AWS re:Invent

ANT320-R

What's new with Amazon Redshift, featuring Workday

Michalis Petropoulos
Director, Engineering
Amazon Redshift
Amazon Web Services

Erol Guney
Architect, Data Platform
Workday Inc.

AWS re:Invent

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

aws

7 years ago…



AWS re:Invent 2012

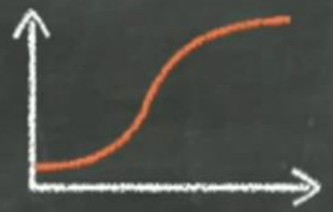
Andy Jassy
Senior Vice President, Amazon Web Services

What Would Data Warehousing Look Like If We Did It the AWS Way

Easy to provision and scale up massively



No upfront costs, pay as you go



Really fast performance at a really low price



Open & flexible with support for popular tools





AWS re:Invent 2012

Andy Jassy
Senior Vice President, Amazon Web Services

Introducing Amazon Redshift

Data Warehousing the AWS Way



- Easily and rapidly analyze petabytes of data
- 1/10 the cost of traditional data warehouses
- Automated deployment & administration
- Compatible with popular BI tools

7 years later: Tens of thousands of customers use Amazon Redshift



Amazon Redshift

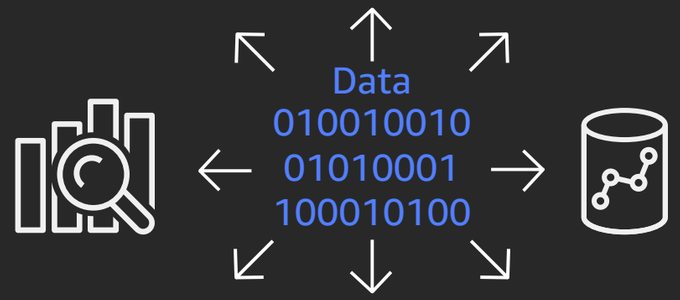
is the most popular cloud data warehouse

23andMe, amazon, AMGEN, ancestry, AstraZeneca, CIVIS, coinbase, coursera, DOW JONES, edmunds, EA, EQUINOX, FT FINANCIAL TIMES, FOURSQUARE, HEARST, iflix, INTUIT, Jack in the box, Kimberly-Clark, Klarna, Liberty Mutual, MANTHAN, MERCK, Nasdaq, The Ohio State University, PADDYPOWER, betfair, Pfizer, Prezi, REDFIN, robinhood, SCHOLASTIC, Schumachergroup, SCOPELY SEIZE THE PLAY, SOUNDCLOUD, Sysco, tinder, yelp

Our customers are asking for…



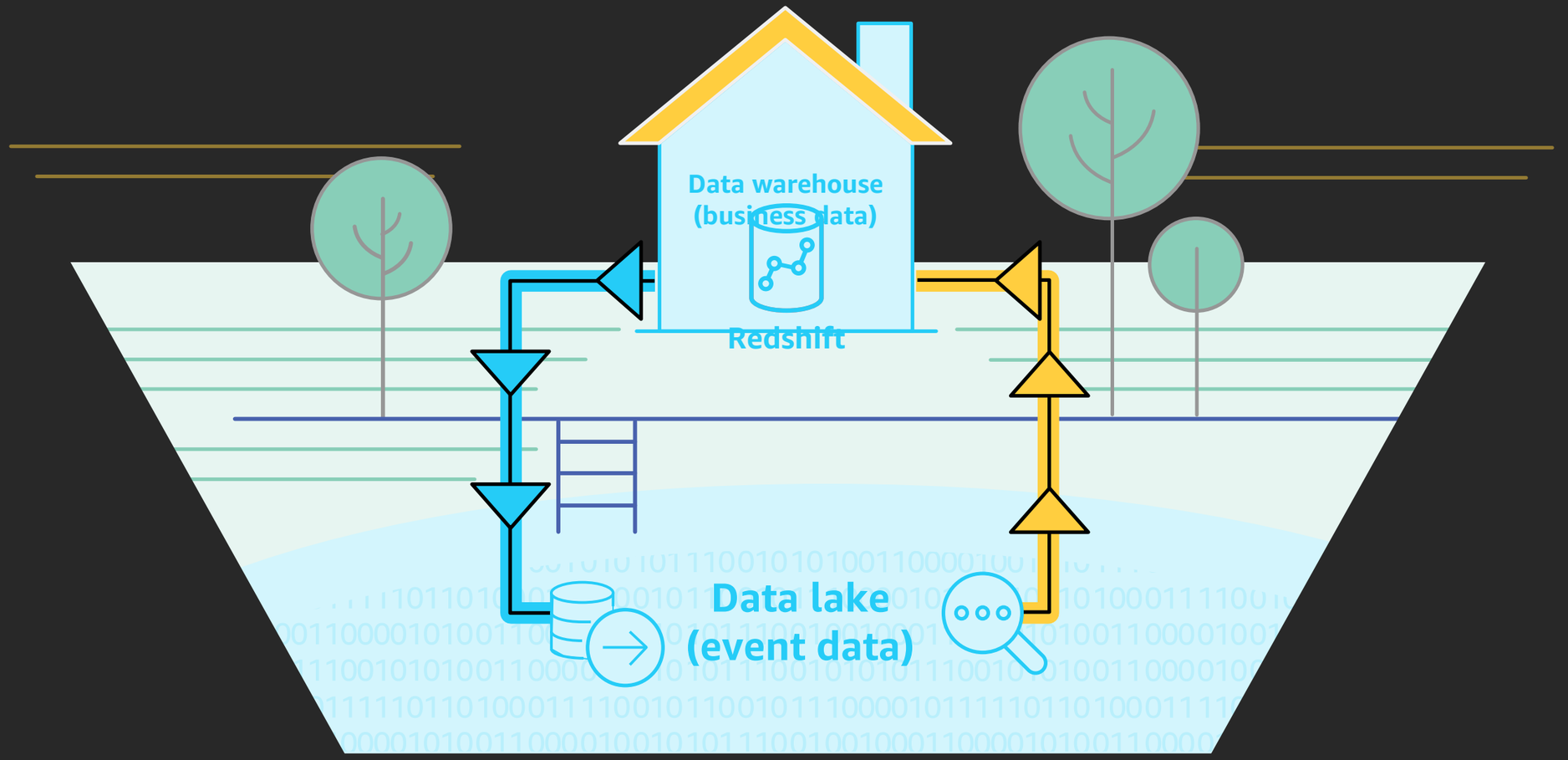
Migrations to the cloud



Exponential growth of event data



End-to-end insights from analyzing all your data



Customers moving to data lake architectures

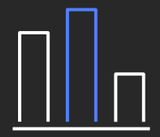
Amazon Redshift enables you to take a lake house approach

Amazon Redshift: Data warehouse of choice for the following reasons



Data lake & AWS integrated

AWS Lake Formation catalog & security, Exabyte scale query (spectrum & federated), AWS integrated (AWS DMS, Amazon CloudWatch)



Best performance

Up to 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 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

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

Auto analyze

Health and performance monitoring w/Amazon CloudWatch

Automatic table distribution style

Amazon 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 & enhancement in the past 18 months

Manage multi-part query in AWS console

Auto analyze for incremental changes on table

DC1 migration to DC2

Resiliency of ROLLBACK processing

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

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

NEW!

Concurrency Scaling

Data lake & AWS integration

PREVIEW!

Federated query across Amazon Redshift & RDS/Aurora

NEW!

Data lake export in Parquet

NEW!

Spectrum Request Accelerator

NEW!

AWS Lake Formation integration

Fully managed

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!

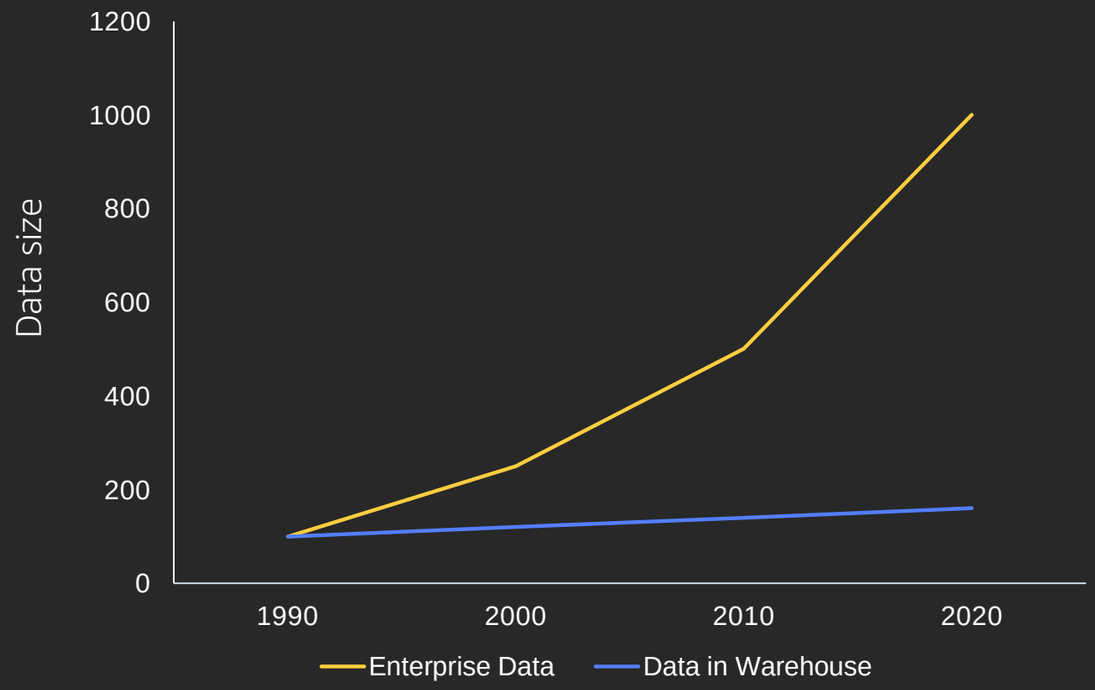
Deferred Maintenance

NEW!

