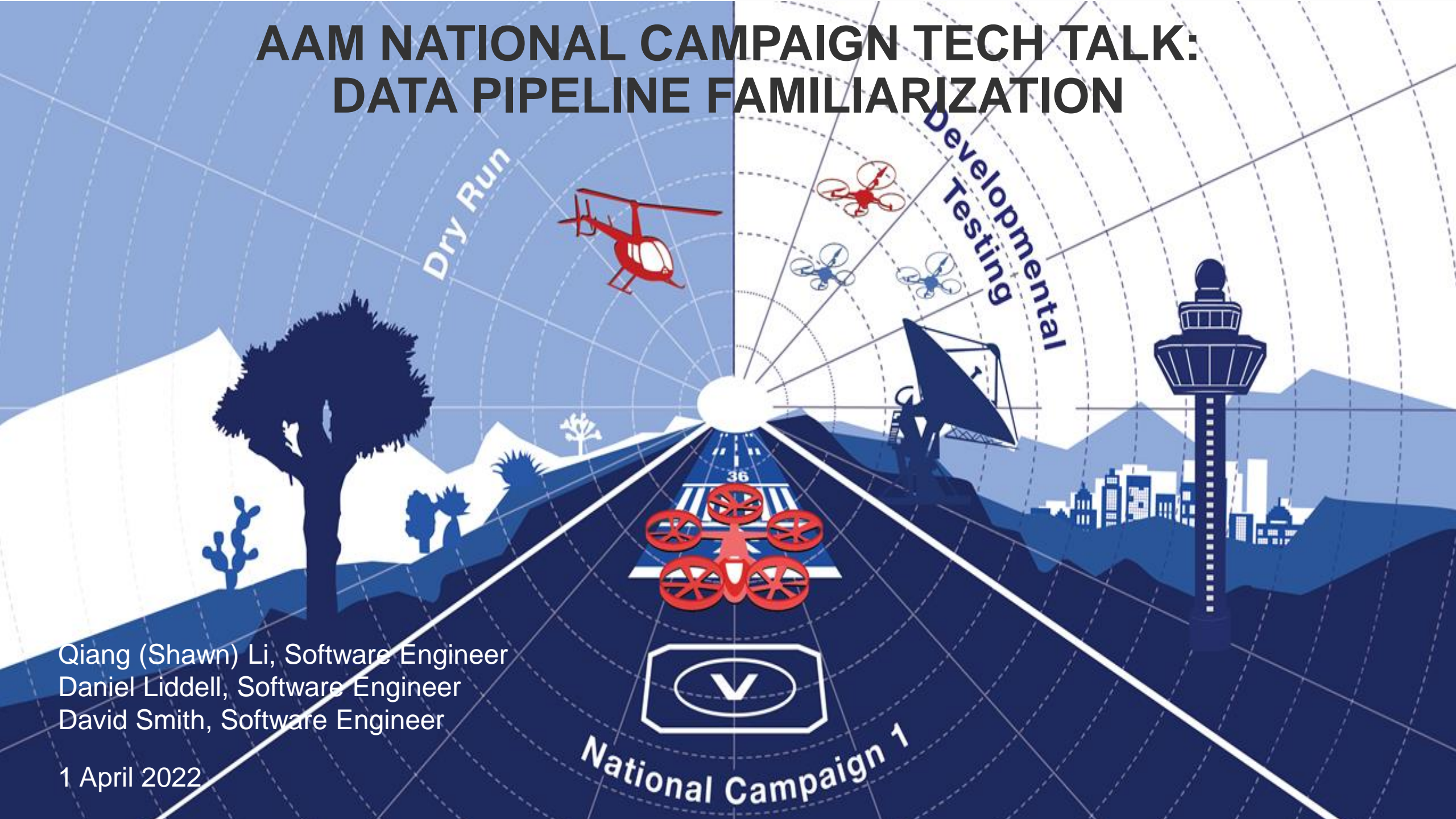
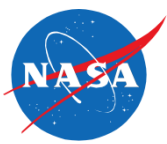


# AAM NATIONAL CAMPAIGN TECH TALK: DATA PIPELINE FAMILIARIZATION

Qiang (Shawn) Li, Software Engineer  
Daniel Liddell, Software Engineer  
David Smith, Software Engineer

1 April 2022





# Intro to Tech Talks

---

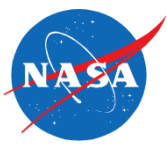
- Purpose of these talks is to engage with the community on types of technologies we are using and developing, there are many more planned
- Ground rules for talk
  - Answers to questions you have may be in upcoming slides
  - It's okay to ask an important question on a slide, but if it can wait then please do so
  - Mute your mike unless you need to talk
  - We'll keep an issues Parking Lot to keep the Tech Talk on point and on time
  - Remember the Tech Talk is being recorded for NASA and its Partners
- Recording. We are recording these Tech Talks and will post online, once approved for external release



# Outline

---

- AAM Project Background, Requirements, Collection, and Relations
- Data Pipeline And Its Functions
- Data Pipeline Major Cloud Components (all in AWS), Benefits
- Benefits of Cloud-Based Application
- Current ATI Data Pipeline Design/Implementation
- Data Pipeline Monitoring by AWS CloudWatch
- Results and Summary

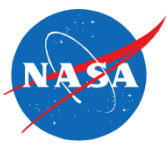


# AAM Project Background and Pipeline Motivation

---

- Data Pipeline initially emerged from lessons learned from UTM
  - Post-event data submission was tardy and manual
  - Data ingestion problems discovered post-event (too late)
  - Data quality or quantity problems discovered in analysis solved by renegotiation/resubmitting data, reducing data quality
  - Difficult to determine partner compliance with requirements

