



NEW UPDATE FROM RE:INVENT

Data Analytics on AWS

Hung Gia

Head of Solutions Architect - Cambodia

Agenda

- **OVERVIEW OF AWS DATA ANALYTICS SERVICES**
- **NEW UPDATES FROM RE:INVENT**
- **GETTING STARTED**
- **QUIZ**

Overview of AWS Data Analytics services



*Every type
of app*



*Every type
of data*

Put data to work



*Make better
decisions*



*Respond
faster*

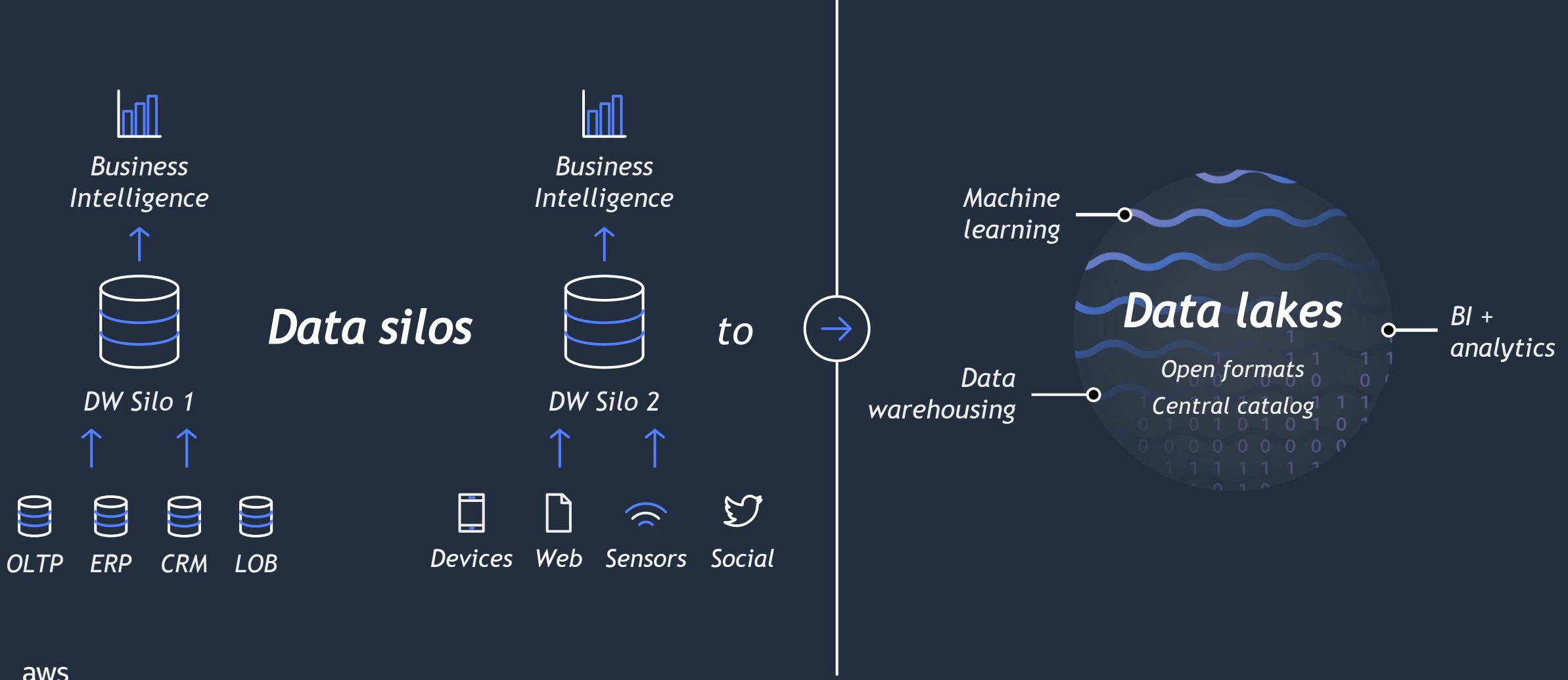


*Improve
efficiencies*

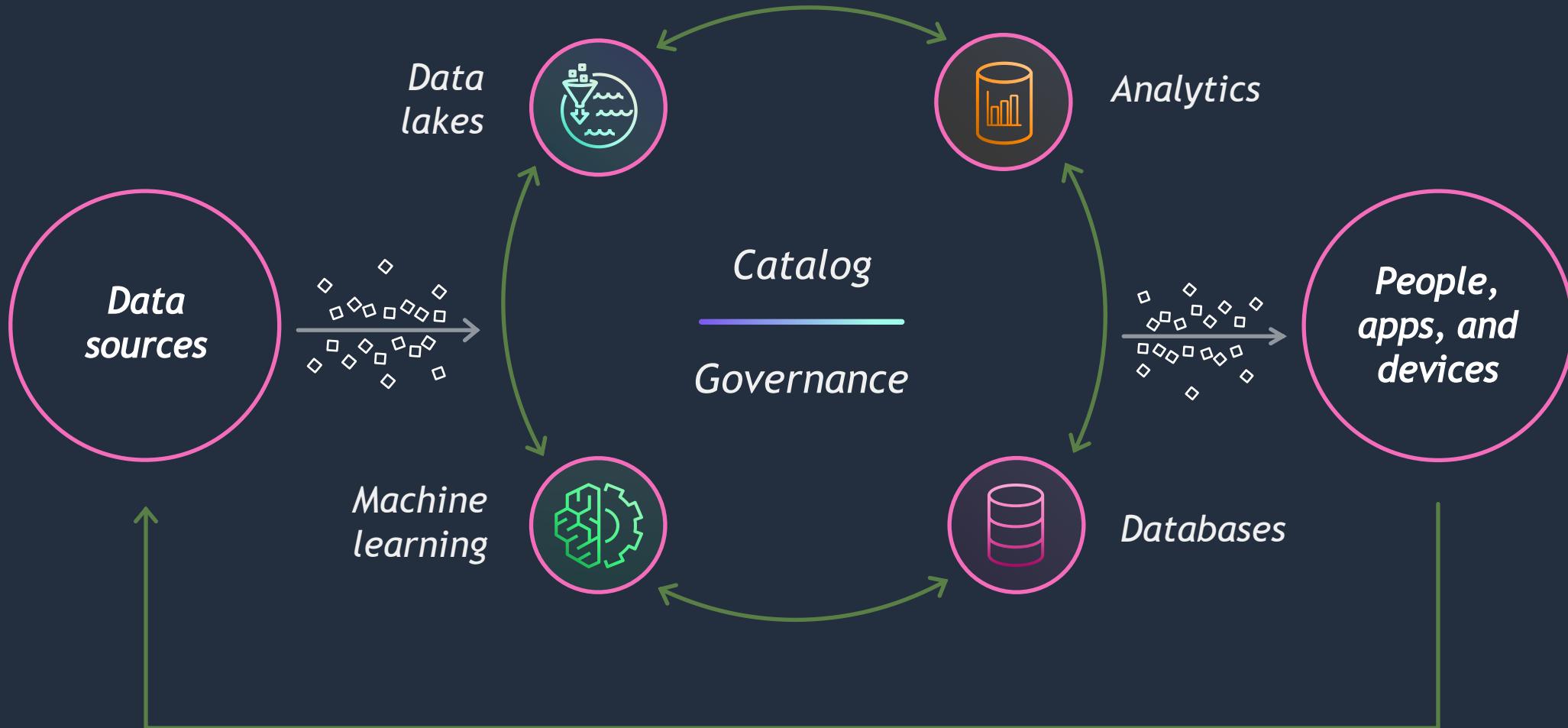


*Uncover
opportunities*

Traditional data warehousing approaches don't scale



Modern data strategy



AWS modern data strategy components



*Scalable
data lakes*



*Unified data
access, security,
and governance*



*Purpose-built
for performance
and cost*



*Serverless
and easy to
use*



*Built-in
machine
learning*



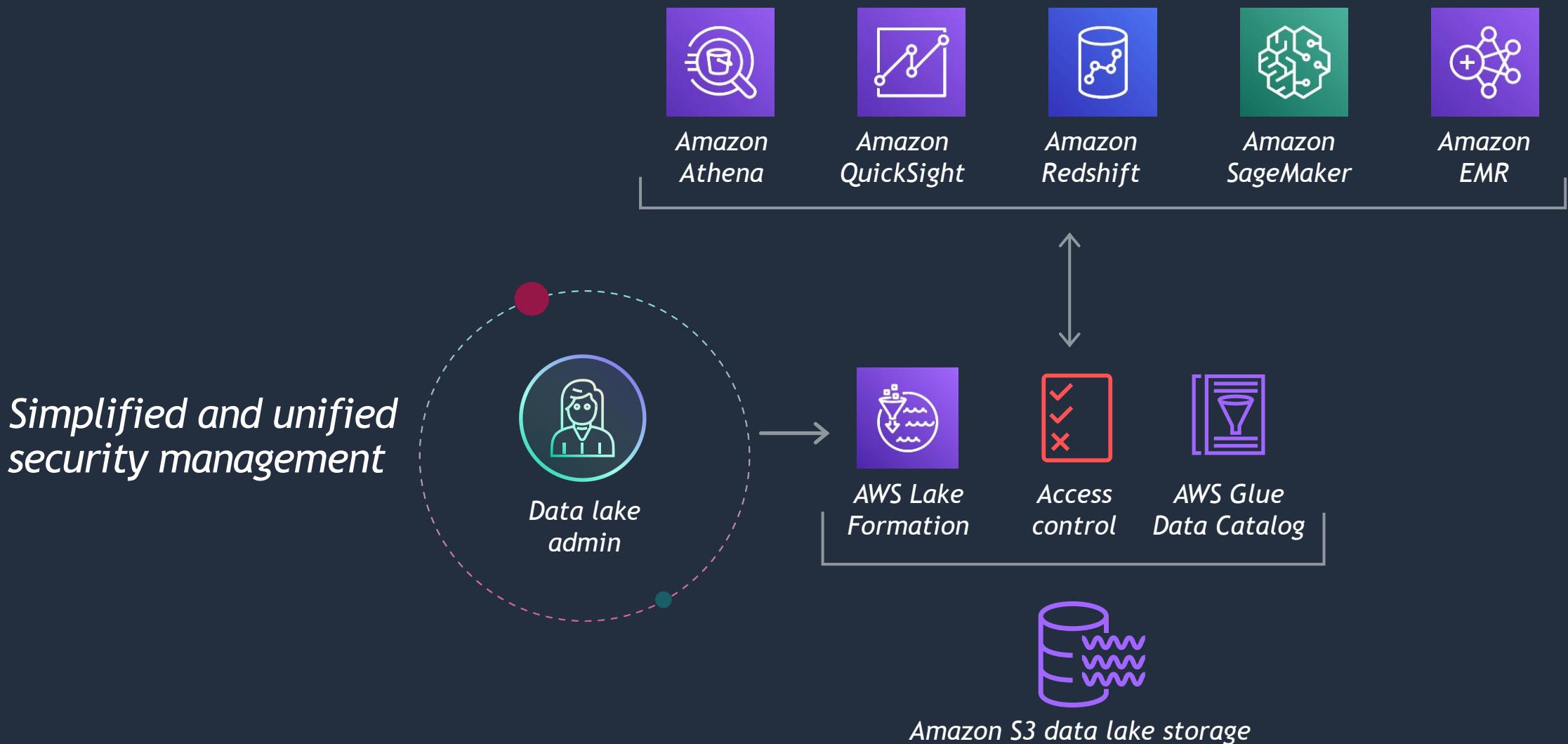
Scalable data lakes

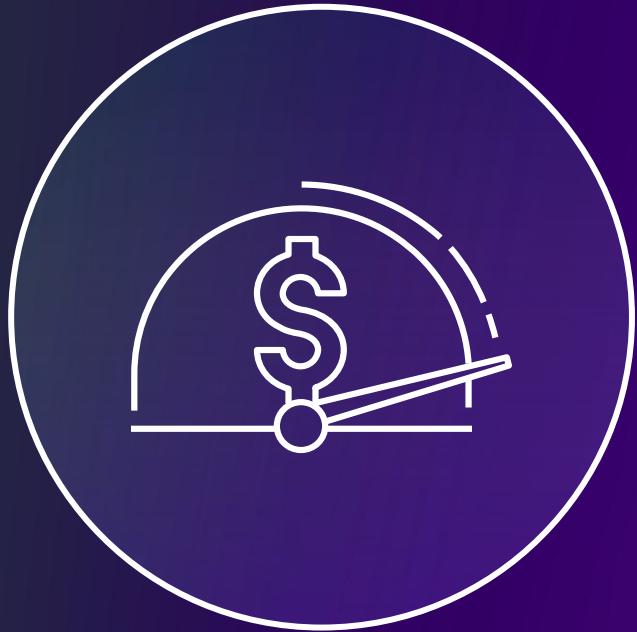
Amazon S3: Data lakes on AWS

- *Store unlimited data in open file formats*
- *Unmatched durability, availability, and scalability*
- *Decouple storage from compute*
- *Choice of analytical and ML engines*
- *Pay as you go*



