# aws inforce

JUNE 10 - 12, 2024 | PHILADELPHIA, PA

IAM301

# Users and their data: Modern access and audit patterns on AWS

**Becky Weiss** 

Vice President/Distinguished Engineer AWS



Data

















Data



















#### Data lake built on a strong data foundation





#### A strong data foundation positions you for generative AI



#### Today's topic: Trusted identity propagation



#### **AWS IAM Identity Center: Identity support in AWS**

Identities in your directory



#### **AWS IAM Identity Center: Identity support in AWS**



#### **AWS IAM Identity Center: Identity support in AWS**



#### IAM Identity Center for users and their data



#### IAM Identity Center for users and their data



#### IAM Identity Center: Trusted identity propagation



## Agenda

- Myths and facts: Identity in AWS
- Common challenges with role-based access control for data
- Deep dive into trusted identity propagation in AWS (If you are a builder, you have come to the right place)
- Users and generative AI on the AWS identity foundation: How Amazon Q integrates with identities

## Identity in AWS: Myths and facts

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

## IAM Identity Center: Identity support in AWS



#### Myth: IAM Identity Center replaces my IdP





#### IAM Identity Center



#### Fact: IAM Identity Center integrates with your IdP



#### Fact: IAM Identity Center integrates with your IdP

#### EXAMPLE: MICROSOFT ENTRA ID



#### IAM Identity Center as an enterprise application

## Myth: IAM Identity Center replaces my AWS account federation solution

"I ALREADY HAVE A SOLUTION FOR THAT; I DON'T NEED OR WANT A NEW SOLUTION"



#### Myth: IAM Identity Center replaces my AWS account federation solution

"I ALREADY HAVE A SOLUTION FOR THAT; I DON'T NEED OR WANT A NEW SOLUTION"



## Fact: Synchronizing identities is independent of your AWS account federation

IF YOUR SOLUTION IS WORKING FOR YOU, KEEP USING IT



# Common challenges with role-based access control in a data lake

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

#### Group-based access in a data lake





Red data





Blue data



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

#### Group-based access in a data lake











d



User in blue group







Blue data



#### Group-based access in a data lake



















Blue data













#### **Approach 1: Separate the compute**

User in red group

User in both groups

User in blue group





Red data



Blue data










### **Approach 1: Separate the compute**







Red data

Split infrastructure challenge 3: Cost

User in blue group





Blue data







© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.











Shared EMR cluster



AWS Lake Formation



Blue data



Red data











### Data lake with trusted identity propagation



### Data lake with trusted identity propagation



# Deep dive: Trusted identity propagation in Amazon EMR

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

	💱 EMR Studio > Workspaces 🛛 🗙 🤇	my-data-explore - JupyterLab × +
4	→ C = e-2finp	-emmotebooks-prod-eu-central-1.amazonaws.com/workspace/lab/tree/my-data-explorer.ipvnb
<u> </u>	File Edit View Run Kernel Git	Tabs Settings Help
	Community.	🛙 Launcher X 📕 my-data-explorer.ipynb 🔸
-	Compute	B + % □ □ ▶ ■ C → Code ∨ Ø
÷.	A	
-%*	This Studio	DataFrame[]
0	uses trusted	
-	identity	[9]: Spark.std( Show (Roles ).show() Last executed at 2924-06-09 14:33:36 in 1.295
<b>.</b>	propagation The following	
v	features aren't	++
	supported from a	namespace tableName isTemporary
$\odot$	Studio with	<pre>/tip-demo-db'  adult_census_income  false </pre>
<i>.</i> .	trusted identity	`tip-demo-db`  nyc_listings  false
{}	propagation:	`tip-demo-db` us_cities_demogra  talse   `tip-demo-db`  vehicles  false
	creating an EMR	÷
i	on EC2 cluster	[18]: # Show columns in table vehicles
	template using	<pre>spark.sql("SELECT id, region, price, year, manufacturer, model FROM vehicles WHERE year IS NOT NULL LIMIT 5").show()</pre>
<u></u>	an EMR	Last executed at 2024-05-09 14:36:52 in 3.29s
	Serverless	Spark Job Progress
	application,	
	launching an	++
	EMR on EKS	id region price year manufacturer  model
	a runtime role. To	/7316814884 auburn 33590 2014  gmc sierra 1500 crew
	use these	7316814758 auburn 22590 2010  chevrolet  silverado 1500   7316814080 auburn 32590 2020  chevrolet  silverado 1500 cnew
	features, ask	7316743432 auburn 30990 2017  toyota tundra double cab sr
	your	7316356412 auburn 15000 2013  ford  f-150 xlt
	administrator to	
	create a new	<pre>[19]: spark.sql("DESCRIBE adult_census_income").show()</pre>
	Studio without	Last executed at 2024-05-09 14:37:17 in 5.325
	propagation	· · · · · · · · · · · · · · · · · · ·
	Learn more	<pre>col_name data_type comment </pre>
		workclass  string  NULL
	▼ Compute type	fnlwgt  bigint  NULL
	EMR on EC2 cluster 🚯	education.num  bigint  NULL
	spark-cluster (j-5V V	marital.status  string  NULL
	·	occupation  string  NUL    calationshin  string  NUL

