AWS re:Inforce

JUNE 10 - 12, 2024 | PHILADELPHIA, PA

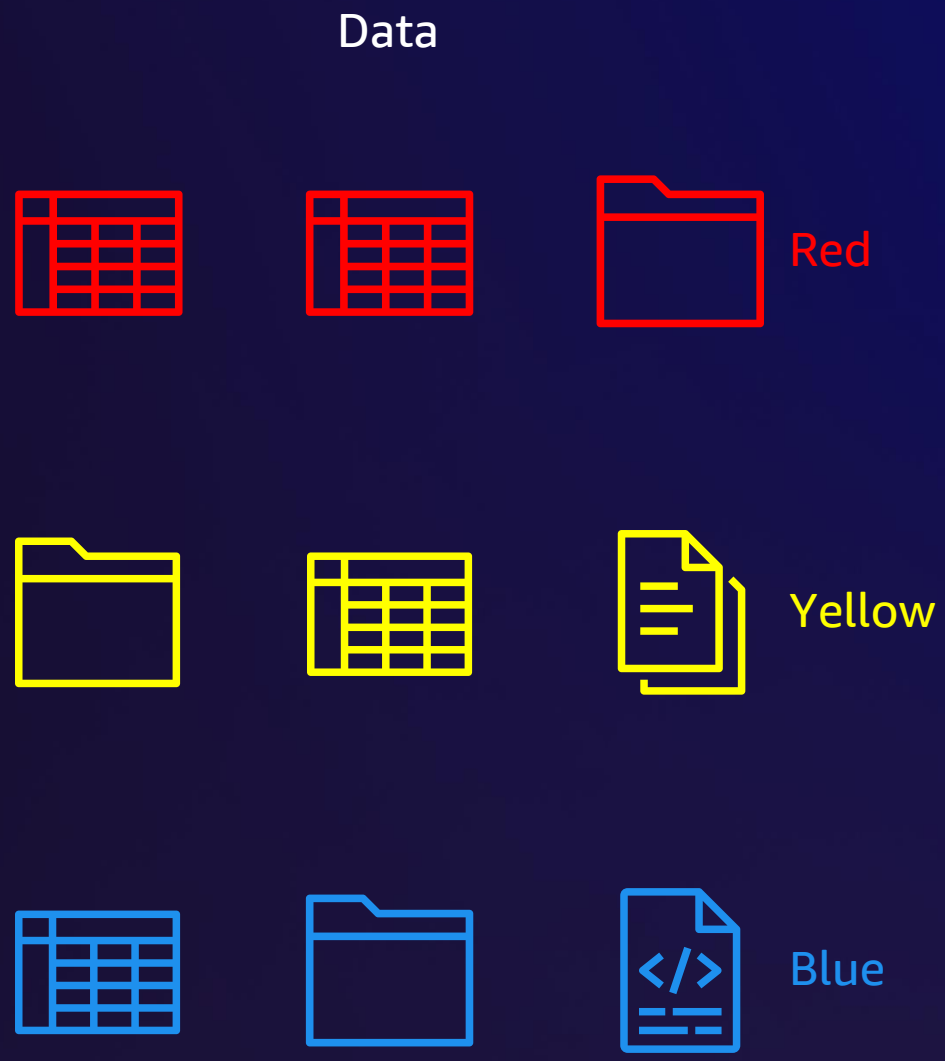
IAM301

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

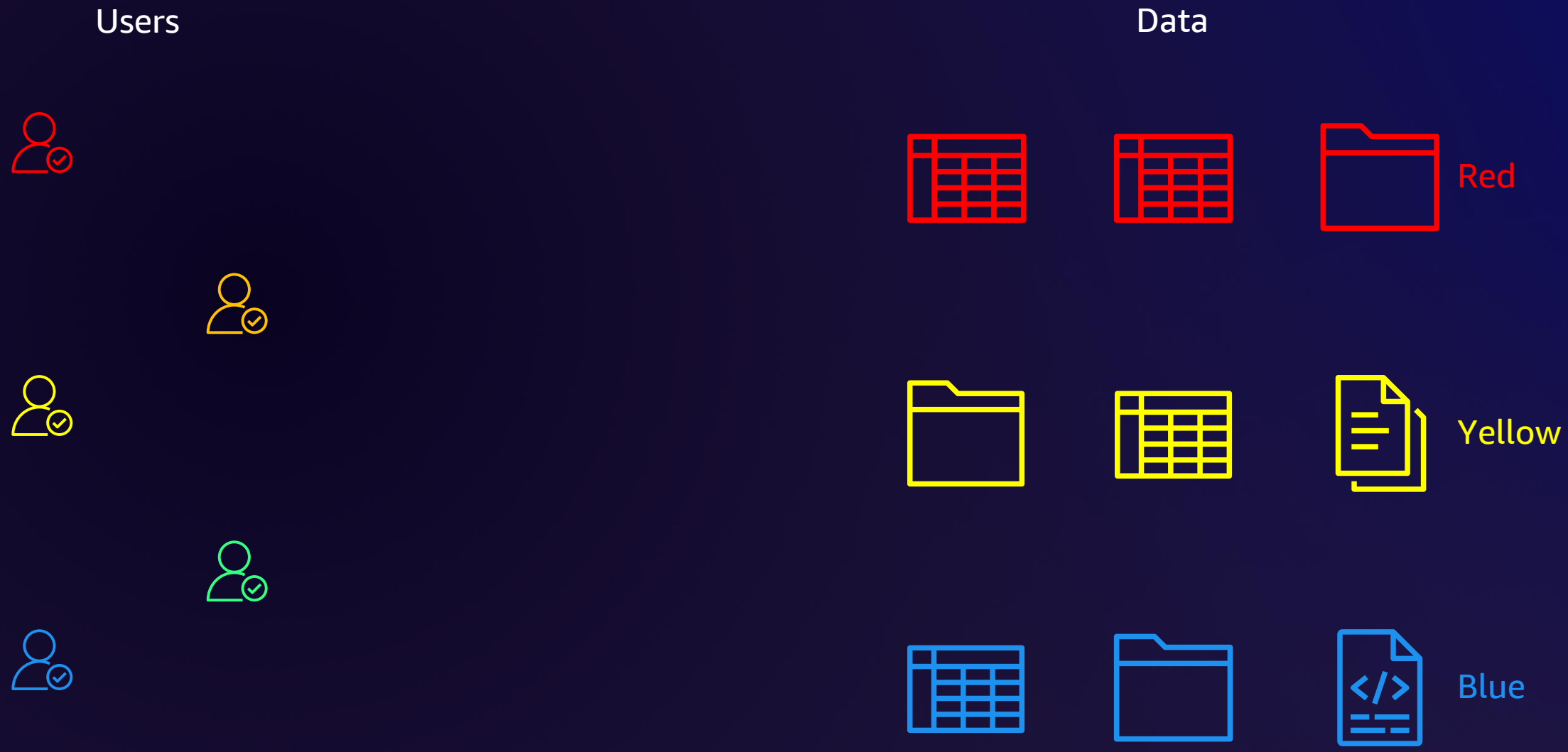
Becky Weiss

Vice President/Distinguished Engineer
AWS

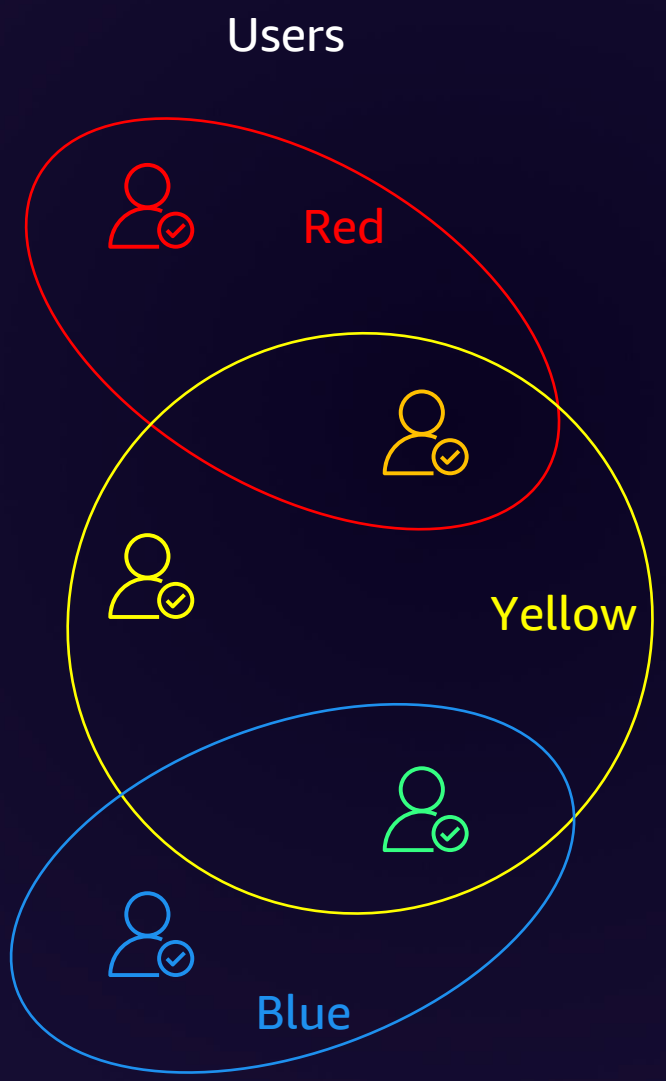
How almost every business thinks about access to data

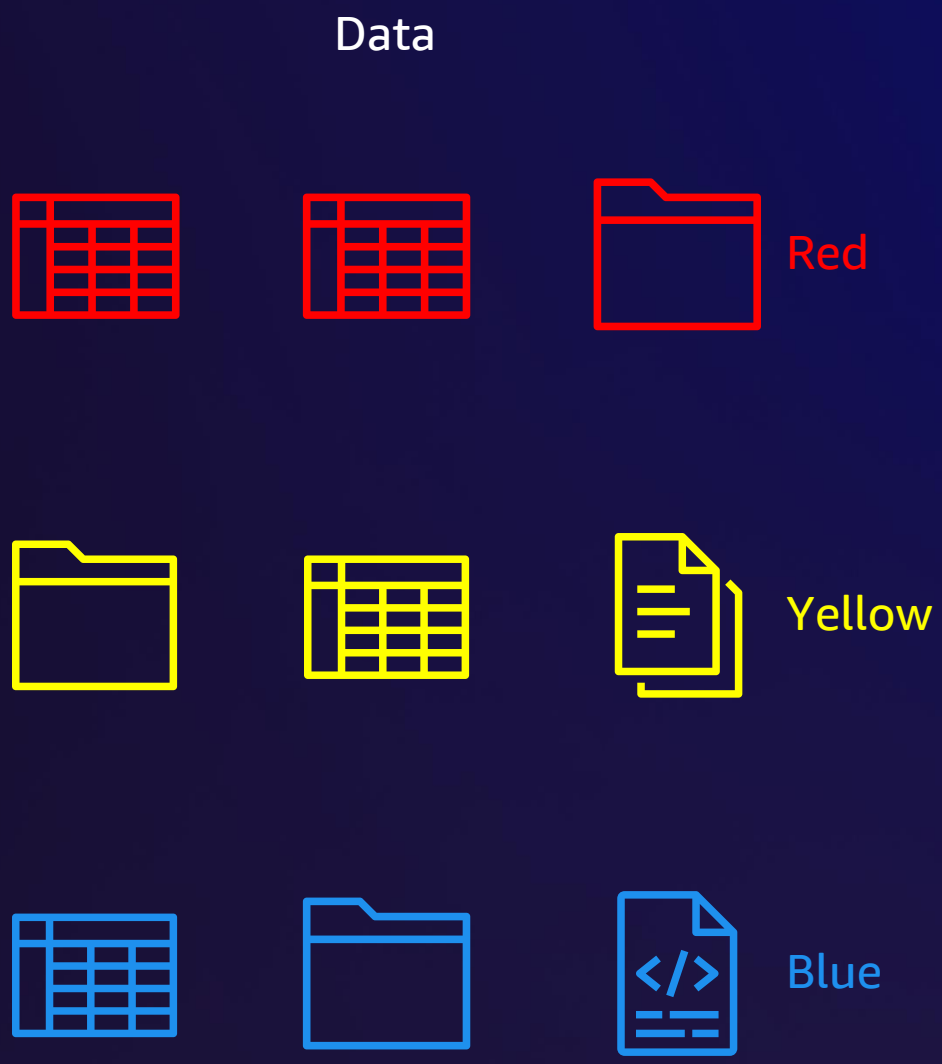


How almost every business thinks about access to data



How almost every business thinks about access to data

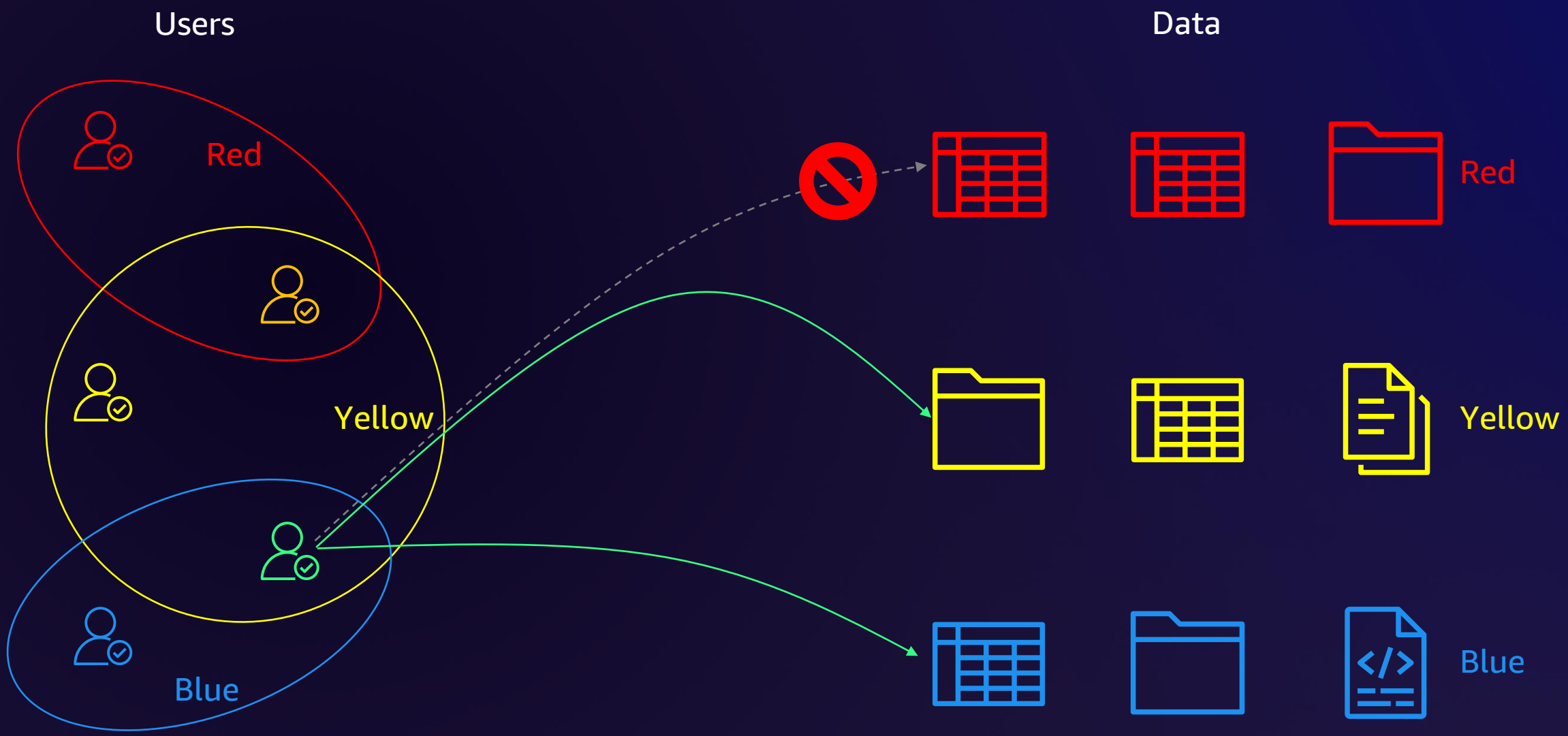




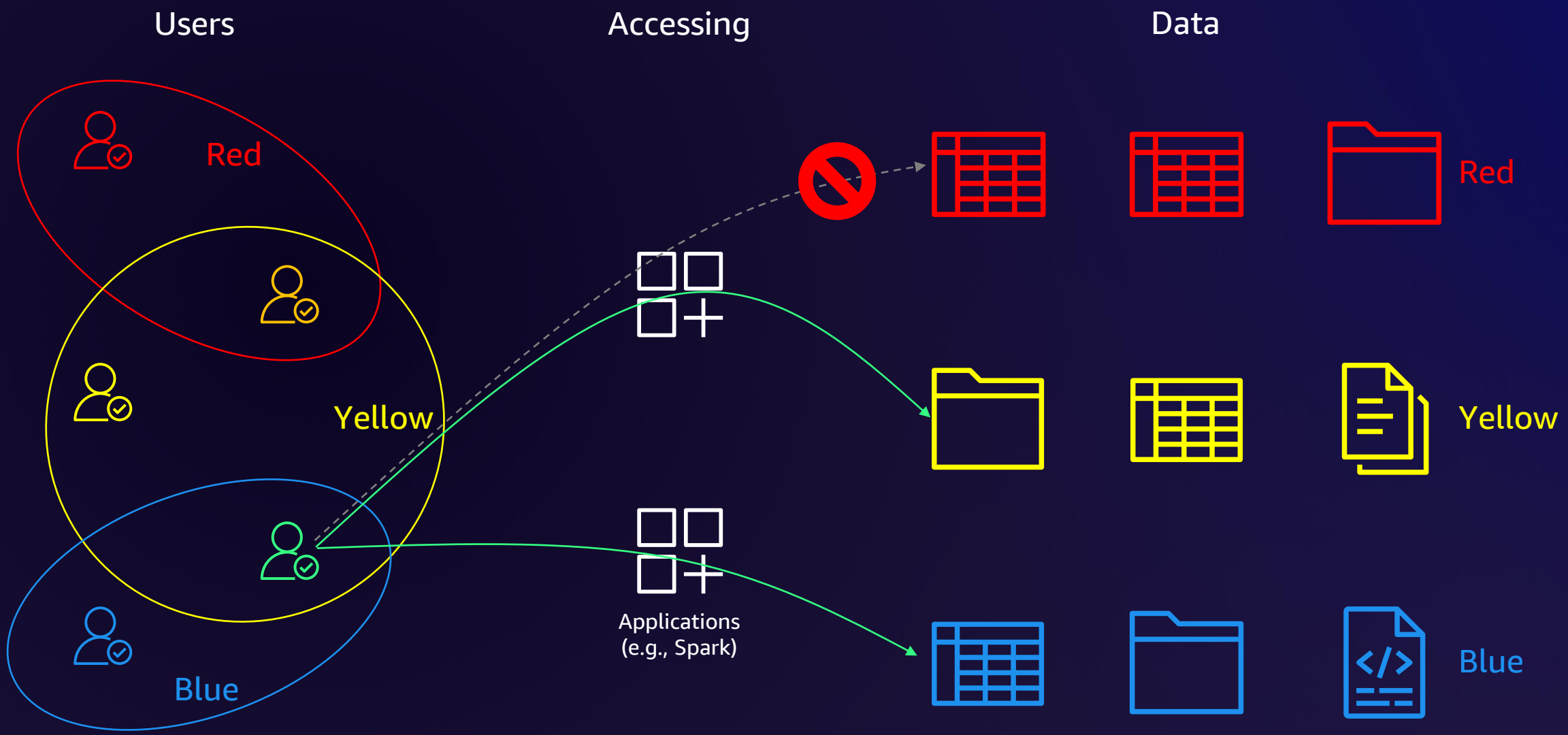
How almost every business thinks about access to data

How almost every business thinks about access to data

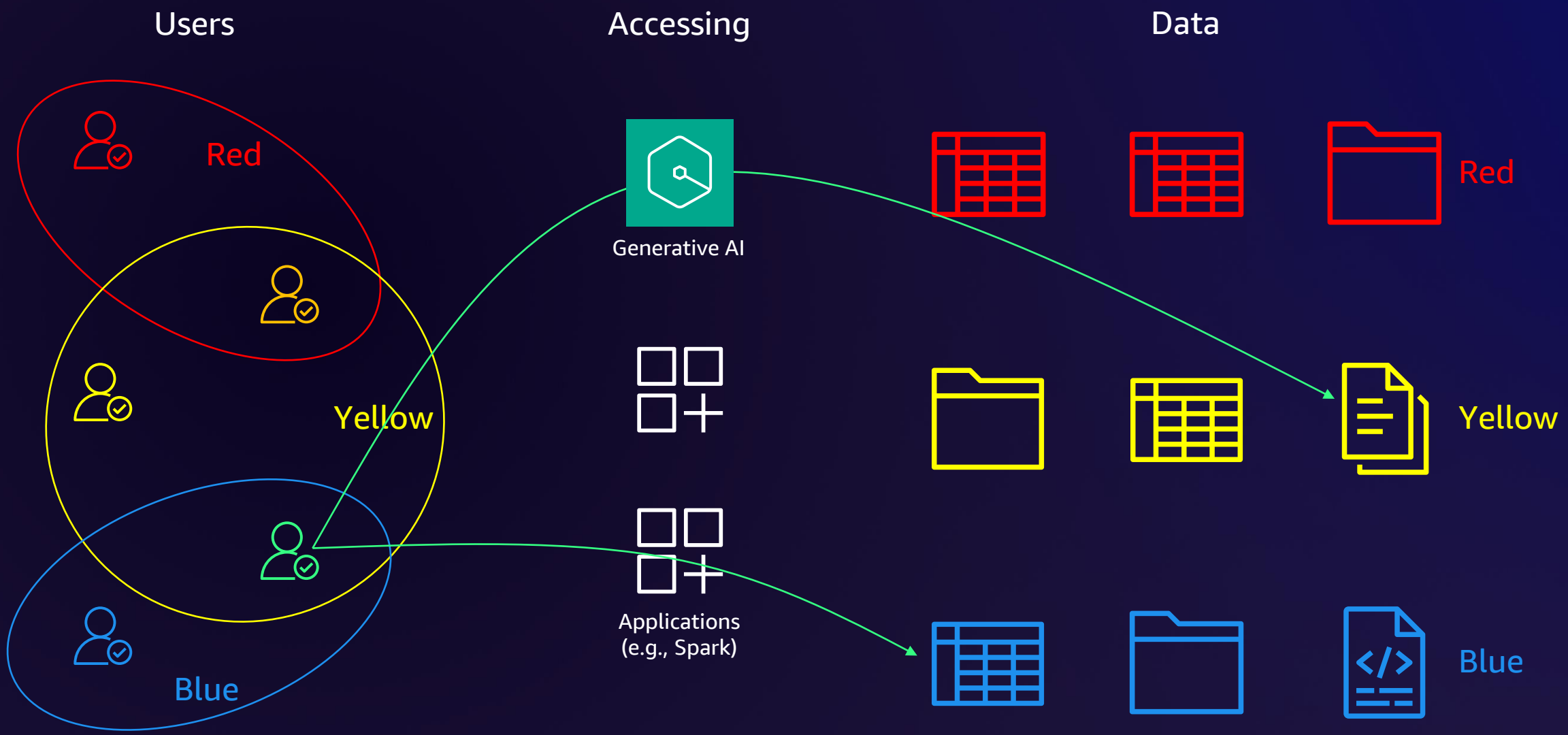
How almost every business thinks about access to data



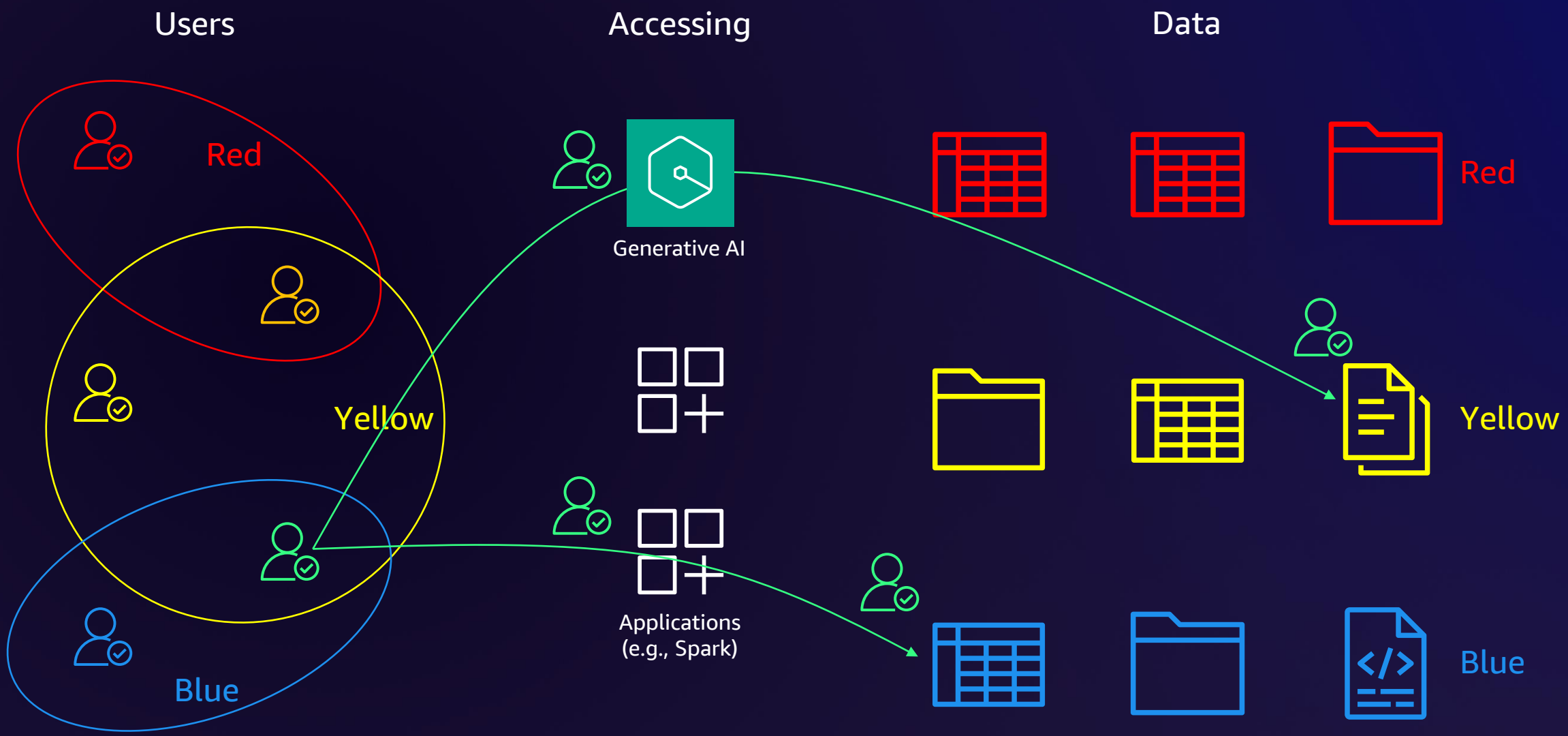
Data lake built on a strong data foundation



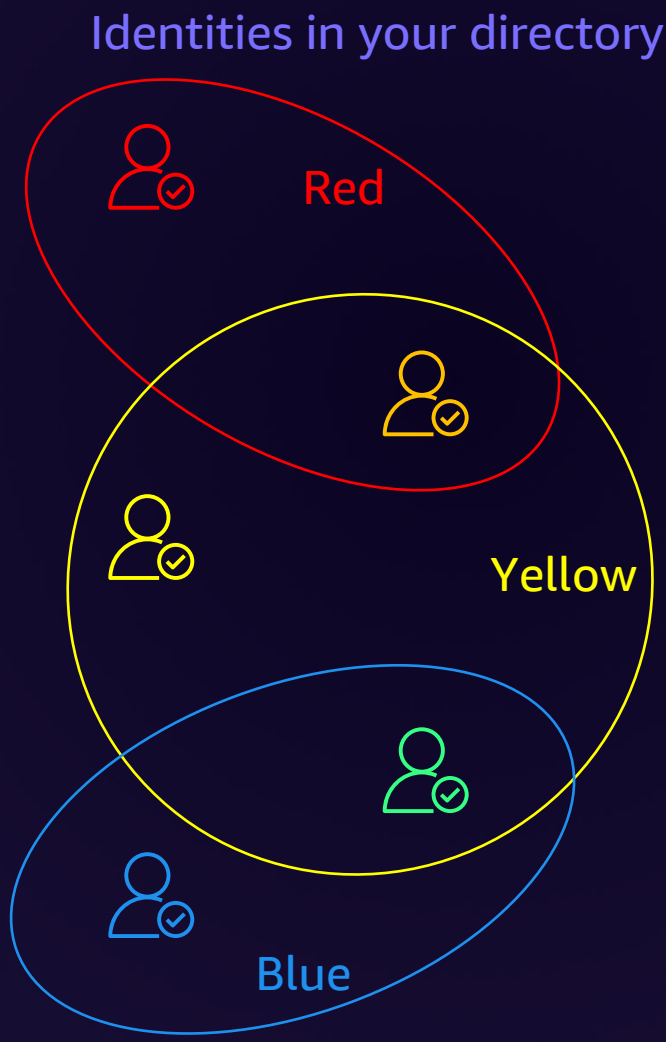
A strong data foundation positions you for generative AI



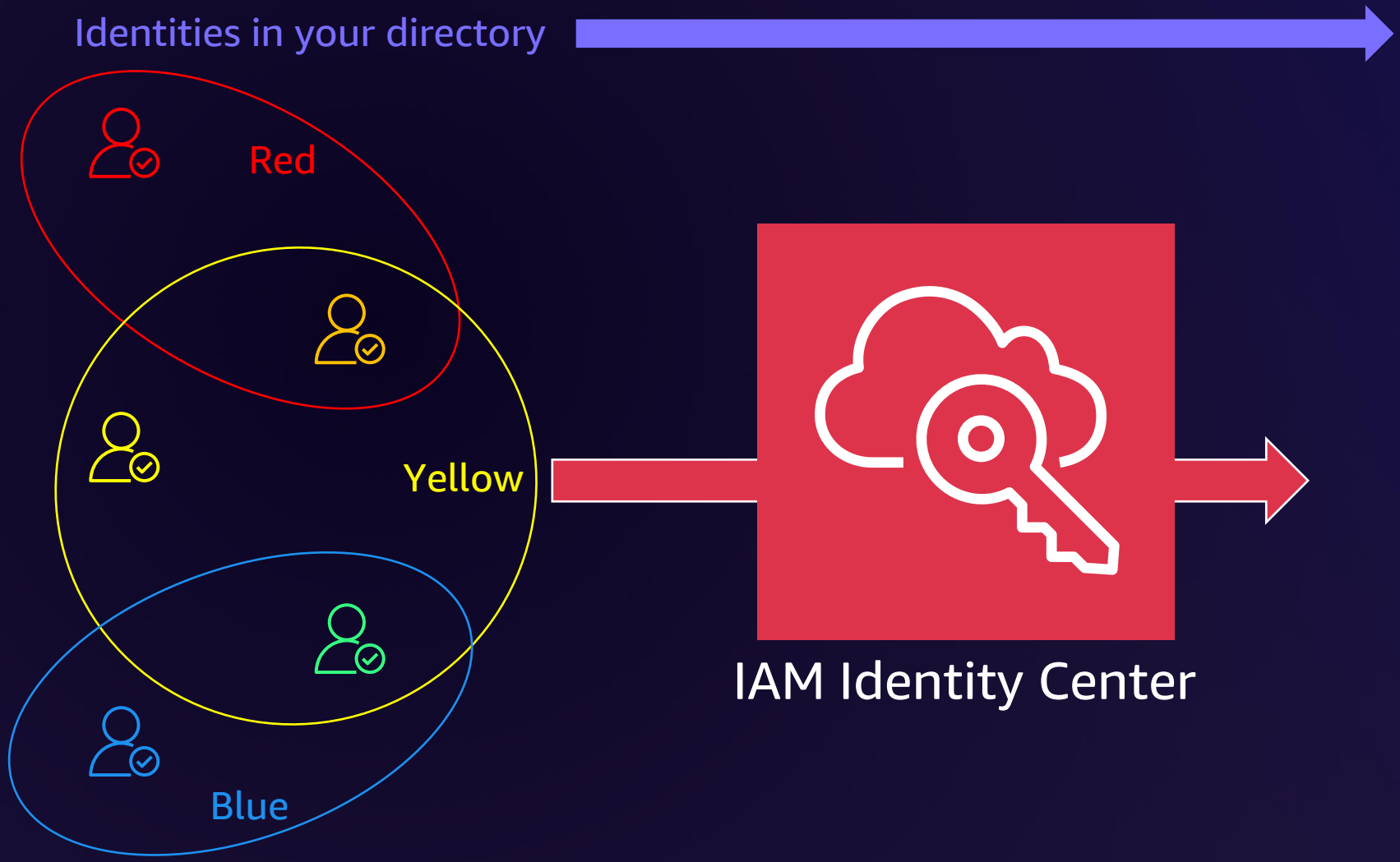
Today's topic: Trusted identity propagation



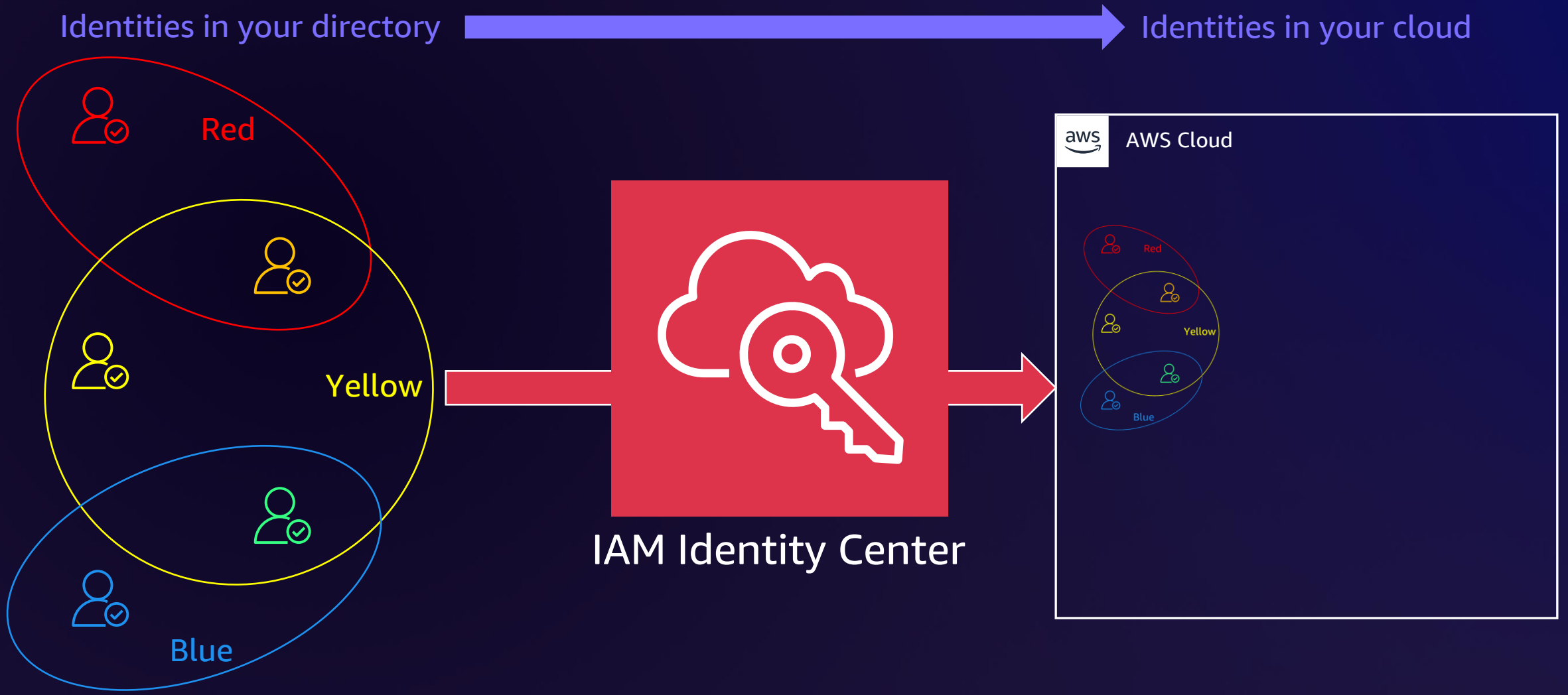
AWS IAM Identity Center: Identity support in AWS



