
AWS
re:Invent

STG302

Data lake security in Amazon S3: Perimeters and fine-grained controls

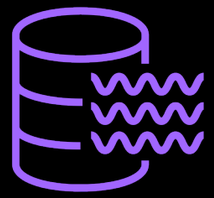
Becky Weiss
Senior Principal Engineer
AWS

Rajeev Sharma
Sr. Application Security Architect
The Vanguard Group

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

What do we mean by data lake?



Data consumers with different use cases and permissions

Analytics, ETL, etc. (AWS and custom)









Data sources











...

What do we mean by data lake?



Data consumers with different use cases and permissions

Analytics, ETL, etc. (AWS and custom)



You are here





Data sources







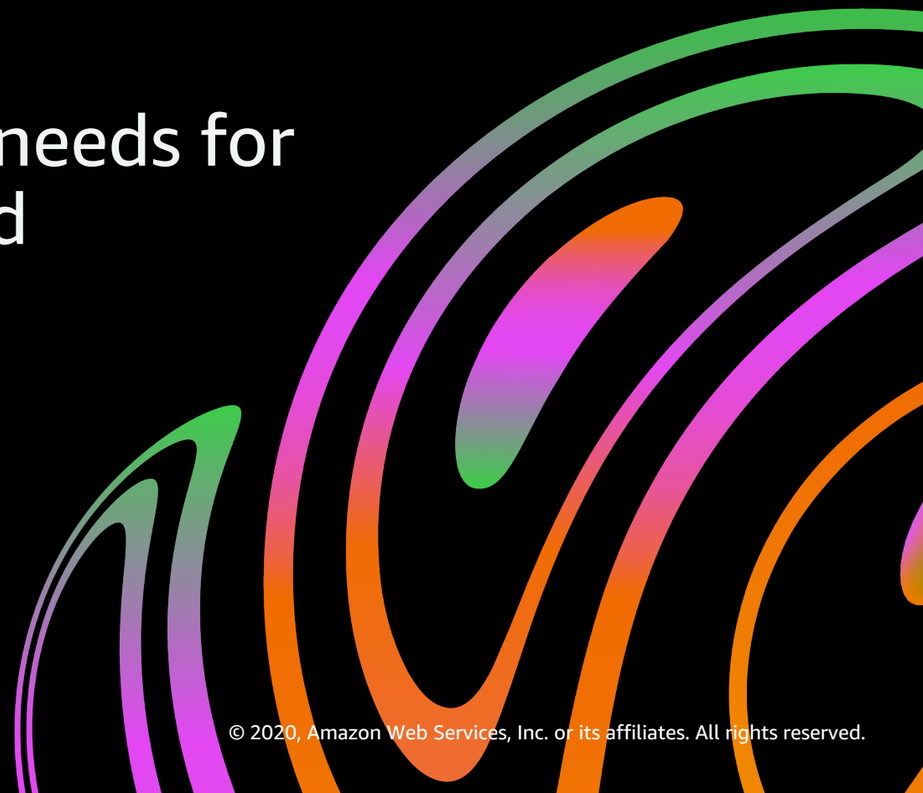




...

You're in the right place if you . . .

- Store data sets in Amazon Simple Storage Service (Amazon S3) that are large, numerous, or both
- Have multiple use cases for people, AWS services, and non-AWS applications to access the data
- Are looking for strategies that will meet your needs for both perimeter-level controls and fine-grained permissions, at scale



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

In this session

➔ Strategies, tips, and tricks for hands-on, security-conscious builders

- Securing the perimeter of your data lake
- Details on data-access authorization patterns in AWS
- Learn about the The Vanguard Group's experience in implementing data security controls in AWS



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

If you do nothing else . . .



© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

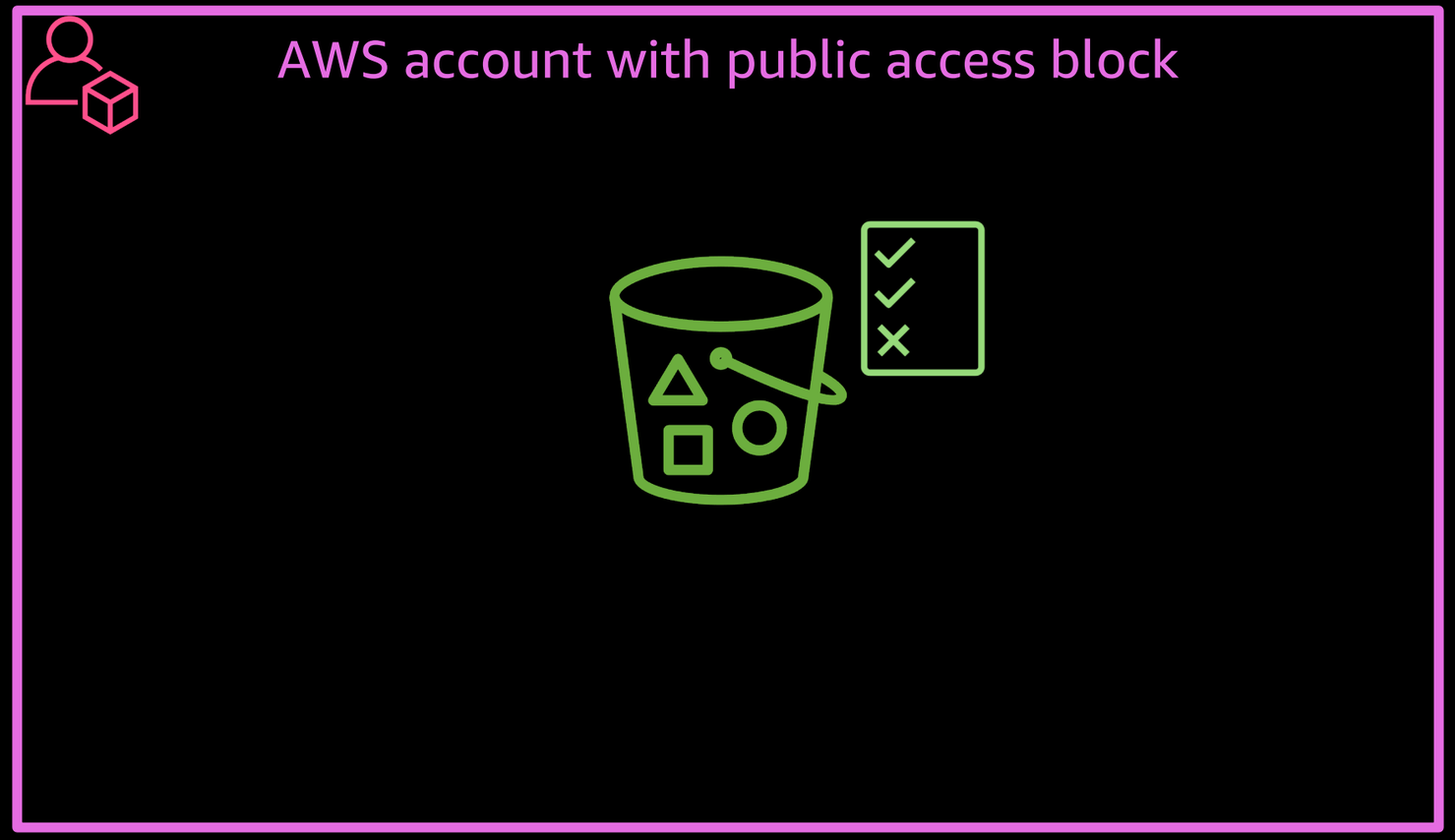
Block public access in Amazon S3: Do it now



Block public access in Amazon S3: Do it now



Block public access in Amazon S3: Do it now



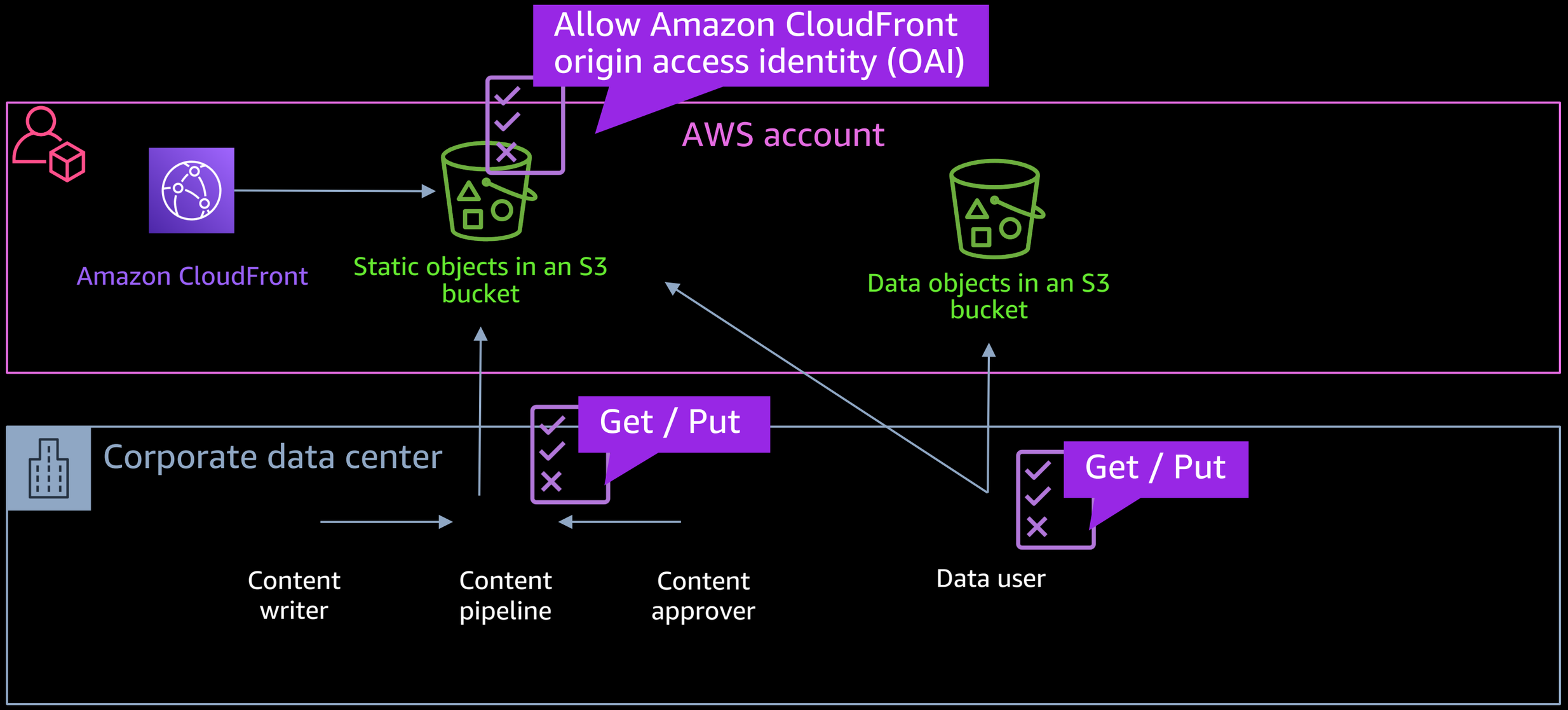
Public and private data in practice: Serving static content



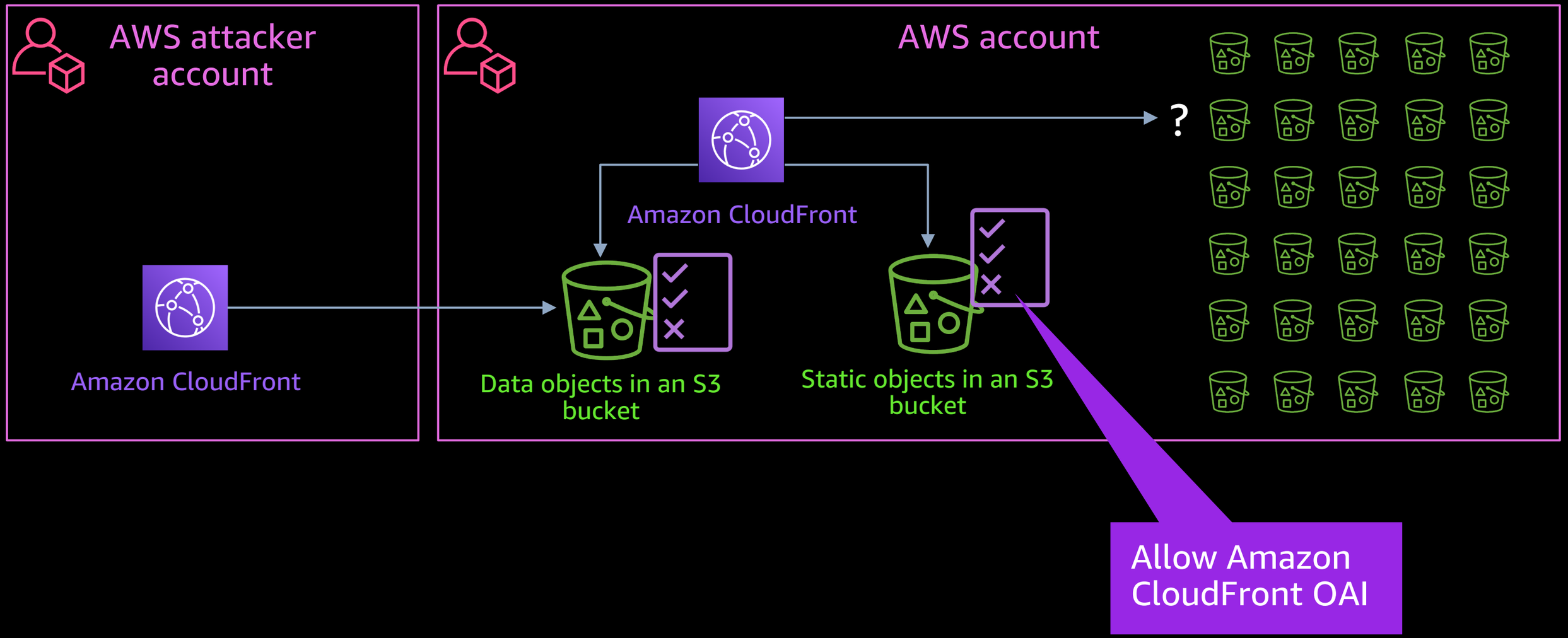
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

