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

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

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

Agenda

Why streaming?

AWS streaming portfolio

What’s new in streaming?

What’s next in streaming?

Q&A

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

Customers using AWS Streaming solutions



Runs business analytics operations using Amazon Kinesis services



Migrated data bus from Self-Managed Apache Kafka to Kinesis



1 billion events per week from connected devices



Real-time personalization insights using Kinesis



> 1 billion travel searches ingested daily



350+ streaming applications on MSK



Runs an event-driven architecture using Amazon MSK



End-end ingestion pipeline on MSK



Processing > 6 billion events per day on Amazon MSK



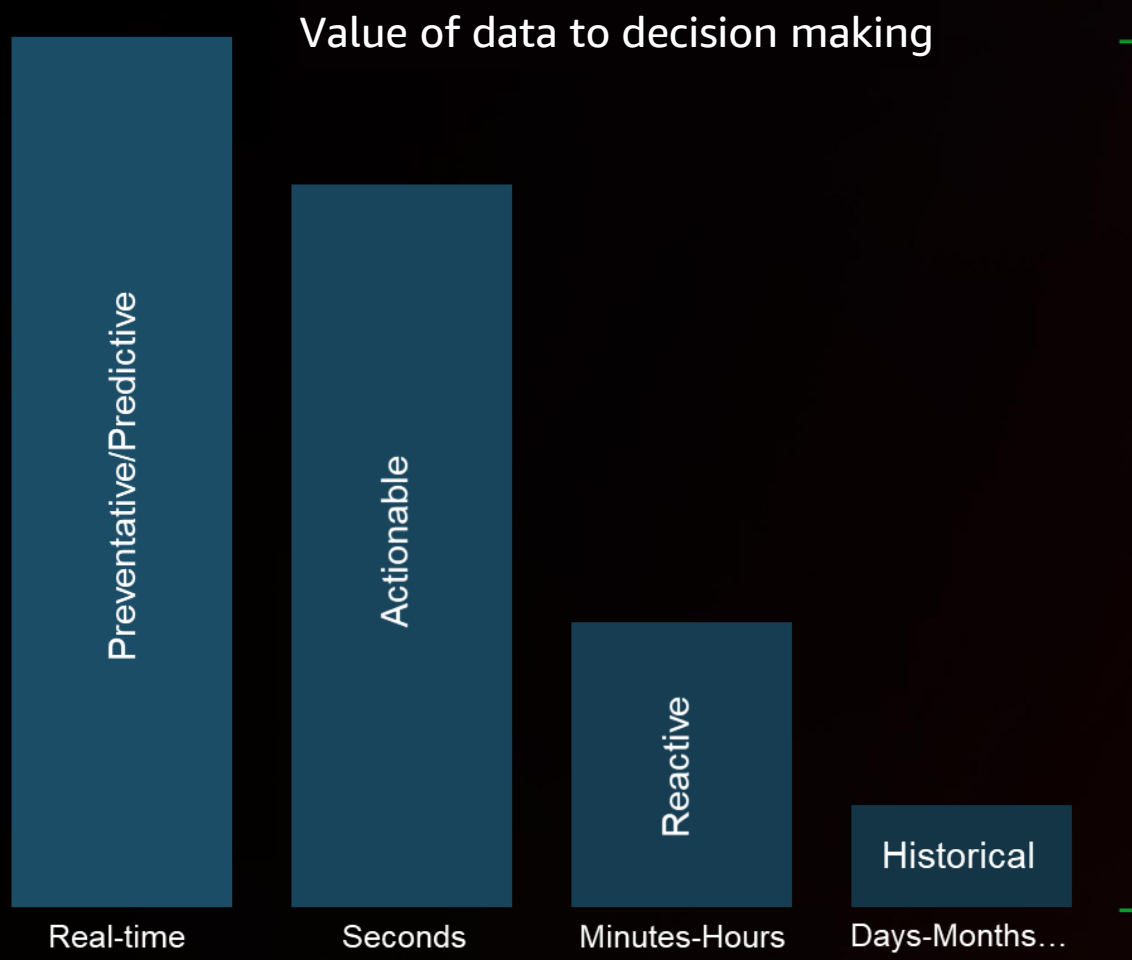
Facilitate communications between 100+ microservices

aws

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

The value of data diminishes over time

Data has a short life of actionability; AWS lets you act on that data as fast as the market dictates