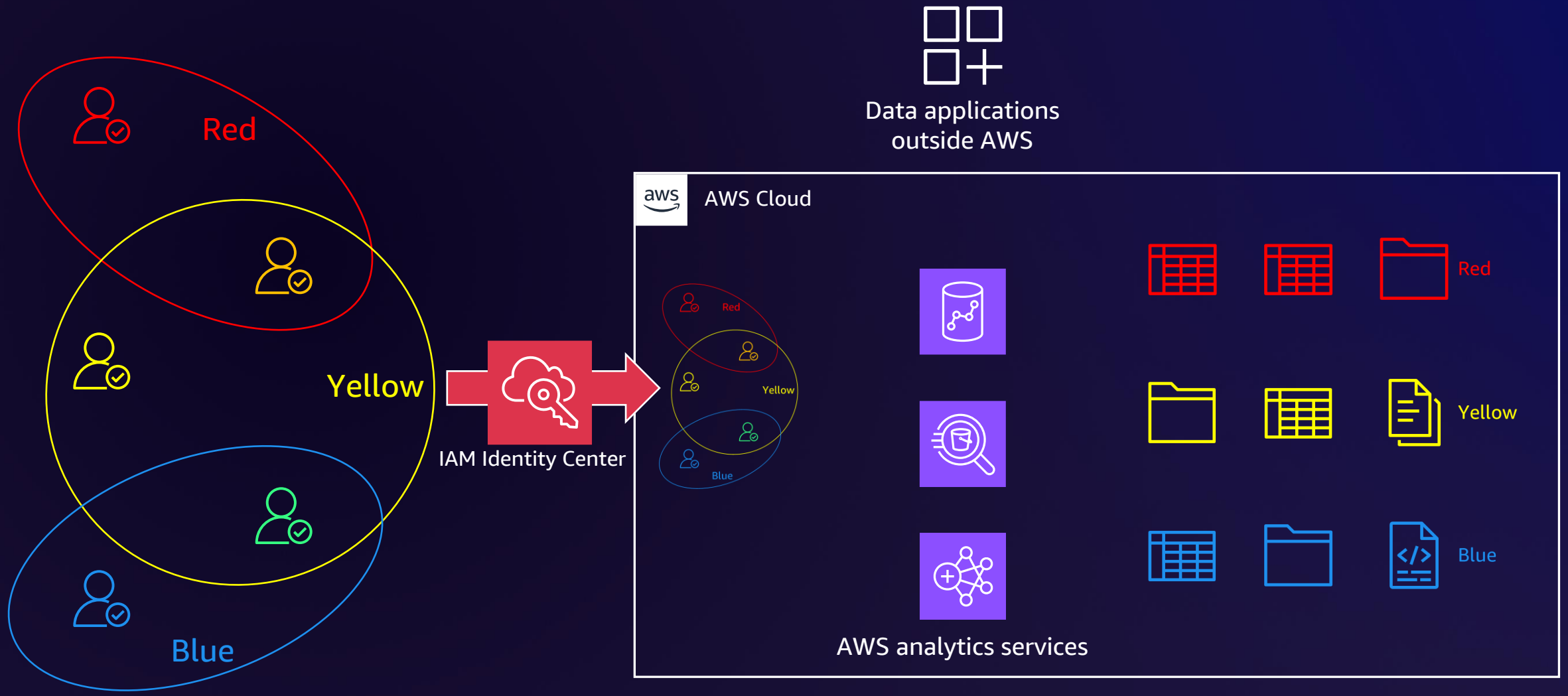
AWS IAM Identity Center: Identity support in AWS



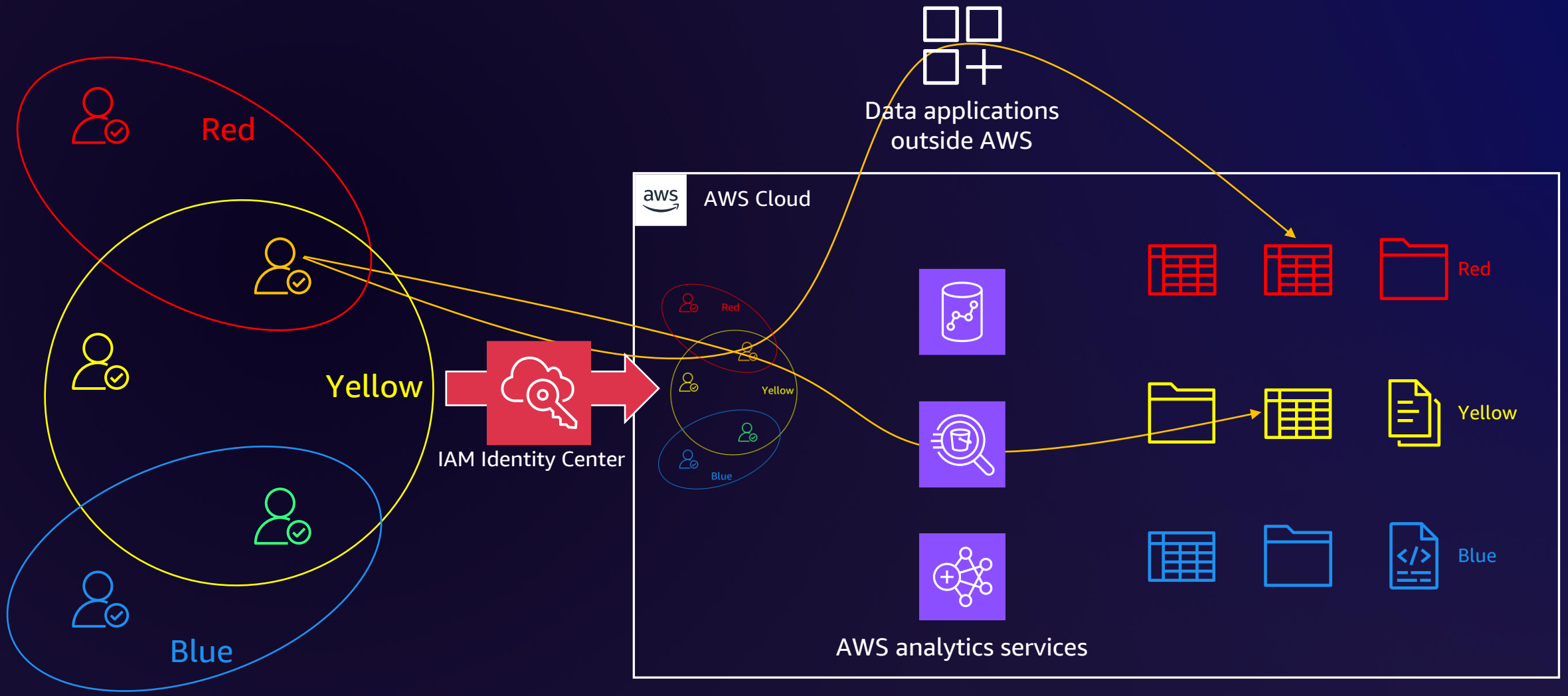
AWS IAM Identity Center: Identity support in AWS



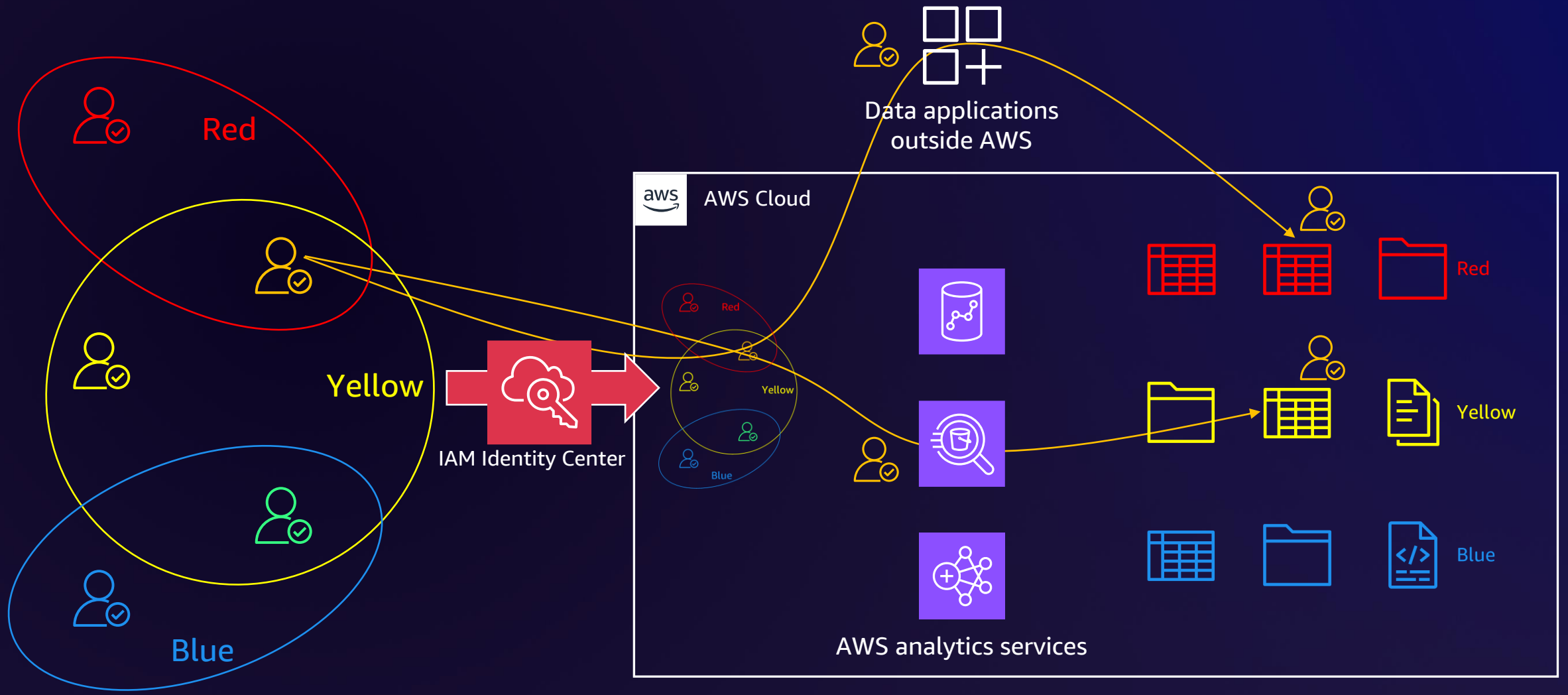
IAM Identity Center for users and their data



IAM Identity Center for users and their data



IAM Identity Center: Trusted identity propagation

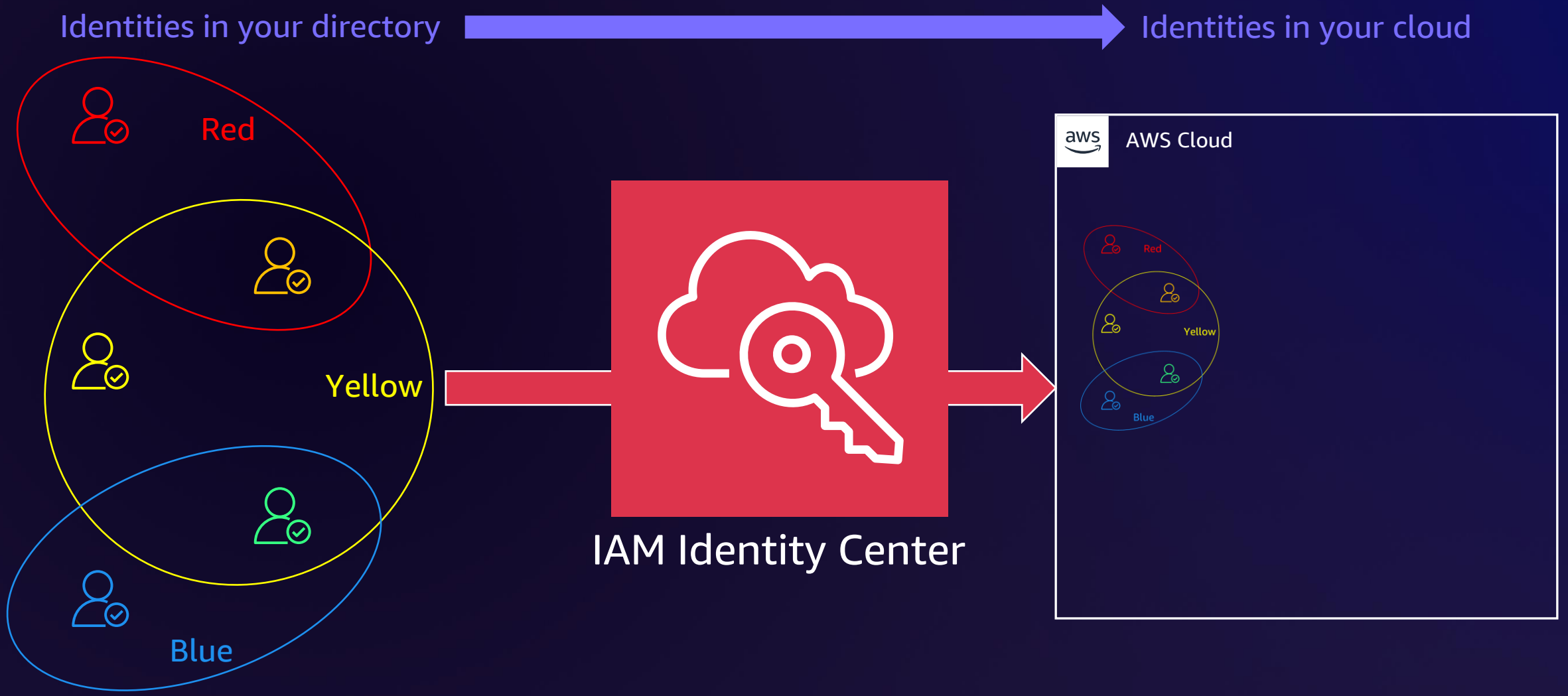


Agenda

- 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

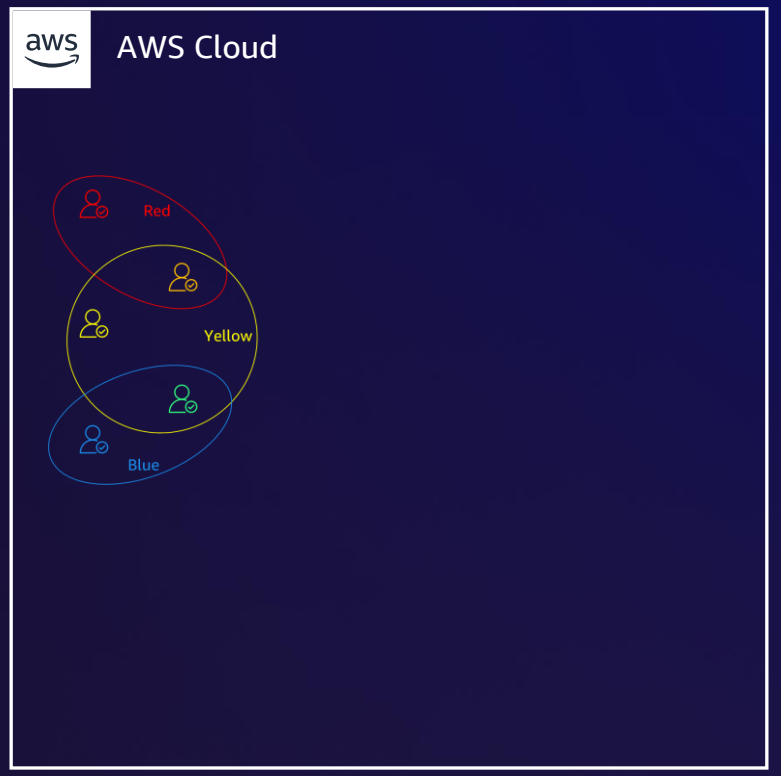
Identity in AWS: Myths and facts

IAM Identity Center: Identity support in AWS

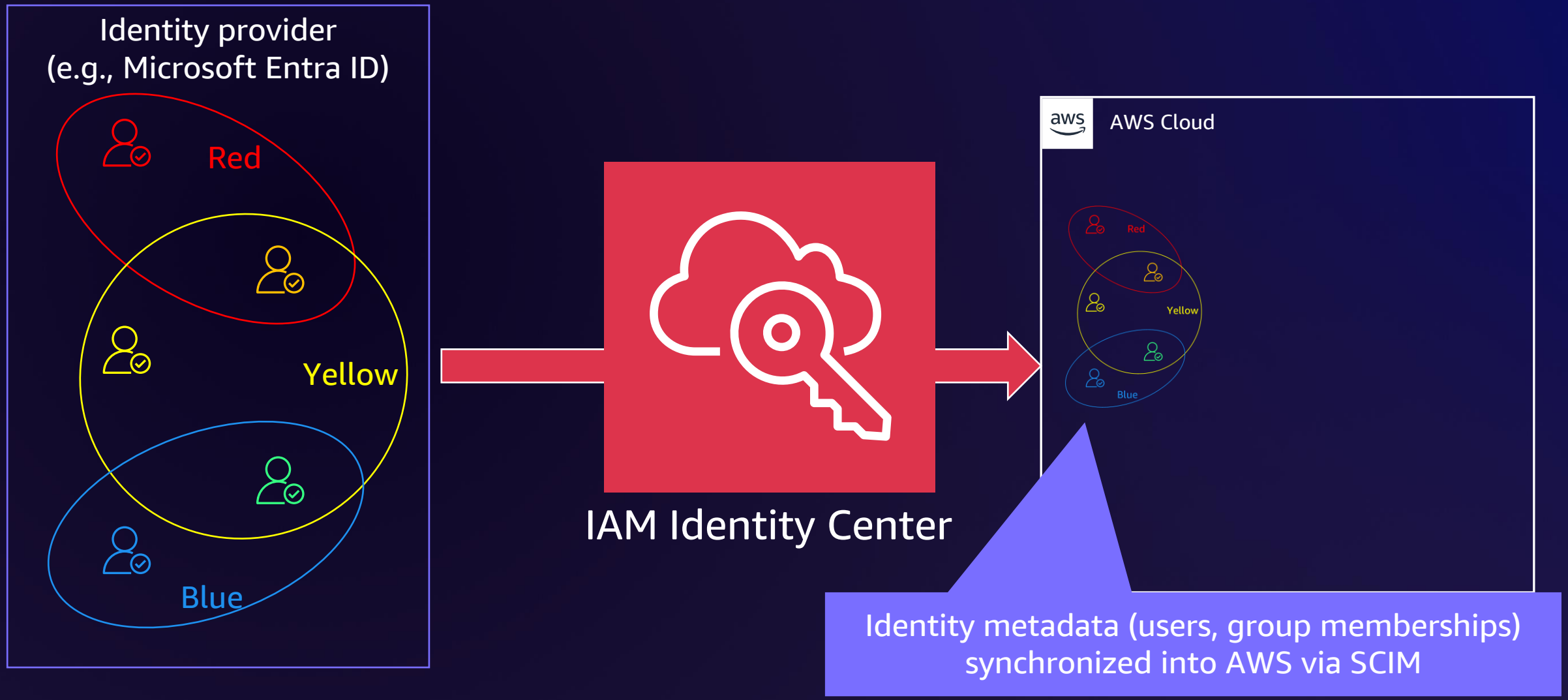


Myth: IAM Identity Center replaces my IdP

IAM Identity Center

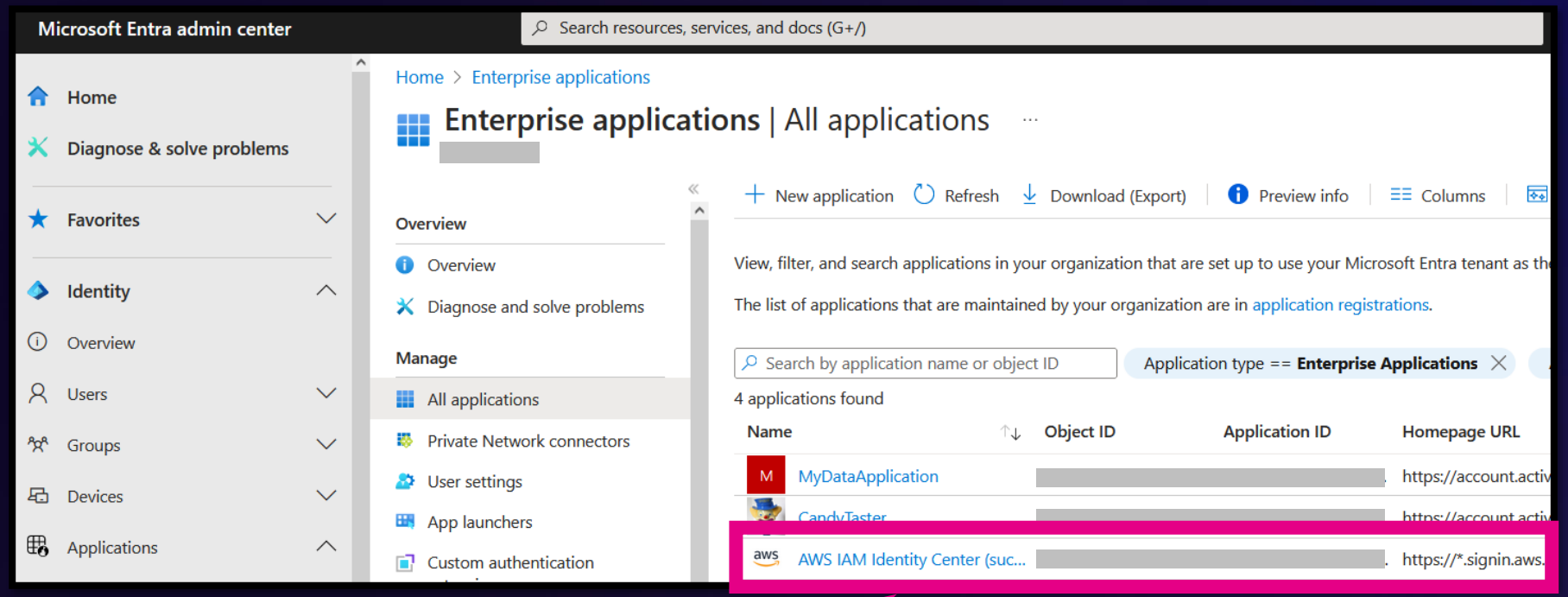


Fact: IAM Identity Center integrates with your IdP



Fact: IAM Identity Center integrates with your IdP

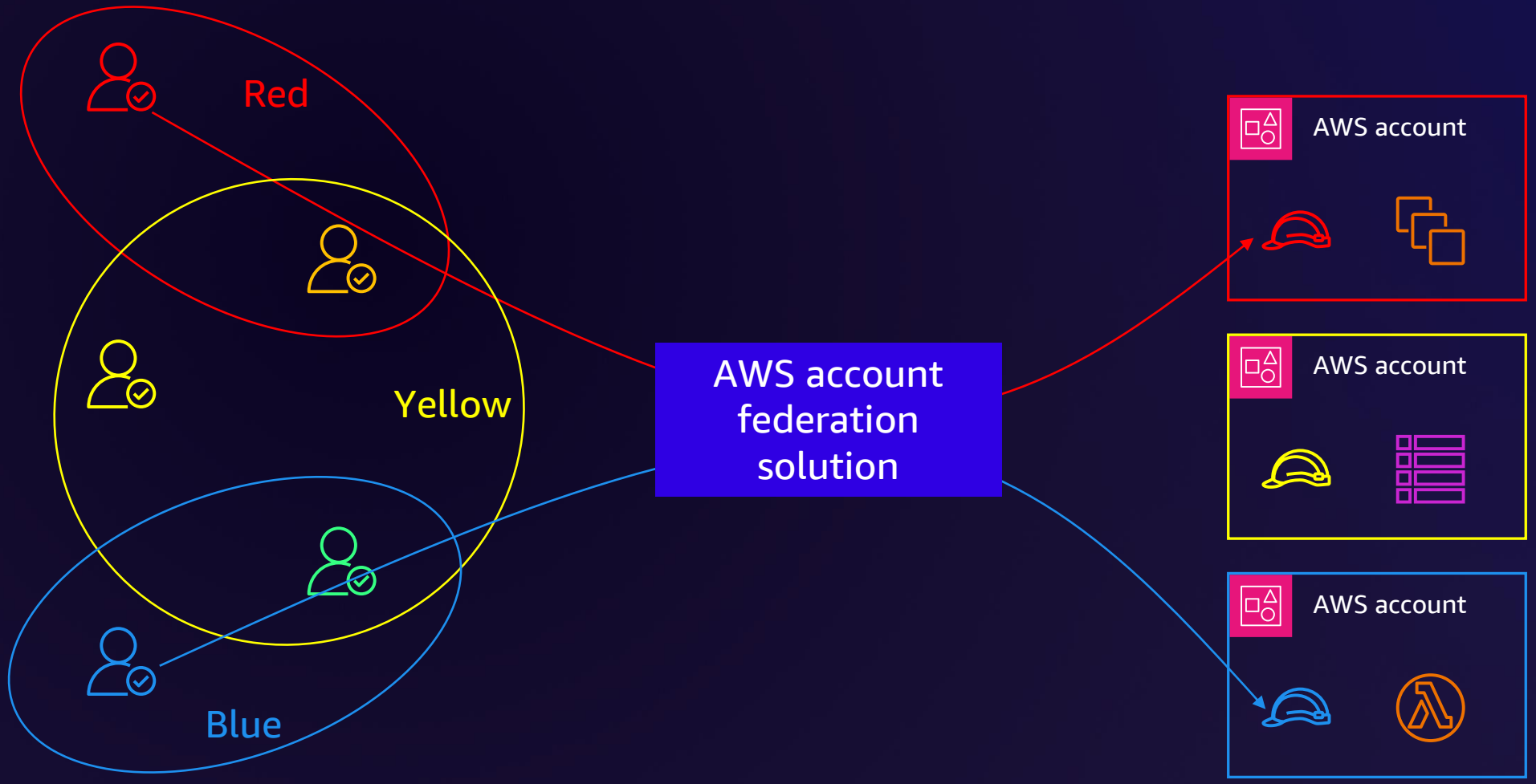
EXAMPLE: MICROSOFT ENTRA ID



IAM Identity Center as an enterprise application

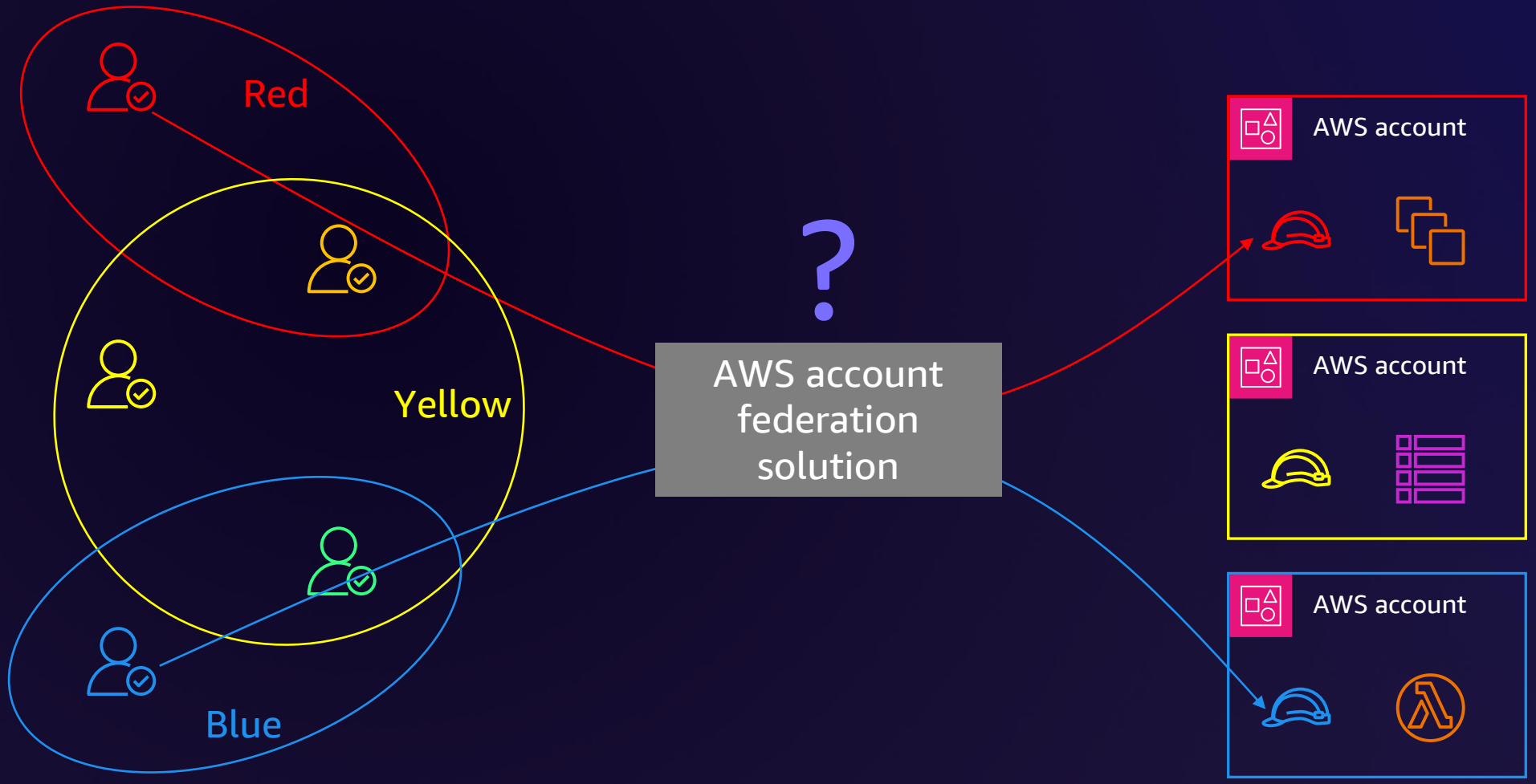
Myth: IAM Identity Center replaces my AWS account federation solution

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



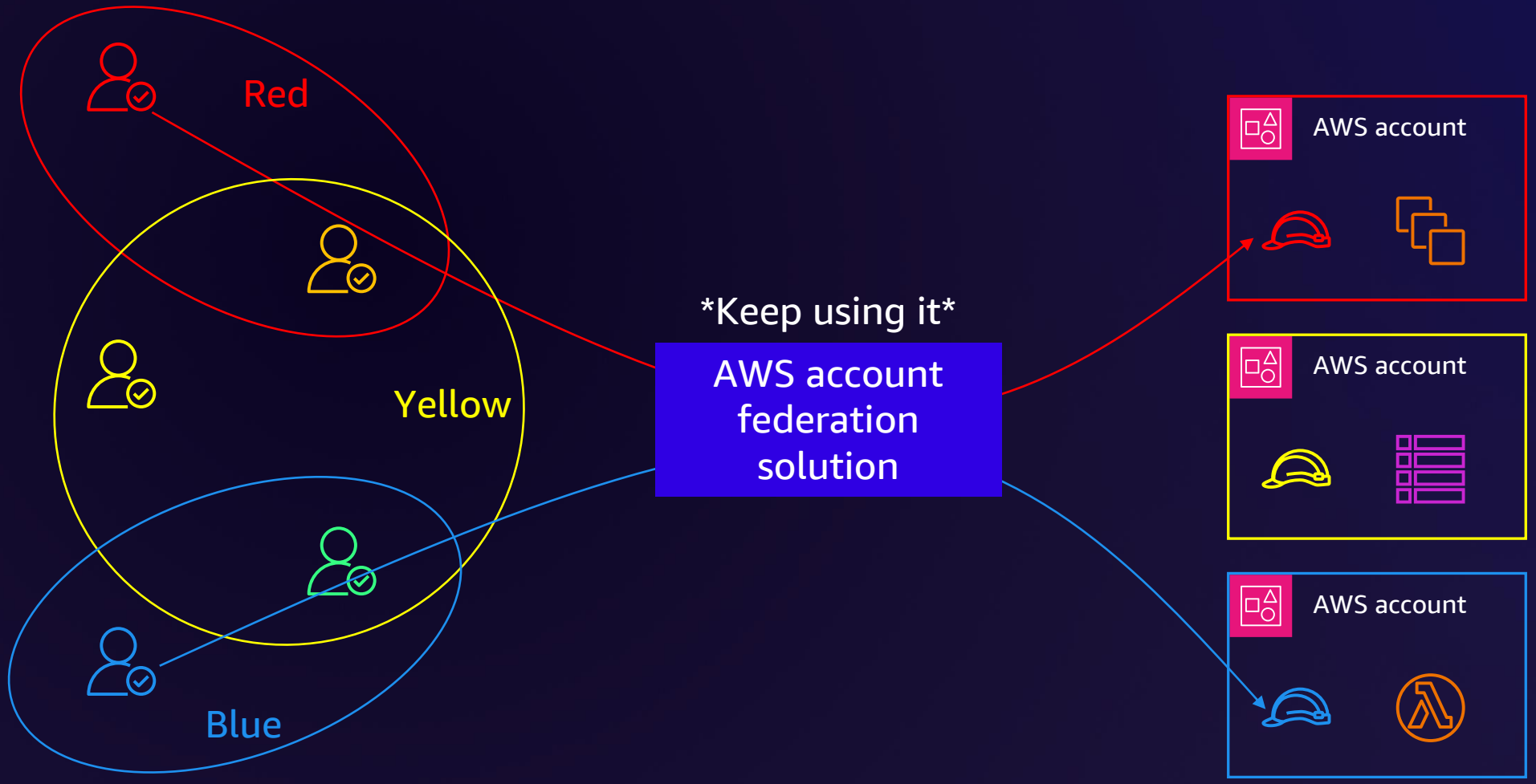
Myth: IAM Identity Center replaces my AWS account federation solution

