Scale Your Data Governance with Amazon SageMaker Unified Studio by Automating Subscription Approvals

Contents

[Introduction 2](#_Toc204003381)

[Understanding SageMaker Unified Studio's Native Capabilities 2](#_Toc204003382)

[Built-in Subscription Management 2](#_Toc204003383)

[Default Subscription Approval Workflow 3](#_Toc204003384)

[Event-Driven Architecture Solution 4](#_Toc204003385)

[Technical Architecture Overview 4](#_Toc204003386)

[Subscription fulfilment workflow using Event driven architecture 6](#_Toc204003387)

[High Level Architecture: 7](#_Toc204003388)

[Architecture Benefits 7](#_Toc204003389)

[Implementation Guide 7](#_Toc204003390)

[Prerequisites 7](#_Toc204003391)

[Step-by-Step Implementation 8](#_Toc204003392)

[Testing and Validation 9](#_Toc204003393)

[Success Metrics 9](#_Toc204003394)

[Security, Governance and Compliance Considerations 9](#_Toc204003395)

[Advanced Options 9](#_Toc204003396)

[Conditional Logic Implementation 9](#_Toc204003397)

[Integration Possibilities 10](#_Toc204003398)

[Best Practices and Recommendations 10](#_Toc204003399)

[Performance Optimization 10](#_Toc204003400)

[Operational Excellence 10](#_Toc204003401)

[Conclusion 10](#_Toc204003402)

[Next Steps 11](#_Toc204003403)

# Introduction

Amazon SageMaker Unified Studio revolutionizes data and AI development by providing a single, integrated environment where organizations can discover, access, and act on data using best-in-class tools across any use case. This unified platform seamlessly integrates functionality from existing AWS Analytics and AI/ML services, including Amazon EMR, AWS Glue, Amazon Athena, Amazon Redshift, Amazon Bedrock, and Amazon SageMaker AI.

As enterprise organizations manage thousands of data assets requiring publication to data marketplaces, they face a critical challenge: balancing governance requirements with operational efficiency. While manual approval workflows remains essential for sensitive datasets and production systems, there's growing demand for automated approval capabilities that can accelerate access to lower-risk data assets in non-production environments without compromising governance frameworks.

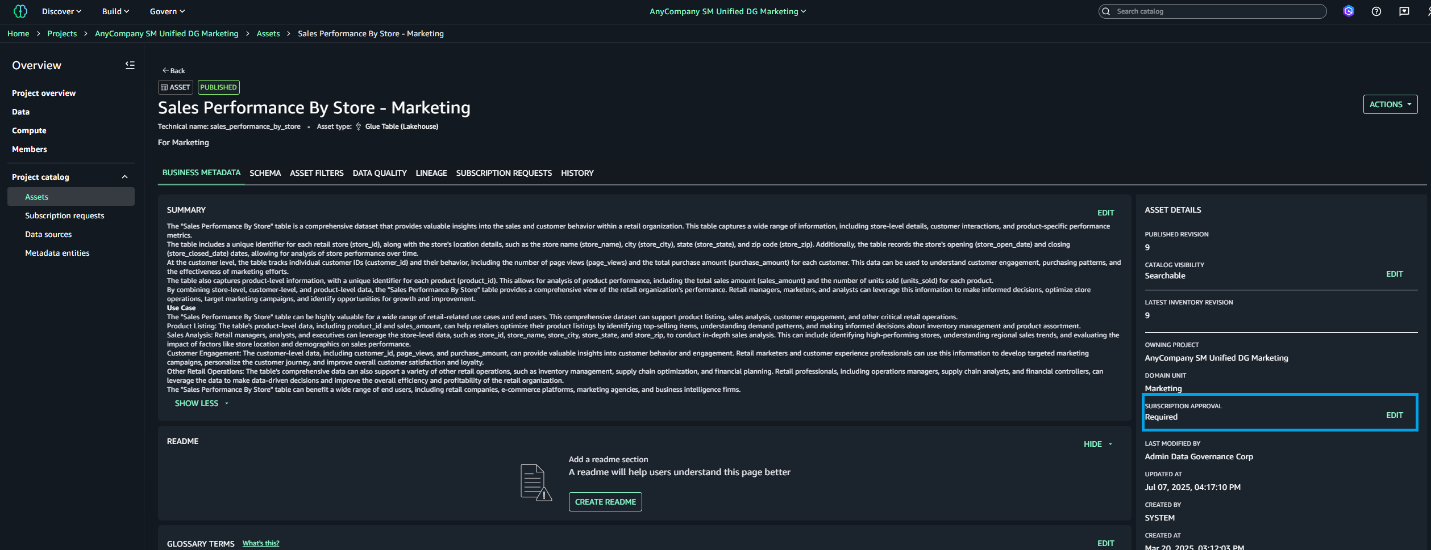
This blog post explores how to implement an event-driven architecture solution that automates subscription request approvals within SageMaker Unified Studio, enabling faster data access while maintaining compliance with established governance standards.

## Built-in Subscription Management in SageMaker Unified Studio

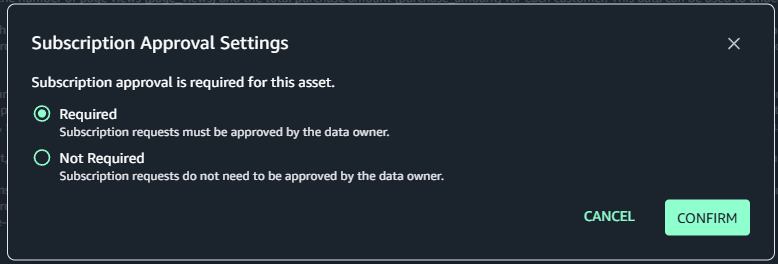
SageMaker Unified Studio provides project owners with granular control over data asset access through configurable subscription requirements. Data producers can easily configure individual assets to require or bypass subscription approval at the asset level.