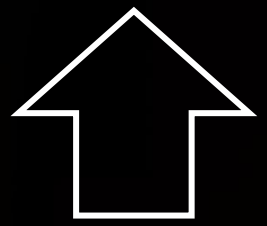


58% Percentage of those surveyed who say their organization saw significant increases in customer retention and loyalty as result of using real-time customer analytics

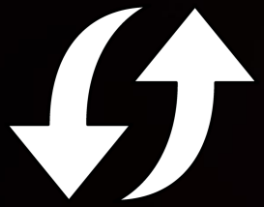
75% Percentage of enterprise organizations surveyed that believe untimely data inhibited business opportunities

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

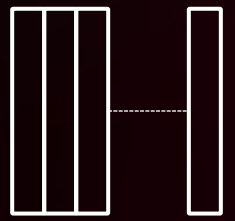
What is streaming data?



High volume



Continuous



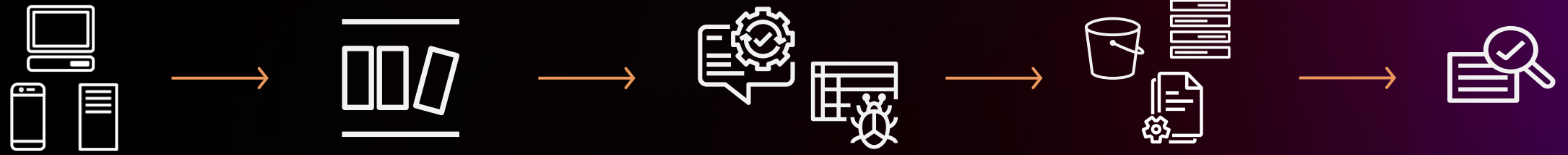
Ordered, incremental



Low latency

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

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.

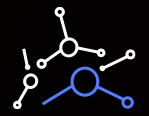
Common streaming-data use cases



Anomaly and fraud detection



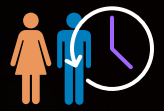
Tailoring customer experience in real time



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

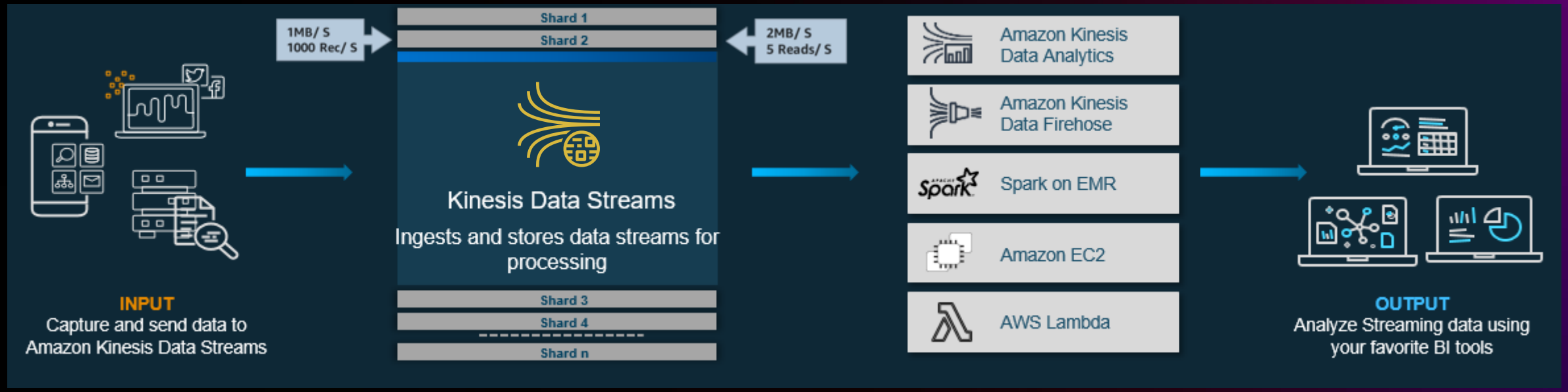


Fully managed, highly available and secure

aws

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

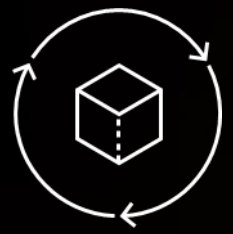
Kinesis Data Streams



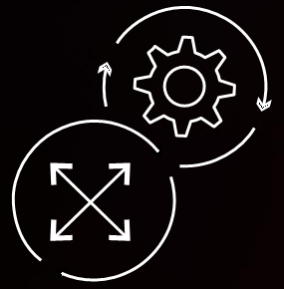
- Secure, durable storage
- Available to multiple real-time analytics applications
- Average latency of 200 milliseconds with one standard consumer
- Enhanced fan out offers typical average latency of 70 milliseconds

