AWS re:Invent

ANT213-R1

# State-of-the-art cloud data warehousing, featuring Asurion

**D. Britton Johnston**
Director of Product Management, Redshift

**Cijo Isaac**
Architect, Enterprise Data Services
Asurion

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws

# What to expect

1

## Hear

about our new Amazon Redshift announcements and recent innovations

2

## Understand

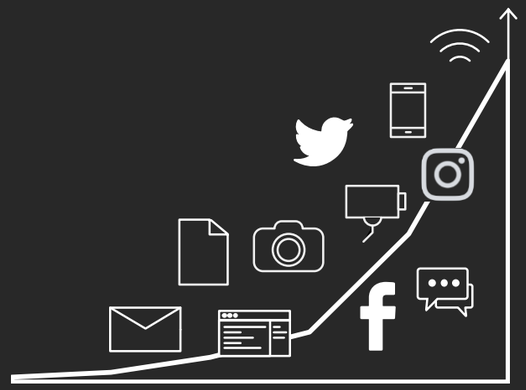
our data warehouse strategy and how it can benefit your organization

3

## Plan

how you can adopt the exciting new innovations in Amazon Redshift

# The new realities customers are facing



Explosion of data

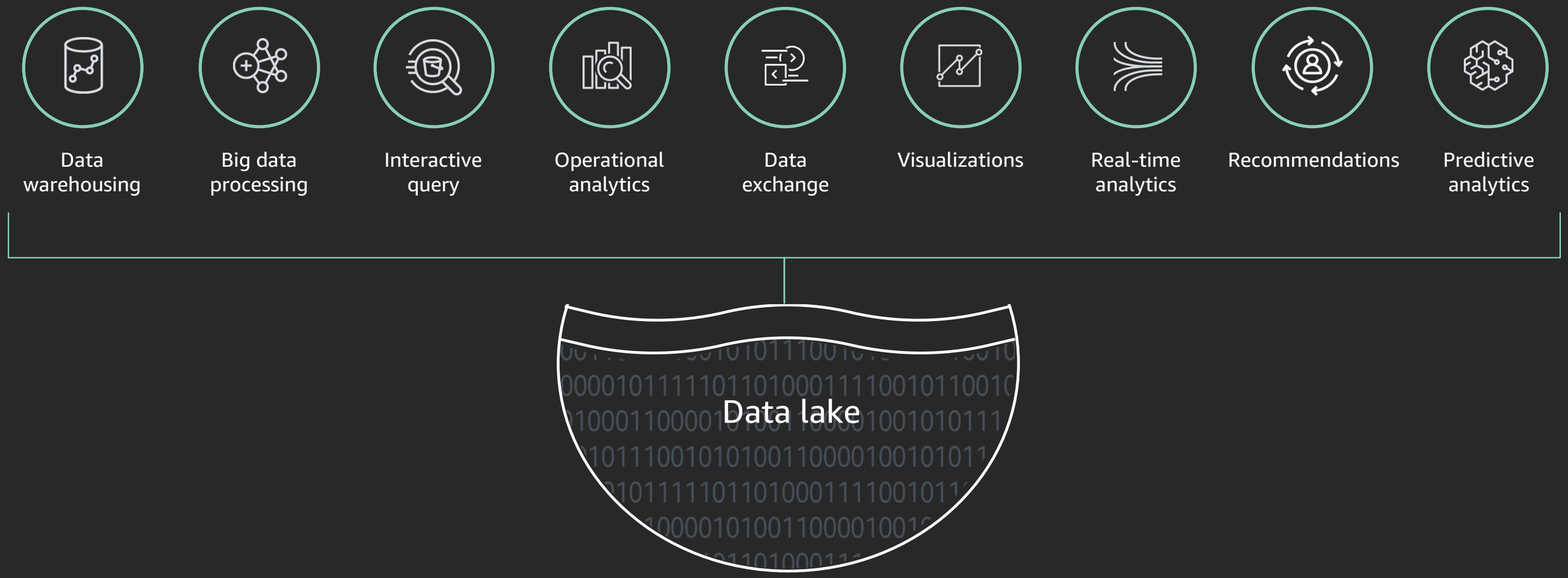


Expansion of personas



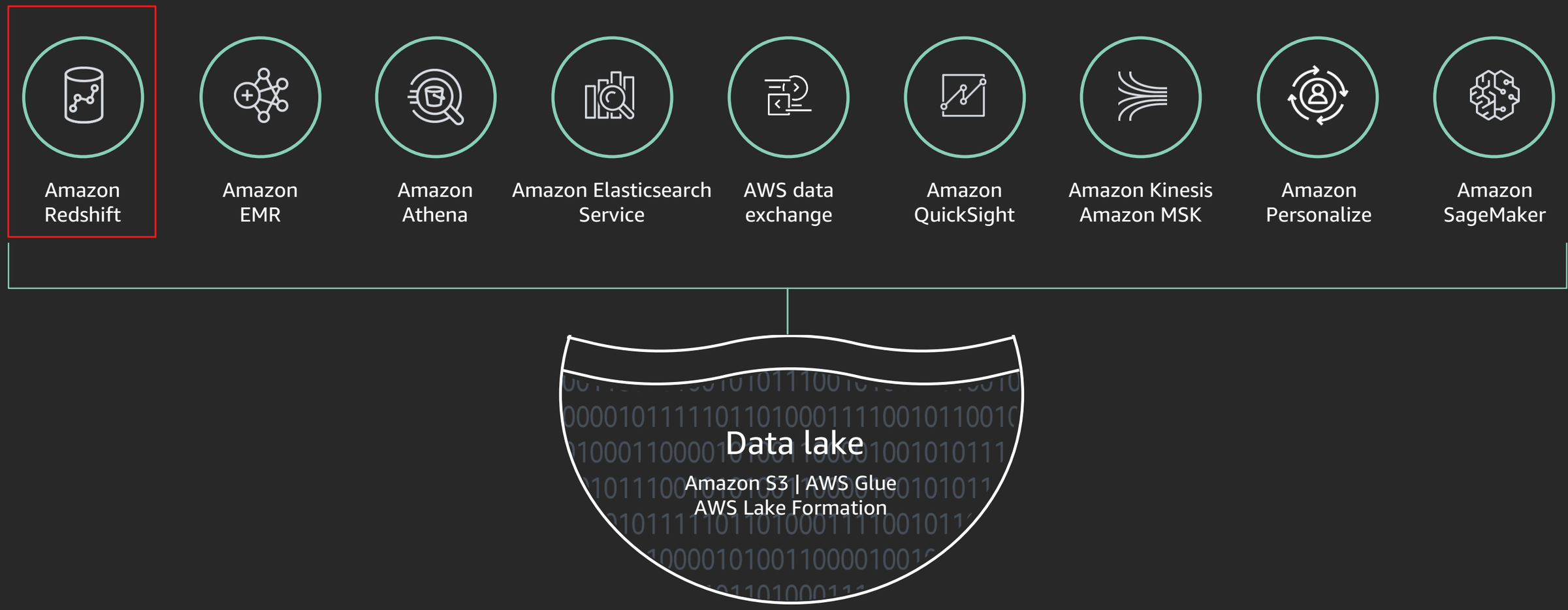
Demand for faster decision making on real-time data