AWS-based Data Pipeline allows **real-time** data submission and ingestion, with **immediate** monitoring of data rate, coverage and ingestion quality. Problems are **immediately** discovered and can be corrected with **agility** by both Partners and NASA ***during the event***



# AAM Project Data Collection Requirements

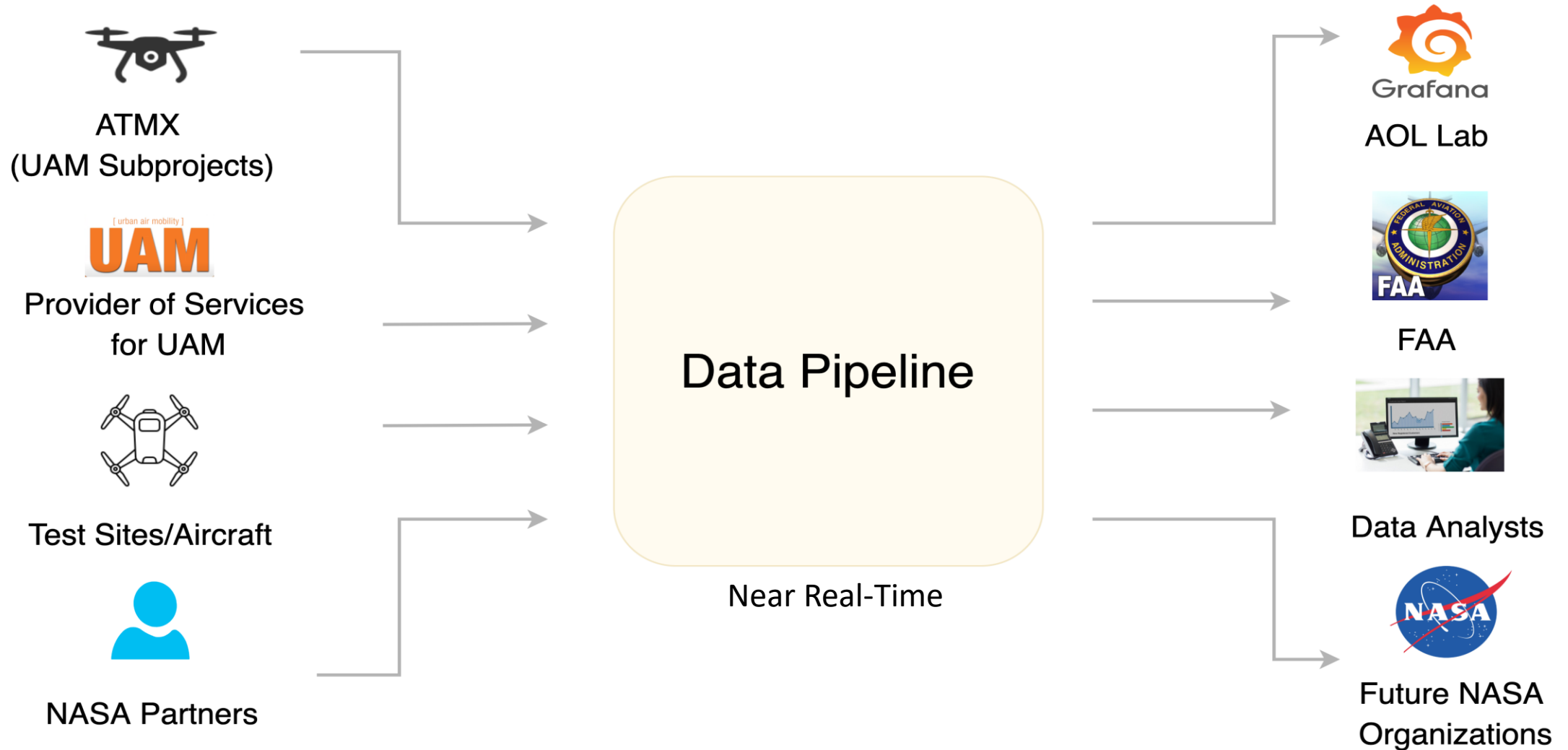
---

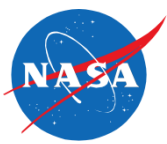
- ATI Data Pipeline must enable collection of many kinds of data (weather, sensor data from vehicles, airports, PSU data, etc.) securely, in a *wide range of formats*, and from different sources in *near real-time*
- *Data integrity* must be confirmed after receipt
- *Data protection* must be assured for the life of the data
- *Secure Access* from outside NASA's network must be supported
- *Near real-time integration* with different services is required
- High *reliability, availability, scalability* are required





