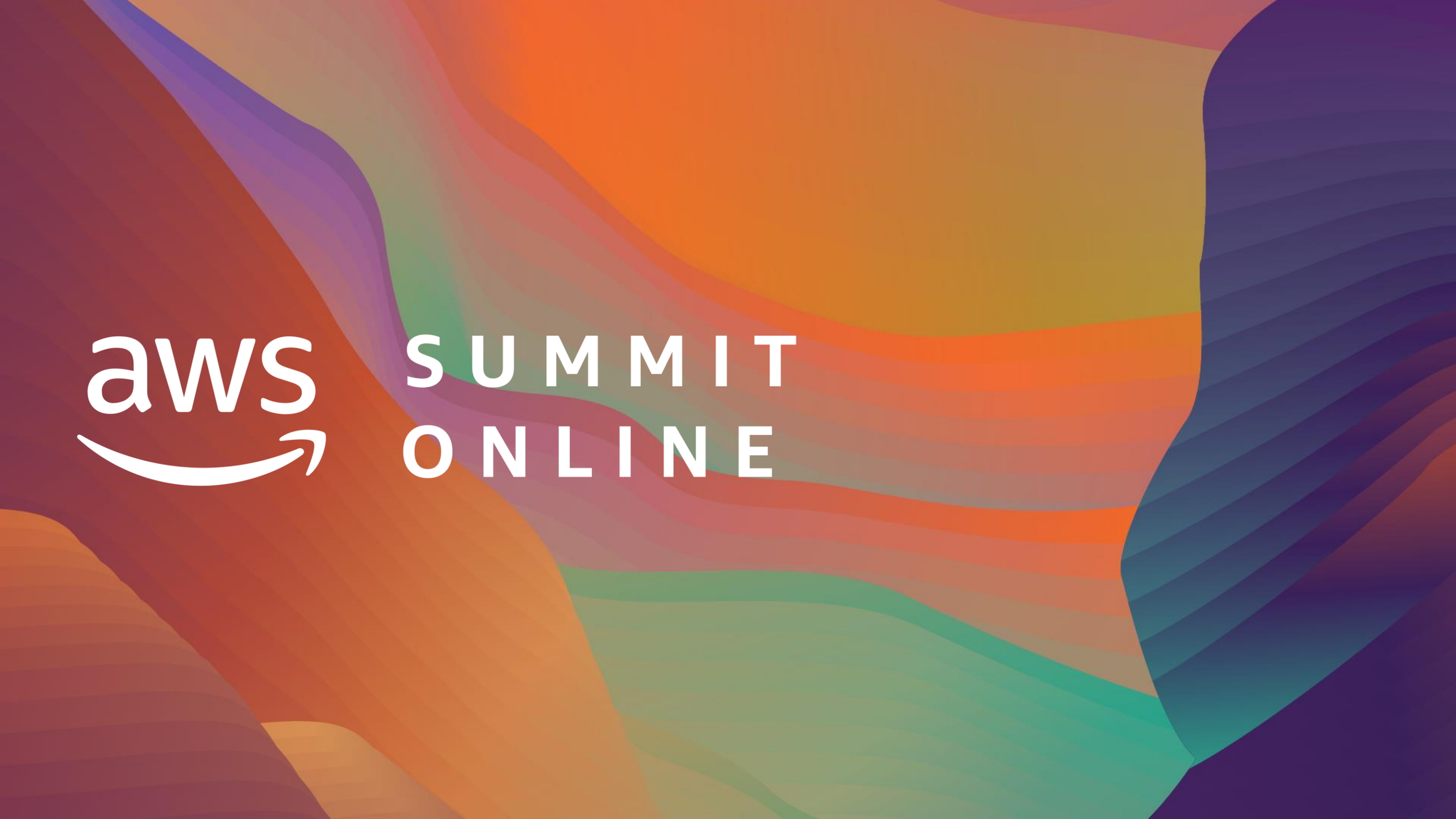

aws
SUMMIT
ONLINE

ANA02

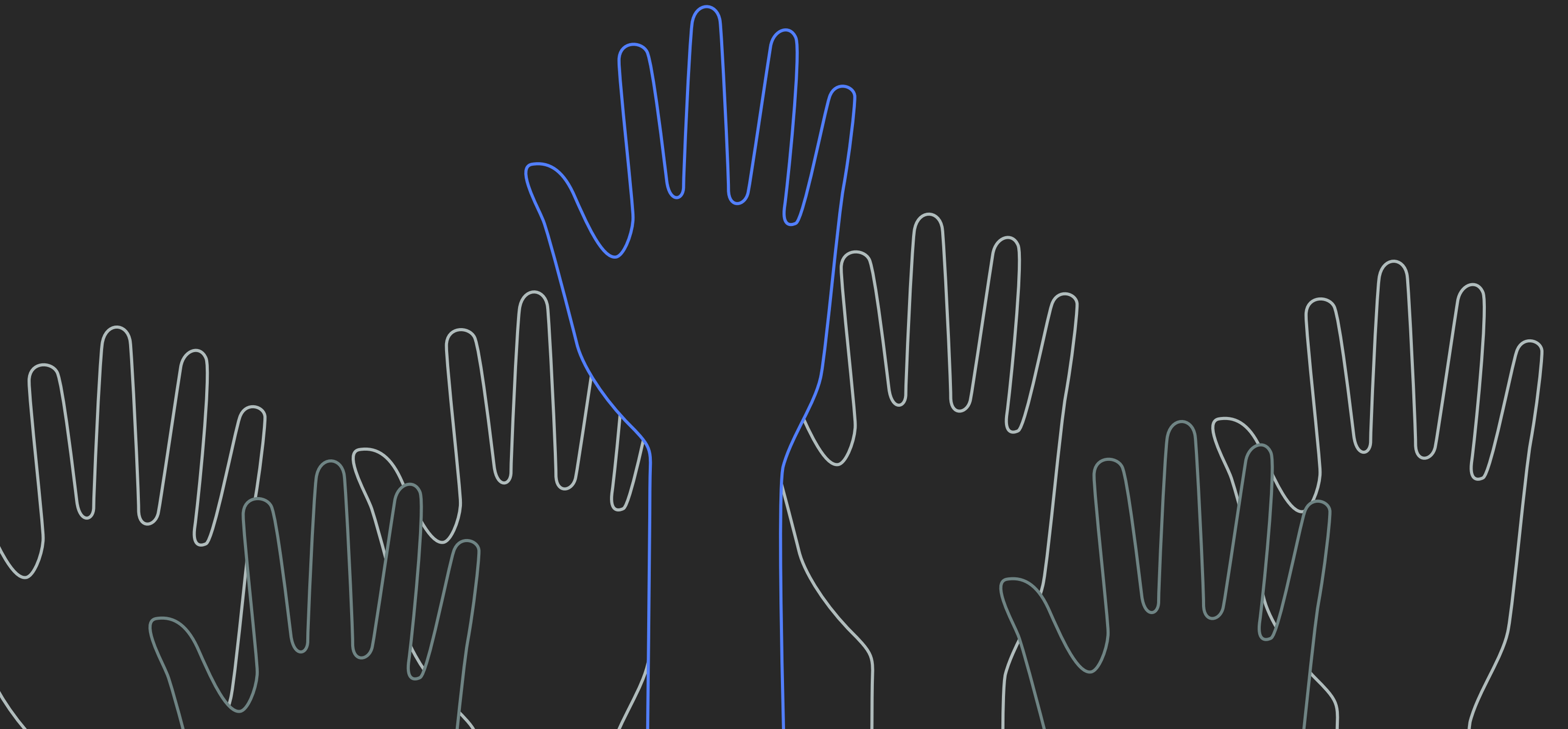
State of the art cloud data warehousing

Anna Coniglio
Data Warehouse Solutions Architect
Amazon Web Services

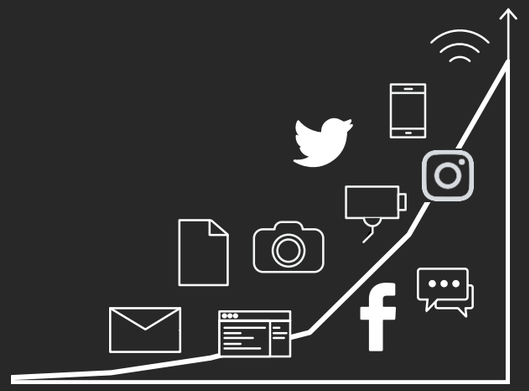
aws SUMMIT ONLINE

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Are you an Amazon Redshift user?