# Any type of analytics on the data lake



# Any type of analytics on the data lake

## Most comprehensive analytics platform



# Data warehousing trends



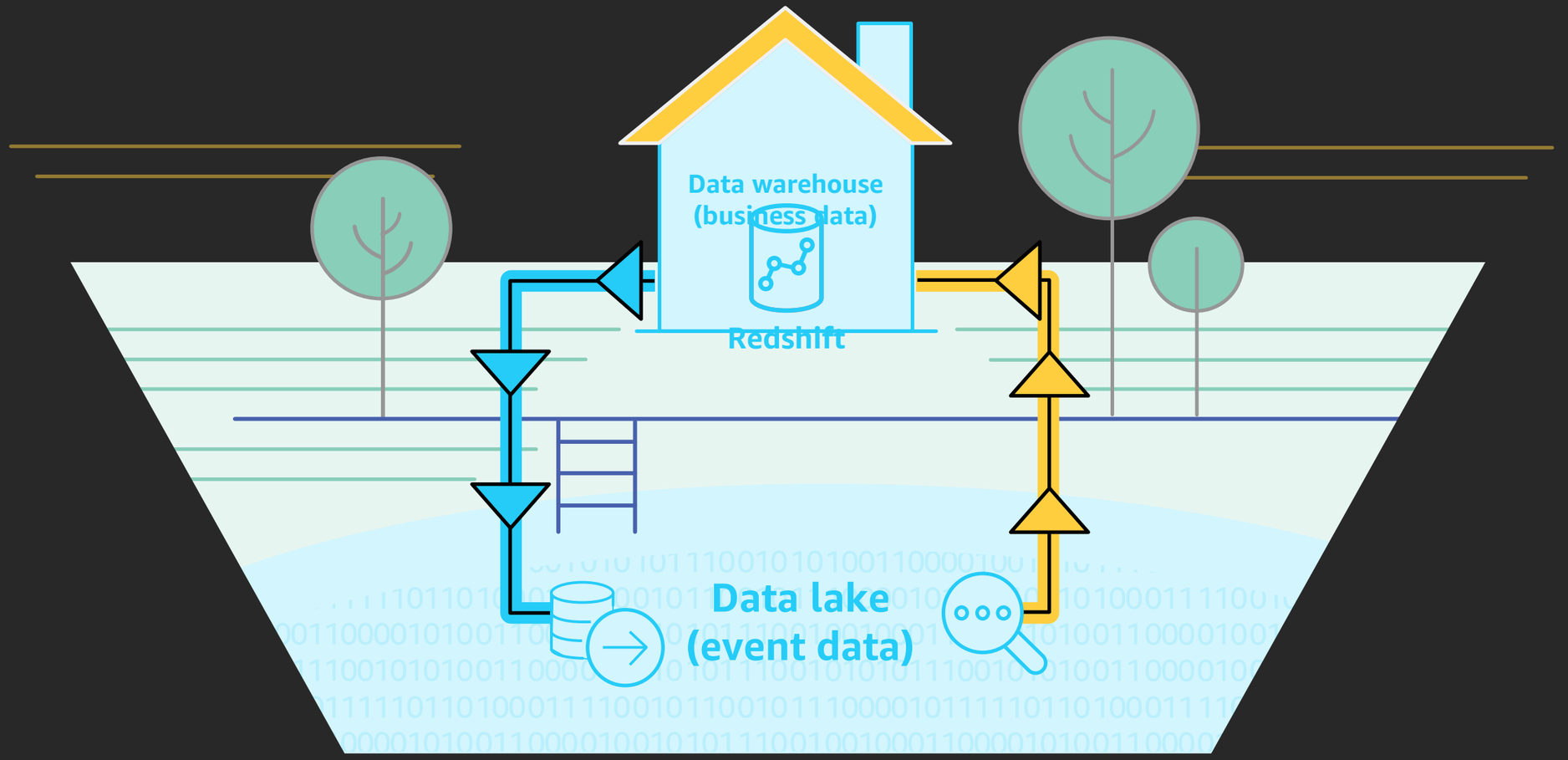
Migrations to the cloud




Exponential growth of event data



End-to-end insights from analyzing all your data




# Customers moving to data lake architectures

**Amazon Redshift enables you to have a lake house approach**

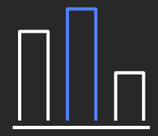
# Amazon Redshift benefits

## Tens of thousands of customers use Amazon Redshift & process exabytes of data per day



### Data lake & AWS integrated

AWS Lake Formation catalog & security, exabyte querying (spectrum & federated), AWS integrated (AWS DMS, Amazon CloudWatch)



### Best performance

3x faster than other cloud data warehouses



### Lowest cost

Up to 75% less than other cloud data warehouses & predictable costs



### Most scalable

Virtually unlimited elastic linear scaling



### Most secure & compliant

AWS-grade security, (e.g., VPC, encryption with AWS KMS, AWS CloudTrail), Certifications such as SOC, PCI, DSS, ISO, FedRAMP, HIPAA



### Fully managed

Easy to provision & manage, automated backups, AWS support, 99.9% SLAs



# Amazon Redshift

is the most popular cloud data warehouse



# Data warehouse: Amazon Redshift

- Industry’s first cloud data warehouse to launch in 2012
- Disrupted data warehousing pricing economics @ $1,000/TB/year
- Highly contested category with lots of followers and imitators
- Remains the most widely used, fastest, highest scale, and lowest cost cloud data warehouse

# Amazon Redshift has been innovating quickly

Robust result set caching

Large # of tables support ~20000

Copy command support for ORC, Parquet

**IAM role chaining**

**Elastic resize**

**Groups**

Redshift Spectrum: date formats, scalar json and ION file formats support, region expansion, predicate filtering

**Auto analyze**

Health and performance monitoring w/Amazon Cloud watch

**Automatic table distribution style**

CloudWatch support for WLM queues

Performance enhancements—hash join, vacuum, window functions, resize ops, aggregations, console, union all, efficient compile code cache

Unload to CSV

**Auto WLM**

~25 Query Monitoring Rules (QMR) support

**AQUA**

Concurrency Scaling

# 200+

## new features in the past 18 months

DC1 migration to DC2

Resiliency of ROLLBACK processing

Manage multi-part query in AWS console

Auto analyze for incremental changes on table

**Spectrum Request Accelerator**

Apply new distribution key

Amazon Redshift Spectrum: Row group filtering in Parquet and ORC, Nested data support, Enhanced VPC Routing, Multiple partitions

Faster Classic resize with optimized data transfer protocol

Performance: Bloom filters in joins, complex queries that create internal table, communication layer

**Amazon Redshift Spectrum**: Concurrency scaling

AWS Lake Formation integration

**Auto-Vacuum sort, Auto-Analyze** and Auto Table Sort

**Auto WLM with query priorities**

**Snapshot scheduler**

Performance: join pushdowns to subquery, mixed workloads temporary tables, rank functions, null handling in join, single row insert

**Advisor recommendations for distribution keys**

**AZ64 compression encoding**

**Console redesign**

**Stored procedures**

**Spatial Processing**

Column-level access control with AWS lake formation

**RA3**

Performance of Inter-Region Snapshot Transfers

**Federated Query**

**Materialized Views**

**Manual Pause and Resume**

# Amazon Redshift: Key 2019 innovations

## Performance & scalability

- NEW! RA3 with Amazon Redshift managed storage
- PREVIEW! AQUA
- NEW! AZ64
- PREVIEW! Materialized views
- NEW! Spatial data support
- PREVIEW! Query federation across Amazon Redshift & Amazon Aurora
- NEW! Concurrency Scaling