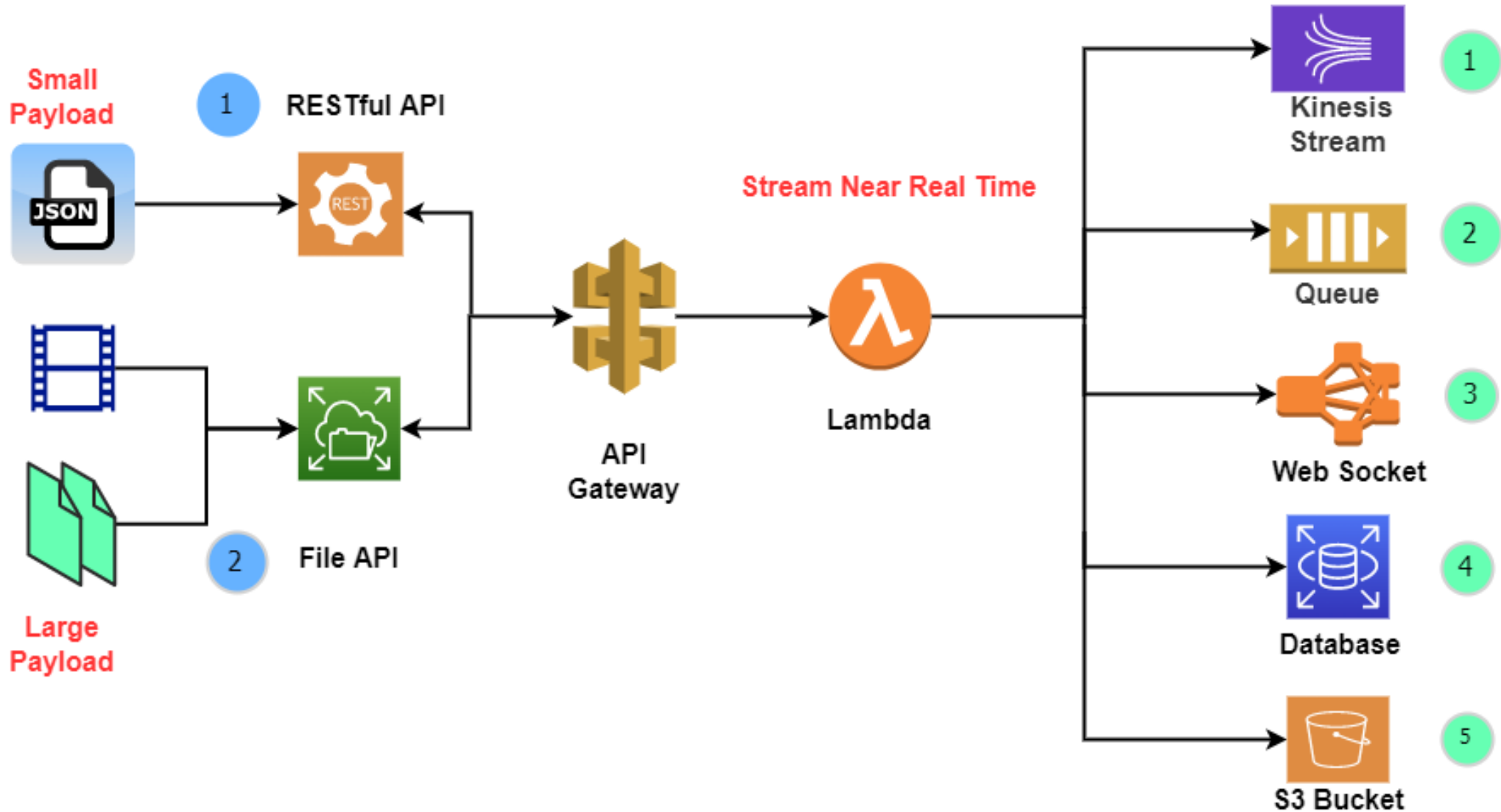
# AAM NC Data Pipeline Relation with Other Parties





# Data Pipeline And Its Functions

- Serverless Cloud collecting, storing, and distributing diverse sets of data in near real-time