The new realities customers are facing



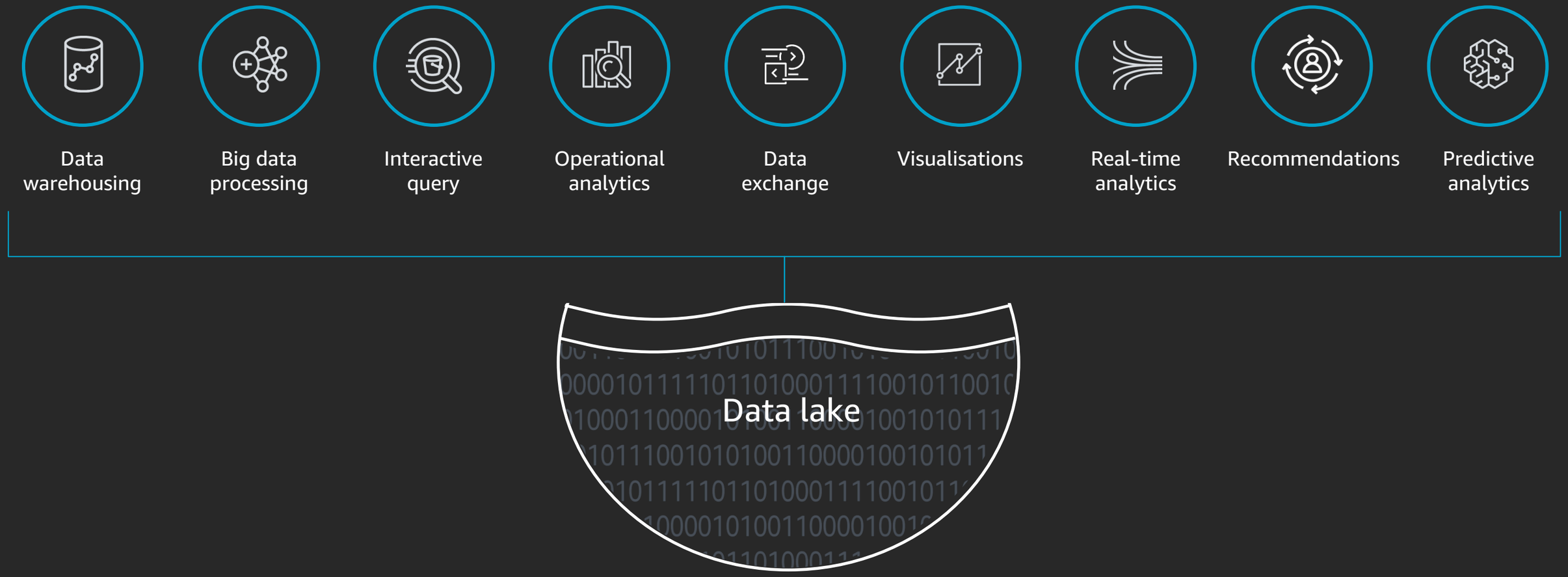
Explosion of data

Expansion of personas

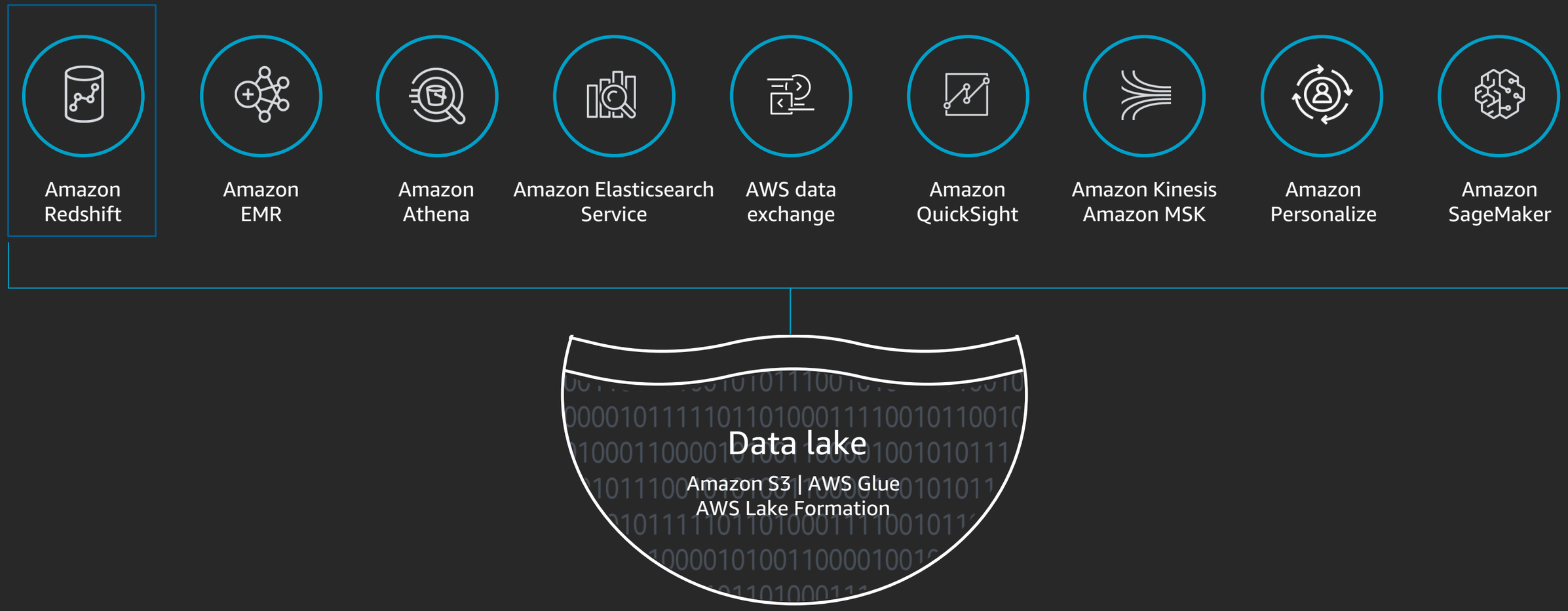
Demand for faster decision making on real-time data