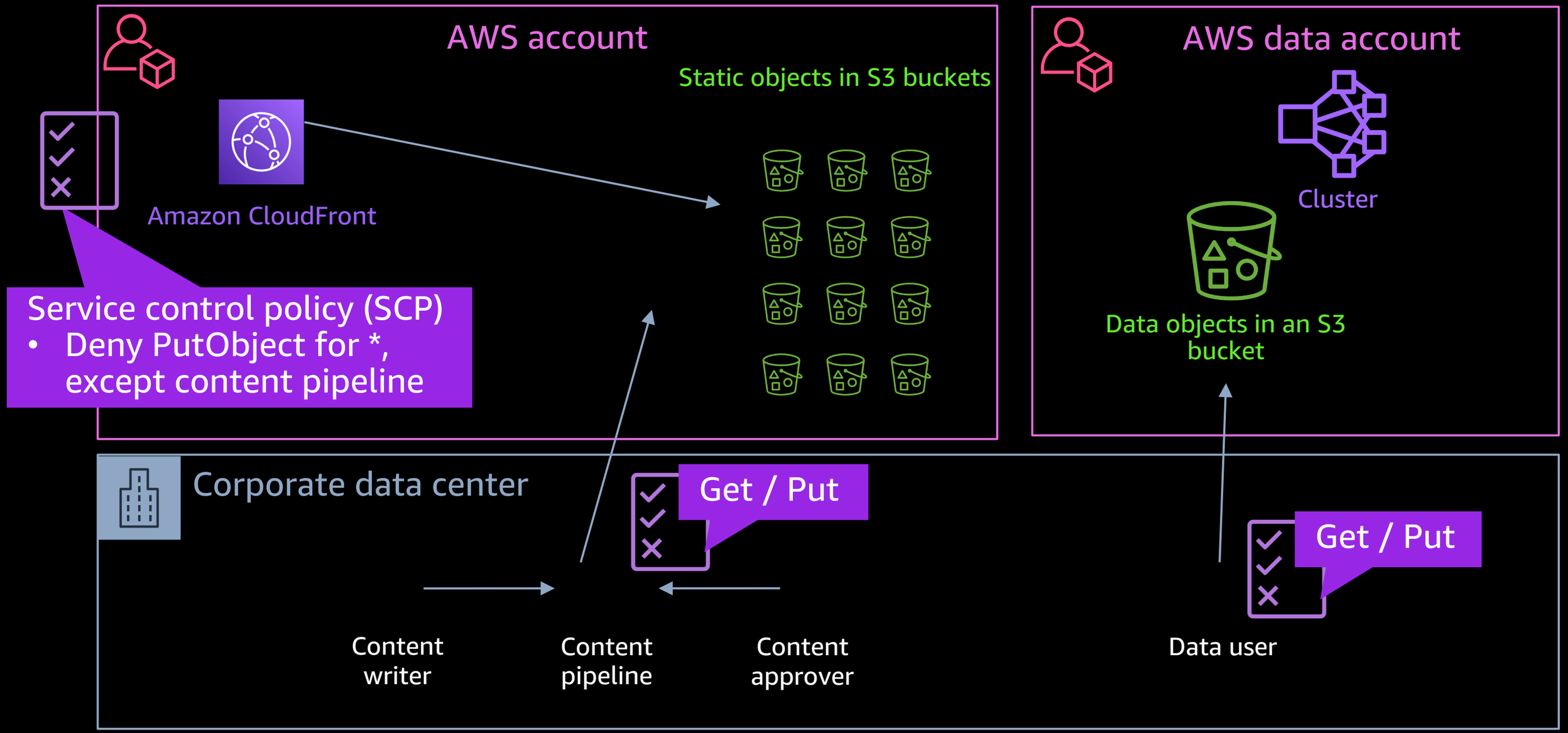
Serving static content



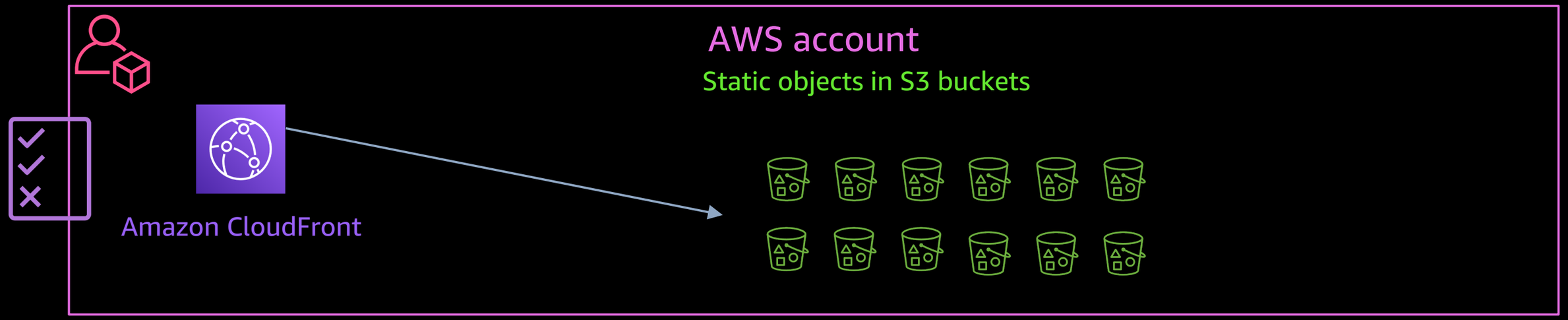
Serving static content: Scale issues



Serving static content: Scale solution

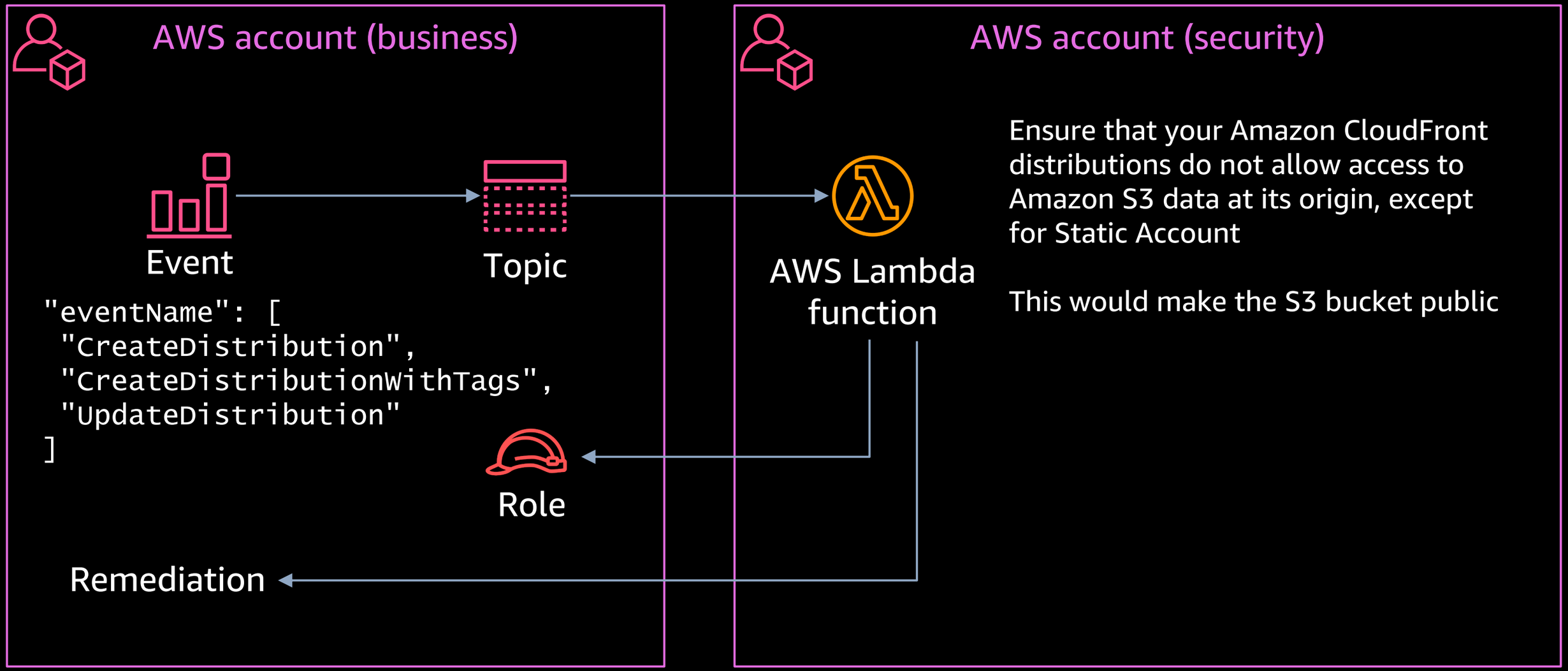


Serving static content: SCP




```
"Effect": "Deny",
"Principal": "*",
"Action": "s3:PutObject",
"Resource": "arn:aws:s3:::contentbucket/*",
"Condition": {
  "ArnNotEquals": {
    "aws:PrincipalArn": [
      "arn:aws:iam::111111111111:role/ContentPipeLine"
    ]
  }
}
```