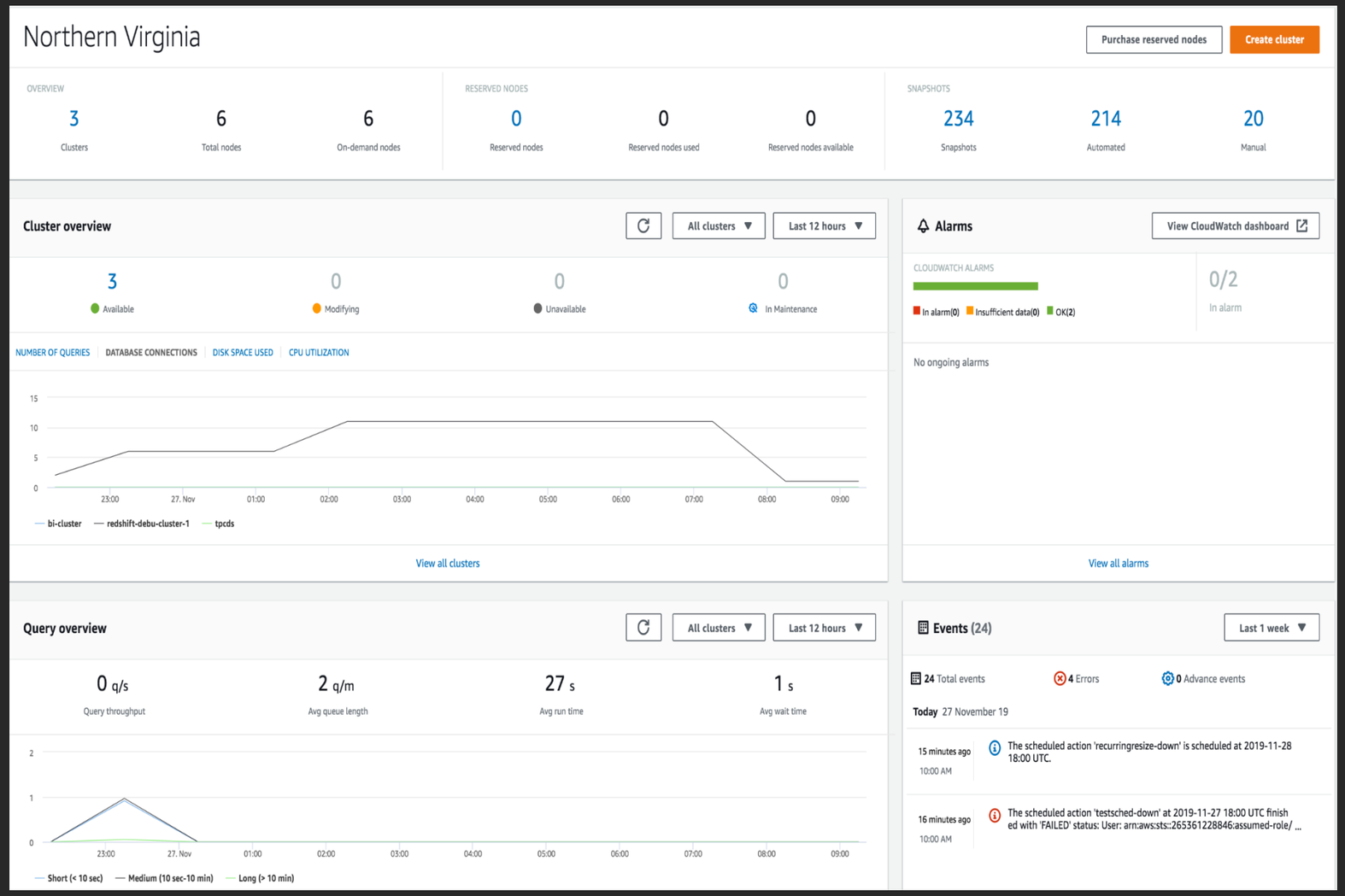
## Ease of use

- NEW! New management console
- NEW! Auto WLM: Query priorities
- NEW! Elastic resize scheduler
- NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort
- NEW! Stored procedures
- NEW! Distribution and sort key advisor
- NEW! Cross-instance restore
- NEW! Faster cross regional copy and change
- NEW! Auto Data Distribution
- NEW! Deferred Maintenance
- NEW! Elastic resize

## Data lake integration

- NEW! Data lake export in Parquet
- NEW! Spectrum Request Accelerator
- NEW! AWS Lake Formation integration

# New management console
## Intuitive, fast, and feature-rich




- Manage multiple data warehouses with command center-like dashboard
- Reduced time to diagnose when unexpected happens
- Share Query Editor with non-admin users
- Visually analyze your query results right in the query editor

# Amazon Redshift out-of-the-box performance is now >2x faster

Maintaining performance and scalability leadership through continuous telemetry and benchmarking

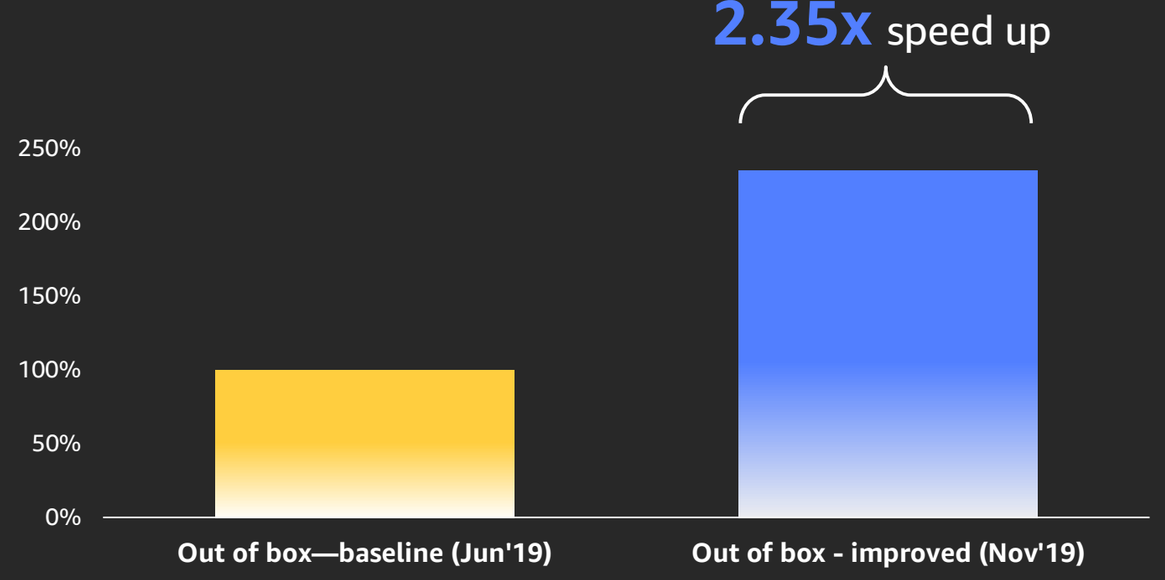
Performance innovations:

- AZ64 encoding
- Bloom filters for collocated & broadcast JOIN queries
- Enhanced planner for modern hardware CPUs and networking
- HLL (HyperLogLog) statistics
- Cache-optimized aggregation and join processing

***Cloud DW benchmark, 30TB**

**Normalized queries/hour (QPH)**

**(Higher is better)**




* Cloud DW benchmark is based on TPC-DS (v2.10) with no query modifications done

# AZ64—A new compression encoding for optimized storage and high query performance

New compression encoding for high compression ratio and improved performance

Efficient compression on small groups of data

Leverages SIMD instructions for parallel processing

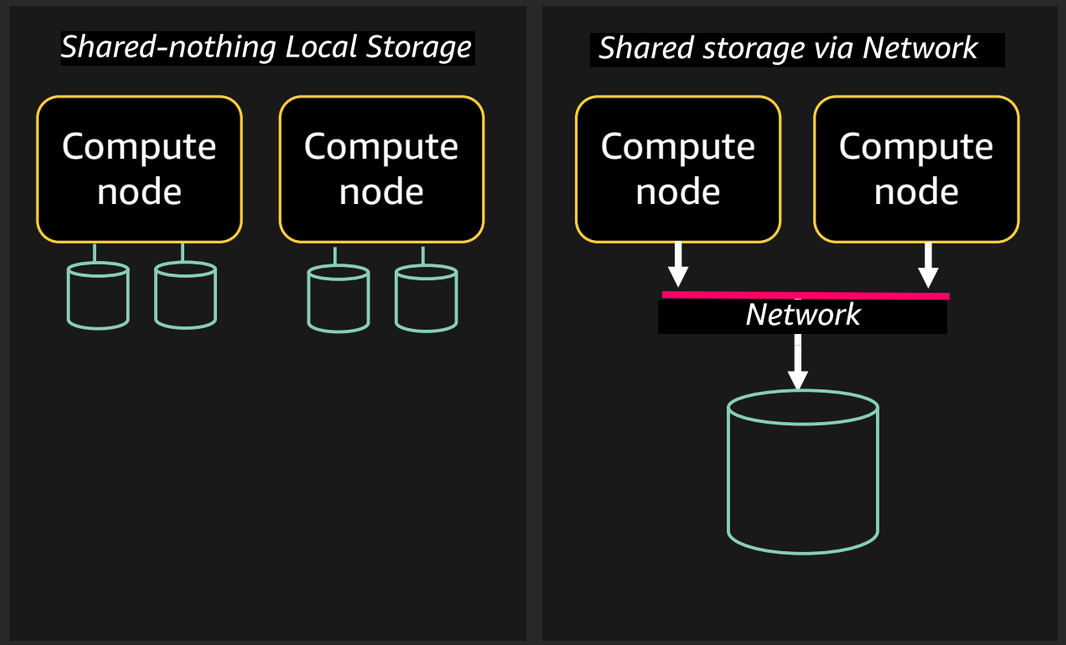
Applicable for all numeric and Amazon Redshift date/time datatypes

Leveraged with Create Table and Alter Table statements

**AZ64 offers better storage savings and high query performance compared to other Amazon Redshift encodings**

| | AZ64 storage savings | AZ64 performance speed ups |
|---|---|---|
| RAW | 60–70% less storage | 25–30% faster |
| LZO | 35% less storage | 40% faster |
| ZSTD | Comparable footprint | 70% faster |

# Evolution – shared nothing to disaggregated compute




*There has be a better way?*

Cloud-based local storage enabled DW in the cloud until today has provided unequaled performance

As cloud evolved, shared storage enabled a change providing flexibility at the cost of performance

What if we could get the performance of local storage, the flexibility of shared storage without a network performance penalty?