# Data Pipeline Major Cloud Components



**API Gateway**  
API Endpoint and Service



**Cognito**  
User Identity Service



**DynamoDB**  
No-SQL Database



**Kinesis**  
Streaming Data Service



**Lambda**  
Serverless Compute Service



**Simple Storage Service**  
Shared Across Servers



**Simple Queue Service**  
Messaging Queue



**PostgreSQL Server**  
Traditional Database



**CloudWatch**  
Monitoring Services and Dashboard

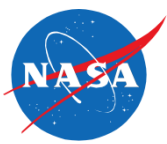




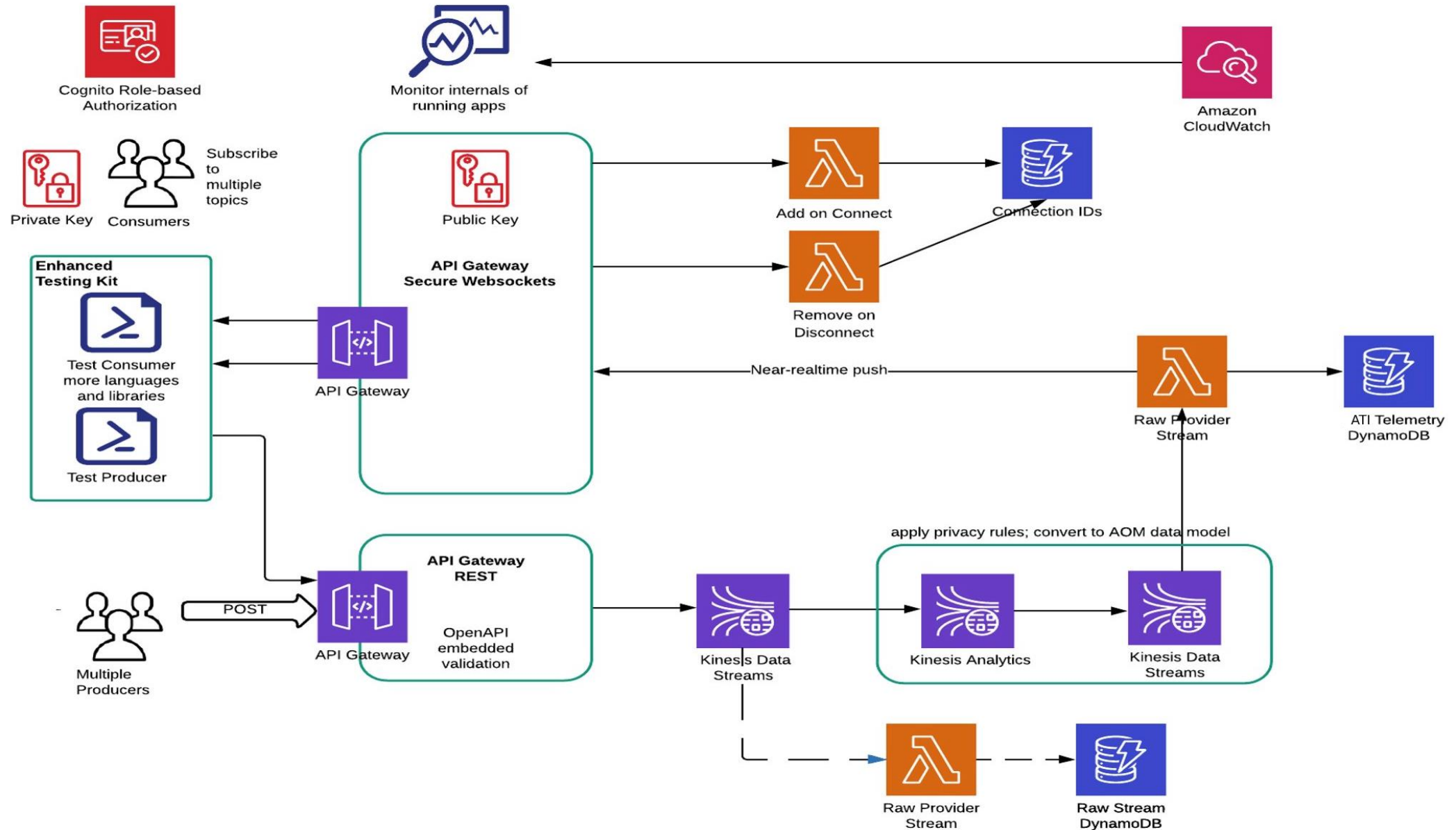
# Benefits of Cloud-Based Application

---

- Fully cloud-based – no on-premises hardware
- High availability
- High scalability
- Serverless means low maintenance and usage cost
- Lambda is pay-as-you-go with charges to millionths of a second
- No worrying about Operating System updates
- Built in support for DLQ (dead letter queue)
- High performance
- Restful endpoint got 500 RPS (Requests Per Second) by default--AWS quote



# Ingress Restful and Egress Web Socket Streaming Application





# Data Pipeline Infrastructure As Code (IAC)

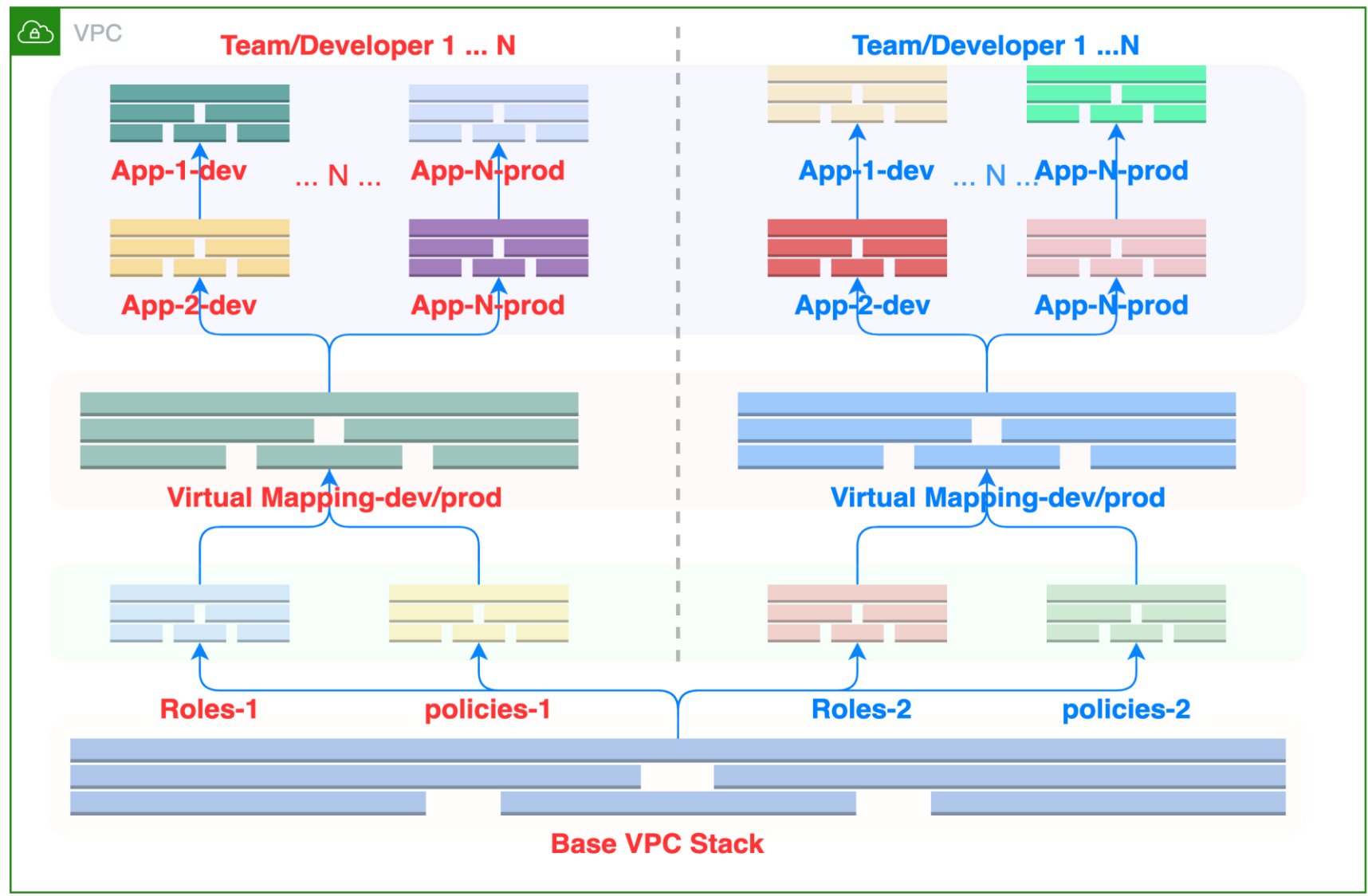
---

AWS CloudFormation lets you model, provision, and manage AWS and third-party resources by treating infrastructure as code