\* EMR Studio > Workspaces
 × C my-data-explore - JupyterLab
 × +

#### c.emrnotebooks-prod.eu-central-1.amazonaws.com/workspace/lab/tree/my-data-explorer.ipynb



💙 🔯 EMR Studio > Workspaces 🗙	K 🔵 my-data-explore - JupyterLab X +
← → C 😁 e-2flnp	emrnotebooks-prod.eu-central-1.amazonaws.com/workspace/lab/tree/my-data-explorer.ipynb
C File Edit View Run Kernel	Git Tabs Settings Help
	☑ Launcher × 🕅 my-data-explorer.ipynb ●
Compute	
Compute	DataFrame[]
	<pre>[9]: spark.sql("SHOW TABLES").show()</pre>
tion	Last executed at 2024-05-09 14:33:36 in 1.29s
ving	++
This Studio	namespace tableName isTemporary
	++  `tip-demo-db`  adult_census_income  false
uses trusted entity	`tip-demo-db`  nyc_listings  false   `tin-demo-db` us_cities_demografalse
identity on:	`tip-demo-db`  vehicles  false
luster	++
propagation	[18]: # Show columns in table vehicles snark.sol("SELECT id. region. price. year. manufacturer. model EROM vehicles WHERE year IS NOT NULL LIMIT 5").show()
	Last executed at 2024-05-09 14:36:52 in 3.29s
an EMR Serverless	A Court Into Deserver
application,	· Spark Job Progress
launching a	··········
EMR on EK US	ers are authenticated to
a runtime ro	0 crew
use these	YOUR IDENTITY PROVIDER Prado 1500 1500 (1500 crew)
features, as	ble cab sr
your administrator to	
create a new	<pre>[19]: spark.sql("DESCRIBE adult_census_income").show()</pre>
Studio without	Last executed at 2024-05-09 14:37:17 in 5.32s
trusted identity	
propagation.	
	++
	age  bigint  NULL    workclass  string  NULL
▼ Compute type	fnlwgt  bigint  NULL education  string  NULL
EMR on EC2 cluster ()	education.num bigint NULL
spark-cluster (j-5V 🔻 🔿	marital.status  string  NULL    occupation  string  NULL
	relationship string NULL

<ul> <li>If EMR Studio &gt; Workspaces</li> </ul>	⊖ my-data-explore - JupyterLab × +					
← → C == e-2flnp	:emmotebooks-prod.eu-central-1.amazonaws.com/workspace/lab/tree/my-data-explorer.ipynb					
C File Edit View Run Kernel Git	Tahs Settings Help					
	Pi Launcher X R mv-data-explorer involu					
Compute	B + X □ □ → Code ∨ Ø					
This Studio uses trusted identity propagation	DataFrame[] [9]: spark.sql("SHOW TABLES").show() Last executed at 2024-05-09 14:33:36 in 1.295					
The following features aren't	++   namespace  tableName isTemporary					
An EMR cluster is running the submitted	<pre>tip-demo-db'  adult_census_income  false   `tip-demo-db'  nyc_listings  false   `tip-demo-db'  us_cities_demogra  false   `tip-demo-db'  vehicles  false  ++</pre>					
Spark commands	<pre>[18]: # Show columns in table vehicles</pre>					
	spark.sql("SELECT id, region, price, year, manufacturer, model FROM vehicles WHERE year IS NOT NULL LIMIT 5").show() Last executed at 2024-05-09 14:36:52 in 3.29s					
Serverless application, launching an	Spark Job Progress					
EMR on EKS cluster, and using	id region price year manufacturer  model  ++					
Compute type	7316814884 auburn  33590 2014        gmc sierra 1500 crew                7316814758  auburn  22590 2010        chevrolet       silverado 1500          7316814989  auburn  39590 2020        chevrolet       silverado 1500         73168143432  auburn  39590 2020        chevrolet       silverado 1500         7316743432  auburn  39990 2017        toyota tundra double cab sr         7316356412  auburn  15000 2013        ford       f-150 xlt					
EMR on EC2 cluster (i)	<pre>[19]: spark.sql("DESCRIBE adult_census_income").show()</pre>					
	Last executed at 2024-05-09 14:37:17 in 5.325					
spark-cluster (j-5V ▼ C	++   col_name data_type comment  ++   age  bigint  NULL					
▼ Compute type	workclass  string  NULL    fnlwgt  bigint  NULL					
EMR on EC2 cluster (3)	education  string  NULL    education.num  bigint  NULL					
spark-cluster (J-5V 🔻 📿	marital.status  string  NULL    occupation  string  NULL    relationshin  string  NULL					