AWS Lake Formation: Unified data governance





*Purpose-built for
performance and cost*

A family of purpose-built data services

Business intelligence and machine learning



Analytics



Databases



Blockchain



Data lake



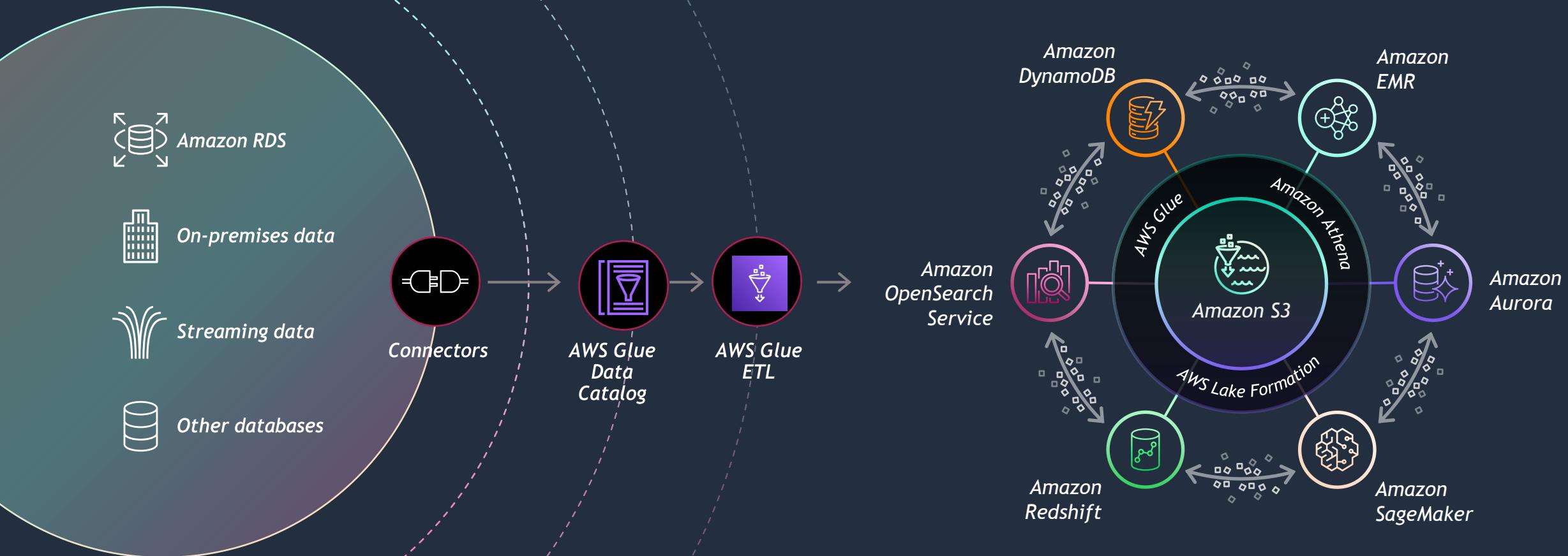
Data movement

[AWS DMS](#) | [AWS Snowball](#) | [AWS Snowmobile](#) | [Amazon Kinesis Data Firehose](#) | [Amazon Kinesis Data Streams](#) | [Amazon MSK](#)



© 2024, Amazon Web Services, Inc. or its affiliates.

Seamless data integration



Out-of-the-box benefits

PERFORMANCE, SCALE, AND COST



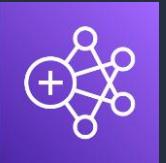
AWS Glue

AWS Glue -
<1 min.
startup time



Amazon S3

Amazon S3
Intelligent-Tiering
saves up to 70% on
storage costs



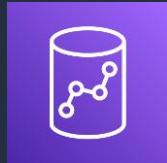
Amazon EMR runtimes

- *Queries run 3.1x faster compared to standard Apache Spark at 40% of the cost*
- *Queries run up to 2.6x faster than standard Presto*



Amazon
OpenSearch Service

- *Store 9x more log data at a lower cost*
- *UltraWarm storage tier reduces costs by up to 90%*



Amazon Redshift

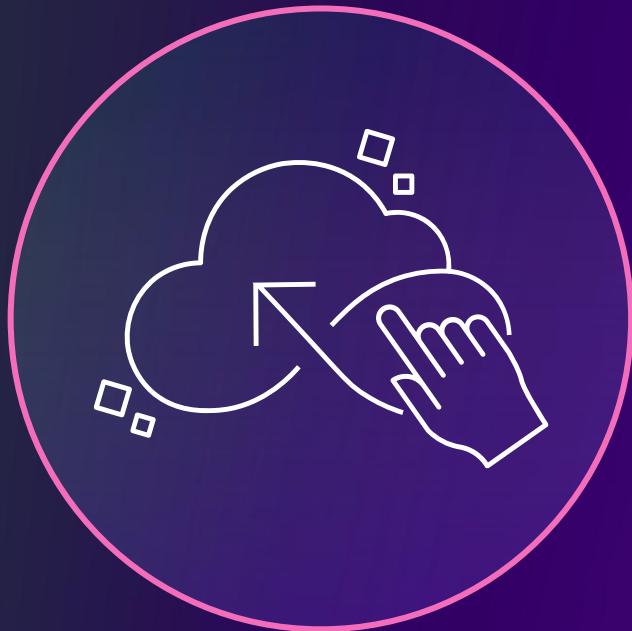
- *Amazon Redshift Serverless pricing reduced by 25% in GA*
- *Up to 5x better price performance compared to other cloud data warehouses*



With AWS Graviton2 instances, customers save 25.7% for typical workloads



© 2024, Amazon Web Services, Inc. or its affiliates.



*Serverless and
easy to use*

Benefits of serverless

1

Think big

2

Start small

3

Go fast

Faster time to market

Zero infrastructure management

Pay for what you use

Automatic scaling

Compute provisioning

Automated patching

Automatic failover

Advanced monitoring

Backup and recovery

Routine maintenance

Security and industry compliance

YOU
focus on insights

aws
*takes care of
the rest*

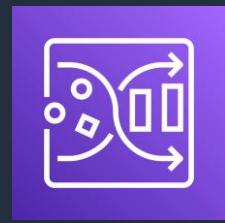
Serverless options for data analytics in the cloud



Low-code/no-code environments



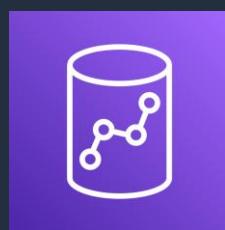
*Amazon EMR
Studio*



*AWS Glue
DataBrew*



*AWS Glue
Studio*



*Amazon Redshift
query editor v2*



*AWS Data
Exchange*



Built-in machine learning

AWS brings ML closer to data



Databases



*Data warehouses
and data lakes*



*Business
intelligence tools*

*Amazon
Aurora ML*



*Amazon
Neptune ML*



*Amazon
Redshift ML*



Amazon Athena ML



*Amazon
QuickSight ML*



Demo

Building a serverless modern data architecture

Demo: Use case

TRANSFORM, ANALYZE, AND VISUALIZE NEW YORK CITY TRANSIT DATA IN REAL TIME



Metropolitan Transportation Authority



Sign up for MTA API developer resources:
<https://api.mta.info/#/landing>

Train Routes	MTA API Endpoints
ACE	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-ace
BDFM	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-bdfm
G	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-g
JZ	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-jz
NQRW	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-nqrw
L	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-l
123456	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs
7	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-7
SIR	https://api-endpoint.mta.info/Dataservice/mtagtfsfeeds/nyct%2Fgtfs-si

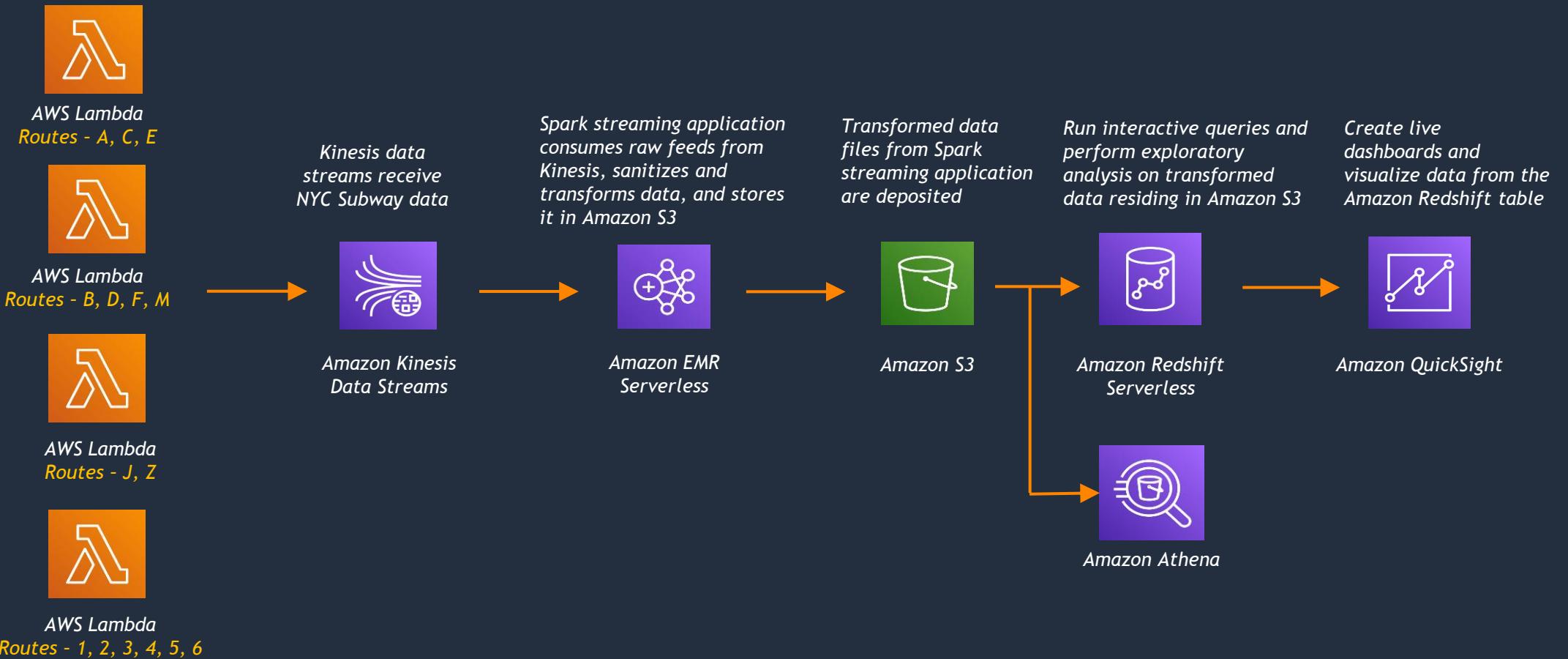


© 2024, Amazon Web Services, Inc. or its affiliates.