- **Manage infrastructure with DevOps**
  - Automate, test, and deploy infrastructure templates with continuous integration and delivery (CI/CD) automations
  - Version control automation 'scripts' as you would software
- **Scale production stacks**
- **Visualization application infrastructure**
- These capabilities are not readily available in traditional, on-premises software systems



# Data Pipeline AWS Multiple Environments And Developers Architecture



**Applications Stack Layer**

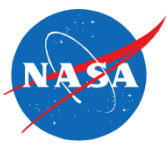
Support multiple environments (dev, production), developers and parallel applications

**Virtual Mapping Stack Layer**

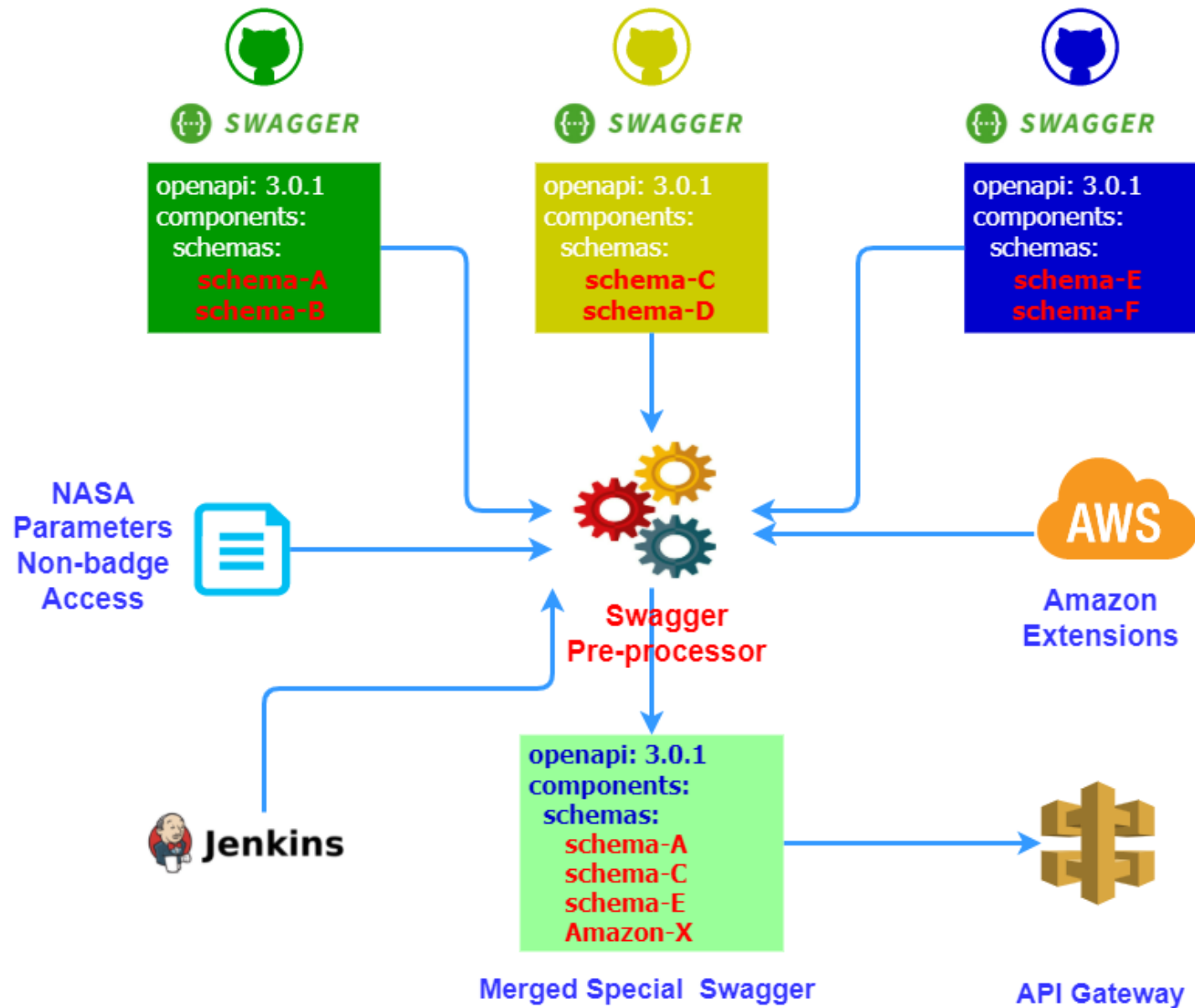
Decouple application layer with infrastructure layer by different mappings

