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

ANT329

Turn data to insights in seconds with secure and reliable Amazon Redshift

Gal Barnea

Senior Manager Database Engineering
AWS

Yanzhu Ji

Product Manager, Amazon Redshift
AWS

Zafrir Babin

VP Product Development
ADP

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

Agenda

- Why become a data-driven organization
- Amazon Redshift to turn data into insights securely
- ADP use case
- Q&A

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

Why become a data-driven organization?



Make better decisions



Respond better to the unexpected



Improve efficiency



Uncover new opportunities



Enhance customer experience and engagement

aws

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

Modern data architecture

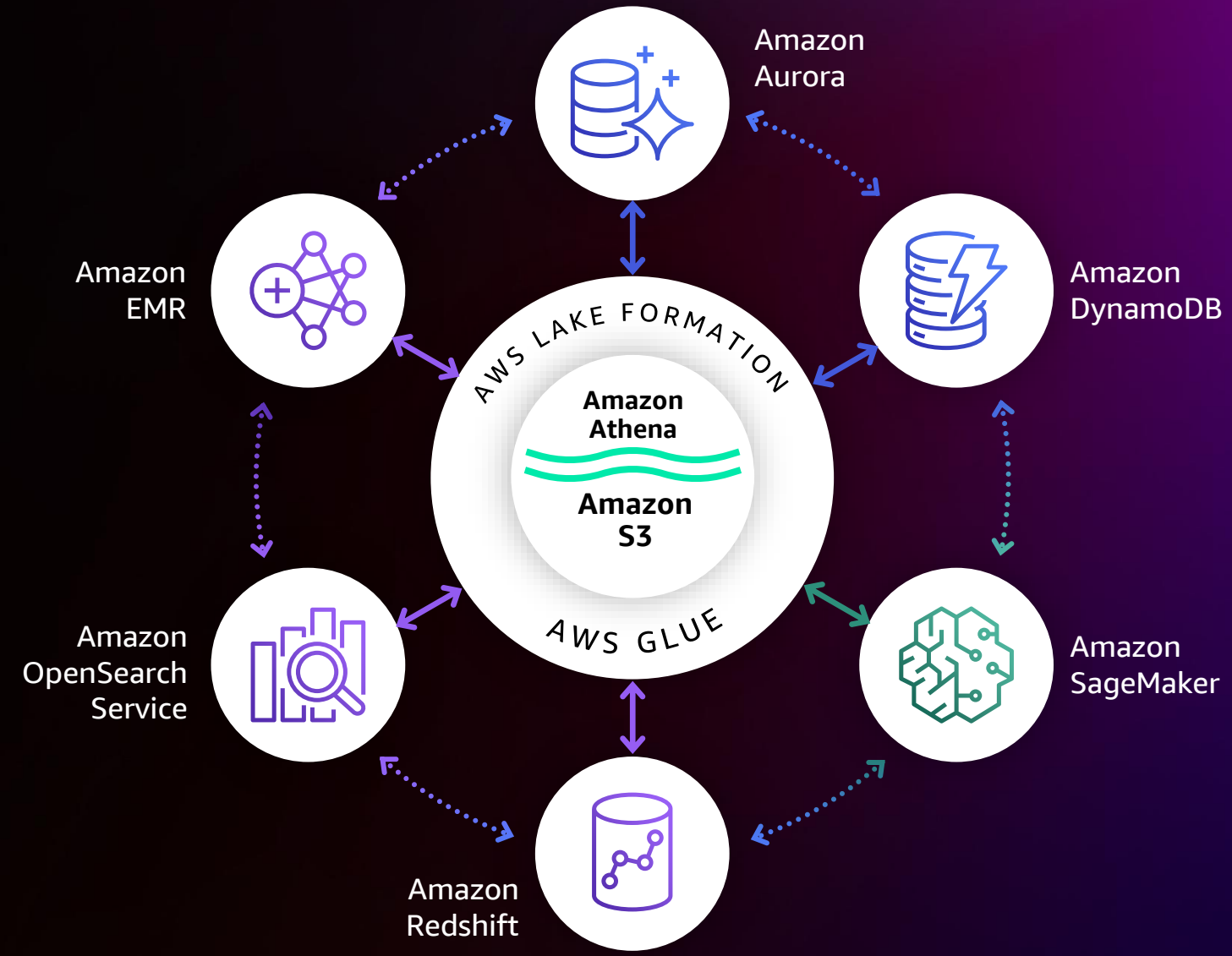
Scalable data lakes

Purpose-built data services

Seamless data movement

Unified governance

Performant and cost-effective

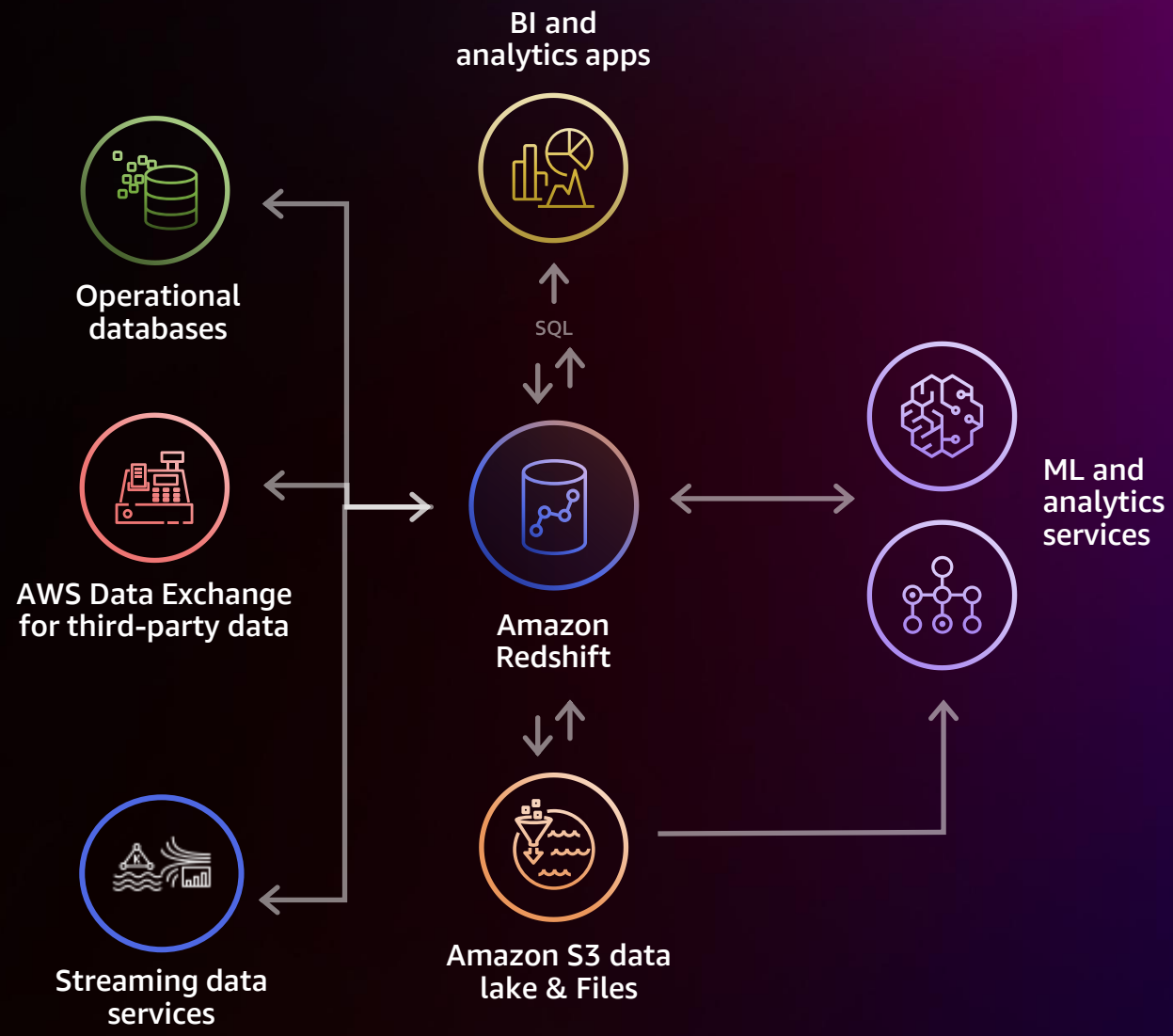


aws

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

Amazon Redshift

ANALYZE ALL YOUR DATA



Zero ETL | No data movement, data copying | Spark application development | Secure and consistent data sharing | Data Lake analytics | Auto Ingestion of S3 files | Use your favorite BI tool | ML in SQL

Built-in security and compliance

SECURITY AND COMPLIANCE FEATURES WITH NO EXTRA COSTS WITH AMAZON REDSHIFT

Authentication	Access control	Audit	Encryption	Helps achieve compliance
IAM integration	Role-based access control	AWS CloudTrail integration	Encrypted data in motion, data at rest	SOC
IDP integration and multifactor integration	Column-level & Row-level security	Amazon Macie integration	AWS KMS integration	PCI
	NEW Dynamic data masking (preview)	Audit logging to Amazon CloudWatch	Faster encryption for resize/restore	FedRAMP
	NEW AWS Lake Formation integration for data sharing (preview)		Tokenization with Lambda UDFs and third-party tools	HIPAA and others