Elastic resize

Exponential data growth causing management, performance, and scaling challenges

What about data growth?



Compute and storage needs can vary independently

Compute scaling:

- To handle increased query load during peak times
- To ingest data fast and make it available for querying in near real time

Storage scaling:

- To store data for querying or auditing
- To pay separately for data storage at cheaper rates, independent of compute usage

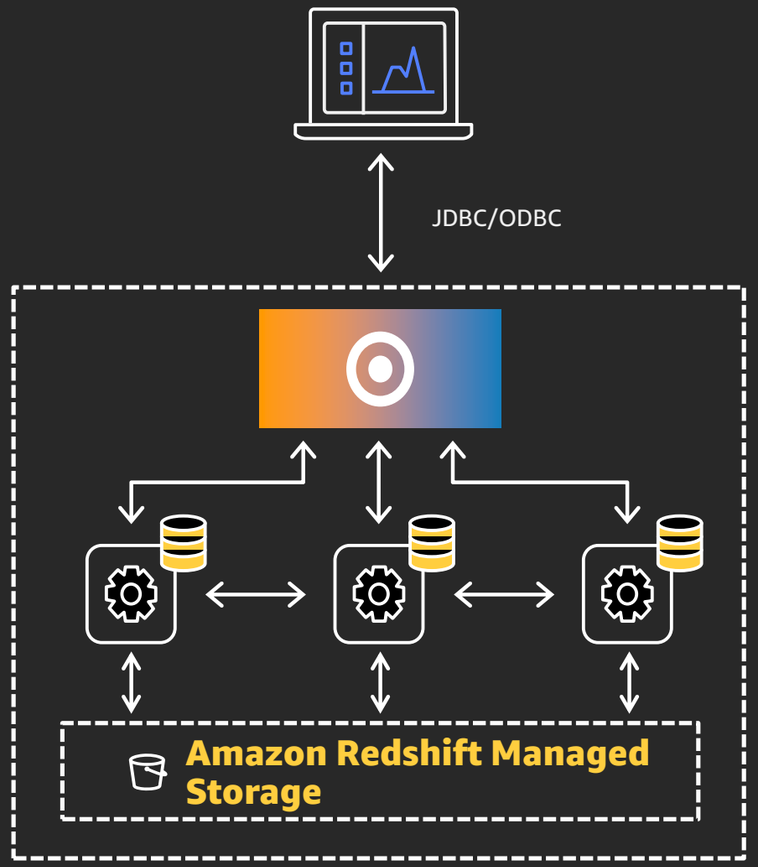
New 3rd-generation Amazon Redshift compute instance: RA3

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

Scale data warehouse only based on steady-state compute needs

Pay separately for storage and compute

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



RA3: Unmatched performance at an unbeatable price

RA3.16xl

On-demand price—$13.04/node/hr

For storage, pay as you use at $0.024/GB-Month

Can scale to tens of PB of data (8PB compressed)

Coming soon
RA3.4xl

2x Performance and 2x storage capacity compared to DS2.8XL for the same price

Up to 3x price performance compared to any other Cloud DW

The minimum size RA3.16XL cluster scales up to 128TB (compressed), size only for your compute need

Customer experience with RA3

Western Digital

"Our data is nearly ***doubling*** every year and we run ***6 Redshift clusters with total 78 nodes and 631+ TB compressed data*** stored to get insights that our business analysts and leadership depend on. The new Redshift ***RA3 instances*** offer us the ability to process our growing data more cost-effectively while we ***doubling our storage capacity compared to our previous Redshift cluster.***"

Fayaz Syed, Sr. Manager, Big Data Platform



"We load ***billions of events per day*** into Amazon Redshift and have ***hundreds of terabytes of data that is expected to double every year***. While we store and process all our data, most of the analysis only ***uses a subset of the data***. The new Redshift RA3 instance with managed storage ***delivered 2x performance*** for most of our queries compared to our previous DS2 instance based Redshift cluster."

Jonathan Burket, Senior Software Engineer



"[We're] thrilled with the new RA3 instance type. We have observed a ***1.9x performance improvement over DS2*** and ***1.5x performance improvement over DC2*** in our workload while ***keeping the same costs and providing scalable managed storage***. This allows us to keep pace with explosive data growth and have the necessary fuel to train our machine learning systems."

Stephen Moy, Software Engineer

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

NEW!
Concurrency Scaling

Data lake & AWS integration

PREVIEW!
Federated query across Amazon Redshift & RDS/Aurora

NEW!
Data lake export in Parquet

NEW!
Spectrum Request Accelerator

NEW!
AWS Lake Formation integration

Fully managed

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

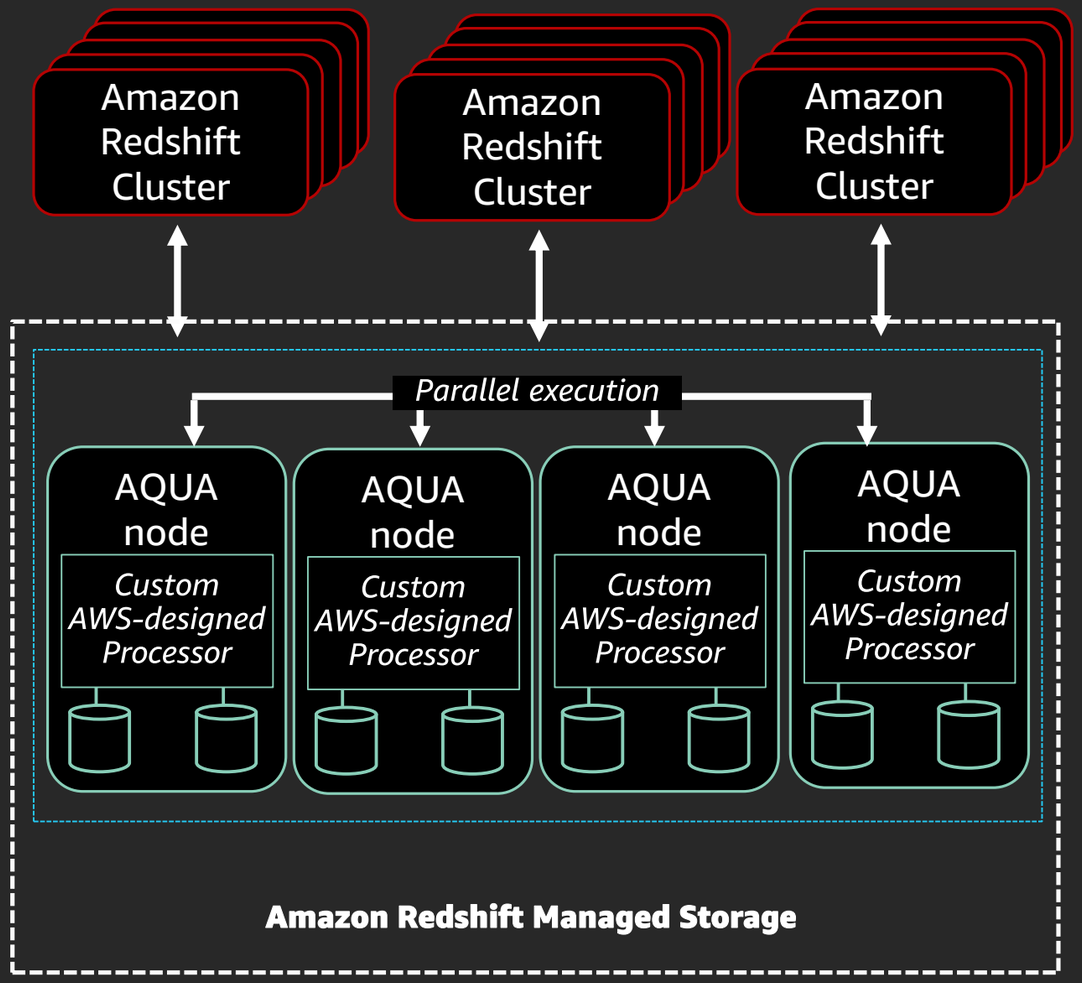
NEW!
Faster cross regional copy and change

NEW!
Auto Data Distribution

NEW!
Deferred Maintenance

NEW!
Elastic resize

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

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

Available in Preview with RA3. No code changes required.