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

Kinesis Data Streams on demand



Hands-free capacity management



Seamless



Like always, same high performance, availability, and durability



Pay for what you use

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

Kinesis Data Firehose overview

MANAGED, NEAR REAL-TIME STREAMING ETL

Extract
Sources (20+)

- AWS SDK: direct put
- Kinesis Data Streams
- 20+ AWS services
  (Amazon CloudWatch, AWS IoT, AWS WAF, Amazon SNS, VPC Flow Logs, and so on)
- Amazon Kinesis Agent
- Open-source agents
  (such as Fluentd, FluentBit, Apache Nifi)

Transform

AWS Lambda custom transformation

Built-in transformation

- Parquet, ORC
- Amazon S3 dynamic partitioning
- Record delimiters
- Compression
- Buffer and batch

Load
Outputs (15+)

AWS destination

- Amazon S3
- Amazon Redshift
- Amazon OpenSearch Service
- Amazon API Gateway

Third-party integration destination

- Splunk
- HTTP endpoint
  (8+ AWS Partners, including Datadog)

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

Why Kinesis Data Firehose?

Easy to use	No code required	Serverless with elastic scale	Low cost
With just a few clicks, create your delivery stream Connect to your destination in under a minute	Handles data delivery failures Built-in transformations 30+ sources and destinations	Firehose automatically provisions, manages, and scales resources with no ongoing administration required	Data can be prepared and delivered at pennies per delivered GB with any scale

aws

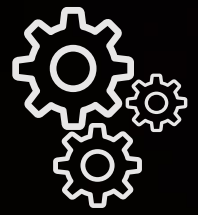
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Typical use cases for Kinesis Data Analytics for Apache Flink



Real-time analytics

- Key business, operational, and customer experience metrics
- Refresh dashboard content



Streaming ETL applications

- Transform, aggregate, filter streaming data
- IoT and log analysis



Stateful event processing

- Trigger conditions and external actions
- Patterns and anomaly detection

aws

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

Easily transform and analyse data in real time



Kinesis Data Analytics

Rapidly create, scale & inspect sophisticated applications

Multiple sources



Amazon Kinesis Data Streams



Amazon MSK



Amazon MQ



Amazon S3

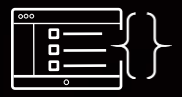


Custom connectors



Additional streaming sources





KDA studio
SQL/Python/Scala





Apache Flink
Stateful, real-time, low latency processing

Fully managed, serverless infrastructure

Multiple sinks



Amazon Kinesis Data Streams



Amazon MSK



Kinesis Data Firehose



Amazon OpenSearch Service



JDBC end points



Amazon S3

Output





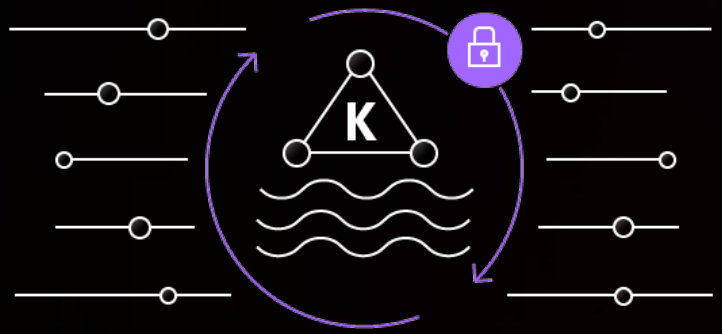
Send processed data to analytics tools so you can create alerts and respond in real time

aws

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

Amazon MSK

FULLY MANAGED SERVICE FOR APACHE KAFKA AND KAFKA CONNECT



Raw Apache Kafka performance at any scale

Deep AWS integrations

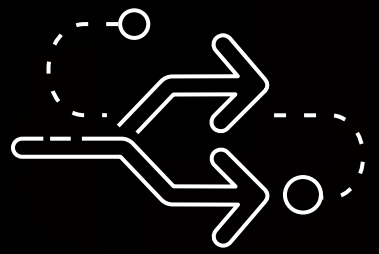
Up to 60% lower cost than next managed Kafka provider