Amazon CloudFront: Custom enforcement



Nailing the fundamentals: Establishing data perimeters for Amazon S3

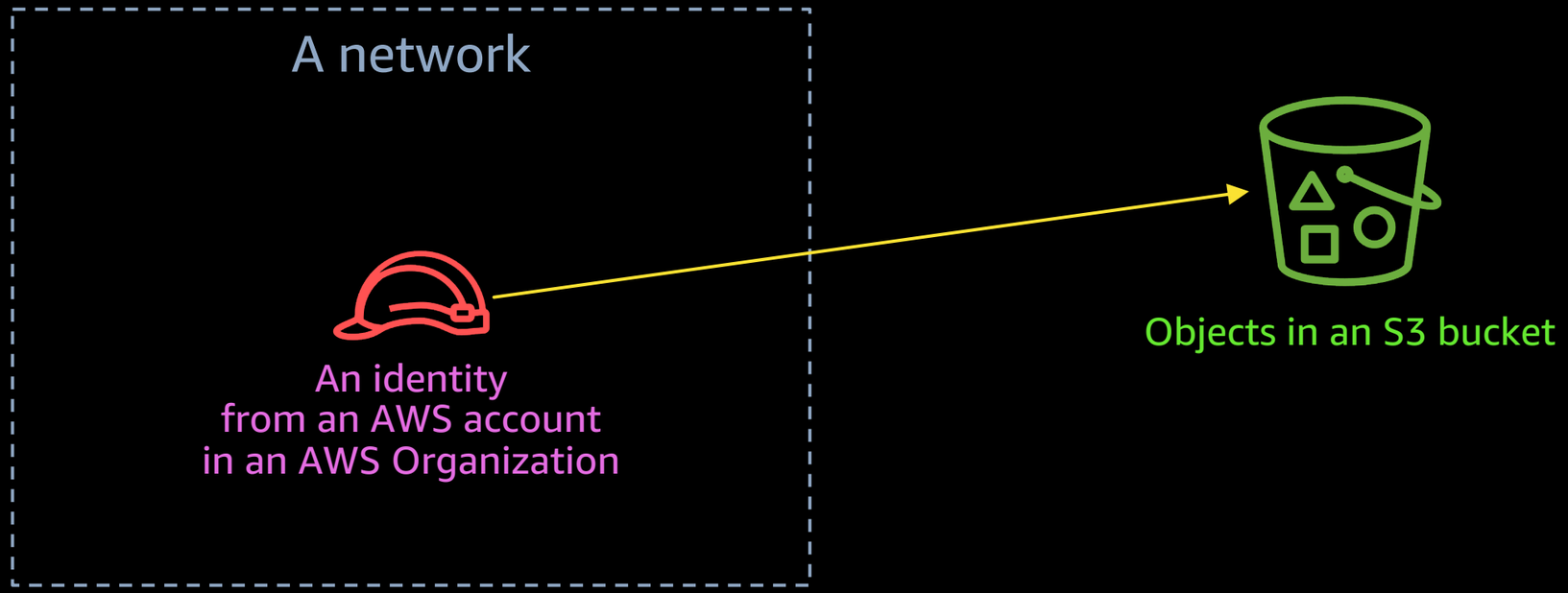


AWS re:Invent

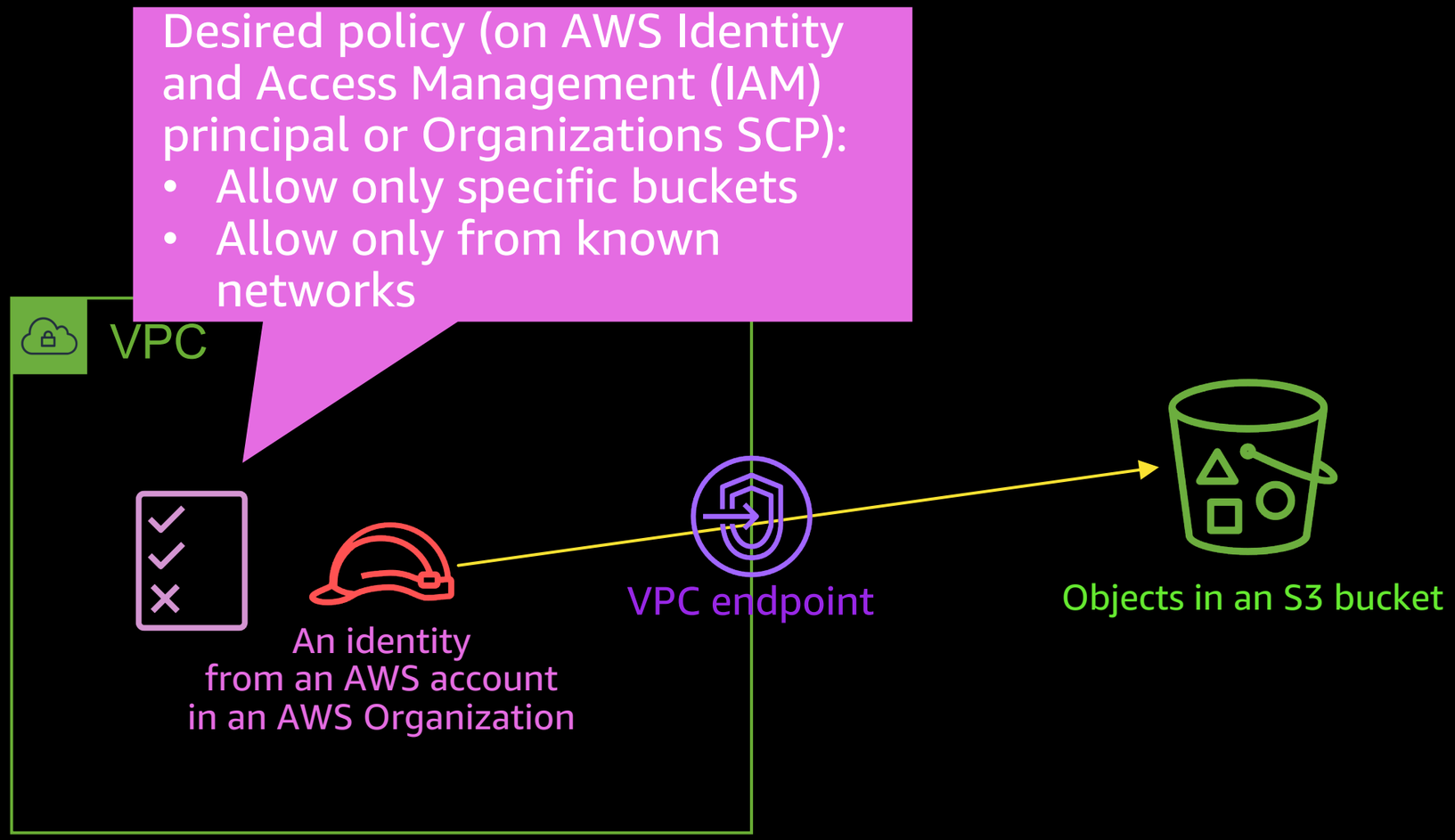
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Perimeter security: Objectives

MY IDENTITIES, ACCESSING MY DATA, FROM MY NETWORKS



Identity-based controls

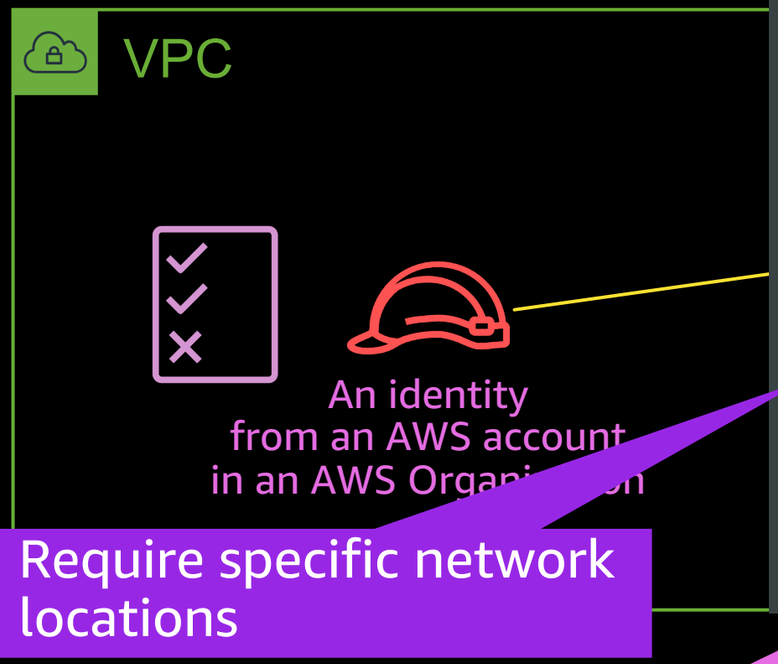


Identity-based controls: Details

IAM policy

Allow specific buckets

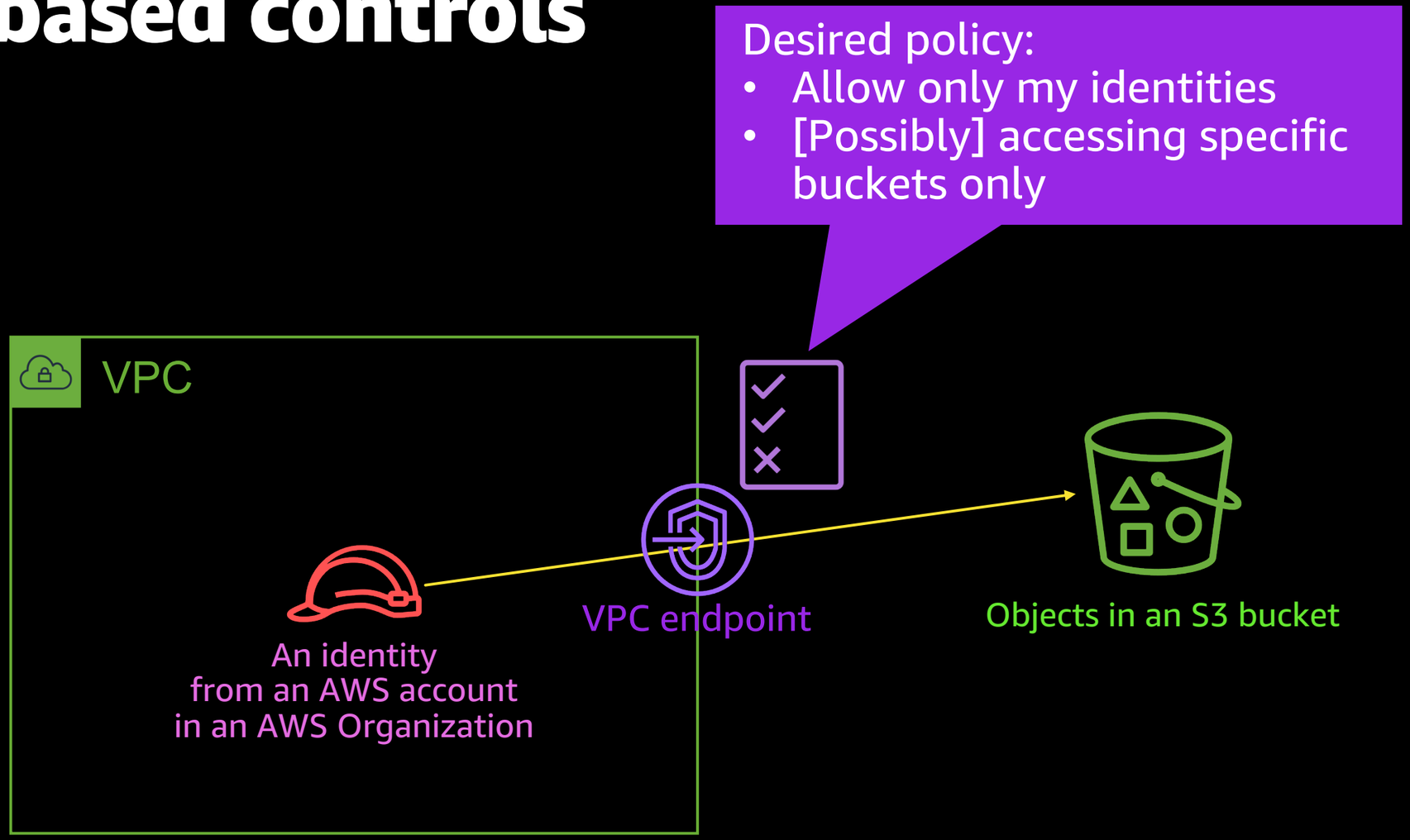
```
{
  "Sid": "AllowSpecificActionsAndBuckets",
  "Effect": "Allow",
  "Action": ["s3:GetObject", "s3:PutObject"],
  "Resource": [
    "arn:aws:s3:::example-bucket-1/*",
    "arn:aws:s3:::example-bucket-2/*"
  ]
},
{
  "Sid": "RequireExpectedNetwork",
  "Effect": "Deny",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "StringNotEqualsIfExists": {
      "aws:SourceVpc": $VPC_ID
    },
    "Bool": {
      "aws:ViaAWSService": "false"
    }
  }
}
```



Require specific network locations

Can be applied at scale using AWS Organizations Service Control Policies

Network-based controls

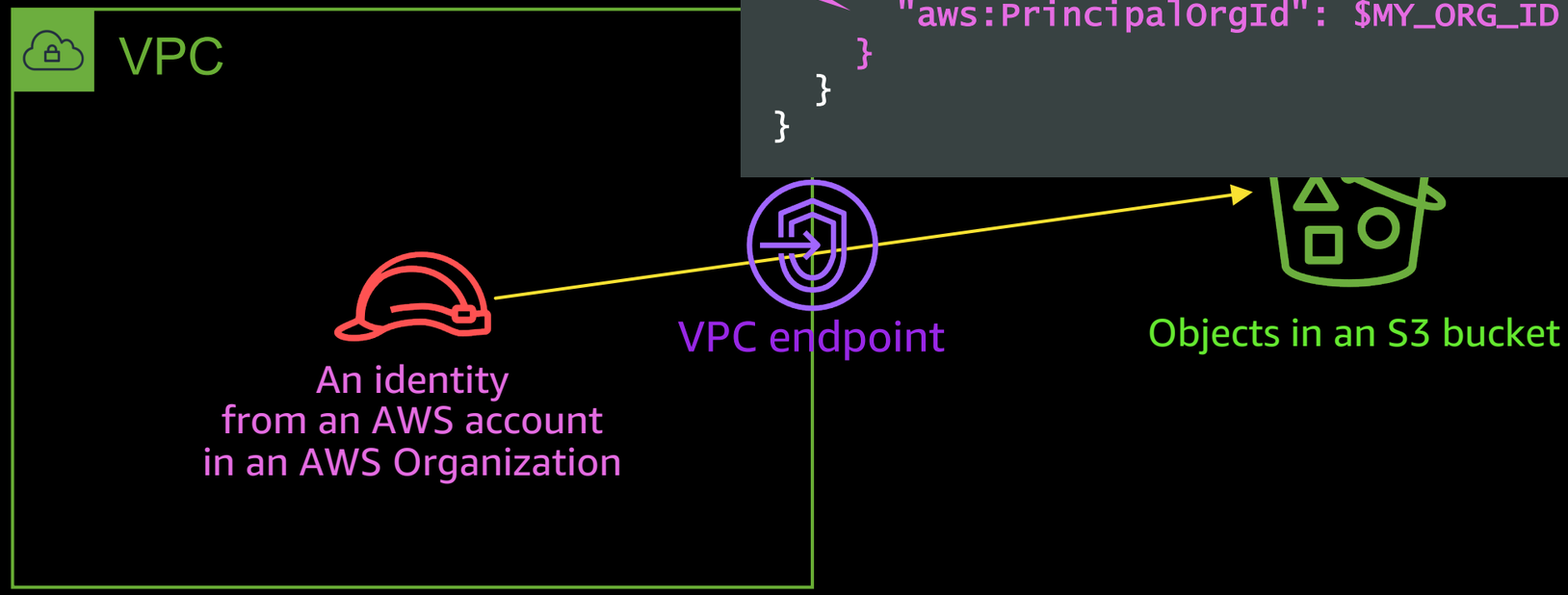


Network-based controls: Details

VPC endpoint policy

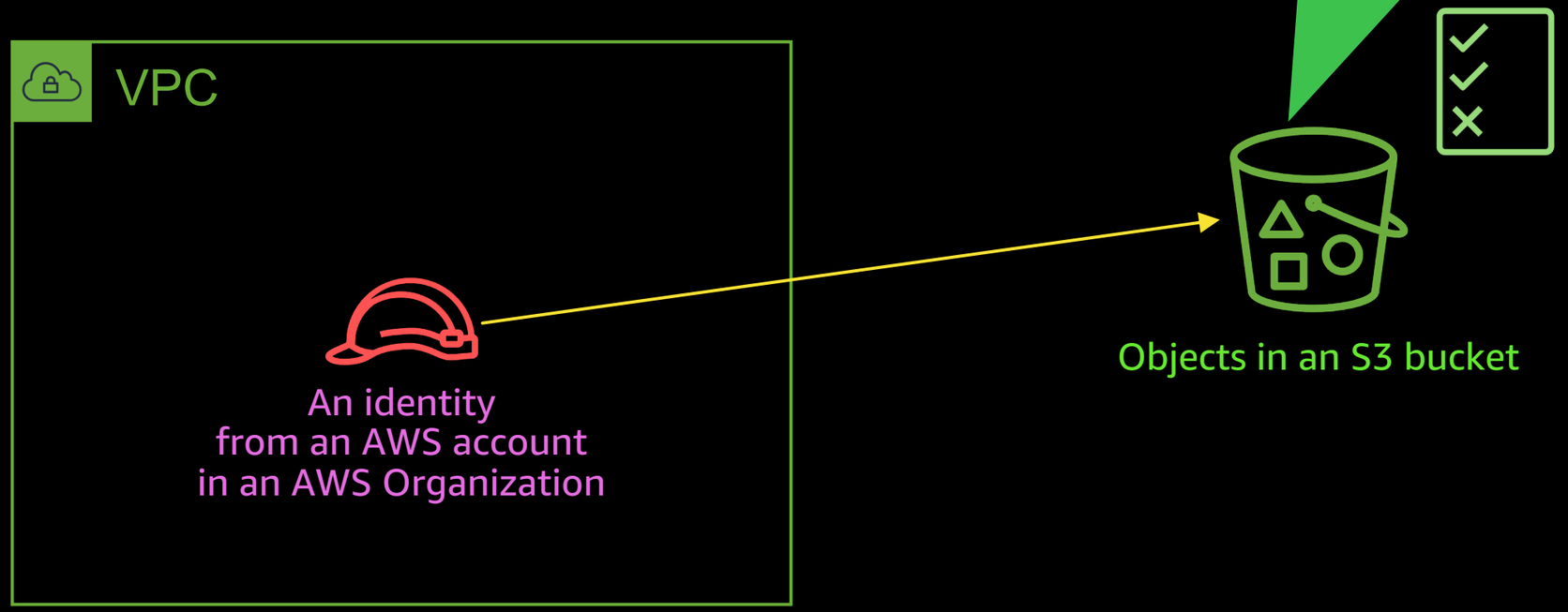
```
{
  "Sid": "AllowMyOrganization",
  "Effect": "Allow",
  "Principal": "*",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "StringEquals": {
      "aws:PrincipalOrgId": $MY_ORG_ID
    }
  }
}
```