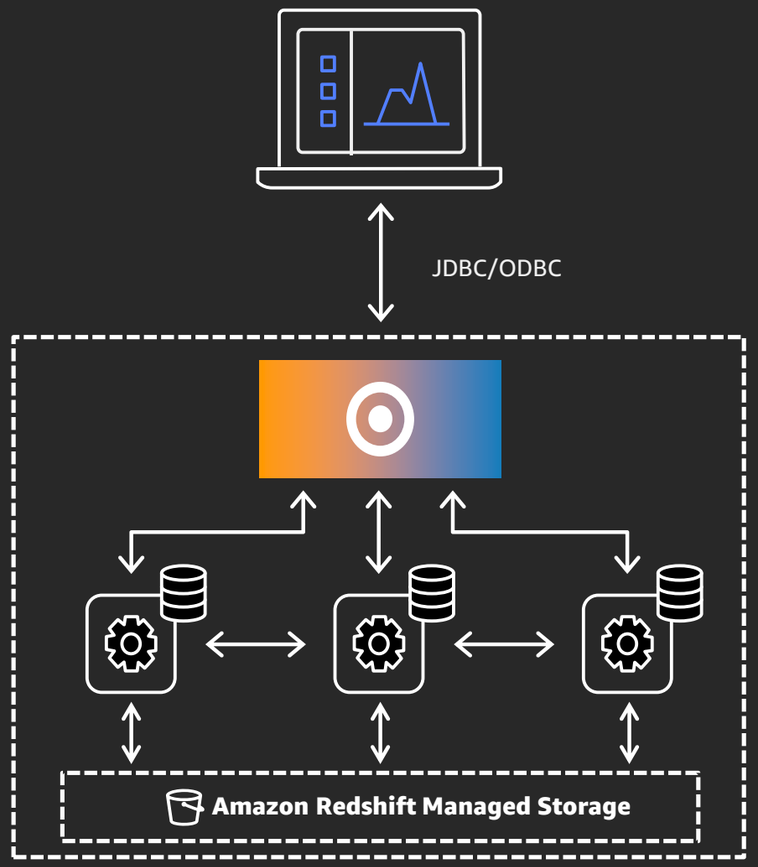
# RA3: New 3rd generation Amazon Redshift compute node

Managed storage | High-speed cache | High-bandwidth networking

Scale data warehouse based on typical workload and scale for peak needs

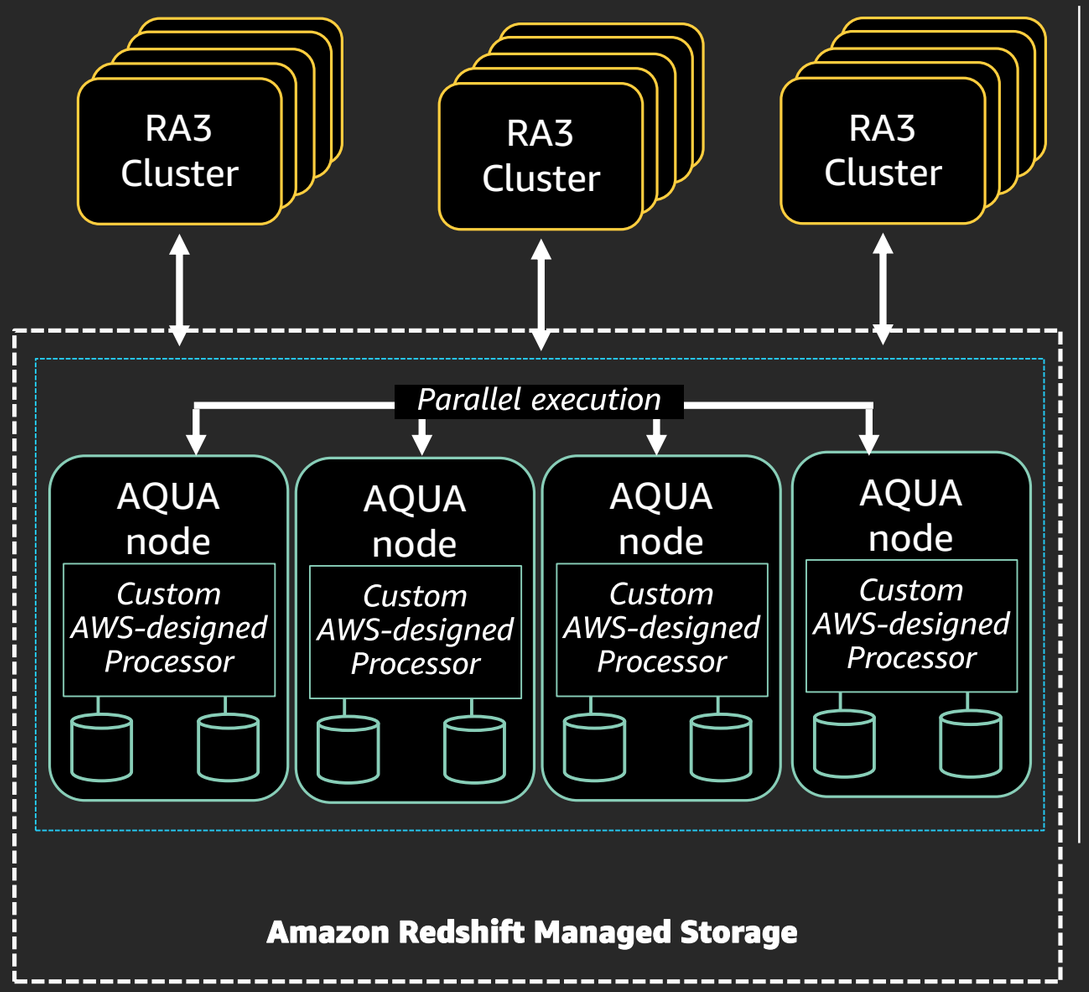
Pay separately for storage and compute

Automatic, no changes to any workflows, no need to manage storage



# AQUA - Advanced Query Accelerator

## A new distributed and hardware-accelerated processing layer that will make Amazon Redshift **10x faster** than any other cloud data warehouse without increasing cost




- Minimize data movement over the network by pushing down operations to AQUA nodes

- AQUA nodes with custom AWS-designed analytics processors to make operations (compression, encryption, filtering, and aggregations) faster than traditional CPUs

- Available in Preview only with RA3 with no code changes required

# Amazon Redshift Federated Query

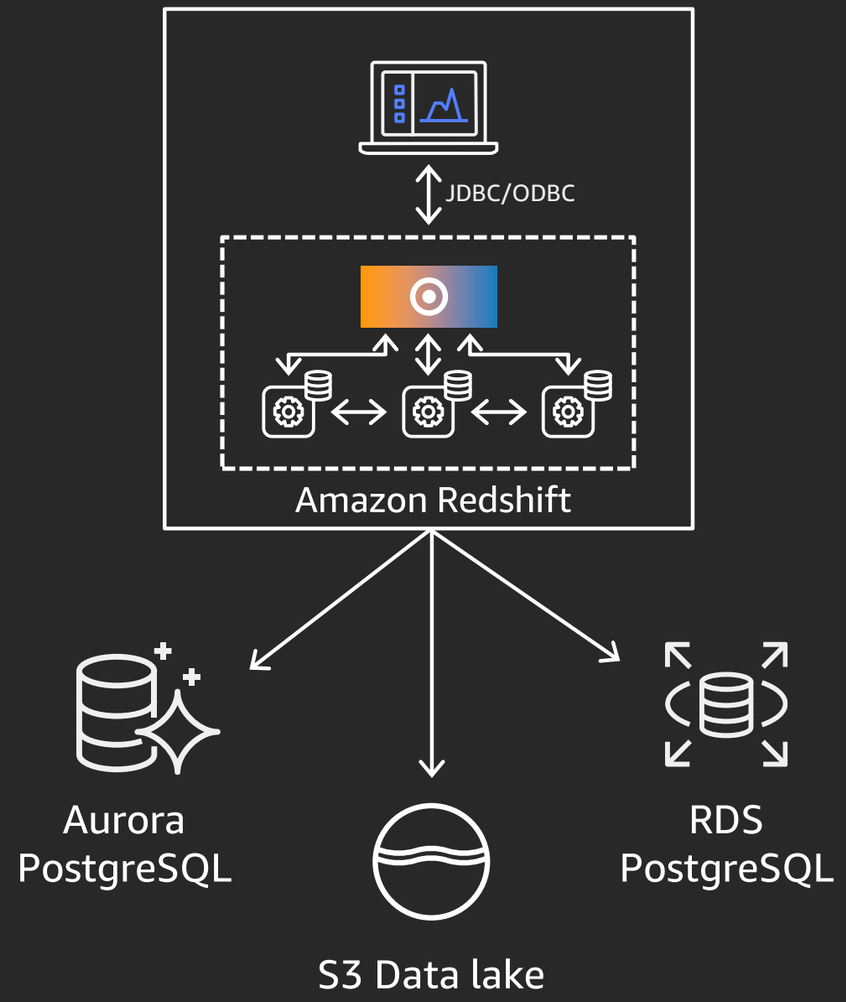
Queries on RDS and Aurora PostgreSQL databases