**Configuration Steps:**

1. **Access Asset Configuration**: Data producers log into the SageMaker Unified Studio console and navigate to their specific project → Assets → Select the target asset for configuration.
2. **Edit Subscription Settings**: In the Asset details page, locate the right pane and click 'Edit' on 'Subscription Required'.



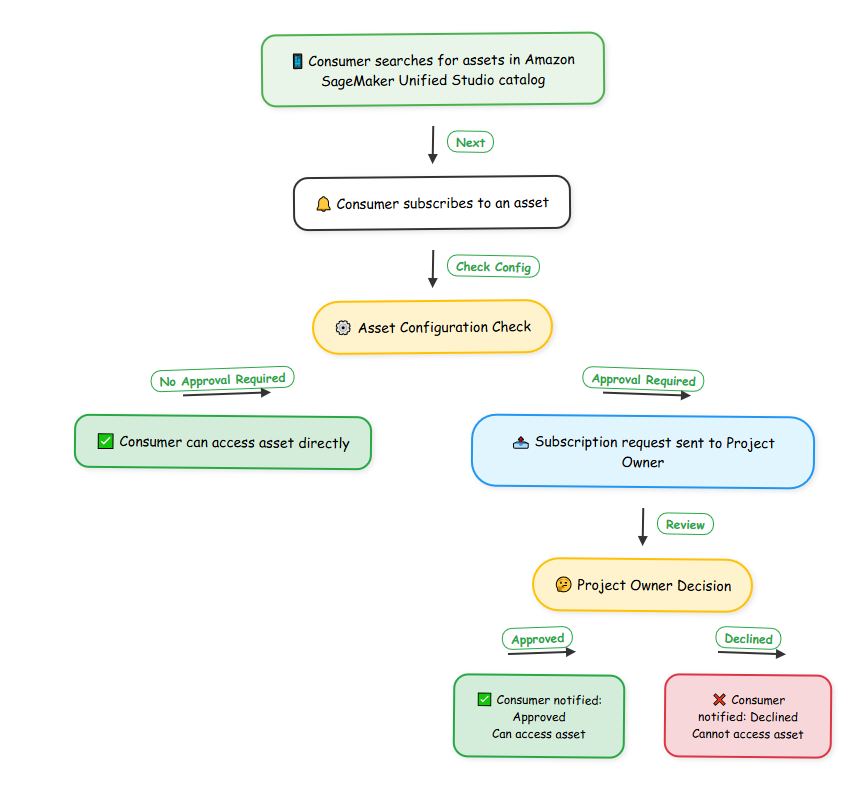
1. **Apply Configuration**: Select the appropriate configuration based on your use case and confirm the changes.



## Default Subscription Approval Workflow

The out-of-the-box subscription request and approval process provides comprehensive governance control:

1. **Request Submission**: Data consumers submit subscription requests for desired assets
2. **Thorough Review**: Data producers or designated approvers carefully evaluate requests to ensure compliance
3. **Informed Decision**: Approvers make well-considered decisions based on established governance policies
4. **Secure Access**: Upon approval, requesters gain secure access to the subscribed data assets



In the above workflow, automated approvals can accelerate data accessibility for users working with low-risk assets in development environments.

# Event-Driven Architecture Solution

## Technical Architecture Overview

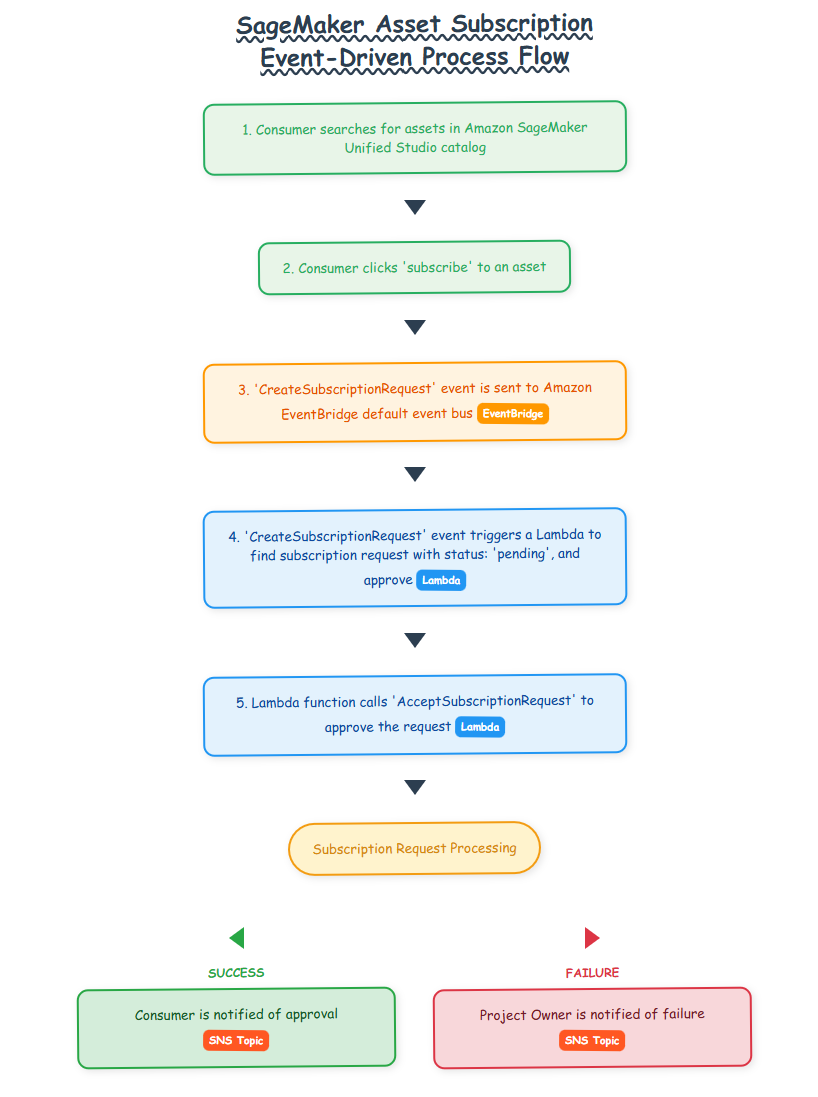
To automate subscription request approvals for lower-risk environments and assets, we implement an event-driven architecture leveraging the following AWS services:

**Core Components:**

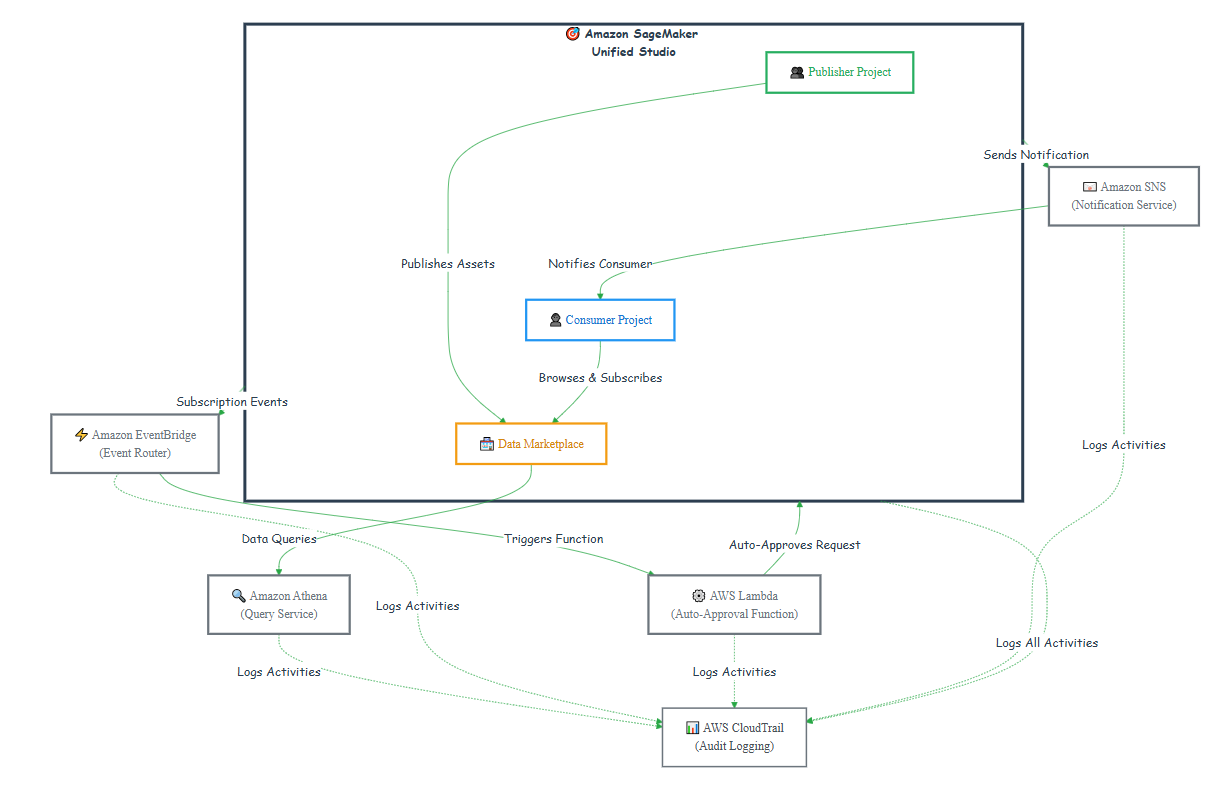
1. **Amazon EventBridge**: Provides real-time access to changes in data in AWS services, your own applications and software as a service (SaaS) applications without writing code. In addition to sending messages to your dedicated inbox in the data portal, DataZone also sends these messages to your Amazon EventBridge default event bus in the same AWS account where your Amazon DataZone root domain is hosted. This enables event-driven automation, such as subscription fulfillment or custom integrations with other tools.
2. **AWS Lambda**: Executes serverless code to process subscription events and automatically approve qualifying requests. Lambda's pay-per-use model ensures cost-effectiveness with zero server management overhead. The Lambda function will process incoming subscription request events and automatically approve them based on predefined criteria such as asset classification, requesting user permissions, and environment type.

* **Amazon Simple Notification Service (SNS)**: Delivers immediate notifications to subscribers confirming their subscription approvals through a scalable publish-subscribe messaging paradigm. In this solution, we leverage SNS to send subscriber an email notification confirming their subscription is approved.
* **AWS CloudTrail**: Provides comprehensive auditing, security monitoring, and operational troubleshooting capabilities by tracking all user activity and API usage related to the automated approval process. In this solution, we leverage Cloud Trail to access all SageMaker Unified Studio events (that are logged under DataZone) in CloudTrail.

## Subscription fulfilment workflow using Event driven architecture



## High Level Architecture:



## Architecture Benefits

This serverless architecture delivers several key advantages:

* **Scalability**: Automatically handles high volumes of subscription requests
* **Cost Efficiency**: Pay-only-for-use model with no idle resource costs
* **Minimal Maintenance**: Serverless components require no infrastructure management
* **Flexible Triggering**: Supports event-driven, scheduled, and manual invocation modes
* **Audit Compliance**: Complete logging and traceability through CloudTrail

# Implementation Guide

## Prerequisites

Before implementing the automated approval solution, ensure the following requirements are met:

1. **SageMaker Unified Studio Setup**: Projects, domains, and users are properly configured
2. **IAM Role Configuration**: Add the following as Contributors in SageMaker Unified Studio Projects:
   * Lambda execution role
   * DataZone user role (derived from Project ARN)
3. **Lambda Permissions**: Configure appropriate IAM permissions for the Lambda execution role.