Run your existing Apache Kafka applications without code changes

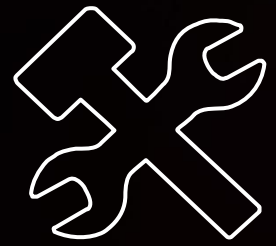
aws

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

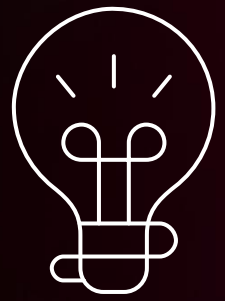
What Amazon MSK does for you



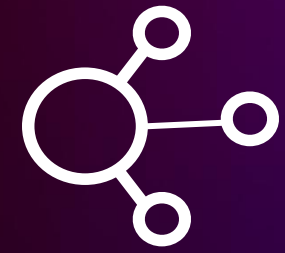
Ease of use and access



Best practices and automation



Focus on what matters



Contribute to Apache Kafka

aws

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

Self managed vs. fully managed

On premises	Amazon EC2	Amazon MSK
App dev/optimization	App dev/optimization	App dev/optimization
Scaling	Scaling	Scaling*
High availability	High availability	High availability
Kafka install/patching	Kafka install/patching	Kafka install/patching
Rolling version upgrades	Rolling version upgrades	Rolling version upgrades
Broker/ZK maintenance	Broker/ZK maintenance	Broker/ZK maintenance
Within-cluster data transfer cost	Within-cluster data transfer cost	Within-cluster data transfer cost
Encryption	Encryption	Encryption
OS patching	OS patching	OS patching
OS install	OS install	OS install
Hardware maintenance	Hardware maintenance	Hardware maintenance
Hardware lifecycle	Hardware lifecycle	Hardware lifecycle
Power/network/HVAC	Power/network/HVAC	Power/network/HVAC

Self-managed Kafka

AWS managed

More focus on creating streaming applications than managing infrastructure

aws

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

What’s new in streaming?

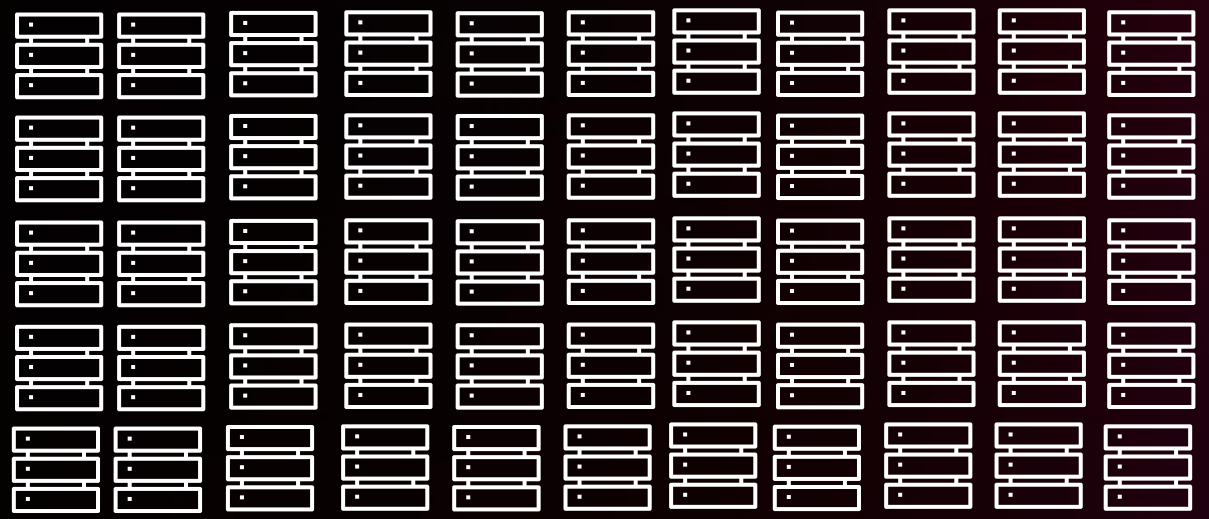
- Hands-off capacity management
- Ease of use
- Deep AWS integrations
- Open source contributions