Must be my Organization in order for authorization to continue



Resource-based controls

Desired policy:
- Allow only my identities
- From my networks only

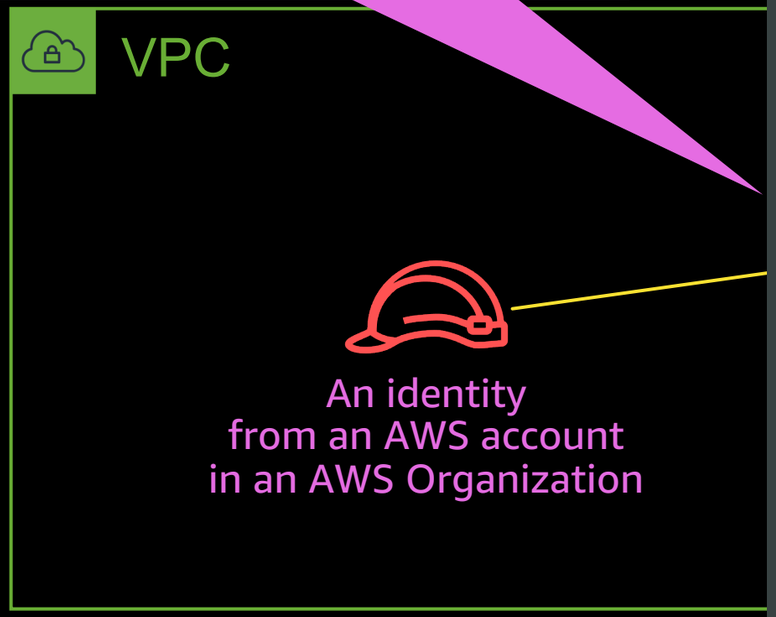


Resource-based controls: Details

S3 bucket policy

Allow specific accounts

Allow specific AWS services, if applicable



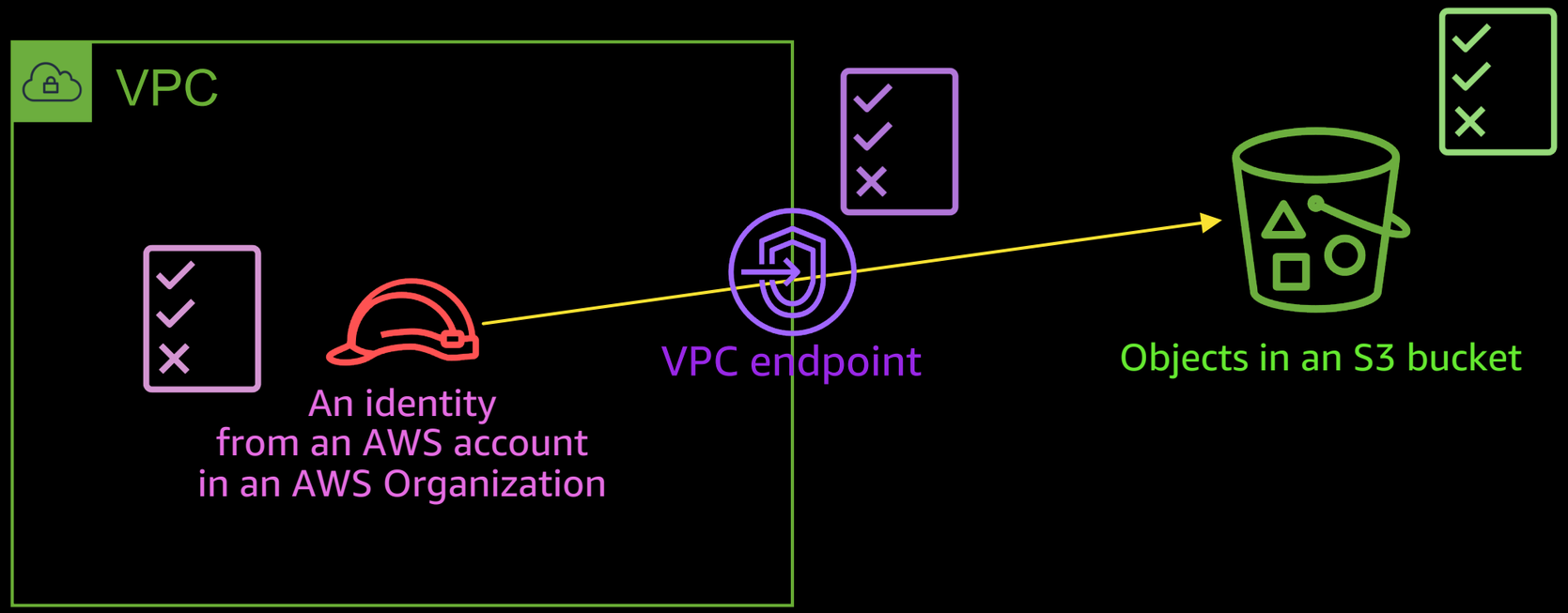
Require specific network locations



```
{
  "Sid": "AllowSpecificAccounts",
  "Effect": "Allow",
  "Principal": {
    "AWS": [ $ACCT_1, $ACCT_2 ]
  },
  "Action": ["s3:GetObject", "s3:PutObject"],
  "Resource": "arn:aws:s3:::example-bucket/*"
},
{
  "Sid": "AllowSpecificAWSServices",
  "Effect": "Allow",
  "Principal": {
    "Service": "someservice.amazonaws.com"
  },
  "Action": "s3:PutObject",
  "Resource": "arn:aws:s3:::example-bucket/*"
},
{
  "Sid": "DenyReadFromWrongNetwork",
  "Effect": "Deny",
  "Principal": "*",
  "Action": ["s3:GetObject"],
  "Resource": "arn:aws:s3:::example-bucket/*",
  "Condition": {
    "StringNotEqualsIfExists": {
      "aws:SourceVpc": $VPC_ID
    },
    "Bool": {
      "aws:ViaAWSService": "false"
```

Perimeter security: Summary

MY IDENTITIES, ACCESSING MY DATA, FROM MY NETWORKS



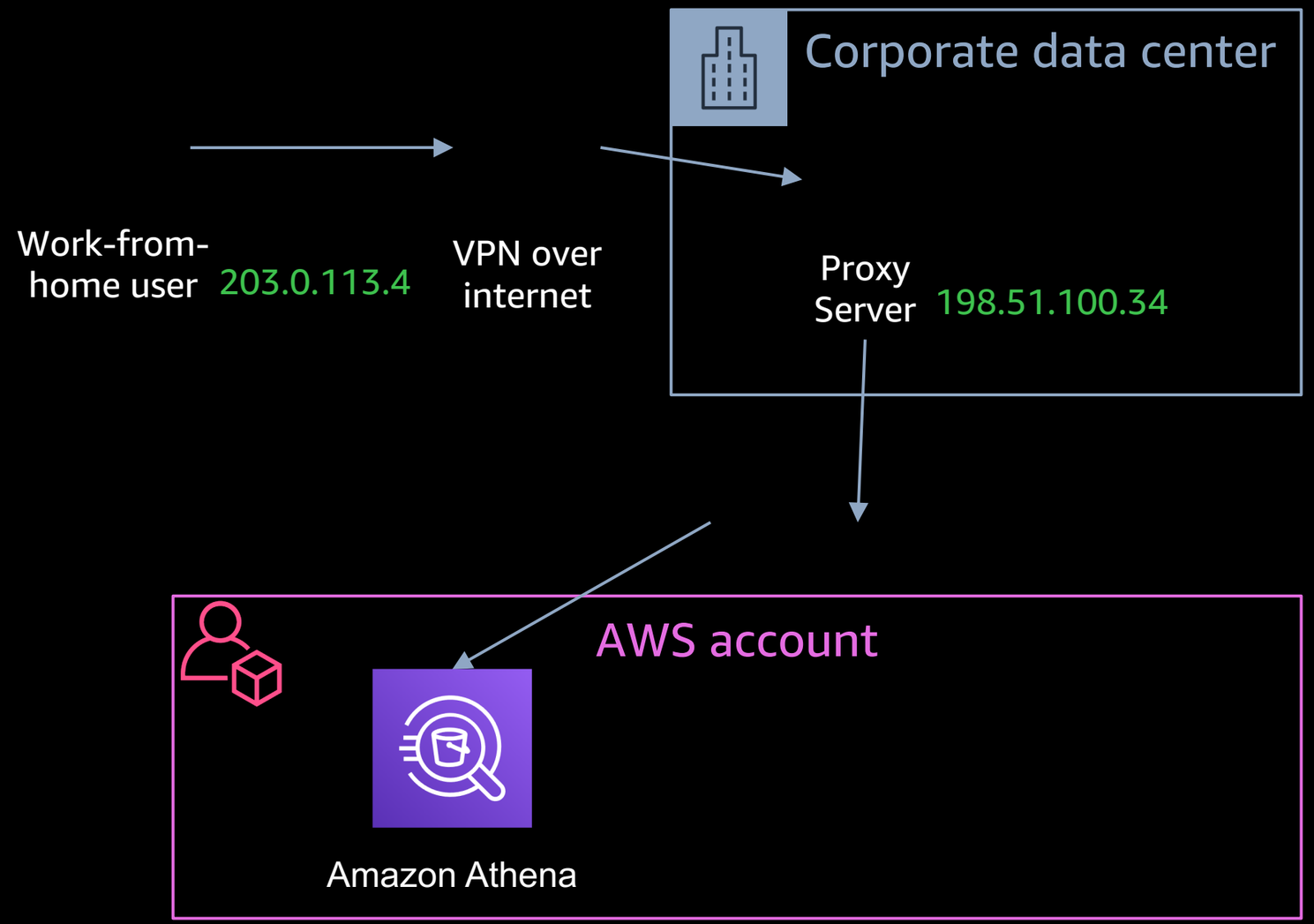
Perimeter security in practice: Work-from-home scenario



AWS re:Invent

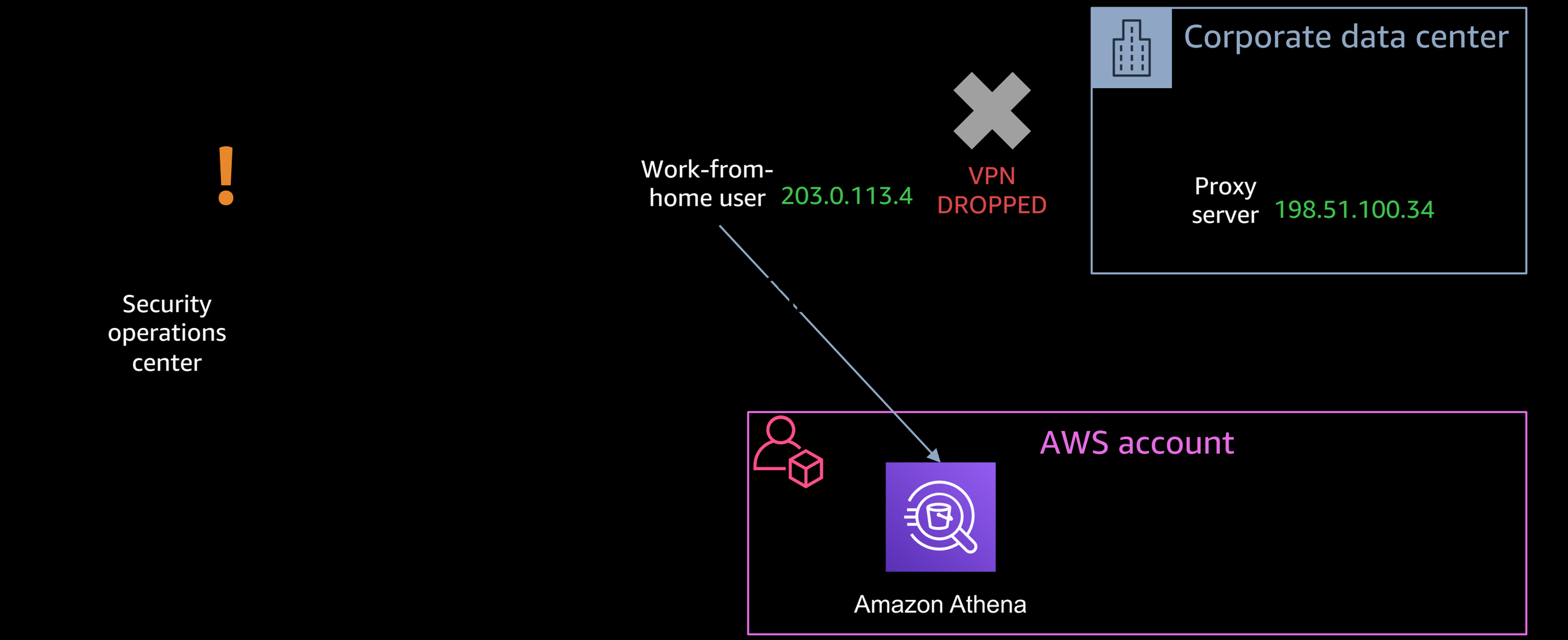
© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Work-from-home scenario



