
AWS
re:Invent
NOV. 28 – DEC. 2, 2022 | LAS VEGAS, NV

SAS301

Multi-tenant meets ML: Building ML-based SaaS environments

Michael Beardsley

Principal Solutions Architect
AWS SaaS Factory
AWS

Anthony McClure

Sr. Solutions Architect
AWS SaaS Factory
AWS

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Agenda

SaaS and machine learning

Security concerns

Cost optimization

Data residency

Sample architectures

Open Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

“Software-as-a-Service (SaaS) is a **business** and software **delivery model** that enables organizations to offer their solution in a low-friction, service-centric approach.”

- AWS SaaS Factory

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

SaaS value proposition

Benefits for customers



Simplified IT management



Time to value/ productivity



Lower switching costs



Pay for what you use

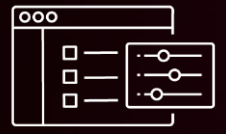
Benefits for software providers



Agility and rapid innovation



Cost optimization



Operational efficiencies



Recurring revenue

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

"The use and development of computer systems that are able to learn and adapt without following explicit instructions, by using algorithms and statistical models to analyze and draw inferences from patterns in data."

- Definition of *machine learning*, Oxford English Dictionary

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning

Model creation and ML ops

- Data collection
- Data preparation
- Feature engineering
- Model training
- Model evaluation
- Model optimization
- Model deployment and management

Inference

- Model access
- Results generation and storage

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Machine learning and SaaS considerations



Security



Cost optimization



Data residency



Model creation and ML ops

Inference

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Security

Model creation and ML ops

- Tenant isolation
- Access to training data
- Model creation pipeline(s)
- Deployment

Inference

- Tenant isolation
- Deployed model access



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Cost and other optimizations

Model creation and ML ops

- The noisy neighbor problem
- Pool vs. silo vs. bridge
- Training compute resources
- Cost allocation
- Model deployment and updates
- Auditing and tracking
- AWS Marketplace

Inference

- Scaling
- Performance
- Noisy neighbor



aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data residency and concerns

Model creation and ML ops

- Where does the training data live?
- Cross account considerations?
- Data quality

Inference

- Isolation of model inference results
- Inputs
- Parameters



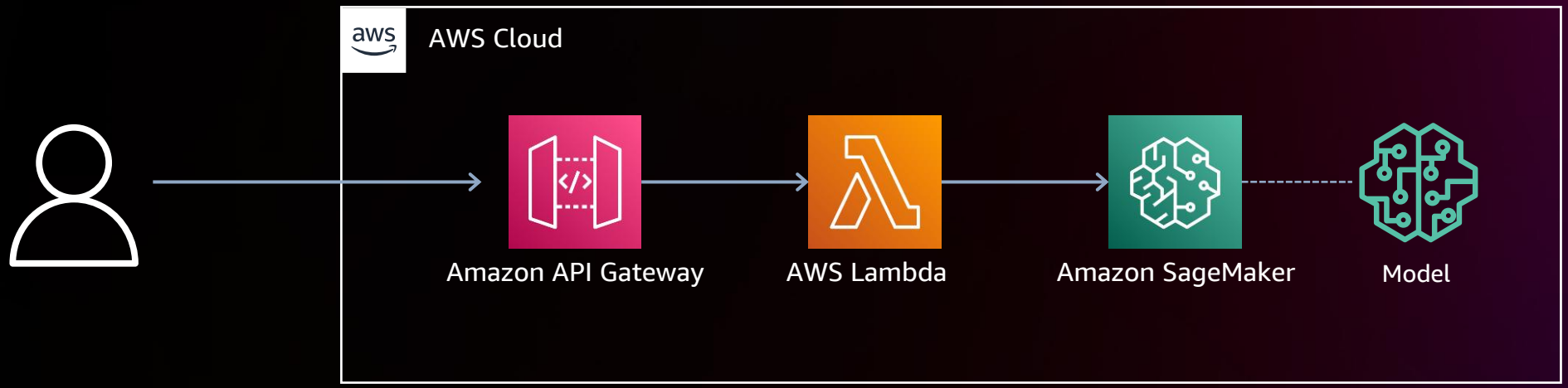
aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Sample architectures