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

1: Hands-off capacity management

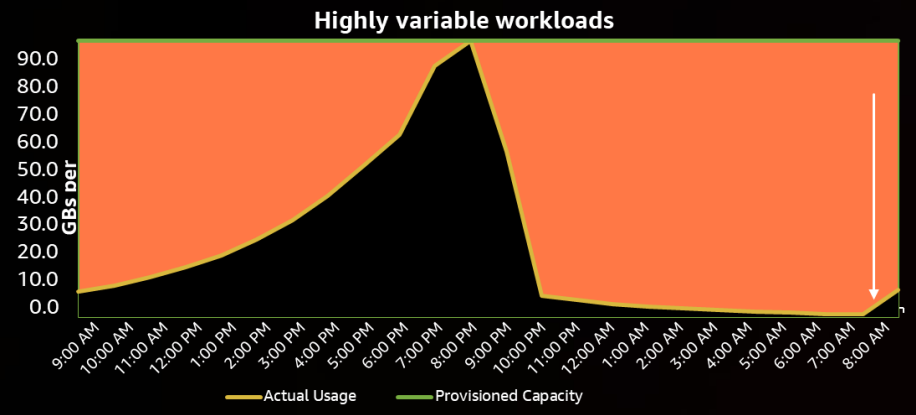
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Customers want to offload capacity management

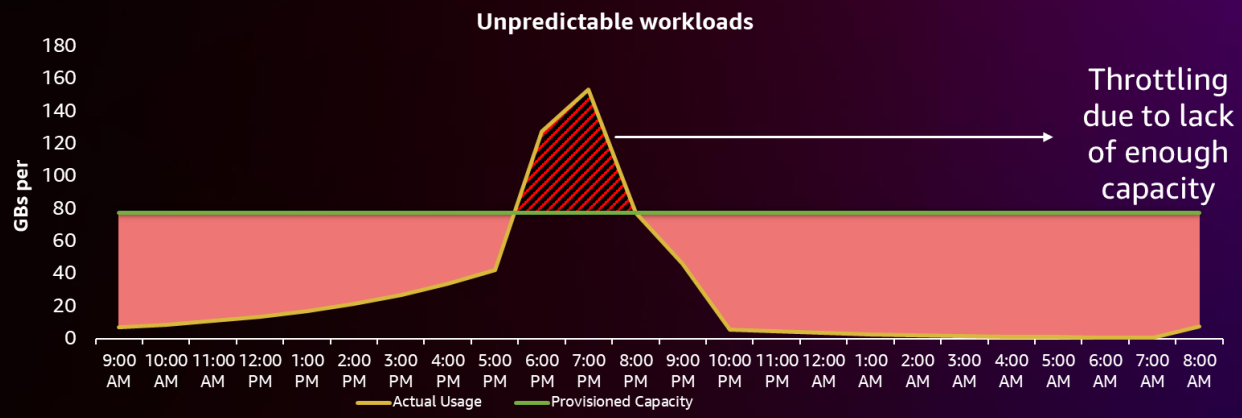


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Highly variable workloads



Unpredictable workloads

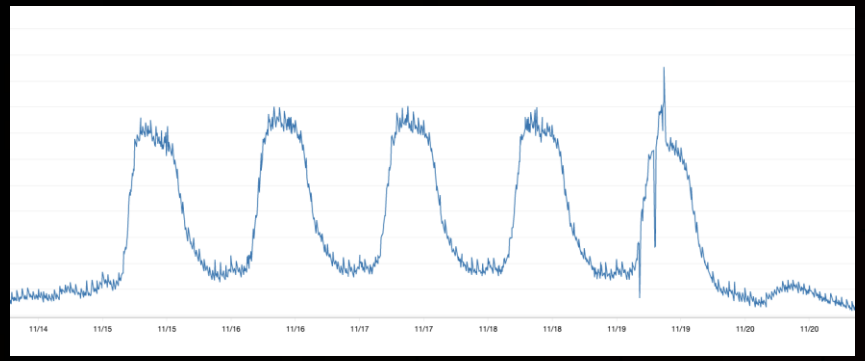


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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.

Choosing between provisioned and serverless offerings across MSK and Kinesis Data Streams

Amazon MSK Serverless/ Kinesis Data Streams on demand	Amazon MSK and Kinesis Data Streams provisioned
Unpredictable traffic	Predictable traffic
Prefer the ease of hands-free management	When you want to have fine-grained control
Lower cost of ownership	Predictable costs