Demo: Architecture

TRANSFORM, ANALYZE, AND VISUALIZE NEW YORK CITY TRANSIT DATA USING FULLY SERVERLESS SERVICES

Kinesis producers running on AWS Lambda functions publish NYC Subway raw data feeds from MTA APIs to Kinesis data streams



*New updates from
Re:Invent*



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

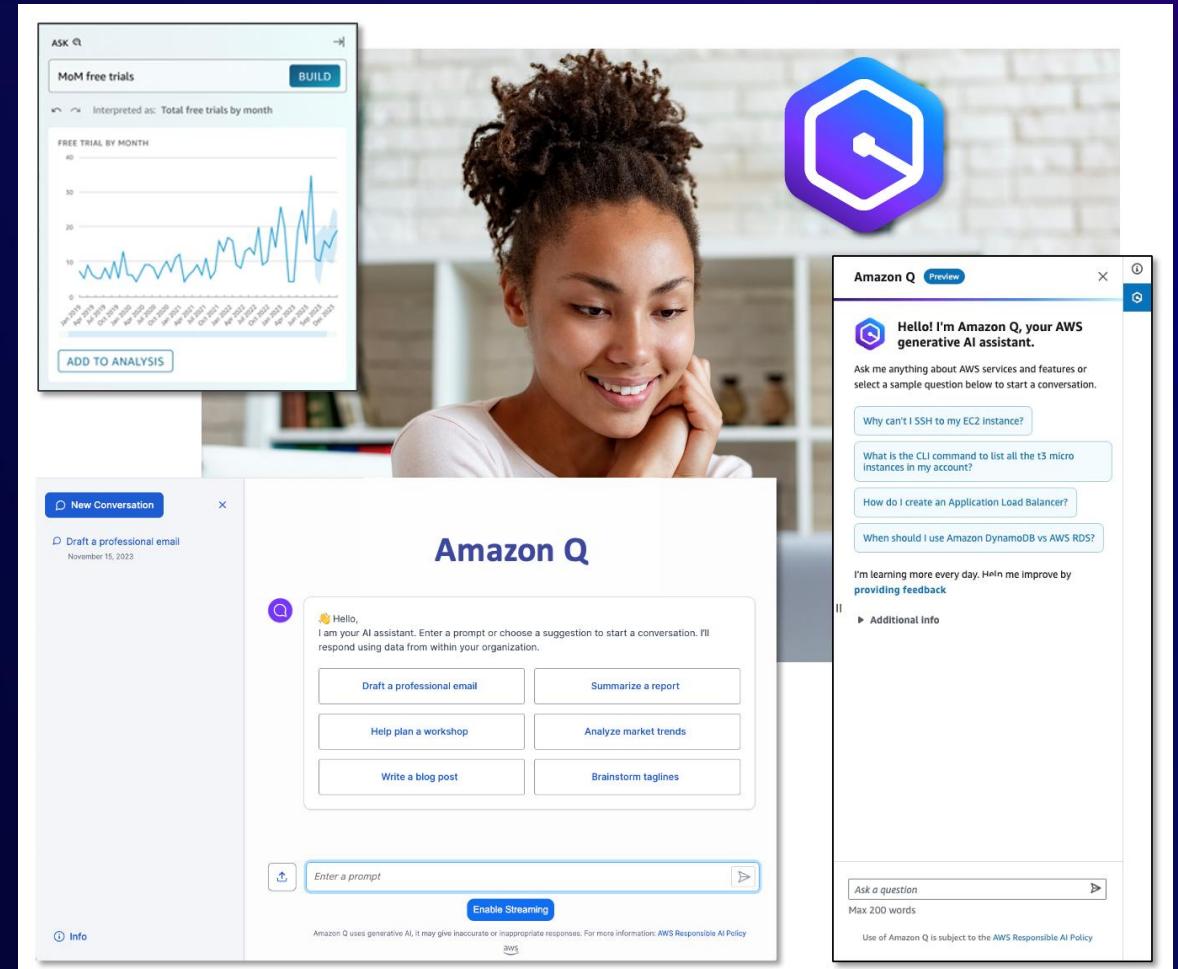
- *Amazon Q - Your generative AI assistant designed for work*
- *Zero ETL - Amazon Redshift and OpenSearch*
- *Data Lake - S3 Express One Zone*

* *More on Appendix*

Amazon Q

YOUR GENERATIVE AI ASSISTANT DESIGNED FOR WORK

- > Engages in conversations to solve problems, generate content, and take action
- > Understands your company information, code, and system
- > Personalizes interactions based on your role and permissions
- > Built to be secure and private



Amazon Q areas of expertise

*Your
business*



*Building
on AWS*



Amazon Quick Sight



*Amazon
Connect*



*AWS
Supply Chain*



Amazon Q with AWS Analytics

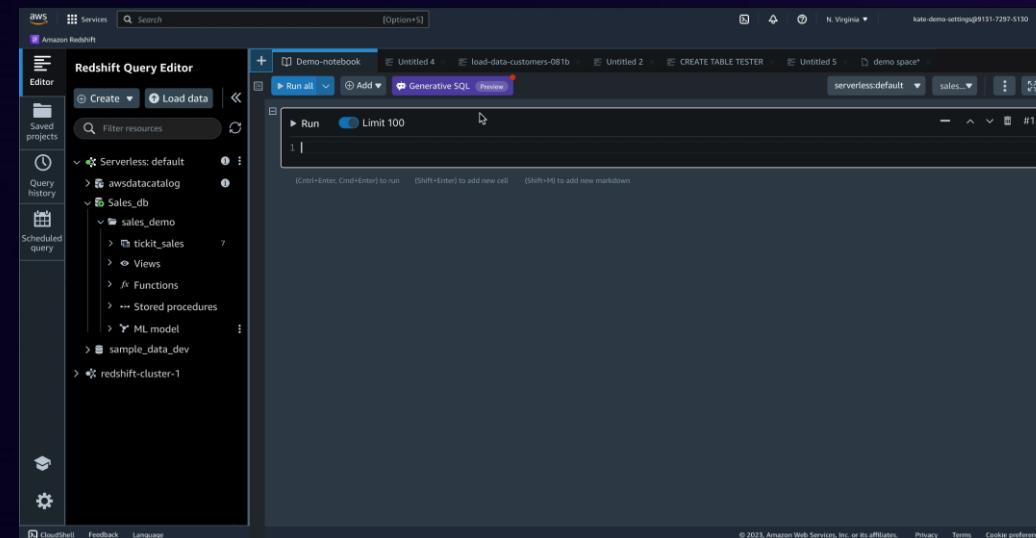


© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

PREVIEW

Amazon Redshift Q for SQL

GENERATE SQL CODE FROM NATURAL LANGUAGE



Increase user productivity

Author *SQL queries faster with a conversational interface by submitting query requests in plain English by using Generative SQL*

Get started fast

Generate custom *SQL code without needing to understand your organization's complex database*

Securely access your data

Experience *consistent data governance compliance and user permissions*.

Continually improve SQL accuracy

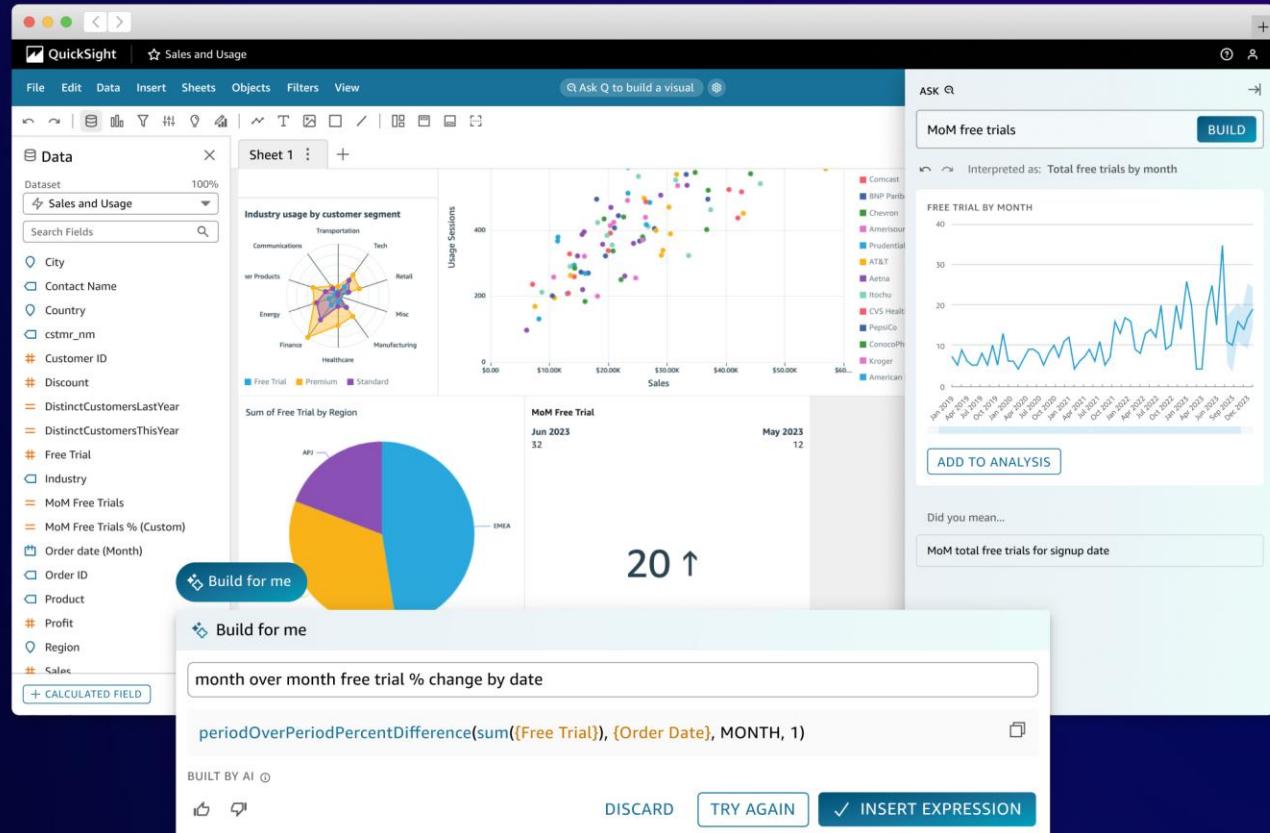
Enhance the *accuracy of account specific SQL code recommendations with admin-controlled query history access, without compromising data privacy*



Amazon Q in QuickSight (Preview)

YOUR GENERATIVE BI ASSISTANT THAT MAKES IT EASY TO BUILD AND CONSUME INSIGHTS

- > Enables business analysts/users' to quickly build and refine compelling visuals, summarize insights, and answer data questions.
- > Accelerate analysis by easily creating calculations without looking up or learning specific syntax
- > Interpret and share insights with visually compelling narratives with simple UI-based workflows

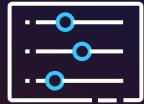


Zero ETL



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Real-time insights on operational data are critical for modern enterprises



Personalization



Fraud detection



Churn prevention



Gaming leader boards



Location optimization



Inventory optimization



Customer relationship management



Scoring



Internet of Things (IoT)



Anomaly monitoring



Sales operations



Marketing optimization



Product insights



More...