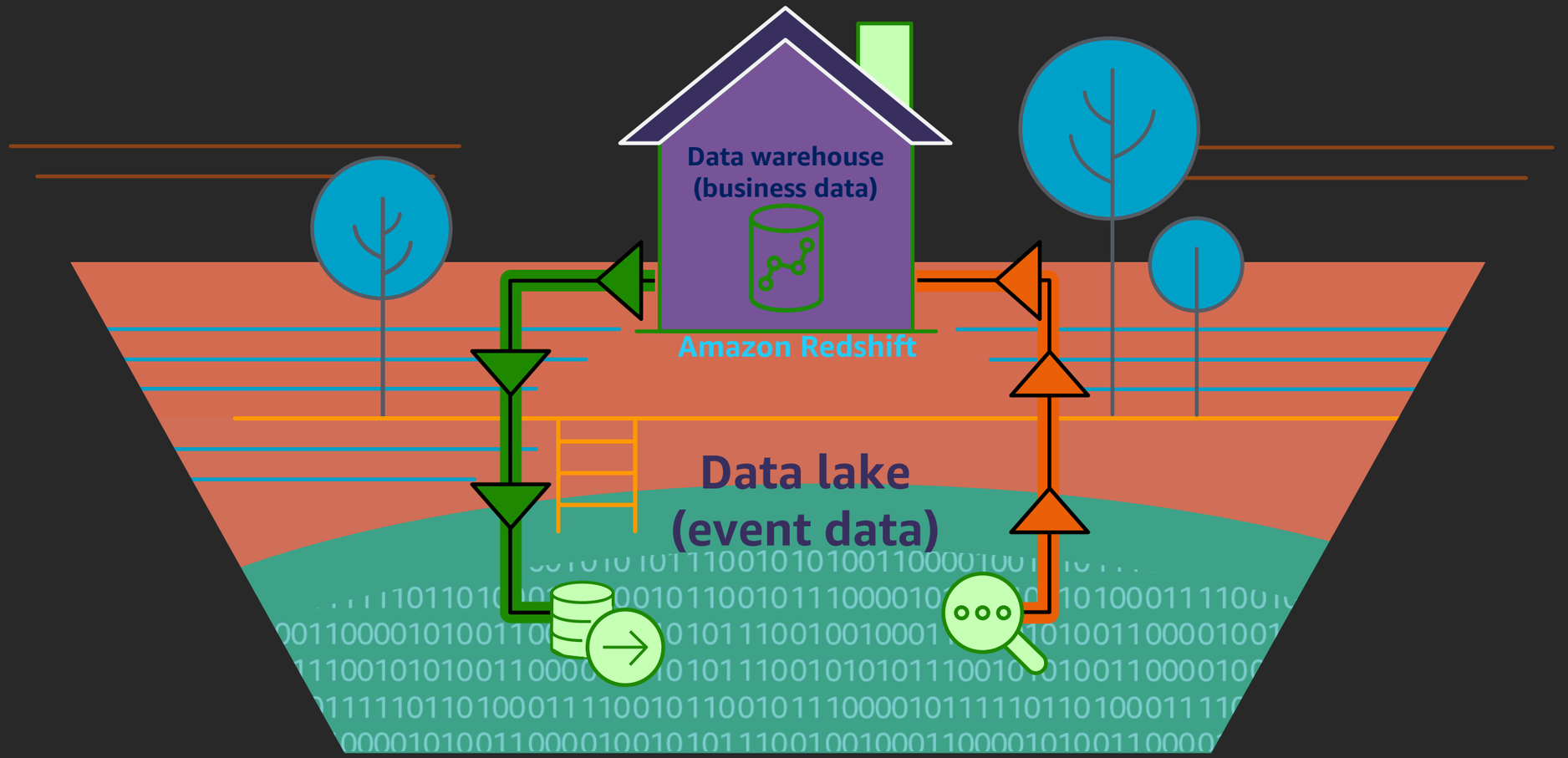
Any type of analytics on the data lake

Most comprehensive analytics platform



Any type of analytics on the data lake





Customers moving to data lake architectures

Amazon Redshift enables you to have a lake house approach

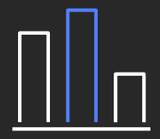
Why Amazon Redshift

Tens of thousands of customers use Amazon Redshift and process over 2 exabytes of data per day