EMR Studio > Workspaces × C my-data-explore - JupyterLab × +							
← → C (s)	25 e-2flnpi ≥emrnote	ebooks-prod.eu-central-1.amazonaws.co	m/workspace/lab/tree/my	r-data-explorer.ipynb			
		Help					
Authenticated user writin	ng PySpark	× my-data-explorer.ip	ynb • Ø				
code in a lunyter not	tehook						
	LEDUUK	taFrame[]					
	Las	ark.sql("SHOW TABLES").show() t executed at 2024-05-09 14:33:36 in 1.2	95				
•							
	+	+	+				
{···}	Query th	ne vehicles	s table				
		unor the sector of		manufactures and 1 c	nou	SDE WARE TO NOT NULL LITE	
A.	[20]: spark.sql( St	elect 10, region, p	rice, year,	manutacturer, model F	ROM Venicles who	ERE year IS NOT NULL LIM	11 5 ).snow()
	Last executed a	t 2024-05-09 14:46:28	in 13.52s				
	Country In	h Dua avaar					
	• Spark Jo	b Progress					
	++-	+++-	+	+			
	1a r	region price year m	anutacturer	model			
	7316814884	auburn 33590 2014	gmc	sierra 1500 crew	-		
	7316814758	auburn 22590 2010	chevrolet	silverado 1500			
	7316814989	auburn 39590 2020	chevrolet	silverado 1500 crew			
	7316743432	auburn 30990 2017	toyota	tundra double cab sr			
▼ Com	7316356412	auburn 15000 2013	ford	f-150 xlt			
EMR on	++-	+++-	+	+			
spark-d	· · · · · · · · · · · · · · · · · · ·	occupation  string  NULL					

- Amazon EMR Studio
- Amazon EMR cluster
- AWS Glue Data Catalog
- AWS Lake Formation
- Amazon S3
- AWS Identity and Access Management (IAM)



→ User's identity propagates through all of these



#### Amazon EMR Studio

- UX application Browser-based notebook environment
- Authenticates the user at the entry point



**EMR Studio** 

#### Amazon EMR Cluster

- Managed cluster of EC2 instances running Spark
- Needs to read table metadata and data on behalf of the authenticated user



#### AWS Glue Data Catalog

- Hive-compatible metastore
- Contains metadata on databases, tables, schema (e.g., columns), etc.
- Contains location of underlying structured data (e.g., in Amazon S3)



#### AWS Glue Data Catalog

- Hive-compatible metastore
- Contains metadata on databases, tables, schema (e.g., columns), etc.
- Contains location of underlying structured data (e.g., in Amazon S3)

Tabl	<b>es</b> (4)						
View an	nd manage all available tables.						
QF	Q Filter tables						
Data	Database = tip-demo-db X Clear filters						
	Name 🔺	Database	$\nabla$	Location	~	Classification	
	adult_census_income	tip-demo-db		s3://tip-demo-data-eu-c	entral-	CSV	
	nyc_listings	tip-demo-db		s3://tip-demo-data-eu-c	entral-	Parquet	
	us_cities_demographics	tip-demo-db		s3://tip-demo-data-eu-c	entral-	CSV	
	vehicles	tip domo dh		sZ://tip.domo.doto.ou.c	ontral	CSV	