Existing operational analytics solutions are hard

SEPARATE SYSTEMS REQUIRE COMPLEX DATA MANAGEMENT PIPELINES

*Operational
databases*



Manual data pipelines



*Expensive and cumbersome
to build and maintain ETL jobs*

*Complex reconstruction of the data,
especially with schema changes*

*Incomplete, inconsistent, and stale
views of data, limiting insights*

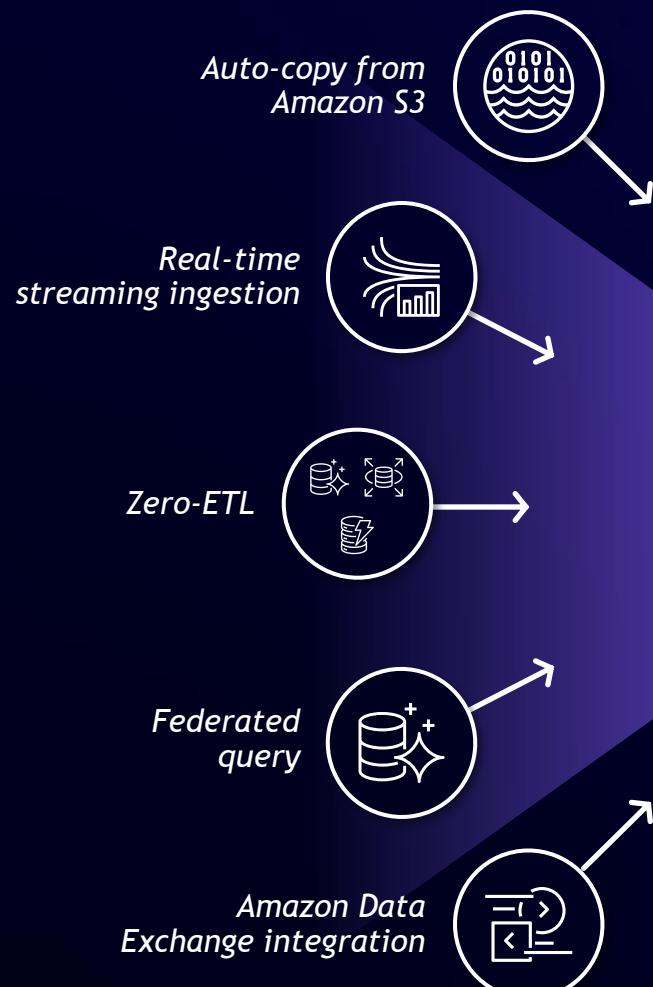
Analytics



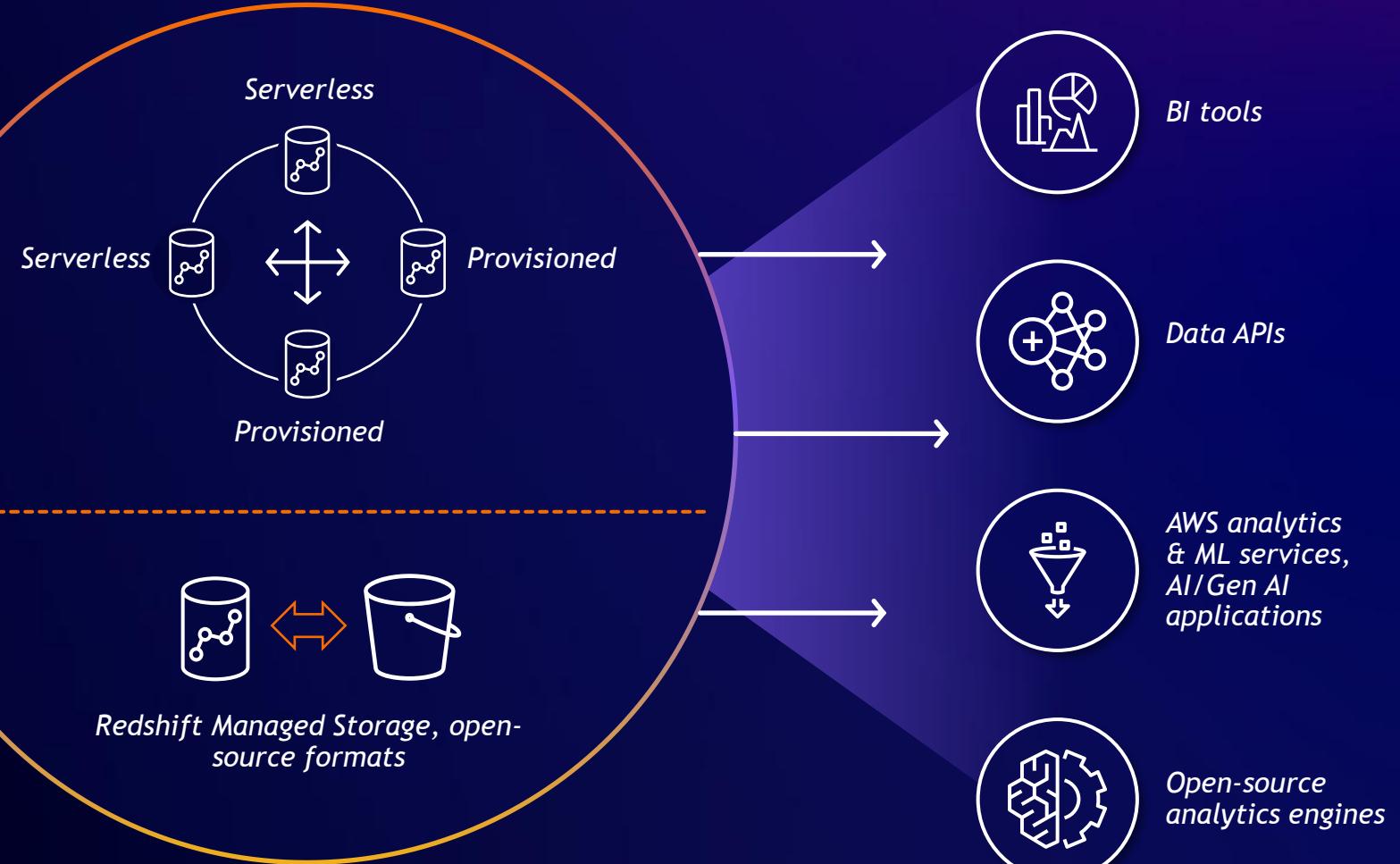
Welcome to a zero-ETL world!

Easy ingestion into the warehouse | Access data in place | No manual pipelines

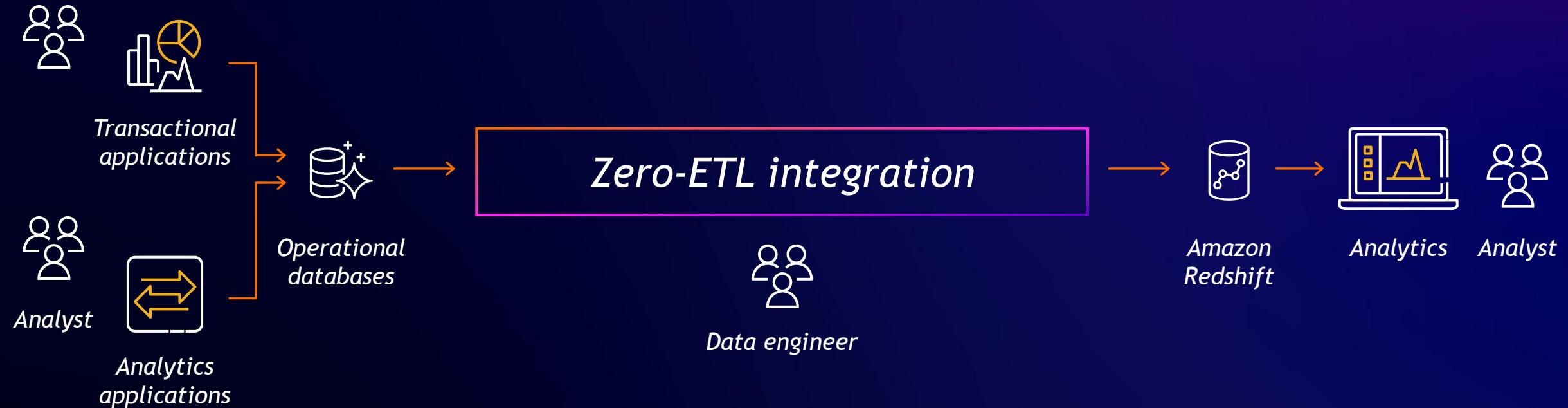
Analytics on all data



Analytics for all users



Amazon Redshift Zero-ETL vision



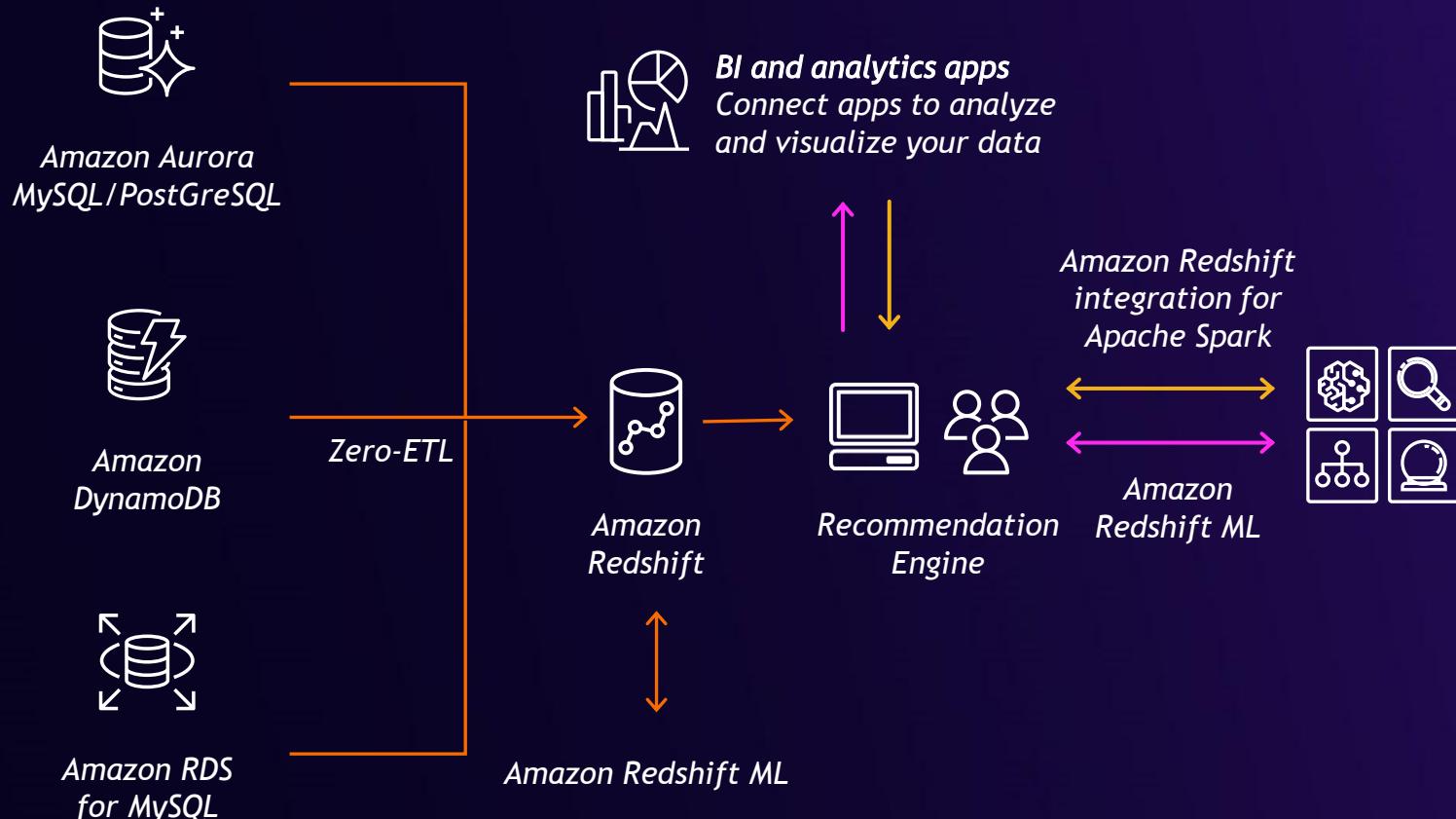
An **easy and secure** way to enable **near real-time analytics** on **petabytes of transactional data with no pipeline management**

NEW!