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

NEW!
Concurrency Scaling

Data lake & AWS integration

PREVIEW!
Federated query across Amazon Redshift & RDS/Aurora

NEW!
Data lake export in Parquet

NEW!
Spectrum Request Accelerator

NEW!
AWS Lake Formation integration

Fully managed

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

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

- High compression ratio and improved query performance
- Efficient compression on small groups of data
- Leverages SIMD instructions for parallel processing
- Applicable for all numeric and date/time datatypes

	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

*Cloud DW benchmark, 30TB

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

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

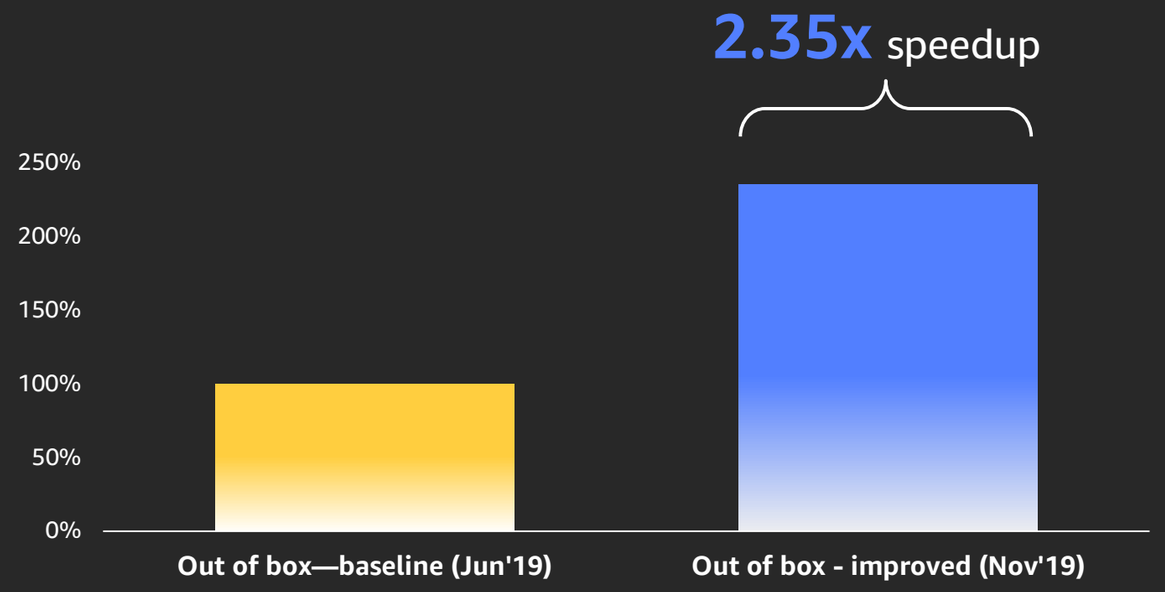
Performance innovations:

- 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

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

Introducing Materialized Views:
Compute once, query many times

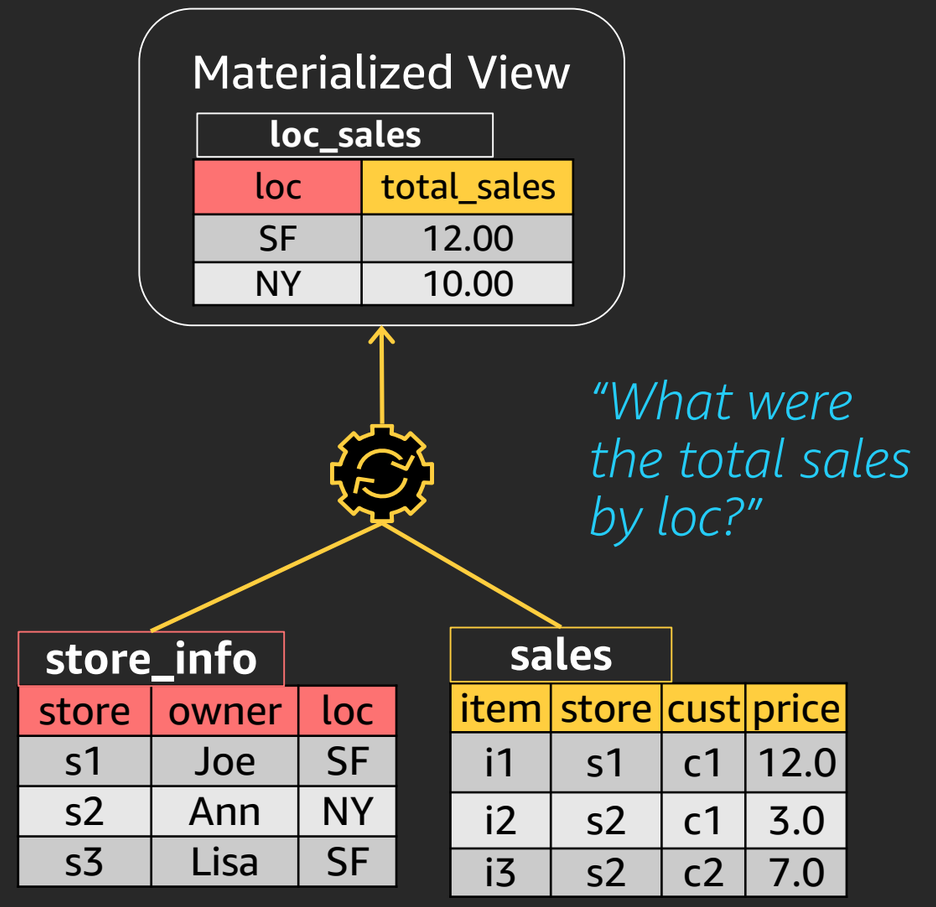
Speed up queries by orders of magnitude

- Joins, filters, aggregations, and projections

Simplify and accelerate ETL/BI pipelines

- Incremental refresh
- User triggered maintenance

Easier and faster migration to Amazon Redshift



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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