Data Catalog







#### AWS Lake Formation

- Permissions service for the AWS Glue Data Catalog
- Who (users/groups) has access to what (data)









Who (users/groups) has access to what (data)



- Storage for underlying table data
- AWS CloudTrail records include user identifier



#### AWS Identity and Access Management (IAM)

- Every step is authenticated and authorized with IAM
- User identity is an overlay on IAM principal identity
- Data perimeters and all other IAM features in effect





#### Trusted identity propagation in Amazon EMR: Our goal



#### Trusted identity propagation in Amazon EMR: Our goal




## **Trusted identity propagation in Amazon EMR: Our goal**





### AWS Lake Formation grants

0374d892-f081-7060-eebc-a9306f	IAM Identi	Table	tip-demo-db	us_cities_demographics
9384b892-e011-7096-f3cf-923bb1	IAM Identi	Table	tip-demo-db	vehicles



Blue data

## **Trusted identity propagation in Amazon EMR: Our goal**

Red data

Blue data



## **Trusted identity propagation in Amazon EMR: Our goal**



### **Trusted identity propagation at work in Amazon EMR**



[5] spark.sql("SELECT region, year, manufacturer, model, price FROM Vehicles WHERE year < 2004 LIMIT 5").show()

region year	manufacturer	model price
auburn 1992  auburn 2001  auburn 1968  auburn 2003  auburn 1966	jeep   ford   volvo  chrysler	cherokee  4500  f450 22500   12990  town & country  9500  1966 C-30 1 ton  2500





#### AWS Lake Formation grants

0374d892-f081-7060-eebc-a9306f	IAM Identi	Table	tip-demo-db	us_cities_demographics
9384b892-e011-7096-f3cf-923bb1	IAM Identi	Table	tip-demo-db	vehicles

### Trusted identity propagation at work in Amazon EMR

User in poth groups

User in blue group

aws

#### No matching grant

[5] spark.sql("SELECT city, state, avg(`total
population`) AS population FROM
US\_Cities\_demographics GROUP BY city,
state ORDER BY population DESC LIMIT 5").show()

An error was encountered: An error occurred while calling o113.sql. : java.io.IOException: com.amazon.ws.emr.hadoop.fs.shaded.com.amazonaws.serv ices.s3.model.AmazonS3Exception: Access Denied...





#### AWS Lake Formation grants

<b>¢</b>	0374d892-f081-7060-eebc-a9306f	IAM Identi	Table	tip-demo-db	us_cities_demographics
)	9384b892-e011-7096-f3cf-923bb1	IAM Identi	Table	tip-demo-db	vehicles

### Trusted identity propagation at work in Amazon EMR



# Trusted identity propagation: What we got

- Permissions are expressed in direct grants to users/groups
- Scalable permissions in one place: AWS Lake Formation
- User identity propagated fully through: Simplifies audit
- And these permissions will apply to other engines too: Amazon Athena, Amazon Redshift, Amazon QuickSight, etc.



# Deeper dive: A word on IAM's role in trusted identity propagation

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Identity propagation example: Amazon EMR

AMAZON EMR, AWS LAKE FORMATION, AND AMAZON S3



# Identity propagation: IAM roles

AMAZON EMR, AWS LAKE FORMATION, AND AMAZON S3



### Amazon EMR Studio

- UX application notebook environment
- Authenticates the user at the entry point



**EMR Studio** 

### Amazon EMR Studio



### Amazon EMR Studio

• UX application – notebook environment



**EMR Studio** 

# How IAM roles and user identities work together

#### **EMR Studio**



### IAM role trust policy (who can assume this role)

```
"Effect": "Allow",
"Principal": {
    "Service": "elasticmapreduce.amazonaws.com"
},
"Action": [
    "sts:AssumeRole",
    "sts:SetContext"
]
```

# How IAM roles and user identities work together



#### IAM role trust policy (who can assume this role)

```
{
    "Effect": "Allow",
    "Principal": {
        "Service": "elasticmapreduce.amazonaws.com"
    },
    "Action": [
        "sts:AssumeRole",
        "sts:SetContext"
    ]
}    New permission: Include
    authenticated user identifier
```



#### IAM principal policy: What it can do

"Effect": "Allow", "Action": "elasticmapreduce:GetClusterSessionCredentials"

"Resource": "arn:aws:elasticmapreduce:111122223..."