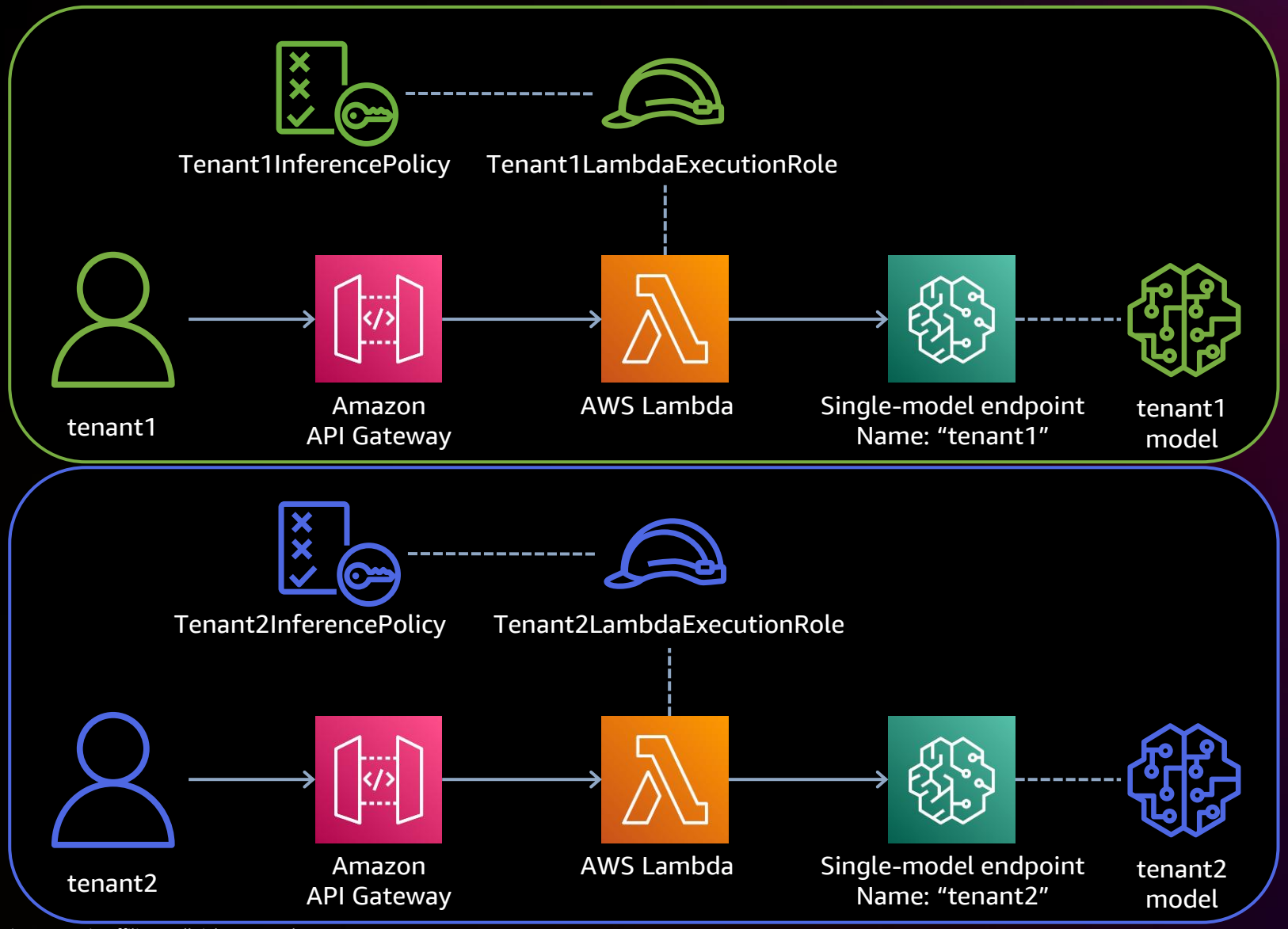
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Silo, pool, and bridge inference