"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

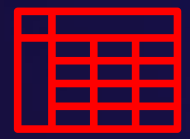
IF YOUR SOLUTION IS WORKING FOR YOU, KEEP USING IT

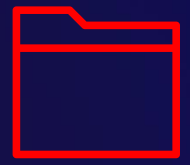


Common challenges with role-based access control in a data lake

Group-based access in a data lake

User in red group





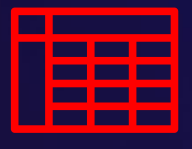
Red data

User in blue group

Blue data

Group-based access in a data lake

User in red group



Red data

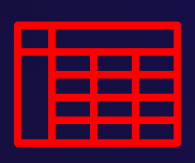
User in blue group

Blue data

Group-based access in a data lake

User in red group

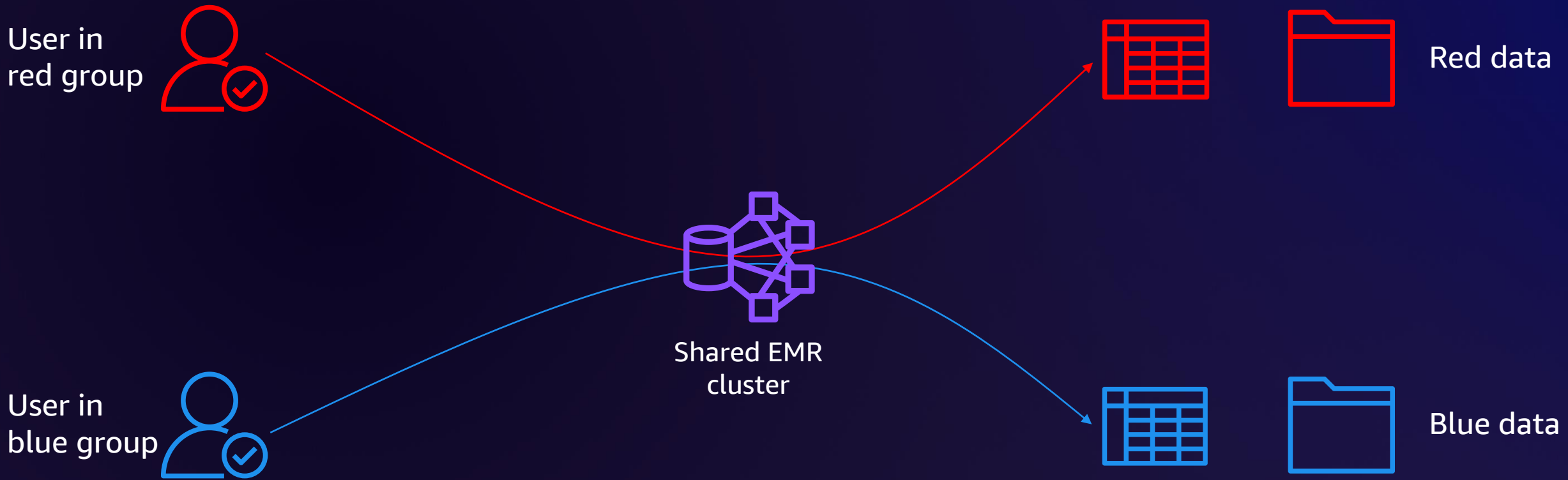
User in blue group



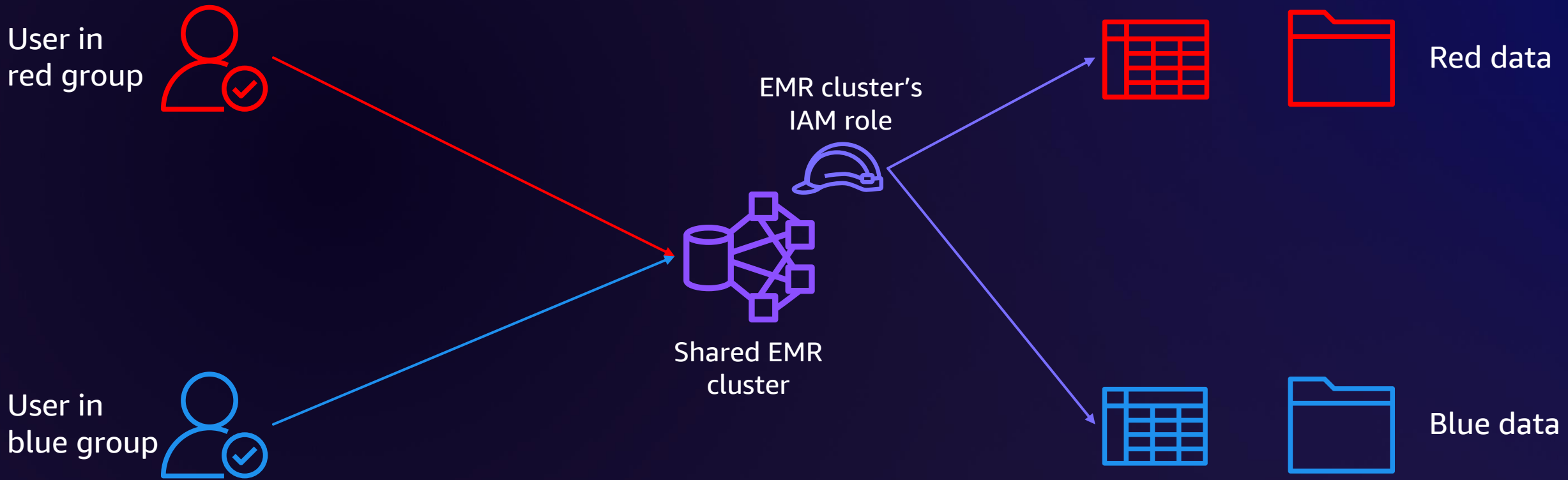
Red data

Blue data

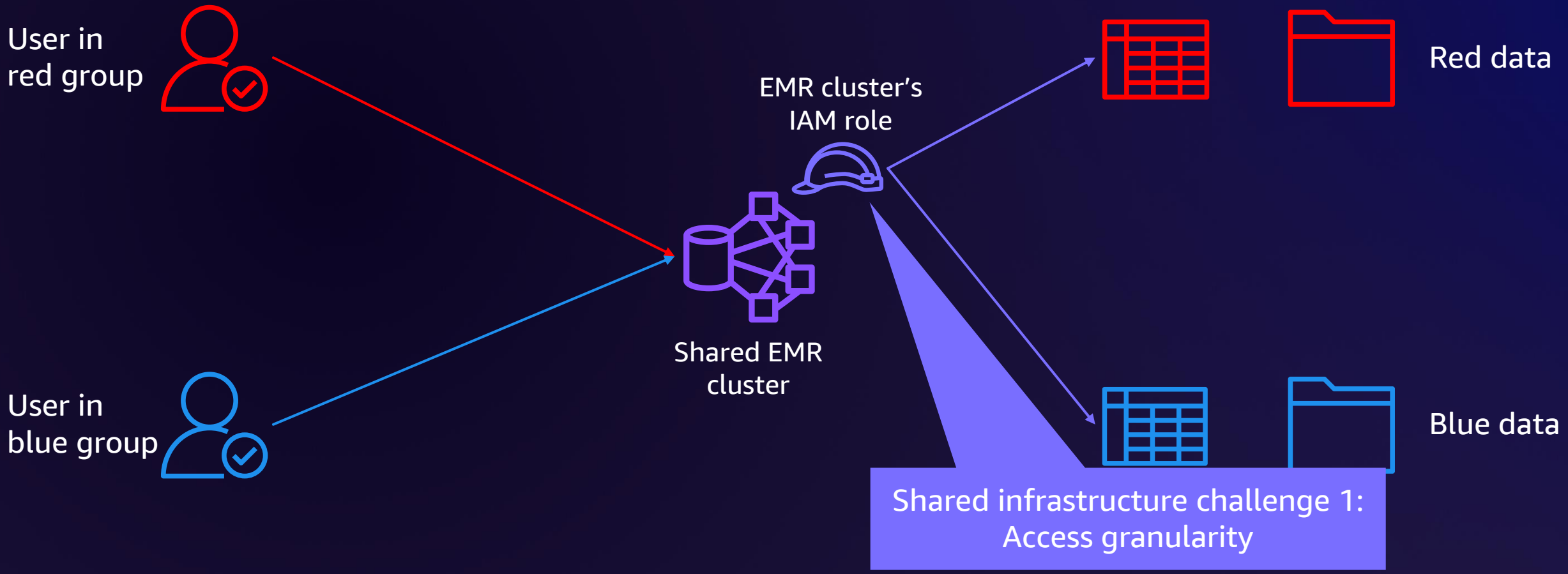
Group-based access to data via Amazon EMR: Without identity support



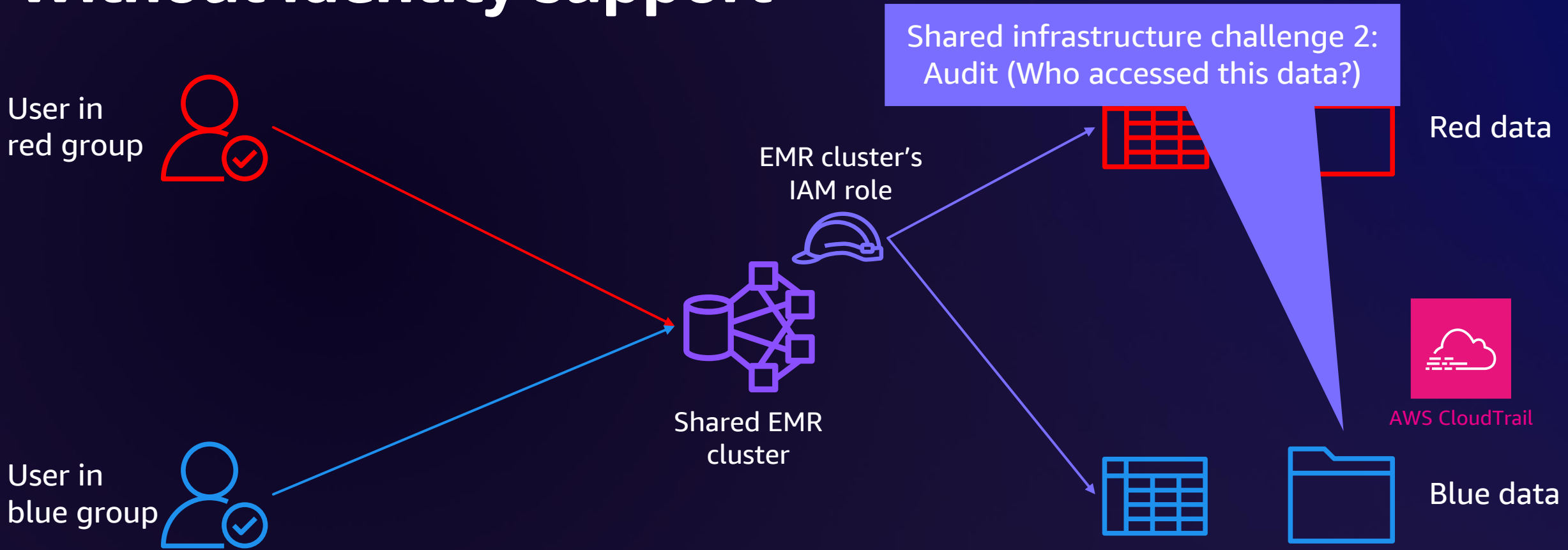
Group-based access to data via Amazon EMR: Without identity support



Group-based access to data via Amazon EMR: Without identity support



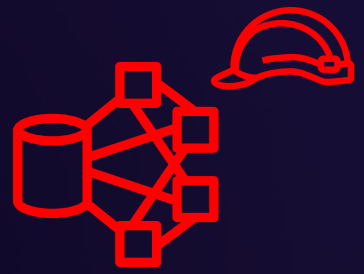
Group-based access to data via Amazon EMR: Without identity support



Approach 1: Separate the compute

Approach 1: Separate the compute

User in red group



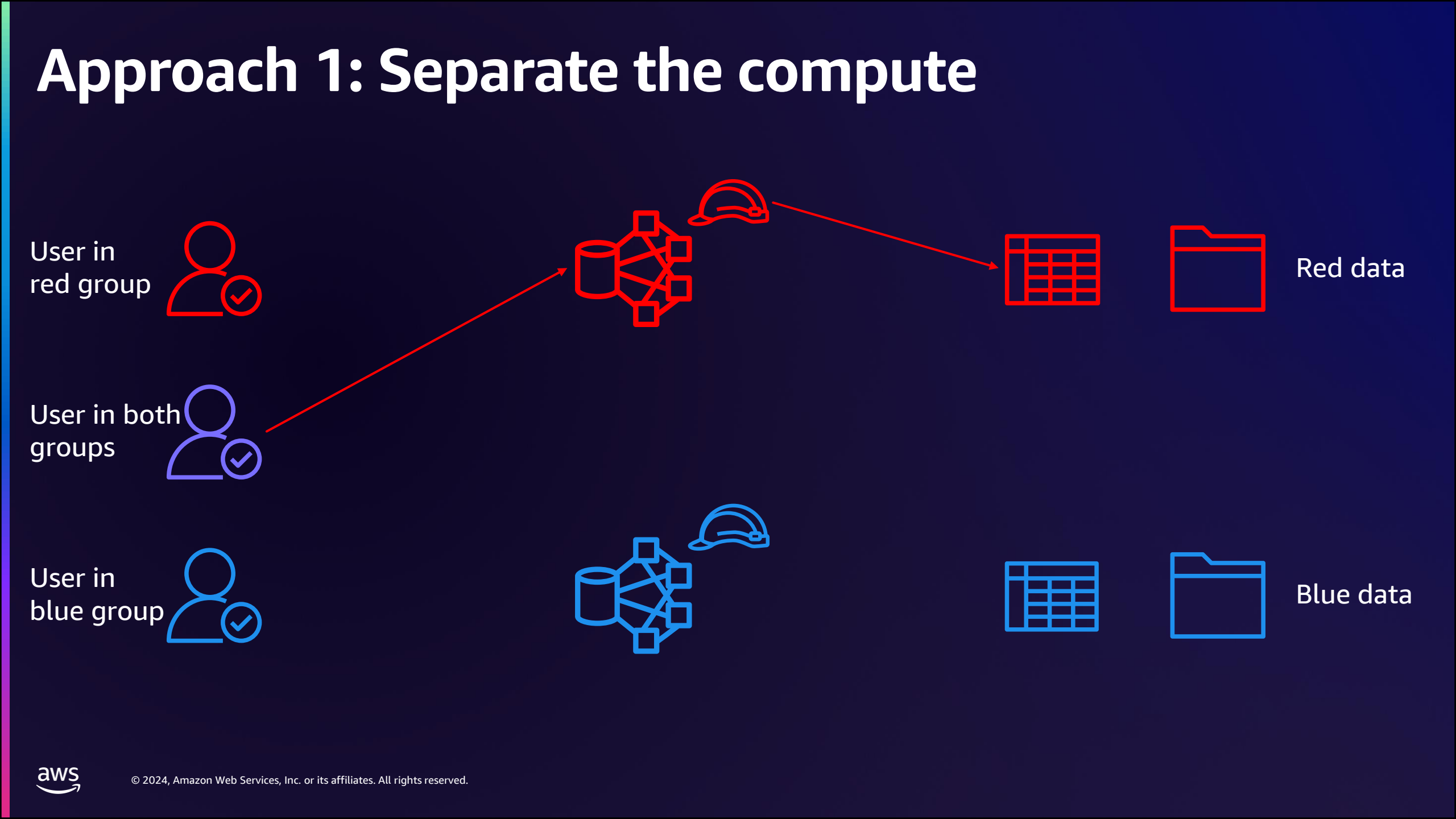
Red data

User in both groups

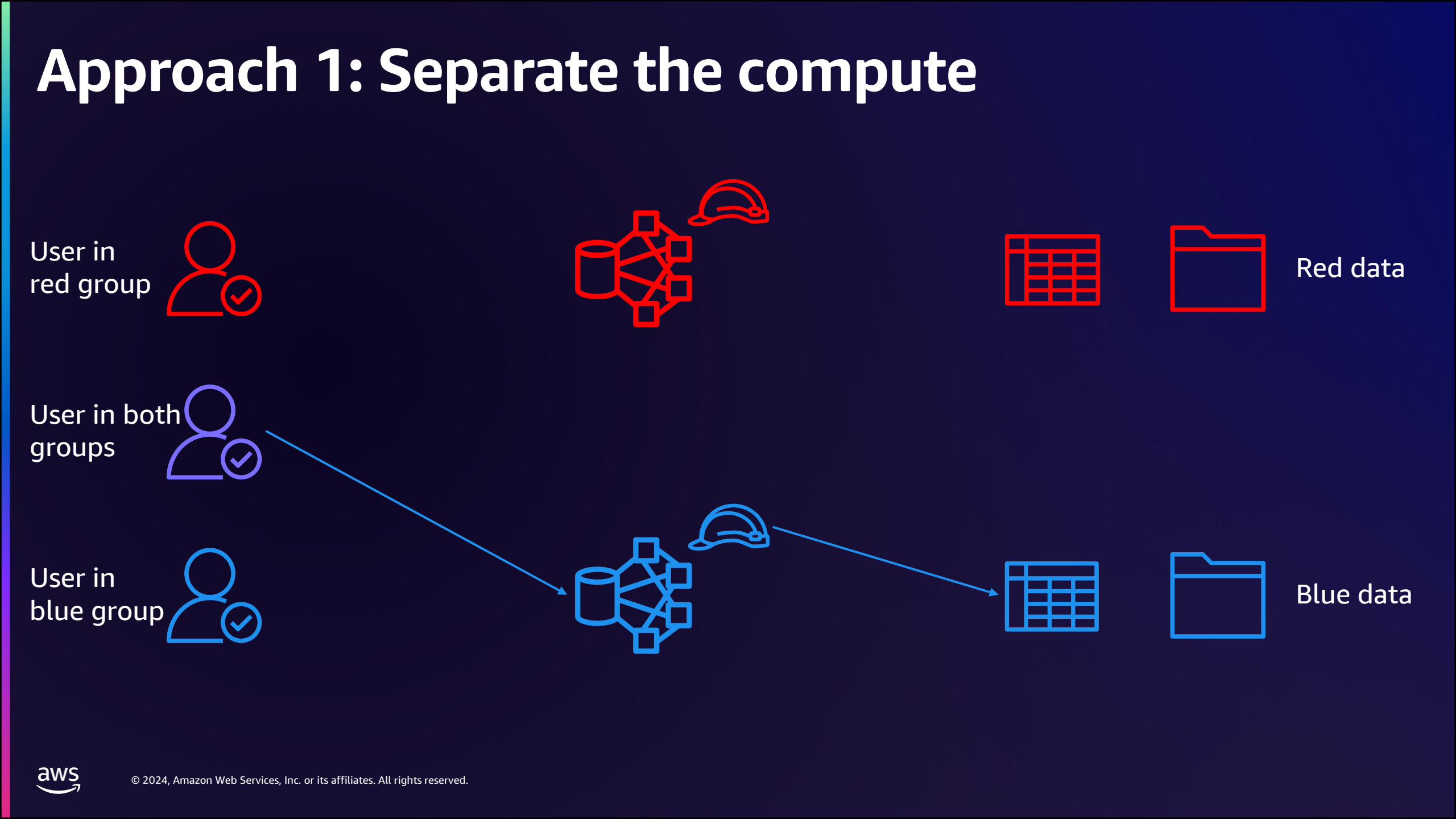
User in blue group

Blue data

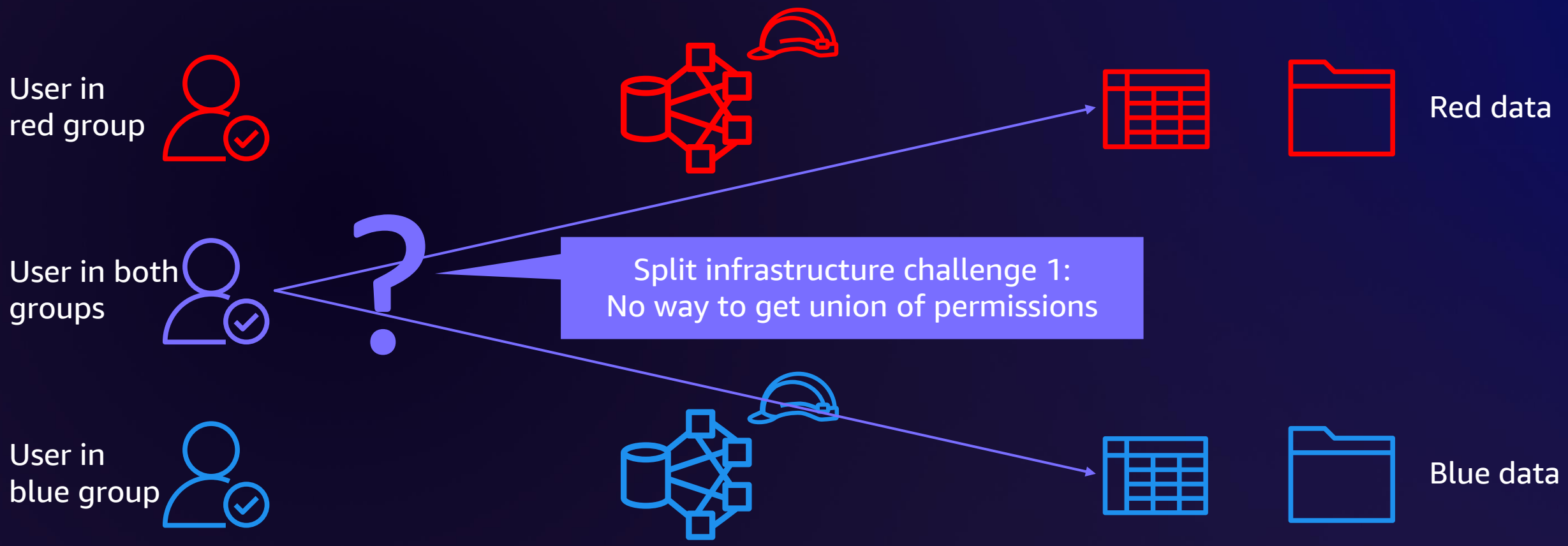
Approach 1: Separate the compute



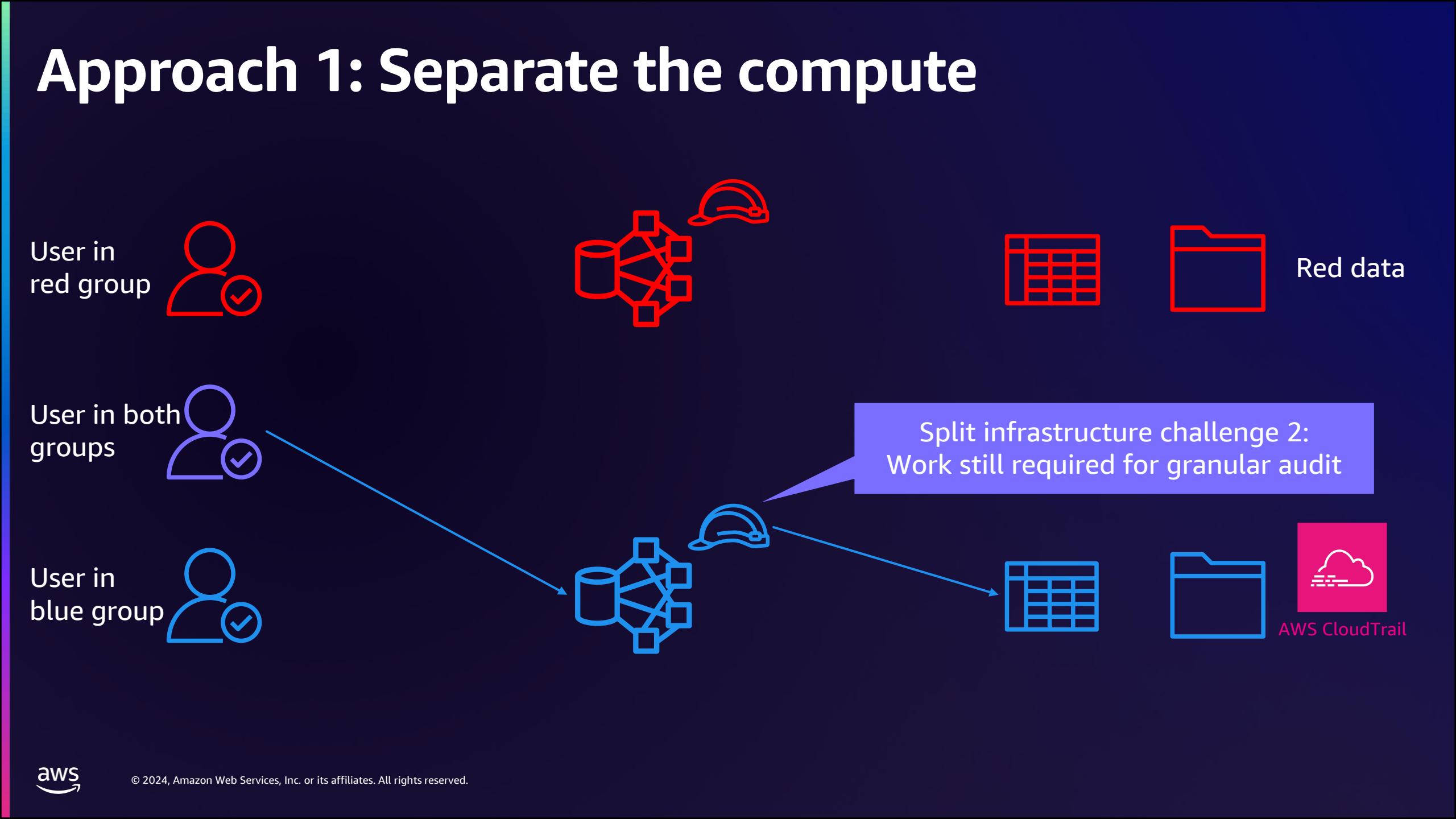
Approach 1: Separate the compute



Approach 1: Separate the compute

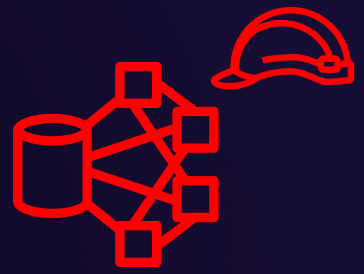


Approach 1: Separate the compute



Approach 1: Separate the compute

User in red group



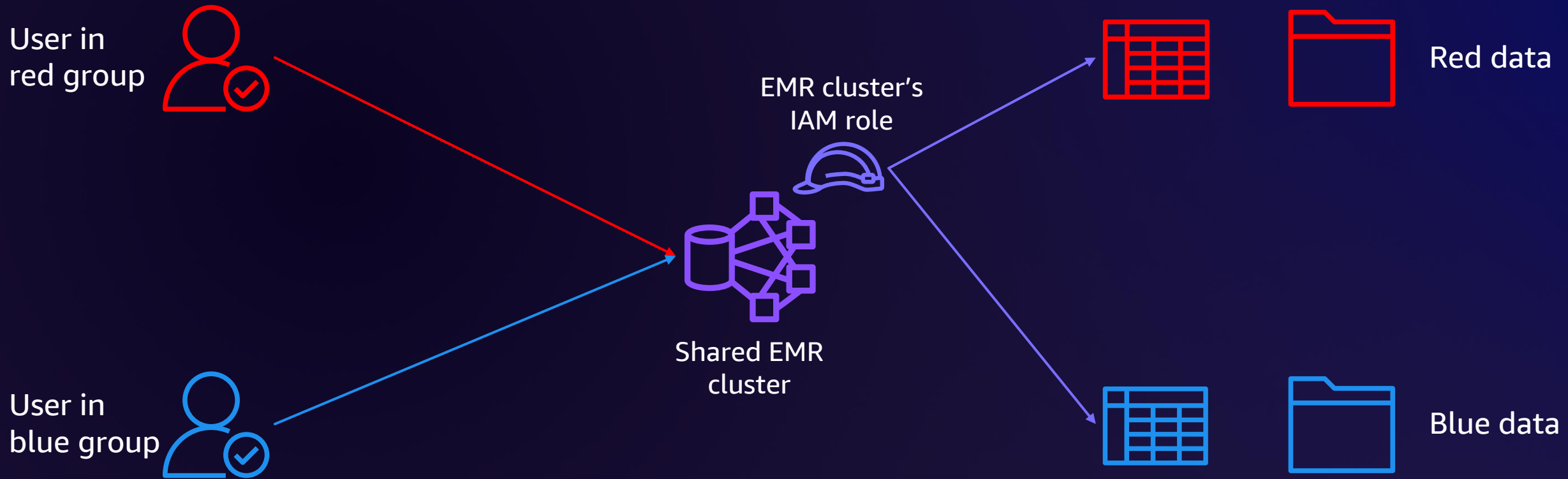
Red data

Split infrastructure challenge 3: Cost

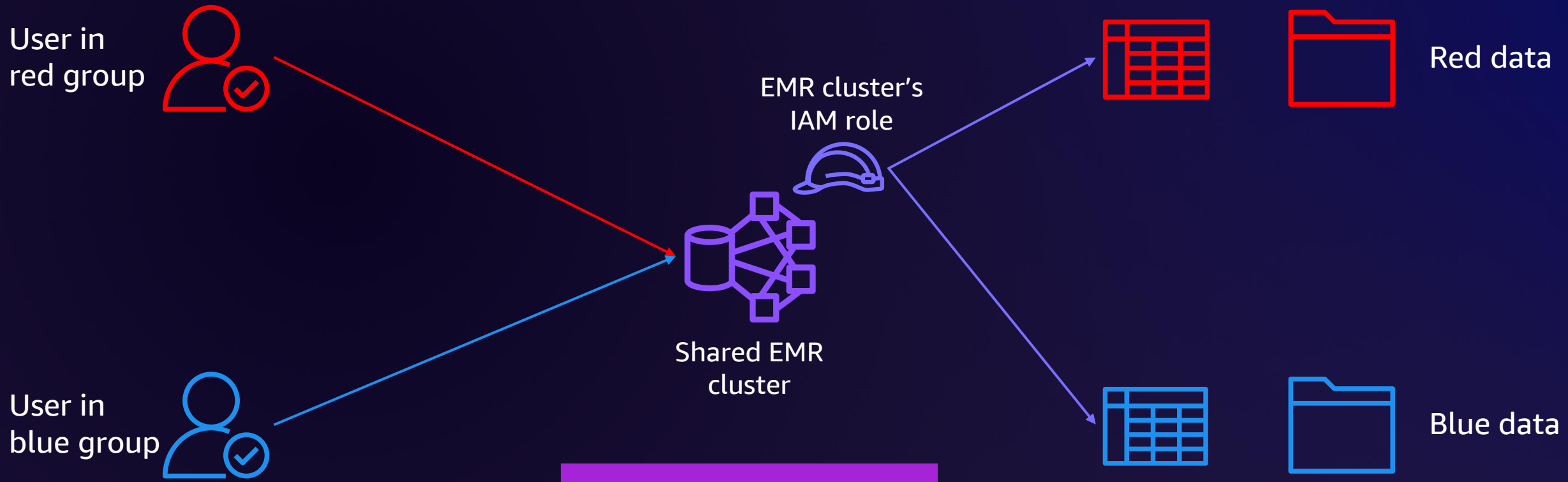
User in blue group

Blue data

Approach 2: Gate access locally

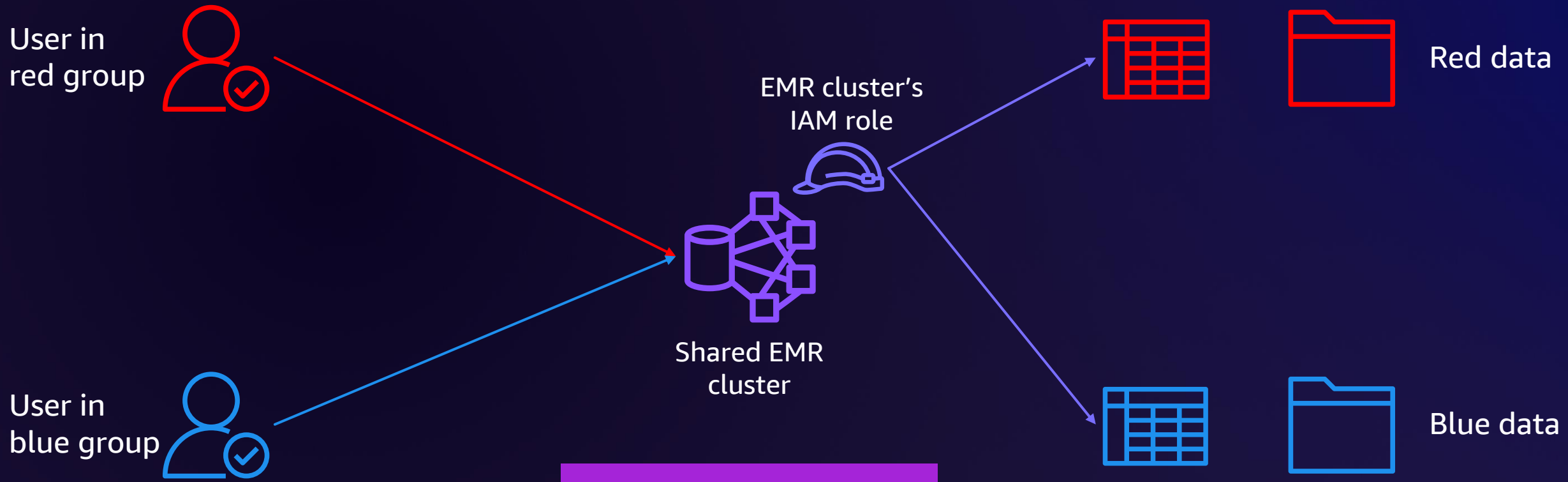


Approach 2: Gate access locally



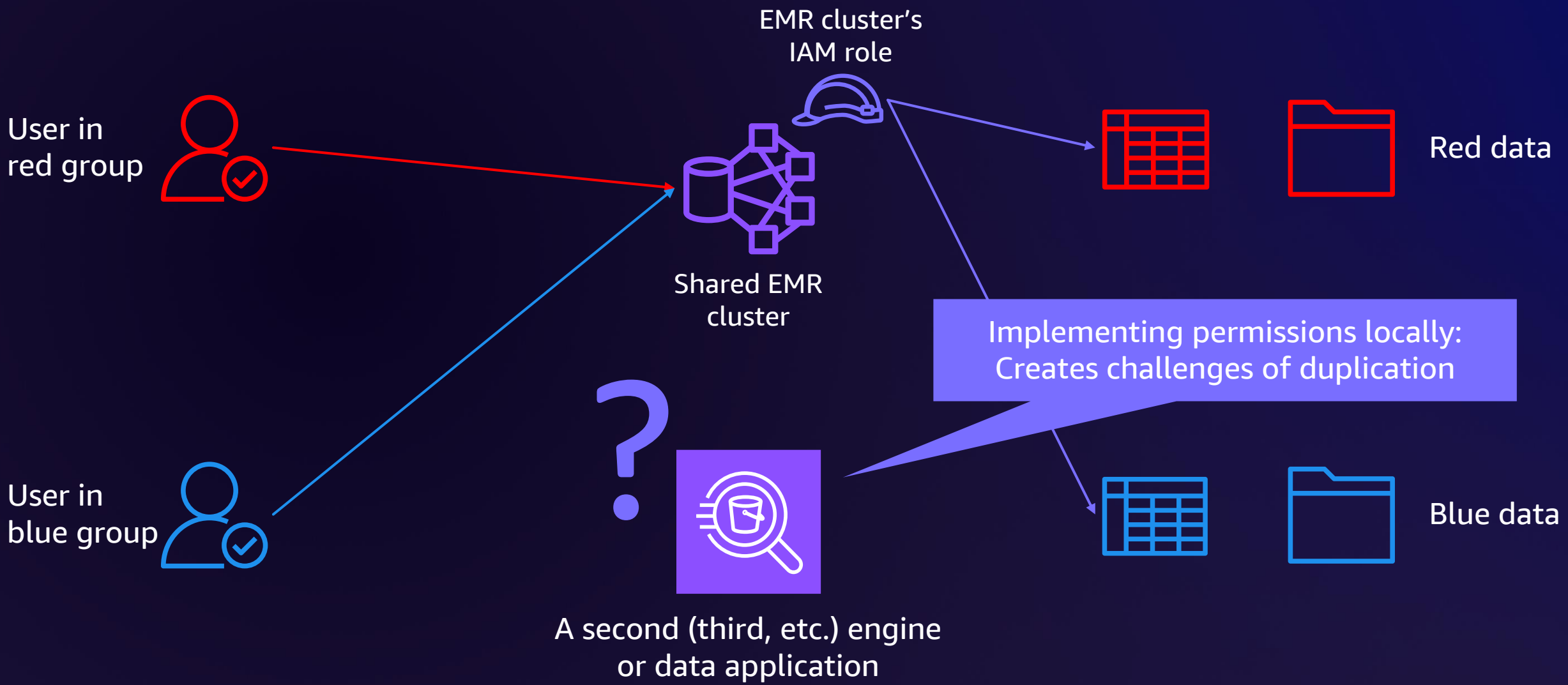
Problem solved . . .