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

Hands-off storage capacity management with Amazon MSK

© 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

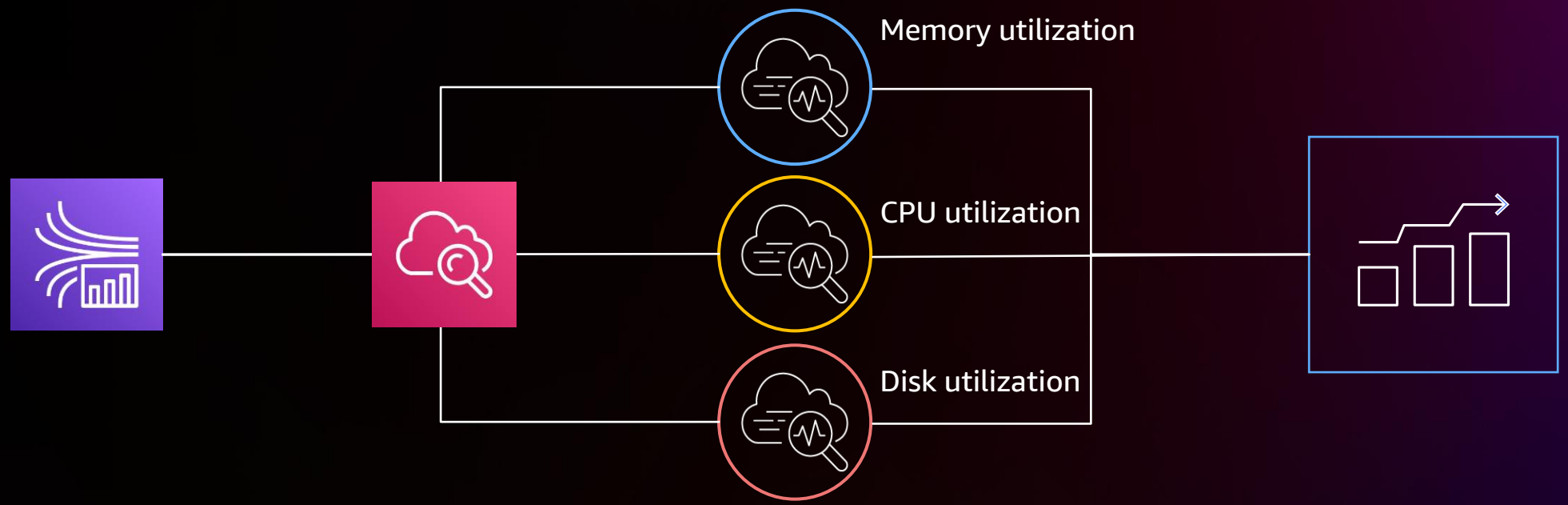
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Hands-off capacity management with Kinesis Data Analytics