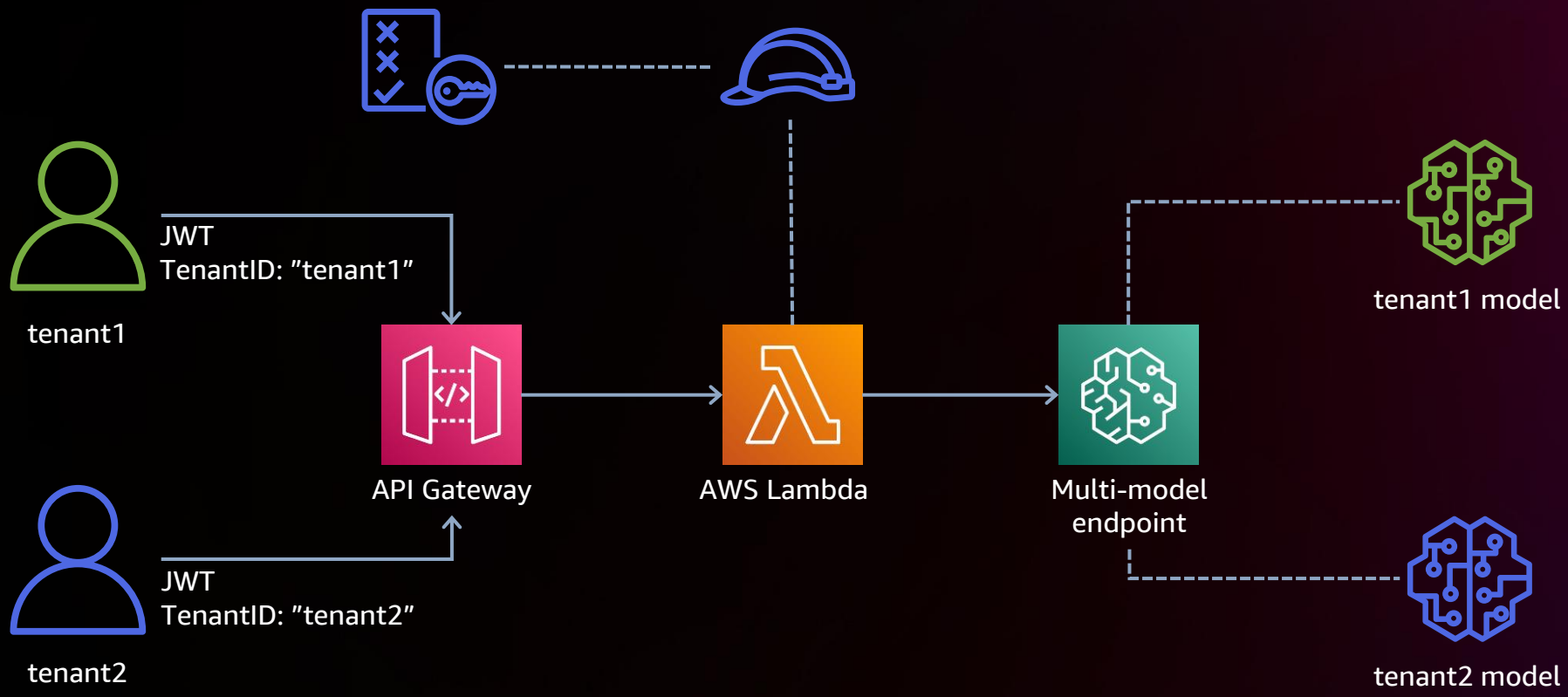
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-tenant silo inference