Approach 2: Gate access locally

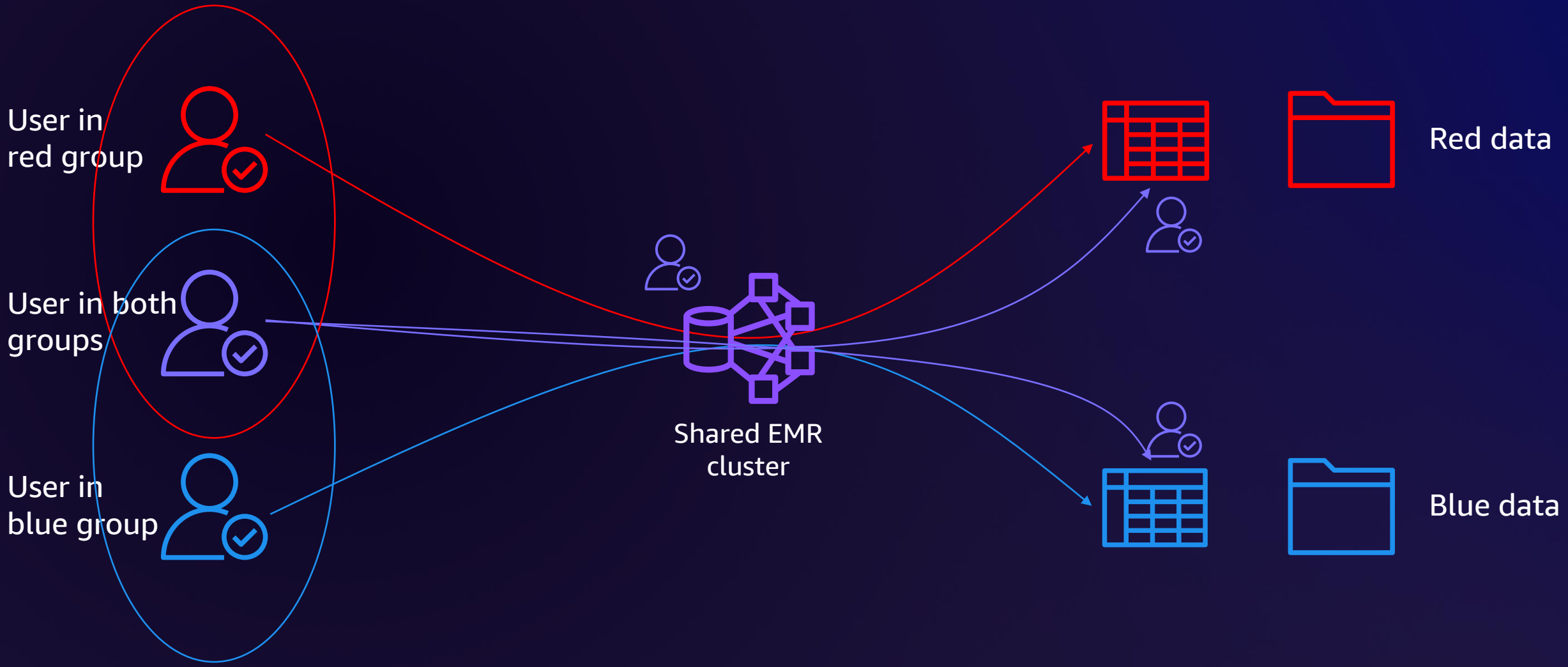


Problem solved . . . Right?

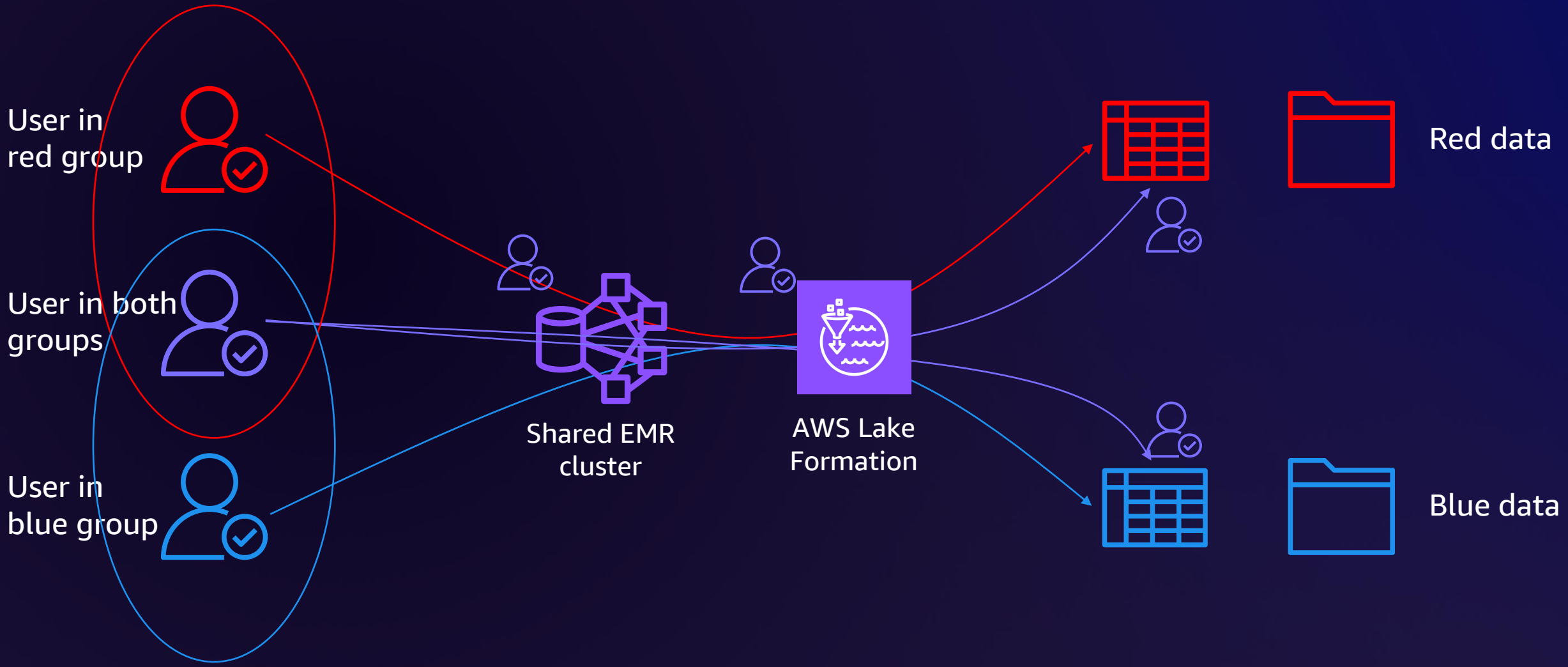
Approach 2: Gate access locally



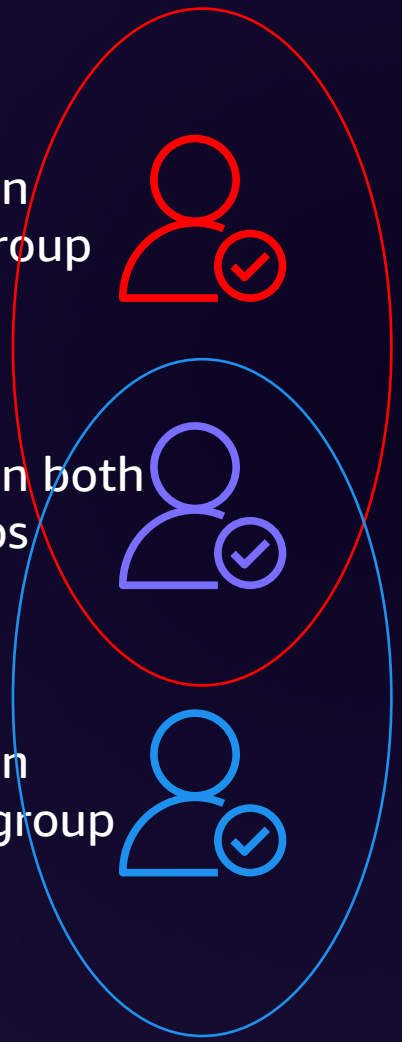
Amazon EMR with trusted identity propagation

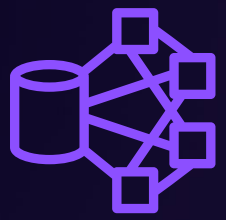


Amazon EMR with trusted identity propagation



Amazon EMR with trusted identity propagation





Shared EMR cluster

AWS Lake Formation

Red data

Blue data

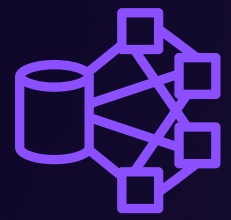
Amazon EMR with trusted identity propagation

Grantee	Permission	Resource
Red (Directory Group)	SELECT	Red Table
Blue (Directory Group)	SELECT	Blue Table

User in red group

User in both groups

User in blue group



Shared EMR cluster

AWS Lake Formation

Red data

Blue data

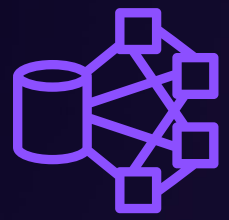
Amazon EMR with trusted identity propagation

Grantee	Permission	Resource
Red (Directory Group)	SELECT	Red Table
Blue (Directory Group)	SELECT	Blue Table

User in red group

User in both groups

User in blue group



Shared EMR cluster

AWS Lake Formation

Red data

Blue data

Number of grants scales with number of datasets; not combinatoric

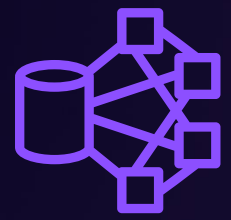
Amazon EMR with trusted identity propagation

Grantee	Permission	Resource
Red (Directory Group)	SELECT	Red Table
Blue (Directory Group)	SELECT	Blue Table

User in red group

User in both groups

User in blue group



Shared EMR cluster

AWS Lake Formation

Red data

Blue data

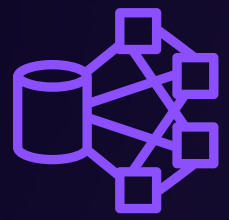
Amazon EMR with trusted identity propagation

Grantee	Permission	Resource
Red (Directory Group)	SELECT	Red Table
Blue (Directory Group)	SELECT	Blue Table

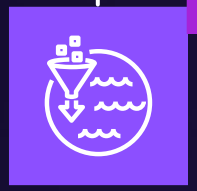
User in red group

User in both groups

User in blue group



Shared EMR cluster



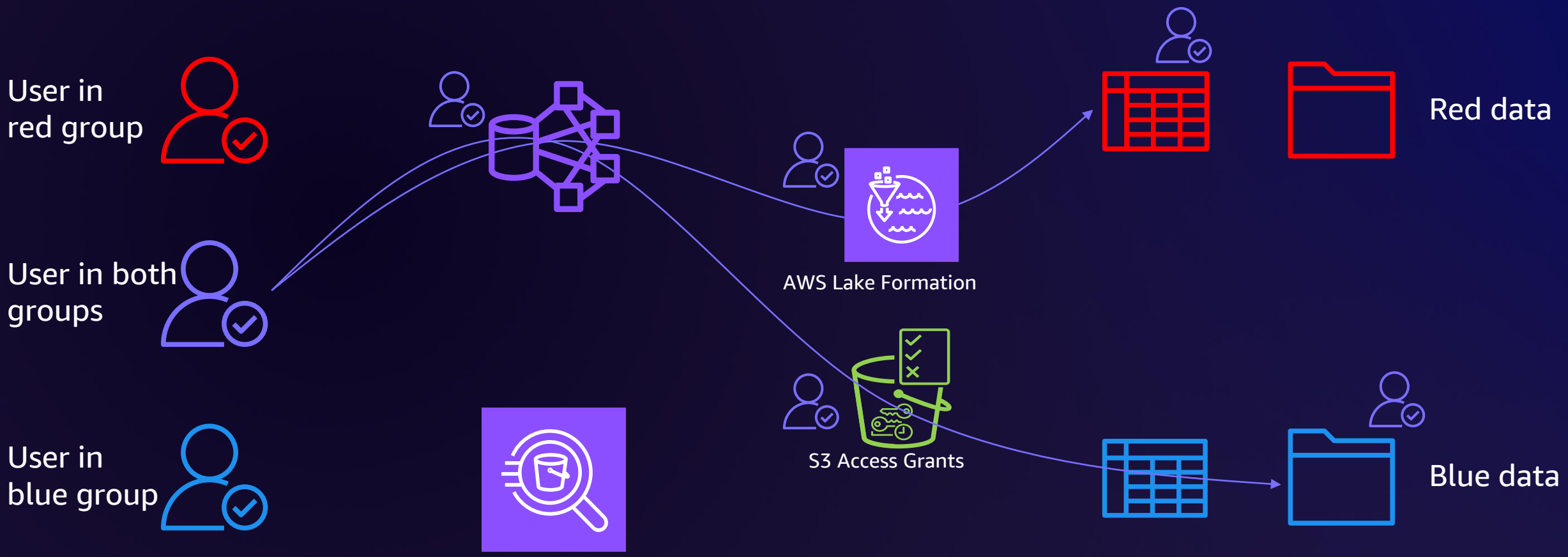
AWS Lake Formation

Permissions are at the data; **independent** of application used

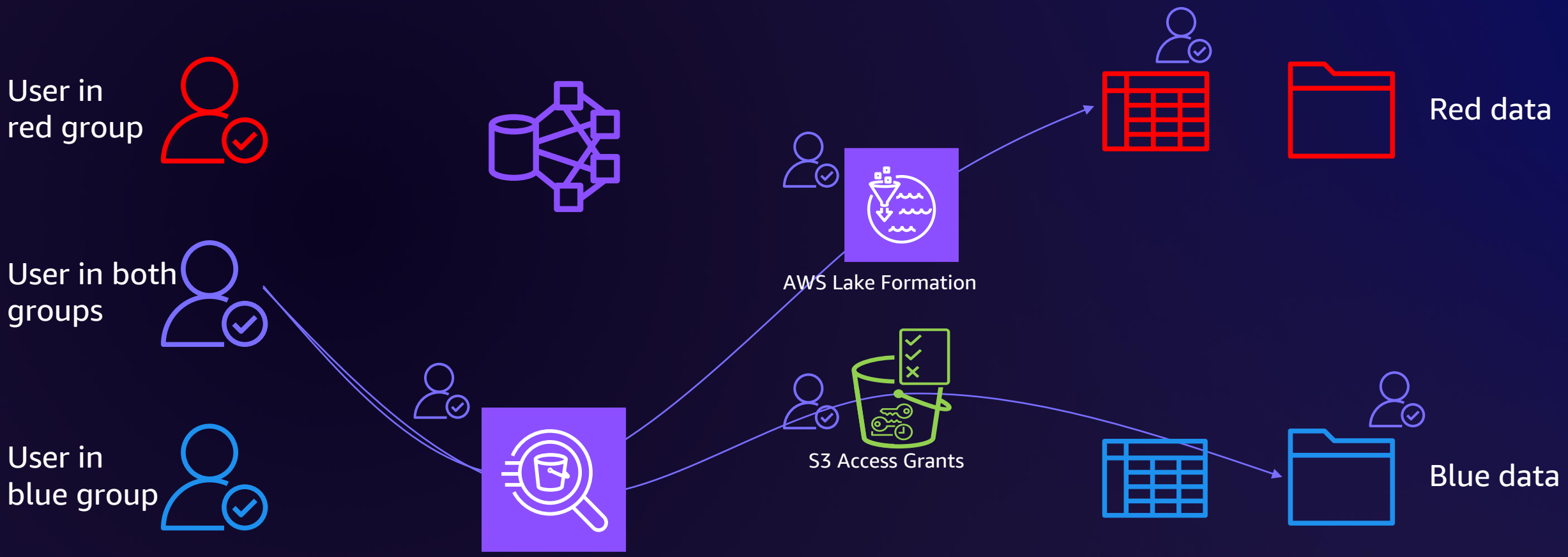
Red data

Blue data

Data lake with trusted identity propagation

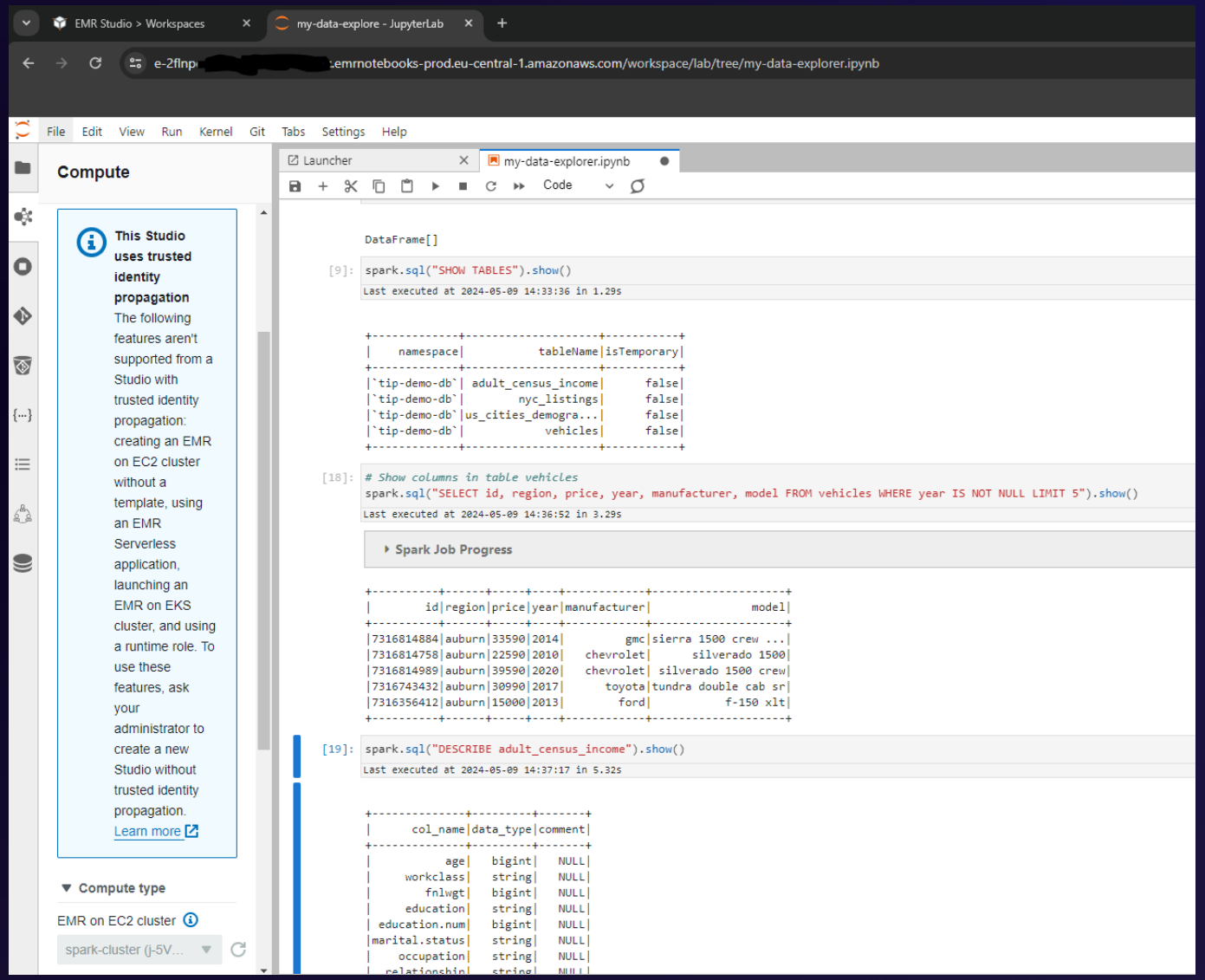


Data lake with trusted identity propagation

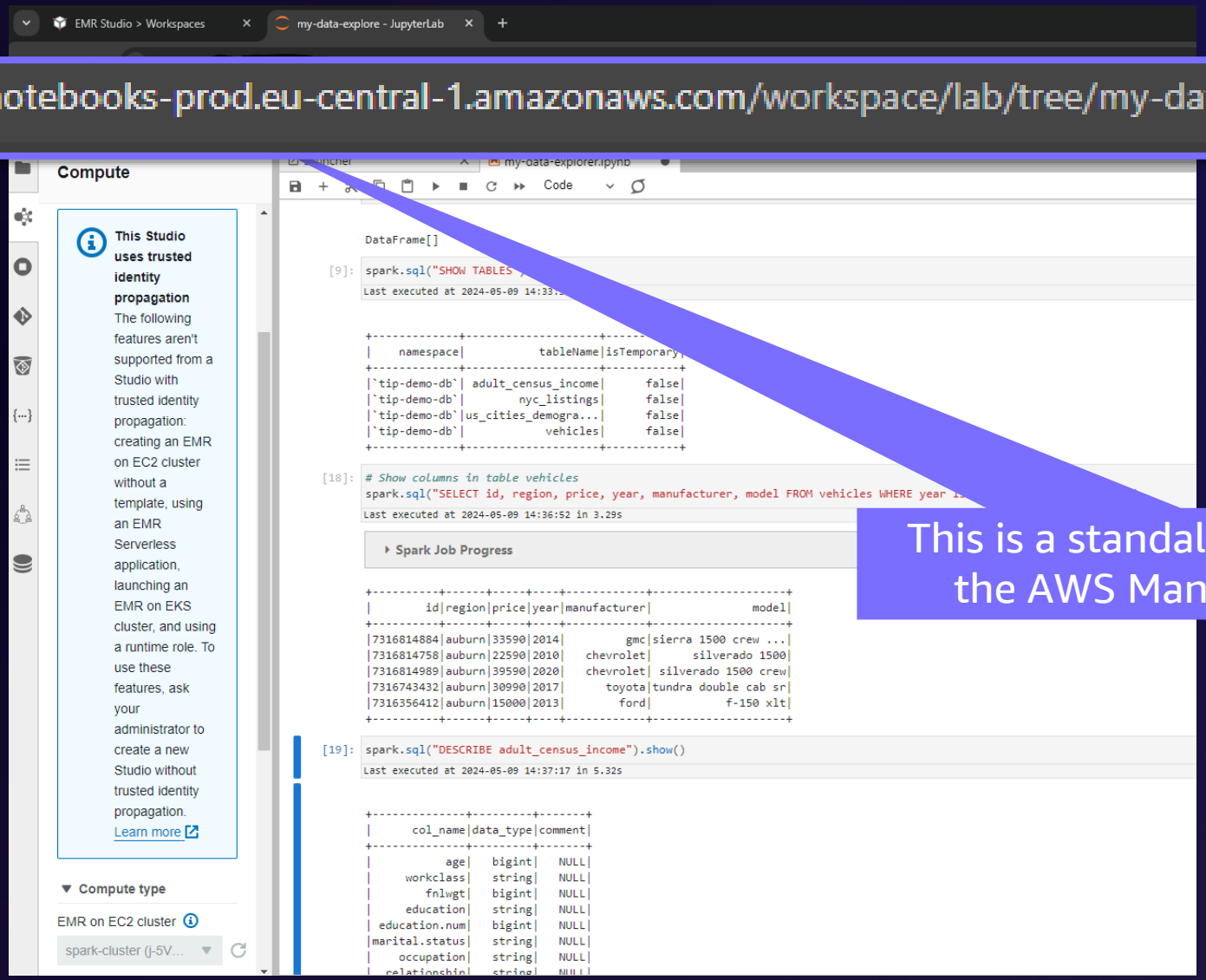


Deep dive: Trusted identity propagation in Amazon EMR

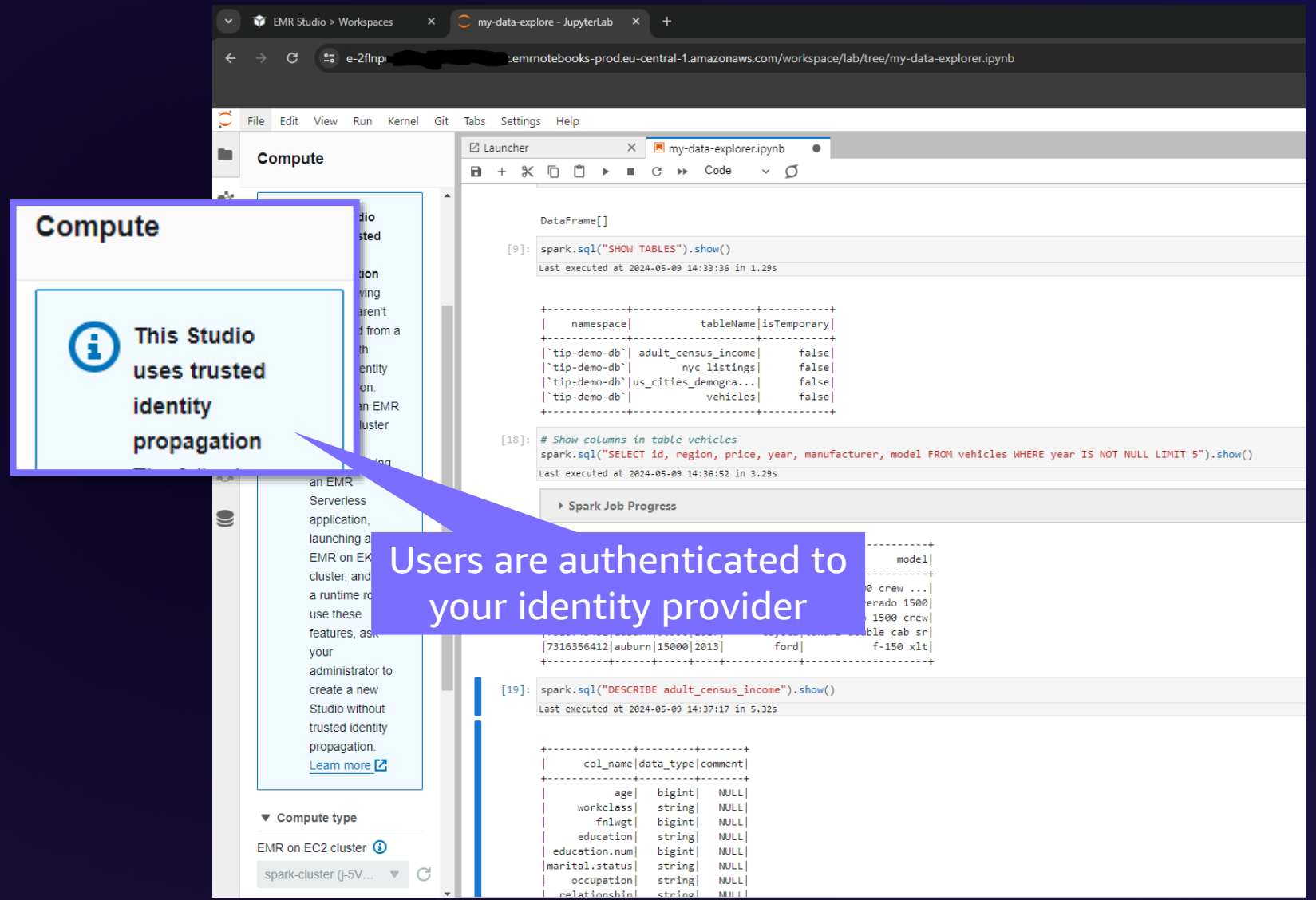
Quick tour – EMR Studio: Managed notebook environment



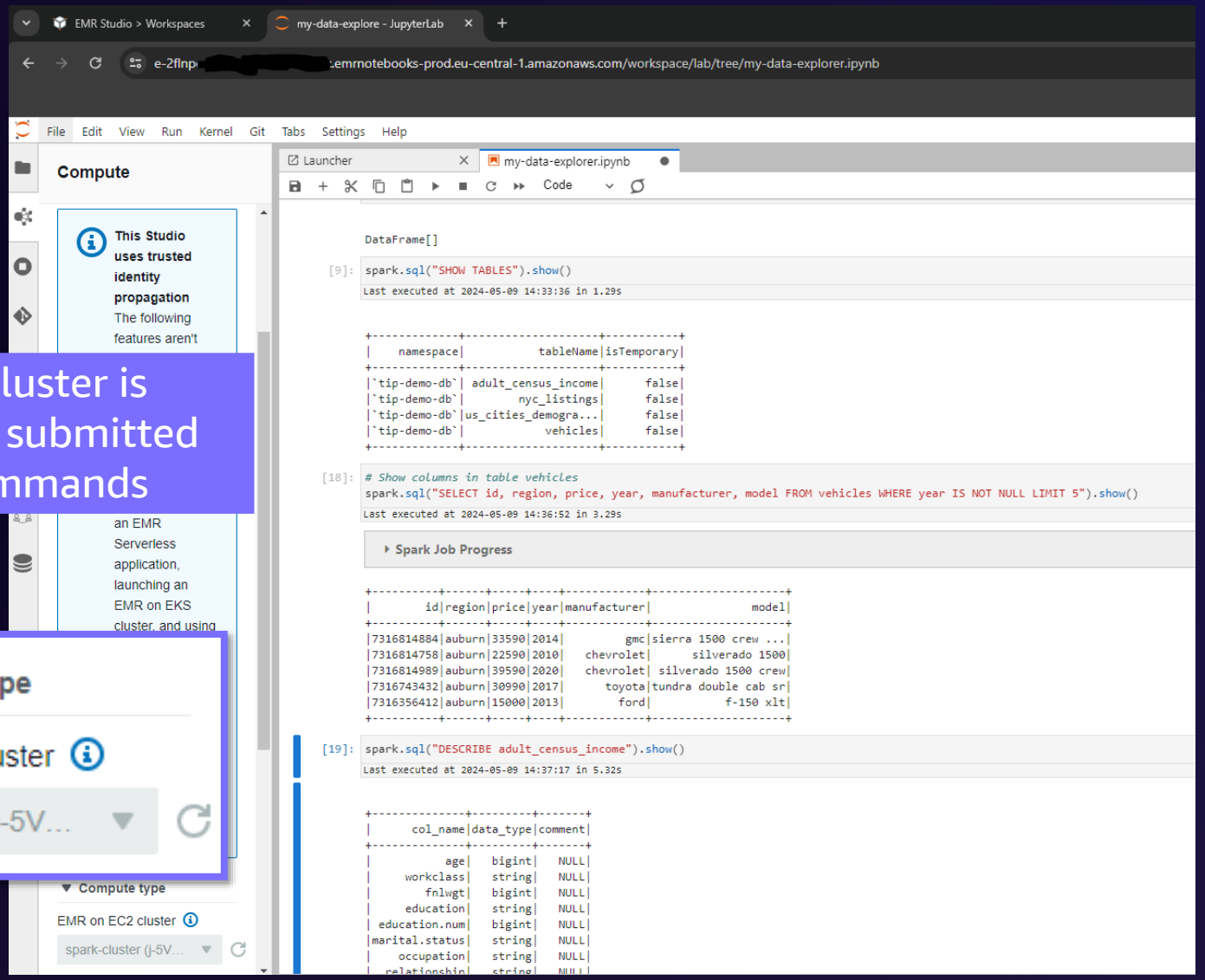
Quick tour – EMR Studio: Managed notebook environment



Quick tour – EMR Studio: Managed notebook environment



Quick tour – EMR Studio: Managed notebook environment



An EMR cluster is running the submitted Spark commands

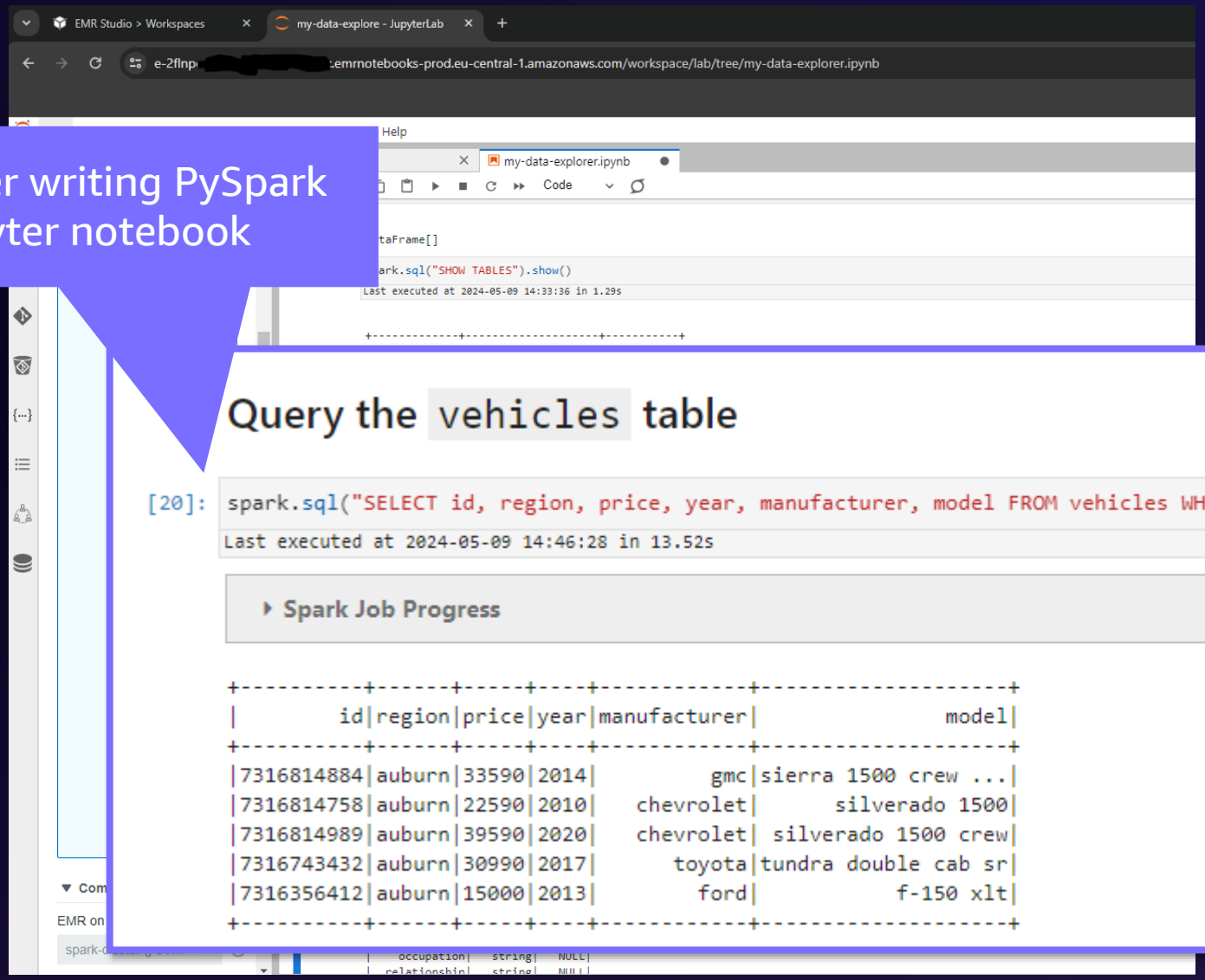
Compute type

EMR on EC2 cluster

spark-cluster (j-5V...