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

Kinesis Data Analytics for Apache Flink enhanced metrics

New



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

2: Ease of use

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

Better observability

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

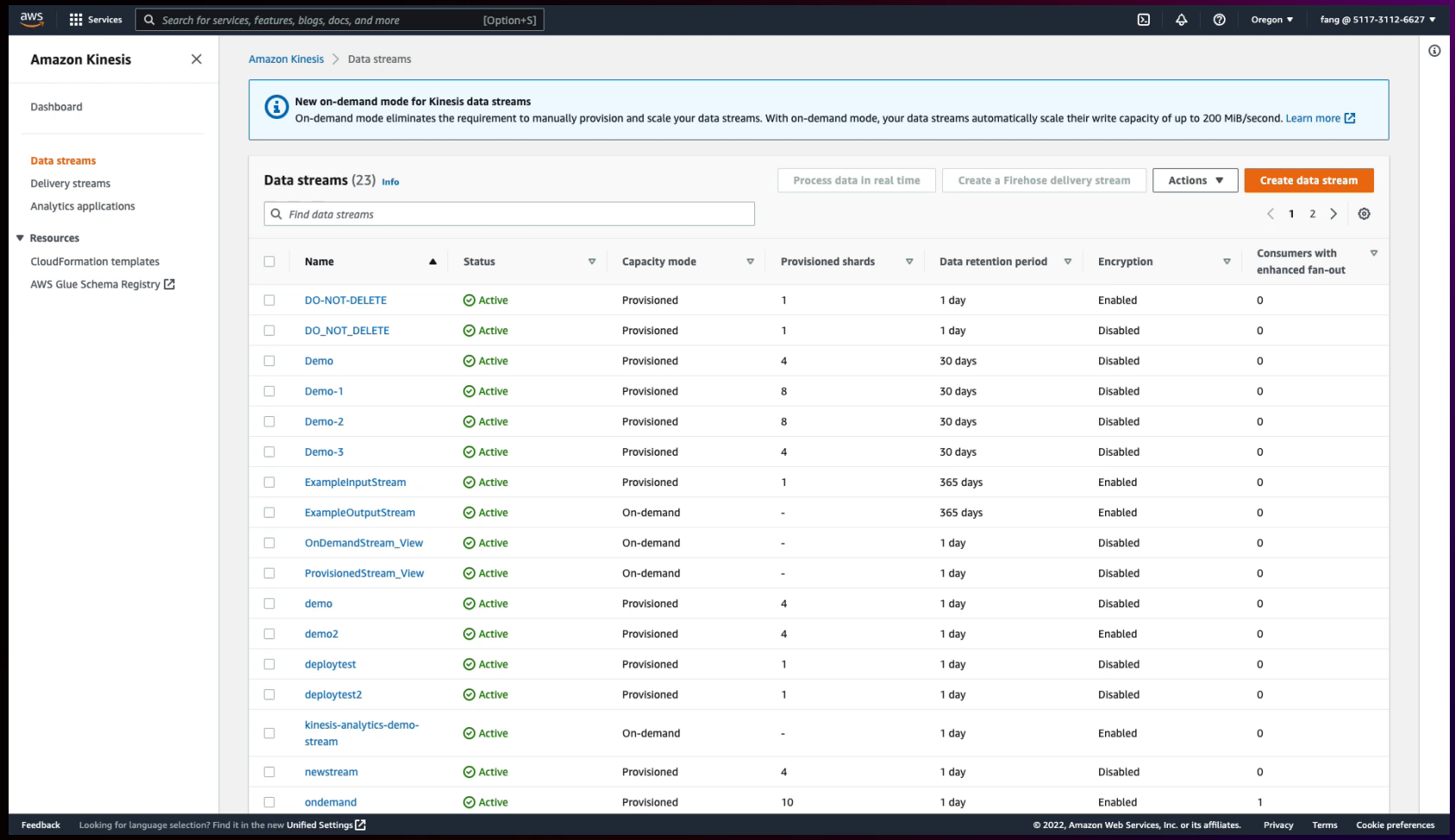
Kinesis Data Streams data viewer in AWS Console

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.

Improved availability

© 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.

Improved performance

© 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

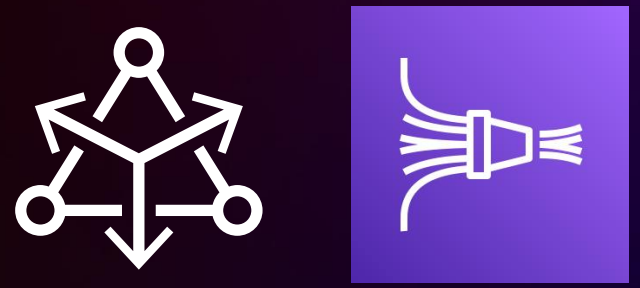
aws

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

3: Deep AWS integrations

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

Customers want an easy way to move data



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

Kinesis Data Firehose now supports VPC Flow Logs and OpenSearch Serverless

Extract
Sources (20+)

Streaming as a source



Amazon Kinesis Data Streams

Direct PUT



AWS SDK

- Customer Application Producers (e.g., mobile app)
- 20+ AWS services (e.g., AWS IoT, CloudWatch, AWS WAF, SNS, **VPC Flow Logs**)
- Agents (e.g., Kinesis Agent, Fluentd, FluentBit, Apache Nifi)

New

Transform

Lambda Custom Transformation



Built-in Transformation



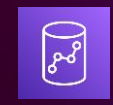
- Parquet, ORC
- S3 Dynamic Partitioning
- Record Delimiters
- Compression
- Buffer and batch

Load
Outputs (15+)

AWS destination



Amazon S3



Amazon Redshift



Amazon OpenSearch and **Amazon OpenSearch Serverless**

New



Amazon API Gateway

Third-party integration destination



Splunk

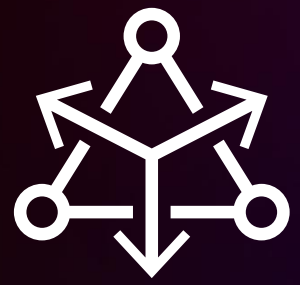


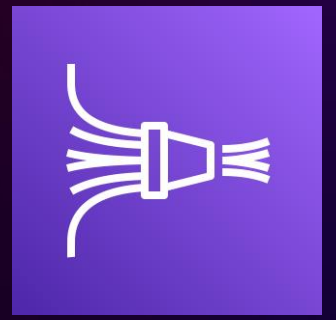
HTTP endpoint
(8+ partners such as Datadog, etc…)

