ANT207

What’s new in AWS streaming

Anusha Dasarakothapalli (she/her)

Principal Software Engineer, AWS Messaging and Streaming

Divya Sridhar (she/her)

Sr. Product Manager, Amazon Managed Streaming for Apache Kafka

ANT207

What's new in AWS Streaming

Speakers

- Divya Sridhar, Senior Product Manager Technical, Amazon
- Anusha Dasarakothapalli, Principal Software Engineer, Amazon

Announcements

- Lots of incremental announcement. See the What's New Section and 4th Slide in Summary.

Takeaways

- Focus on Offloading Capacity Management and Ease of Use.

Azure

- Streaming services on Azure are primarily categorized under the Messaging Services category.

< Section Header
< Table of Contents

Why Streaming

- Data is a perishable resource, streaming data real-time can help you identify problems before they occur.
- Common streaming-data use cases
  - Anomaly and Fraud Detection
  - Tailoring customer experience in real-time
  - Empowering IoT Analytics
  - Nourishing marketing campaigns
  - Real-time personalization
  - Supporting Healthcare and Emergency Services

Streaming Portfolio

- AWS Data Streaming, Processing and Analytics
  - Kinesis Data Streams
  - Kinesis Video Streams
  - Kinesis Data Firehose
  - Kinesis Data Analytics
  - Managed streaming for Apache Kafka

What's New

- Hands Off Capacity Management w/MSK Serverless
  - New – Use MSK Connect with MSK
  - New – MSK Serverless expanded Region Capability
  - New – Hands Off Capacity for MSK Tiered Storage
  - New – Additional Metrics for Kinesis Data Analytics for Apache Flink
- Ease of Use
  - New – Kinesis Data Streams Viewer in AWS Console
  - New – Improved Visibility in Kinesis Data Analytics for Apache Flink with Customized Mtx Windows & 5x Performance Increase in Logging
  - New – MSK Connect now supports private DNS
  - New – Move data directly from Kinesis and MSK to Redshift
  - New – Kinesis Data Analytics launches support for Apache Flink 1.15.2

Data streaming workflow

Source
Devices and or applications that produce real-time data at high velocity

Stream ingestion
Data from tens of thousands of data sources can be written to a single stream

Kinesis Data Streams
Amazon MSK

Stream transformation
Records are read in the order they are produced, enabling real-time analytics or streaming ETL

Kinesis Data Analytics
Kinesis Data Firehose

Stream destination
Data warehouse
Data store
Data lake
Analytical tool

Stream insights
Amazon Athena
Presto
Amazon QuickSight

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

Common streaming-data use cases

- Anomaly and fraud detection
- Tailoring customer experience in real time
- Empowering IoT analytics
- Nourishing marketing campaigns
- Real-time personalization
- Supporting healthcare and emergency services

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

Real-time data streaming, processing, and analytics on AWS

Amazon Kinesis Data Streams

Collect and store data streams for analytics

Amazon Kinesis Video Streams

Collect and store video streams for analytics

Amazon Kinesis Data Firehose

Load data streams into AWS date stores

Amazon Kinesis Data Analytics

Analyze data streams with Apache Flink

Amazon Managed Streaming for Apache Kafka

Fully managed, highly available and secure

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

What's new in AWS Streaming

Managed Streaming for Apache Kafka

- ✓ Support for tiered storage
- ✓ MSK Serverless support for MSK Connect
- ✓ MSK Serverless available in 11 AWS Regions
- ✓ MSK Connect support for private DNS
- ✓ Native integration with Amazon Redshift

Kinesis Data Streams

- ✓ Kinesis Data Streams Data Viewer in AWS Management Console
- ✓ Native integration with Amazon Redshift

Kinesis Data Firehose

- ✓ Kinesis Data Firehose supports OpenSearch Serverless as a destination
- ✓ Kinesis Data Firehose supports VPC Flow Logs as a source

Kinesis Data Analytics

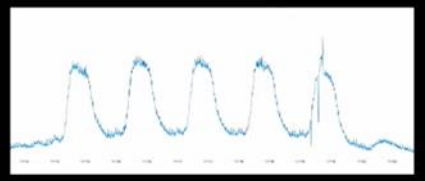
- ✓ Enhanced metrics
- ✓ Custom Maintenance Windows
- ✓ Improved application performance with logging
- ✓ Support for Apache Flink 1.15.2

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

Amazon MSK Serverless

EASILY STREAM DATA WITH AMAZON MSK WITHOUT MANAGING CLUSTER CAPACITY

"Our banking application gets used mostly during business hours. We have to scale our cluster up and down to stay cost-efficient."



"We have sales days every month, and we have to keep rescaling our clusters based on demand. Having an autoscaling feature will reduce our operational overhead."