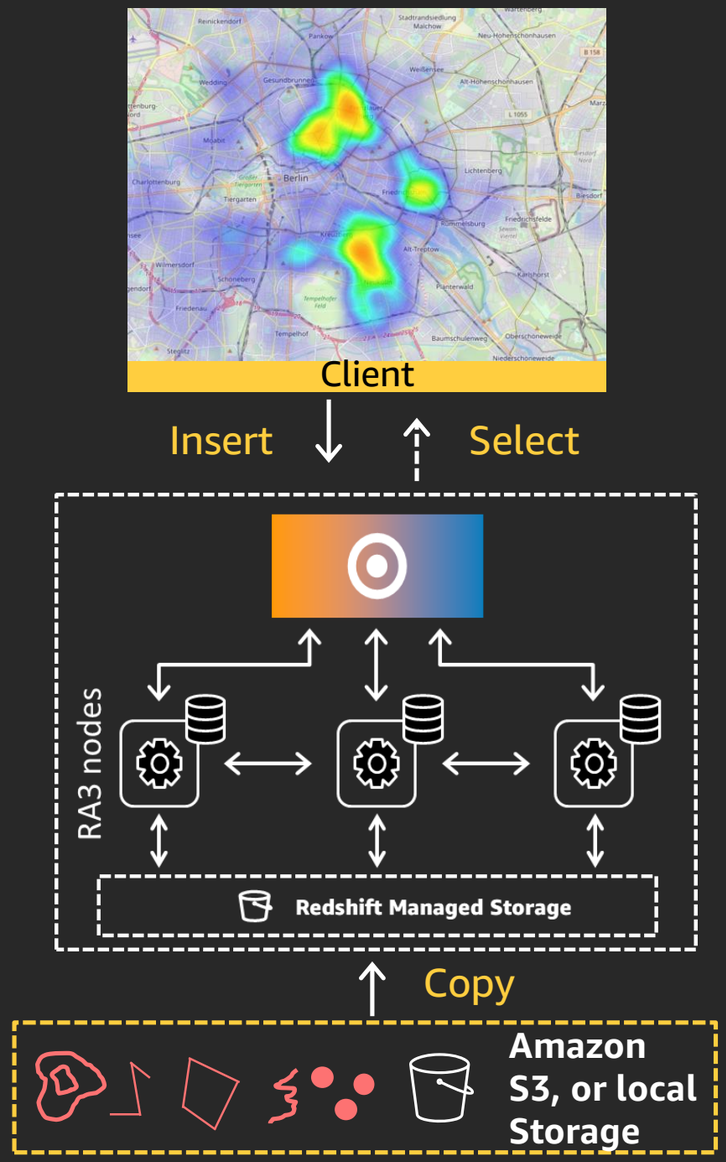
NEW! Elastic resize

New data type geometry: Ingest, store, and analyze spatial data

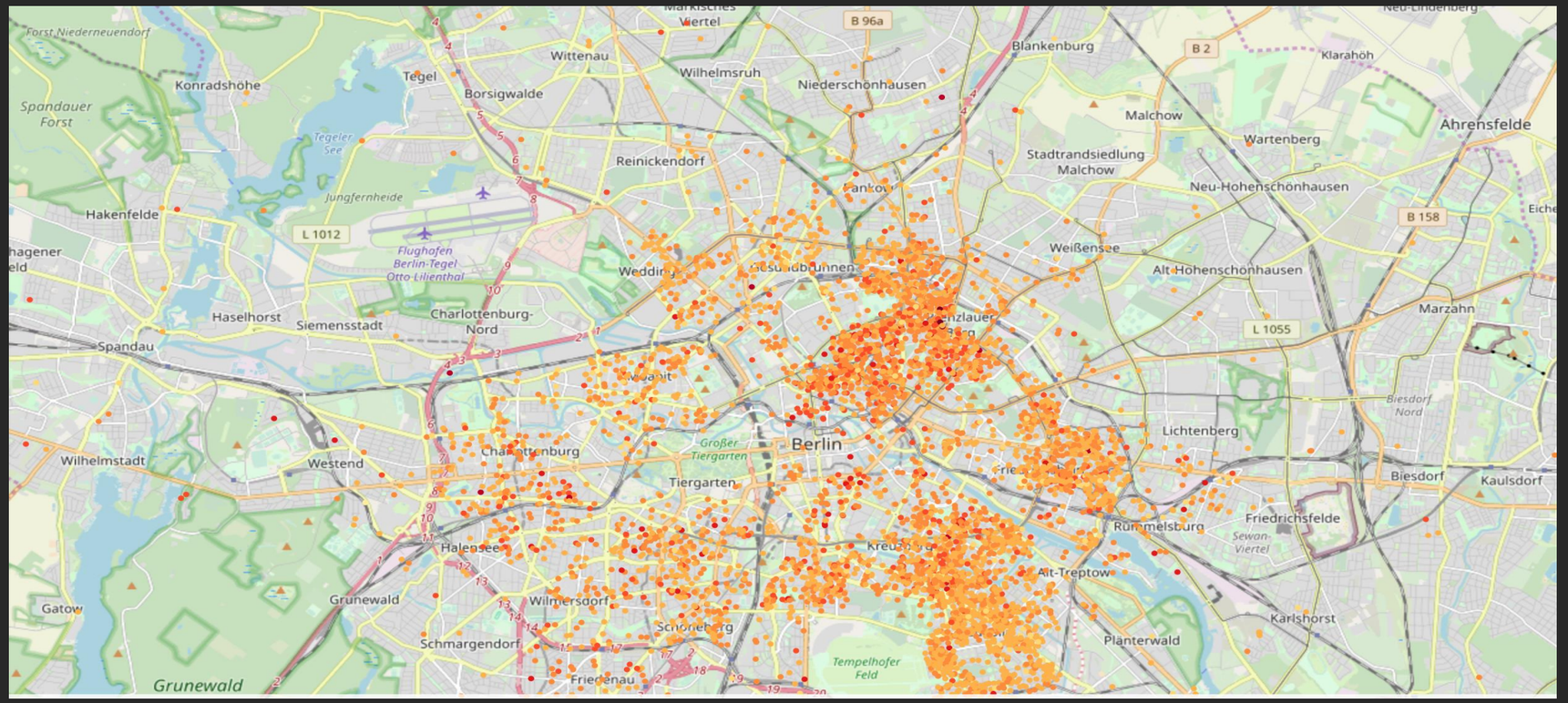
Seamlessly integrate spatial and business data

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



Spatial processing: Sample query




```
SELECT name, ST_X(shape) as lng, ST_Y(shape) as lat, price
FROM accommodations
WHERE ST_Within(shape, ST_GeomFromText( 'POLYGON((13.111839294433596
52.4285942596063, 13.111839294433596 52.60117089057946, 52.4285942596063))', 4326))
LIMIT 5000
```

Data Types

GEOMETRY
Point, Linestring, Polygon,
MultiPoint, MultiLinestring,
MultiPolygon,
GeometryCollection

Spatial Accessors

ST_NumGeometries,
ST_GeometryType,
ST_Dimension, …

Spatial Predicates

ST_Covers, ST_Equals,
ST_Within, ST_DWithin,…

Spatial Functions

ST_Distance,
ST_Azimuth, …

Spatial Formats

WKT/WKB, EWKT/EWKB, GeoJSON
Ingestion: CSV

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

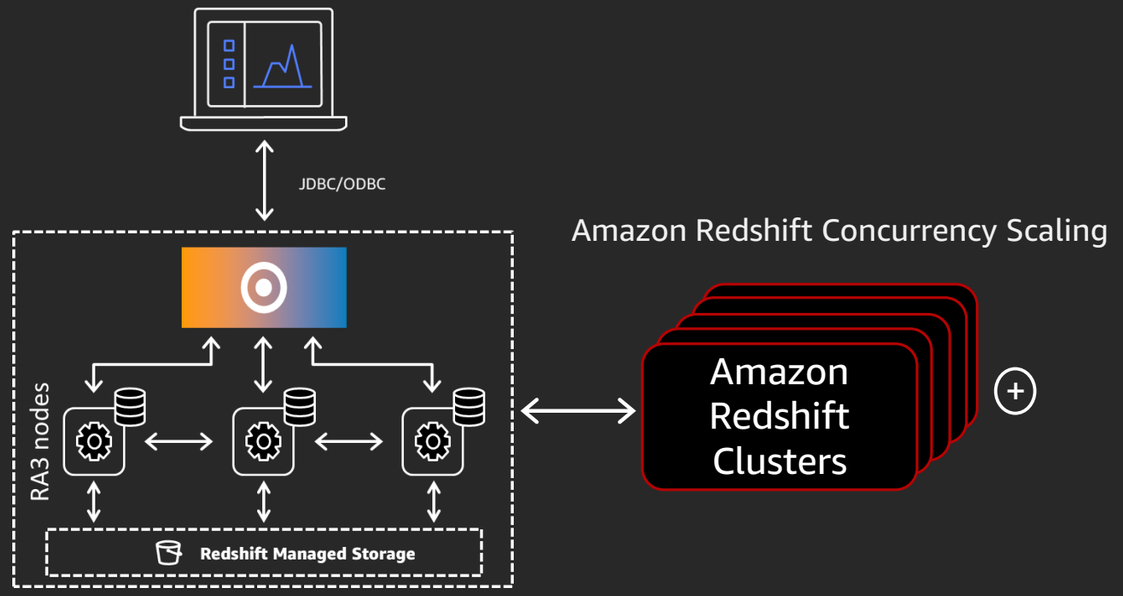
NEW! Faster cross regional copy and change

NEW! Auto Data Distribution

NEW! Deferred Maintenance

NEW! Elastic resize

Compute elasticity and scalability with concurrency scaling



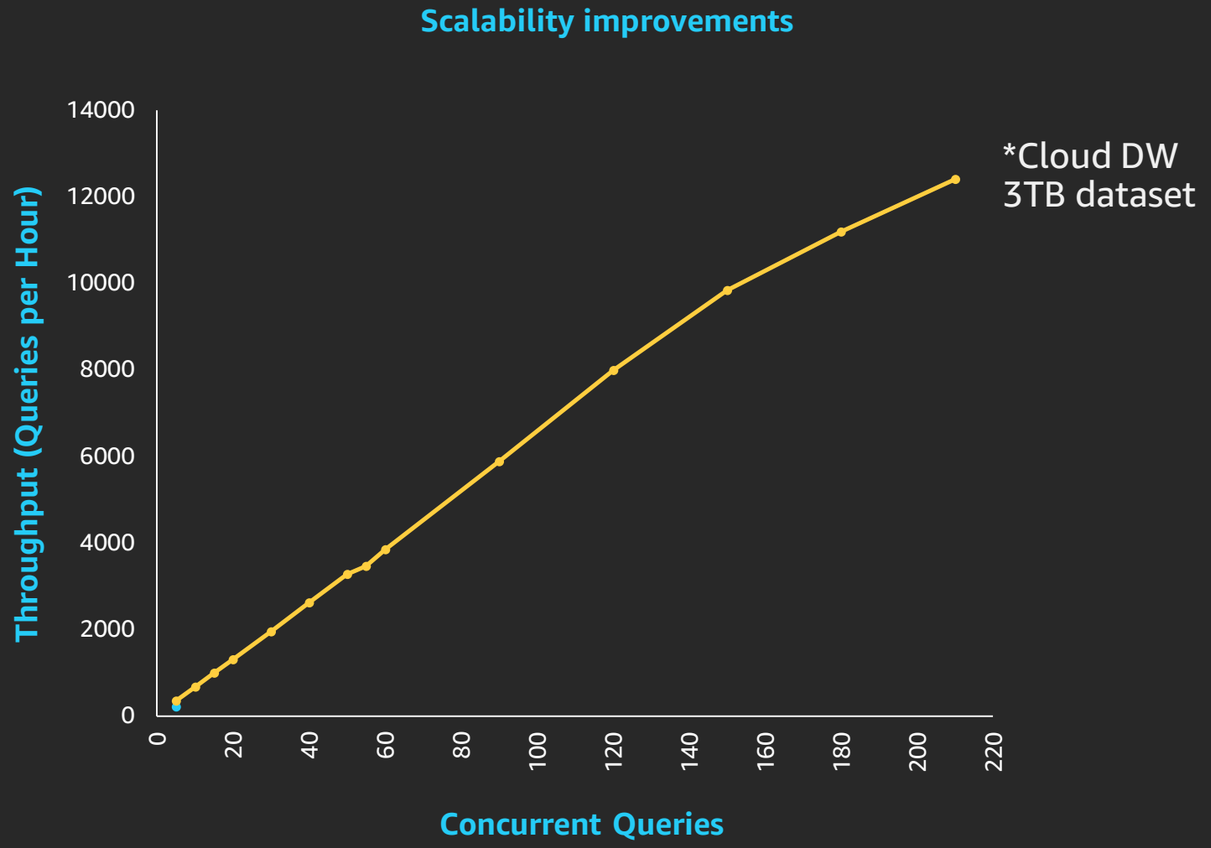
Scale out to multiple Amazon Redshift clusters from a single endpoint in seconds