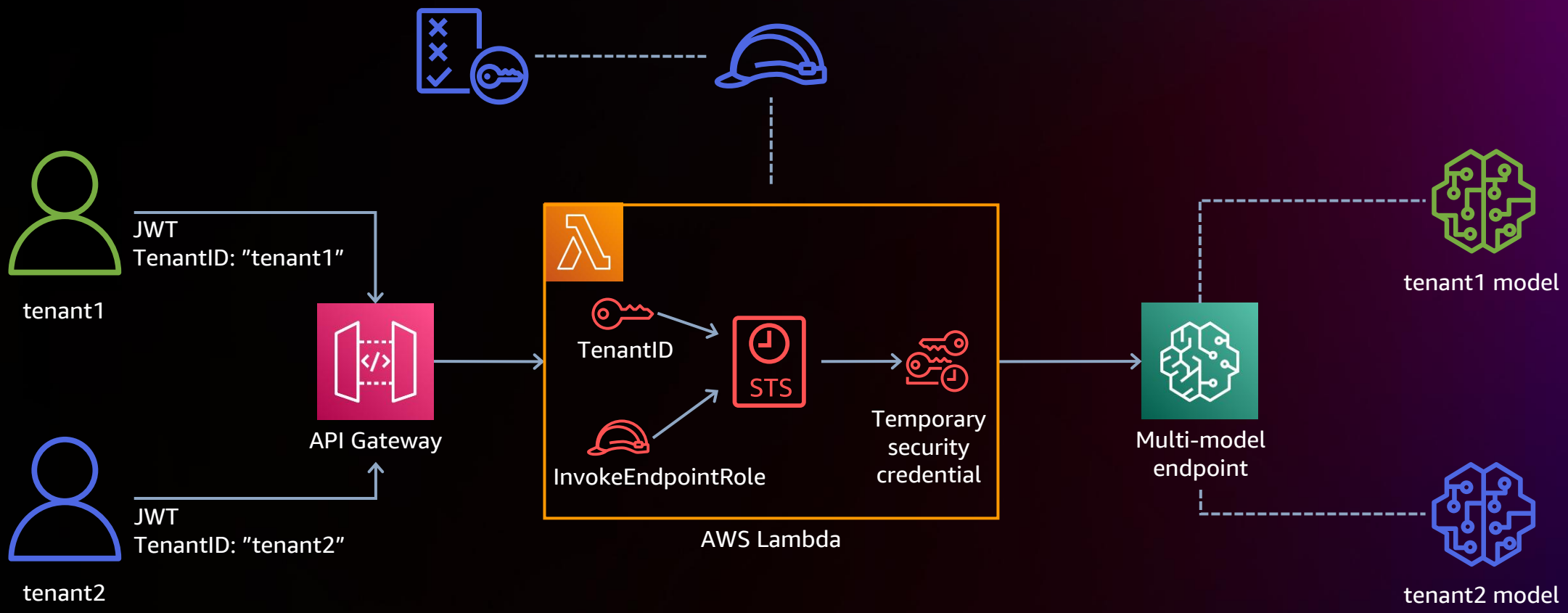
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-tenant pool inference