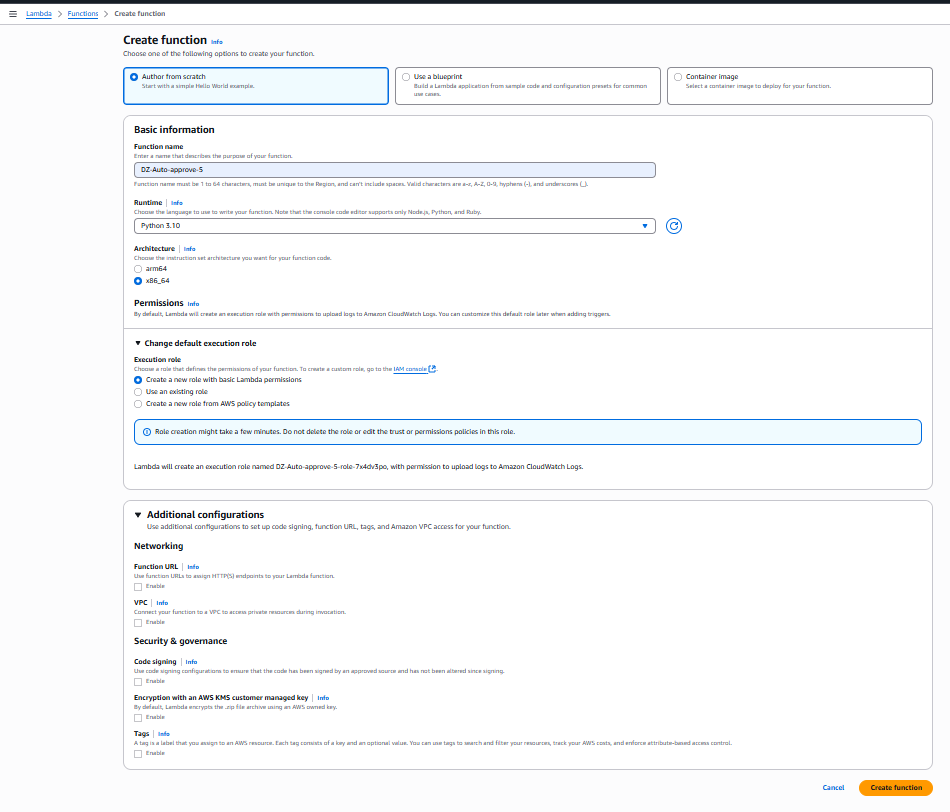
**Insert Lambda permissions json**

## Step-by-Step Implementation

**Step 1: Lambda Function Configuration**

Create a Lambda function with the following specifications:

* **Runtime**: Python 3.9 or later
* **Memory**: 128 MB (adjustable based on processing requirements)
* **Timeout**: 5 minutes



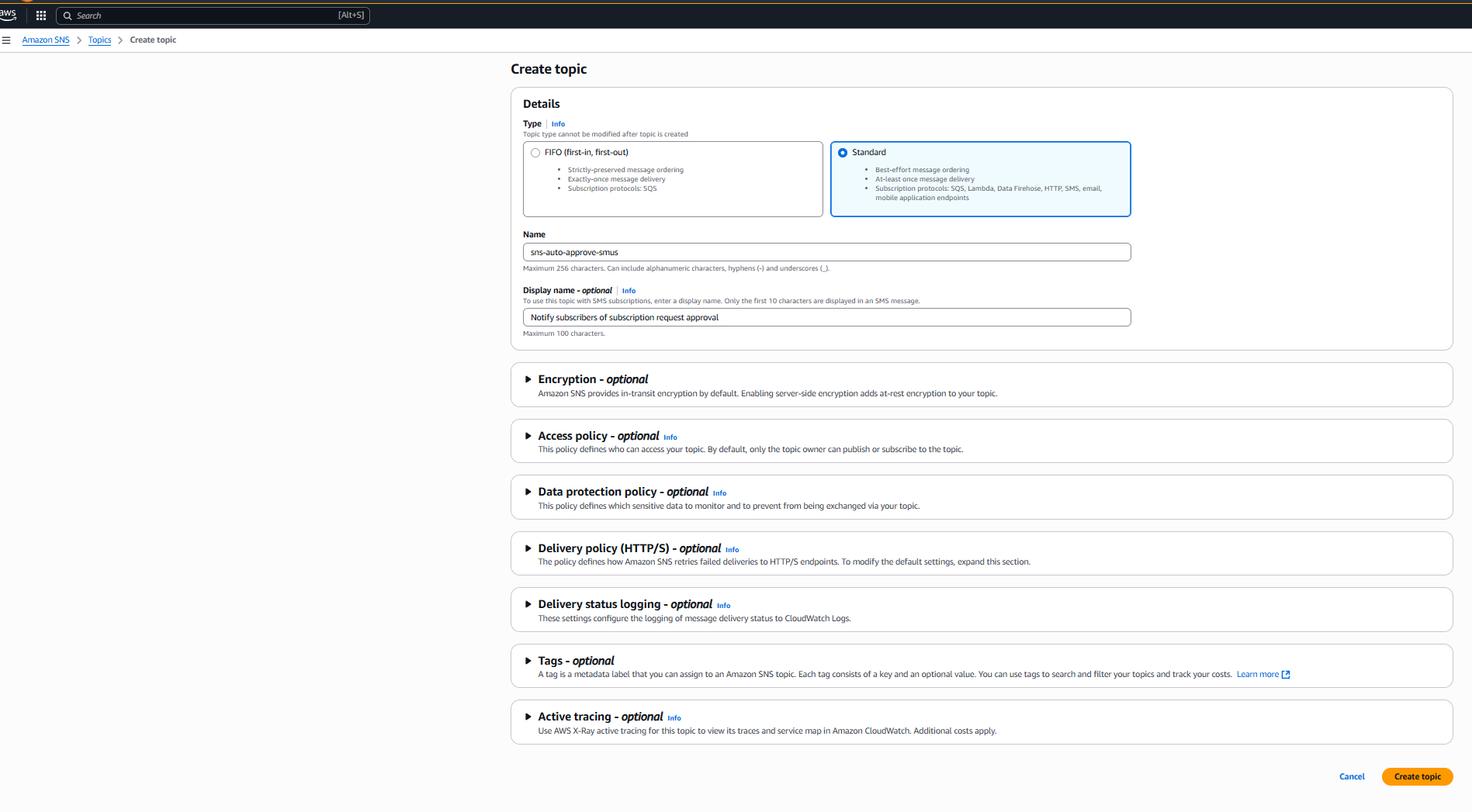
**Step 2: Deploy code in Lambda.**

Refer to GitHub repo for Lambda code:

**Step 3: SNS Topic Creation**

Establish an SNS topic to handle notification delivery:

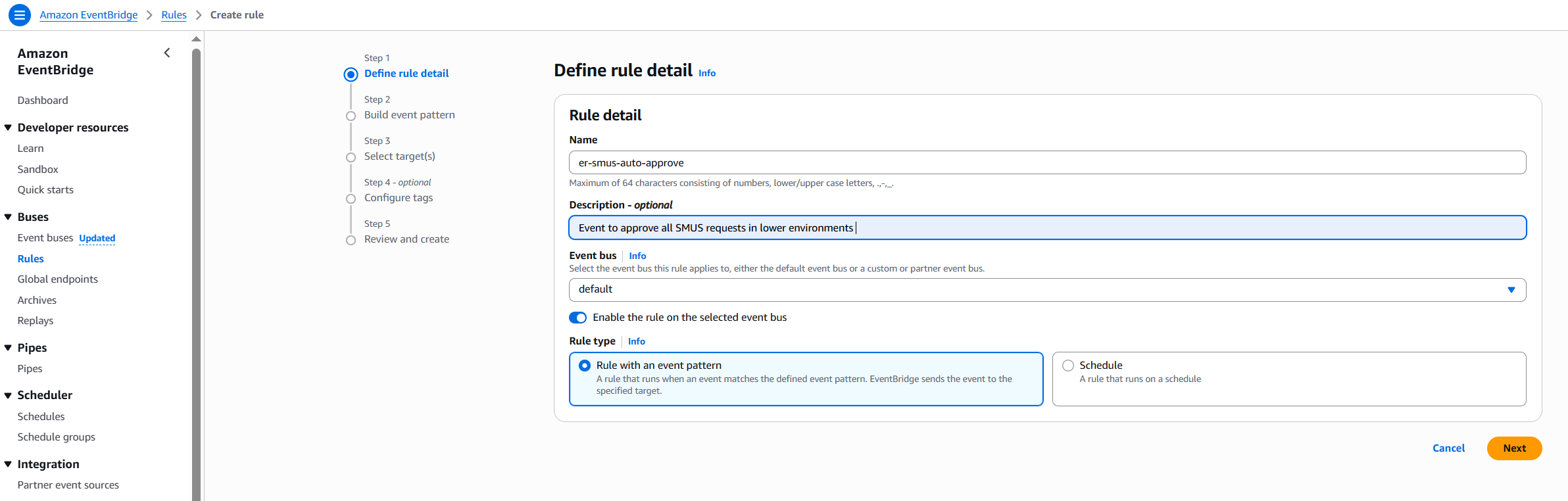
* **Topic Type**: Standard (for most use cases)
* **Subscribers**: Configure email endpoints for relevant stakeholders