Support virtually unlimited concurrent users and queries while maintaining SLAs

Per-second billing for additional clusters used

Free 1 hr. 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

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

Amazon Redshift federated query (preview)

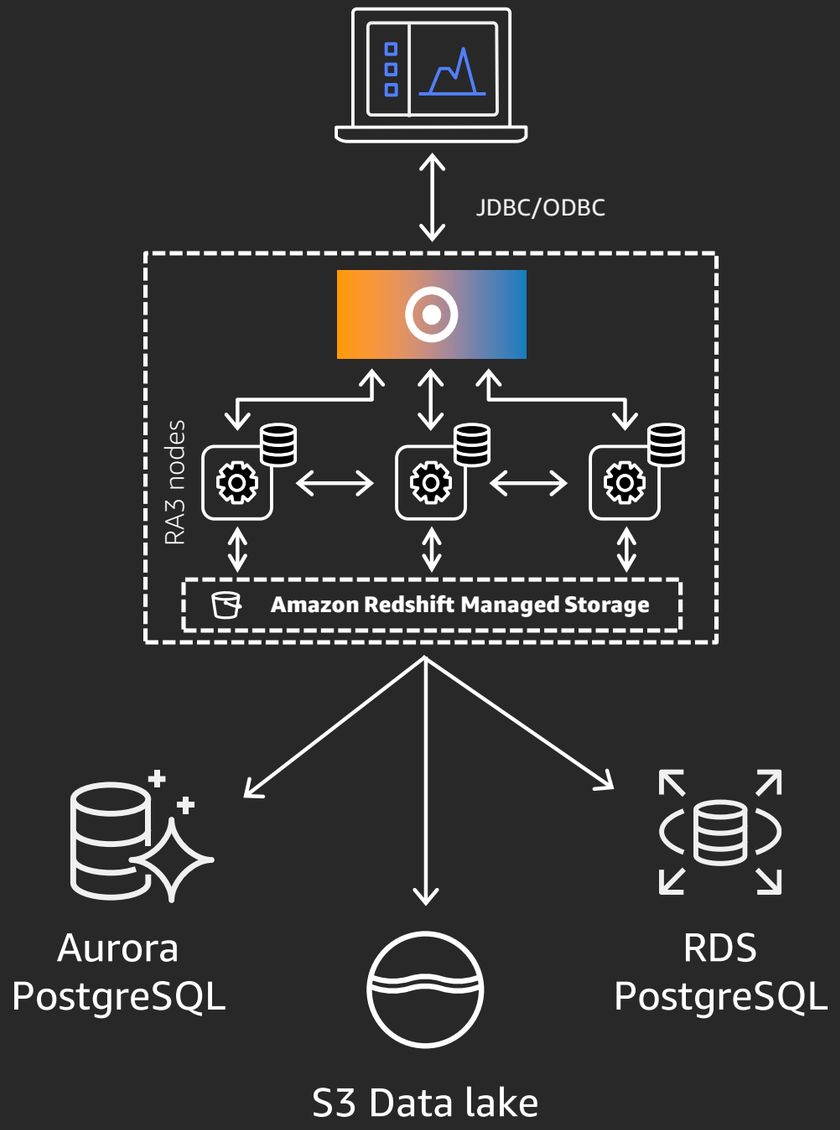
Queries on Amazon RDS and Aurora PostgreSQL databases

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



Amazon Redshift federated query: Most up-to-date data, no ETL required

```
-- Aurora Postgres has Hot Data (2019)
-- Redshift has Recent Data (2016-2018)
-- S3 has Archival Data (1992-1998)
-- Declare a view across all backends

CREATE VIEW lineitem_all AS
SELECT * FROM s3.lineitem_1t_part
UNION ALL
SELECT * FROM public.lineitem
UNION ALL
SELECT * FROM apg.lineitem
WITH NO SCHEMA BINDING
```

```
-- Find #sales with 1 item in Jan of each year
-- Predicates are being pushed down
-- Partition pruning on the S3 data
-- Aggregates are being pushed down
-- Very intuitive syntax

SELECT EXTRACT(year FROM l_shipdate) AS year,
       EXTRACT(month FROM l_shipdate) AS month,
       COUNT(*) AS orders
FROM lineitem_all
WHERE extract(month FROM l_shipdate) = 1
AND l_quantity < 2
GROUP BY 1,2
ORDER BY 1,2;
```

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

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

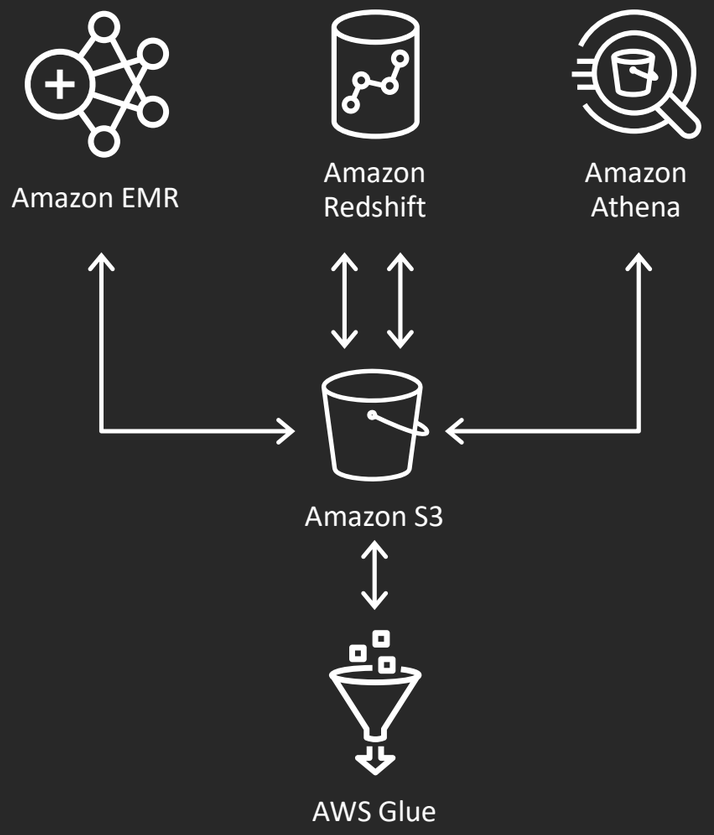
NEW! Faster cross regional copy and change

NEW! Auto Data Distribution

NEW! Deferred Maintenance

NEW! Elastic resize

Data lake export: Share data in Parquet format



Parquet is an efficient open columnar storage format for analytics

Analyze your data with Amazon Redshift Spectrum and other AWS services such as Amazon Athena and Amazon EMR

```
UNLOAD
('select * from lineitem')
TO
's3://mybucket/unload/lineitem/'
FORMAT as PARQUET
PARTITION BY (l_shipdate);
```

Self-optimizing and easy to maintain



Fully managed

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



Automatic optimizations

Self-learning and self-optimizing to turbocharge performance based on usage patterns



Easy to customize

Get the best out of the cluster hardware and spend where it matters to the business

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort

NEW! New management console

NEW! Auto WLM: Query priorities

NEW! Elastic resize scheduler

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

Amazon Redshift turbocharges 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