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

Role-based access control

Simplifies database permissions management

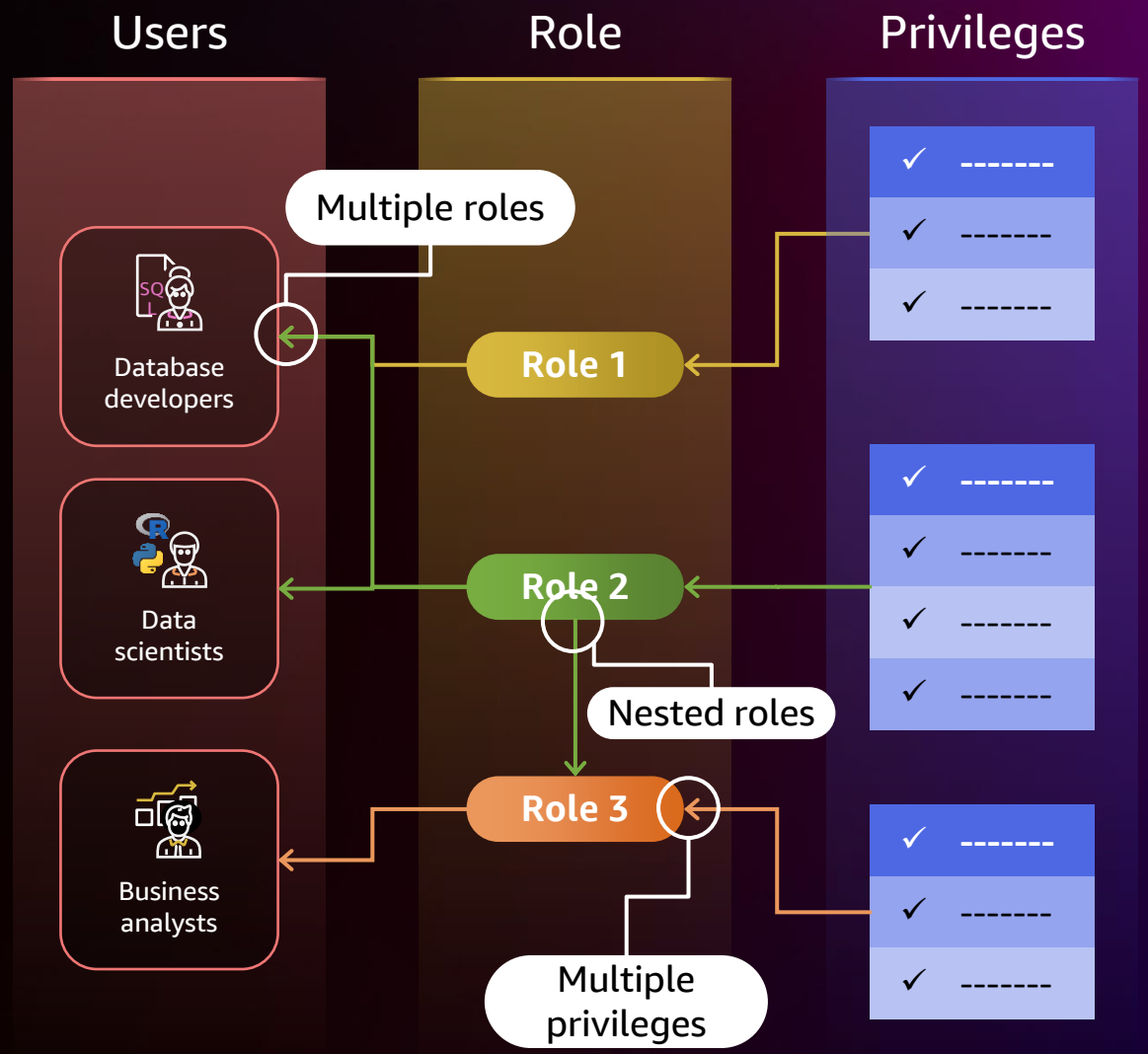
A role is a collection of database privileges. Assign one or more roles to a user

Fine-grained permissions for users

Assign both system and object privileges to a role. System privileges for commands

Improved security for different levels of admin users

Out-of-the-box system roles for root user, DBA, operator, security admins, or create your own role



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

Row-level security

Restrict row access based on roles

Security admin defines if a role can access specific records of data within a table based on a security policy