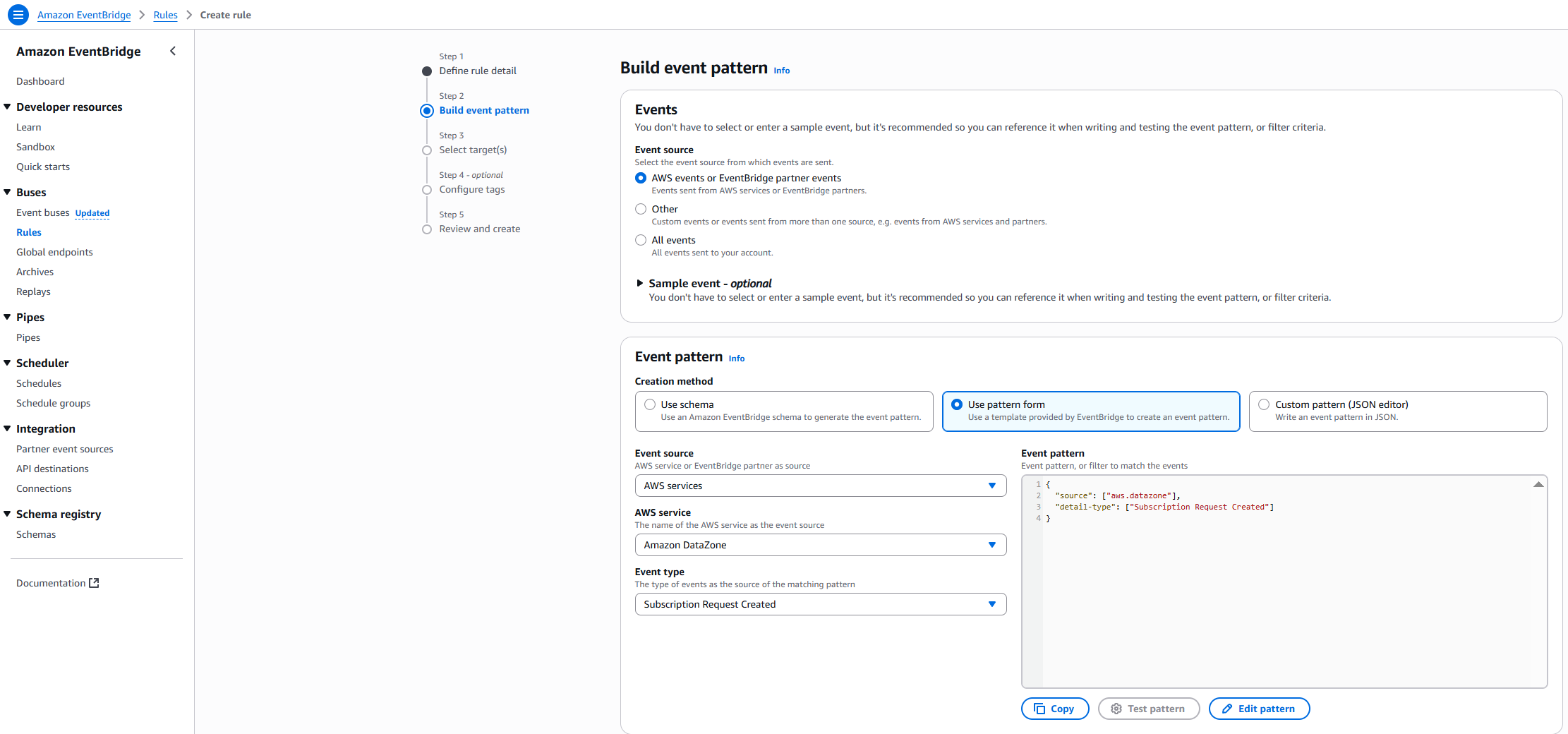


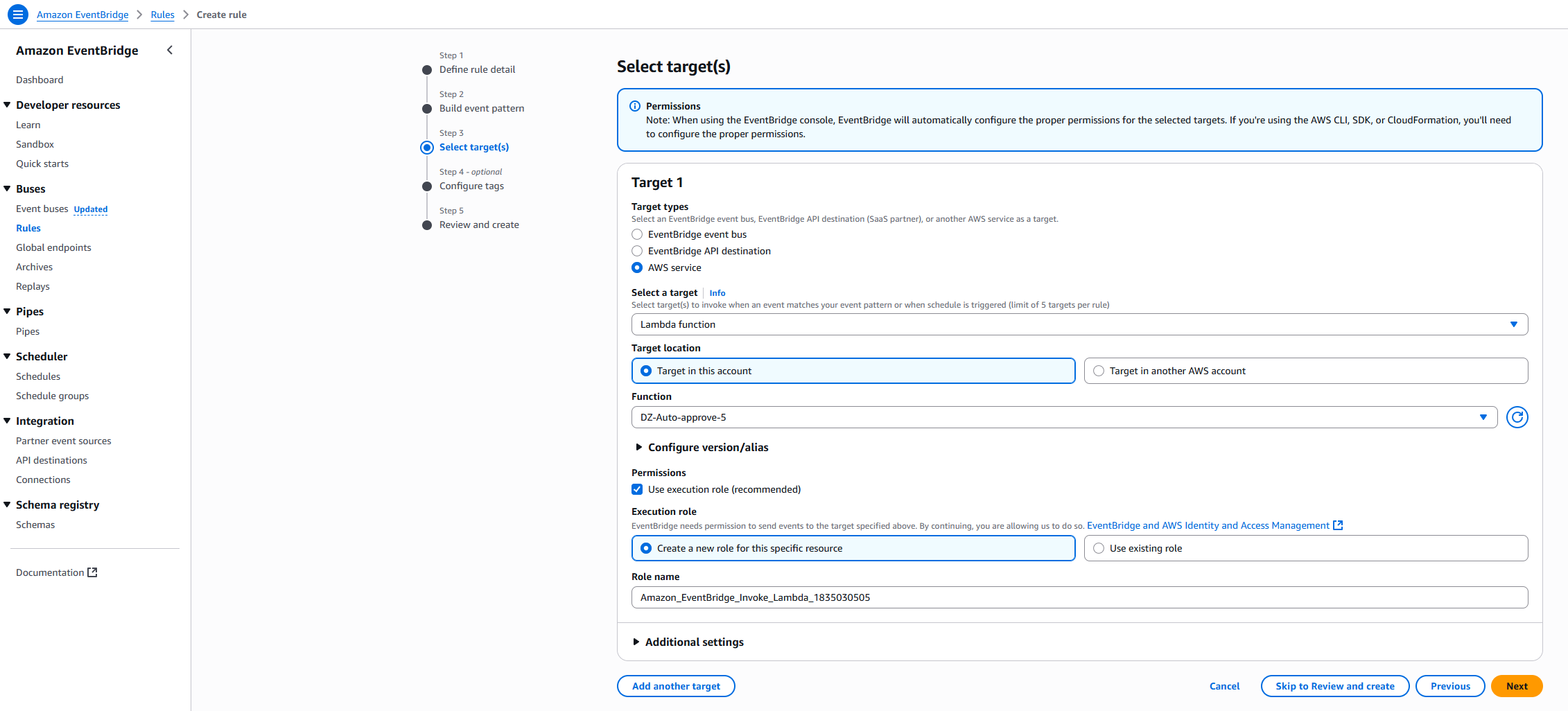
**Step 4: EventBridge Rule Configuration**

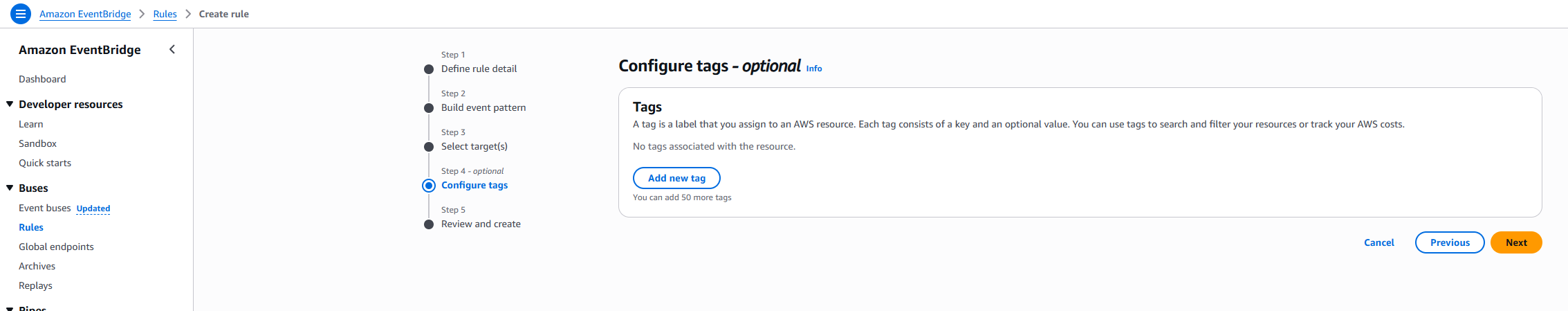
Create an EventBridge rule to capture subscription request events:

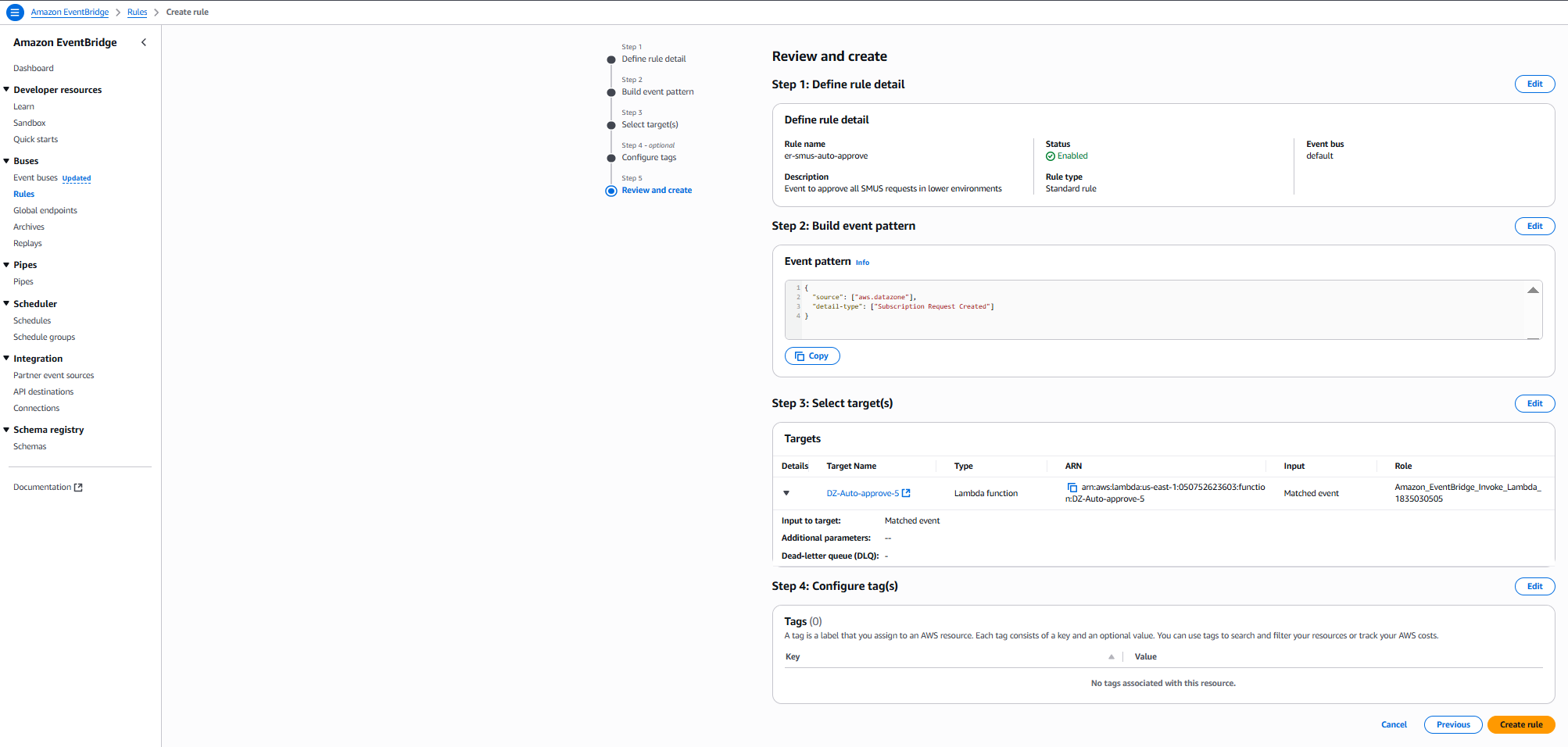
* **Event Pattern**: Target specific SageMaker Unified Studio subscription request events
* **Event Source**: Configure to match aws.datazone events
* **Detail Type:** Subscription request creation events
* **Targets**: Route events to both Lambda function and SNS topic