Automatic table sort

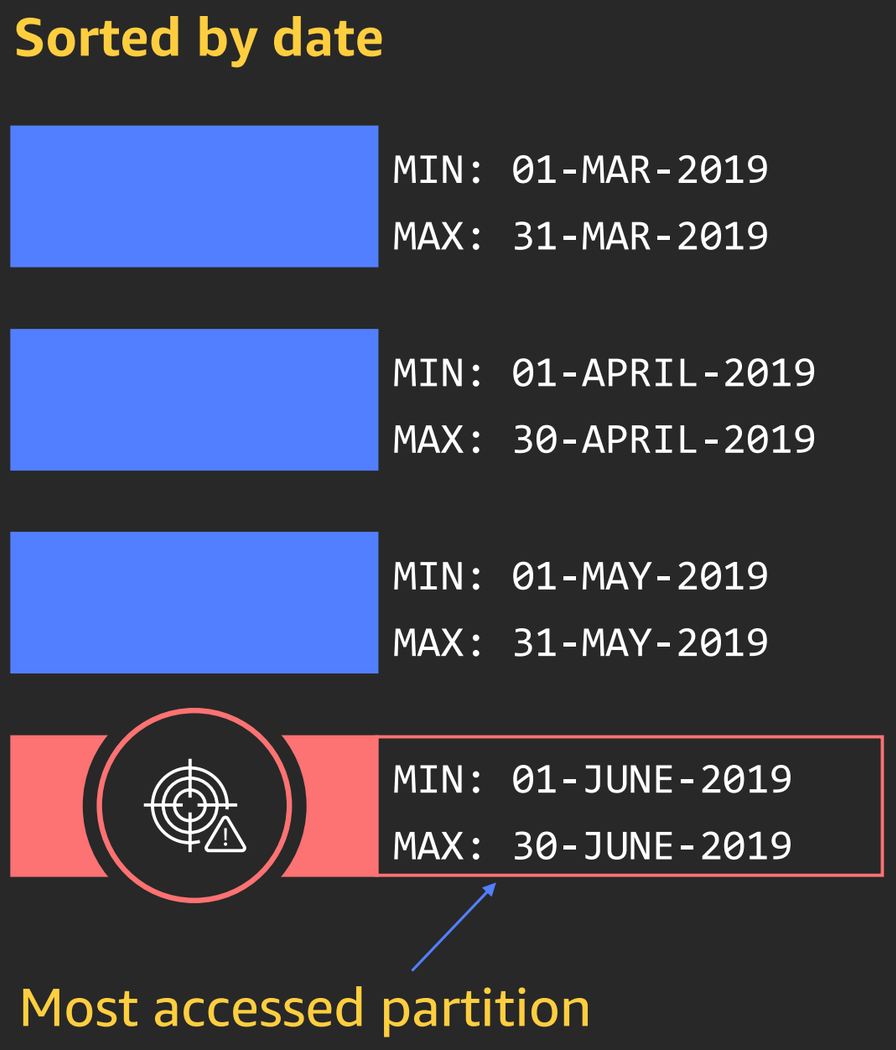
Eliminates DBA-managed sorting

Automates sort-order maintenance

Maximizes efficient use of compute

Leverages machine learning

No impact on performance or accessibility



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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort

NEW! New management console

NEW! Auto WLM: Query priorities

NEW! Elastic resize scheduler

NEW! Stored procedures

NEW! Distribution and sort key advisor

NEW! Cross-instance restore

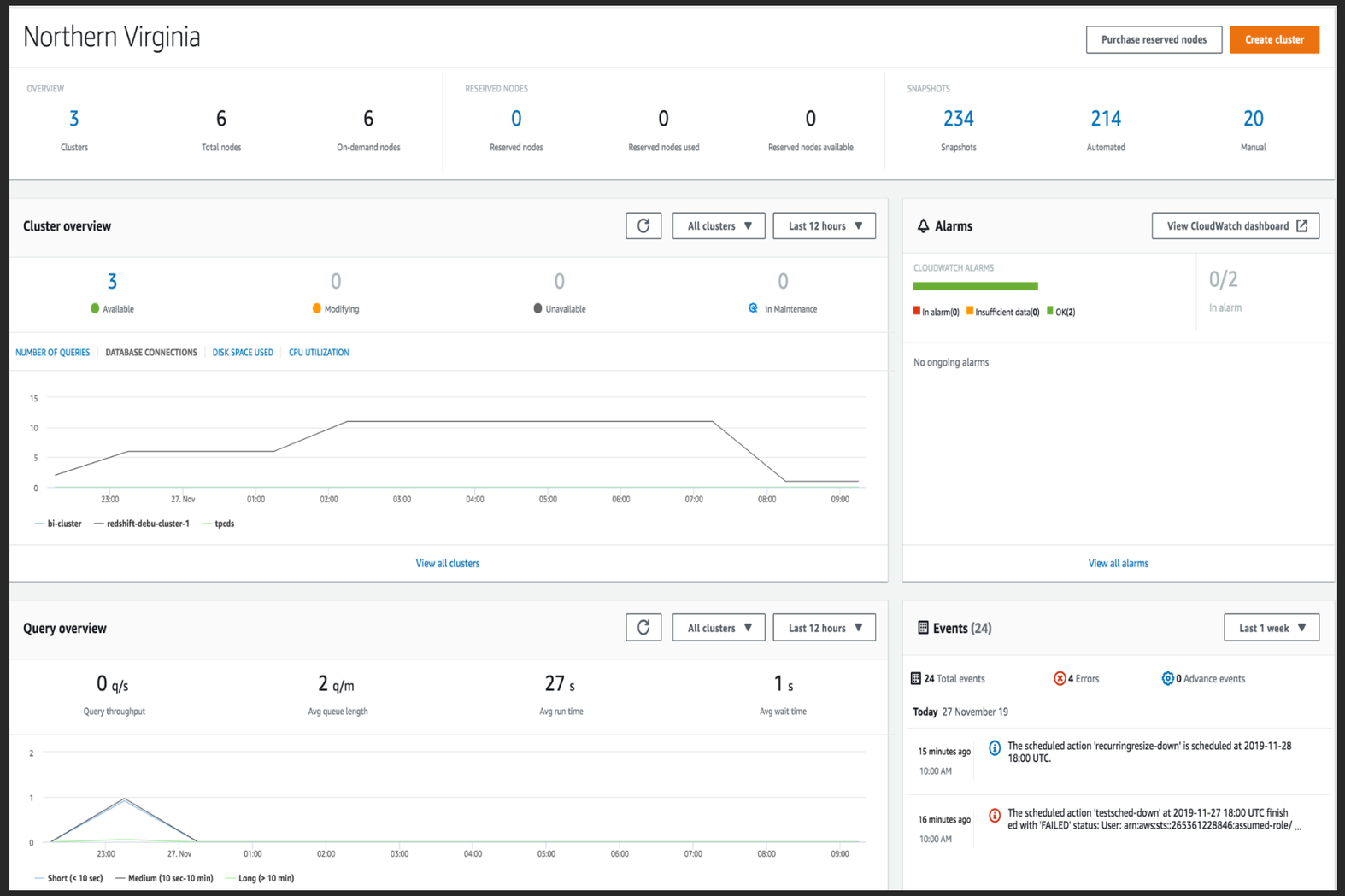
NEW! Faster cross regional copy and change

NEW! Auto Data Distribution

NEW! Deferred Maintenance

NEW! Elastic resize

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: Key 2019 innovations

Performance & scalability

NEW! RA3 with Amazon Redshift managed storage

PREVIEW! AQUA

NEW! AZ64

PREVIEW! Materialized views

NEW! Spatial data support

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort

NEW! New management console

NEW! Auto WLM: Query priorities

NEW! Elastic resize scheduler

NEW! Stored procedures

NEW! Distribution and sort key advisor

NEW! Cross-instance restore

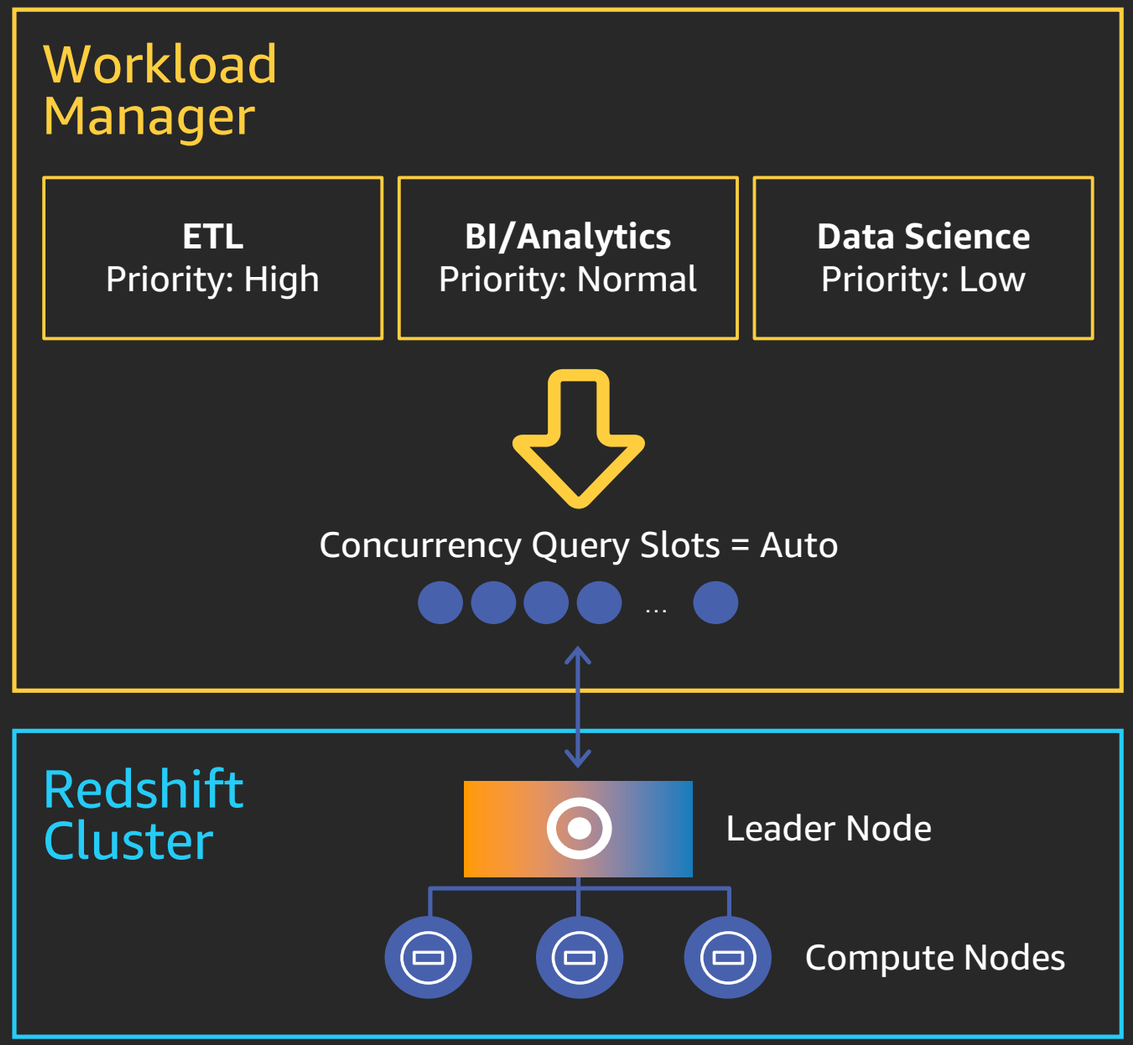
NEW! Faster cross regional copy and change