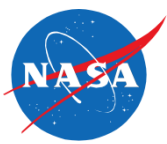
**Roles and Policies Stack Layer**

**VPC, Subnets, NAT Gateway and Security Groups Stack Layer**



# Data Pipeline Swagger Pre-processors





# Data Pipeline Demo using CloudFormation

CloudFormation > Stacks

## Stacks (17)

View nested

< 1 >

Stack name	Status	Created time	Description
<input type="radio"/> arc-af-aom-tdp-etl-sbox-demo	CREATE_COMPLETE	2022-01-24 13:23:05 UTC-0800	AOM consumer lambda stack
<input type="radio"/> arc-af-aom-tdp-web-socket-api-sbox-demo	CREATE_COMPLETE	2022-01-24 13:19:59 UTC-0800	AOM-Web-Socket-Egress-API-Gateway
<input type="radio"/> arc-af-aom-tdp-apigateway-sbox-demo	CREATE_COMPLETE	2022-01-24 13:13:16 UTC-0800	Data Pipeline telemetry data provider
<input type="radio"/> arc-af-aom-tdp-kinesis-sqs-sbox-demo	CREATE_COMPLETE	2022-01-24 12:55:04 UTC-0800	Data Pipeline telemetry data provider kinesis stream
<input type="radio"/> arc-af-aom-tdp-roles-policies-sbox-demo	CREATE_COMPLETE	2022-01-24 12:51:11 UTC-0800	AOM project role allow full access for API Gateway,DynamoDB,kinesis, S3,cloudwatch,VPCendpoint and lambdaExecution
<input type="radio"/> arc-af-aom-tdp-cognito-sbox	CREATE_COMPLETE	2022-01-23 20:12:27 UTC-0800	create aom cognito user pool and clients
<input type="radio"/> arc-af-aom-tdp-python-lib-layer	CREATE_COMPLETE	2022-01-16 23:03:39 UTC-0800	python library layer
			VM-base layer exports infrastructure





# Data Pipeline Security Considerations

## Networking Level Security

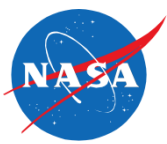
AWS VPC, Subnets, Security Group, Route table, NAT Gateway, etc which are controlled by NASA IT/EMCC

## AWS Identity and Access Management (IAM) Level

AWS IAM Roles and policies which control access of the different AWS resources  
Eg. Lambda connections only have read permission from connection table

## Application-Level Security

Access applications endpoints and API calls between applications. Eg. Cognito token is required for producer and consumer endpoints



# Results and Summary

---

- ATI Data Pipeline is a powerful platform for real-time data integration, for both AAM and ATM-X Projects
- Data Pipeline leverages advanced AWS capabilities to achieve requirements with far less software development and IT management than traditional approaches
- Data Pipeline benefits from NASA Cloud initiatives and is a pathfinder for them as well

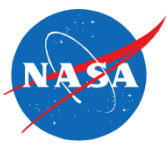


# Acknowledgements

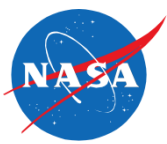
---

- Qiang (Shawn) Li (ARC-TI), KBR
- Daniel Liddell (ARC-AF), KBR
- David R. Smith (ARC-AF), KBR
- Michelle M. Eshow (ARC-AFS)[MOSAIC ATM INC]
- Dennis A. Rivera (ARC-AFO)
- Jody Null (ARC-AFO), USRA
- Lawrence Z. Markosian (ARC-TI), KBR
- Irene Smith (ARC-AFO)