Connect to the cluster for an interactive session

aws

# EMR cluster, IAM, and user identity

### Amazon EMR cluster

- Managed cluster of EC2 instances running Spark
- Two IAM roles:
  - EMR EC2 instance profile role: Represents the cluster
  - EMR identity role: Represents the user





## EMR cluster: EC2 instance profile role



EC2 instance role

## EMR cluster: EC2 instance profile role



# EMR security configuration for identity support

Amazon EMR <	Amazon EMR > EMR on EC2: Security configurations > emr-trusted-identity-security-configuration emr-trusted-identity-security-configuration Info				
EMR Serverless ▼ EMR on EC2 Clusters Notebooks and Git repos Events Block public access Security configurations	Encryption At-rest encryption for Turned off	Amazon S3	At-rest encryption for local d Turned off	isk	In-transit encryption Certificate provider type PEM PEM certificate location s3://emr-encryption-in-transit-option-in-
<ul> <li>▼ EMR on EKS Virtual clusters</li> <li>▼ EMR Studio Getting Started</li> </ul>	EC2 Instance metadata service         Minimum instance metadata service         version         Only allow IMDSv2				
Workspaces (Notebooks) What's New Video tour 🖸	Authentication AWS IAM Identity Center				
Compact mode	IAM Identity Center connect Central IAM Identity Center instance arn:aws:sso::583	e application/ssoins-69878138990e6	e9d/apl-b4a4bf47f56d8474	Identity Center Access to a ws:iam::767 :ro	Amazon EMR on EC2 le/EMR-Security-Configuration-Role

# EMR security configuration for identity support

	Amazon EMR <	Amazon EMR emr-trus	Amazon EMR > EMR on EC2: Security configurations > emr-trusted-identity-security-configuration emr-trusted-identity-security-configuration Info			
EMR Serverless		Encrunt	Francestion			
-	EMR on EC2 Clusters Notebooks and Git repos Events Block public access Security configurations	At-rest en Turned off	1cryption for Amazon S3	At-rest encryption for local dis	sk In-transit encryption Certificate provider type PEM PEM certificate location ① s3://emr-encryption-in-tran	nsit-(
• Authentication	EMR on EKS			The EMR clust authentica	ter will operate on be ted user, using this IA	ehalf of an AM role
AWS IAM Identity Cer	nter					
Amazon EMR o	on EC2 to Ident	ity Center conne	ction	IAM Identity Center A	Access to Amazon EMR on EC2	
Central				arn:aws:iam::	:role/EMR-Security-Configura	ation-Role
AM Identity Center	instance					
arn:aws:sso::	:application/ss	oins-69878138990e6e	9d/apl-b4a4bf47f56d8474			









#### EMR cluster

EMR identity role (security configuration)

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



EMR identity role (security configuration)

#### IAM role trust policy (who can assume this role)

```
"Effect": "Allow",
"Principal": {
    "AwS": "111122223333"
},
"Action": [
    "sts:AssumeRole",
    "sts:SetContext"
],
"Condition": {
    "ArnEquals": {
        "aws:PrincipalArn": "arn:aws:iam:...:EMR_EC2_Instance_Role"
```







#### IAM principal policy: What it can do

```
"Effect": "Allow",
"Action": "sso-oauth:CreateTokenWithIAM",
"Resource": "arn:aws:sso:...:application/.../apl-123"
,
"Effect": "Allow",
"Action": ["glue:GetDatabase", "glue:GetTable", ...],
"Resource": "..."
,
"Effect": "Allow",
"Action": "lakeformation:GetDataAccess",
"Resource": "*"
```





(security configuration)

figuration)

"Resource":

Pass user identity onward

#### IAM principal policy: What it can do

```
"Effect": "Allow",
"Action": "sso-oauth:CreateTokenWithIAM",
"Resource": "arn:aws:sso:...:application/.../apl-123"
,
"Effect": "Allow",
"Action": ["glue:GetDatabase", "glue:GetTable", ...],
"Resource": "..."
,
"Effect": "Allow",
"Action": "lakeformation:GetDataAccess",
```







AWS Glue with metadata permissions in AWS Lake Formation

IAM principal policy: What it can do

Access DB/table metadata on behalf of the user

```
"Effect": "Allow",
"Action": "sso-oauth:CreateTokenWithIAM",
"Resource": "arn:aws:sso:...:application/.../apl-123"
```

```
"Effect": "Allow",
"Action": ["glue:GetDatabase", "glue:GetTable", ...],
"Resource": "..."
```

```
"Effect": "Allow",
"Action": "lakeformation:GetDataAccess",
"Resource": "*"
```