NEW! Auto Data Distribution

NEW! Deferred Maintenance

NEW! Elastic resize

Efficiency with automatic workload management and query priorities



- Manages concurrency dynamically
- Supports efficient sharing of cluster resources
- Maximizes throughput and performance
- Leverages machine learning to classify queries based on resources needs
- Ability to influence workload performance based on business priorities
- Ensures low priority queries make progress

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

NEW! Concurrency Scaling

Data lake & AWS integration

PREVIEW! Federated query across Amazon Redshift & RDS/Aurora

NEW! Data lake export in Parquet

NEW! Spectrum Request Accelerator

NEW! AWS Lake Formation integration

Fully managed

NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort

NEW! New management console

NEW! Auto WLM: Query priorities

NEW! Elastic resize scheduler

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

Stored procedures support to simplify migrations



Bring your existing **stored procedures** to Amazon Redshift

Migrating to Amazon Redshift is even easier!

Amazon Redshift supports Stored Procedures in PL/pgSQL format

Support for stored procedures provides the ability to run code where the data is, to efficiently run ETL, data validation, and custom business logic

```
CREATE OR REPLACE PROCEDURE test_spl(f1 int, f2 varchar)
AS $$
BEGIN
  RAISE INFO 'f1 = %, f2 = %', f1, f2;
END;
$$ LANGUAGE plpgsql;

Call test_spl(5, 'abc');
INFO: f1 = 5, f2 = abc
CALL
```

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
- NEW! Concurrency Scaling

Data lake & AWS integration

- PREVIEW! Federated query across Amazon Redshift & RDS/Aurora
- NEW! Data lake export in Parquet
- NEW! Spectrum Request Accelerator
- NEW! AWS Lake Formation integration

Fully managed

- NEW! Auto-Vacuum, Auto-Analyze & Auto Table Sort
- NEW! New management console
- NEW! Auto WLM: Query priorities
- NEW! Elastic resize scheduler
- 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

Workday Data-as-a-Service

Erol Guney
Architect, Data Platform
Workday Inc.

AWS re:Invent

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

aws

Session objectives



Understand the Workday Data-as-a-Service (DaaS) **vision and strategy**.



Learn about how Amazon Redshift allowed the Workday DaaS product team to **focus on architecture** goals and **business logic**.

Workday DaaS overview

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

Workday Safe Harbor statement

This presentation may contain forward-looking statements for which there are risks, uncertainties, and assumptions. If the risks materialize or assumptions prove incorrect, Workday's business results and directions could differ materially from results implied by the forward-looking statements. Forward-looking statements include any statements regarding strategies or plans for future operations; any statements concerning new features, enhancements or upgrades to our existing applications or plans for future applications; and any statements of belief. Further information on risks that could affect Workday's results is included in our filings with the Securities and Exchange Commission which are available on the Workday investor relations webpage: www.workday.com/company/investor_relations.php

Workday assumes no obligation for and does not intend to update any forward-looking statements. Any unreleased services, features, functionality or enhancements referenced in any Workday document, roadmap, blog, our website, press release or public statement that are not currently available are subject to change at Workday's discretion and may not be delivered as planned or at all.

Customers who purchase Workday, Inc. services should make their purchase decisions upon services, features, and functions that are currently available.

DaaS Benchmarking announcement

"Workday will be unveiling a new Data-as-a-Service offering, for which Benchmarking will be the first service."

The Benchmarking service lets customers select metrics that they wish to contribute to the service. In return, they gain access to Benchmarking data on the same metrics from peer groups.

Rising 2016

Workday announced

Data-as-a-Service (DaaS)

Benchmarking



Opt-in service

Metrics & trends

De-identified data

Workday Benchmarking



Workday Benchmarking enables Workday customers to compare their company's important business metrics with **other Workday customers** — all **within their tenant**.

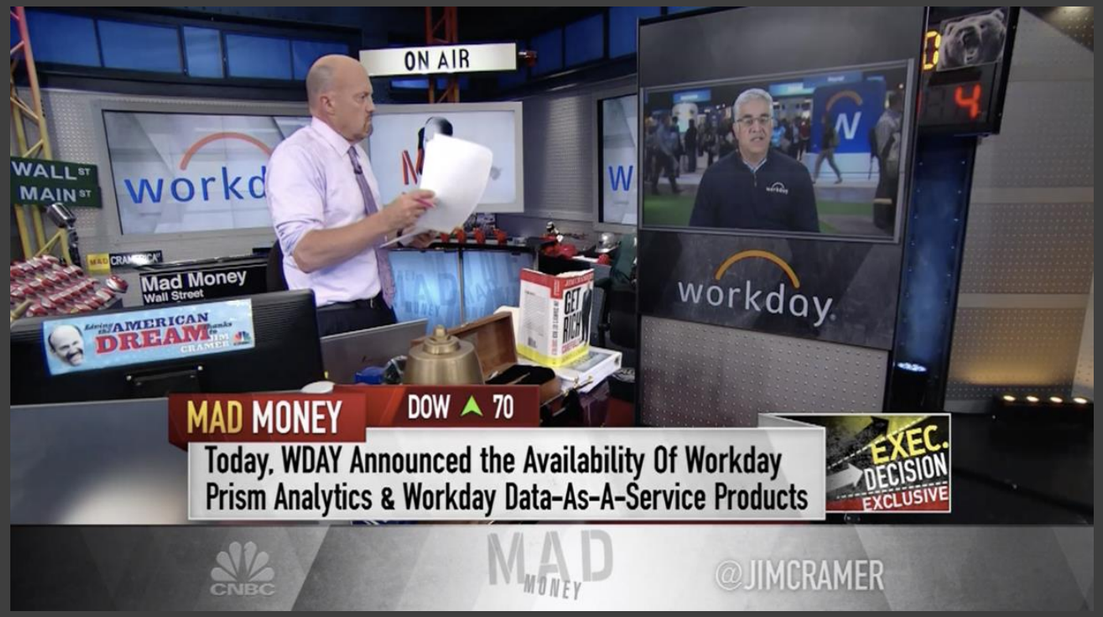


Customers can view benchmarks for the metrics in which they contribute their own data — **give to get it**.

Rising 2017

Workday delivered

Data-as-a-Service



100+
countries



DaaS Benchmarking delivered

100+ customers in less than a year

Customer comments:

- We are addicted to Workday Benchmarking!
- This program is such an innovative and valuable service offering to customers. Love how it leverages and expands the "Power of One" from your own tenant with unified products to the greater Workday ecosystem

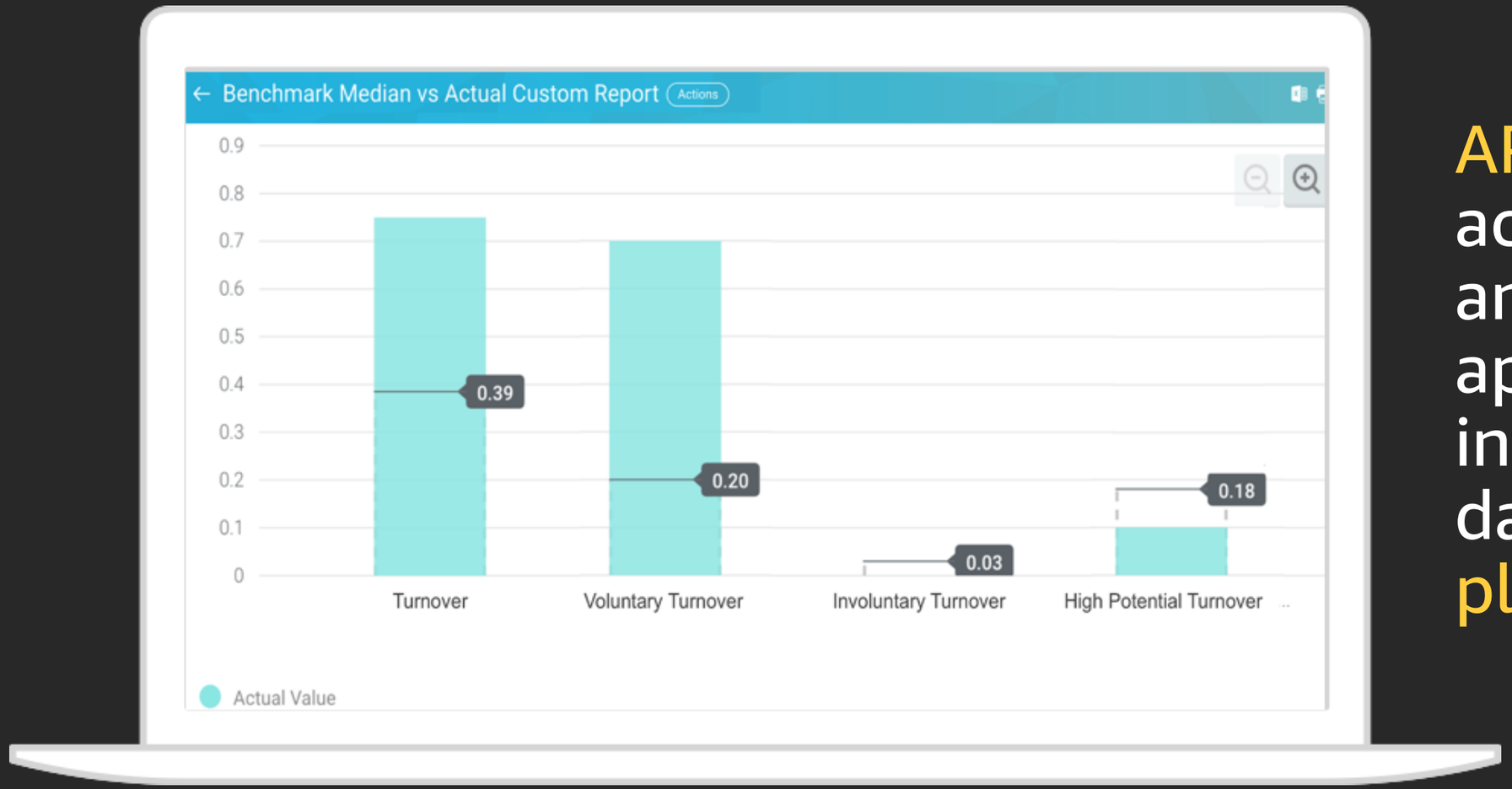
Today, we have enough customers to report benchmarks across 100+ countries.