# Questions



# APPENDIX



# Acronyms

---

AAM-Advanced Air Mobility

API-Application Programming Interface

ATI-Airspace Testing & Integration

AWS-Amazon Web Services

CI/CD-Continuous Integration and Delivery

DLQ-Dead Letter Queue

EMCC-NASA Enterprise Managed Cloud Computing

IAC-Infrastructure As Code

IAM-Identity Access and Management

IOT-Internet of Things

NAT-Network Address Translation

PSU-Provider of Services

RDS-Relational Database Service

RPS-Requests Per Second

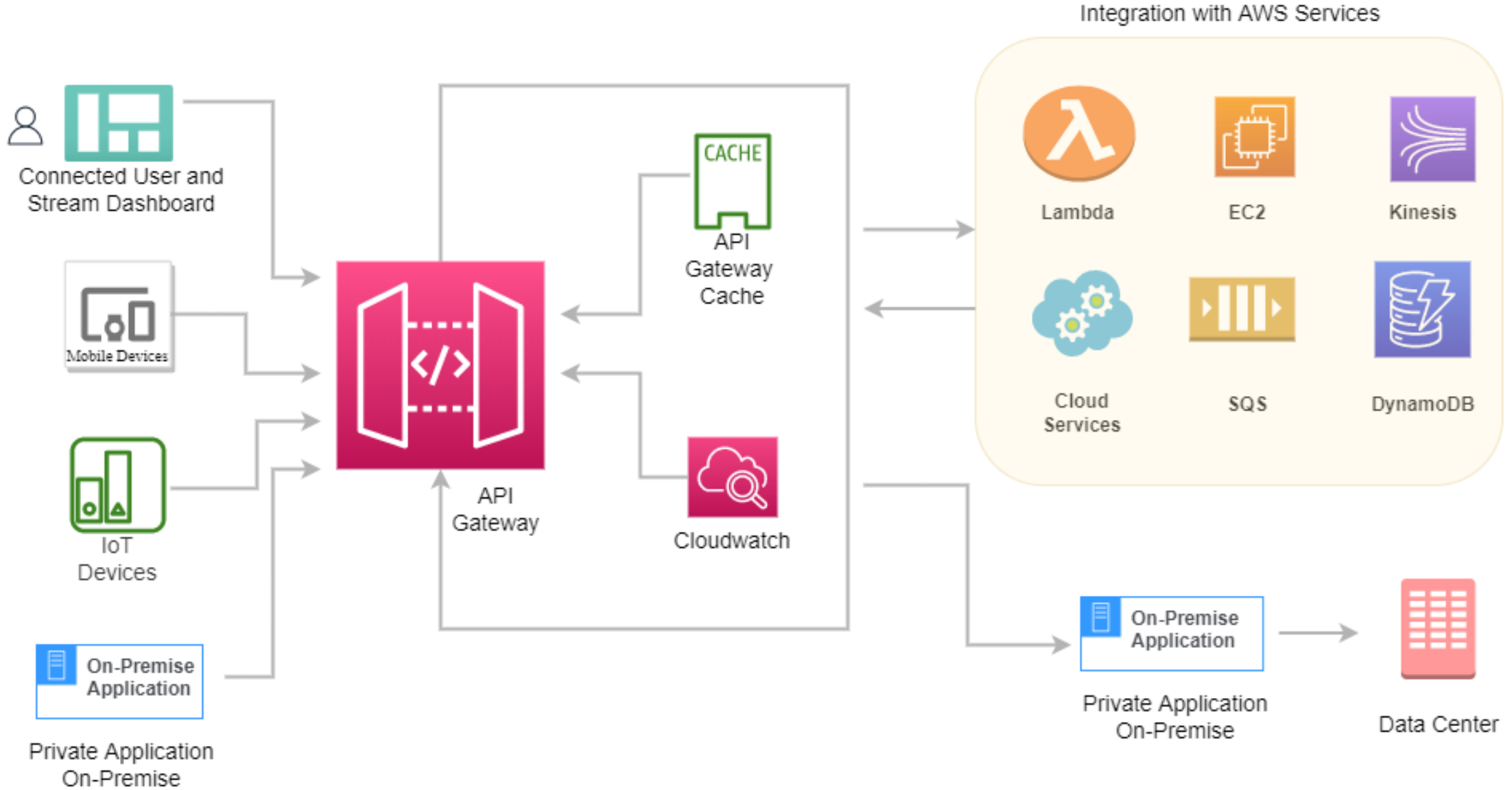
SQS-Sequel Server

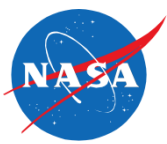
VPC-Virtual Private Cloud



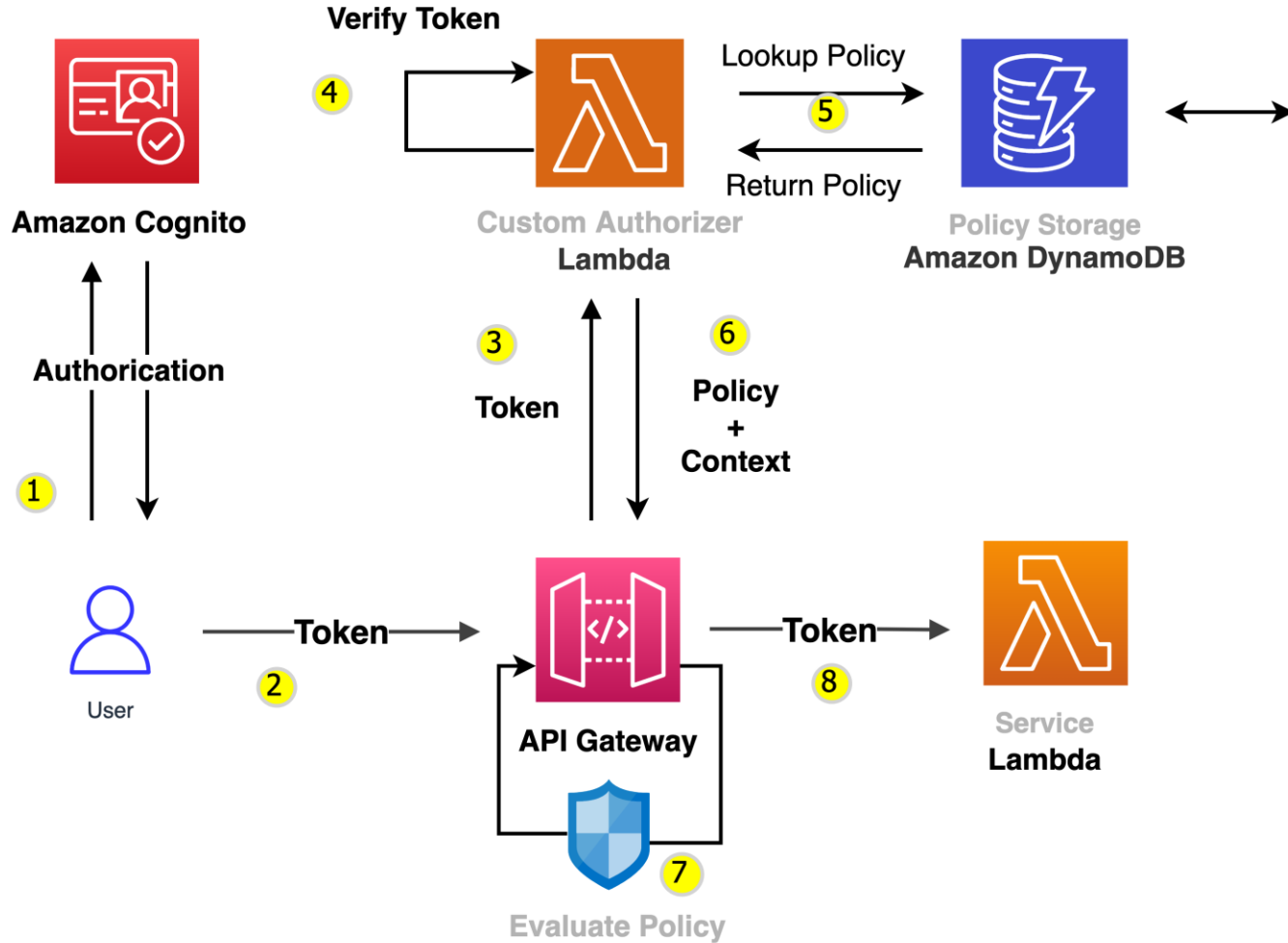


# AWS API Gateway





# Secure API Gateway with Cognito



```
{
  "principalId": "1tonibb18ftb94smrmi0f4lcf",
  "policyDocument": {
    "Version": "2012-10-17",
    "Statement": [
      {
        "Action": "execute-api:Invoke",
        "Effect": "Deny",
        "Resource": "arn:aws:execute-api:us-east-1:xxxxxxx:xhwci1bx2/prod/$connect"
      }
    ]
  }
}
```