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-tenant pool inference



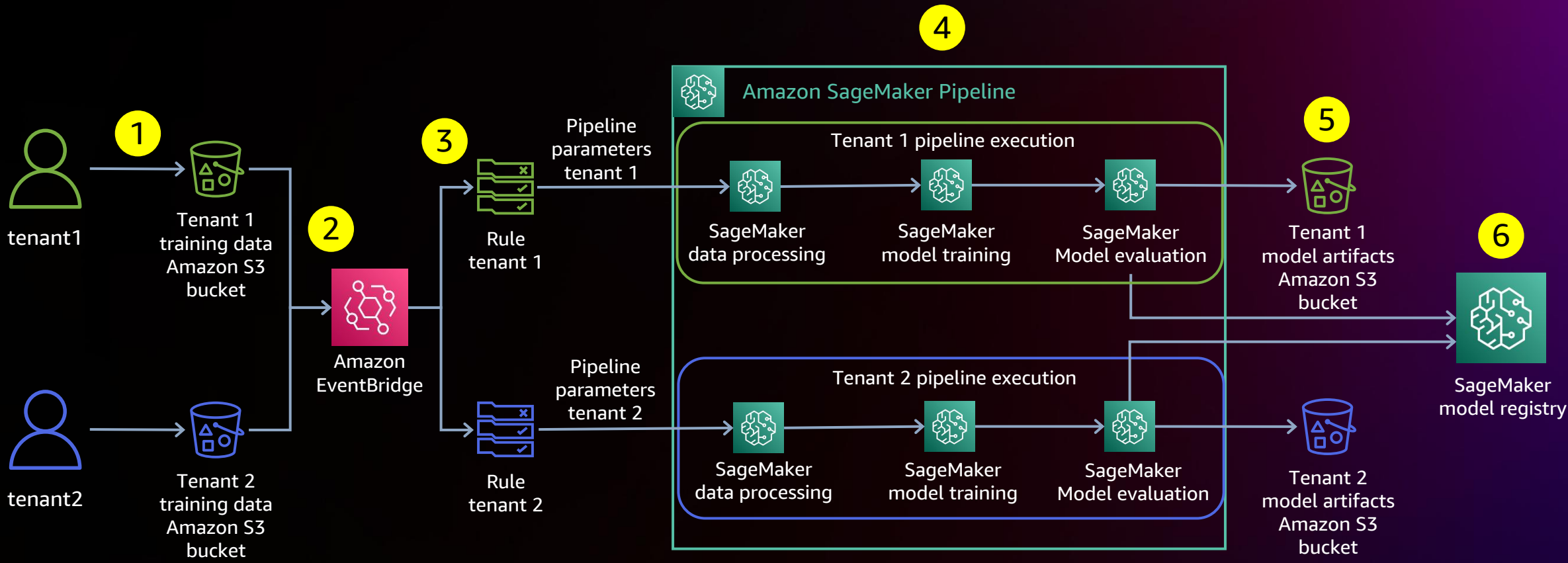
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

InvokeEndpointPolicy example

```
{
  "Version": "2012-10-17",
  "Statement": [
    {
      "Sid": "VisualEditor0",
      "Effect": "Allow",
      "Action": "sagemaker:InvokeEndpoint",
      "Resource": "*",
      "Condition": {
        "StringEquals": {
            "sagemaker:TargetModel":"${aws:PrincipalTag/TenantID}.tar.gz”
        }
      }
    }
  ]
}
```

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Multi-tenant model creation pipeline



© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Open Q&A

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

More SaaS sessions

Breakout sessions

- SAS305 – SaaS architecture patterns: From concept to implementation
- SAS405 – SaaS microservices deep dive: Simplifying multi-tenant development
- SAS306 – SaaS migration: Inside a real-world multi-tenant transformation
- SAS302 – Supporting extensibility in SaaS environments
- PEX310 – Optimizing your multi-tenant SaaS architecture

Workshops

- SAS403 - SaaS microservices deep dive: Multi-tenancy meets microservices
- SAS402 – Serverless meets SaaS: Inside a real-world serverless SaaS solution
- SAS401 – Amazon EKS SaaS: Building a working multi-tenant environment

Business session

- PEX209 – Building your SaaS journey on AWS

aws
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

More SaaS sessions

Chalk talks

- SAS307 – DevOps and SaaS: Applying automation in multi-tenant environments
- SAS303 – SaaS anywhere: Building SaaS solutions that runs in hybrid models
- SAS301 – Multi-tenant meets ML: Building ML-based SaaS environments
- SAS304 – Solving the SaaS compliance puzzle
- PEX313 – The SaaS control plane: The heart of SaaS growth
- ARC403 – Amazon EKS SaaS deep dive: Inside a multi-tenant EKS solution
- ARC323 – Designing a multi-tenant SaaS tiering and throttling strategy
- SVS315 – Building multi-tenant applications with AWS Lambda and AWS Fargate

Builder session

- ARC327 – How to optimize cost in your multi-tenant architecture

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Additional resources

1

Subscribe to AWS SaaS insights

Get monthly emails with bite-sizes advice and the latest updates

2

Explore the SaaS on AWS hub

Check out the SaaS on AWS page for more resources and insights

3

Discover resources for builders

Access our curated list of SaaS reference solutions, demos, tech events, and more

aws

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Michael Beardsley
mibeard@amazon.com

Anthony McClure
antmcl@amazon.com



Please complete the session survey in the **mobile app**

© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.