Customer

- On-demand streaming capacity
- Fully compatible
- Throughput based pricing
- Same security as Amazon MSK
- Auto partition management
- Same high availability

New

- Use Amazon MSK Connect with Amazon MSK
- MSK Serverless expanded Region availability

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

Amazon MSK support for tiered storage

New

VIRTUALLY UNLIMITED AND COST-EFFECTIVE STORAGE TIER FOR LONG-LIVED DATA IN APACHE KAFKA APPLICATIONS

Tiered storage provides a new low-cost option that scales to virtually unlimited storage

- Increase your topics' storage without needing to increase the number of brokers
- Reprocess old data in the exact same order in which it was produced
- Save with a low-cost storage tier

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

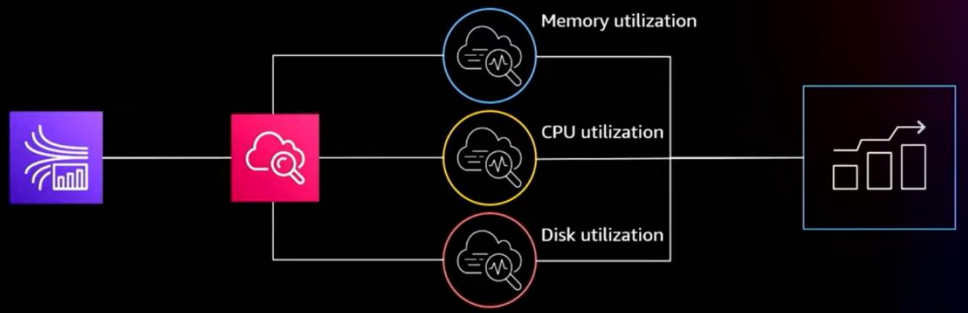
Kinesis Data Analytics for Apache Flink enhanced metrics

New

Kinesis Data Analytics for Apache Flink task managers

Amazon CloudWatch

Use metrics to scale Kinesis Data Analytics Apache Flink applications



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

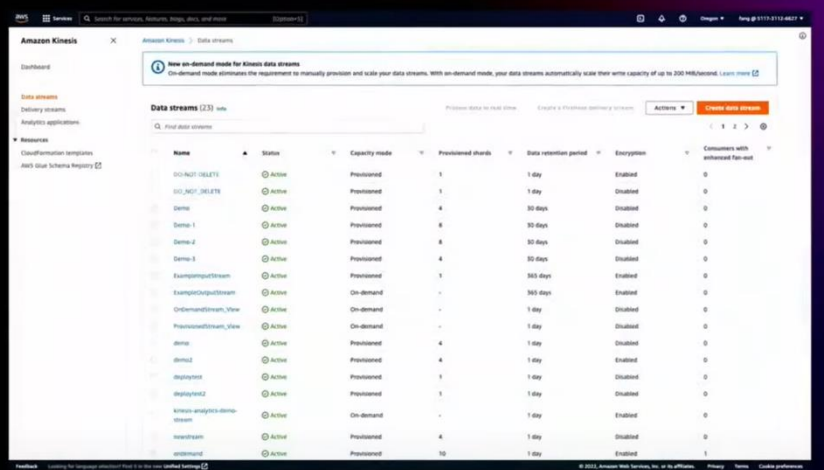
Kinesis Data Streams data viewer in AWS Console

New

INTERACTIVE, NO-CODE DATA EXPLORER

Use cases

1. Preview data quickly
2. Understand record schema
3. Data debugging



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

Kinesis Data Analytics for Apache Flink: Improved usability

New

Customize maintenance windows

5x improvement in application performance with logging turned on

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

Amazon MSK Connect

EASILY MOVE DATA BETWEEN APACHE KAFKA CLUSTERS AND OTHER SYSTEMS

Fully managed | Fully compatible | Serverless | Autoscaling

New

Amazon MSK Connect private DNS support

- ✓ Keep databases, data warehouses, and other systems like the secrets manager in your own VPC, inaccessible to the internet
- ✓ Use Amazon MSK Connect with Amazon MSK Serverless

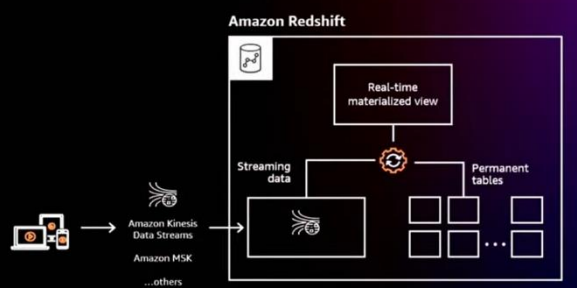
aws © 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Move data from MSK and Kinesis directly to Amazon Redshift

New

NATIVE INTEGRATION WITH STREAMING ENGINES FOR FAST AND EASY DATA INGESTION

- Create materialized views on top of streams directly
- Achieve low latency in the seconds
- Support up to hundreds of MBps of throughput
- Entirely configurable with SQL
- Define a schema or ingest semi-structured data with SUPER & VARBYTE data types



Learn more: https://go.aws/3eG5kZF

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

Amazon Kinesis Data Analytics launches support for Apache Flink version 1.15.2

Kinesis Data Analytics championed Apache Flink Version 1.15.2 release

Focus feature: Async Sink, an AWS contributed feature

A new framework allowing customers to easily create custom sinks

AWS facilitating community project to create a DynamoDB sink using the framework

Build a custom sink in under 20 minutes: https://bit.ly/3VyAkeJ

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