# Data Pipeline Monitoring by AWS CloudWatch

## All requests

All HTTP requests made to this bucket.

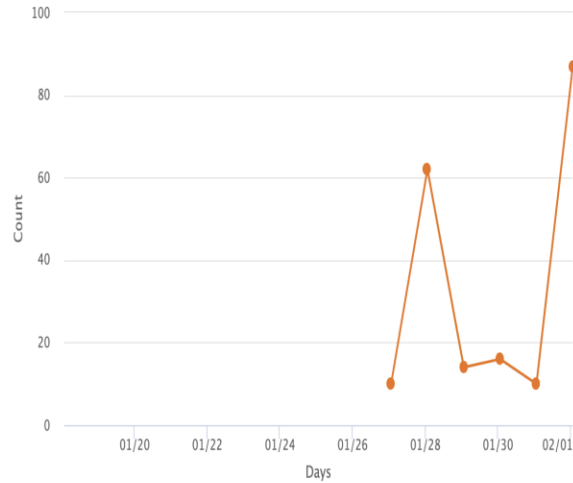
[View in CloudWatch](#)



## Get requests

HTTP GET requests for objects in bucket.

[View in CloudWatch](#)



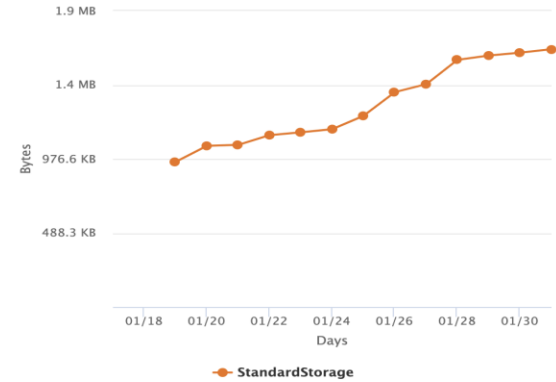
## Bucket metrics

Explore metrics for usage, request, and data transfer activity within your bucket. Metrics are also available in Amazon CloudWatch. [Learn more](#)

### Total bucket size

Amount of data in bytes stored in this bucket.

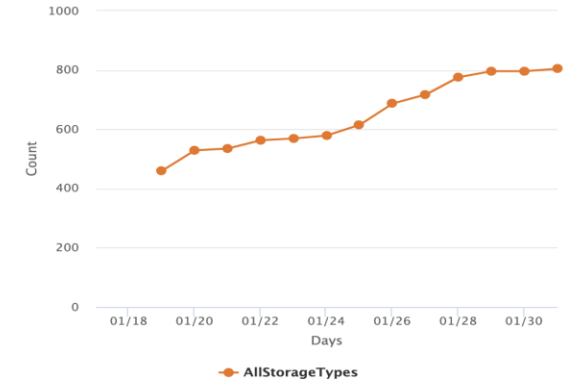
[View in CloudWatch](#)



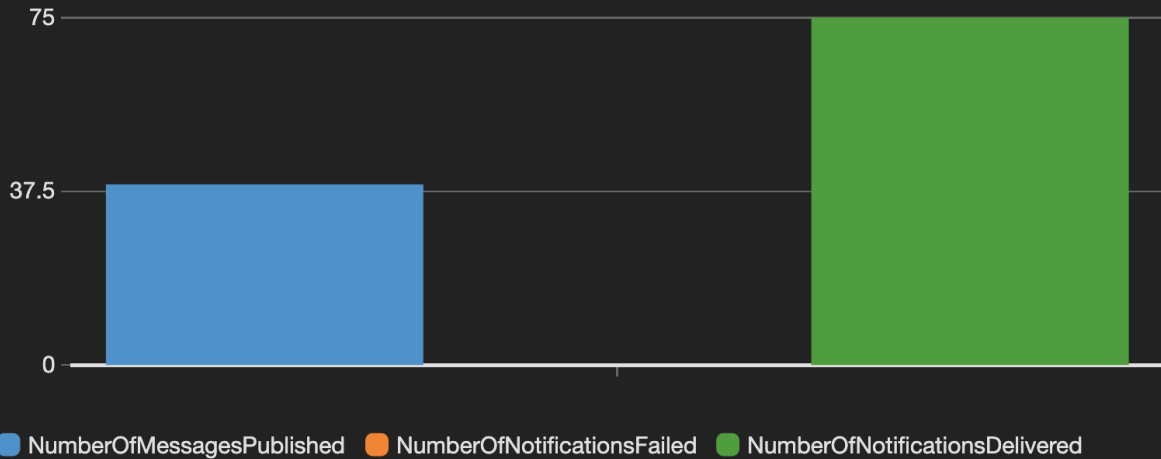
### Total number of objects

Total number of objects stored in this bucket for all storage classes.

[View in CloudWatch](#)



## SNS Published-Failed-Delivered



## SNS Published-Failed-Delivered Ratio



NumberOfNotificationsFailed

Success rate (%)