results:

- Query 1: Completed
- Time in queue: 72 ms
- Run time: 242 ms
- Data scanned: -

Query successful.

Amazon Athena Query Editor interface showing the execution of a SQL query.

Header: AWS logo, Search bar, [Alt+S], Account ID: 7725-4885-8659, sakshi-aws, Europe (Stockholm).

Navigation: Amazon Athena > Query editor tabs. Editor, Recent queries, Saved queries, Settings. Workgroup: primary.

Data Panel:

- Data source: AwsDataCatalog
- Catalogue: None
- Database: pectest
- Tables and views: Create, Filter tables and views
- Tables (2): partitionedproduct (Partitioned), producttest
- Views (0)

Query Editor:

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS pectest.producttest (  
2   'product_id' int,  
3   'product_name' string,  
4   'category' string,  
5   'price' double,  
6   'stock_quantity' int  
7 )  
8 ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'  
9 WITH SERDEPROPERTIES (  
10   'separatorChar' = ',',  
11   'quoteChar' = '"',  
12   'escapeChar' = '\\'  
13 )  
14 LOCATION 's3://aditya-athena-pract6-data/non_partitioned_data/'  
15 TBLPROPERTIES ('skip.header.line.count'='1');
```

Query Results:

- Query 4: Completed
- Time in queue: 76 ms
- Run time: 947 ms
- Data scanned: -

Query successful.

Amazon Athena Query Editor interface showing the execution of a SQL query to create an external table.

Data Source: AwsDataCatalog
Catalogue: None
Database: pectest

Tables and views: partitionedproduct, producttest

Query 4:

```
1 CREATE EXTERNAL TABLE IF NOT EXISTS pectest.partitionedproduct (  
2   'product_id' int,  
3   'product_name' string,  
4   'price' double,  
5   'stock_quantity' int  
6 )  
7 PARTITIONED BY('category' string)  
8 ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.OpenCSVSerde'  
9 WITH SERDEPROPERTIES (  
10   'separatorChar' = ',',  
11   'quoteChar' = '\"',  
12   'escapeChar' = '\\'  
13 )  
14 LOCATION 's3://aditya-athena-pract6-data/partitioned_data/'  
15 TBLPROPERTIES ('skip.header.line.count'='0');
```

Query results: Completed
Time in queue: 83 ms Run time: 925 ms Data scanned: -

Query successful.

Amazon Athena Query Editor interface showing the execution of a SQL query to repair a table.

Data Source: AwsDataCatalog
Catalogue: None
Database: pectest

Tables and views: partitionedproduct, producttest

Query 7:

```
1 MSCK REPAIR TABLE pectest.partitionedproduct;
```

Query results: Completed
Time in queue: 168 ms Run time: 1.249 sec Data scanned: -

Partitions not in metastore: partitionedproduct:category=Books partitionedproduct:category=Clothing partitionedproduct:category=Electronics
Repair: Added partition to metastore partitionedproduct:category=Books
Repair: Added partition to metastore partitionedproduct:category=Clothing
Repair: Added partition to metastore partitionedproduct:category=Electronics

Previewing the partitioned table

The screenshot shows the Amazon Athena console interface. On the left, the 'Data' sidebar is visible with 'Data source' set to 'AwsDataCatalog', 'Catalogue' set to 'None', and 'Database' set to 'pectest'. Under 'Tables and views', the 'partitionedproduct' table is listed as 'Partitioned'. The main editor area shows a SQL query: `SELECT * FROM pectest.producttest LIMIT 10;`. Below the query editor, the 'Query results' section shows a 'Completed' status with a green progress bar. The 'Results (10)' section displays a table with 10 rows of product data.

#	product_id	product_name	category	price	stock_quantity
1	101	Classic Laptop	Electronics	1200.0	50
2	102	Smart-Home Hub	Electronics	150.0	120
3	103	4K Ultra-HD Monitor	Electronics	350.0	75
4	104	Gaming Mouse	Electronics	80.0	300
5	105	Wireless Earbuds	Electronics	199.0	250
6	301	The Great Catcher	Books	14.0	400

The screenshot shows the Amazon Athena console interface. On the left, the 'Data' sidebar is visible with 'Data source' set to 'AwsDataCatalog', 'Catalogue' set to 'None', and 'Database' set to 'pectest'. Under 'Tables and views', the 'partitionedproduct' table is listed as 'Partitioned'. The main editor area shows a SQL query: `SELECT * FROM pectest.partitionedproduct LIMIT 10;`. Below the query editor, the 'Query results' section shows a 'Completed' status with a green progress bar. The 'Results (10)' section displays a table with 10 rows of product data.

#	product_id	product_name	price	stock_quantity	category
1	101	Classic Laptop	1200.0	50	Electronics
2	102	Smart-Home Hub	150.0	120	Electronics
3	103	4K Ultra-HD Monitor	350.0	75	Electronics
4	104	Gaming Mouse	80.0	300	Electronics
5	105	Wireless Earbuds	199.0	250	Electronics
6	301	Marvin's Dangerous Character	60.0	400	Classics

Querying the regular table [data scanned - 0.62 kb]

The screenshot shows the Amazon Athena console interface. On the left, the 'Data' sidebar shows the 'Data source' as 'AwsDataCatalog', 'Catalogue' as 'None', and 'Database' as 'pectest'. Under 'Tables and views', 'partitionedproduct' and 'producttest' are listed. The main editor shows 'Query 4' with the following SQL:

```
1 SELECT category, COUNT(*) as item_count
2 FROM pectest.producttest
3 GROUP BY category;
```

Below the SQL editor, the 'Query results' section shows the query is 'Completed'. The 'Results (5)' table is displayed:

#	category	item_count
1	Books	5
2	Electronics	5
3	Clothing	4

Metadata: Time in queue: 101 ms, Run time: 364 ms, Data scanned: 0.62 KB.

Querying the partitioned table [data scanned - 0.15 kb]

The screenshot shows the Amazon Athena console interface. On the left, the 'Data' sidebar shows the 'Data source' as 'AwsDataCatalog', 'Catalogue' as 'None', and 'Database' as 'pectest'. Under 'Tables and views', 'partitionedproduct' and 'producttest' are listed. The main editor shows 'Query 4' with the following SQL:

```
1 SELECT product_name, price
2 FROM pectest.partitionedproduct
3 WHERE category = 'Books';
```

Below the SQL editor, the 'Query results' section shows the query is 'Completed'. The 'Results (5)' table is displayed:

#	product_name	price
1	The Great Gatsby	15.0
2	A Tale of Two Cities	12.0
3	Learning Python	45.0
4	A Brief History of Time	22.0
5	Dune	18.0

Metadata: Time in queue: 108 ms, Run time: 570 ms, Data scanned: 0.15 KB.

The non-partitioned table `producttest` required Athena to scan the entire `products_full.csv` file to find rows `WHERE category = 'Books'`. The partitioned table `partitionedproduct` only scanned the data inside the `category=Books/` folder. This is shown in the 'Data scanned' metric, which was significantly lower for the partitioned query. This demonstrates how partitioning improves query performance and reduces cost by pruning data.