Enhance granular access control

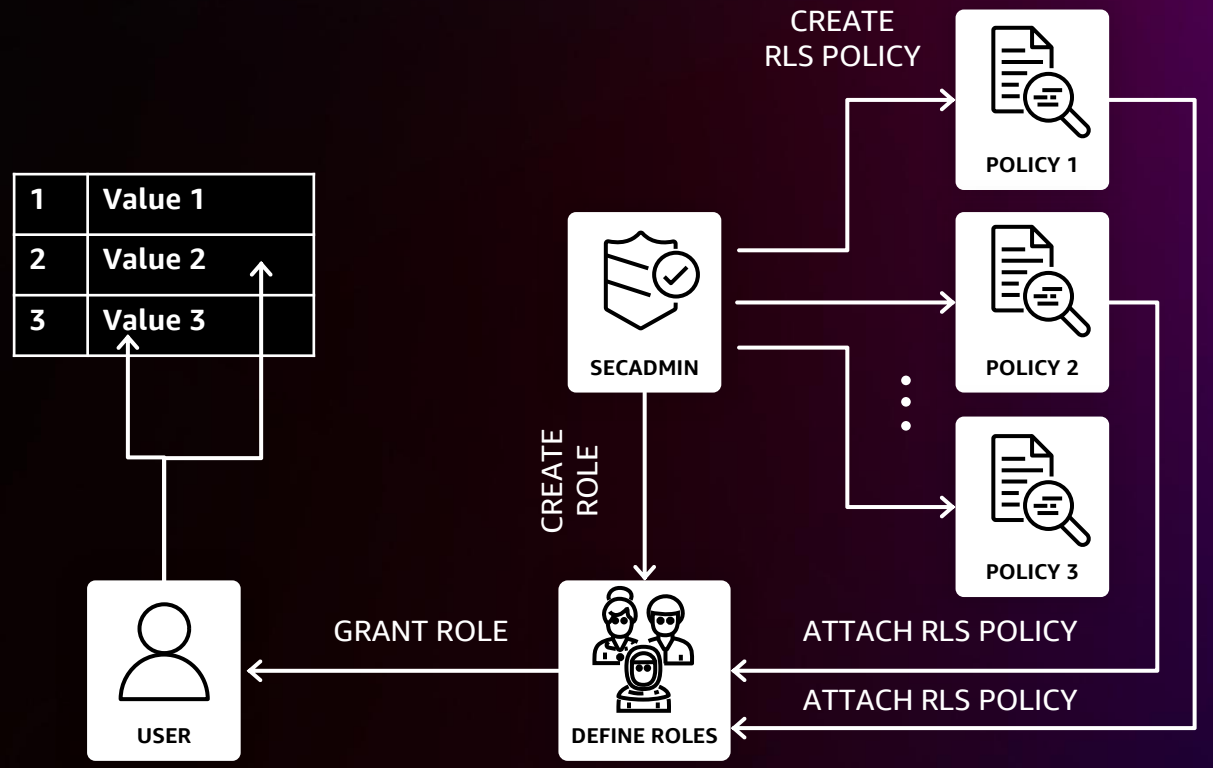
Built based on RBAC, and can work interactively with column-level access control

Combine multiple policies per user

When there are multiple policies defined per user, we will apply all the policies **with AND syntax**

No access if no policy applied

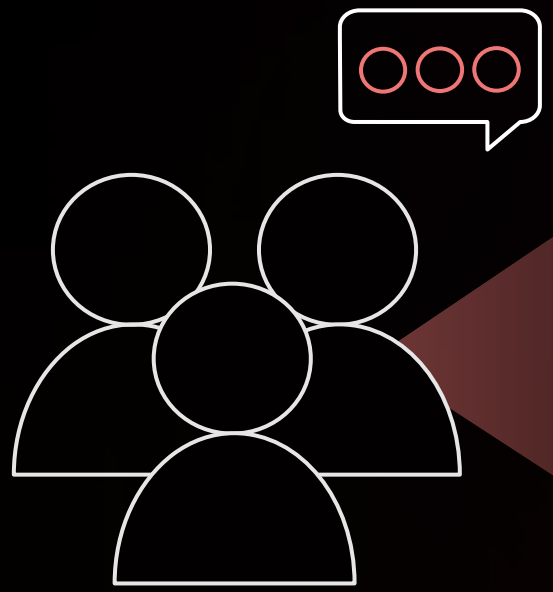
All data access is blocked when there is no applicable policy on an RLS-protected table



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

Accelerate encryption with faster cluster resize

WHAT WE ARE HEARING FROM CUSTOMERS



We need to encrypt our Amazon Redshift cluster based on security, compliance, and data consumption requirements

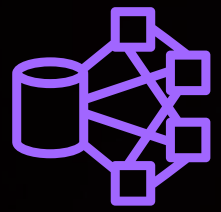
An encrypted cluster is required when we want to conduct cross-account, cross-Region data sharing, or integrations with AWS Data Exchange

We have to use classic resize technology to update encryption setting, which can take a few hours to days depending on the dataset size; during the resize process, the cluster will be read-only.

aws

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

Accelerate encryption with faster cluster resize



Amazon Redshift clusters

Process queries in minutes after being encrypted from an unencrypted state

Wait time is reduced from up to several days to minutes

Can be encrypted or changed to different encryption key in snapshot restore process