PREVIEW

Amazon Redshift zero-ETL integrations

THREE MORE AWS DATABASES NOW SUPPORT ZERO-ETL INTEGRATION WITH AMAZON REDSHIFT



- Combine data from many operational databases into a single data warehouse
- Amazon Aurora MySQL (GA), Amazon Aurora PostgreSQL (preview)
- Amazon DynamoDB (preview)
- Amazon RDS MySQL (preview)



Amazon OpenSearch ZERO ETL capabilities

- *Amazon OpenSearch Service zero-ETL integration with Amazon S3 (preview)*
- *AWS announces Amazon DynamoDB zero-ETL integration with Amazon OpenSearch Service*



Amazon Athena ZERO ETL Capabilities

- *AWS CloudTrail Lake data now available for zero-ETL analysis in Amazon Athena*

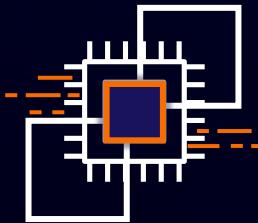


Amazon S3 Express One Zone storage class



Available now

Amazon S3 Express One Zone storage class



For compute-intensive workloads



Single-digit millisecond latency

One zone S3 storage class that delivers the fastest data access speed and highest performance of any cloud object storage for customers' most latency-sensitive applications

Most frequently accessed data

Designed for request intensive applications - ML training and inference, interactive analytics, media content creation

10x faster + 50% lower request costs

Data access speeds up to 10x faster, and request costs up to 50% lower than S3 Standard. Fully elastic with no storage provisioning

The newest Amazon S3 storage class



S3 Intelligent-Tiering



S3 Standard



S3 Standard-IA



S3 Glacier Instant Retrieval



S3 Glacier Flexible Retrieval
(formerly S3 Glacier)



S3 Glacier Deep Archive

AWS Region ≥ 3 Availability Zones

Changing access patterns

- Milliseconds access
- No retrieval charge
- Archive Instant Access tier

Frequently accessed data

- Milliseconds access
- No retrieval charge

Infrequently accessed data

- Milliseconds access
- Per-GB retrieval charge

Rarely accessed data

- Milliseconds access
- Per-GB retrieval charge

Archive data

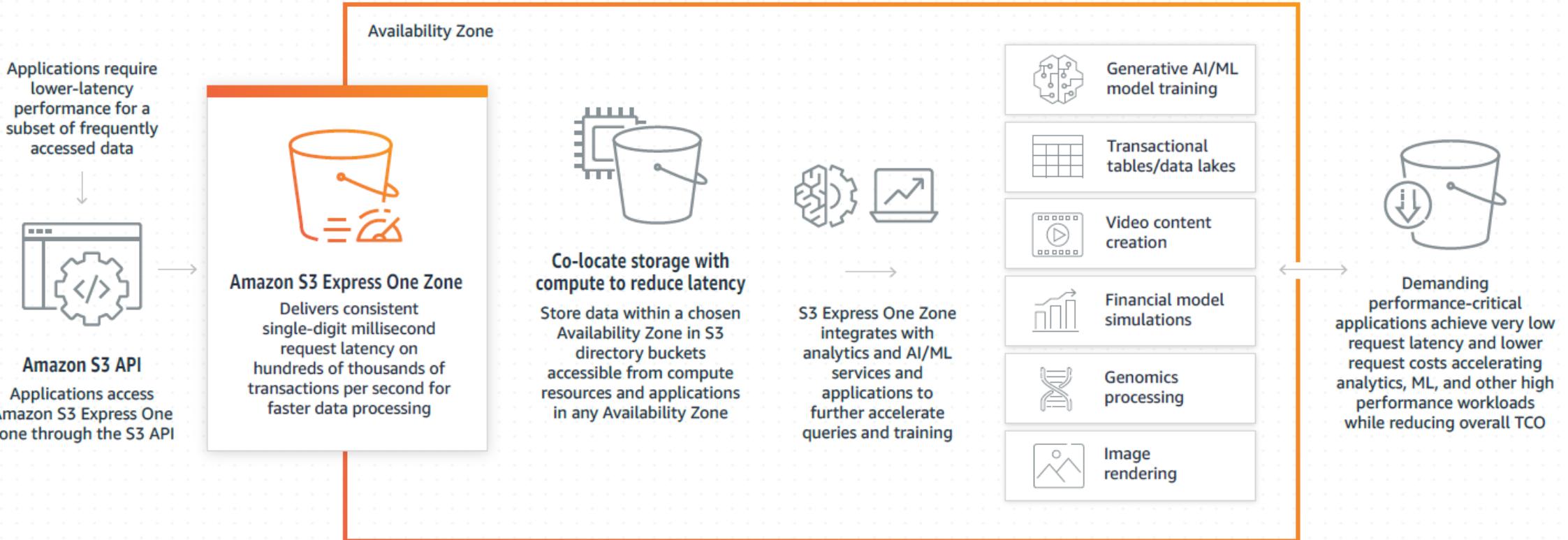
- Retrieval options from minutes to hours
- Free bulk retrievals

Long term archive data

- Retrieval in hours

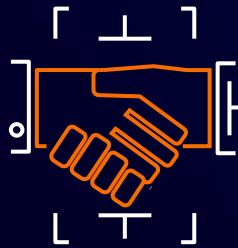
Purpose built for fastest data processing for performance critical applications

How S3 Express One Zone works



Getting Started

Accelerate your modern data strategy on AWS



MODERNIZE

DB Freedom Program

*Migration Assistance
Program*

*ProServe/Partner
Accelerators*

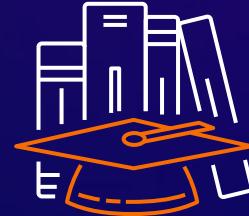


UNIFY

MAP for Analytics

*Data-Driven Everything
(D2E)*

Immersion Days



INNOVATE

*AWS Machine Learning
Embark Program*

ML Solutions Lab

*ProServe/Partner
Accelerators*

Get your Hands-on!

Access **130+ free** digital hands-on labs.

Achieve **3.000.000 views**

Explore resources with a variety of skill levels

Build in **English**

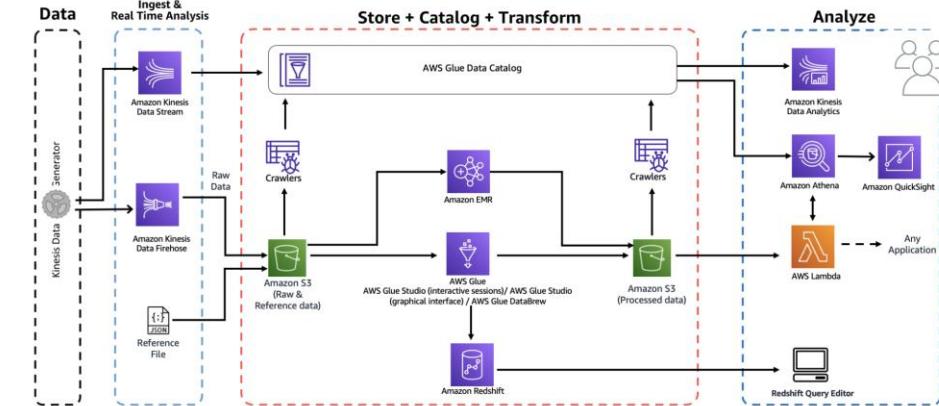
Build as **road map** and update continuously



ANALYTICS ON AWS WORKSHOP

Welcome to the Analytics workshop on AWS. The purpose of this workshop is to acquaint you with the different analytics services offered in the AWS Analytics portfolio.

The workshop will cover various modules that discuss different aspects of constructing an analytics platform on AWS. Participants will acquire knowledge on data ingestion, storage, transformation, and utilization using various analytics services including AWS Glue, Amazon Athena, Amazon Kinesis, Amazon EMR, Amazon QuickSight, as well as AWS Lambda and Amazon Redshift. The architecture diagram below provides a more detailed representation of the design:



Learning outcomes from this workshop:

1. Designing a serverless architecture for data lakes (serverless data lake).



<https://cloudjourney.awsstudygroup.com/6-dataandanalytic/>



Learn in-demand AWS Cloud skills



AWS Skill Builder

Access **500+ free** digital courses and Learning Plans

Explore resources with a variety of skill levels

Deepen your skills with digital learning on demand



Train now

AWS Classroom Training

Live classes led by expert AWS instructors

Learn **best practices** based on real-world problems and solutions from building in the AWS Cloud



AWS Certifications

Earn an industry-recognized credential

Receive Foundational, Associate, Professional and Specialty certifications



Access exam guides



Thank you

HUNG GIA

Head of Solutions Architect - Cambodia



Service Specific Announcement



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift

FULLY MANAGED, AI-POWERED CLOUD DATA WAREHOUSING



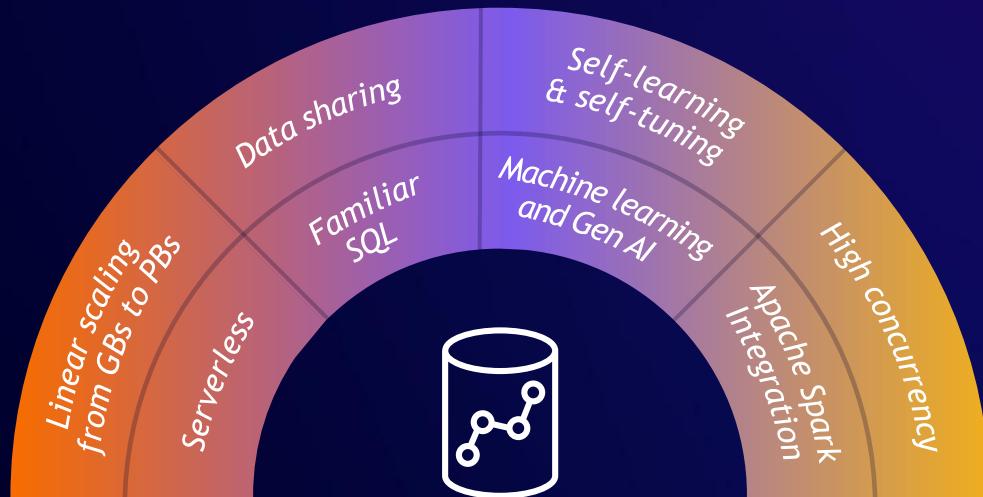
Data

Transactional data

Clickstream

IoT telemetry

Application logs



Amazon Redshift

Unify data across databases, data lakes and data warehouses with a zero-ETL approach



Best-in-class security, governance, and compliance



Insights

Analyze and visualize data

Deliver real-time & predictive analytics

Build data-driven applications

Amazon Redshift evolution

Best price
performance
cloud DW

Serverless

Automatic scaling
for consistent
performance

Real-time
data ingestion
and zero-ETL

Streaming
Operational
databases

BI tools

Data API

Query Editor

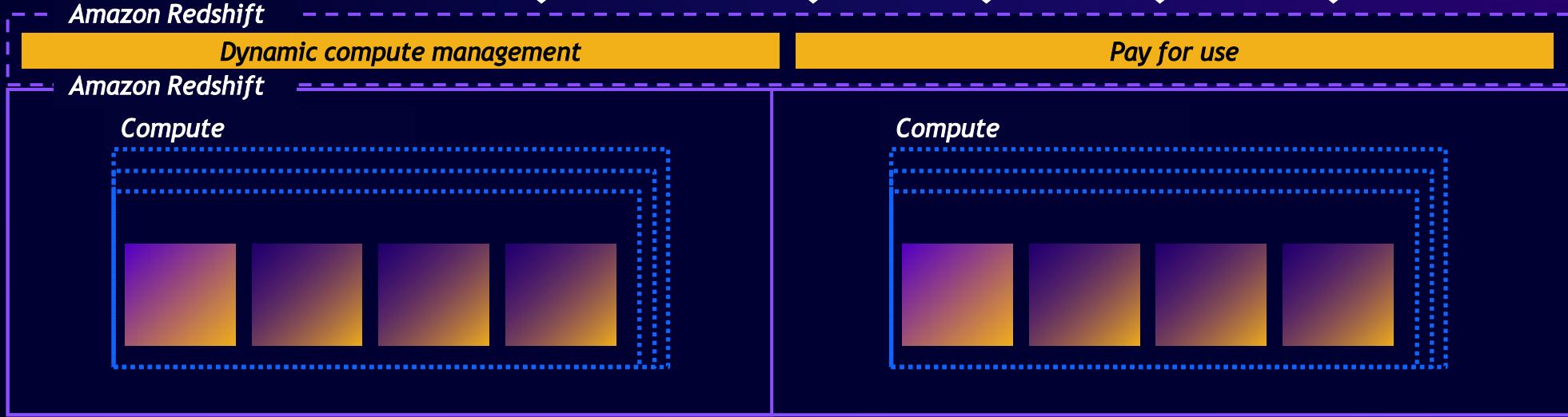
Amazon
EMR

Amazon
Glue

Amazon
Sage Maker

Amazon
Data Exchange

Third party
data exchanges



Scale and pay for
compute and storage
independently

Workload isolation and charge-
ability with data sharing
Native data lake querying

