



Build a Serverless Analytics Framework

Bob Maus

Head, WW Data & Streaming GTM Team

AWS Analytics WWSO

Put data to work



Make better
decisions



Improve
efficiencies



Respond
faster



Uncover
opportunities

Most comprehensive set of services for the entire end-to-end data, analytics, and ML journey for all workloads and all types of data

Ingest + store

Analyze + visualize

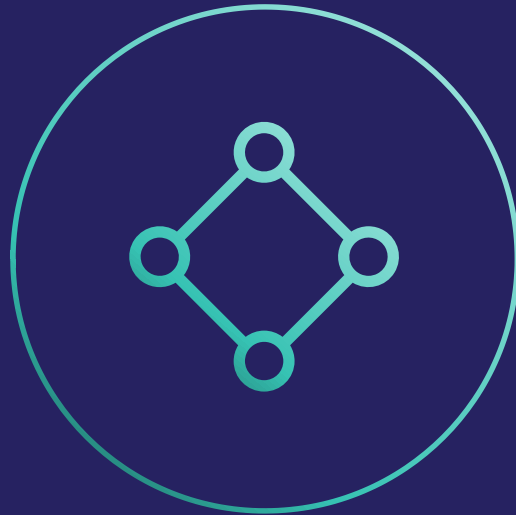
Predict



**Security +
Governance +
Access control**



MODERNIZE



UNIFY



INNOVATE



Serverless
and easy to use

Why go Serverless?



Simple to use



Innovate
Faster



Automatically
Scale



Cost-effective

AWS Analytics Goes Serverless

NEW
as of
reinvent
2021



AMAZON
REDSHIFT

Data
warehousing

NEW
as of
reinvent
2021



AMAZON
EMR

Big data
processing

NEW
as of
reinvent
2021



AMAZON
KINESIS

Real-time
analytics

NEW
as of
reinvent
2021



AMAZON
MSK

Real-time
analytics

AVAILABLE IN PREVIEW!

NEW

Amazon Redshift Serverless



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

New **serverless** **option** in Amazon Redshift

Focus on insights

YOU

aws

Manage and scale the
data warehouse



Serverless brings together all the benefits of Amazon Redshift with an easy-to-start experience



No compromises

SQL features

Performance at scale

Scalability

Analyze all your data

Security

Easy to get started



Developer and business analyst

Activate Amazon Redshift Serverless for your AWS account



BI tool

Connect from your favorite BI tool or Amazon Redshift Query Editor



Amazon Redshift

Load data seamlessly and Amazon Redshift executes queries by automatically provisioning capacity



Pay

Pay for compute and storage used during analysis



Roche

The largest pharmaceutical company in the world and the leading provider of cancer treatments globally

“ Amazon Redshift Serverless helps us complete our data management without having to manage clusters and optimizes our cost by provisioning just the right amount of capacity to meet demand. Amazon Redshift Serverless is reducing the operational burden, lowering costs, and enabling scale for the Roche Go-to-Market domain. This simplification is a game changer, helping us rapidly onboard and support a variety of analytics-heavy use cases without friction. ”

Dr. Yannick Misteli

Lead Cloud Platform & ML Engineer, Roche



Roche

Benefits of Amazon Redshift Serverless

Simplified user experience



Run and scale analytics without having to manage data warehouse clusters

All Redshift functionality and performance



Leverage Amazon Redshift's rich SQL capabilities, seamless data lake integration, as well as industry-leading price performance at scale

Intelligent and dynamic compute



Automatically provisions and scales data warehouse capacity to deliver consistently fast performance

Pay for use



Pay for the compute capacity only for the workload duration on a per-second basis. No charges for idleness

AVAILABLE IN PREVIEW!

NEW

Amazon **EMR Serverless**

Run petabyte-scale data analytics in the cloud without managing and operating clusters

aws.amazon.com/emr/serverless



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

NEW
PREVIEW

NEW

Amazon EMR Serverless



Simple to use

No servers to manage. Amazon EMR Serverless provisions, configures, and dynamically scales the compute and memory resources needed at each stage of your data processing application.



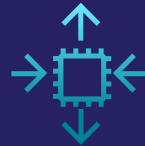
Fast

Performance optimized runtime that is compatible with and over 2X faster than standard open source



Cost effective

Pay only for the compute time and resources that you use.