Analytics on live data without data movement

Unified analytics across data warehouse, data lake & operational databases

Flexible and easy way to ingest data

Performant and secure access to data



# Amazon Redshift Materialized Views

## Compute once, query many times

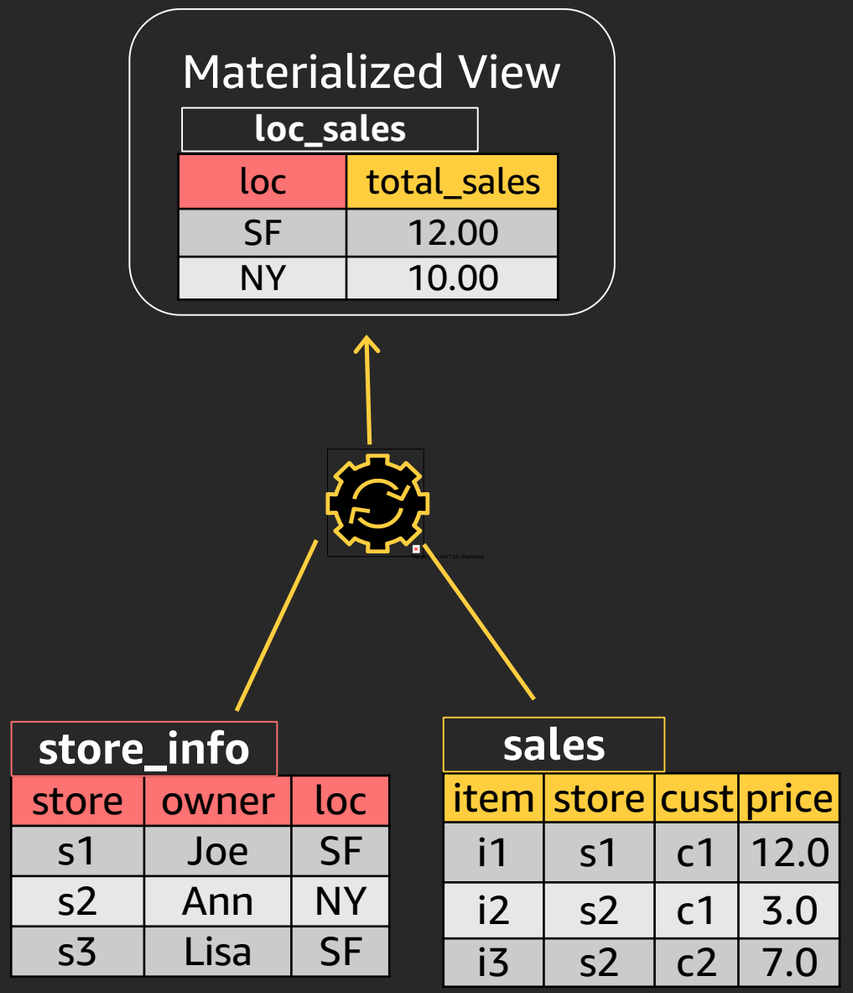
Speed up queries by orders of magnitude

- Joins, filters, aggregations, and projections

Simplify and accelerate ETL/BI pipelines

- Incremental refresh
- User triggered maintenance

Easier and faster migration to Amazon Redshift



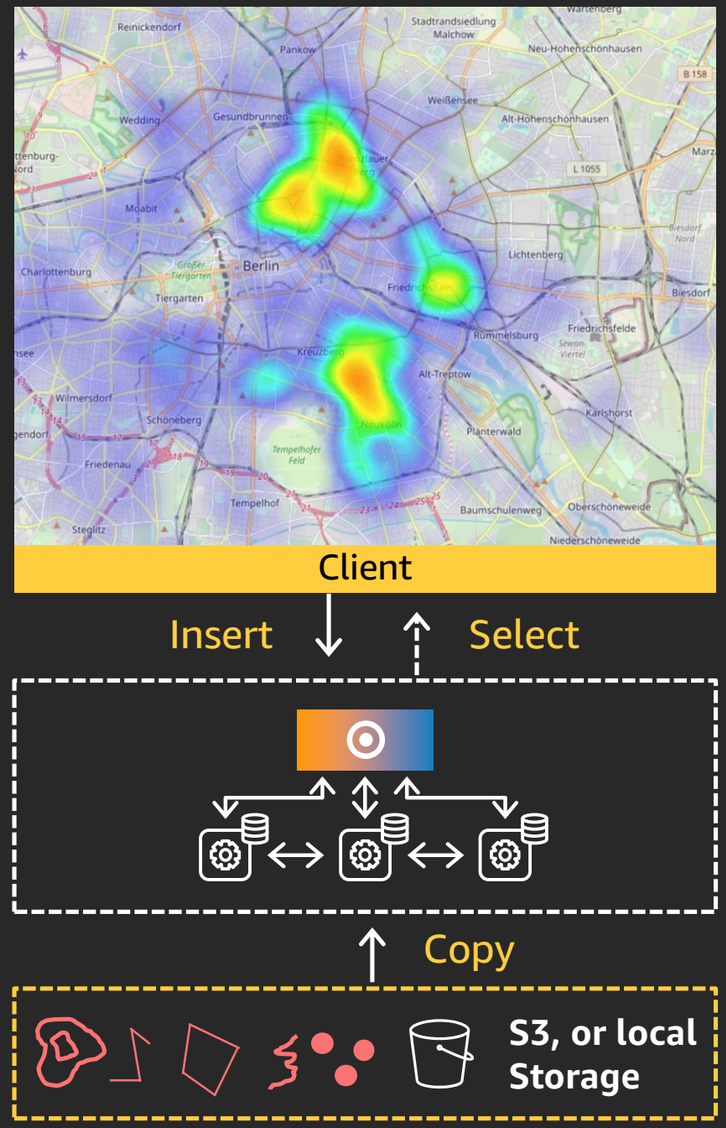
# New data type Geometry

## Ingest, store, and analyze spatial data

Seamlessly integrate spatial and business data

Geometry data types supports shapes such as MultiPoint, MultiLinestring, MultiPolygon, and GeometryCollection

40+ SQL spatial functions to construct geometric shapes, import, export, access and process the spatial data