Managed Storage

Amazon S3

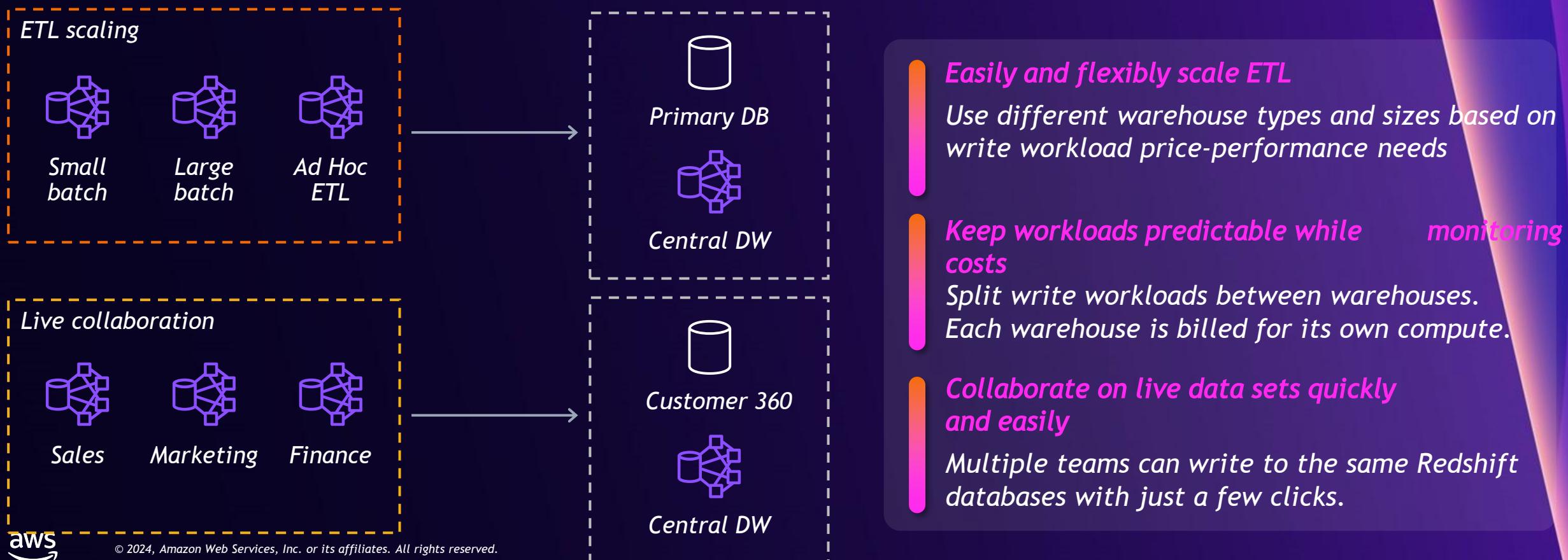
Streaming
Operational
databases

Parquet ORC JSON



Multi-data warehouse writes through data sharing

WRITE WORKLOAD ISOLATION AND LIVE COLLABORATION



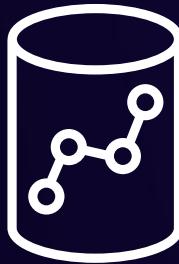
With new AI-driven scaling and optimizations, Amazon Redshift Serverless is smarter than ever

AUTO-SCALE ALONG MORE DIMENSIONS | MEET YOUR PRICE-PERFORMANCE TARGETS

NEW!

PREVIEW

Up to 10x
better price
performance



OPTIMIZED
FOR COST



BALANCE

OPTIMIZED FOR
PERFORMANCE

Intelligent auto-scaling for dynamic workloads along changing workloads dimensions such as amount of data and query complexity

Experience better price-performance with holistic ML driven optimizations that maintain consistent performance while optimizing costs

Adjust your desired price-performance targets for your workload with flexible options ranging from optimized for cost, balanced, to optimized for performance.

General availability for Apache Iceberg support

QUERYING THE APACHE ICEBERG OPEN TABLE FORMAT FROM REDSHIFT



- Access to Apache Iceberg tables in AWS Glue data catalog*
- Query data lake with transactional consistency*
- Join data warehouse data and data lake data easily*
- Improved performance*

PREVIEW

Redshift ML support for bring your own LLMs

BRING THE POWER OF LARGE LANGUAGE MODELS TO YOUR DATA ON REDSHIFT

```
CREATE MODEL my_llm_model  
FUNCTION my_llm_model(supер)  
RETURNS super  
SAGEMAKER '<endpoint\_name>'  
IAM_ROLE default/'<iam\_role>';
```

SQL powered LLM inferences

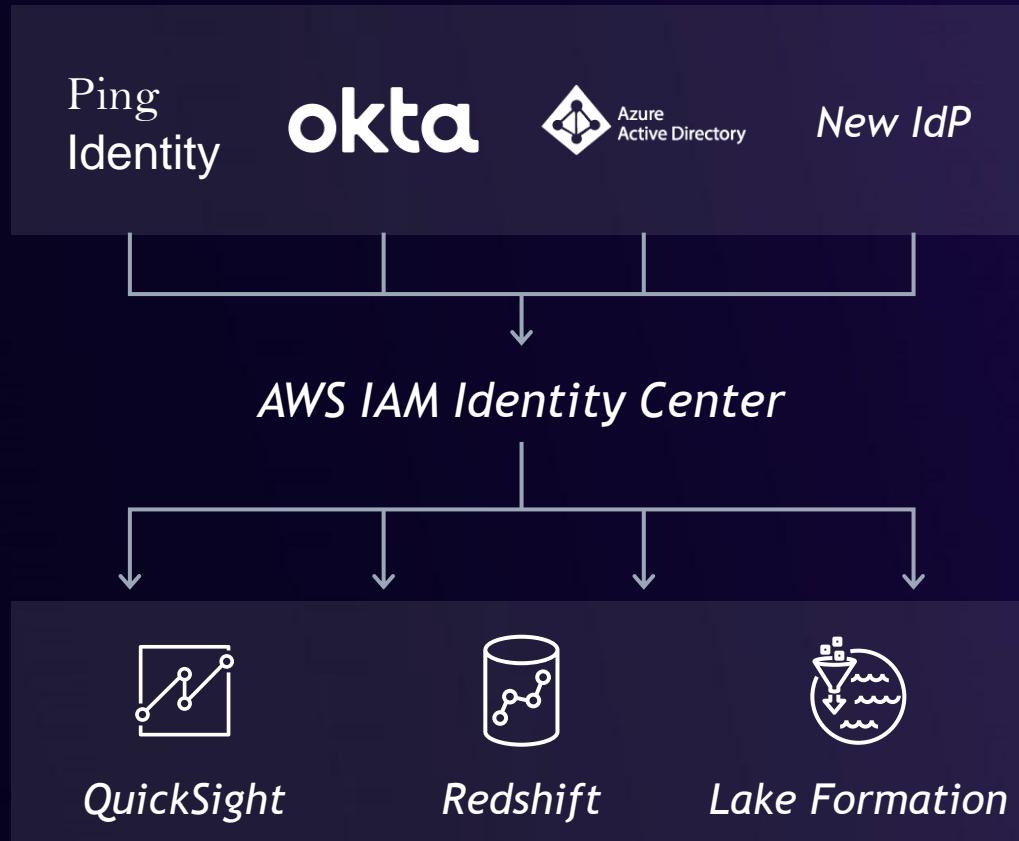
Create and make inferences on LLM models using SQL

Utilize SageMaker Jumpstart large language models for remote inference



AWS IAM Identity Center Integration

CENTRALIZE IDP FOR AUTHENTICATION AND AUTHORIZATION

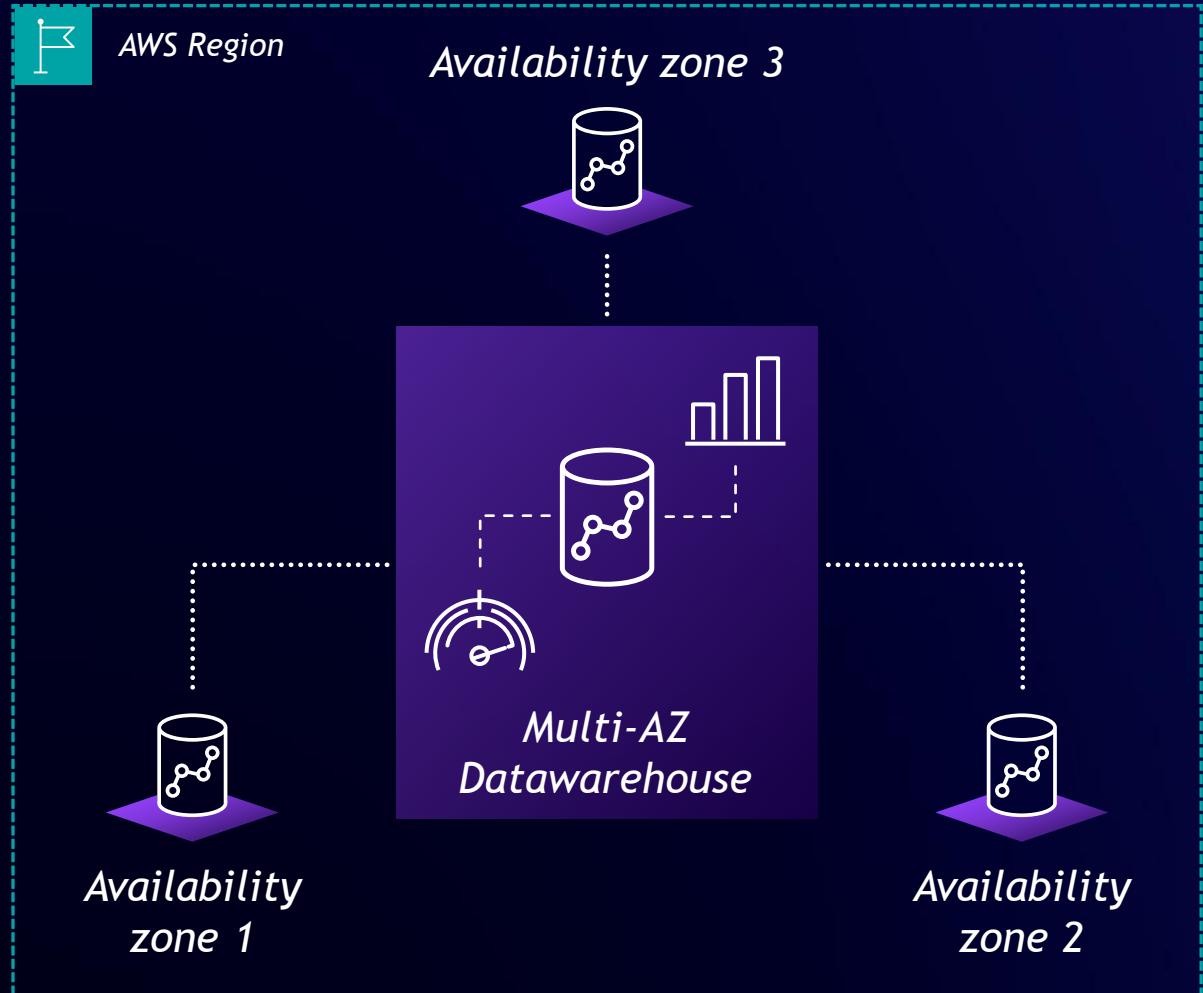


- Use organization's end-user identity mechanisms (e.g., Azure AD, Ping Federate) in AWS Analytics & ML services
- Consistent, familiar auth experience for end-users
- Seamless end-to-end access across AWS services with no re-authentication needed (e.g., QuickSight to Redshift)
- Authorization using group membership by Redshift and Lake Formation
- Easy, consistent auditability across services

Amazon Redshift Multi-AZ

GA

HIGHLY RESILIENT DATA WAREHOUSE FOR MISSION-CRITICAL WORKLOADS



99.99%
for mission-critical workloads

Amazon Redshift cont...

- Amazon Redshift now supports metadata security to simplify multi-tenant applications
- Amazon Redshift announces general availability of row-level security enhancements
- Amazon Redshift adds support for incremental refresh for materialized views on data lake tables (preview)
- Amazon Redshift announces new fine-grained access control capabilities to nested objects (preview)
- Announcing Amazon Redshift Serverless with AI-driven scaling and optimizations (Preview)
- Amazon Redshift announces Multidimensional Data Layouts to optimize your query performance (preview)



Amazon Redshift cont...