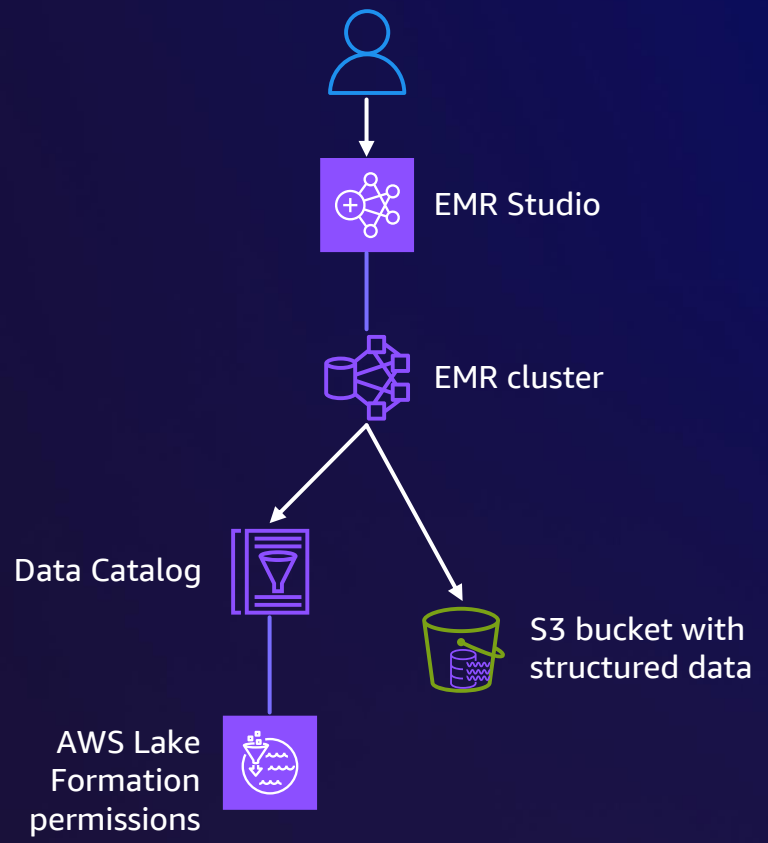
Quick tour – EMR Studio: Managed notebook environment

Authenticated user writing PySpark code in a Jupyter notebook



Rundown of AWS components

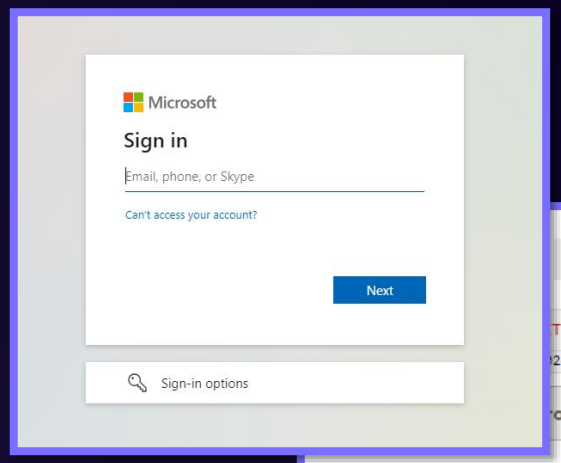
- Amazon EMR Studio
- Amazon EMR cluster
- AWS Glue Data Catalog
- AWS Lake Formation
- Amazon S3
- AWS Identity and Access Management (IAM)

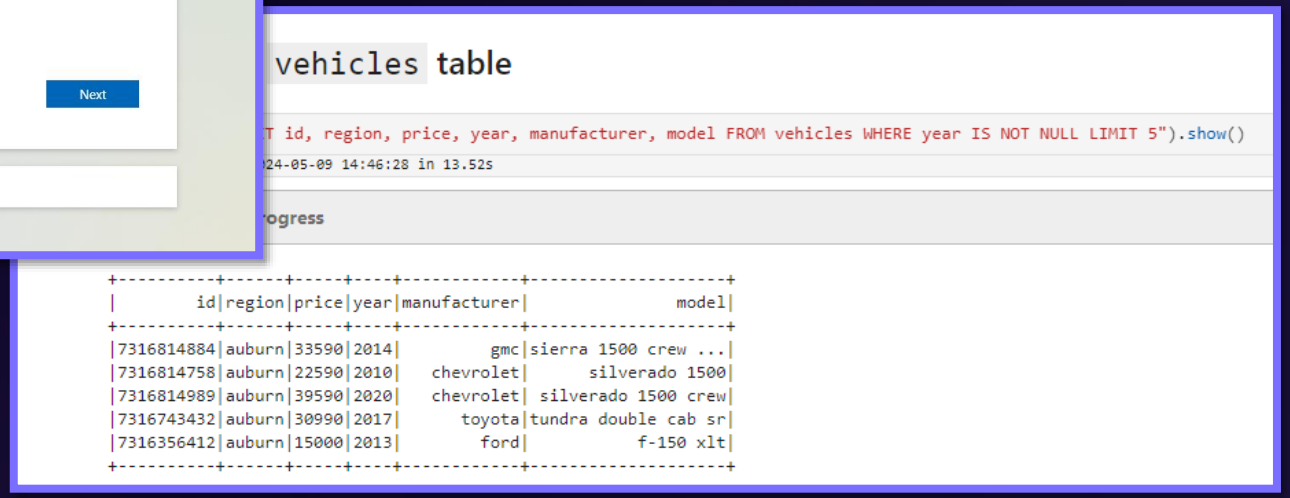


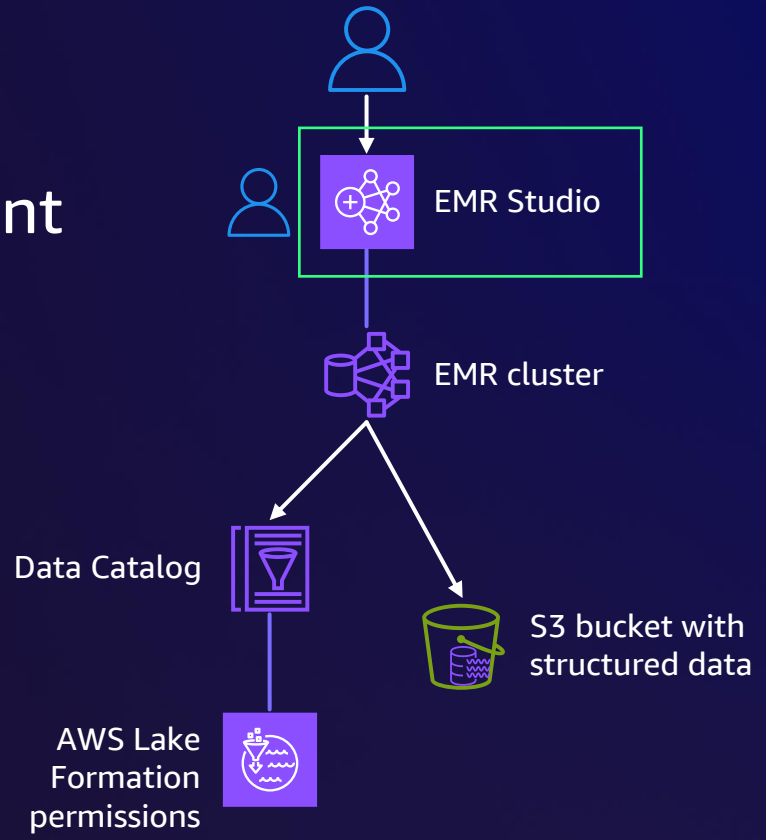
➔ User's identity propagates through all of these

Rundown of AWS components

- **Amazon EMR Studio**
  - UX application – Browser-based notebook environment
  - Authenticates the user at the entry point

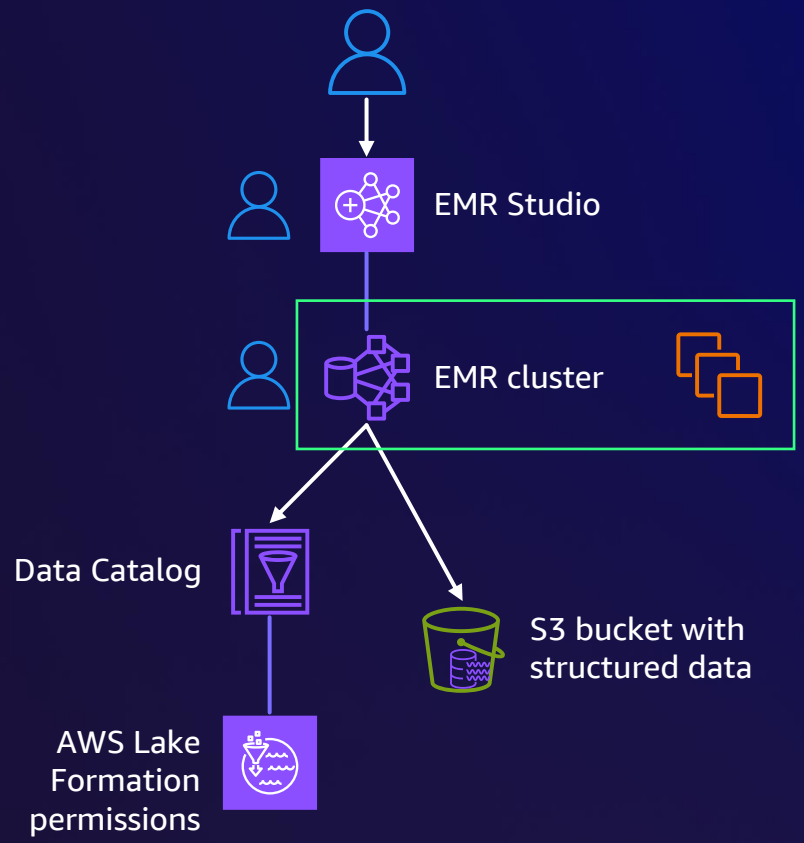






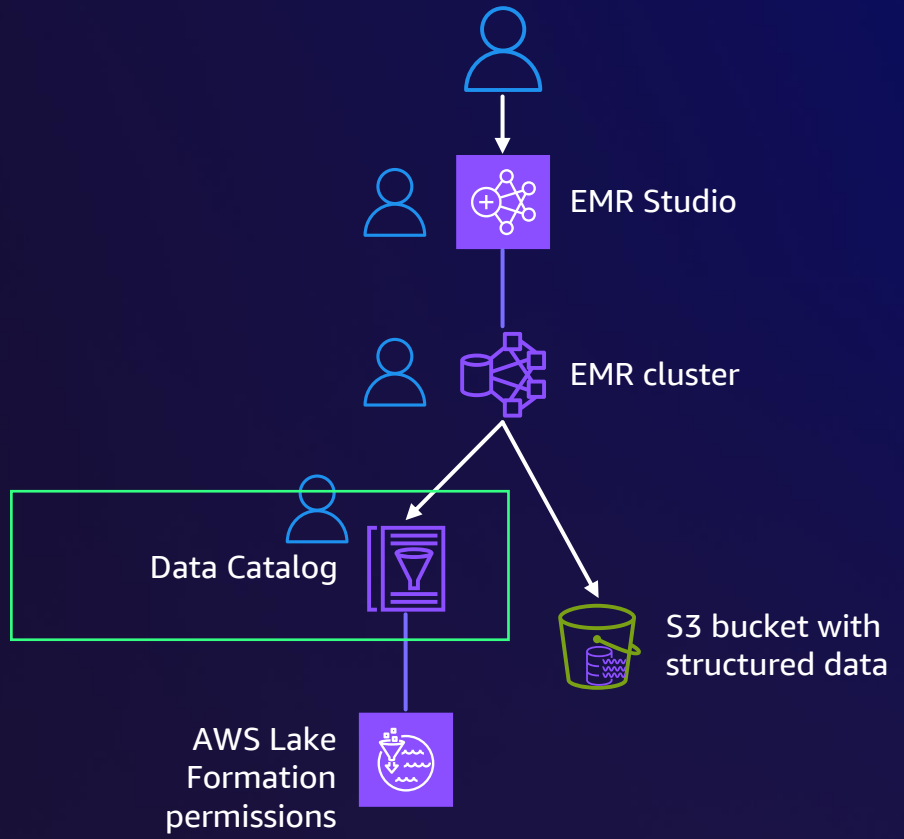
Rundown of AWS components

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



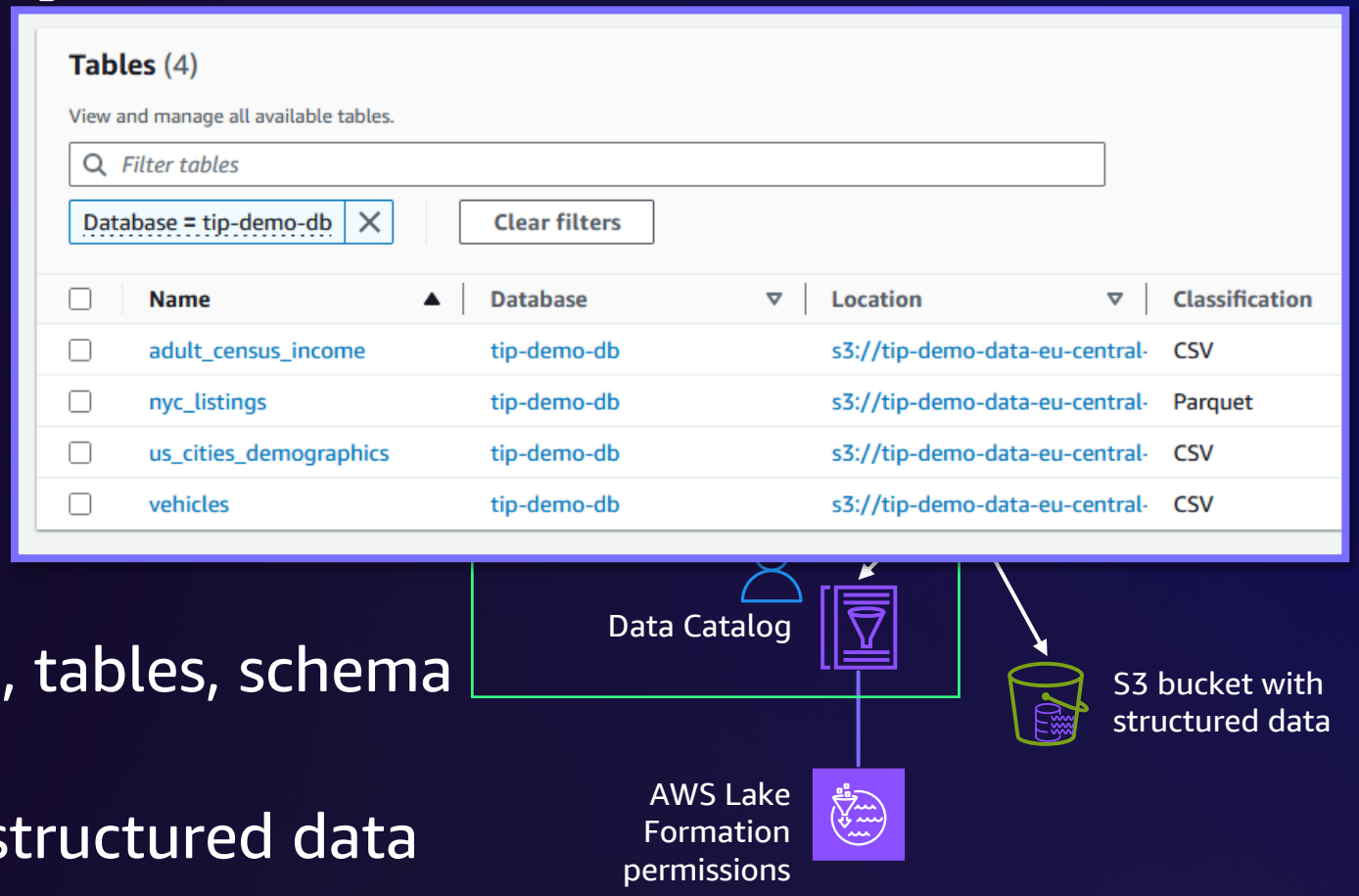
Rundown of AWS components

- **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)



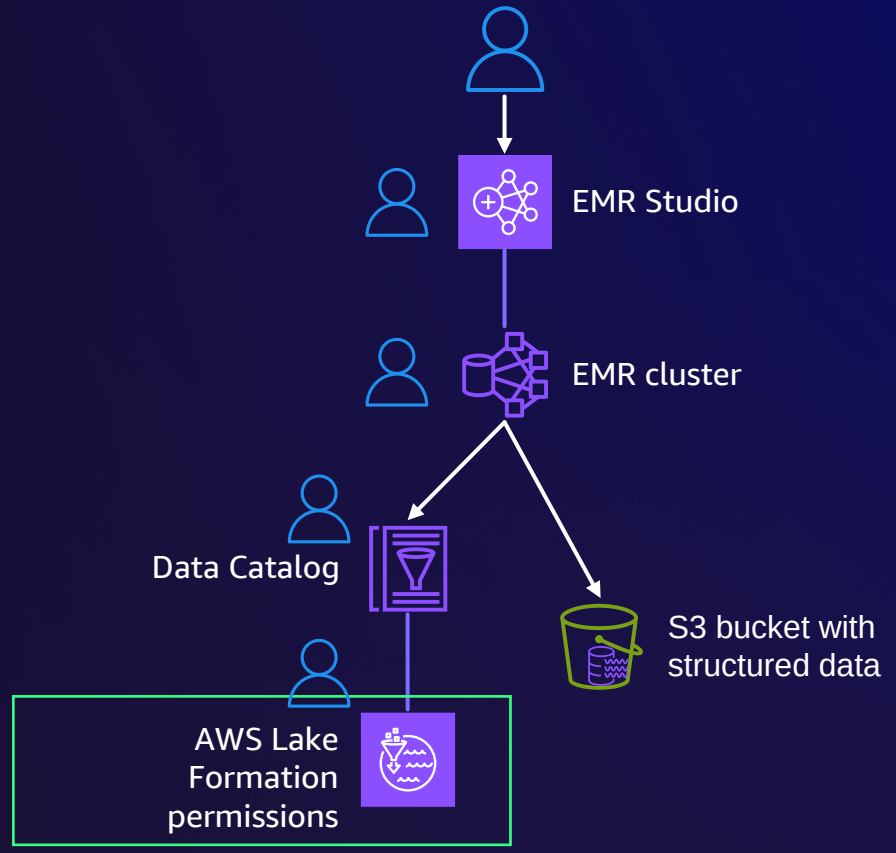
Rundown of AWS components

- **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)



Rundown of AWS components

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



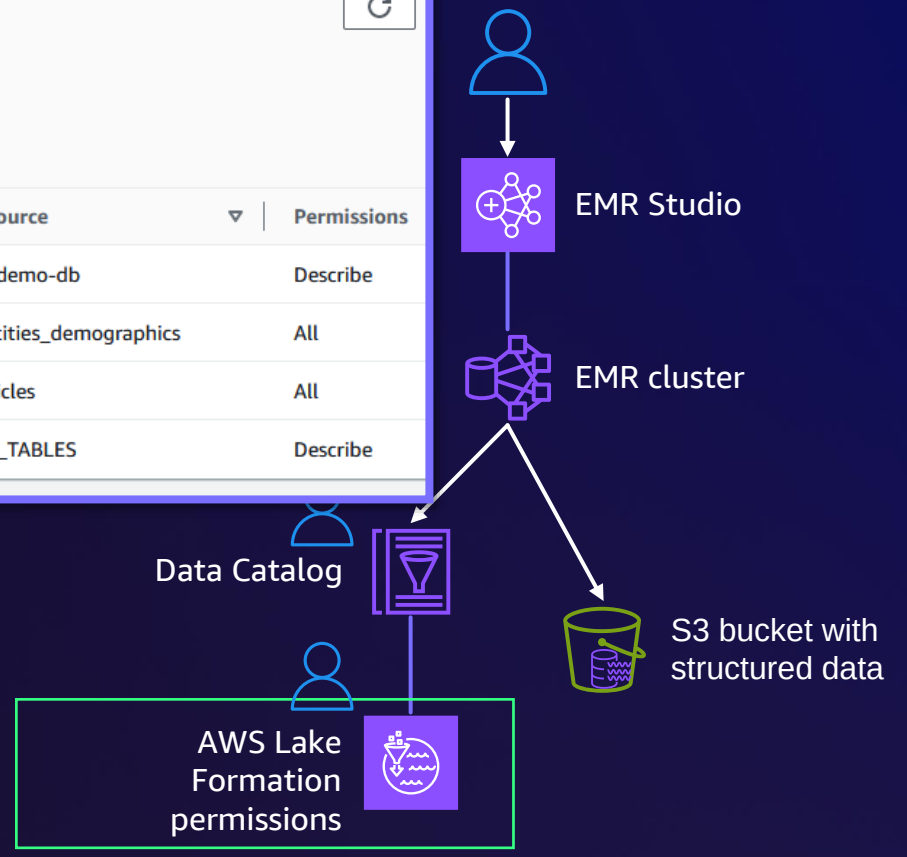
Rundown of AWS components

Data lake permissions (4)

Filter permissions by property or value — 4 matches

Database = tip-demo-db | and | Principal type = IAM Identity Center group | Clear filters

Principal	Princip...	Resour...	Database	Table	Resource	Permissions
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Database	tip-demo-db	-	tip-demo-db	Describe
0374d892-f081-7060-eebc-a9306f...	IAM Identi...	Table	tip-demo-db	us_cities_demographics	us_cities_demographics	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	vehicles	vehicles	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	ALL_TABLES	ALL_TABLES	Describe



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

Rundown of AWS components

Data lake permissions (4)

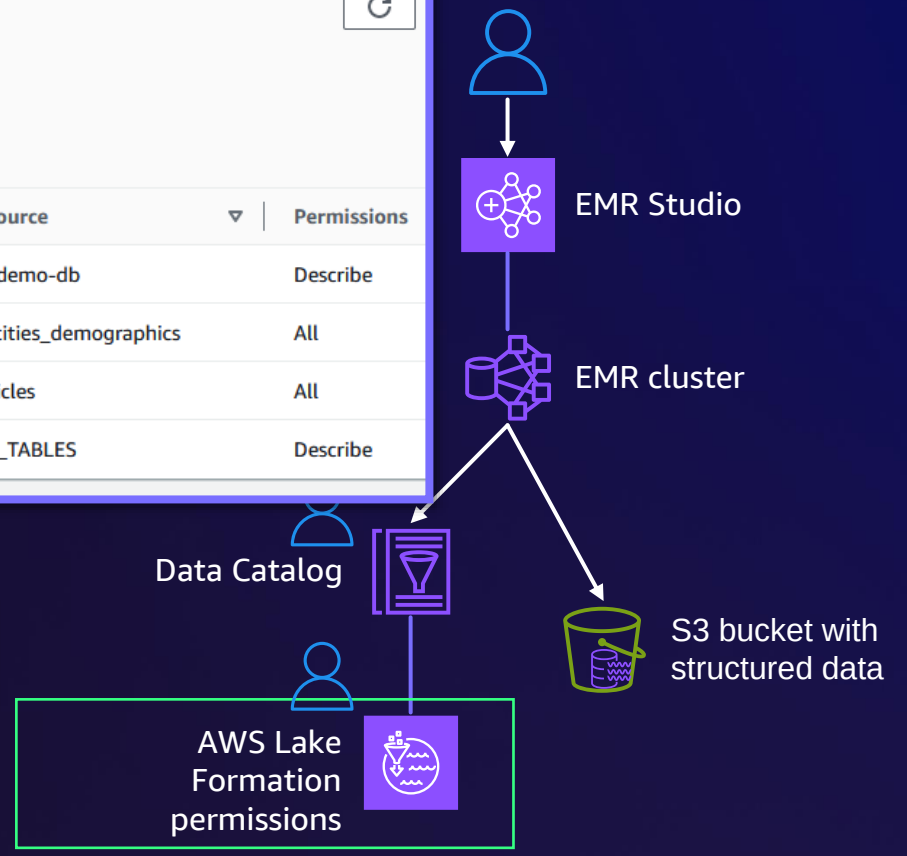
Filter permissio

Database = tip-der

Clear filters

Identifiers of directory users and groups from IAM Identity Center

Principal	Princip...	Resour...	Database	Table	Resource	Permissions
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Database	tip-demo-db	-	tip-demo-db	Describe
0374d892-f081-7060-eebc-a9306f...	IAM Identi...	Table	tip-demo-db	us_cities_demographics	us_cities_demographics	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	vehicles	vehicles	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	ALL_TABLES	ALL_TABLES	Describe



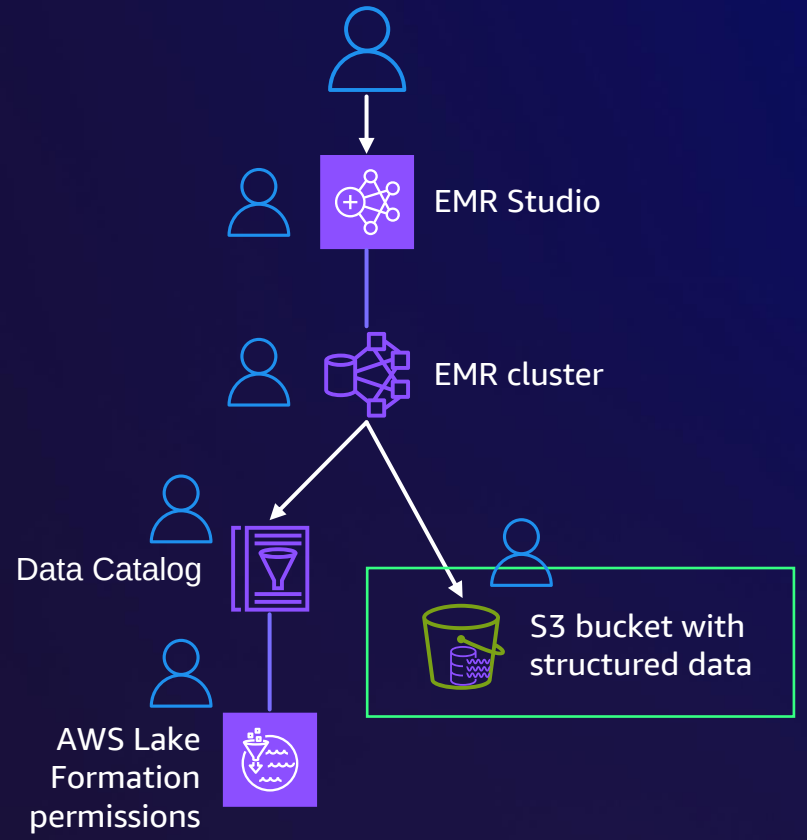
- **AWS Lake Formation**
  - Permissions service for the AWS Glue Data Catalog
  - Who (users/groups) has access to what (data)

Rundown of AWS components

AWS CloudTrail data event from Amazon S3

```
{
  "userIdentity": {
    "type": "AssumedRole",
    "arn": "arn:aws:sts::…:RoleForS3Access…"
      "onbehalfof": {
        "userid": "63548842-…"
      }
  },
  "eventSource": "s3.amazonaws.com",
  "eventName": "GetObject",
  …
}
```

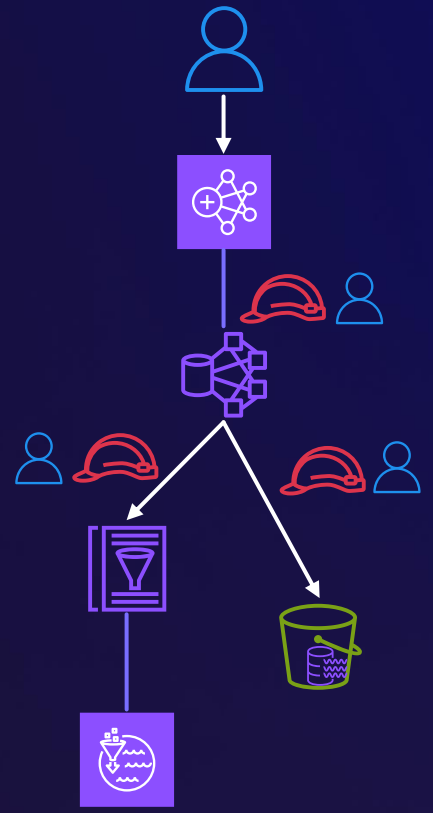
Access attributed to end user



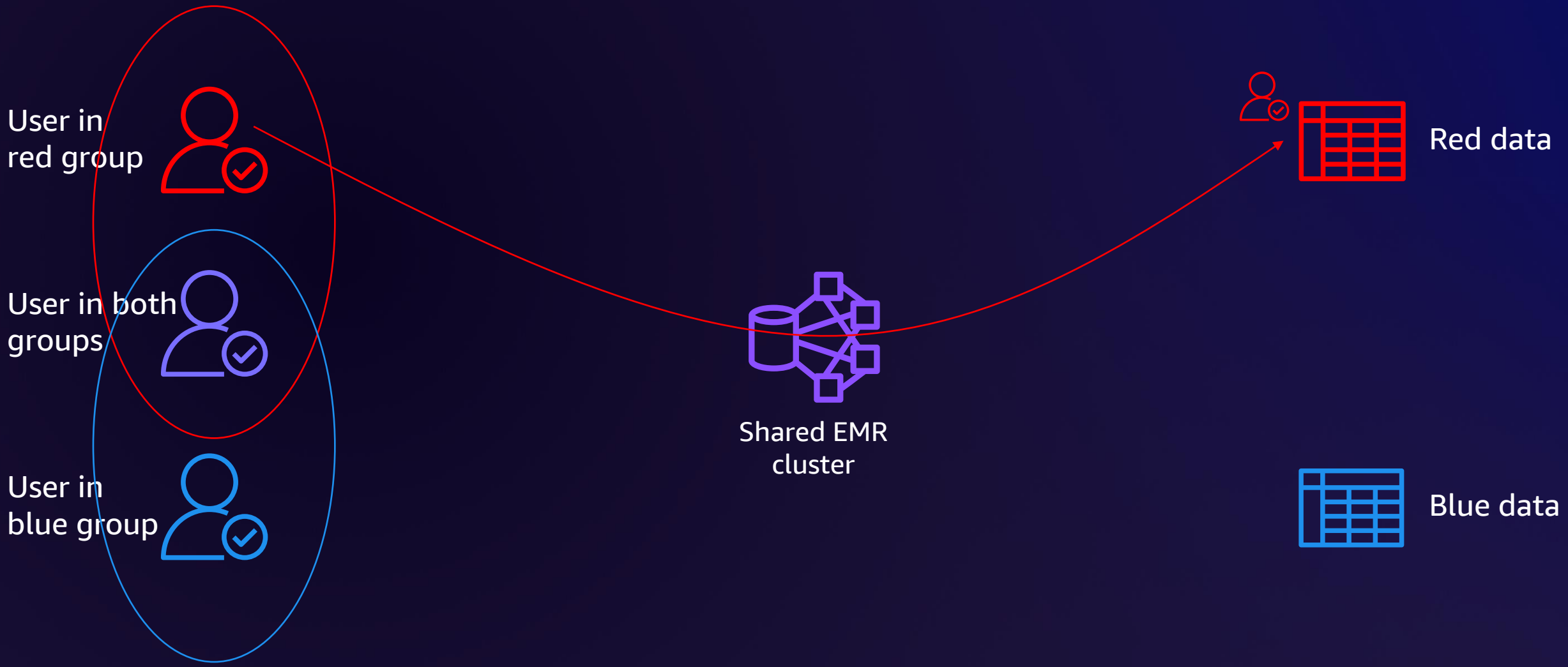
- **Amazon S3**
  - Storage for underlying table data
  - AWS CloudTrail records include user identifier

Rundown of AWS components

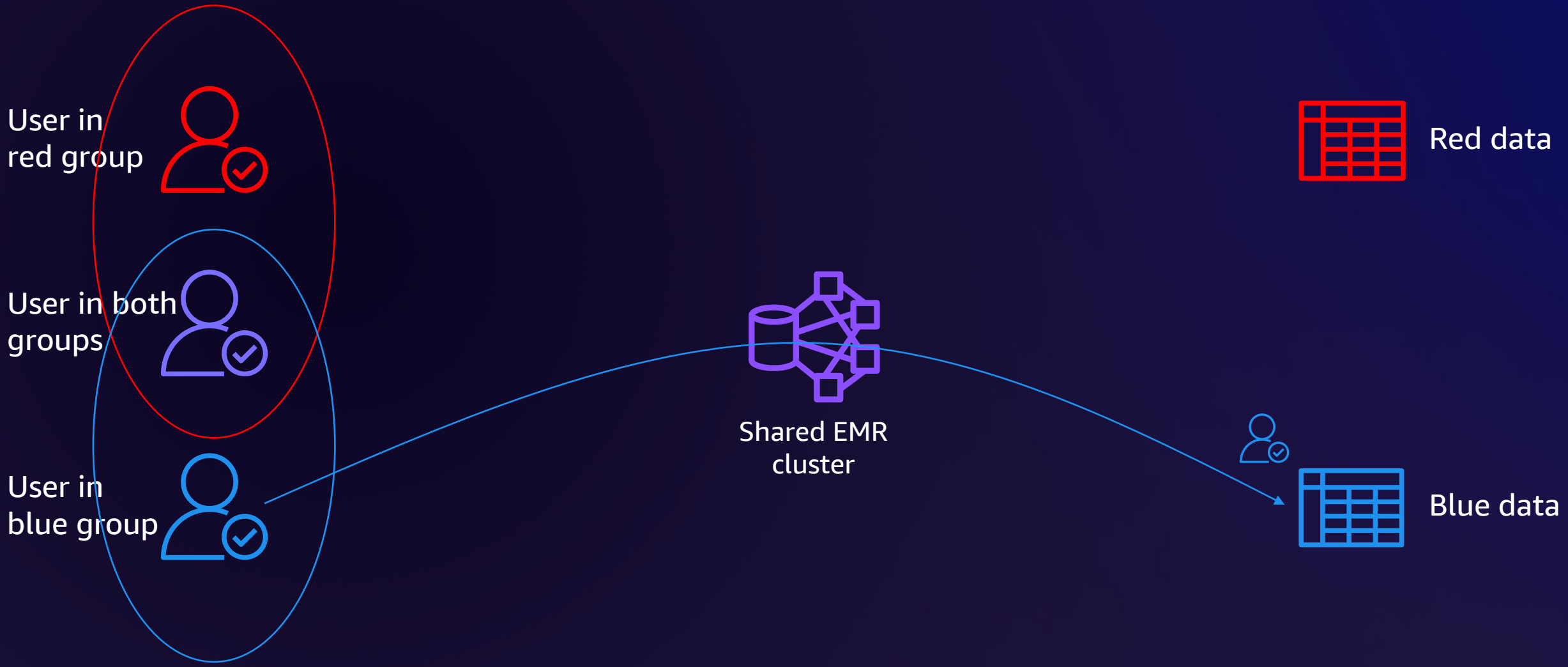
- **AWS Identity and Access Management (IAM)**
  - Every step is authenticated and authorized with IAM
  - User identity is an overlay on IAM principal identity
  - Data perimeters and all other IAM features in effect



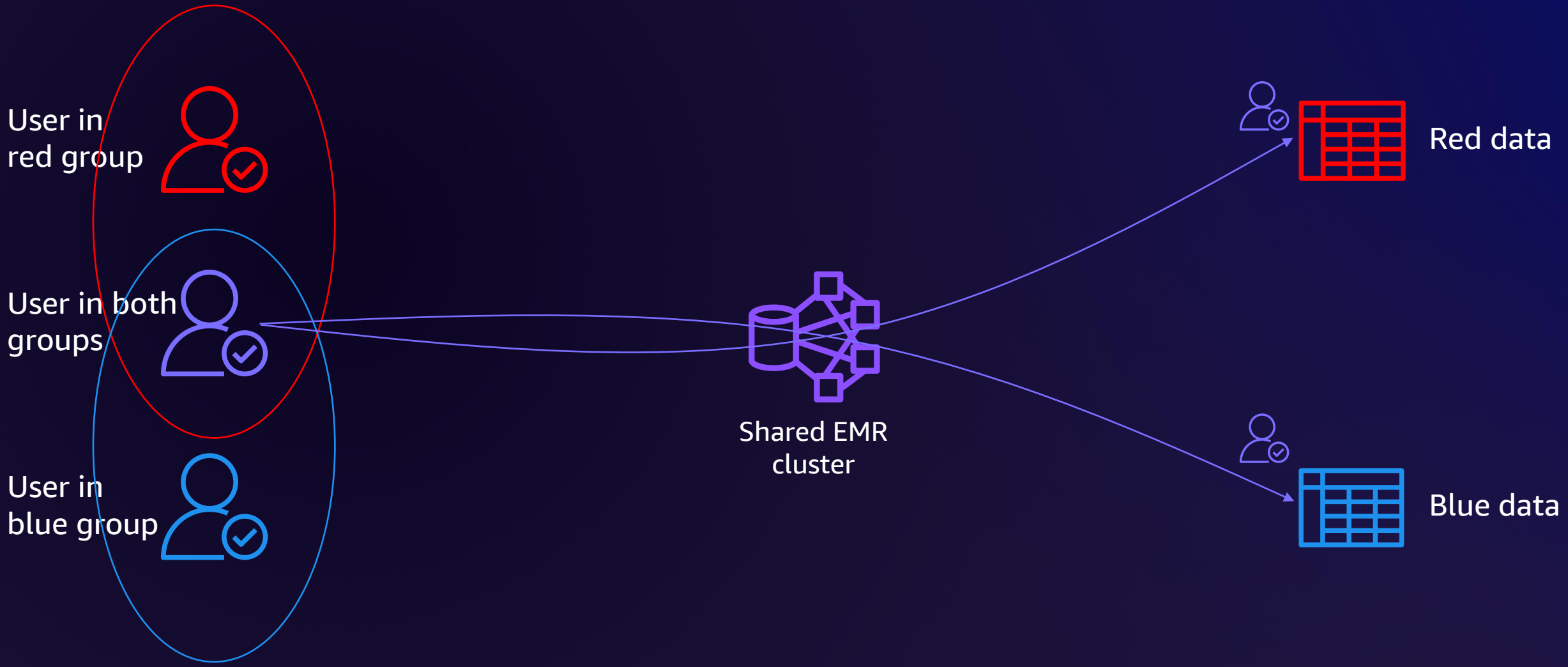
Trusted identity propagation in Amazon EMR: Our goal