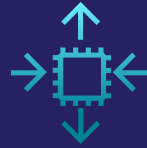
Comprehensive

Includes Amazon EMR Studio with Notebooks and familiar open source tools to easily develop, visualize, and debug applications



Amazon EMR Serverless

Use cases



Build scalable data pipelines

Extract and transform data at scale from a variety of sources



Accelerate data science and ML

Develop, debug, and visualize big data applications with EMR Studio



Process real-time data streams

Analyze events from streaming data sources in real time



Query any dataset

Query datasets interactively using Spark, Hive, and Presto

Benefits of Amazon EMR Serverless



No clusters to manage



Interactive data analytics and machine learning



Super simple to use



Bring your own customizations

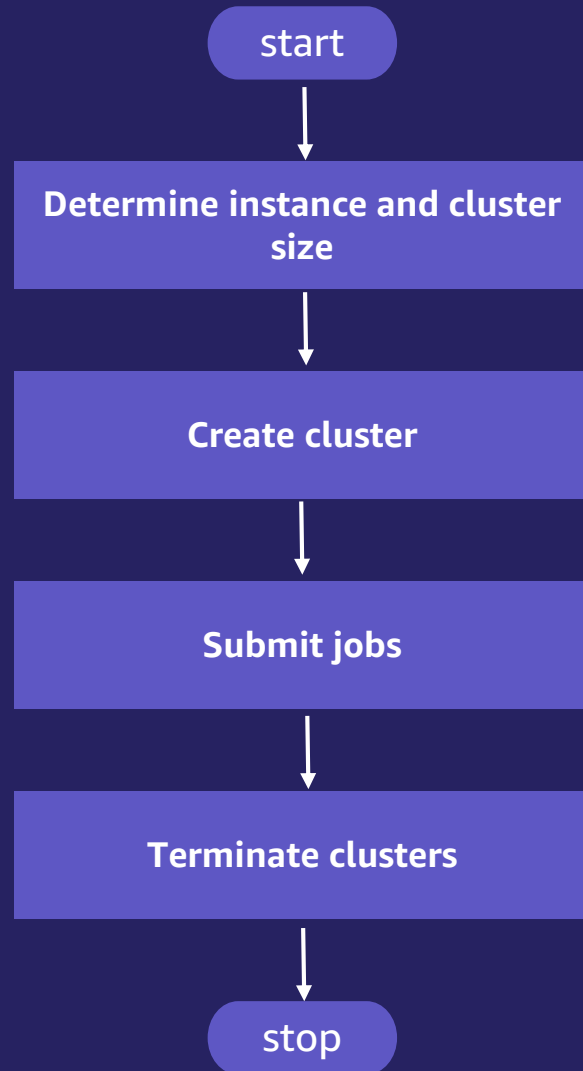


Over 2x faster than open source

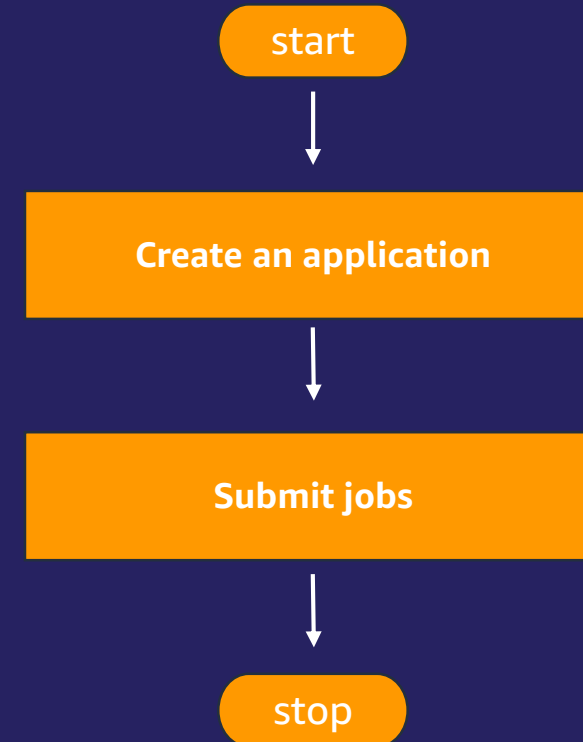


Amazon S3 data lake integration

Working with Amazon EMR made simple with Serverless



Data pipelines on Amazon EC2



Data pipelines on Amazon EMR Serverless

Amazon EMR Serverless brings flexibility to compute environments



Application =
OSS framework & version

Use **different open-source frameworks** (for example, Hive or Spark)

Use **different versions of open-source frameworks** for different use cases

Run **A/B testing** when upgrading from one version to another

Maintain **separate logical environments** for test and production scenarios

Allows **independent cost controls and usage tracking**

Separate different line-of-business applications (for example, finance vs. marketing)

AVAILABLE IN PREVIEW!