- *Redshift provisioned concurrency scaling and serverless autoscaling now supports Create Table As Select (CTAS)*
- *Amazon Redshift announces integration with AWS Glue column-level statistics*
- *Amazon Redshift announces support for CNAME*
- *Amazon Redshift launches MaxRPUs, a new compute cost control setting for Redshift Serverless*

Amazon QuickSight



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon QuickSight

UNIFIED BI SERVICE AT HYPERSCALE



Unified BI for all your analytics needs

*Lower costs with
consumption-based pricing*

*Consistent high performance with
auto scaling*

*Augmented analytics with ML-insights
and NLQ*

Generative BI capabilities

AI-powered dashboard authoring experience

enables business analysts to build faster

AI-powered insights on demand

simplifies extracting insights for business users

AI-assisted data storytelling

enables business users to discover and share findings

**Powered by
Amazon Bedrock**

AI-powered dashboard authoring experience

Build visuals

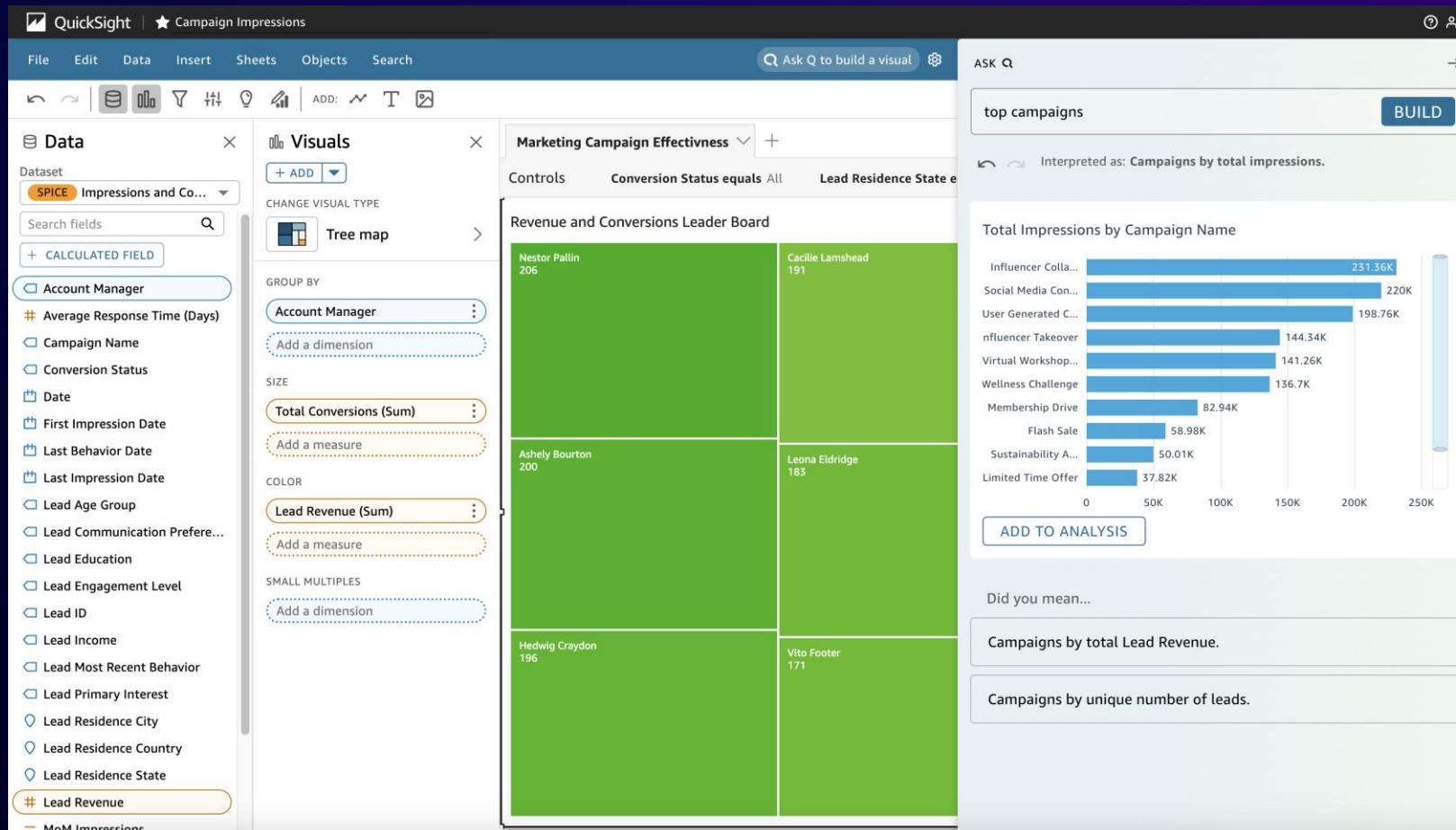
Use natural language to quickly build visuals for dashboards and reports

Build calculations

Easily create calculations using natural language without looking up or learning specific syntax

Refine visuals

Quickly update visuals by describing desired formats using natural language



AI-answers to questions of data on-demand

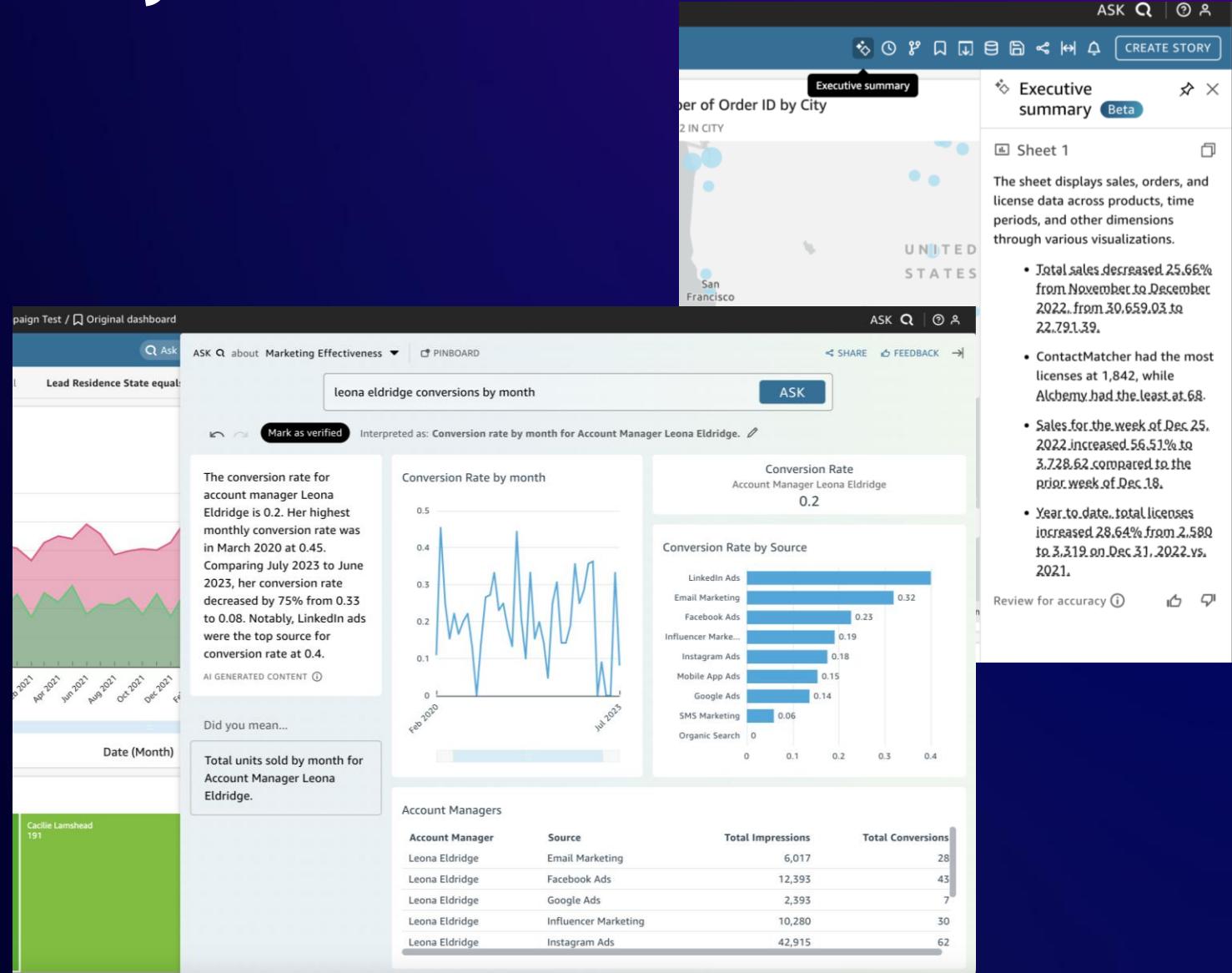
Executive summaries of dashboards

Instant summaries of key dashboard insights in natural language explaining top movers, outliers, and more

Easy powerful Q&A for non-experts
Suggested questions and what's in my data show what can be asked

Multi-visual answers with narrative insight summaries explain answer context

Vague question support and did-you-mean alternatives enable iterative fact finding



AI-assisted storytelling

Interpret data for others

Help others derive meaning from data and reach conclusions to drive decisions

Generate stories using AI

Generate cohesive, powerful, and insightful narratives analyzing data with only a few words

Create refined content

Control AI verbosity, customize narrative text, and apply stunning visual themes to bring content to life

Governed and always up to date

Quickly and easily share, update data any time

A Data-Driven Journey: Optimizing Marketing Performance Through Insights

Prepared by [Name]

Introduction

As we examine how to maximize success across our marketing efforts, we will analyze campaign results over time to understand what strategies have been most effective. By identifying the goal is to optimize processes and targeting approaches to continue impro

Revenue and Conversions Leader Board

Revenue and Conversions Leader Board				
Nestor Pallin 206	Leona Eldridge 183	Tasha Doug... 156	Karol Brown... 152	Celka Dui... 139

Build a story about marketing campaign performance over time. Describe top campaigns and account managers. How can we improve overall campaign success?

Revenue and Co... × Revenue by State ×
Impressions and... × Conversions per ... ×

+ ADD VISUALS ⓘ

BUILD

Amazon QuickSight

BI

- *Announcing the Cost and Usage Dashboard powered by Amazon QuickSight*
- *Amazon QuickSight launches a new redesigned analysis experience*
- *Amazon QuickSight supports SPICE capacity auto-purchase*
- *Amazon QuickSight now supports export and import of asset permissions and tags*
- *Amazon QuickSight now supports runtime theming for embedded dashboards and visuals*
- *Amazon QuickSight now supports asset events using Amazon EventBridge*
- *Amazon QuickSight now supports runtime filtering for embedded dashboards and visuals*
- *Amazon QuickSight supports SPICE capacity auto-purchase*



Amazon QuickSight Cont...

- *Amazon QuickSight now Supports Connectivity to Google BigQuery*
- *QuickSight launches Custom Week Start*
- *Amazon QuickSight launches custom access to capabilities based on roles*
- *Amazon QuickSight now supports programmatic user access management by assigning groups to roles*
- *Amazon QuickSight launches shared restricted folders and a folder Contributor role to govern asset sharing*
- *QuickSight launches Custom Time Zone*
- *QuickSight launches FLOAT data type support for SPICE datasets*

AWS Glue



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Glue

DISCOVER, PREPARE, AND INTEGRATE ALL YOUR DATA AT ANY SCALE



All-in-one data integration service



Cost effective, serverless, and scalable



Tailored tools to support all data users



Support all workloads in one place

AWS Glue

- *AWS Glue Data Quality announces anomaly detection and dynamic rules*
- *AWS Glue Data Catalog supports multi engine views with AWS Analytics Engines*
- *AWS Glue announces entity-level actions to manage sensitive data*
- *AWS Glue Data Catalog now supports generating column-level statistics*
- *AWS Glue Data Catalog supports automatic compaction for Apache Iceberg tables*
- *AWS Glue launches native connectivity to 6 databases*
- *AWS Glue for Apache Spark announces native connectivity for Amazon OpenSearch Service*
- *Glue Studio Visual now supports interactive data previews*

Amazon EMR



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

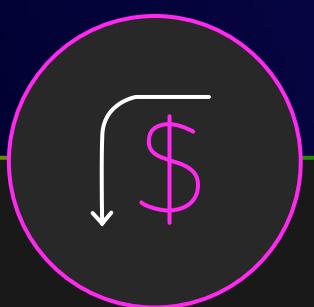
Amazon EMR

EASILY RUN SPARK, HIVE, PRESTO, HBASE, FLINK, AND MORE BIG DATA APPS ON AWS

Latest versions



Great performance at lower cost



Use S3 storage



Multiple deployment options



Updated with latest open source frameworks within 60 days

Support for popular OSS like Spark, Flink, Hudi, Iceberg etc.