Work-from-home scenario

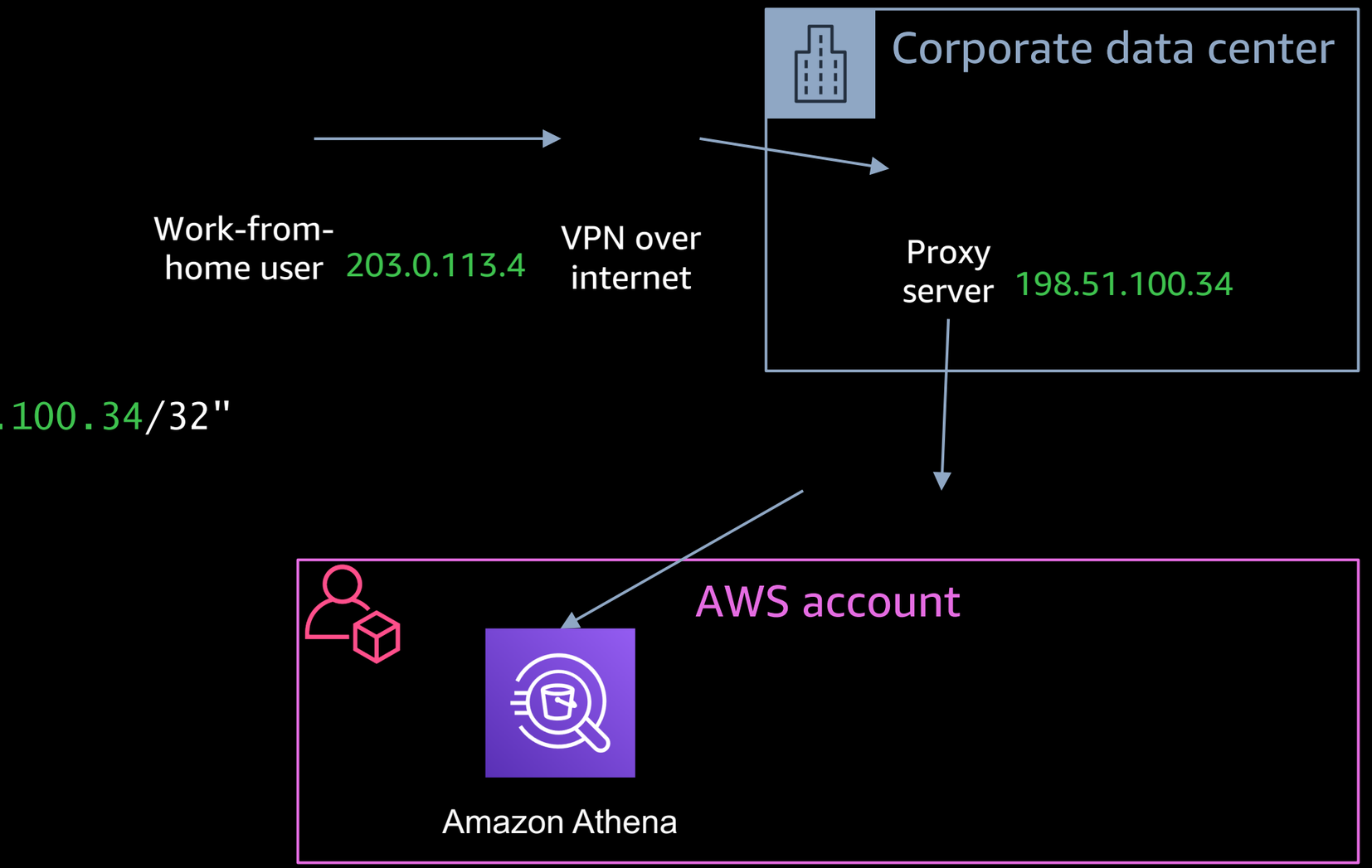
DROPPED VPN APPEARS LIKE A STOLEN CREDENTIAL



Boundary policy IP address

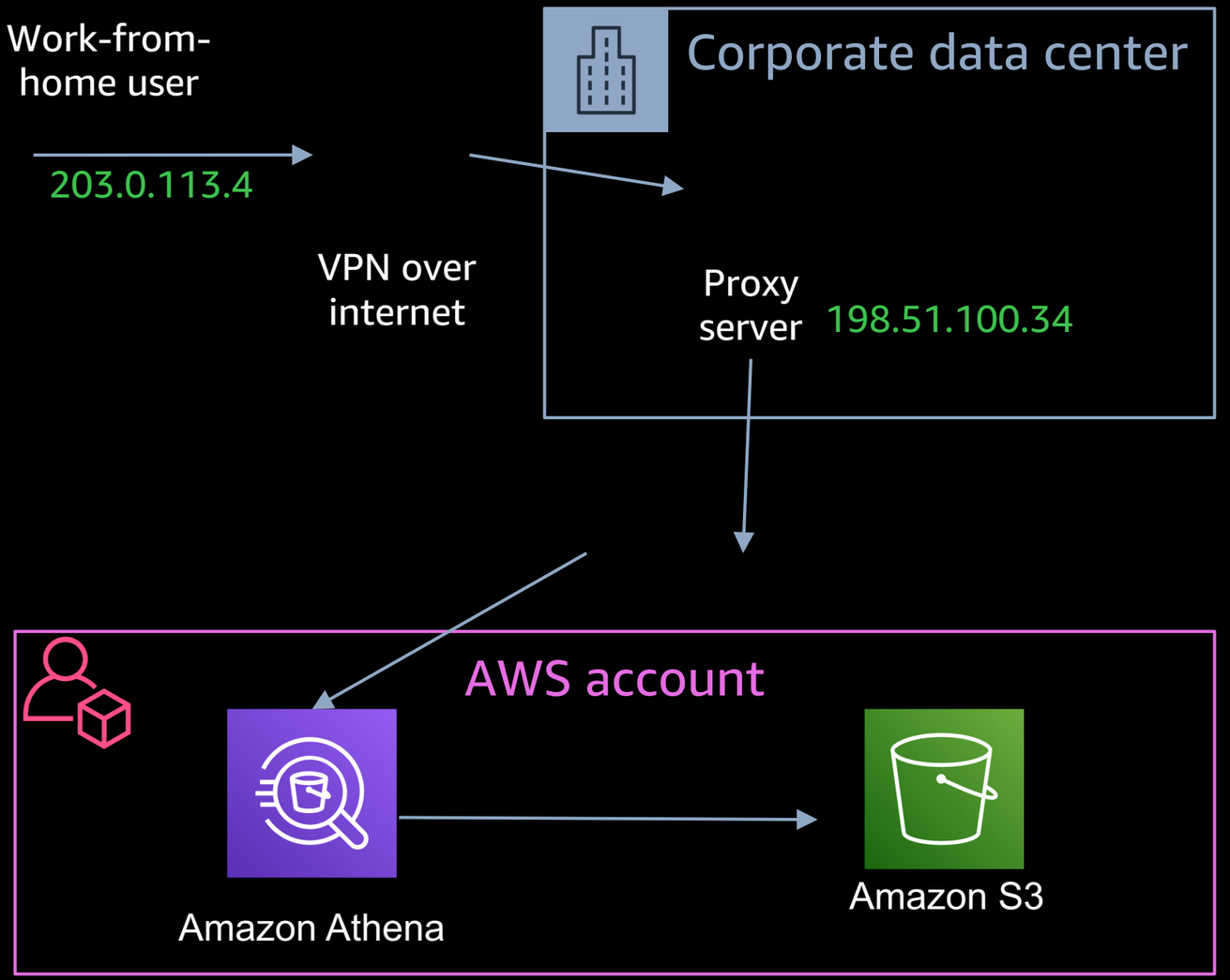
LIMITS STS TOKENS GENERATED BY AWS TO ONLY VANGUARD'S NETWORK

```
// IAM Permission Boundary
{
  "Effect": "Allow",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "IpAddress": {
      "aws:SourceIp": "198.51.100.34/32"
    }
  }
}
```



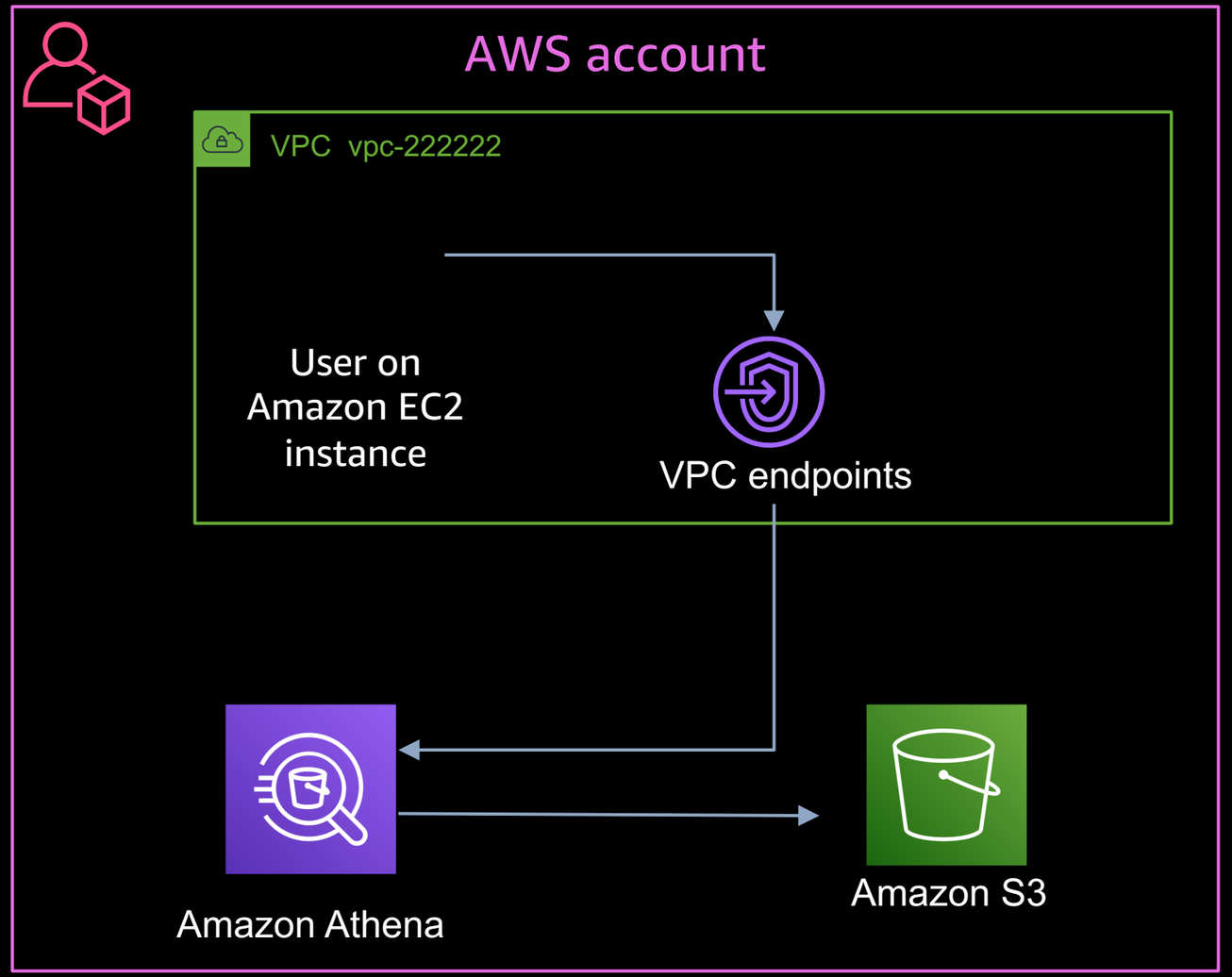
Boundary policy IP address + ViaAWSService

```
// IAM Permission Boundary
{
  "Effect": "Allow",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "IpAddress": {
      "aws:SourceIp": "198.51.100.34/32"
    },
    "Bool": {
        "aws:ViaAWSService": "false"
    }
  },{
    "Effect": "Allow",
    "Action": "*",
    "Resource": "*",
    "Condition": {
      "Bool": {
        "aws:ViaAWSService": "true"
      }
    }
  }
```