Data lake and AWS integrated

Lake formation catalog and security

Exabyte querying and AWS integrated



Best performance

3x faster than other
cloud data warehouses

Best value

Up to 75% less than other
cloud data warehouses

Predictable costs

Most scalable

Virtually unlimited
elastic linear scaling

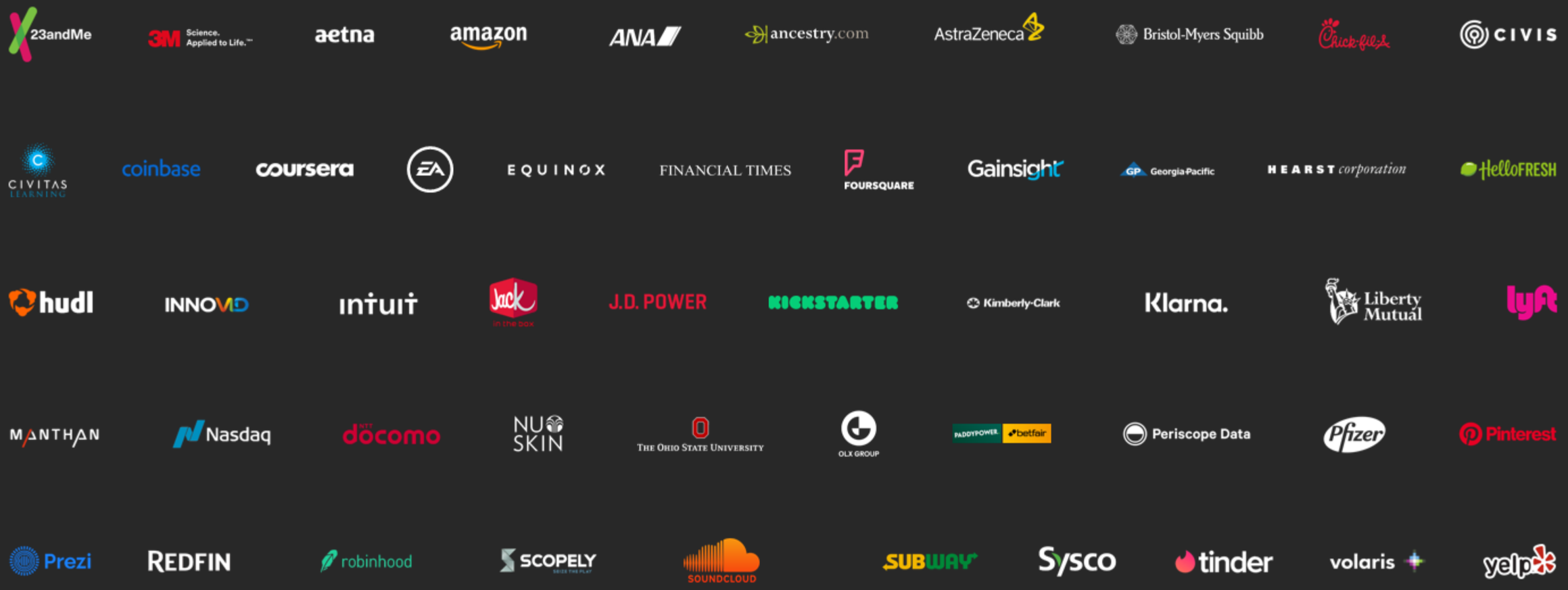
Most secure & compliant

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

Easy to manage

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

Tens of thousands of customers use Amazon Redshift



Features delivered to meet customer needs

Robust result set caching

Large # of tables support ~20,000

Copy command support for ORC, Parquet

IAM role chaining

Elastic resize

Groups

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

Auto analyse

Health and performance monitoring w/Amazon CloudWatch

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 (Advanced Query Accelerator)

Concurrency scaling

200+

DC1 migration to DC2

Resiliency of ROLLBACK processing

Manage multi-part query in AWS console