# Amazon Redshift turbo charges query performance with machine learning based automatic optimizations

Automates table maintenance

Optimizes for peak performance as data and workloads scale

Leverages machine learning

Offers prescriptive recommendations with ability to apply changes dynamically



Automatic Analyze



Automatic Table Distribution Style



Distribution/Sort Key Advisors



Automatic Vacuum Delete



Automatic Table Sort

# Amazon Redshift Automatic Table Sort

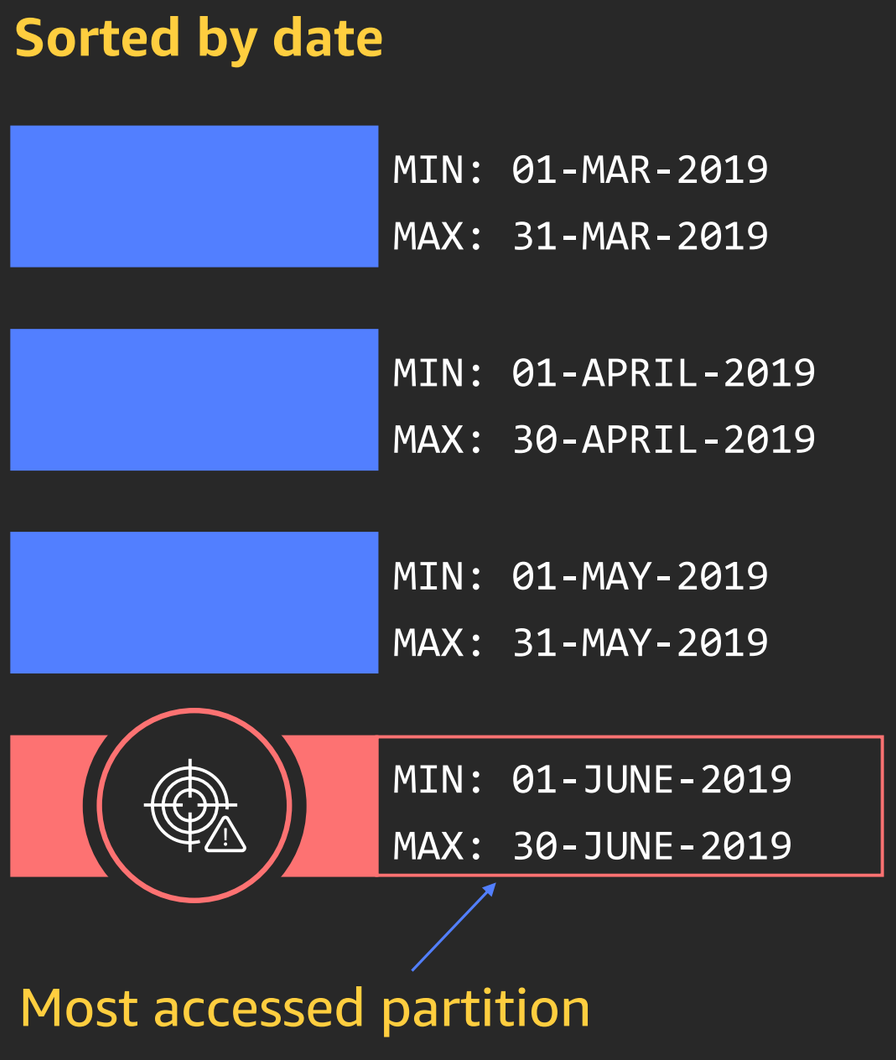
Eliminates DBA-managed sorting

Automates sort-order maintenance

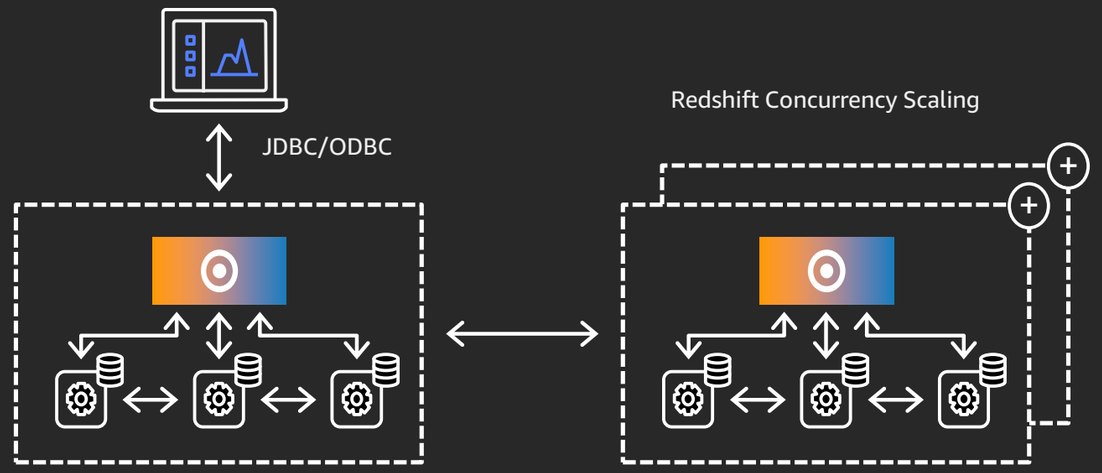
Maximizes efficient use of compute

Leverages machine learning

No impact to performance or accessibility



# Compute elasticity and scalability with Concurrency Scaling



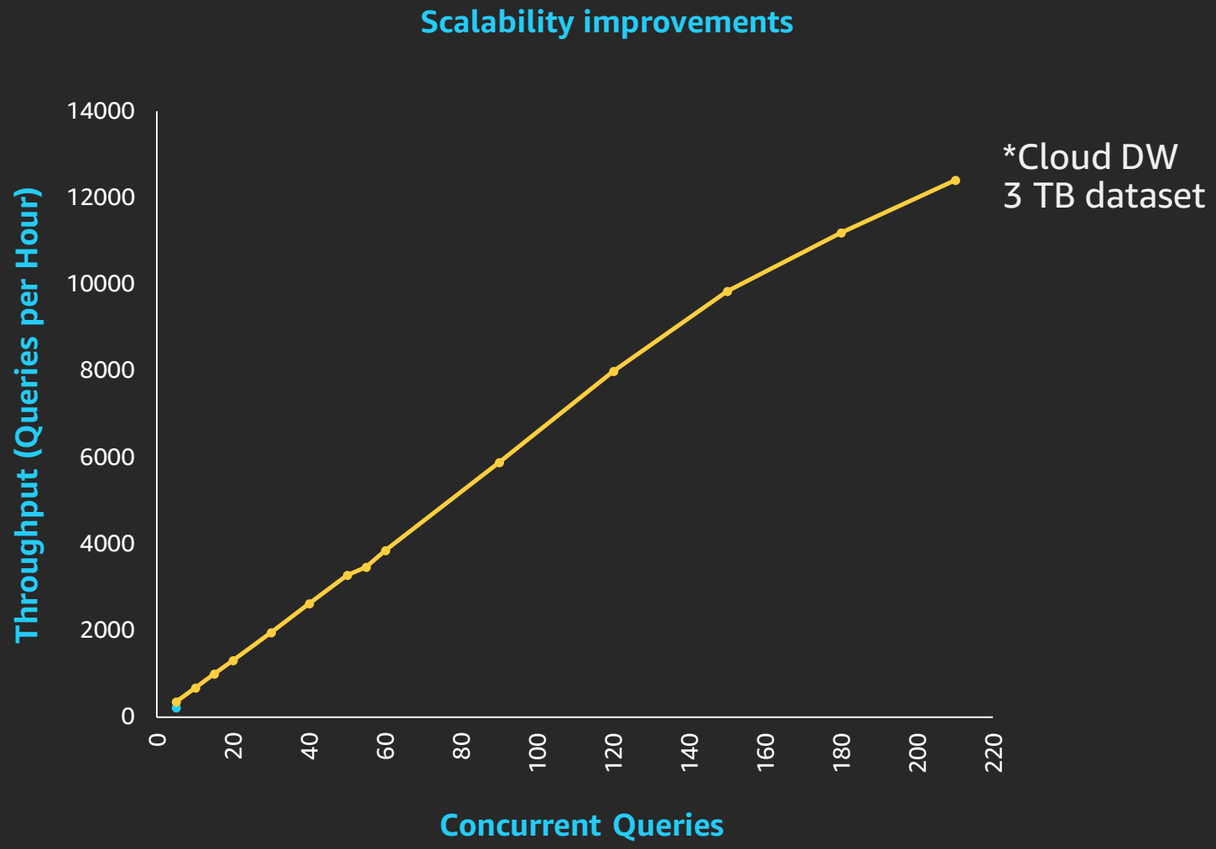

Scale-out to multiple Amazon Redshift clusters from a single endpoint in seconds

Support virtually unlimited concurrent users and queries while maintaining SLAs

Per-second billing for additional clusters used

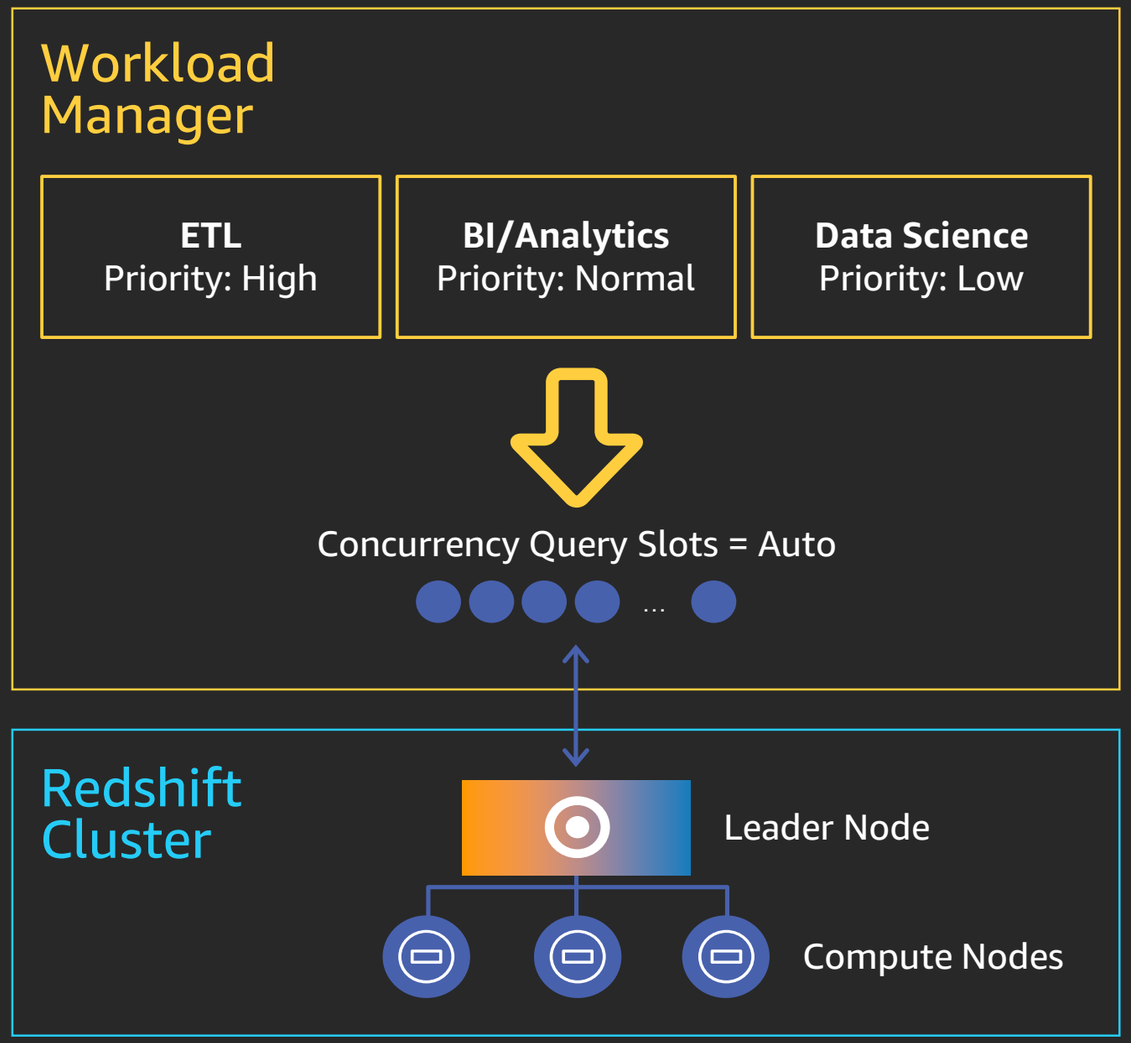
Free 1hr usage per day
(free for 97% of clusters)

## >35x improvement in throughput in 2019




* Cloud DW benchmark is based on TPC-DS (v2.10) with no query modifications are done

# Efficiency with automatic workload management and query priorities




Dynamically manage concurrency with no administration overhead

Efficient sharing of cluster between users & business groups

Advanced ML to determine # of queries to process in parallel to optimize throughput and performance

Feedback loop to continuously monitor system utilization and regulate admission

Influence workload performance based on business priorities

Intelligent algorithms to ensure low priority queries continue to make progress

# Asurion

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Key differentiators



**Best-in-class technology**

Highly integrated infrastructure



**Real-time insights and solutions**

Deliver meaningful, real-time solutions to customers



**Scale advantages**

Global footprint enabling scale efficiencies

# Design principles

1

## Advanced analytics

Ability to support both near real-time and batch analytics

Support for predictive (AI/ML) analytics

2

## Scalability

Modular approach to service deployments

Plug and play type ingestion and consumption models

Ability to scale independent layers based on variety, velocity, and volume of data

3

## Security & privacy

Least privilege access model, Role-based Access Control (RBAC)