No node number limit when restoring cluster from snapshot

Volume	Source Cluster	Target Cluster	Classic Resize Duration (min)	Faster Classic Resize Stage1 Duration (min)	% faster
10TB	ra3 4xlarge - 6 nodes	ra3 16xlarge - 8 nodes	78	11	86%
10TB	ra3 16xlarge - 8 nodes	ra3 4xlarge - 2 nodes	738	11	99%
10TB	dc2 8xlarge - 6 nodes	ra3 4xLarge - 2 nodes	706	8	99%
3TB	ra3 4xLarge - 2 nodes	ra3 16xLarge - 4 nodes	53	11	79%
3TB	ra3 16xLarge - 4 nodes	ra3 4xLarge - 2 nodes	244	7	97%
3TB	dc2 8xlarge - 6 nodes	ra3 4xLarge - 2 nodes	251	7	97%

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

** This feature only applies to RA3 target clusters*

Overview:
Amazon Redshift native IdP Federation



Native integration
Amazon Redshift now provides native integration with Microsoft Azure Active Directory (AD)



Simplify administration
Register external Identity Provider(IdP) with Amazon Redshift using a SQL command



Auto-create Redshift roles
Automatically create Redshift Roles



Auto-create Redshift user
Automatically create Redshift User

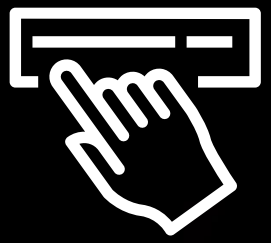


Group membership natively mirrored with Amazon Redshift roles
External user group membership natively mirrored with Amazon Redshift roles

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

New Amazon Redshift Enhancements

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



NEW [PREVIEW]

Amazon Redshift supports dynamic data masking

NEW

Easily protect sensitive data by managing data masking policies through an SQL interface

User can define the way to do the data masking. Modify sensitive or PII data with fictitious content viable for software development, testing, analytics

Restrict different levels of permissions to masked data with Role-Based Access Control

ID	Geo-location	Name	Phone number
123	WA	Ana	123-456-3568
124	NY	Alice	123-457-****
125	WA	Bruce	123-457-3569
126	CA	Chris	123-457-****
130	CA	Sharon	123-457-****
	Condition column		Mask column

baffle

"We are excited about utilizing the Amazon Redshift Dynamic Data Masking capability to allow our customers to achieve the goal of protecting sensitive data throughout the analytics pipeline from secure ingestion to responsible consumption."

Ameesh Divatia

CEO & Co-Founder, Baffle.io

aws

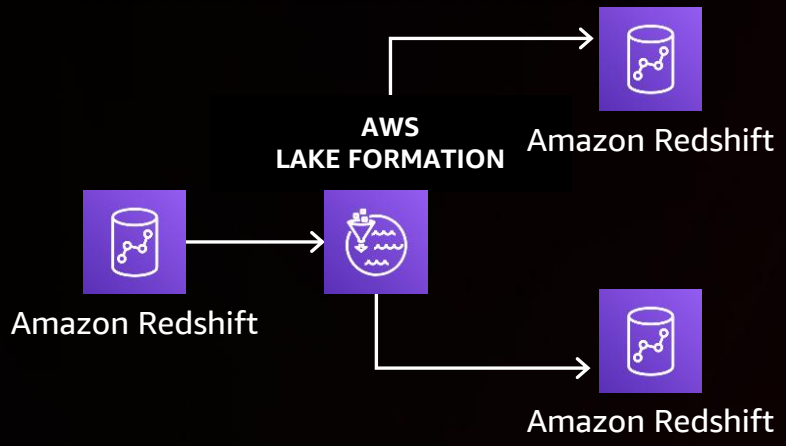
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW

NEW [PREVIEW]

Data sharing access control with AWS Lake Formation

Centrally manage data sharing with AWS Lake Formation



Centrally manage granular access to data across all consuming data services

Improve security and governance with row-level and column-level granular permissions on data sharing

No manual scripting or complex querying

Define policies once and enforce those consistently for multiple consumers

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

NEW



INTRODUCING [PREVIEW]