NEW

Amazon **MSK Serverless**

A serverless option to run Amazon MSK
without managing cluster capacity



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

NEW
PREVIEW



Amazon MSK

Serverless

Kafka

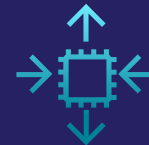
Simple to use

Easily run Apache Kafka clusters without rightsizing cluster capacity



Automated high availability

Provide built-in availability and fault tolerance by default



Flexible scaling

Instantly scale I/O without worrying about scaling capacity up and down or reassigning partitions



Pay for what you use

Pay for the data volume you stream and retain



Key features



No servers to manage



Fully compatible



On-demand streaming capacity



Same security as MSK



Pay-for-throughput pricing

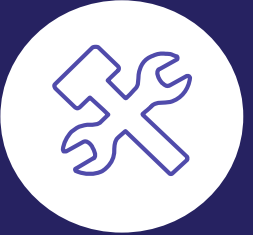


Same high availability

Benefits of Amazon MSK Serverless



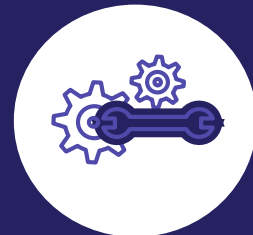
Makes Apache Kafka more secure, highly available, and accessible to your organization



Drives best practices through design, defaults, and automation



Allows developers to focus more on application development and less on infrastructure management



Eliminates rightsizing, scaling, and partition management with MSK Serverless

GENERALLY AVAILABLE TODAY!