Auto analyse for incremental changes on table

new features in the past 18 months

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 optimised data transfer protocol

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

Amazon Redshift Spectrum: Concurrency scaling

Integration with AWS Lake Formation

Auto-vacuum sort, auto-analyse, 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

Materialised views

Manual pause and resume

Amazon Redshift: Key 2019 innovations

Performance and scalability

- New — RA3 with Amazon Redshift managed storage
- Preview — AQUA for Amazon Redshift
- New — AZ64
- New — Materialised views
- New — Spatial data support
- Preview — Federated query
- New — Concurrency scaling

Ease of use

- New — New management console
- New — Auto WLM: Query priorities
- New — Elastic resize scheduler
- New — Auto-vacuum, Auto-analyse & 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 — Integration with Amazon Lake Formation

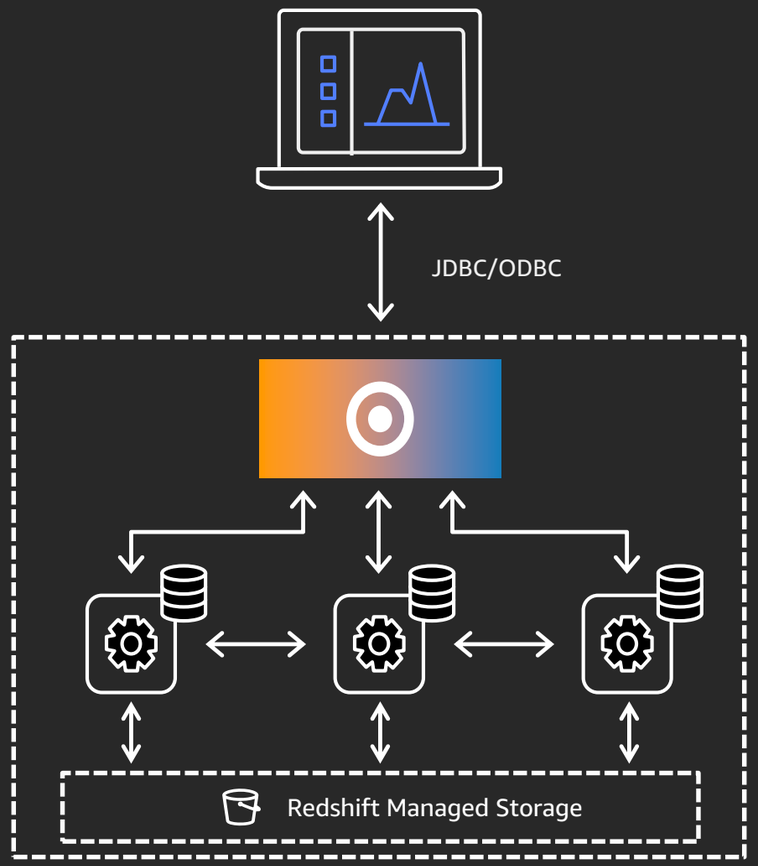
RA3: New 3rd gen Amazon Redshift compute node