AWS Lake Formation data permissions

#### IAM principal policy: What it can do

```
"Effect": "Allow",
"Action": "sso-oauth:CreateTokenWithIAM",
"Resource": "arn:aws:sso:...:application/.../apl-123"
"Effect": "Allow",
"Action": ["glue:GetDatabase", "glue:GetTable", ...],
"Resource": "..."
"Effect": "Allow",
"Action": "lakeformation:GetDataAccess",
"Resource": "*"
```

Request access to Amazon S3 data on behalf of the user

EMR cluster	EMR identity role (security configuration)	AWS Lake Formation data permissions					
	AWS Lake Formation $\qquad  imes$	AWS Lake Formation > Data lake locations					
	Dashboard	Data lake locations (2)					
	▼ Data Catalog	Q s3://tip-demo-data-eu-central-1 X 1 match					
	Databases						
	Tables	Data lake location     ▼     IAM role     ▼     Location Type					
	Data filters	S3://tip-demo-data-eu-central-1       LakeFormation-S3-Location-Role       Amazon S3					
	Data sharing						
	Crawlers 🛂						
	Permissions						
	Data lake permissions						
	LF-Tags and permissions						
	Hybrid access mode						
t access to A	Data locations						
on behalf of	▼ Administration						
	Administrative roles and tasks						
its affiliates. All rights reserved.	Data lake locations						

### Reques data

© 2024, Amazon Web Services, Inc. o



# AWS Lake Formation, IAM, and user identity

### AWS Lake Formation

- Permissions service for the metadata in AWS Glue and data locations in Amazon S3
- Who (users/groups) has access to what (data)







EMR identity role (security configuration)

AWS Lake Formation data permissions





#### IAM role trust policy (who can assume this role)

```
"
"Effect": "Allow",
"Principal": {
    "Service": "lakeformation.amazonaws.com"
},
"Action": [
    "sts:AssumeRole",
    "sts:SetContext"
]
```





#### IAM role trust policy (who can assume this role)

```
"
"Effect": "Allow",
"Principal": {
    "Service": "lakeformation.amazonaws.com"
},
"Action": [
    "sts:AssumeRole",
    "sts:SetContext"
]
```


#### Lake Formation data location role



#### Requesting data from Amazon S3 with user's identity



### IAM authentication and authorization at every step

USER IDENTITY AS AN OVERLAY ATOP IAM



# Adding Amazon S3 "folder" access: S3 Access Grants

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

#### Direct reads and writes to Amazon S3 via Amazon EMR–Spark

[1] oldCarsDF = spark.sql("SELECT region, year, manufacturer, model, price FROM vehicles WHERE year < 2004") **EMR Studio** [2] oldCarsDF.count() 47092 EMR cluster Data Catalog Amazon S3 User in [3] oldCarsDF.write.parquet("s3://accessblue group grants-demo/blue/old-vehicles.parquet") AWS Lake Formation: S3 Access Grants: Metadata and data Data permissions permissions Store the result in Amazon S3

#### Amazon EMR with Lake Formation and direct-to-S3



# Querying a table and writing the result



# Querying a table and writing the result



#### S3 Access Grants: Data access for users and groups

#### Amazon S3 > Access Grants > Europe (Frankfurt) eu-central-1: default

#### Europe (Frankfurt) eu-central-1: default info

S3 Access Grants provides scalable access control to data sets in your S3 buckets. To share an S3 Access Grants instance with external accounts by using the AWS Resource Access Manager console, choose Share instance. With S3 Access Grants, you can use identities from your corporate directory or from AWS Identity and Access Management (IAM) to control access. You can create one S3 Access Grants instance per AWS Region per account.