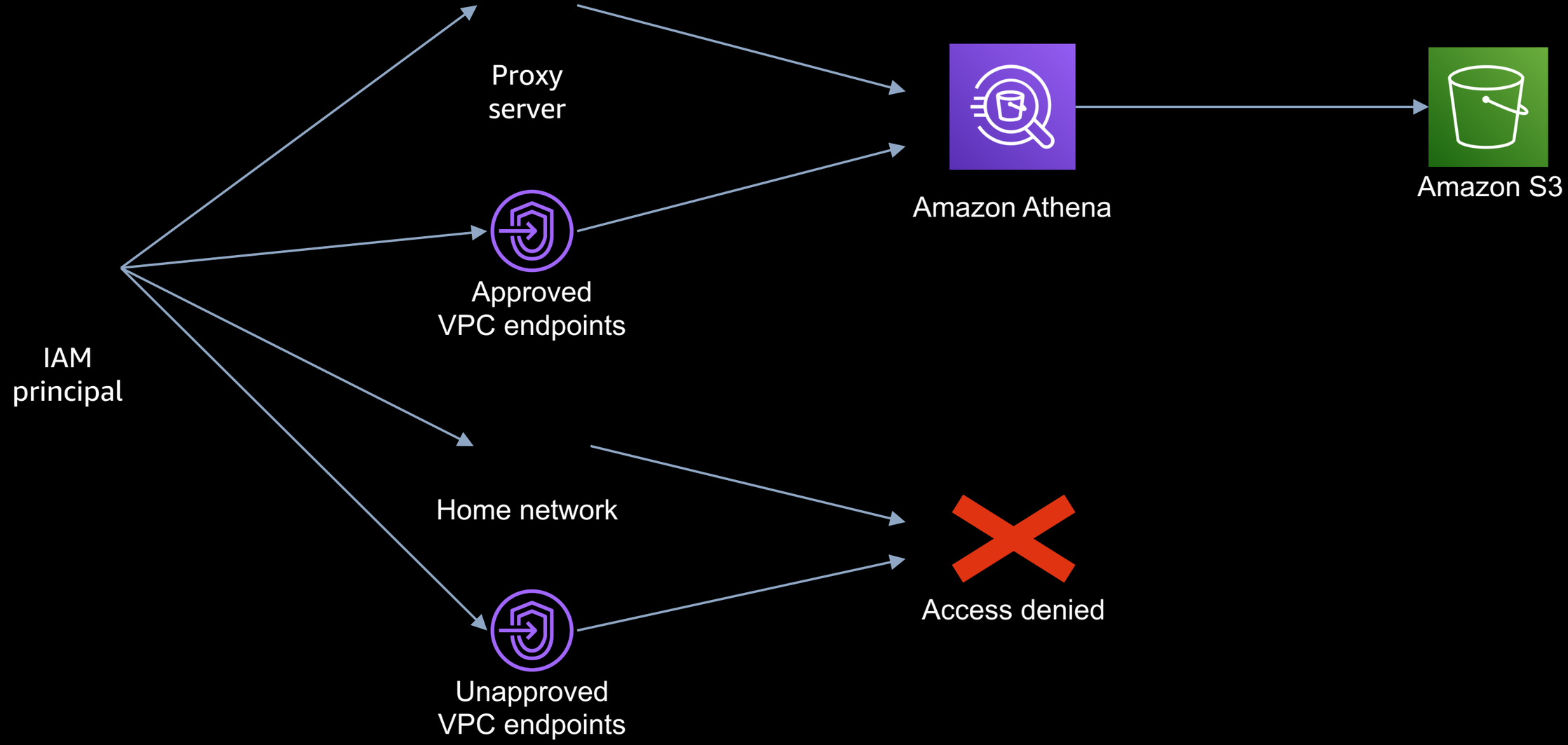


Boundary policy VPC + ViaAWSService

```
// IAM Permission Boundary
{
  "Effect": "Allow",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "StringEquals": {
      "aws:SourceVpc": "vpc-222222"
    },
    "Bool": {
      "aws:ViaAWSService": "false"
    }
},
{
  "Effect": "Allow",
  "Action": "*",
  "Resource": "*",
  "Condition": {
    "Bool": {
      "aws:ViaAWSService": "true"
    }
  }
}
```



Boundary policy summary



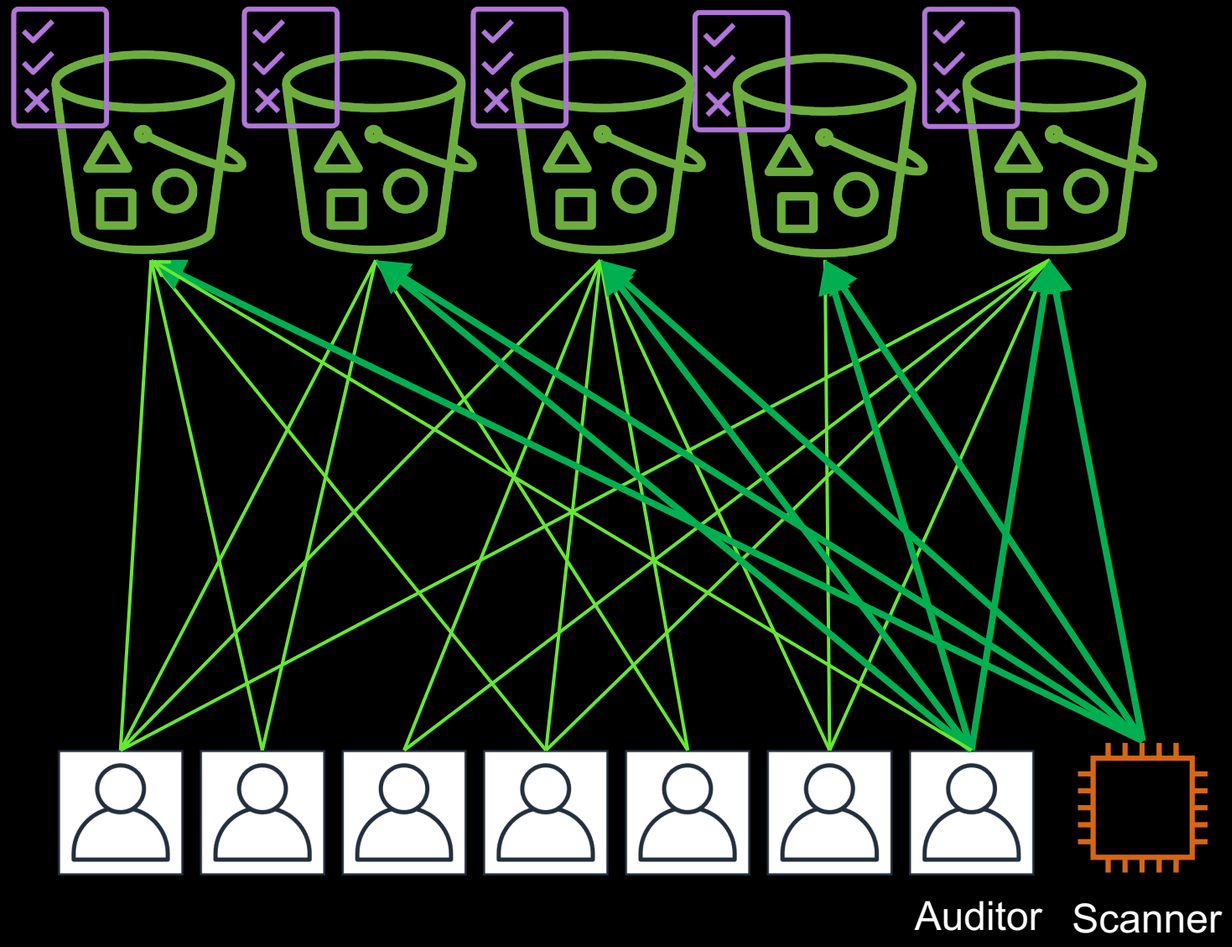
Managing least-privilege access to data in practice



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resource policy scaling issues: Least privilege



A new principal may need access to several buckets

This would require multiple bucket policy updates

Pathed IAM roles



```
arn:aws:iam::111111111111:role/DEVELOPER

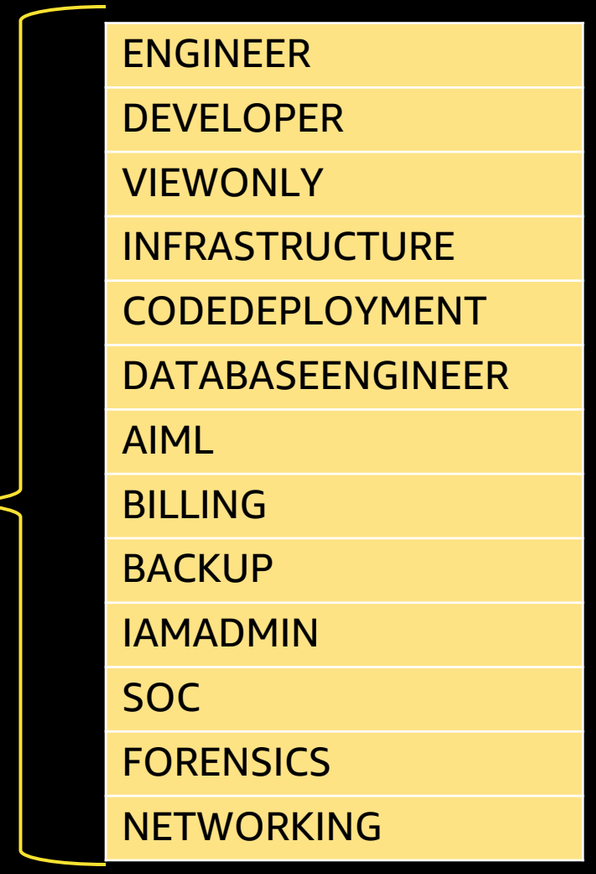
arn:aws:iam::111111111111:role/universal/DEVELOPER
```

Traditional IAM roles: Resource policy

```
"Sid": "RestrictAccessResourcePolicy",
"Effect": "Deny", "Principal": "*", "Action": "s3:*",
"Resource": "arn:aws:s3:::exampleBucket/*",
"Condition": {
    "ForAnyValue:StringNotLike": {
      "aws:PrincipalArn": [
      "arn:aws:iam::111111111111:role/DEVELOPER",
      "arn:aws:iam::111111111111:role/OPERATIONS",
      "arn:aws:iam::111111111111:role/AUDIT",
      "arn:aws:iam::111111111111:role/SCANNER",
      "arn:aws:iam::111111111111:role/GOVASCODE",
      "arn:aws:iam::111111111111:role/FORENSICS",
      "arn:aws:iam::111111111111:role/exampleIProfile1",
      "arn:aws:iam::111111111111:role/aBusinessRole",
      "arn:aws:iam::111111111111:role/……"
      ]
}
```

Pathed IAM roles: Resource policy

```
"Sid": "RestrictAccessResourcePolicy",
"Effect": "Deny",
"Principal": "*",
"Action": "s3:*",
"Resource": "arn:aws:s3:::exampleBucket/*",
"Condition": {
  "ArnNotLike": {
    "aws:PrincipalArn": [
      "arn:aws:iam::111111111111:role/universal/*",
      "arn:aws:iam::111111111111:role/exampleIProfile1"
      ,
      "arn:aws:iam::111111111111:role/aBusinessRole",
      "arn:aws:iam::111111111111:role/……"
    ]
  }
}
```



Pathed IAM roles: Protection with SCP

```
{
    "Sid": "RestrictUniversalAccessPathSCP",
    "Effect": "Deny",
    "Action": "iam:*",
    "Resource": "arn:aws:iam::*:role/universal/*",
    "Condition": {
        "ArnNotLike": {
            "aws:PrincipalArn":
"arn:aws:iam::*:role/IAMENGINEERADMIN"
        }
    }
}
```

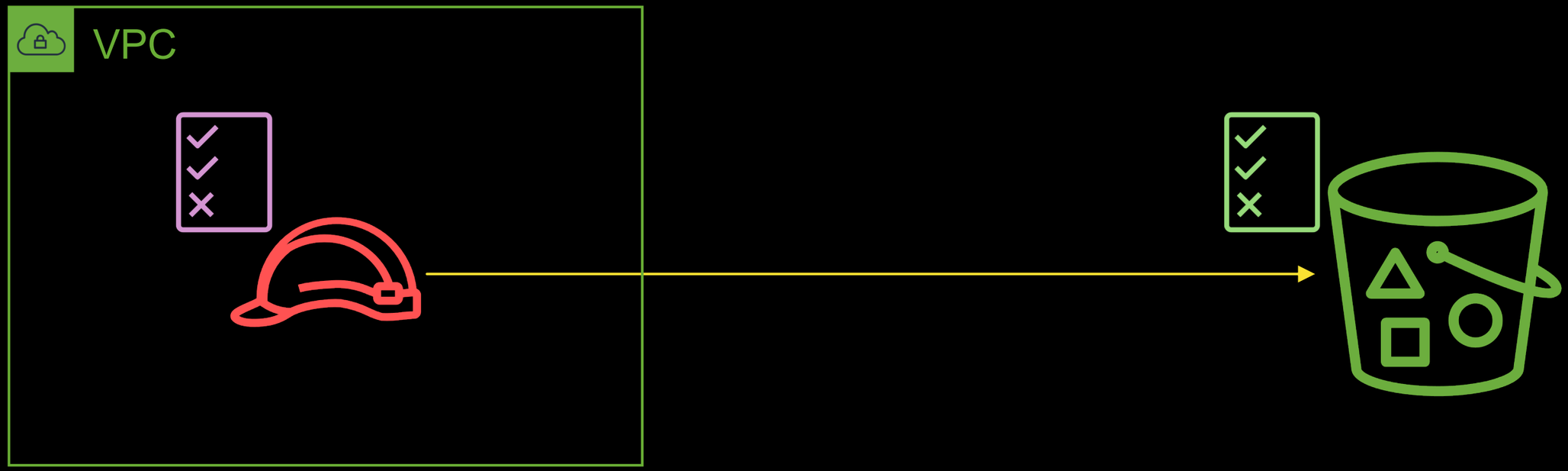
Data access authorization in AWS: Principles and patterns



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Direct access of your identities to data



Direct access to data by way of an AWS service



AWS services with direct access to your buckets

```
// … other policy statements …
{
  "Effect": "Allow",
    "Principal": {
      "Service": "cloudtrail.amazonaws.com"
    },
    "Action": "s3:PutObject",
    "Resource": "arn:aws:s3:::example-bucket/AWSLogs/111122223333/*",
      "Condition": {
        "StringEquals": {
          "s3:x-amz-acl": "bucket-owner-full-control"
```