Data security through encryption at-rest and in-transit

Support GDPR, CCPA, and other privacy requirements

4

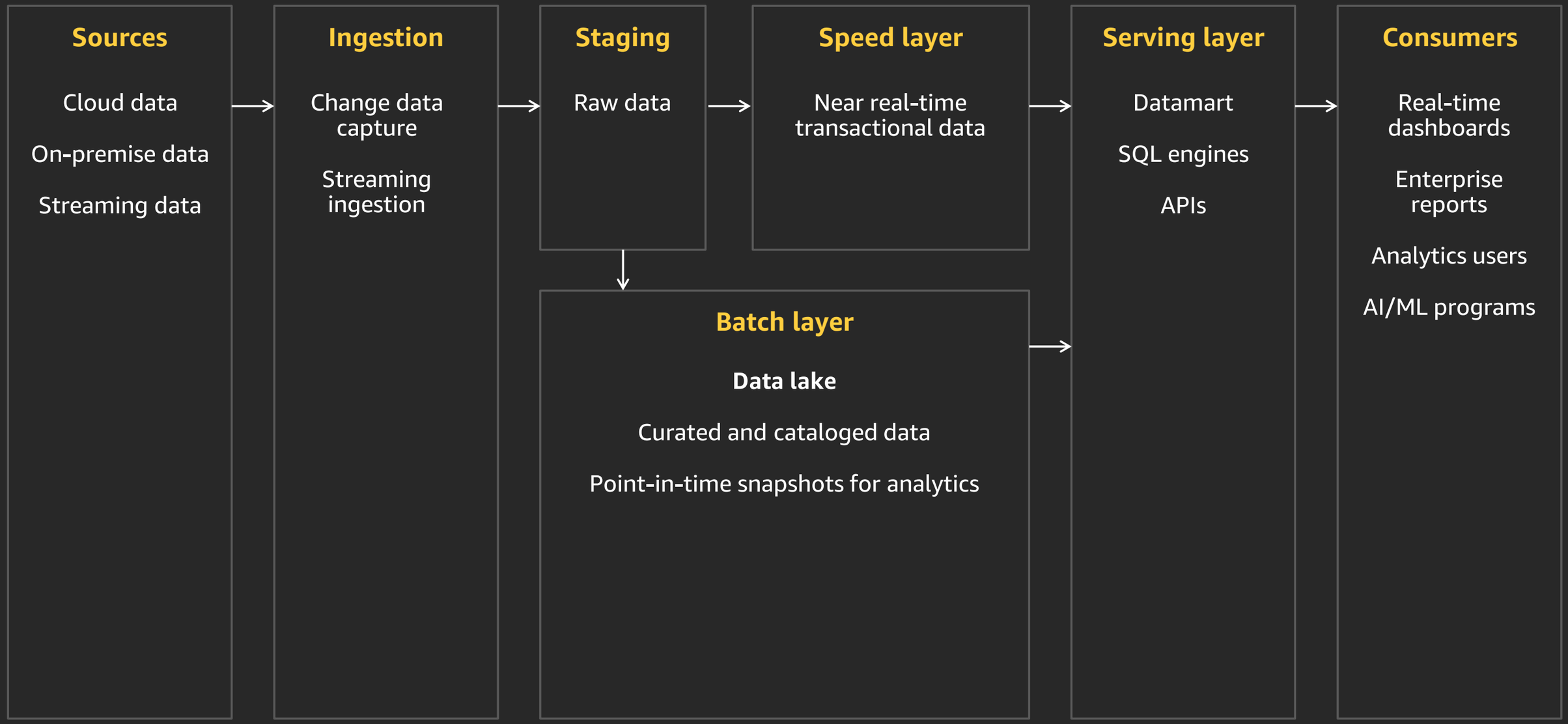
## Operational excellence

Logging and alerting

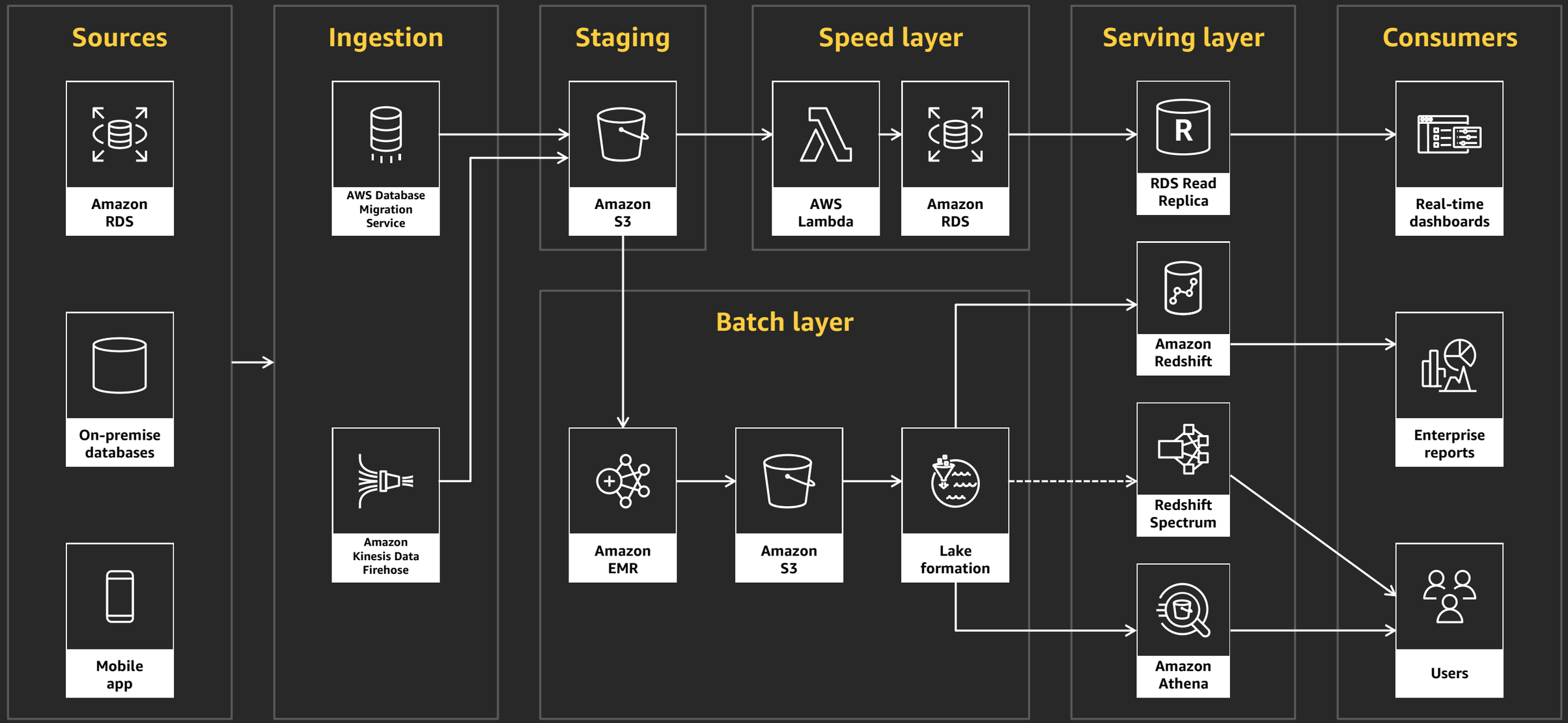
Information lifecycle management and secure archival

End-to-end automation, CI/CD pipeline

# Logical architecture



# Architecture—AWS services



# Amazon Redshift Spectrum

## Extended Data Warehouse

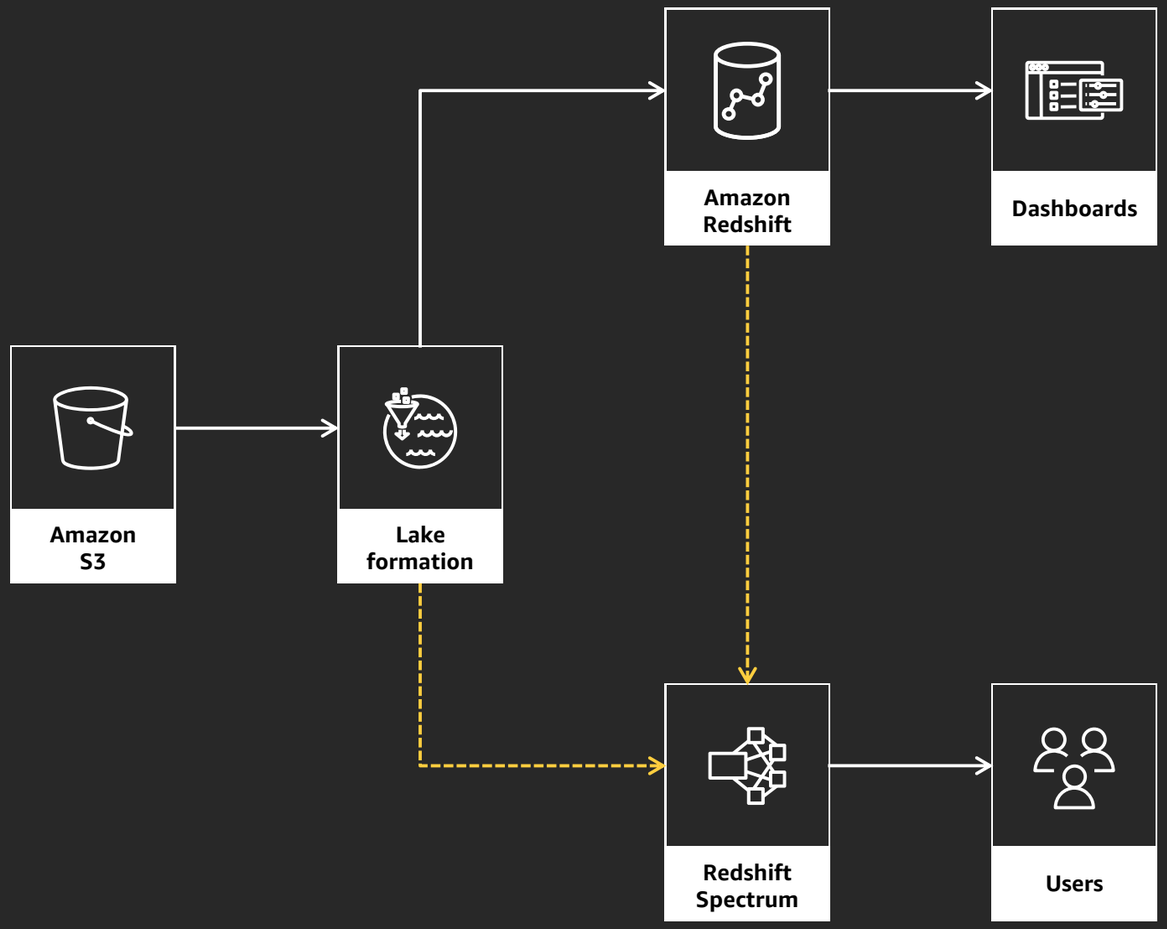
Seamlessly join Redshift with S3

Faster turn around

Performant and cost effective

## Security

Access control through Lake Formation





# Amazon Redshift best practices

## Query performance

Concurrency scaling

Automatic Workload Management (WLM)

## Security

LDAP integration



# Amazon Redshift storage scalability

Cost optimization

Performance gains

# What we have covered

1

## Hear

about our new Redshift announcements and recent innovations

2

## Understand

our data warehouse strategy and how it can benefit your organization

3

## Plan

how you can adopt the exciting new innovations in Redshift

# Related sessions

- ANT320-R What’s new with Amazon Redshift, featuring Workday
- ANT418 Deep dive and best practices for Amazon Redshift
- ANT237-R1 Zero to Hero: Quick Starts with Amazon Redshift
- ANT335-R1 How to build your data analytics stack at scale with Amazon Redshift

# Booth

Come visit us in the Analytics Booth on the expo floor to learn more!



# Please complete the session survey in the mobile app.



© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

# Thank you!

AWS re:Invent

© 2019, Amazon Web Services, Inc. or its affiliates. All rights reserved.

aws