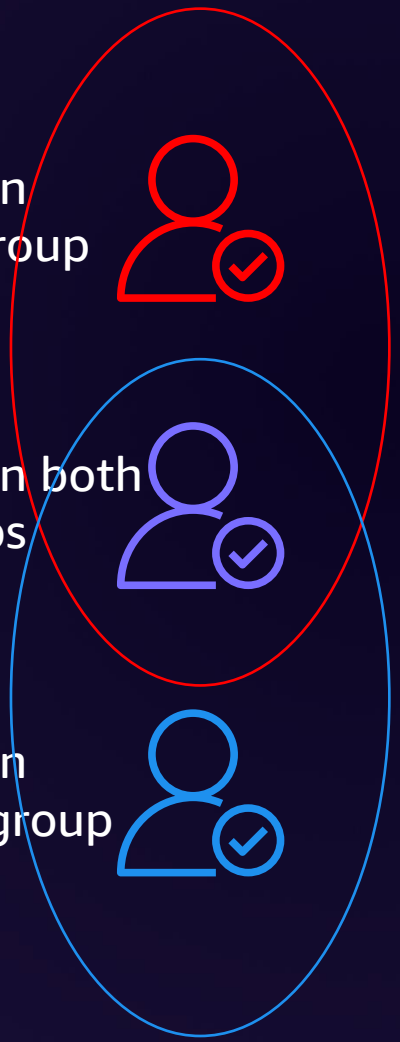
Trusted identity propagation in Amazon EMR: Our goal



Trusted identity propagation in Amazon EMR: Our goal

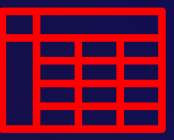


Trusted identity propagation in Amazon EMR: Our goal



AWS Lake Formation grants

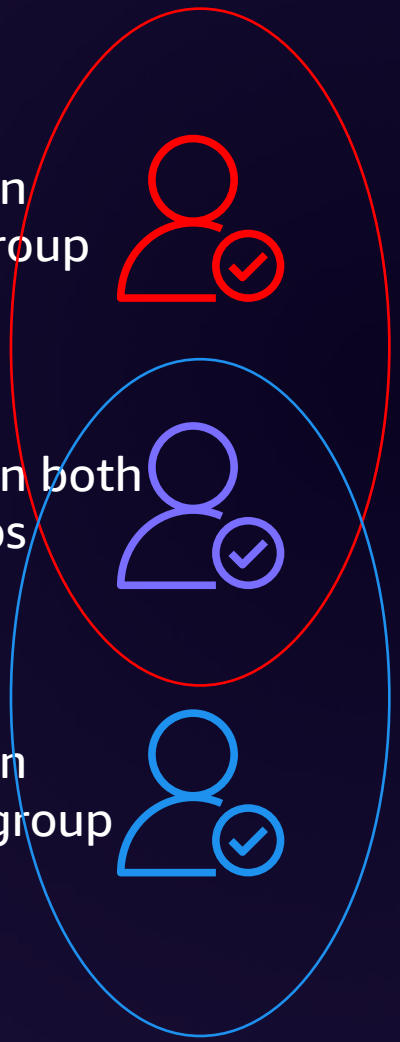
☐	0374d892-f081-7060-eebc-a9306f...	IAM Identi...	Table	tip-demo-db	us_cities_demographics
☐	9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	vehicles



Red data

Blue data

Trusted identity propagation in Amazon EMR: Our goal



Red data

"Red" group with access to "us_cities_demographics" table

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

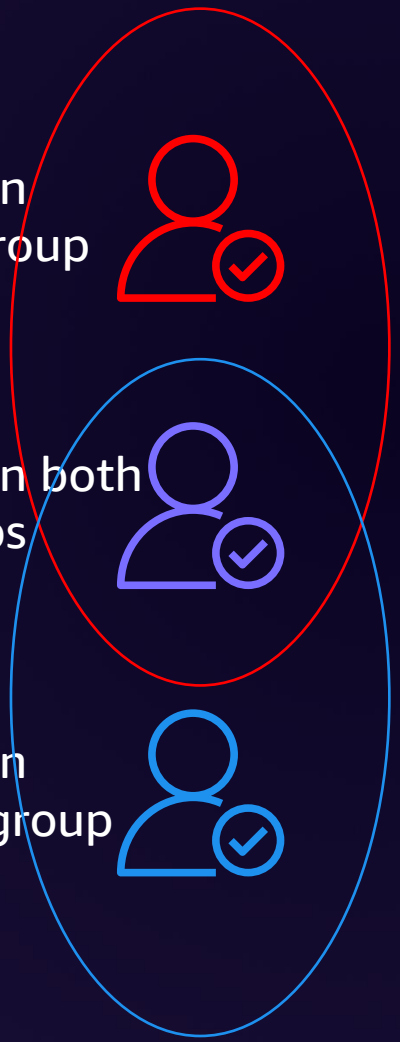
Blue data

Trusted identity propagation in Amazon EMR: Our goal

User in red group

User in both groups

User in blue group



AWS Lake Formation grants

Red data

"Red" group with access to "us_cities_demographics" table

☐	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" group with access to "vehicles" table

Blue data

Trusted identity propagation at work in Amazon EMR

User in red group

User in both groups

User in blue group

```
…

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

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

Red data

Blue data

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 red group

User in both groups

User in blue group

```
…

[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…
```

Red data

Blue data

No matching grant

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 red group

User in both groups

User in blue group

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

+------+----+------------+---------------+-----+
|region|year|manufacturer|          model|price|
+------+----+------------+---------------+-----+
|auburn|1992|        jeep|       cherokee| 4500|
…

[6] 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()

+------------+------------+----------+
|        city|       state|population|
+------------+------------+----------+
|    New York|    New York| 8550405.0|
…
```

Red data

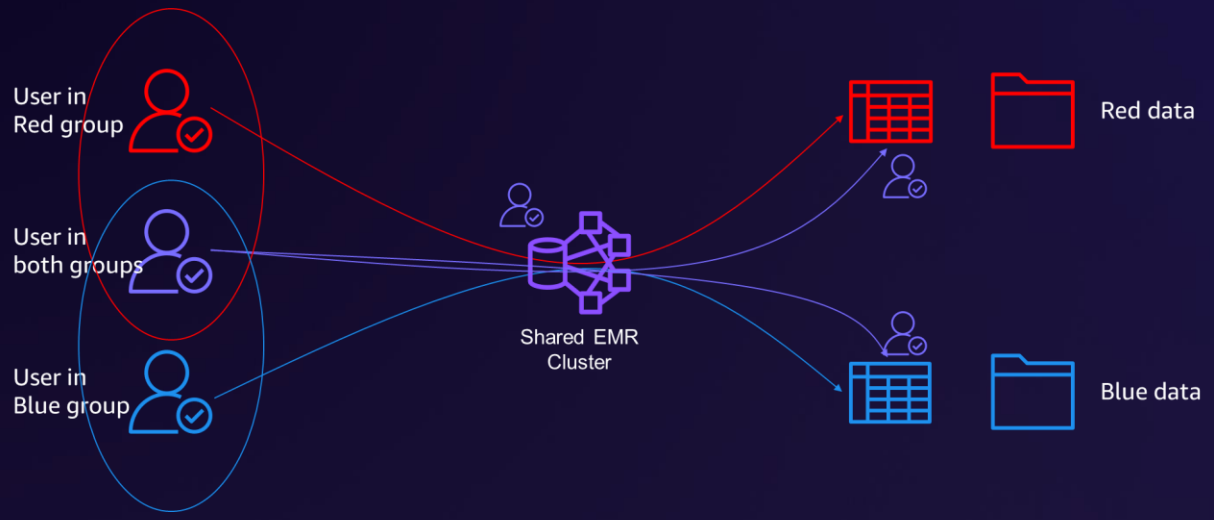
Blue data

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: What we got

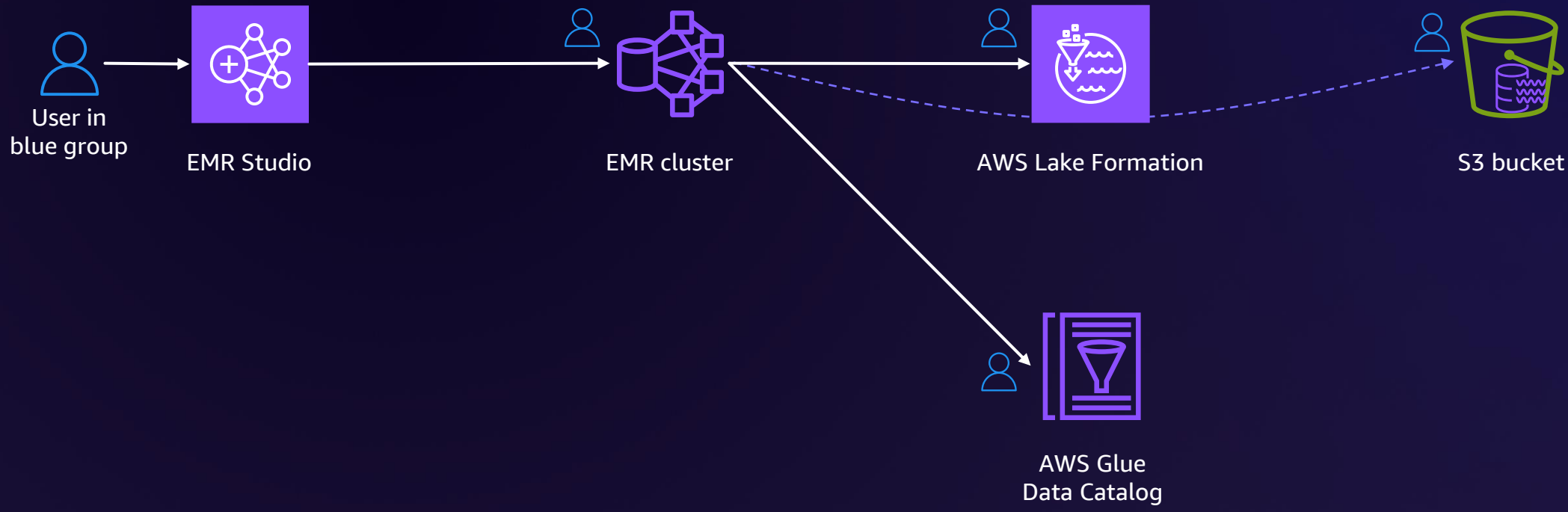
- Permissions are expressed in direct grants to users/groups
- Scalable permissions in one place: AWS Lake Formation
- User identity propagated fully through: Simplifies audit
- And these permissions will apply to other engines too: Amazon Athena, Amazon Redshift, Amazon QuickSight, etc.



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

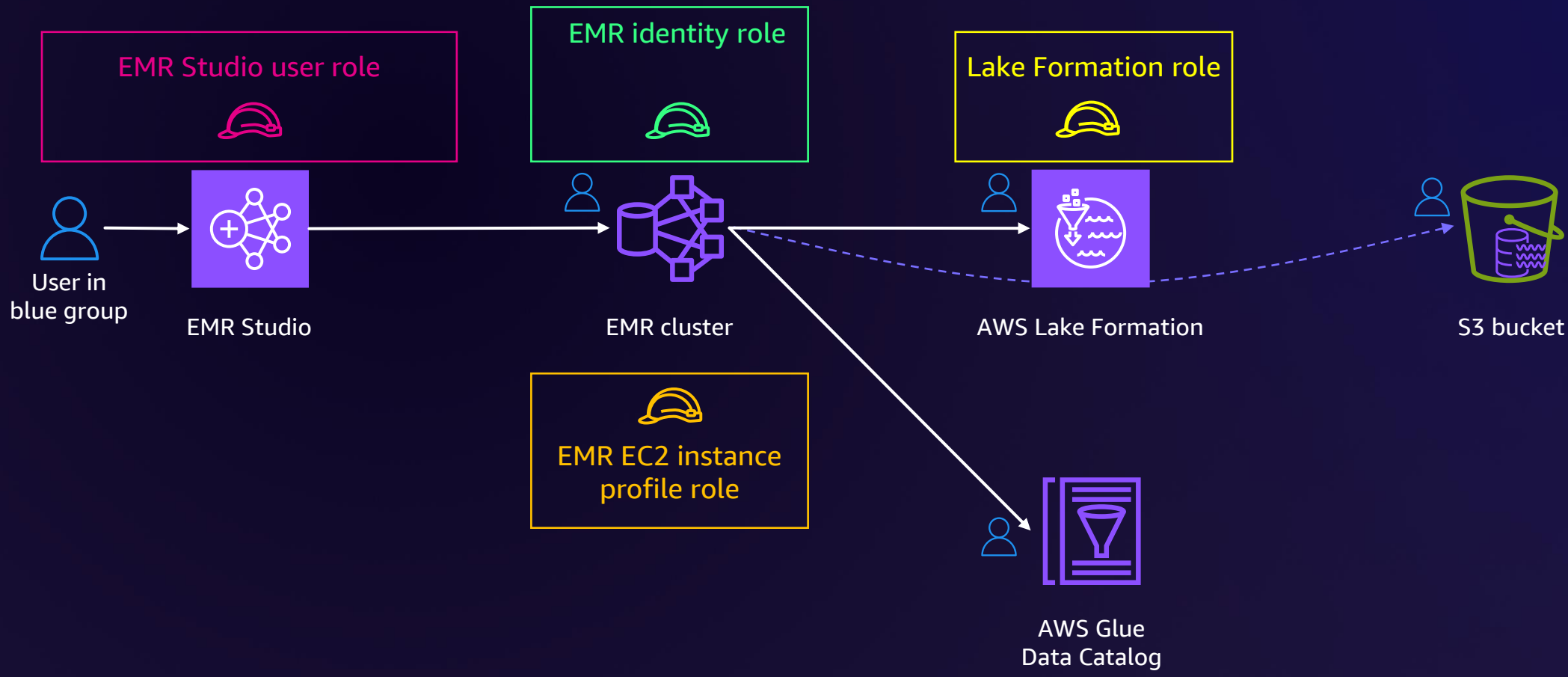
Identity propagation example: Amazon EMR

AMAZON EMR, AWS LAKE FORMATION, AND AMAZON S3



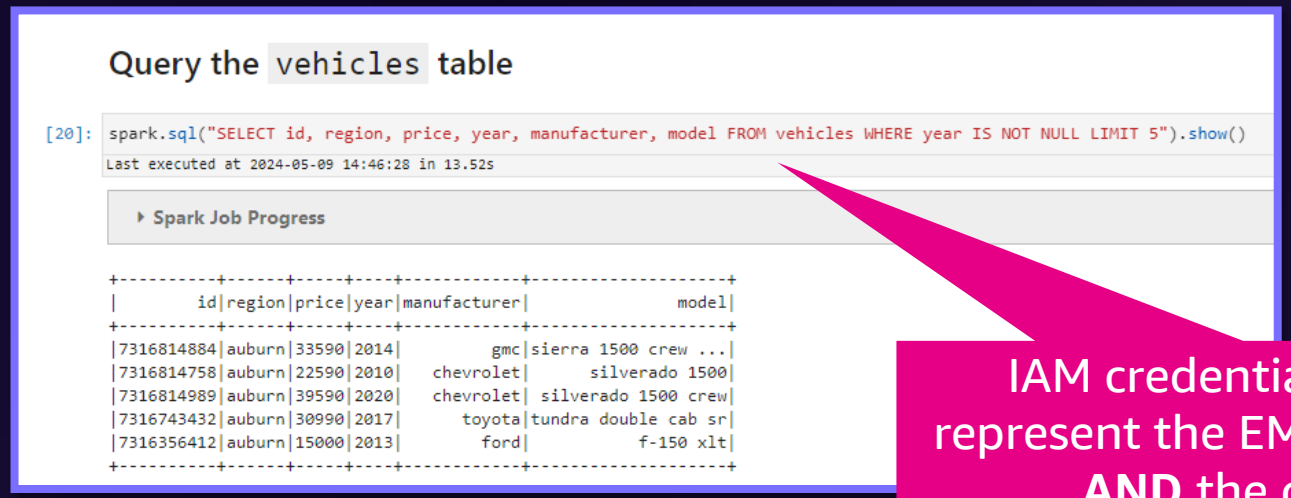
Identity propagation: IAM roles

AMAZON EMR, AWS LAKE FORMATION, AND AMAZON S3

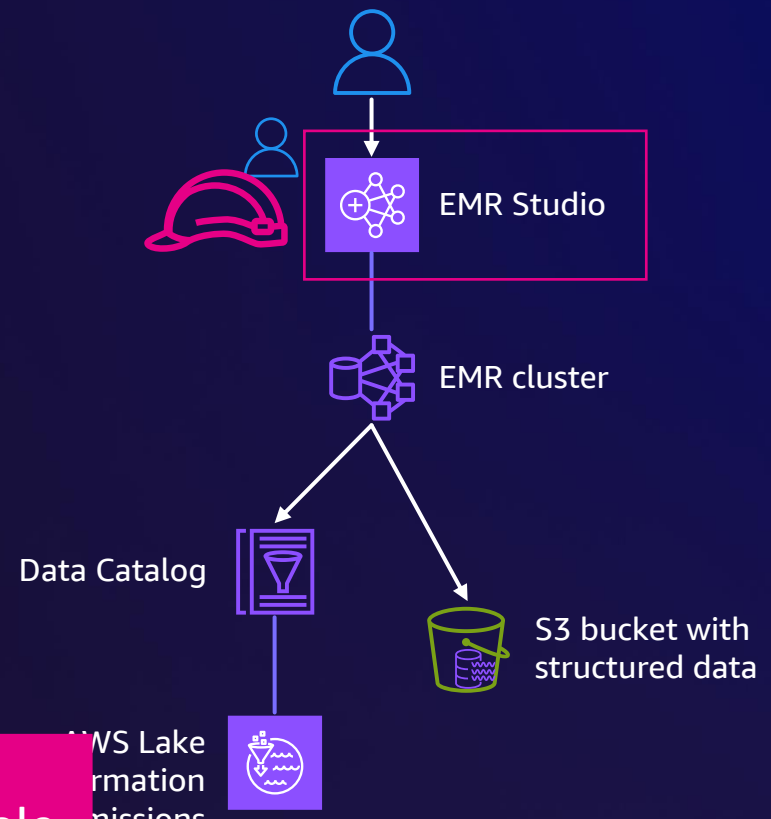


EMR Studio, IAM, and user identity

- **Amazon EMR Studio**
  - UX application – notebook environment
  - Authenticates the user at the entry point

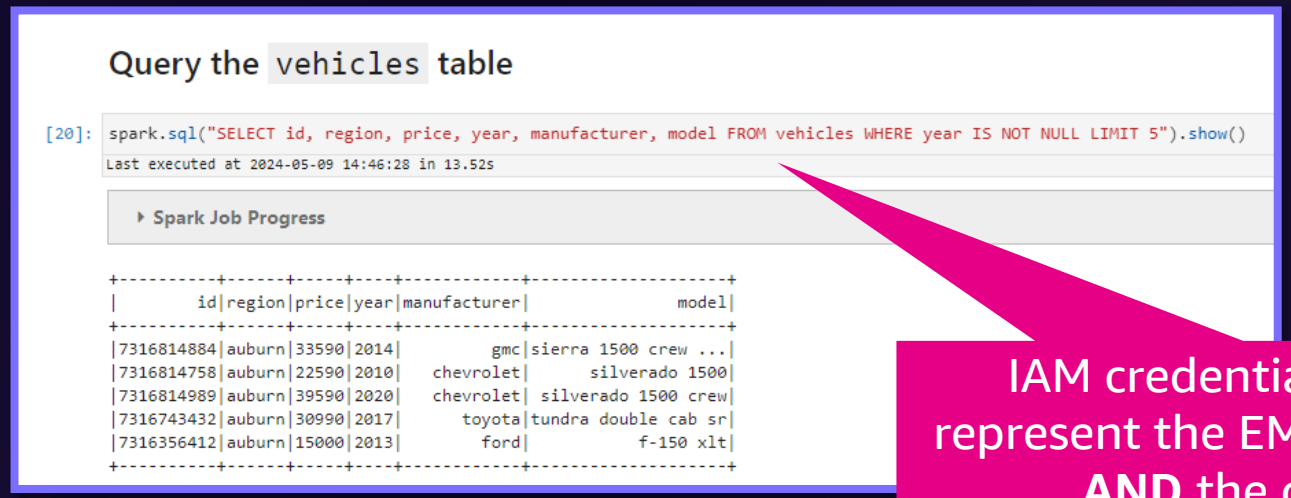


IAM credentials in notebook represent the EMR Studio user role **AND** the current user

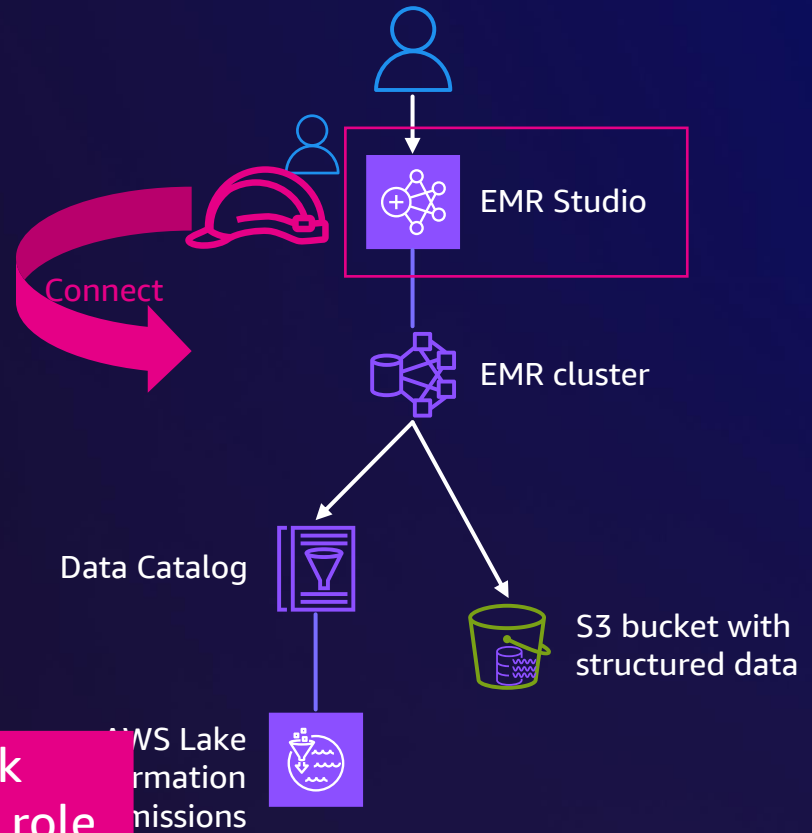


EMR Studio, IAM, and user identity

- **Amazon EMR Studio**
  - UX application – notebook environment
  - Authenticates the user at the entry point

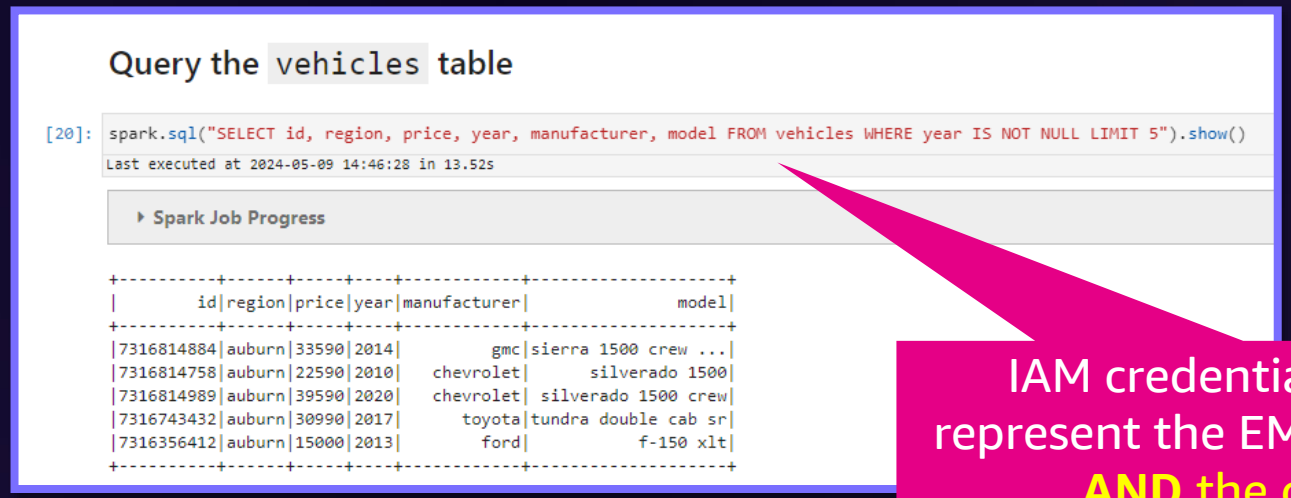


IAM credentials in notebook represent the EMR Studio user role **AND** the current user

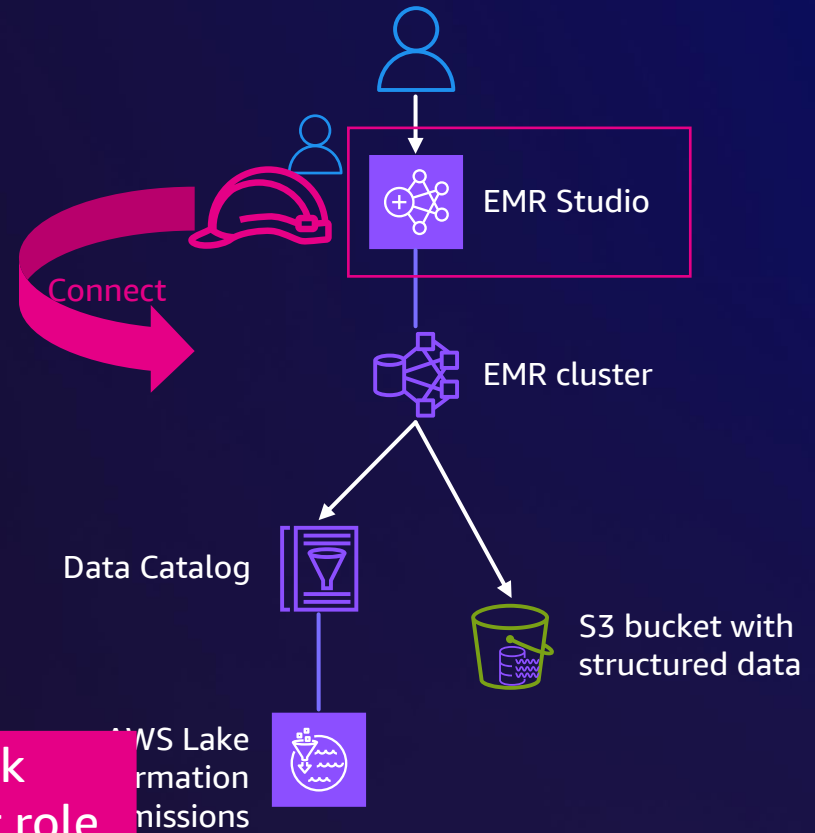


EMR Studio, IAM, and user identity

- **Amazon EMR Studio**
  - UX application – notebook environment
  - Authenticates the user at the entry point



IAM credentials in notebook represent the EMR Studio user role **AND** the current user



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"
  ]
}
```

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

EMR Studio, IAM, and user identity

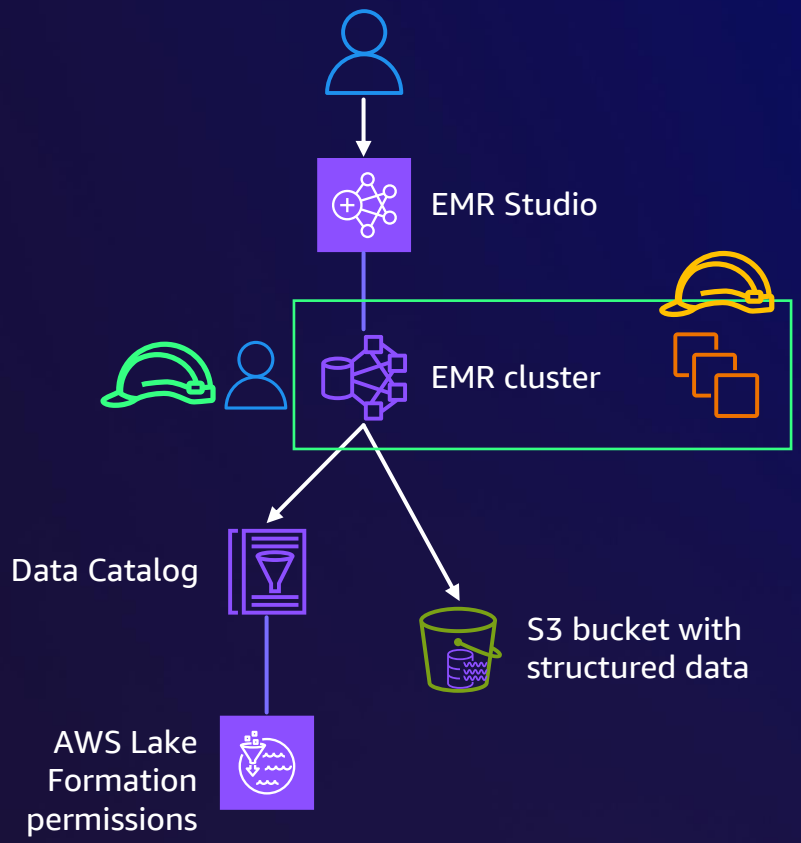
IAM principal policy: What it can do

```
{
  "Effect": "Allow",
  "Action":
    "elasticmapreduce:GetClusterSessionCredentials"
  ,
  "Resource": "arn:aws:elasticmapreduce:111122223…"
}
```

Connect to the cluster for an interactive session

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

IAM role trust policy (who can assume this role)

```
{
  "Effect": "Allow",
  "Principal": {
    "Service": "ec2.amazonaws.com"
  },
  "Action": "sts:AssumeRole"
}
```

EC2 instance role

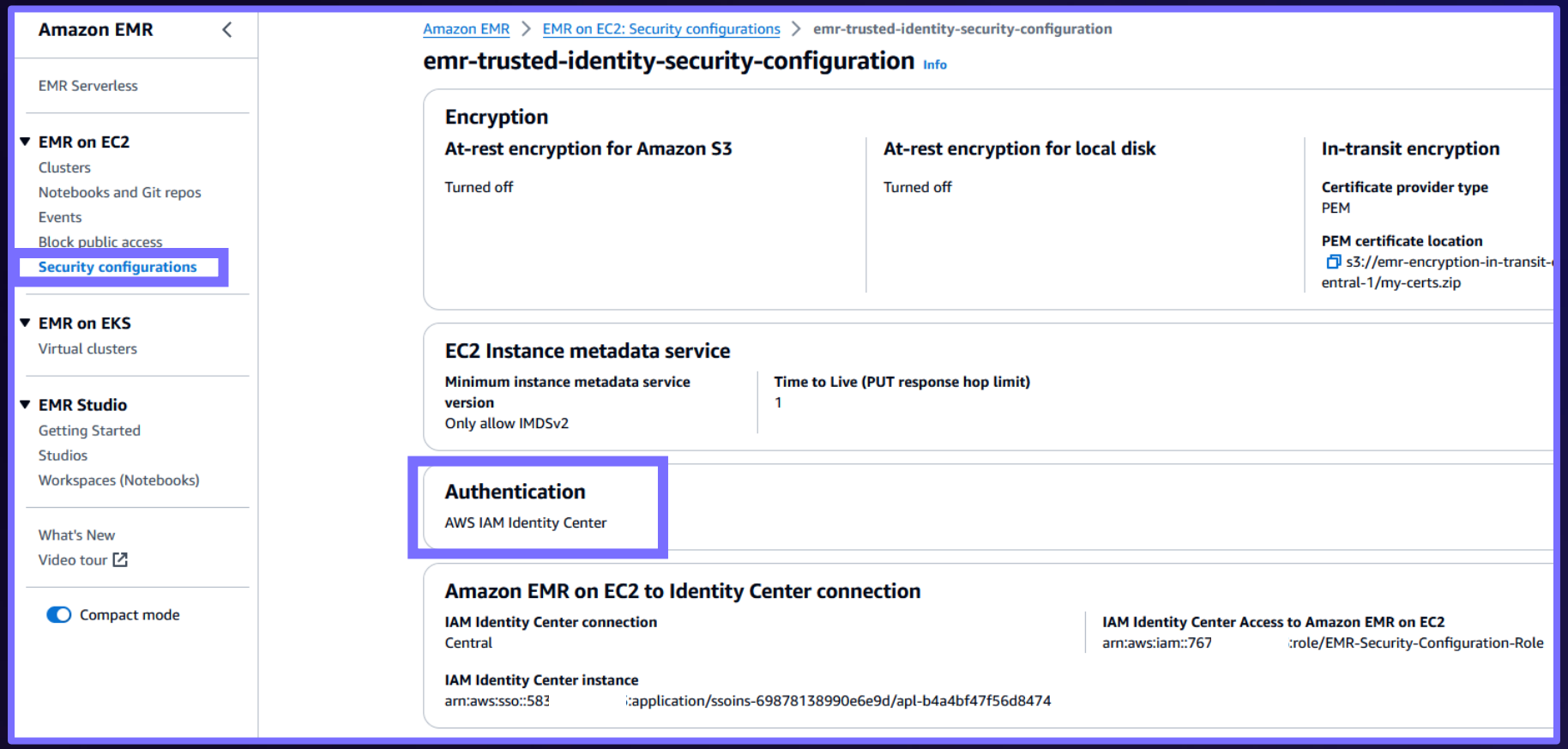
EMR cluster: EC2 instance profile role

Assume the other IAM role, including user identity (EMR identity role)

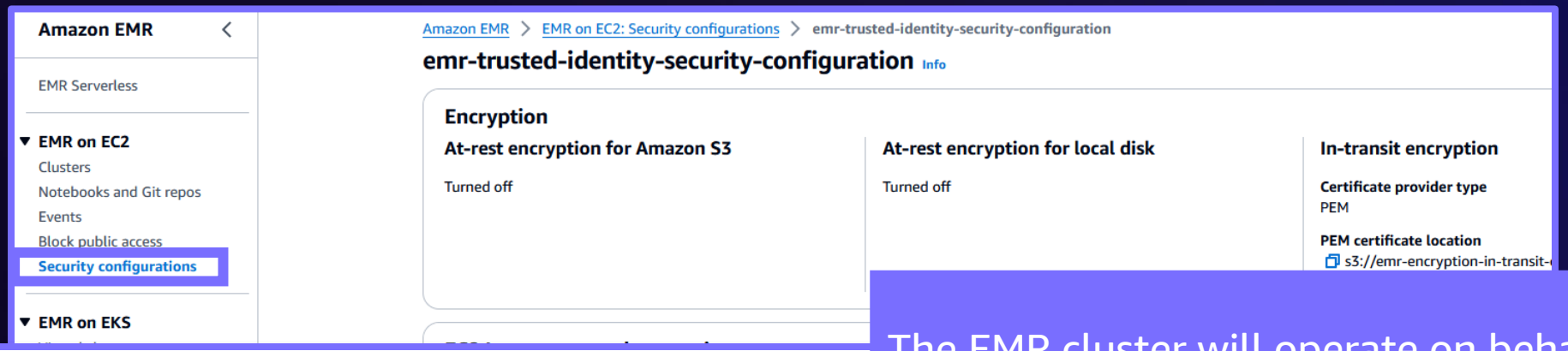
IAM principal policy (what it can do)