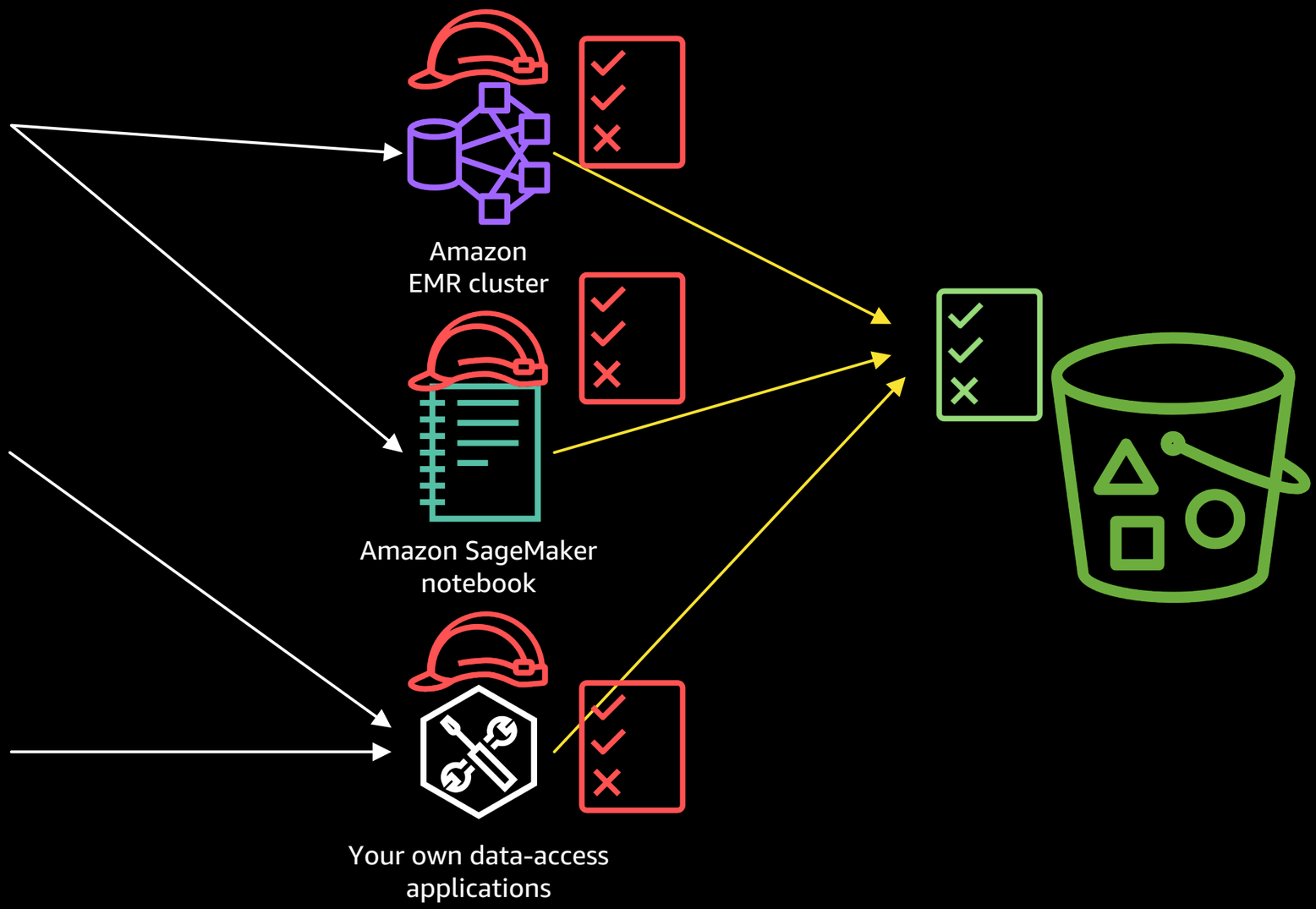


AWS CloudTrail



AWS CloudTrail
logs bucket

Intermediate IAM roles for data access



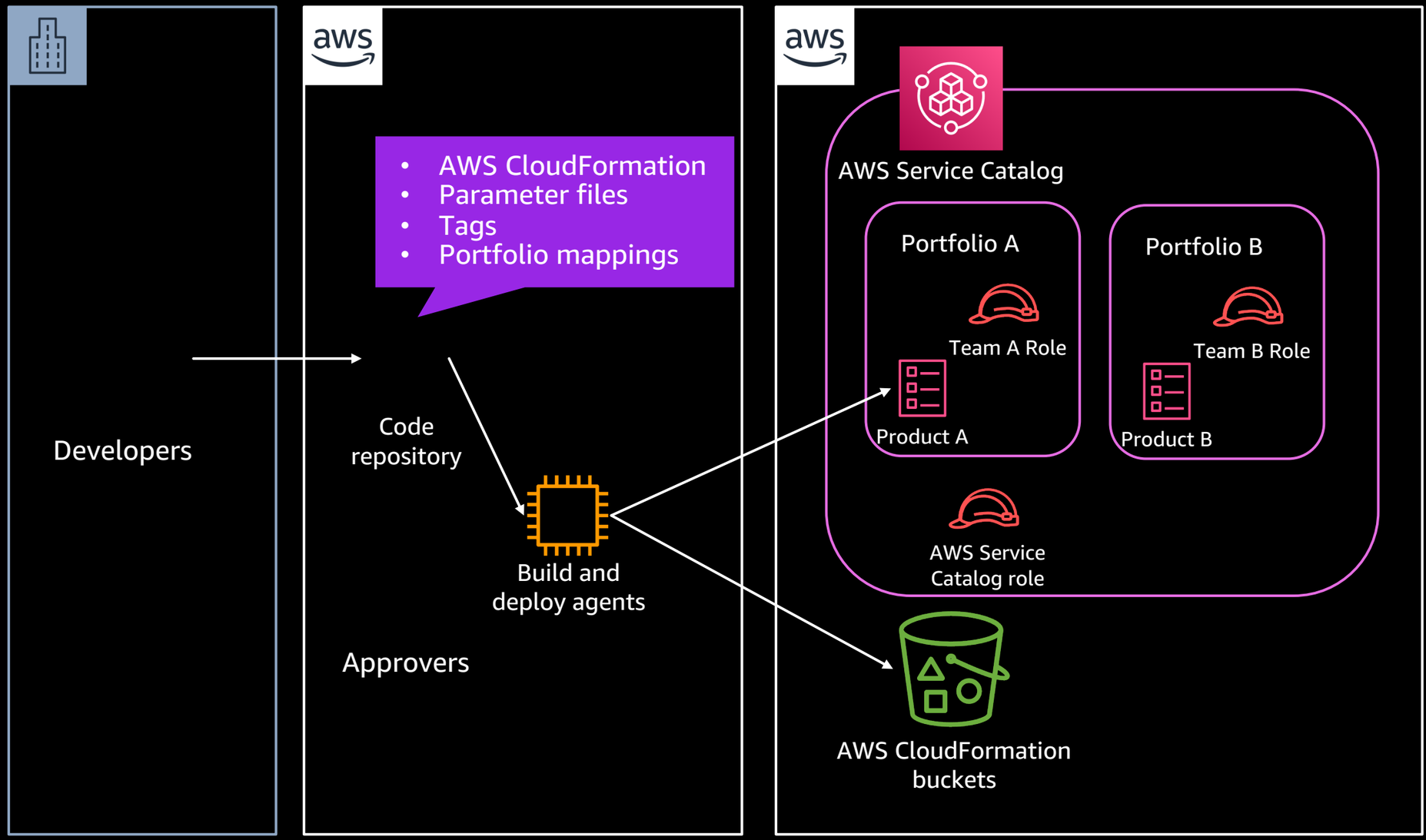
Managing AWS service environment access in practice



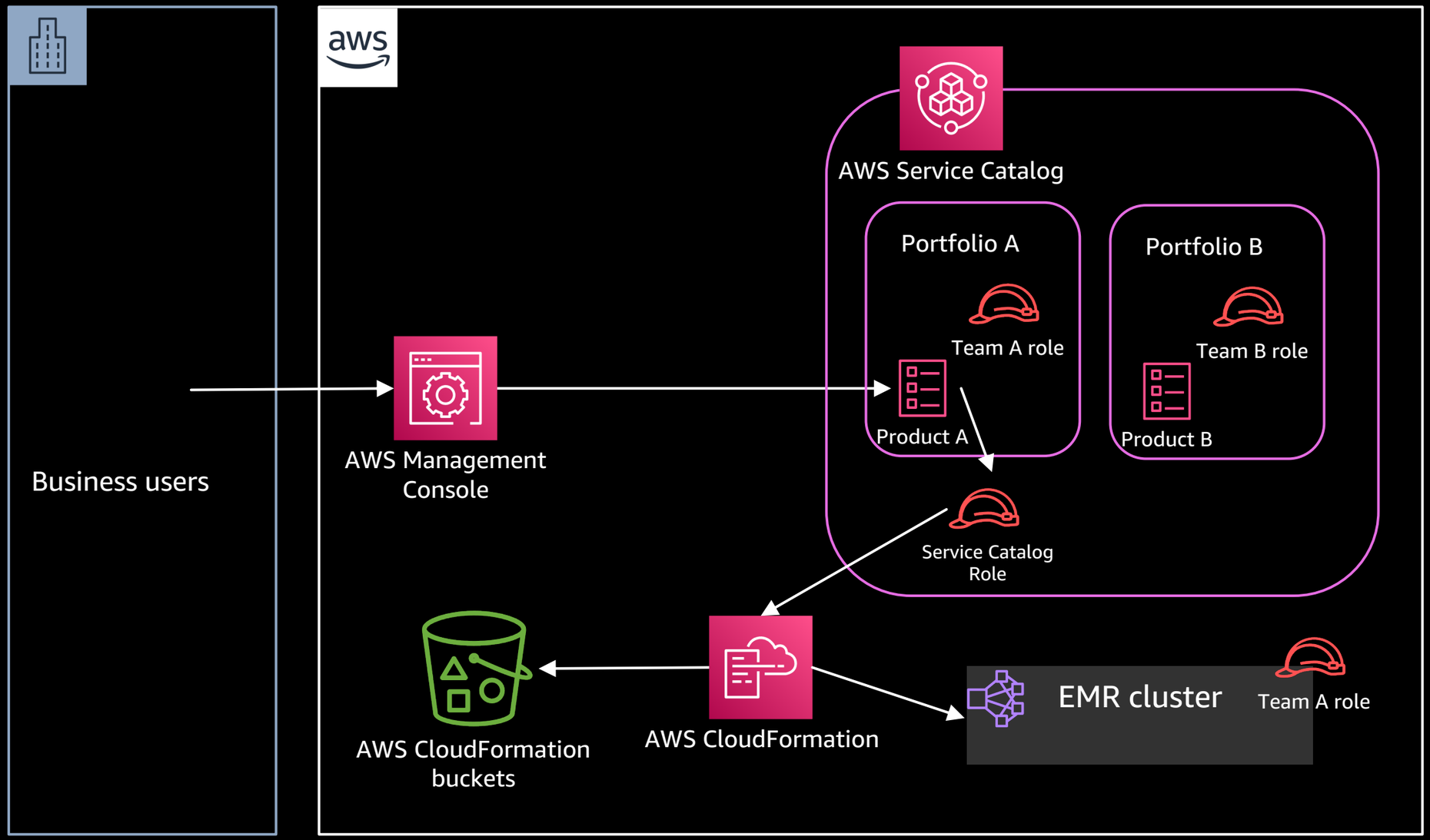
AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

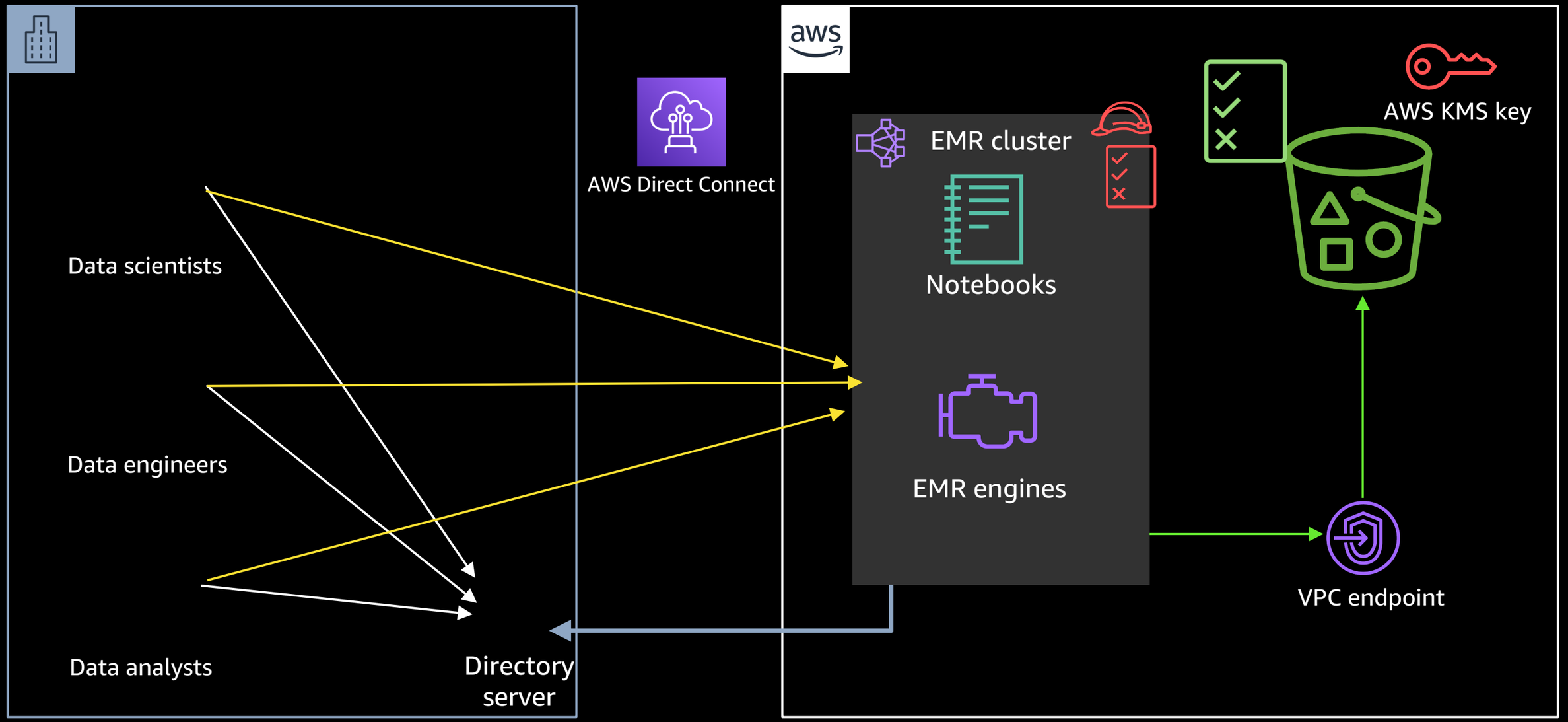
Amazon EMR build: AWS Service Catalog



Amazon EMR deploy: AWS Service Catalog



Amazon EMR: Data access



Fine-grained access control at scale



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Lots of data, smallish number of access patterns