aws

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

Customers want an easy way to move data





© 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

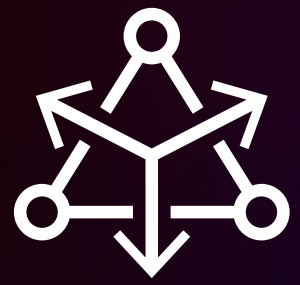
New

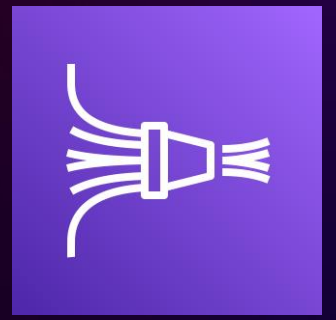
Amazon MSK Connect private DNS support

- ✓ 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

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

Customers want an easy way to move data





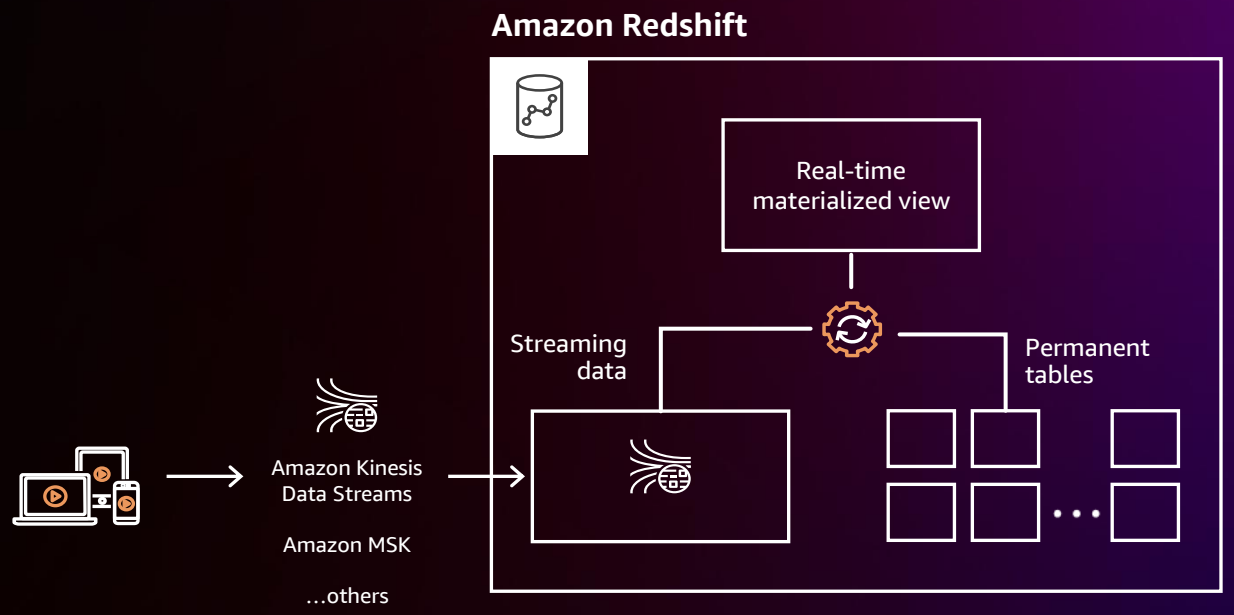
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

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.

4: Open-source contributions

© 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.

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

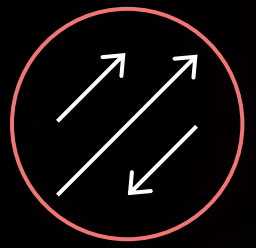


- ✓ Enhanced metrics
- ✓ Custom Maintenance Windows
- ✓ Improved application performance with logging
- ✓ Support for Apache Flink 1.15.2

aws

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

What’s next in streaming?



More scalability



Improved cost performance



Operations and ease of use



Resiliency at scale

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

Resource links

Managed Streaming for Apache Kafka



Kinesis Data Streams



Kinesis Data Firehose



Kinesis Data Analytics



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

Thank you!

Anusha Dasarakothapalli
linkedin.com/in/anusha-dasarakothapalli

Divya Sridhar
linkedin.com/in/sridhardivya



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

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