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



RA3: Two node types for compute granularity

RA3.16xl

48 vCPU, 384 GiB Memory, 8 GB IO

On demand price—$13.04/node/hour

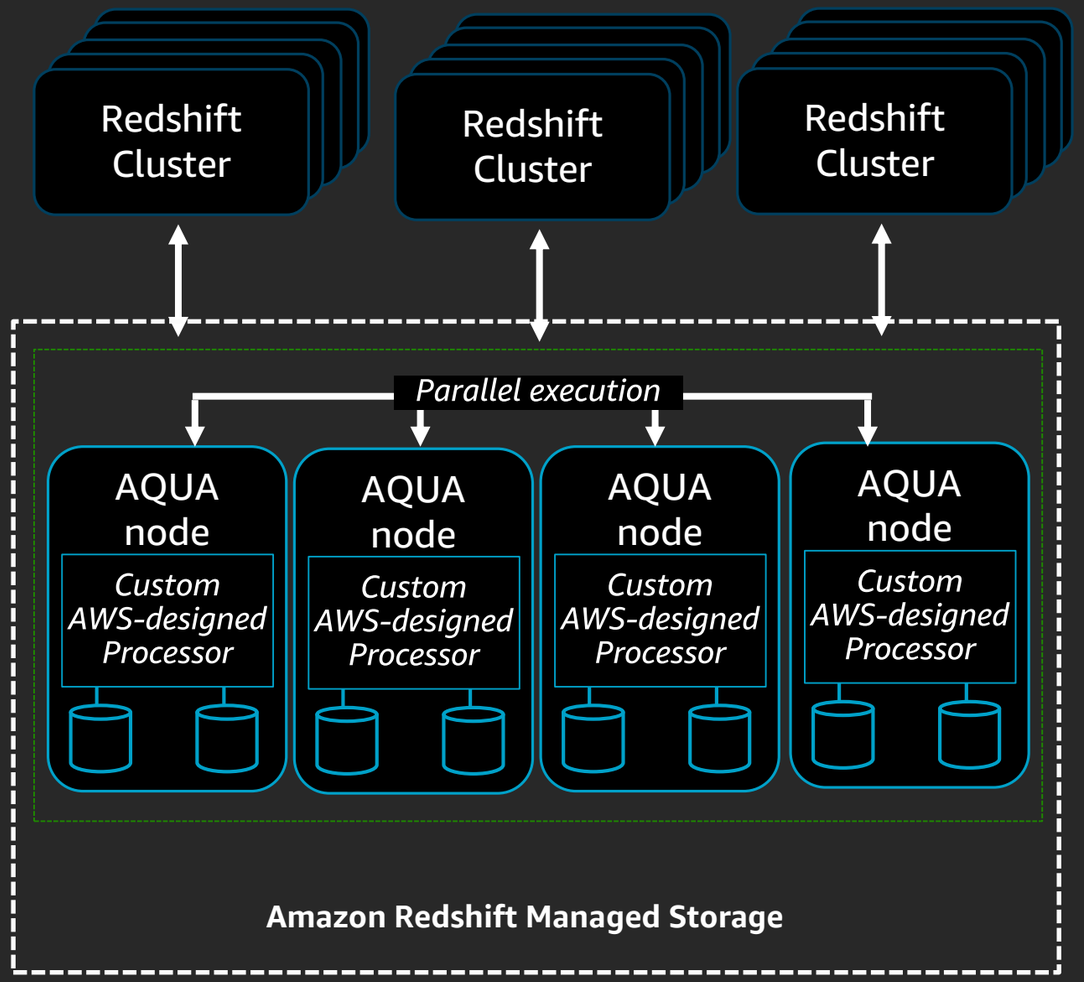
RA3.4xl

12 vCPU, 96 GiB Memory, 2 GB IO

On demand price—$3.26/node/hour

Scale to tens of PB of data, pay as you go at $0.024/GB-month for storage

AQUA (Advanced Query Accelerator) for Amazon Redshift



New distributed & hardware-accelerated processing layer

With AQUA, Amazon Redshift is up to **10x faster** than any other cloud data warehouse, no extra cost