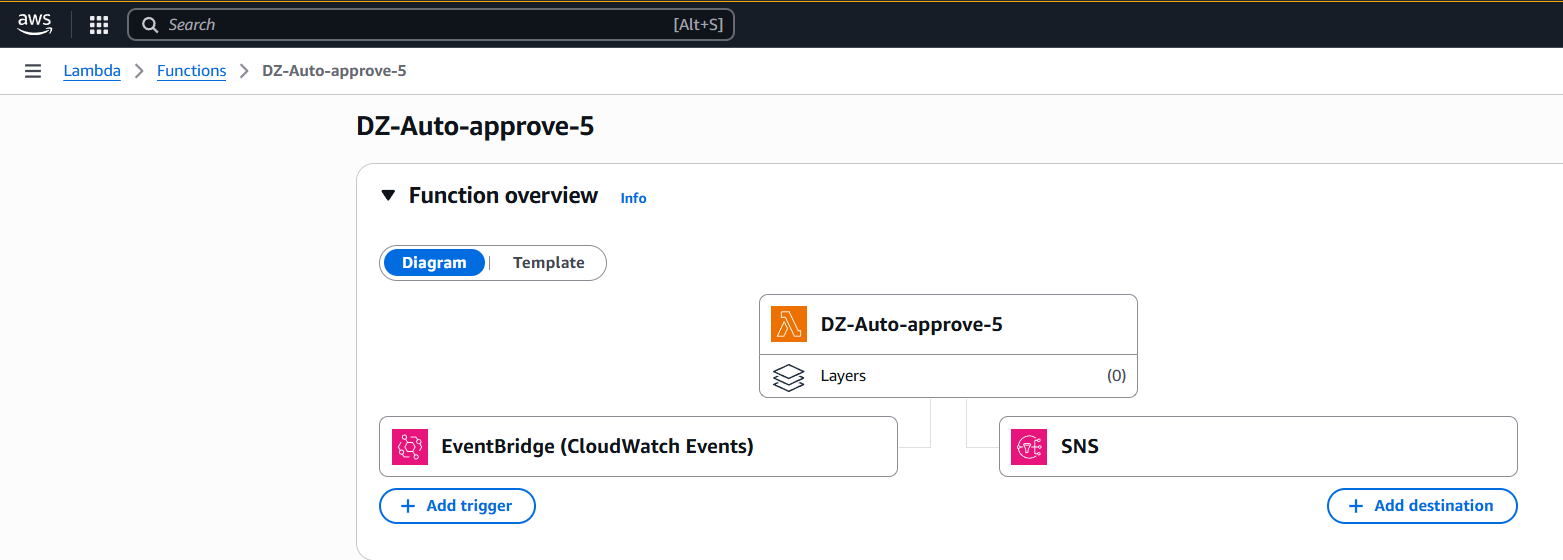
**Sample Event Pattern:**

**Insert event JSON**

**Step 5: Integration Configuration**

Configure the EventBridge rule to trigger both the Lambda function and SNS notifications:

* **Primary Target**: Lambda function for approval processing
* **Secondary Target**: SNS topic for immediate stakeholder notification
* **Error Handling**: Configure dead letter queues for failed processing attempts



## Testing and Validation

**Test Scenario Execution**

1. **Initiate Subscription Request**: Navigate to SageMaker Data Catalog within SageMaker Unified Studio and click 'Subscribe' on a configured asset
2. **Monitor Approval Status**: Check Subscribing Project → Subscription Requests → Outgoing Requests → Approved tab to verify automatic approval
3. **Verify Approval Details**: Click "View Subscription" to confirm the approver appears as the Lambda execution role with "Auto approved by Lambda" reason
4. **Audit Trail Verification**: Navigate to CloudTrail → Event History, search for "AcceptSubscriptionRequest" events to review the automated approval audit trail

# Success Metrics

Monitor the following KPIs to measure solution effectiveness:

* **Approval Time Reduction**: Measure the decrease in average time from request to approval
* **Processing Volume**: Track the number of automated approvals handled
* **Error Rate**: Monitor failed processing attempts and resolution times
* **User Satisfaction**: Gather feedback on improved data access experience

# Security, Governance and Compliance Considerations

The automated approval solution maintains governance standards through:

* **Selective Automation**: Only applies to pre-classified low-risk assets and non-production environments
* **Audit Trail**: Complete logging of all automated decisions through CloudTrail
* **Role-Based Access**: Respects existing IAM and SageMaker Unified Studio permission structures
* **Configurable Criteria**: Allows customization of approval logic based on organizational policies

# Advanced Options

## Conditional Logic Implementation

Enhance the Lambda function to include sophisticated approval logic:

* **Asset Metadata Analysis**: Evaluate data classification, sensitivity levels, and compliance tags
* **User Context Evaluation**: Consider requesting user's role, department, and historical access patterns
* **Time-Based Rules**: Implement different approval criteria based on time of day or business hours
* **Resource Utilization**: Factor in current system load and resource availability

## Integration Possibilities

Extend the solution to integrate with enterprise systems:

* **ITSM Integration**: Connect with ServiceNow or similar platforms for ticket creation
* **Approval Workflows**: Interface with existing approval systems for hybrid automation
* **Monitoring Dashboards**: Create custom dashboards for approval metrics and system health
* **Cost Management**: Implement usage tracking and cost allocation features

# Best Practices and Recommendations

## Performance Optimization

* **Event Filtering**: Use precise EventBridge patterns to minimize unnecessary Lambda invocations
* **Batch Processing**: Consider batching multiple requests for efficiency when volume is high
* **Caching**: Implement caching for frequently accessed metadata to reduce API calls
* **Monitoring**: Set up CloudWatch alarms for key performance metrics

## Operational Excellence

* **Documentation**: Maintain comprehensive documentation of approval criteria and exception handling
* **Version Control**: Use infrastructure as code (CloudFormation/CDK) for solution deployment
* **Testing Strategy**: Implement automated testing for approval logic and edge cases
* **Disaster Recovery**: Plan for solution recovery and data consistency in failure scenarios

# Conclusion

Event-driven architecture perfectly complements SageMaker Unified Studio's native features and capabilities, creating a powerful solution for modern data governance challenges. By implementing automated subscription request approvals, organizations can significantly streamline data accessibility for low-risk assets while maintaining robust compliance with established governance standards.

This solution empowers data teams to achieve faster time-to-insight while ensuring that sensitive data remains protected through appropriate manual oversight. The serverless architecture ensures scalability, cost-effectiveness, and minimal operational overhead, making it an ideal choice for organizations looking to modernize their data governance practices.

As data volumes continue to grow and the pace of business accelerates, solutions like this automated approval system become essential for maintaining competitive advantage while ensuring responsible data stewardship. The combination of Amazon SageMaker Unified Studio's comprehensive data management capabilities with event-driven automation creates a foundation for scalable, efficient, and secure data operations.

# Next Steps

To implement this solution in your organization:

1. **Assessment**: Evaluate your current data governance processes and identify automation opportunities
2. **Pilot Implementation**: Start with a small set of low-risk assets in a development environment
3. **Gradual Expansion**: Progressively extend automation to additional asset types and environments
4. **Continuous Improvement**: Regularly review and refine approval criteria based on operational experience

For additional resources and detailed implementation code, refer to the AWS documentation and best practices guides for SageMaker Unified Studio and event-driven architectures.