Spark workloads run up to 3x faster compared to open source

*50-80% reduction in costs with EC2 Spot, EC2 Savings Plan, Reserved Instances
Per-second billing for flexibility*

Process data in S3 securely with high performance using the EMRFS connector

Scale Compute and Storage independent of each other

Fully managed, choose multiple deployment options based on your need

Amazon EMR

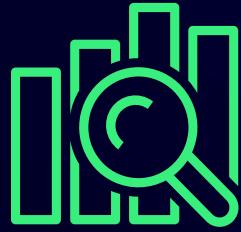
- *Amazon EMR Studio now supports Amazon CodeWhisperer*
- *Customers can now launch their Amazon EMR on EC2 clusters in 5 minutes or less*
- *Accelerate data processing and analysis with Amazon EMR and Amazon S3 Express One Zone*
- *Amazon EMR Studio now supports Amazon CodeWhisperer*
- *Amazon EMR Studio is now available in 4 new AWS Regions*
- *Amazon EMR Studio adds interactive query editor powered by Amazon Athena*
- *Amazon EMR on Amazon EKS is now available in 3 additional regions*
- *Launch low code data preparation for machine learning with Amazon SageMaker Data Wrangler from Amazon EMR Studio*



Amazon OpenSearch



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.



Amazon OpenSearch Service

Amazon OpenSearch Service securely unlocks real-time search, monitoring, and analysis of operational data



Managed: Increase operational excellence and use a popular open source solution



Secure: Audit and secure your data with a data center and network architecture and built-in certifications



Observability: Systematically detect potential threats and react to a system's state through an open source solution for machine learning, alerting, and visualization



Cost conscious: Optimize time and resources for strategic work

Amazon OpenSearch

- *Amazon OpenSearch Service now supports Neural Sparse Retrieval*
- *AWS announces OR1 for Amazon OpenSearch Service*
- *Amazon OpenSearch Service zero-ETL integration with Amazon S3 preview now available*
- *Announcing enhanced manageability and usability features for Amazon Redshift Serverless*
- *AWS announces Amazon DynamoDB zero-ETL integration with Amazon OpenSearch Service*
- *Amazon OpenSearch Service now supports Neural Sparse Retrieval*
- *Amazon QuickSight launches a new redesigned analysis experience*

Amazon OpenSearch

- *Amazon OpenSearch Ingestion now supports migrating your ElasticSearch indexes to Amazon OpenSearch Service*
- *Amazon OpenSearch Ingestion announces support for persistent buffering*

Amazon Athena



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Athena

Query all your data using SQL or Python

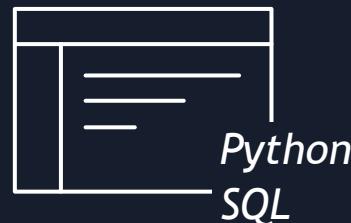


Simple, instant start

Serverless, no setup

*Instant start, optimized
runtimes for fast results*

*Point to Amazon S3 and
start querying*

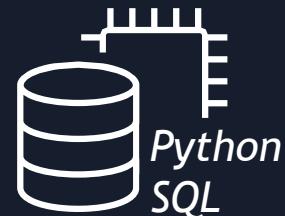


Interactive, advanced analytics

*Federated queries across
35+ data stores*

Use PySpark interface

*Simplified notebooks on
console for PySpark*



Open and flexible

ANSI SQL, Apache Spark

*Multiple formats,
compression types, and
complex joins
and data types*



Cost-effective

Pay only for what you use

*SQL: Save 30%-90% on
per-query costs
through compression*

*Python: Minimize idle
compute charges*

Amazon Athena

- Accelerate data lake queries with Amazon Athena and Amazon S3 Express One Zone
- AWS CloudTrail Lake data now available for zero-ETL analysis in Amazon Athena
- AWS Glue Data Quality announces anomaly detection and dynamic rules
- Amazon Athena adds CloudWatch metrics for Provisioned Capacity
- Amazon Athena adds cost-based optimizer to enhance query performance
- New JDBC driver now available for Amazon Athena
- Amazon Athena announces one hour reservations for Provisioned Capacity

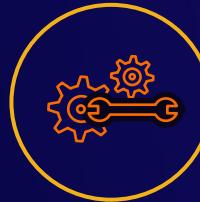
Streaming - MSK & Kinesis





Amazon MSK

Securely stream data with a fully managed, highly available Apache Kafka service



Automate provisioning, configuring, and tuning
Eliminate operational overhead - including the provisioning, configuration, and maintenance of Apache Kafka and Kafka Connect clusters



Fully compatible with open-source Apache Kafka
Use applications and tools built for Apache Kafka out of the box, with no application code changes required



Highly secure
Easily deploy secure, production-ready applications using native integrations to an Amazon VPC for authentication and authorization



Lower cost
Keep costs low with fully managed Apache Kafka, offered at as low as 1/13 the cost of other providers

Amazon MSK

STEAMING

- *Amazon MSK now supports Graviton3-based M7g instances for new provisioned clusters*
- *Amazon MSK Serverless now supports all programming languages*
- *Amazon MSK extends AWS IAM support to all programming languages for new clusters*
- *Amazon MSK adds check for too many partitions to AWS Trusted Advisor*
- *Amazon MSK now sends storage capacity alerts*
- *Amazon MSK adds support for Apache Kafka version 3.6.0 with Tiered Storage*

Amazon Kinesis

Easily collect, process, and analyze data streams in real time

Amazon Kinesis *Data Streams*



Collect and store data streams for analytics

Amazon Kinesis *Data Firehose*



Load data streams onto AWS data stores

Amazon Kinesis *Data Analytics*



Analyze data streams with Amazon Kinesis Data Analytics Studio or Apache Flink

Amazon Kinesis *Video Streams*



Collect and store video streams for analytics

Amazon Kinesis Data

- *Amazon Kinesis Data Streams launches cross-account access with AWS Lambda*
- *Amazon Kinesis Data Streams announces Amazon EventBridge Pipes console integration*

AWS Lake Formation



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

AWS Lake Formation

GOVERNANCE

- *AWS Lake Formation data filters now support permissions on nested data*
- *Enforce fine-grained access control via AWS Lake Formation with Open Table Formats on Amazon EMR*
- *AWS Lake Formation launches Hybrid Access Mode for AWS Glue Data Catalog*



Data Governance on AWS



© 2024, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon DataZone

Unlock data across organizational boundaries with built-in governance



Govern data access
across organizational
boundaries



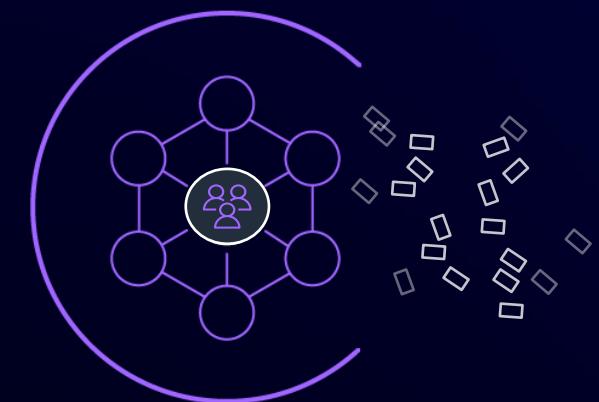
Connect people and
data through shared tools
to drive business insights



Automate data
discovery and cataloging
with machine learning

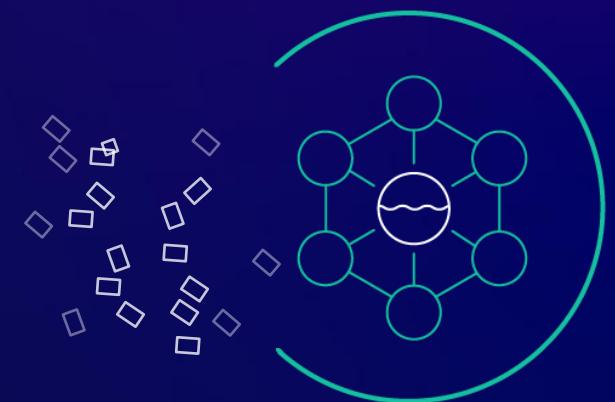
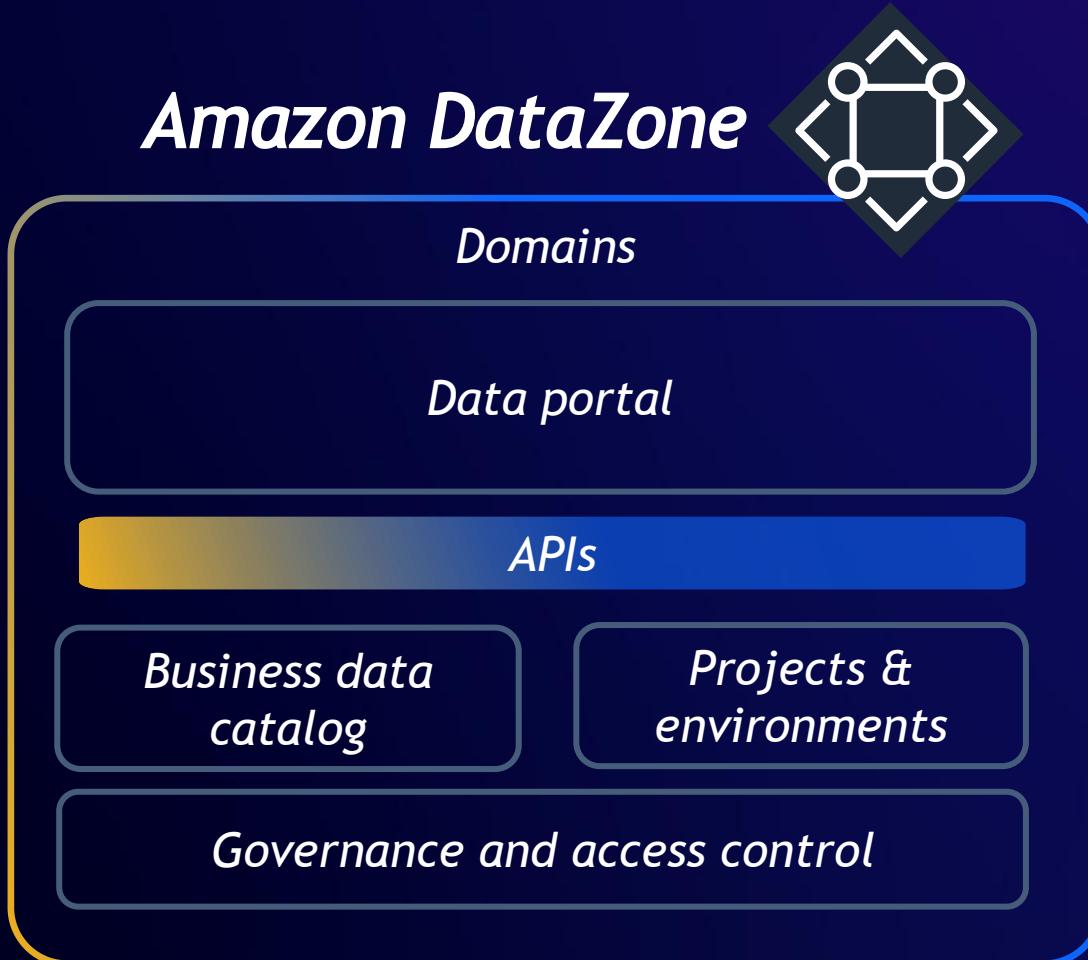
Core components of Amazon DataZone

ENABLE COMPANIES TO TREAT DATA AS A PRODUCT



Data producers

Bring data from different sources, and across accounts/regions



Data consumers

Simplify access, collaboration and consumption using different tools



AI recommendations for descriptions in Amazon DataZone

*Automate labor-intensive process of
data cataloging*



Aid search and discovery of valuable datasets

AI-generated business descriptions to expand datasets discovered for analysis



Guide data usage

AI-driven recommendations on how best to use datasets, boosting appropriate and effective data use



Increase efficiency in data documentation

Reduced manual work of data cataloging

AI recommendations for descriptions at work!

The screenshot shows the AWS Data Catalog interface for a project named "Finance Data". The top navigation bar includes "FINANCE DATA", a search bar, and user account information. The main dashboard displays a "Project overview" section with counts for Published Assets (0), Subscribed Assets (0), Data Sources (1), Environments (1), and Members (1). Below this, there are four cards under "Working with Projects": "Find data" (View Subscribed Data, View Pending Requests), "Publish data" (Create Data Source, View Incoming Requests), "Manage data catalog" (View Metadata Forms, View Glossaries), and "Set up environments" (Create Environment, Create Profile). To the right, an "ANALYTICS TOOLS" sidebar shows an active "Finance Ingest" environment for Amazon Athena. The bottom of the page includes standard footer links like Privacy, Site terms, and Copyright information.