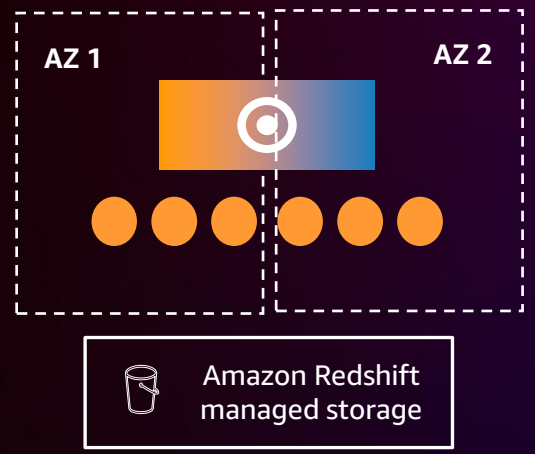
Amazon Redshift Multi-AZ

Highly resilient data warehouse

Auto-failover with zero data loss and no manual intervention

Easy management through a single endpoint

Active-active workload processing across AZs



aws

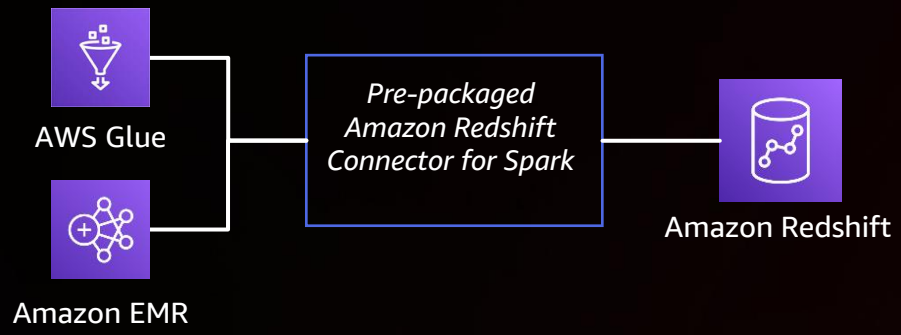
© 2022, Amazon Web Services, Inc. or its affiliates. All rights reserved.

NEW

NEW [GENERAL AVAILABILITY]

Amazon Redshift integration for Apache Spark

Simplify and speed up Apache Spark applications accessing Amazon Redshift data from AWS analytics services



Author Apache Spark applications using Java, Python, Scala, with access to rich, curated data in your data warehouse

No manual setup and maintenance of uncertified versions of Spark-Redshift open-source connectors

Improved performance with only relevant data moved from Amazon Redshift to consuming applications

Improved security with IAM-based credentials

aws

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

About ADP and ADP analytic application

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

ADP: 70+ years of innovations

Invented payroll outsourcing

First in the industry to deliver automation

First HCM vendor in the cloud

First to create an HCM Marketplace

First to provide a mobile HCM app



990K clients

39M workers globally

1 in 6 US workers

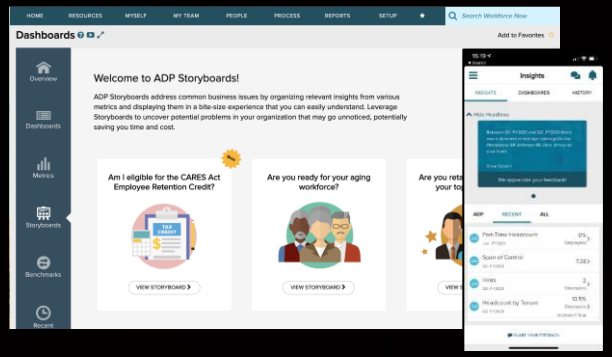
2.8B annual mobile logins

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

Data Cloud People Analytics

PRODUCTS

People Analytics



Metrics: 90+ HCM metrics on desktop

Theme-based dashboards:

question-based approach to solving key business problems (e.g., reduce turnover, manage costs, drive diversity, optimize recruiting)

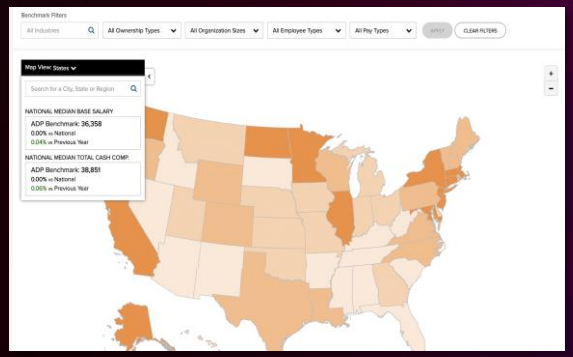
Storyboards:

Story-telling approach to analytics

Executive and Manager Insights:

AI-driven, push insights via ADP mobile app

Enhanced Insights



Benchmarks:

on 30+ metrics (HR, Payroll, Benefits)

Data Mashup:

Tableau connectors & Data Ingestion

Compensation Explorers:

compare employee pay against market either as a single job look-up or a portfolio analysis