YellowProjectRole



GreenProjectRole



BlueProjectRole



```
+ projects
| - yellow/
| - green/
| + blue/
   | blue-1.txt
   | blue-2.txt
   | …
```

Lots of data, smallish number of access patterns

```
{
  "Effect": "Allow",
  "Action": "s3:GetObject",
  "Resource": "arn:aws:s3:::example-bucket/projects/blue/*"
},
{
  "Effect": "Allow",
  "Action": "s3:ListBucket",
  "Resource": "arn:aws:s3:::example-bucket",
  "Condition": {
    "StringLike": {
      "s3:prefix": "projects/blue/*"
```



BlueProjectRole

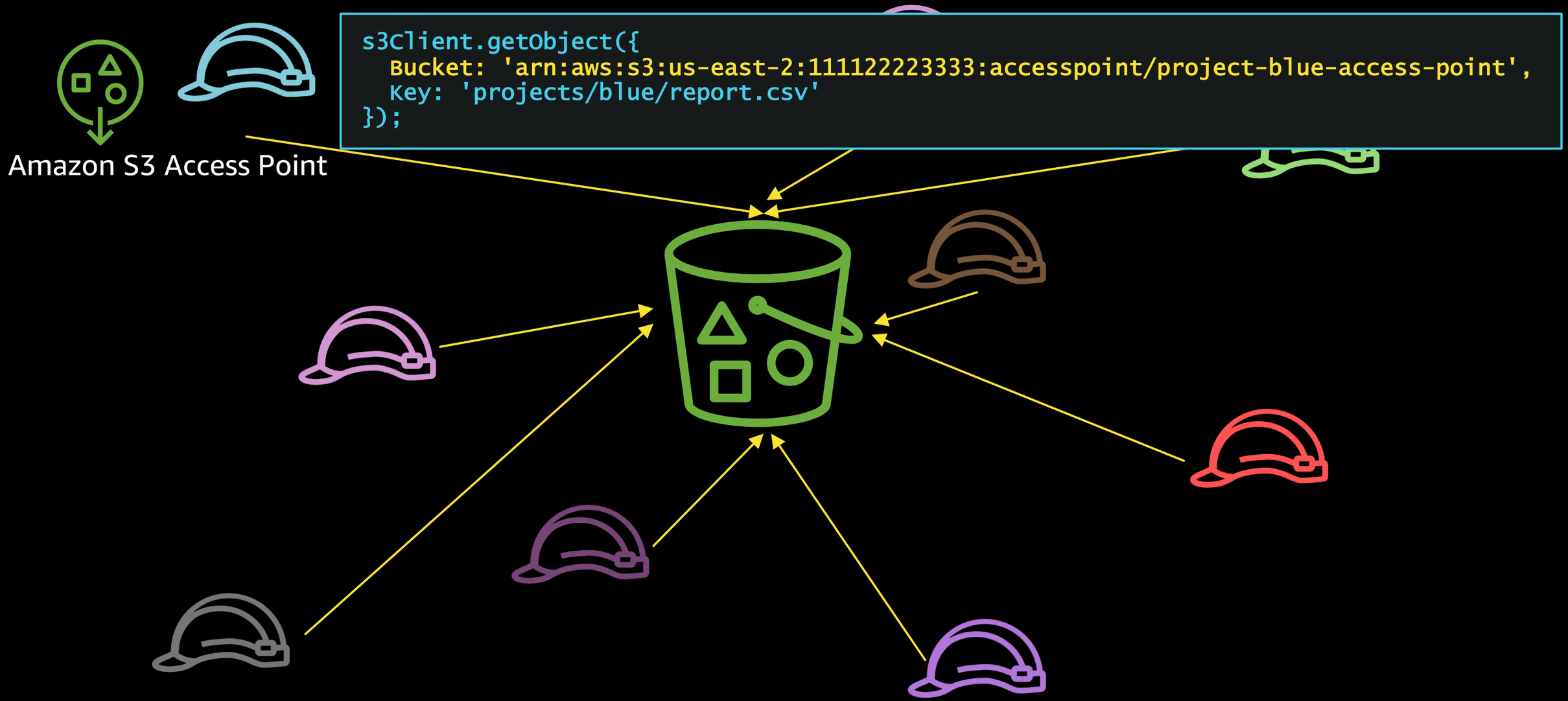


```
+ projects
| - yellow/
| - green/
| + blue/
    | blue-1.txt
    | blue-2.txt
    | …
```

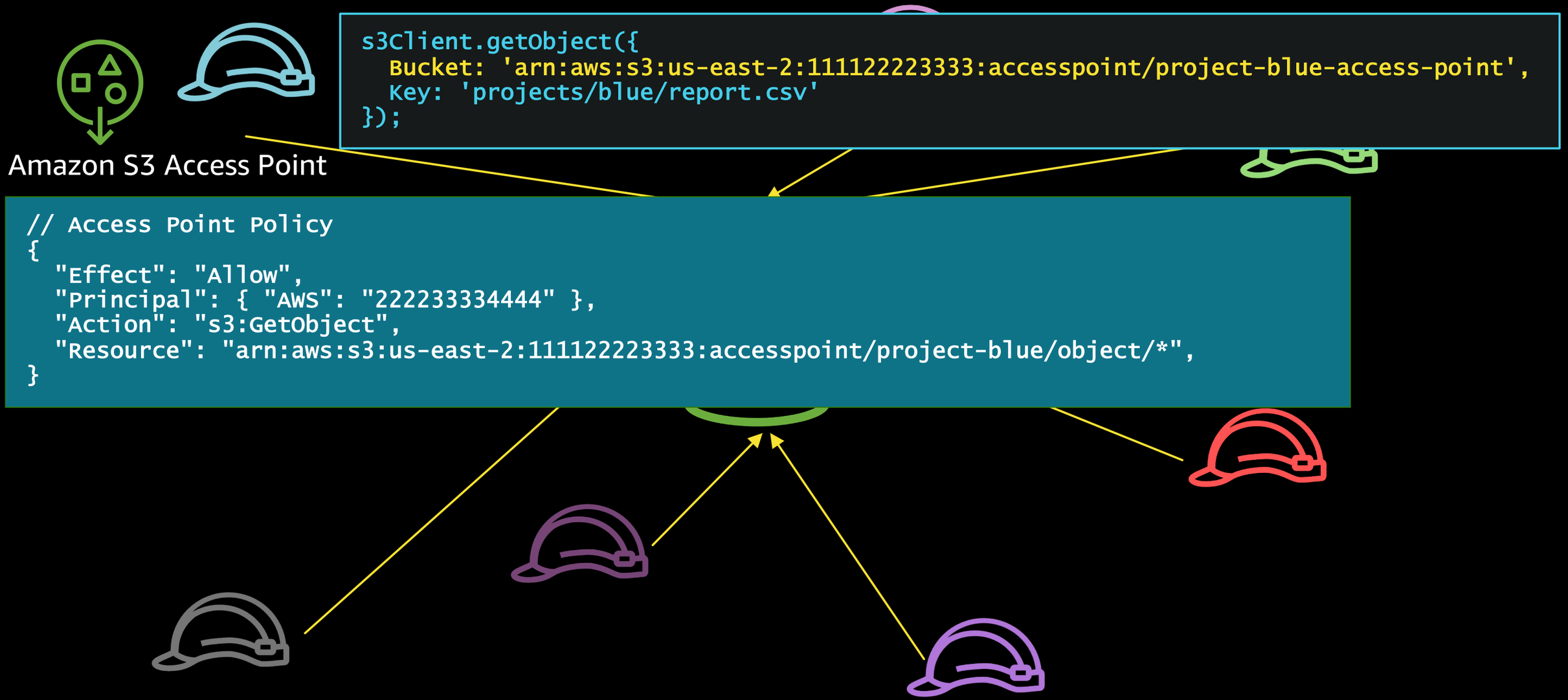
Access points: Lots of data, lots of access patterns



Access points: Lots of data, lots of access patterns



Access points: Lots of data, lots of access patterns



Access points: Lots of data, lots of access patterns



Amazon S3 Access Point

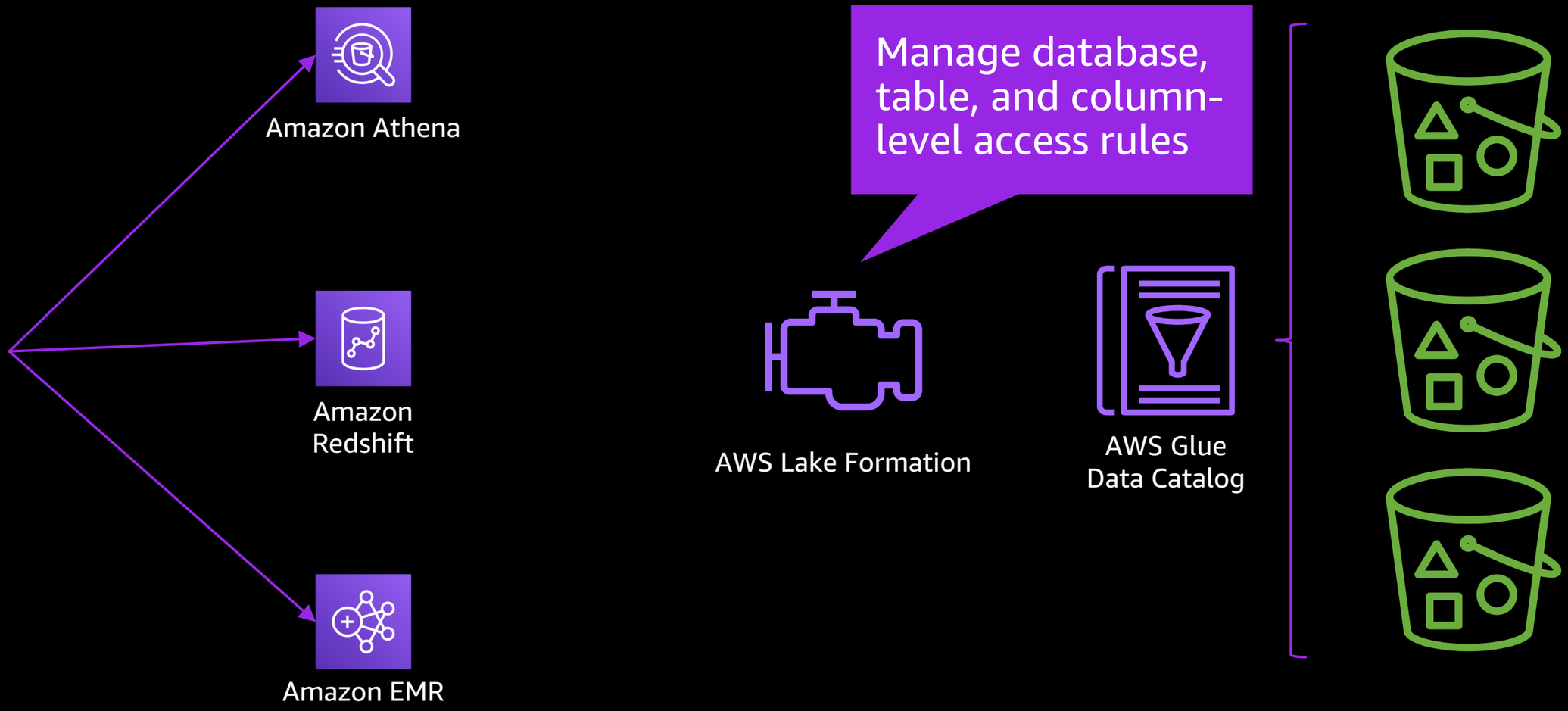
```
s3Client.getObject({
  Bucket: 'arn:aws:s3:us-east-2:111122223333:accesspoint/project-blue-access-point',
  Key: 'projects/blue/report.csv'
});
```

```
// Access Point Policy
{
  "Effect": "Allow",
  "Principal": { "AWS": "222233334444" },
  "Action": "s3:GetObject",
  "Resource": "arn:aws:s3:us-east-2:111122223333:accesspoint/project-blue/object/*",
}
```

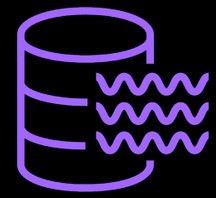


```
// Bucket policy
…
{
  "Effect": "Allow",
  "Principal": "*",  // ← But note the conditions below!
  "Action": "s3:GetObject",
  "Resource": "arn:aws:s3:::example-bucket/*",
  "Condition": {
    "StringEquals": {
      "aws:PrincipalOrgId": "o-a1b2c3"
    },
    "ArnLike": {
      "s3:DataAccessPointArn":
        "arn:aws:s3:us-east-2:111122223333:accesspoint/project-*"
```

Column-level granularity: AWS Lake Formation



Securing the data lake in Amazon S3



Data consumers with different use cases and permissions

Analytics, ETL, etc. (AWS and custom)



You are here





Data sources











...

Thank you!



AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Please complete the session survey

AWS re:Invent

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.