```
{
  "Effect": "Allow",
  "Action": [
    "sts:AssumeRole",
    "sts:SetContext"
  ],
  "Resource": "arn:aws:iam:…:role/EMRIdentityRole"
}
```

EMR security configuration for identity support



EMR security configuration for identity support



Authentication

AWS IAM Identity Center

Amazon EMR on EC2 to Identity Center connection

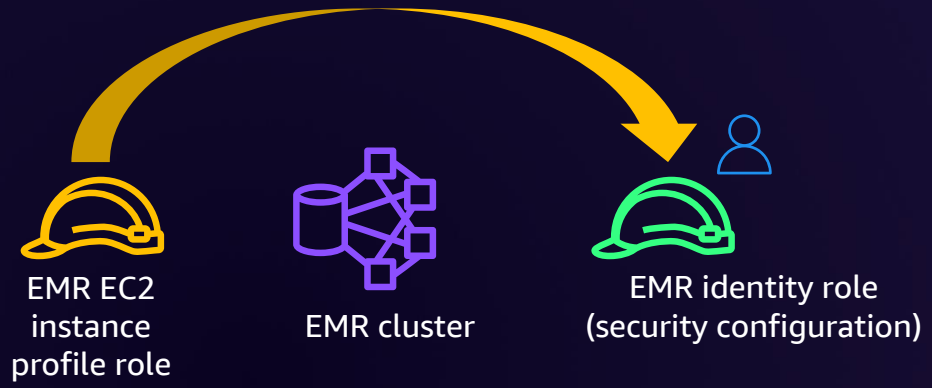
IAM Identity Center connection
Central

IAM Identity Center instance
arn:aws:sso:: :application/ssoins-69878138990e6e9d/apl-b4a4bf47f56d8474

IAM Identity Center Access to Amazon EMR on EC2
arn:aws:iam:: :role/EMR-Security-Configuration-Role

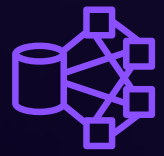
The EMR cluster will operate on behalf of an authenticated user, using this IAM role

EMR cluster: EMR identity role



EMR cluster: EMR identity role

EMR EC2 instance profile role



EMR cluster

EMR identity role (security configuration)

EMR cluster: EMR identity role

EMR cluster

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"
}
```

EMR cluster: EMR identity role

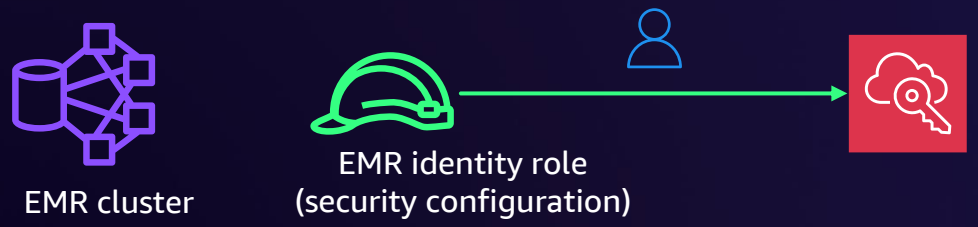
EMR cluster

EMR identity role
(security configuration)

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": "*"
}
```

EMR cluster: EMR identity role



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",
  "Resource": "*"
}
```

EMR cluster: EMR identity role

Access DB/table metadata on behalf of the user

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": "*"
}
```

EMR cluster: EMR identity role

IAM principal policy: What it can do

Data lake permissions (4)

Filter permissions by property or value — 4 matches

Database = tip-demo-db | and | Principal type = IAM Identity Center group | Clear filters

Principal	Princip...	Resour...	Database	Table	Resource	Permissions
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Database	tip-demo-db	-	tip-demo-db	Describe
0374d892-f081-7060-eebc-a9306f...	IAM Identi...	Table	tip-demo-db	us_cities_demographics	us_cities_demographics	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	vehicles	vehicles	All
9384b892-e011-7096-f3cf-923bb1...	IAM Identi...	Table	tip-demo-db	ALL_TABLES	ALL_TABLES	Describe

```
{
                                    ithIAM",
                                    ation/…/apl-123"

                                    lue:GetTable", …],

                                    ccess",
  "Resource": "*"
}
```

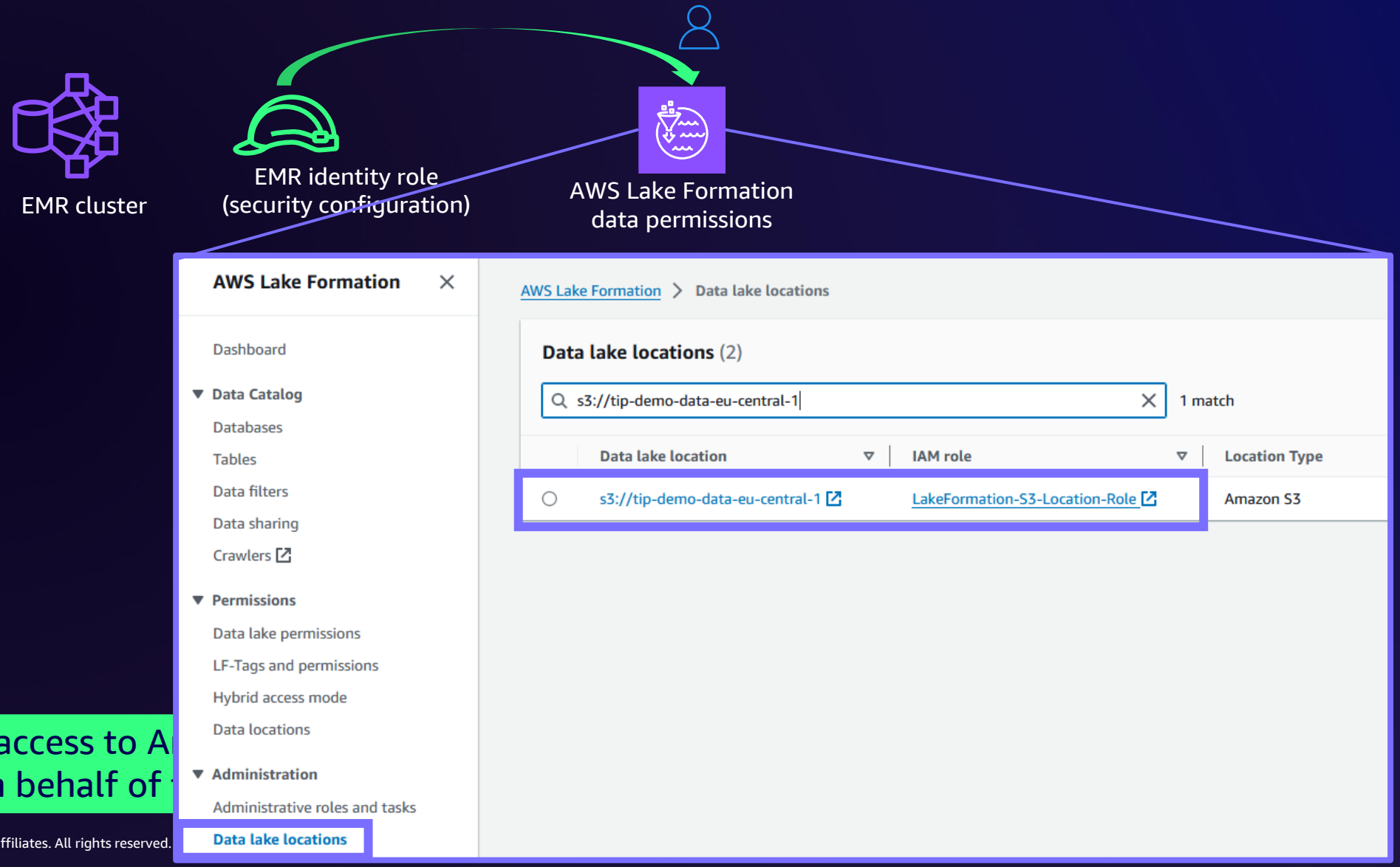
EMR cluster: EMR identity role

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": "*"
}
```

Request access to Amazon S3 data on behalf of the user

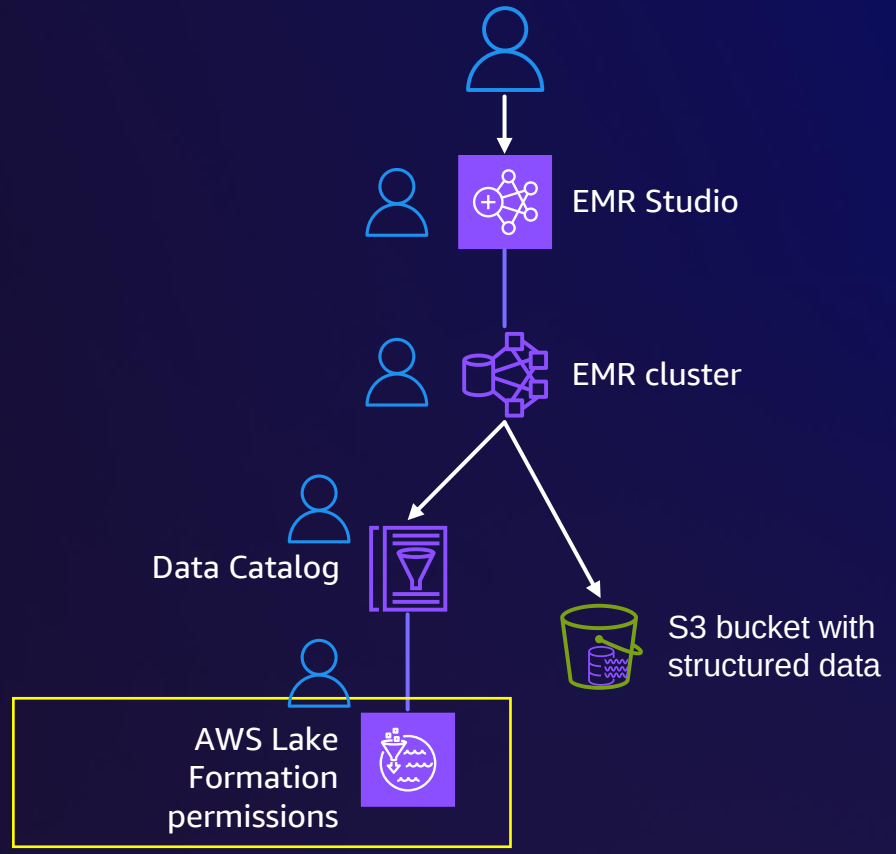
EMR cluster: EMR identity role



Request access to A
data on behalf of

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)**



Lake Formation data location role

Lake Formation data location role

Lake Formation data location role

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

IAM role trust policy (who can assume this role)

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

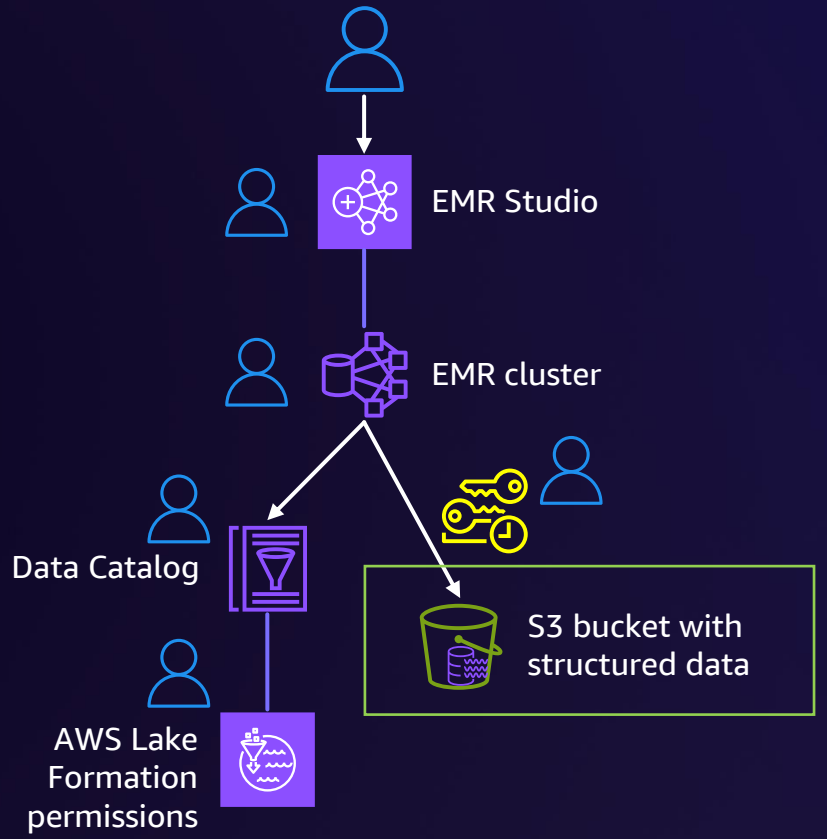
Lake Formation data location role

Short-term IAM credentials with exactly the right Amazon S3 permissions . . . so that the cluster can read the tables' contents

IAM principal policy: What it can do

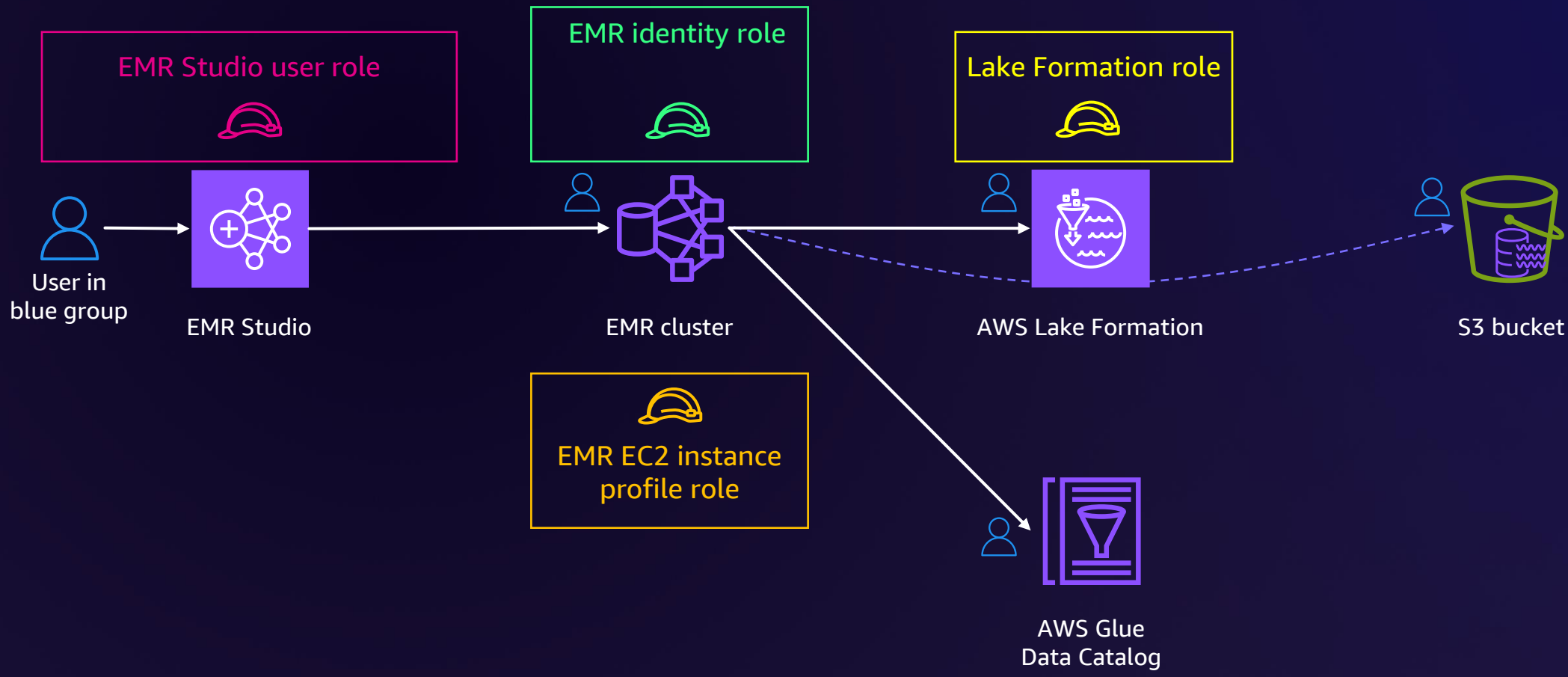
```
{
  "Effect": "Allow",
  "Action": ["s3:GetObject", "s3:PutObject",… ],
  "Resource": "arn:aws:s3:::example-bucket/*"
},
{
  "Effect: "Allow",
  "Action": "s3:ListBucket",
  "Resource": "arn:aws:s3:::example-bucket"
}
```

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

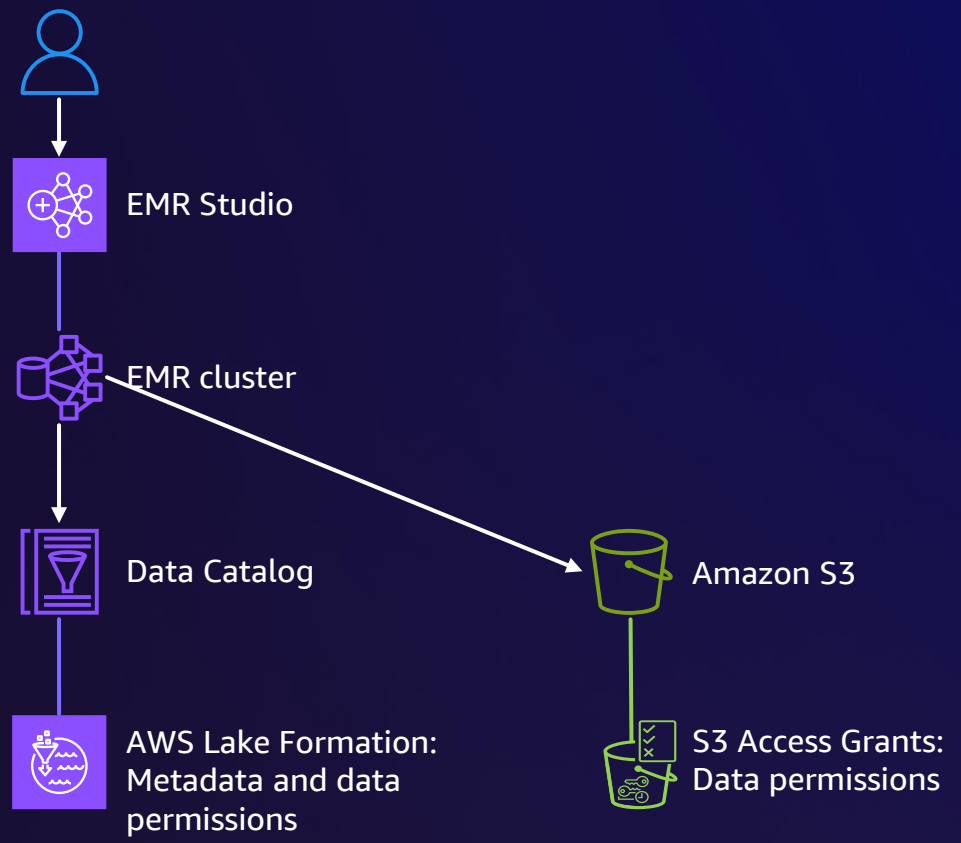
Direct reads and writes to Amazon S3 via Amazon EMR–Spark

```
[1] oldCarsDF = spark.sql("SELECT region,
year, manufacturer, model, price FROM
vehicles WHERE year < 2004")

[2] oldCarsDF.count()
47092
```

User in blue group

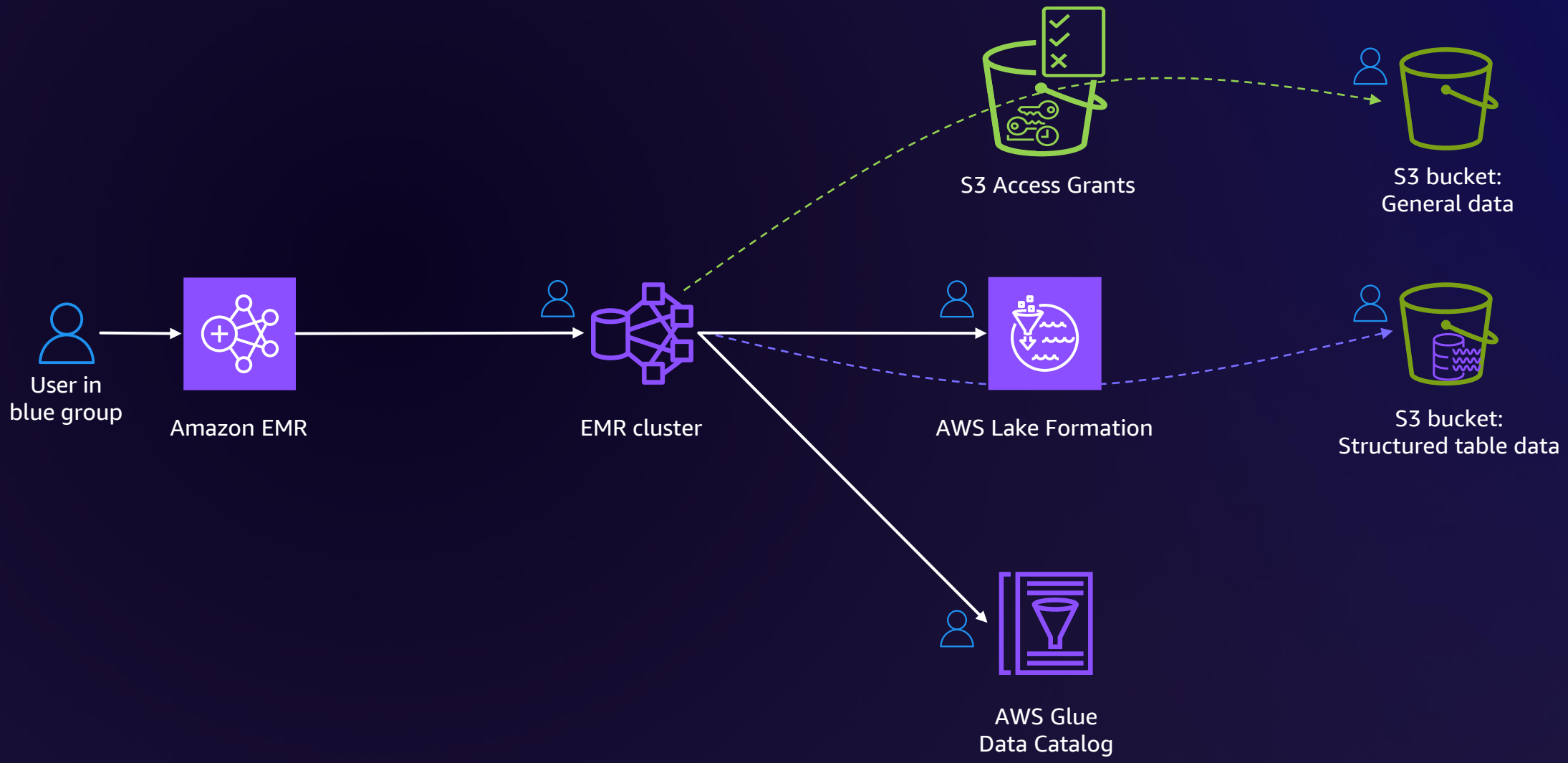
```
[3] oldCarsDF.write.parquet("s3://access-
grants-demo/blue/old-vehicles.parquet")
```

Store the result in Amazon S3



Amazon EMR with Lake Formation and direct-to-S3

AMAZON EMR, AWS LAKE FORMATION, S3 ACCESS GRANTS, AND AMAZON S3



Querying a table and writing the result

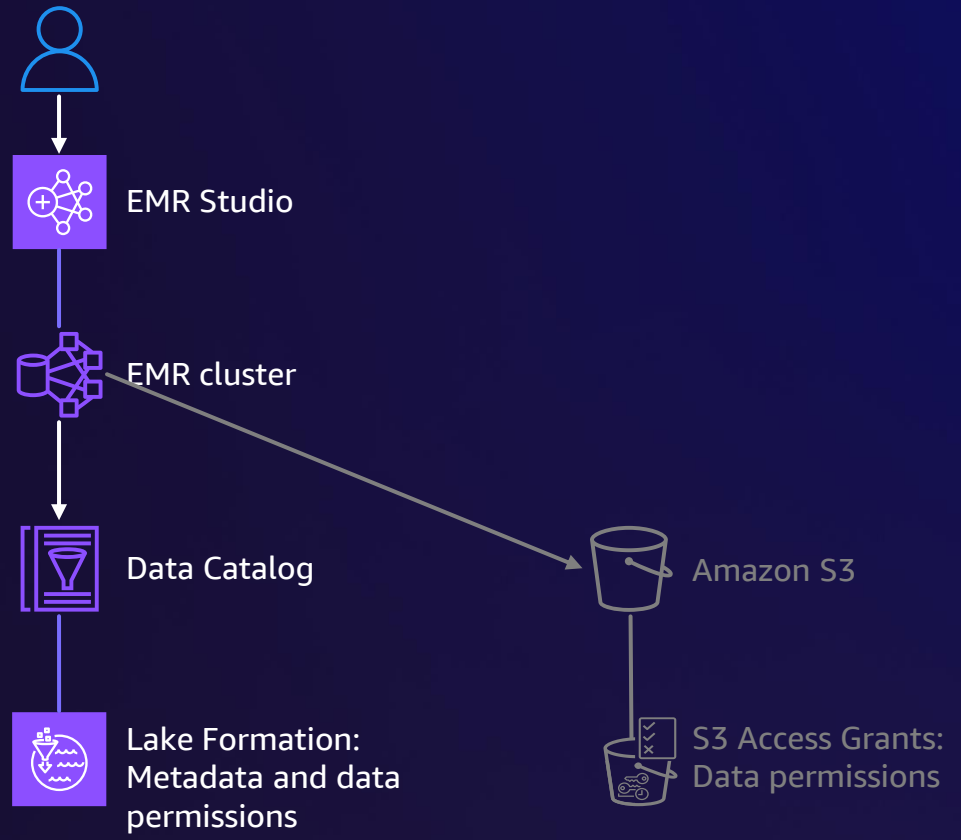
```
[1] oldCarsDF = spark.sql("SELECT region, year,
manufacturer, model, price FROM vehicles WHERE year <
2004")

[2] oldCarsDF.count()
47092
```

Make a new table: Lake Formation permissions

```
[3] oldCarsDF.write.parquet("s3://access-grants-
demo/blue/old-vehicles.parquet")
```

Store the result in Amazon S3

User in blue group



Querying a table and writing the result

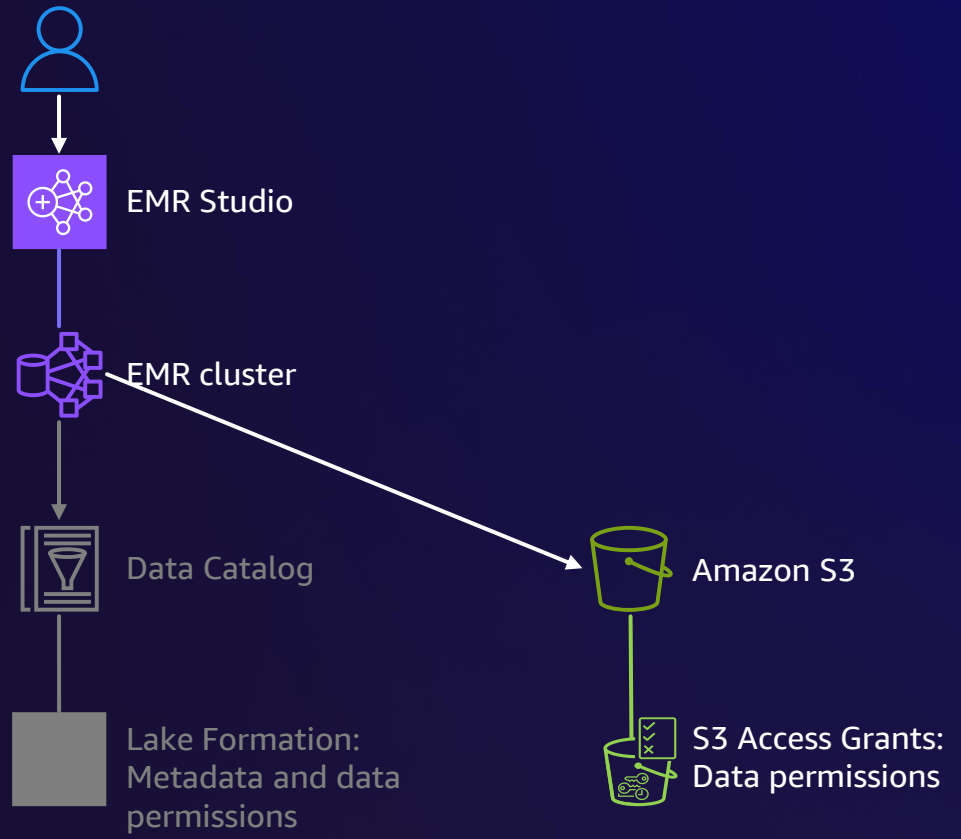
```
[1] oldCarsDF = spark.sql("SELECT region, year,
manufacturer, model, price FROM vehicles WHERE year <
2004")

[2] oldCarsDF.count()
47092
```

Make a new table

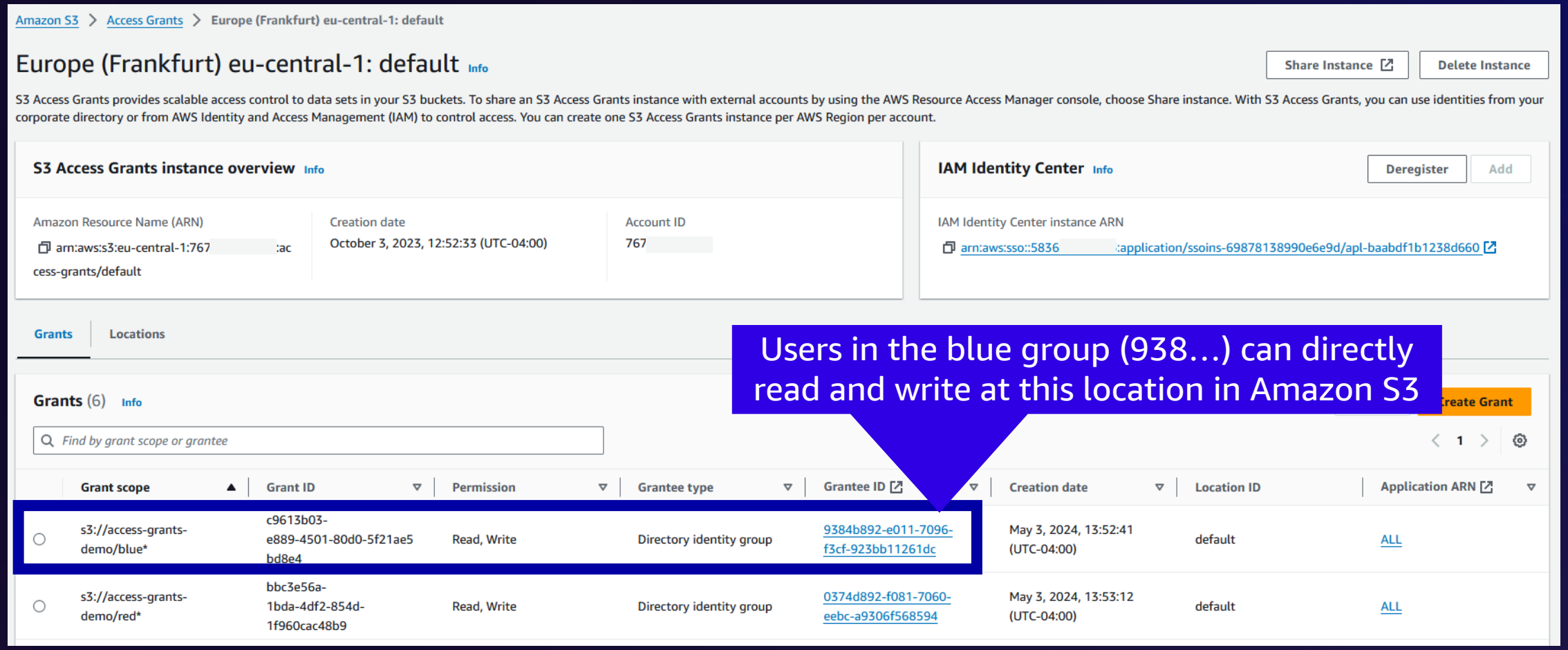
```
[3] oldCarsDF.write.parquet("s3://access-grants-
demo/blue/old-vehicles.parquet")
```

Store the result in Amazon S3

User in blue group



S3 Access Grants: Data access for users and groups



Audit: CloudTrail directly attributes access to user

```
[3] oldCarsDF.write.parquet("s3://access-grants-
demo/blue/old-vehicles.parquet")
```

User in blue group

AWS CloudTrail data event from Amazon S3

```
{
  "userIdentity": {
    "type": "AssumedRole",
    "arn": "arn:aws:sts::…:S3AccessGrantsRole…"
      "onbehalfof": {
        "userid": "63548842-…"
      }
  },
  "eventSource": "s3.amazonaws.com",
  "eventName": "GetObject",
  …
}
```