---

Available in Preview with RA3

No code changes required

Amazon Redshift federated query (preview)

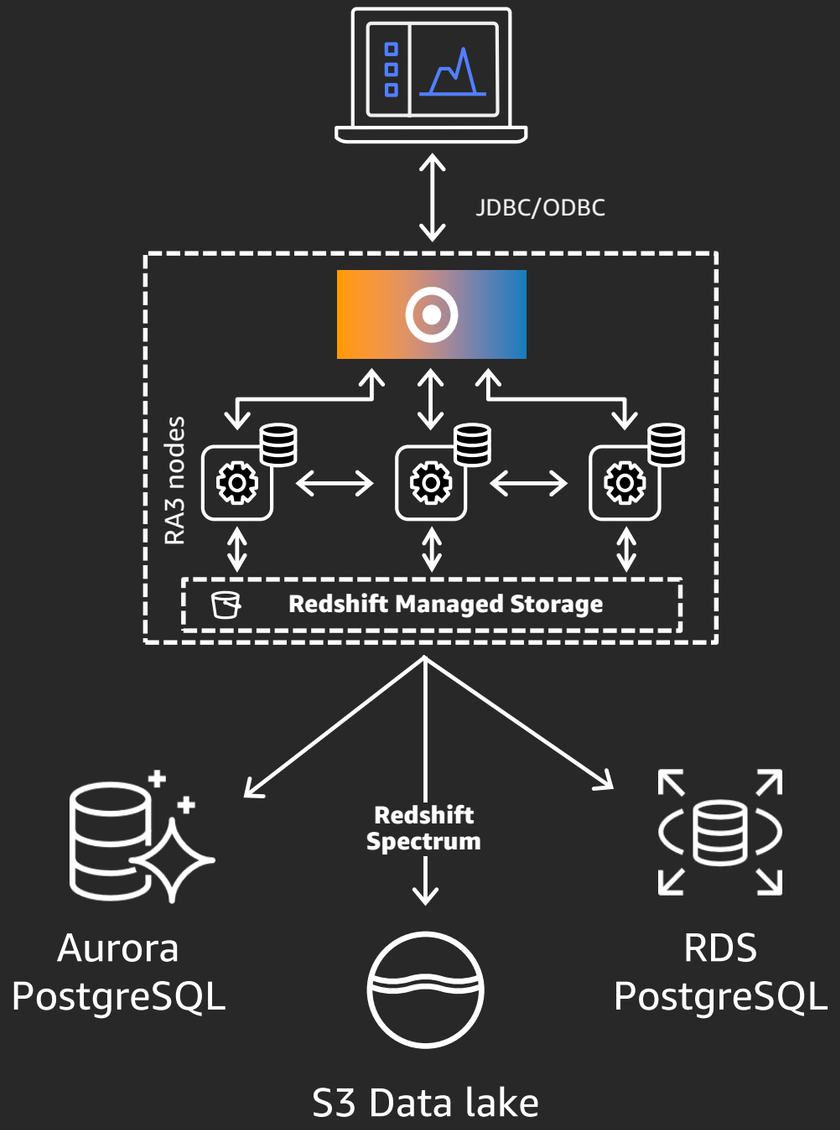
Queries on RDS and Aurora PostgreSQL databases

Analytics on live data without data movement

Unified analytics across data warehouse, data lake and operational databases

Flexible and easy way to ingest data

Performant and secure access to data



Thank you!

Anna Coniglio

annacn@amazon.com

aws SUMMIT ONLINE

© 2020, Amazon Web Services, Inc. or its affiliates. All rights reserved.