Share Instance

ALL

default

Delete Instance

S3 Access Grants instance overview Info						IAM Ide	entity Center Info		<b>Deregister</b> Add
Amaz D a cess-g	on Resource Name (ARN) m:aws:s3:eu-central-1:767 grants/default	Creation date ac October 3, 2023, 12:52	2:33 (UTC-04:00)	Account ID 767		IAM Ident	ity Center instance ARN vs:sso::5836 :applic	cation/ssoins-69878138990e6e	9d/apl-baabdf1b1238d660 🖸
Gran	Locations			Use	ers in the	e blu	e group (9	38) can di	irectly
Grants (6) Info				read	d and wr	r <mark>ite</mark> a	it this locat	tion in Amaz	zon S3 <sub>treate Grant</sub>
Q	Find by grant scope or grantee								< 1 > ©
	Grant scope	Grant ID	Permission	Grantee type 🛛 🔻	Grantee ID 🖸	▽	Creation date	▼ Location ID	Application ARN 🔀 🛛 🗸
0	s3://access-grants- demo/blue*	c9613b03- e889-4501-80d0-5f21ae5 R bd8e4	Read, Write	Directory identity group	<u>9384b892-e011-70</u> f3cf-923bb11261d	2096- dc	May 3, 2024, 13:52:41 (UTC-04:00)	default	ALL
	s3://access-grants-	bbc3e56a-			0374d892-f081-70	060-	May 3, 2024, 13:53:12		

eebc-a9306f568594

(UTC-04:00)

Directory identity group

1bda-4df2-854d-

1f960cac48b9

Read, Write

demo/red\*

0

## Audit: CloudTrail directly attributes access to user



# Adding another analytics engine: Identity support in Amazon Athena workgroups

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

## Adding a second analytics engine: Amazon Athena



## Athena SQL query editor in EMR Studio

MR Studio > Query editor						
Editor Recent queries Saved queries	Settings	Workgroup identity-aware-wor				
Data C <	Query 1 : X Query 3 : X Query 4 : X	+ •				
Data source AwsDataCatalog		1.				
Database	SQL Ln 1, Col 1	≥ ≡ ⊚				
tip-demo-db	Run again Explain 🖸 Cancel Clear Create 🔻	Reuse query results up to 60 minutes ago Z				
Tables and views   Create     Q. Filter tables and views	Query results Query stats					
▼ Tables (4) < 1 >	⊘ Completed Time in queue: 609 ms Run time	me: 1.337 sec Data scanned: 3.81 MB				
adult_census_income	Results (10)	Copy View in S3 🖸				
nyc_listings	Q Search rows	< 1 > @				
us_cities_demographics						
vehicles	# ▼ id  ▼ url					
▶ Views (0) < 1 >	1 7312841671 https://bellingham.craigslist.org/ctd/d/bellingham-2015-toyota-camry-xse-sedan/7312841671.html	l bellingham https://bellingham.				
	2 7312841315 https://bellingham.craigslist.org/ctd/d/bellingham-2012-bmw-x5-xdrive35i/7312841315.html	bellingham https://bellingham.				
	3 7312841222 https://bellingham.craigslist.org/ctd/d/bellingham-2010-bmw-x5-xdrive35d/7312841222.html	bellingham https://bellingham.				





S3 Access Grants



AWS Lake Formation





S3 bucket: Structured table data



AWS Glue Data Catalog







© 2024. Amazon Web Services. Inc. or its affiliates. All rights reserved.



AWS Glue Data Catalog Output location access

S3 bucket:

table data



#### User output locations in Athena with S3 Access Grants

Amazon S3 > Access Grants > Europe (Frankfurt) eu-central-1: default

#### Europe (Frankfurt) eu-central-1: default Info

S3 Access Grants provides scalable access control to data sets in your S3 buckets. To share an S3 Access Grants instance with external accounts by using the AWS Resource Access Manager co Grants, you can use identities from your corporate directory or from AWS Identity and Access Management (IAM) to control access. You can create one S3 Access Grants instance per AWS Reg

S3 A	ccess Grants instance overview Info	IAM Identit	y Center Info						
Amaz	on Resource Name (ARN) Creation date m:aws:s3:eu-central-1: October 3, 2023, 12:52: access-grants/default	IAM Identity Co	enter instance ARN application/s						
Grant	Grants Locations								
Gran	nts (7) Info Find by grant scope or grantee	user has action output	cess to a t folder						
	Grant scope	Grant ID 🛛 🔻	Permission 🗢	Grantee type	Grantee ID 🔀 🛛 🔻				
0	s3://access-grants-demo/athena-workgroup-output /63548842-6061-70f8-2e92-46309552daf8*	5563c71a- 4580-440c- 8534-09503e066b7f	Read, Write	Directory identity user	<u>63548842-6061-70f8</u> -2e92-46309552daf8				
0	s3://access-grants-demo/athena-workgroup-output /b374e8e2-d0d1-70a7-3041-07f336187b3a*	66ce95e5-53d2-43ff- b0f3-ccdc01b1b969		Directory identity user	<u>b374e8e2-</u> d0d1-70a7-3041-07f <u>336187b3a</u>				