Direct attribution of access event to end user

EMR Studio

EMR cluster

Data Catalog

Lake Formation: Metadata and data permissions

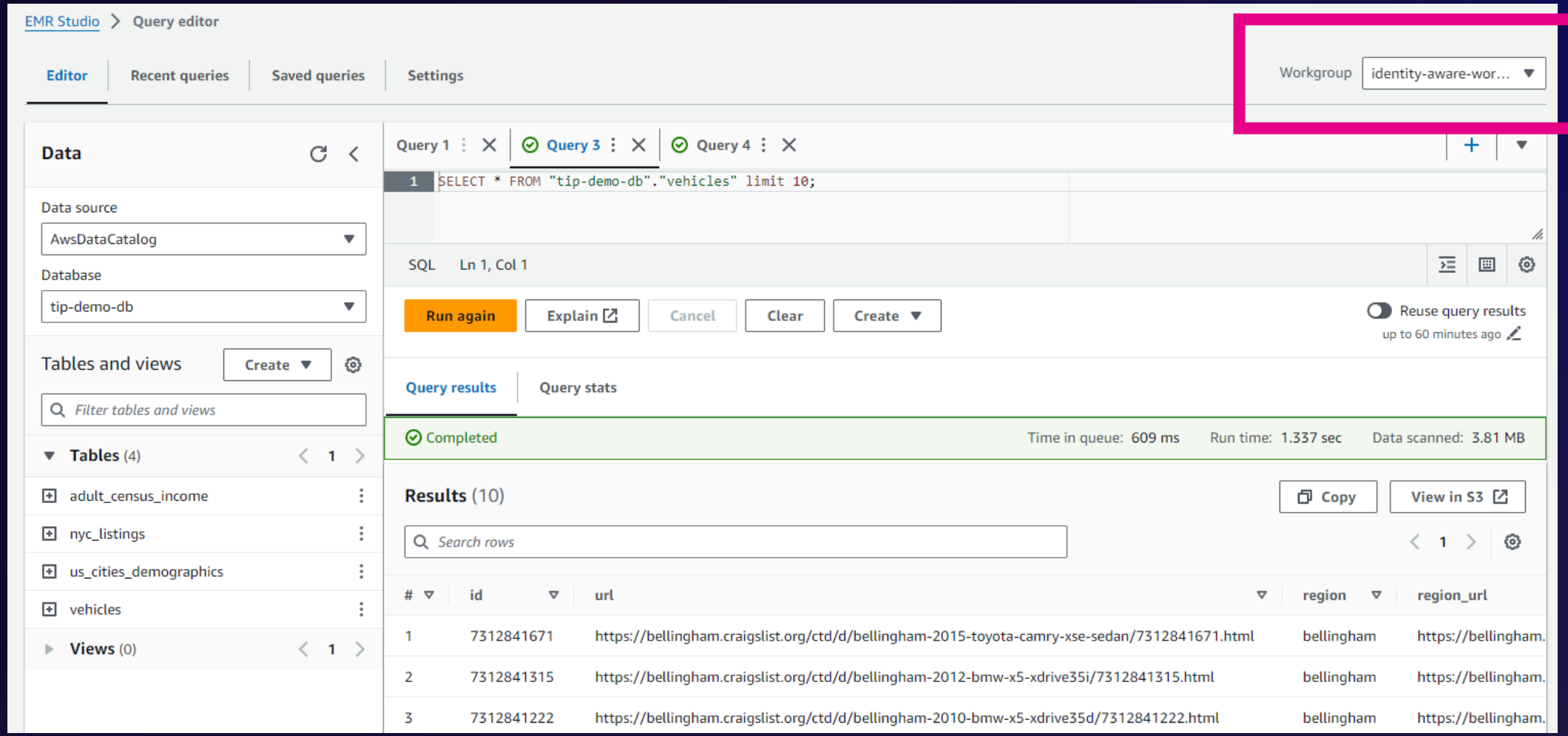
Amazon S3

S3 Access Grants: Data permissions

Adding another analytics engine: Identity support in Amazon Athena workgroups

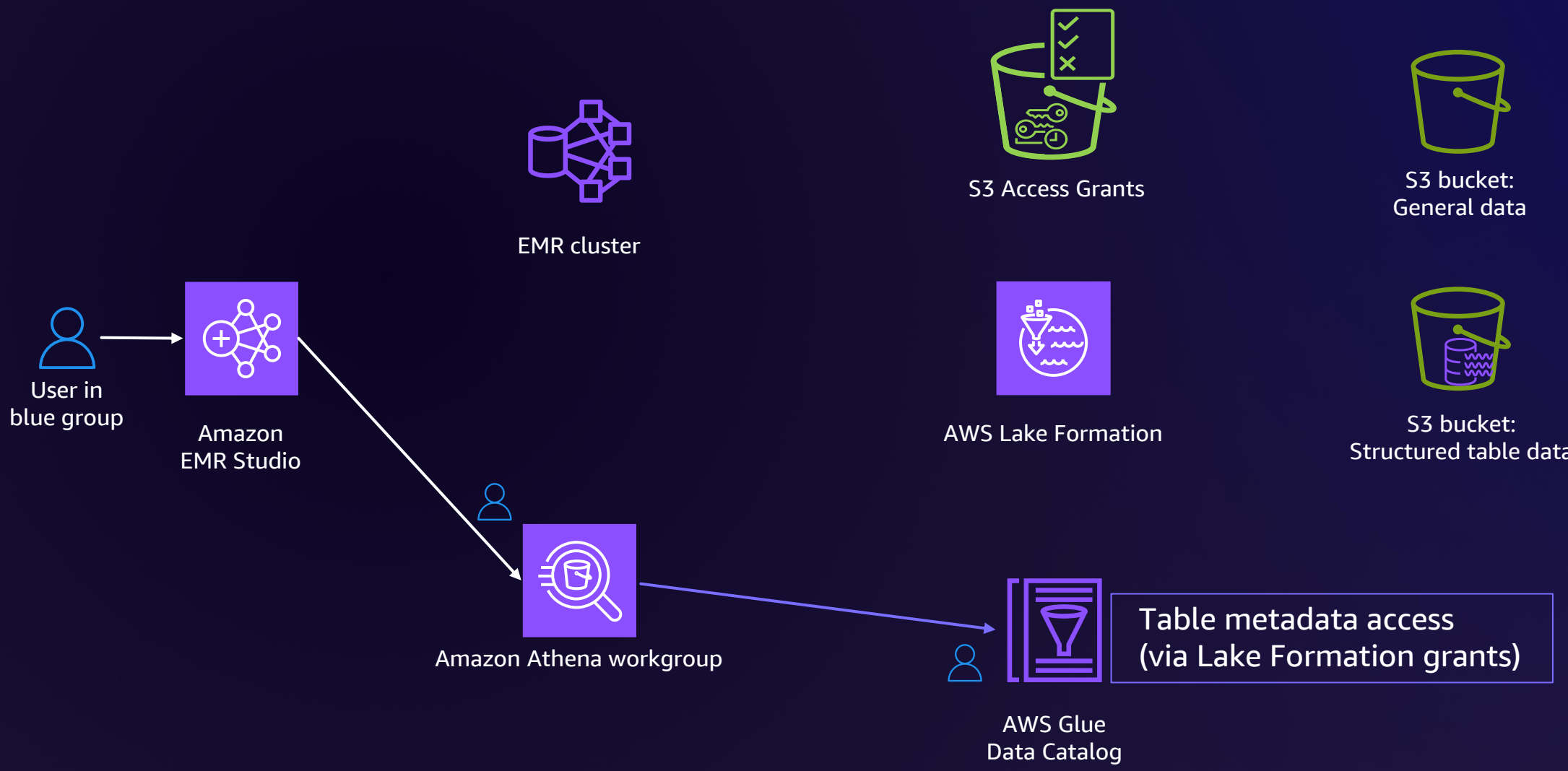
Adding a second analytics engine: Amazon Athena

Athena SQL query editor in EMR Studio



Identity-aware queries through Amazon Athena

Identity-aware queries through Amazon Athena

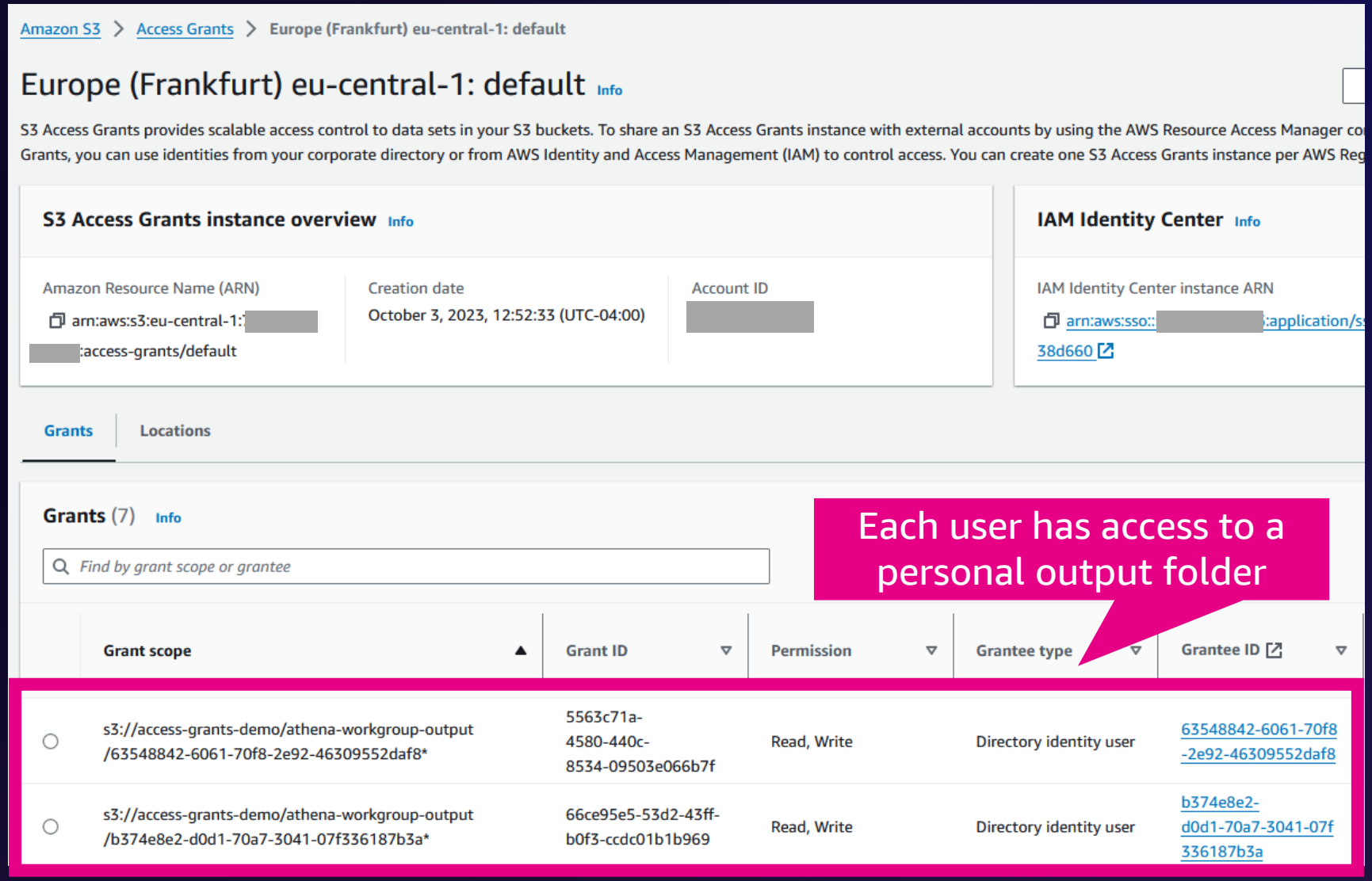


Identity-aware queries through Amazon Athena

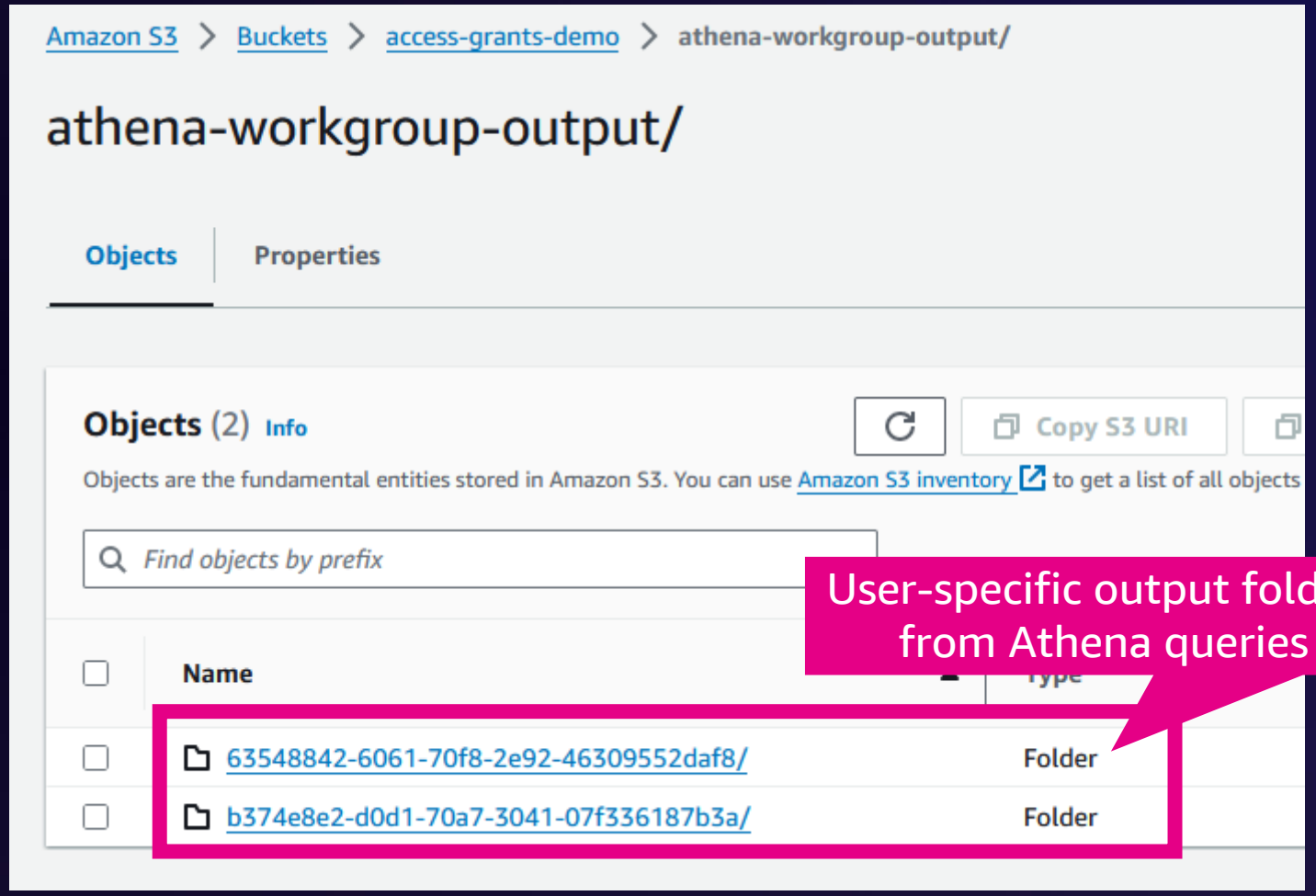
Identity-aware queries through Amazon Athena

Identity-aware queries through Amazon Athena

User output locations in Athena with S3 Access Grants

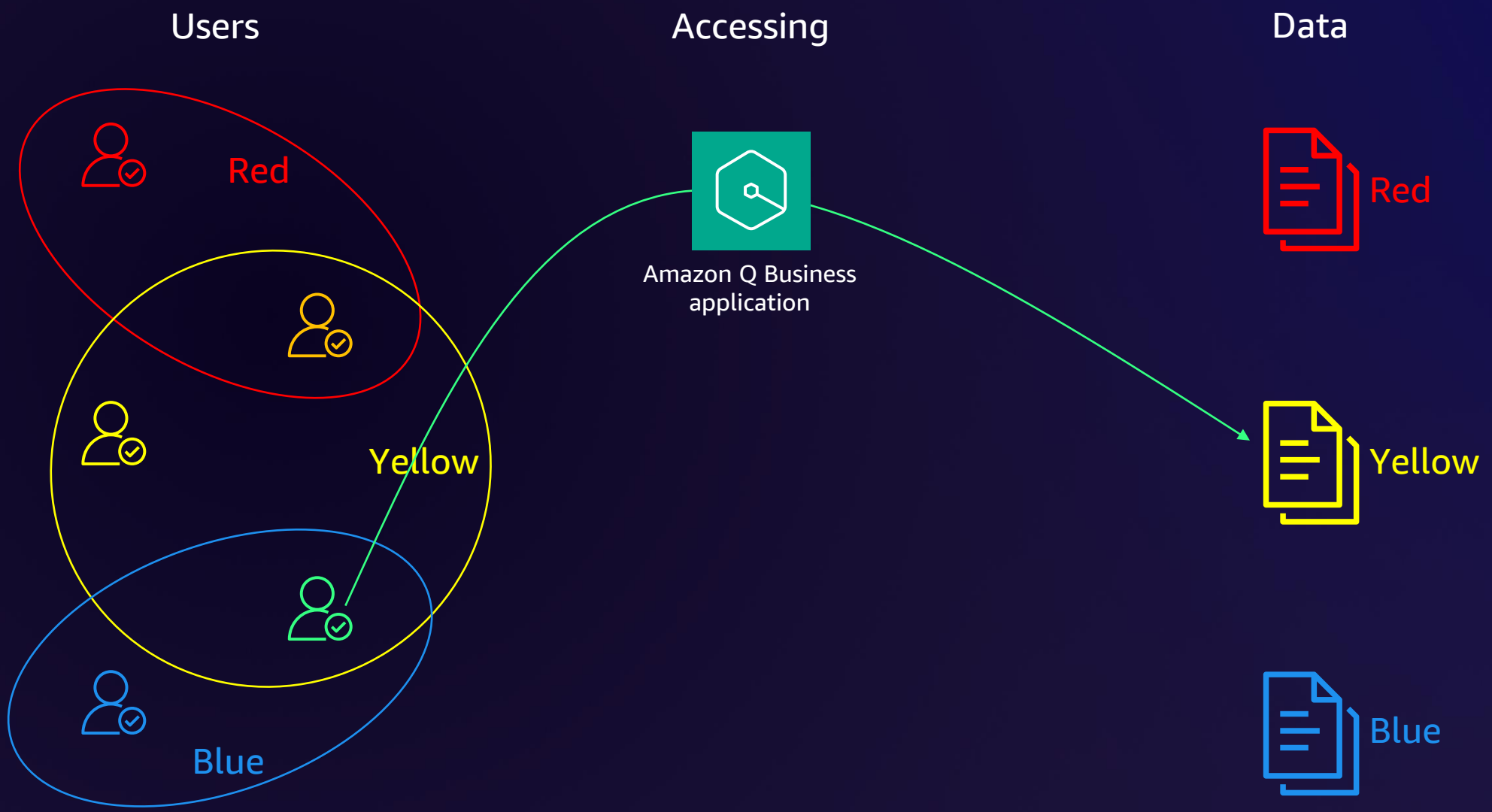


User output locations in Athena with S3 Access Grants

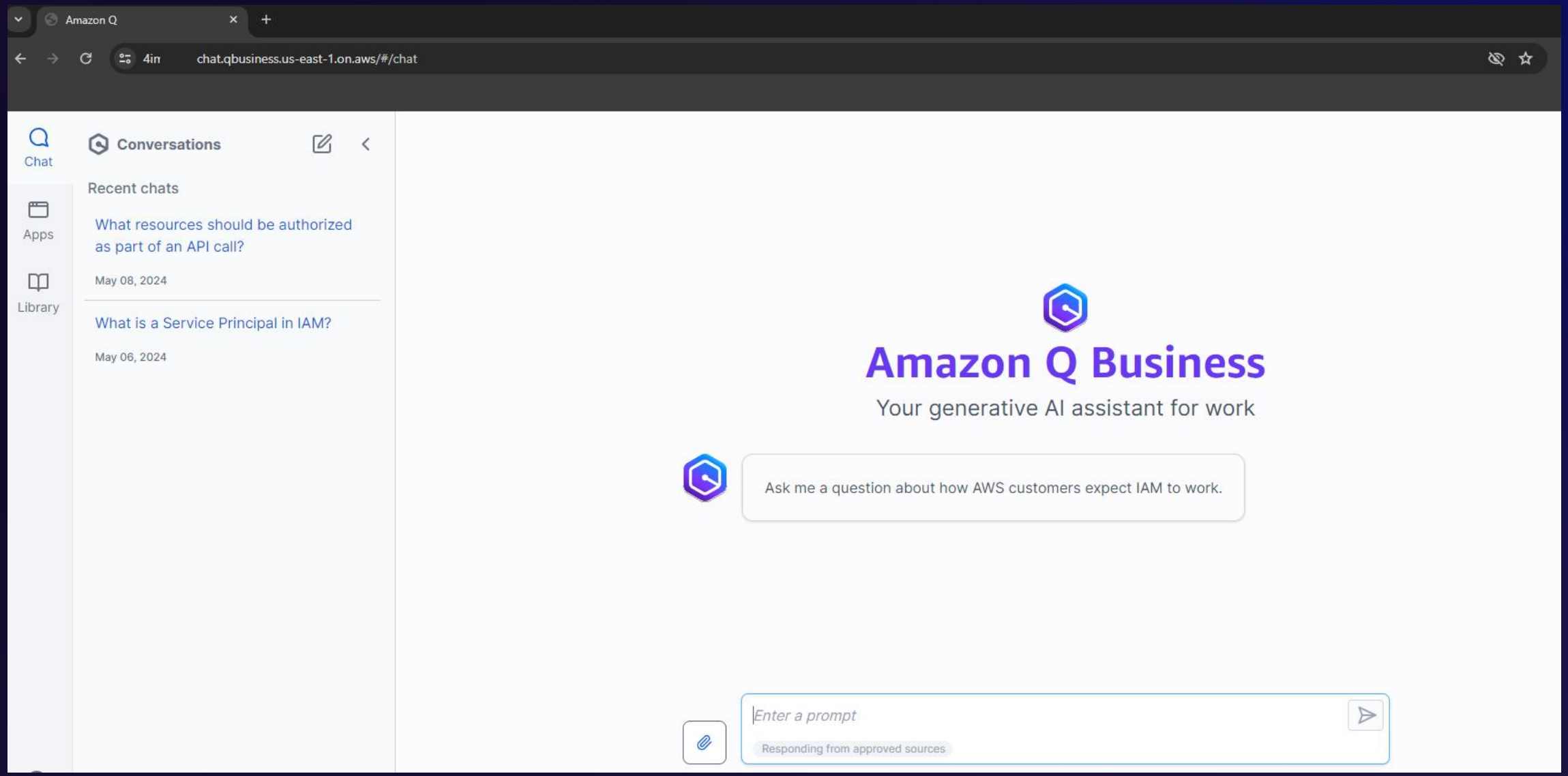


Identity in generative AI: Amazon Q Business and Amazon Q Developer

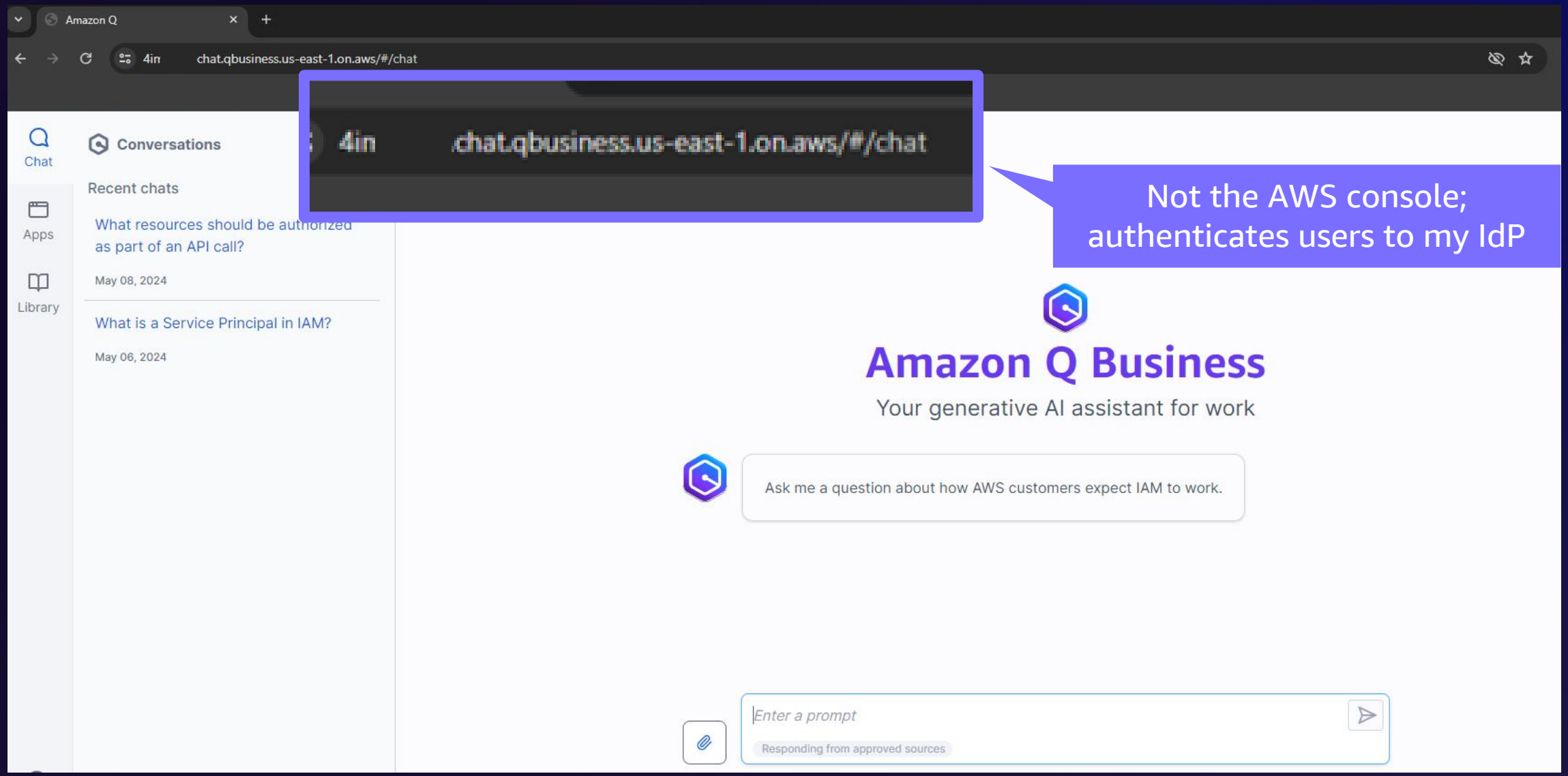
Generative AI atop a strong data foundation



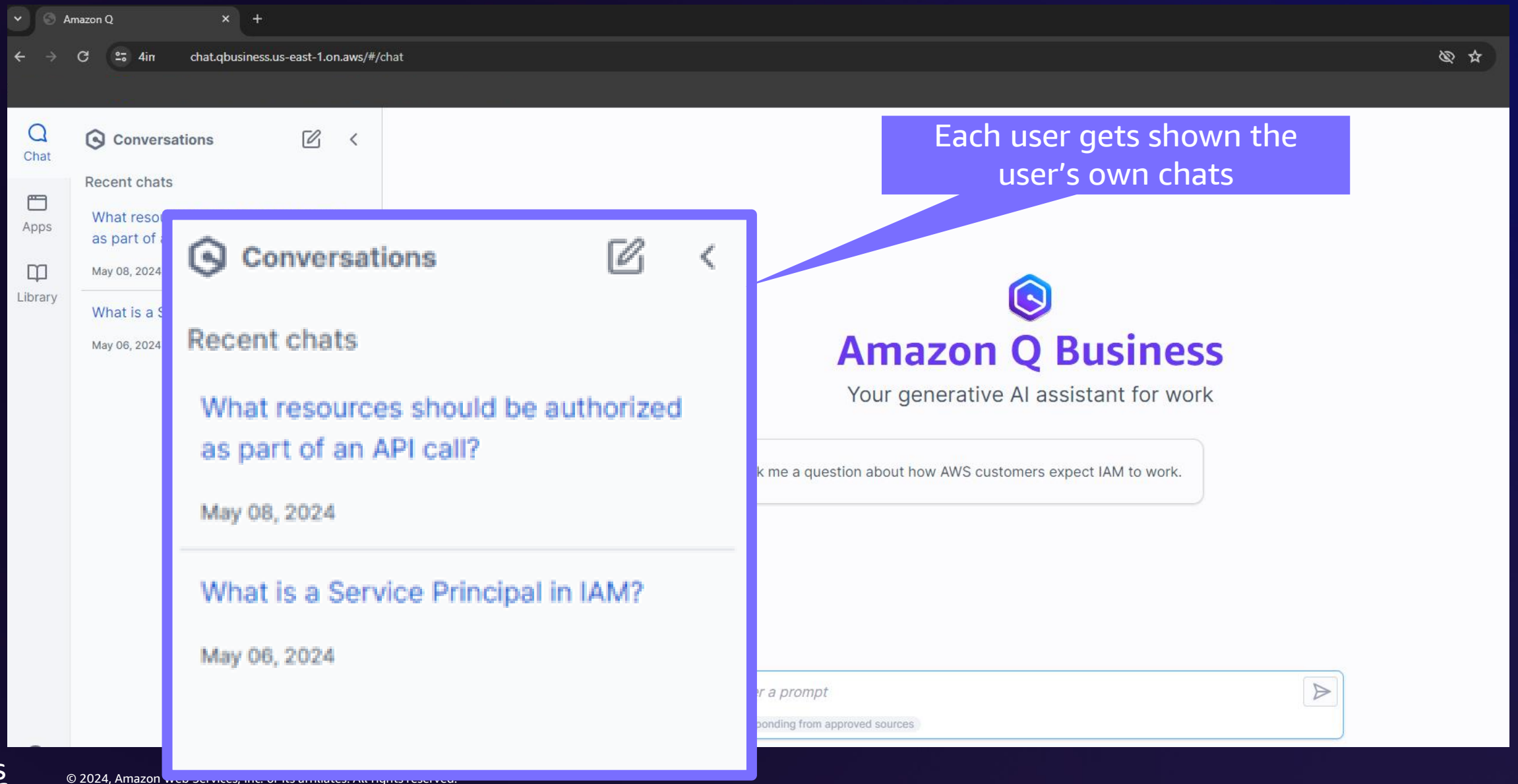
Example: Amazon Q Business application



Example: Amazon Q Business application

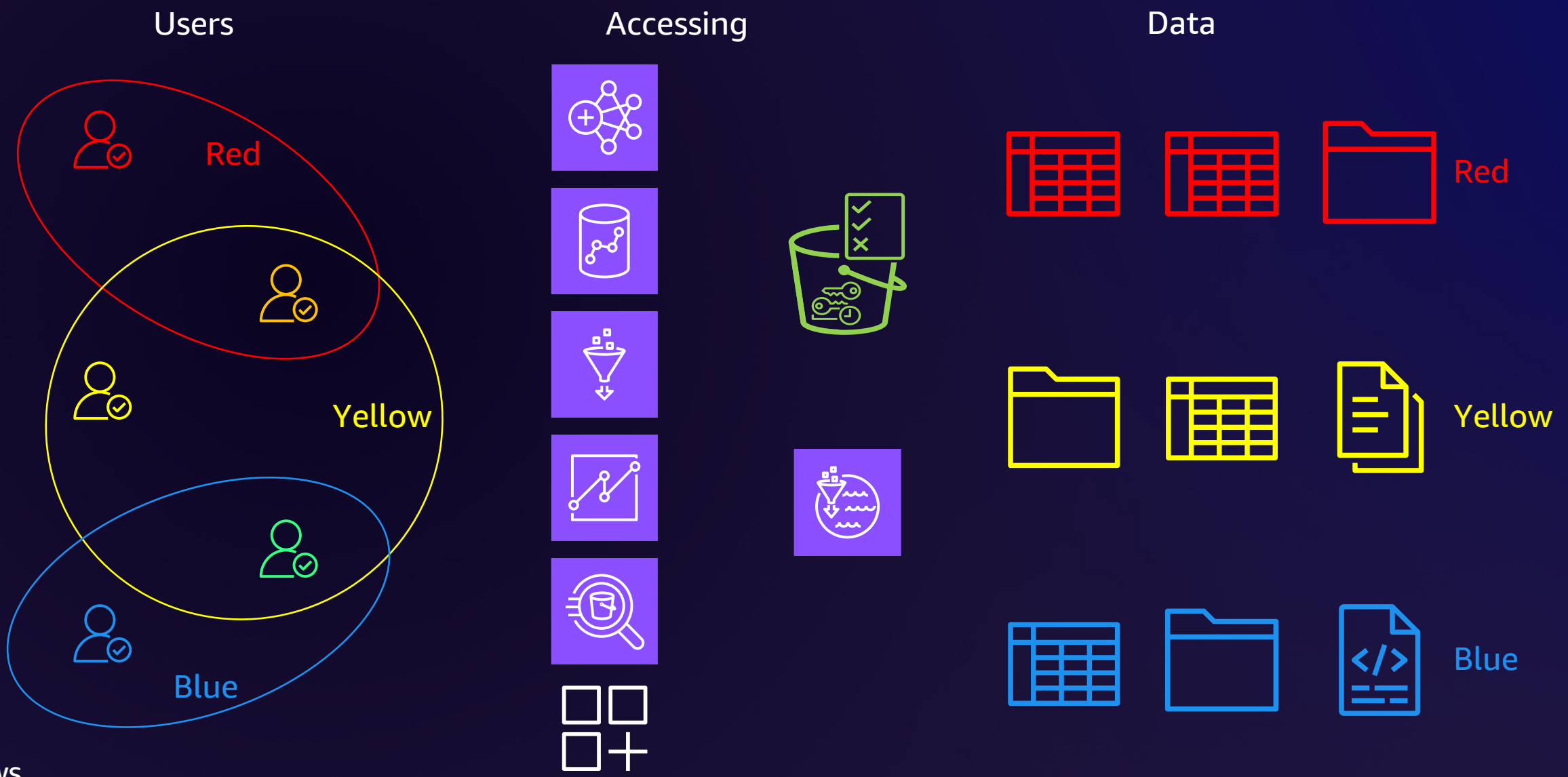


Example: Amazon Q Business application



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

Data lake built on a strong data foundation



Identity support in AWS analytics services

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

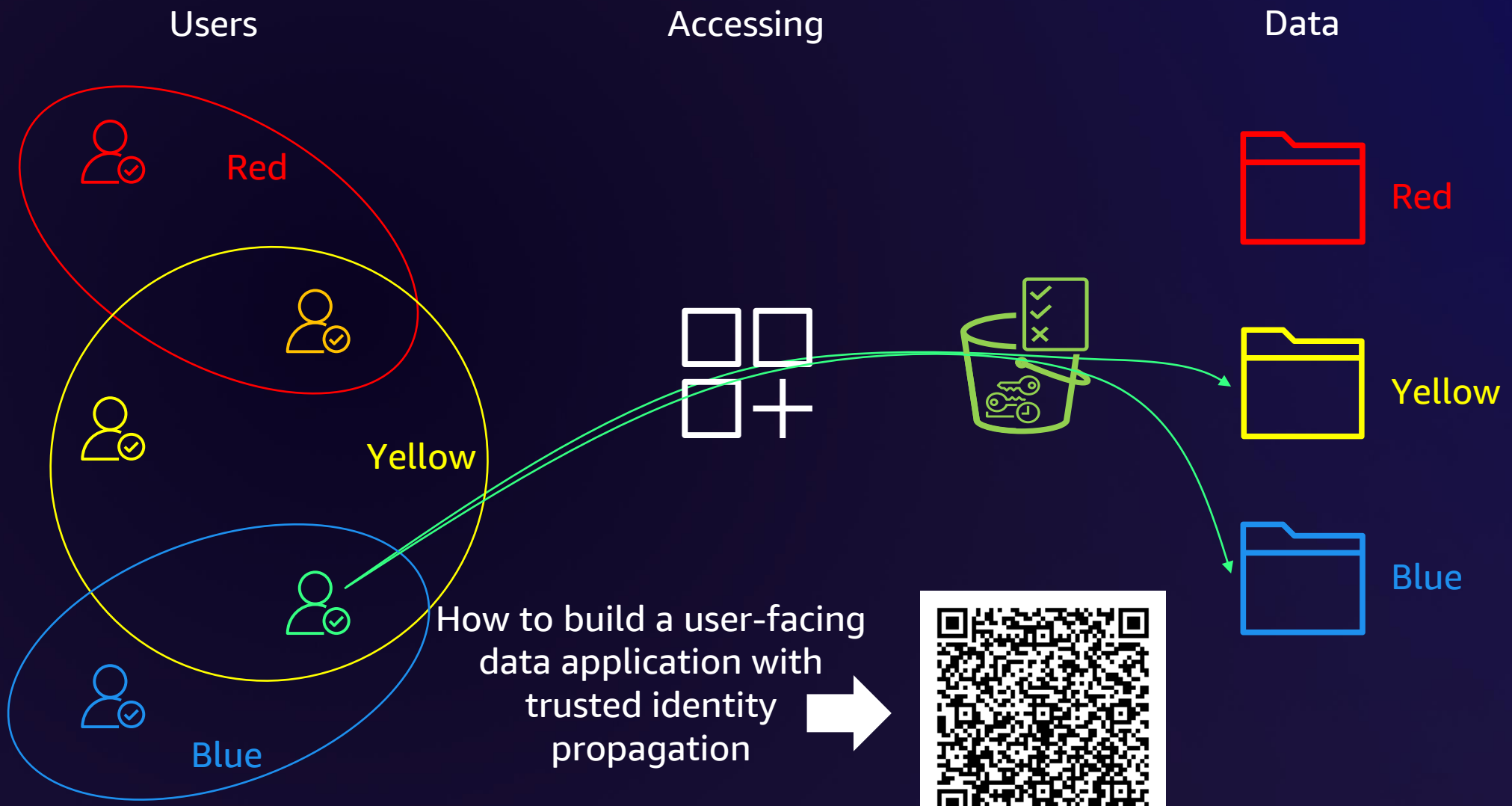
Identity support in AWS data permission services

- AWS Lake Formation
- Amazon S3 Access Grants
- Amazon Redshift

Identity support in AWS generative AI services

- Amazon Q Business
- Amazon Q Developer

Supporting identity from your own applications



How to build a user-facing data application with trusted identity propagation