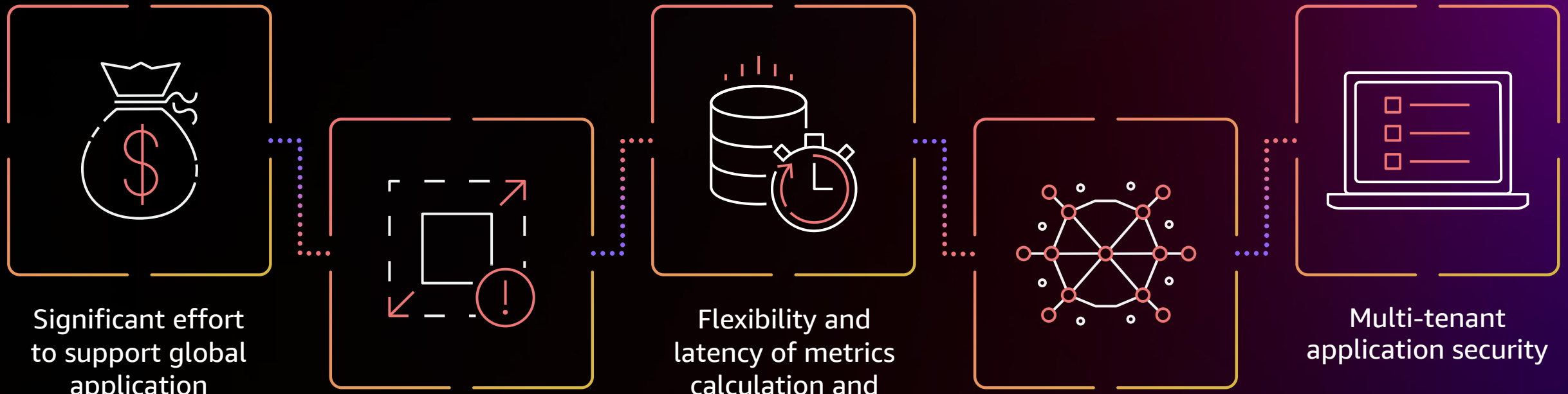
Talent Market Insights:

external labor market (supply-side) insights to help recruiters creatively expand talent pools and generate new leads

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

Challenges with the ADP analytics application

PROBLEMS THAT NEEDED TO BE RESOLVED



Significant effort to support global application deployment and DR

Scalability

Flexibility and latency of metrics calculation and data ingestion

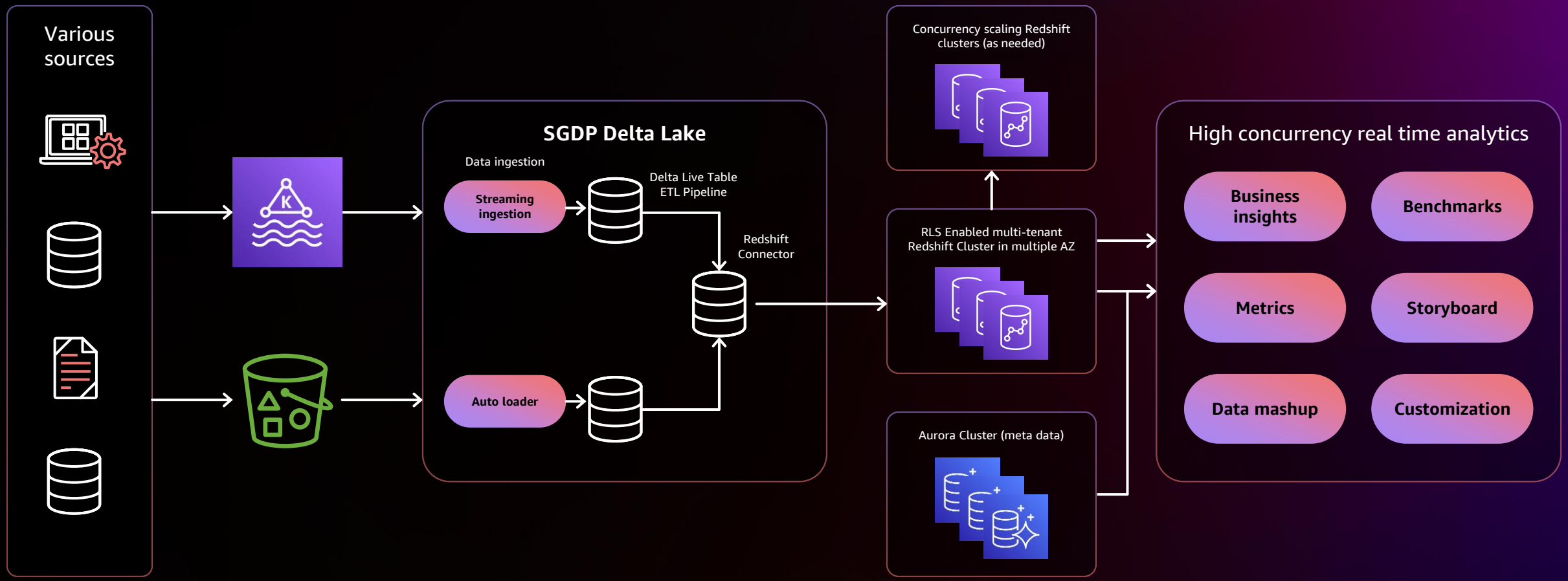
Lower environment and testing

Multi-tenant application security

aws

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

ADP Analytics Application Architecture



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

Next gen AWS Redshift advantage

Redshift performance

Interactive matrices calculation for clients. No more offline calculation

One-click global deployment

An achievable target through CI/CD pipelines

Flexibility

New client metrics can be created within the application

DRS support

With Redshift DR support availability, application is minimal effort and low cost