#### User output locations in Athena with S3 Access Grants

Amazon S	<u>Amazon S3</u> > <u>Buckets</u> > <u>access-grants-demo</u> > athena-workgroup-output/				
athe	na-workgroup-output/				
Objects Properties					
Objects (2) Info C C Copy S3 URI C Objects are the fundamental entities stored in Amazon S3. You can use Amazon S3 inventory C to get a list of all objects					
Q Find objects by prefix User-specific output folders					
	Name	from Athena queries			
	<b>63548842-6061-70f8-2e92-46309552daf8/</b>	Folder			
	<b>b</b> 374e8e2-d0d1-70a7-3041-07f336187b3a/	Folder			

# Identity in generative AI: Amazon Q Business and Amazon Q Developer

© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

#### Generative AI atop a strong data foundation



### **Example: Amazon Q Business application**

✓ S Amazon Q × +					
$\leftarrow \rightarrow$	C 25 4in chat.qbusiness.us-east-1.on.aws/#/	chat		\$ \$	
Q Chat	S Conversations ☑ <				
E Apps	Recent chats What resources should be authorized as part of an API call?				
Ш Library	May 08, 2024 What is a Service Principal in IAM?		$\overline{\mathbf{O}}$		
	May 06, 2024		<b>Amazon Q Business</b>		
			Your generative AI assistant for work		
		$\bigcirc$	Ask me a question about how AWS customers expect IAM to work.		
			Enter a prompt		
		Ø	Responding from approved sources		

#### **Example: Amazon Q Business application**

Amazon Q × +					
$\leftarrow \rightarrow$	C 😅 4in chat.qbusiness.us-east-1.on.aws	chat 🦉	☆		
Q Chat	Conversations	.chat.qbusiness.us-east-1.on.aws/#/chat			
-	Recent chats	Not the AWS console:			
Apps	What resources should be authorized as part of an API call?	authenticates users to my ld	Ρ		
Ф	May 08, 2024				
Library	What is a Service Principal in IAM?				
	May 06, 2024	Amazon O Business			
		Your generative AL assistant for work			
		Four generative Ar desistant for work			
		Ask me a question about how AWS customers expect IAM to work.			
		Enter a prompt			
		Responding from approved sources			
_					

#### **Example: Amazon Q Business application**

C Conversations   Conversations   Conversations   What reso   Awy 08, 2024   What is a Service Principal in IAM?   May 08, 2024   The principal in IAM? May 08, 2024 The principal in IAM? The pri	Y S Amazon Q X +						
Cert Conversations   Wat reso   Wy 08.2024   What is a Spart of an API call?   May 08, 2024   What is a Service Principal in IAM?   May 06, 2024	 ← →	C 🖙 4in	chat.qbusiness.us-east-1.on.aws/#/chat	Ø \$			
Conversations Recent chats What resources should be authorized as part of an API call? What is a Service Principal in IAM? May 06, 2024 Recent chats What is a Service Principal in IAM? May 06, 2024 Recent chats							
May 06, 2024	Chat Apps Library	Converse Recent chats What resol as part of a May 08, 2024 What is a S May 06, 2024	ations I < < Conversations I < < Recent chats What resources should be authorized as part of an API call? May 08, 2024 What is a Service Principal in IAM?	<image/> <section-header><section-header><section-header><section-header><section-header><section-header><section-header><section-header></section-header></section-header></section-header></section-header></section-header></section-header></section-header></section-header>			
r a prompt			May 00, 2024				
bonding from approved sources				ponding from approved sources			

# A strong data foundation on AWS: Your users, your data

#### Data lake built on a strong data foundation

Accessing

















Data







# Identity support in AWS analytics services



\_\_\_\_\_\_ مە

- Amazon EMR
- Amazon Athena
- AWS Glue Data Catalog
- Amazon Redshift
- Amazon QuickSight



# Identity support in AWS data permission services



AWS Lake Formation



مح

- Amazon S3 Access Grants
- Amazon Redshift

# Identity support in AWS generative AI services



aws

ullet

- Amazon Q Business
- Amazon Q Developer

# Supporting identity from your own applications