NEW

Amazon Kinesis **Data** **Streams On-Demand**

Run Amazon Kinesis Data Streams without provisioning and managing capacity

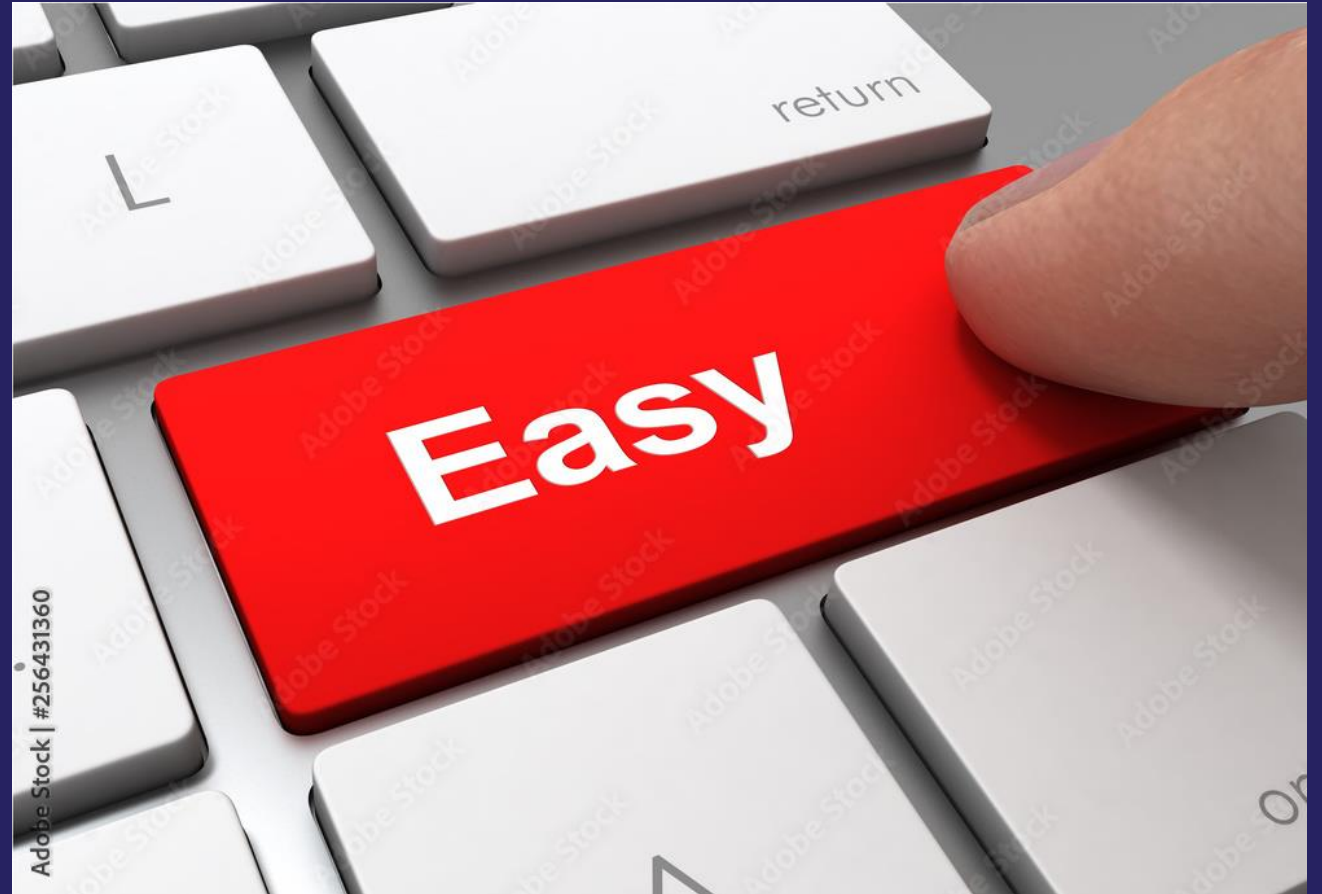


© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

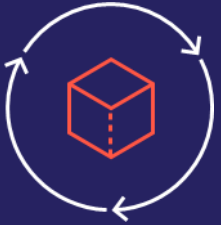
Customers want an “easy” button for their streaming data applications

Capacity planning can become challenging when you are operating dozens of workloads, each with different traffic patterns

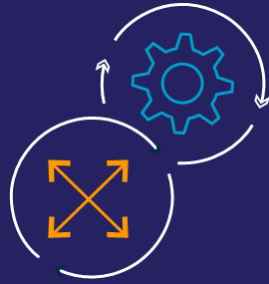
Customer have asked for a default experience that enables them to **get to market faster** and **reduce total cost of ownership**



Benefits Kinesis Data Streams On-Demand



Hands-free capacity
management



Seamless



Like always, same
high performance,
availability, and
durability



Pay for what
you use

NOW
AVAILABLE



Amazon Kinesis **Data Streams** **On-Demand**



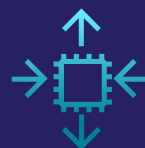
Simple to use

Simplify streaming data processing by eliminating capacity management



Automated high availability

Provide built-in availability and fault tolerance by default



Flexible scaling

Automatically scale capacity in response to changing data volumes



Pay for what you use

Pay per gigabyte of data written, read, and stored



Seamless switch for existing data streams

- You can also switch your existing data streams into on-demand mode
- There is no disruption to your write and read applications and you don't need to make any code changes
- You can switch between both on-demand and provisioned modes
- Shard count remains the same after the switch into either mode

Amazon Kinesis > Data streams > MyHighlyVariableStream_2 > Edit capacity mode

Edit capacity mode of MyHighlyVariableStream_2

Data stream capacity [Info](#)

Capacity mode

On-demand
Use this mode when your data stream's throughput requirements are unpredictable and variable. With on-demand mode, your data stream's capacity scales automatically.

Provisioned
Use provisioned mode when you can reliably estimate throughput requirements of your data stream. With provisioned mode, your data stream's capacity is fixed.

Total data stream capacity
By default, data streams with on-demand mode scale throughput automatically to accommodate traffic of up to 200 MiB per second and 200,000 records per second for the write capacity. If traffic exceeds capacity, your data stream will throttle.

Write capacity

Maximum
200 MiB/second and 200,000 records/second

Read capacity

Maximum (per consumer)
400 MiB/second

Up to 2 default consumers. Use Enhanced Fan-Out (EFO) for more consumers. EFO supports adding upto 20 consumers, each having a dedicated throughput.

[On-demand mode has a pay-per-throughput pricing model. See \[Kinesis pricing for on-demand mode\]\(#\)](#)

[Cancel](#) [Save changes](#)

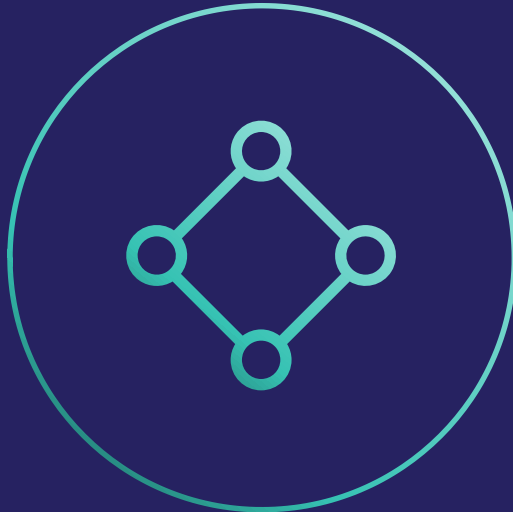
AWS has the most serverless options for data analytics in the cloud



Modern data strategy on AWS



MODERNIZE



UNIFY



INNOVATE



Thank you!