Cost savings

Automatic scale up and down by Redshift saves ADP Money

Scaling

Redshift offers almost linear horizontal scaling

aws

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

Support for multitenant security

ADP multitenant security required complex code to support:



One client does not see other client data



One manager does not see other managers' data

HR practitioners only see data within their span of control



ADP partner portfolio analytics

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

Support for multitenant security



Amazon Redshift Row Level Security

Helps simplify the implementation and prevent costly security incidents due to defects



No cross-tenant data access

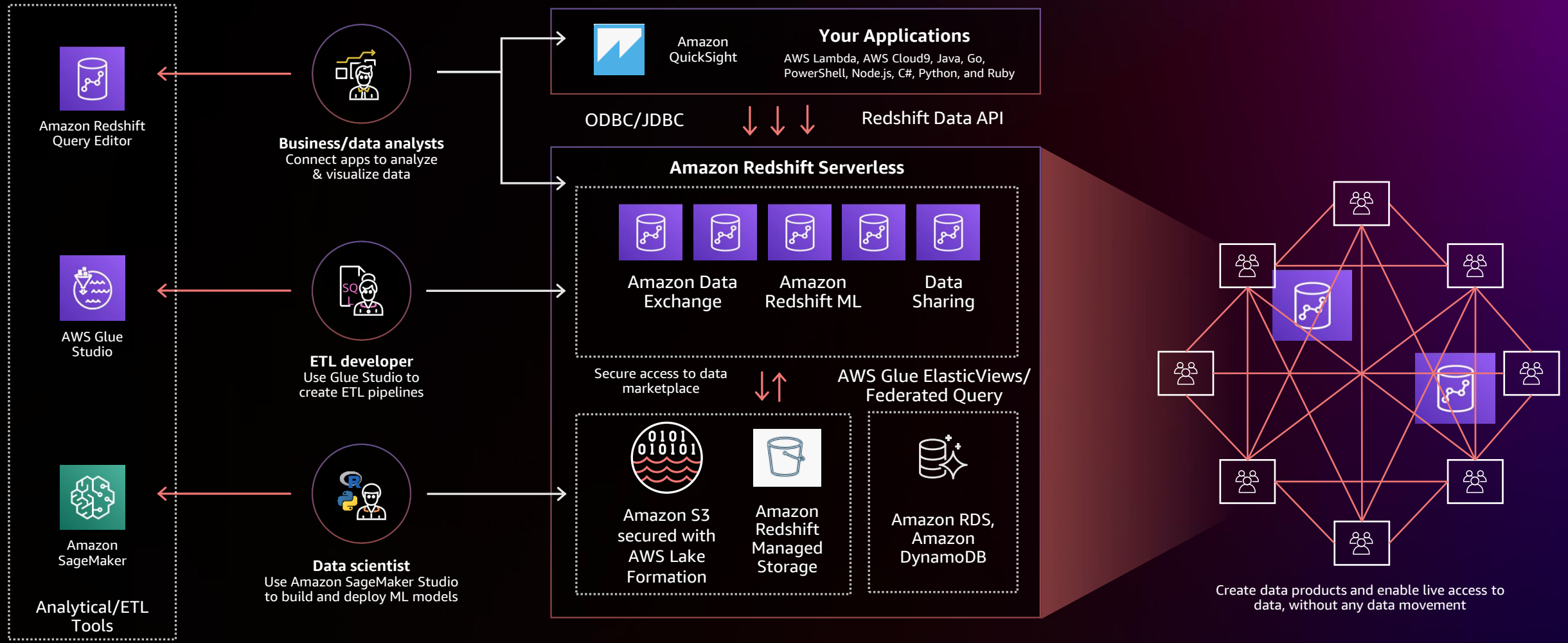
Minimal code changes to enforce RLS policies

Role-based system privileges

No implementation changes with new tenants

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

Modern data architecture



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

Thank you!

Gal Barnea
galbarn@amazon.com

Yanzhu Ji
linkedin.com/in/yanzhu-ji

Zafrir Babin
linkedin.com/in/zafrir-babin-20207b4/



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

aws

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