What does it look like?

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

Benchmark report

Turnover: Comparison with industry



API-driven access allows any Workday application to interact with the data in the DaaS platform

Benchmark insights

Benchmark vs. Actual

Actions

Industry Group

Education, Non-Profit and Public Sector
Energy, Utilities, Transportation and Construction
Financial Services and Insurance
Healthcare
Manufacturing and Consumer Goods
More (3)

Size

<= 1,000
1,001 - 10,000
10,001 - 50,000
> 50,000

Include Benchmarking Details No

Tenure

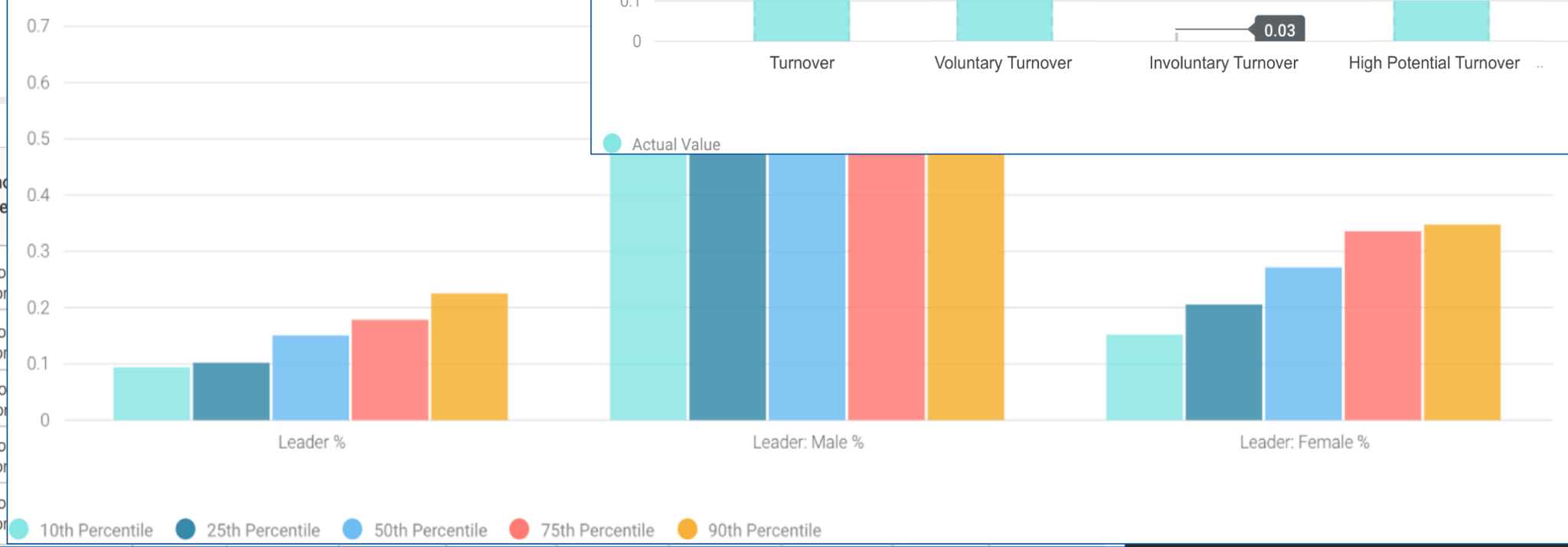
<= 1 years
1 - 2 years
> 2 years

5 items

Benchmark	Benchmark Subcategory	Benc Cate
Average Retirement Age (12 Months)	Age	Wo Cor
Average Workforce Age	Age	Wo Cor
Male to Female Ratio	Workforce Diversity	Wo Cor
Female %	Workforce Diversity	Wo Cor
Male %	Workforce Diversity	Wo Cor

Benchmark Explorer Custom Report

Actio



Benchmark Median vs Actual Custom Report

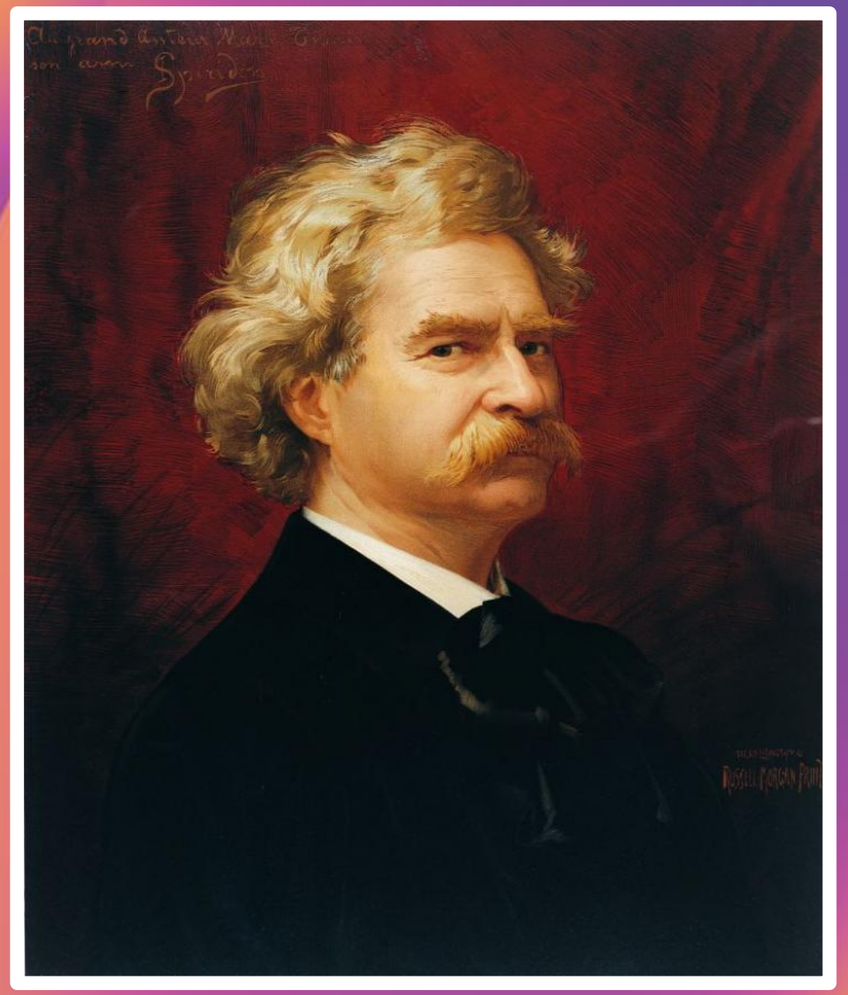
Actions



How did we do it?

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

“The secret of getting ahead is getting started.”

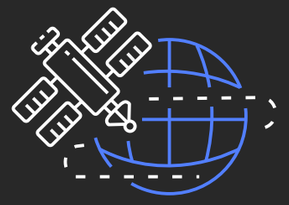


Mark Twain

Writer

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

Architectural must-haves



Global data architecture

Single master for schema and data



Security

Transport layer

Security and encryption at rest



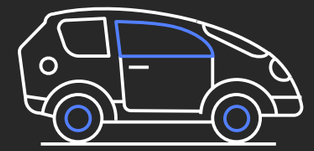
Privacy

Opt-in service, De-identified data GDPR and country opt-out



Architecture flexibility

Only introduce complexity where and when needed



Rationalized definitions

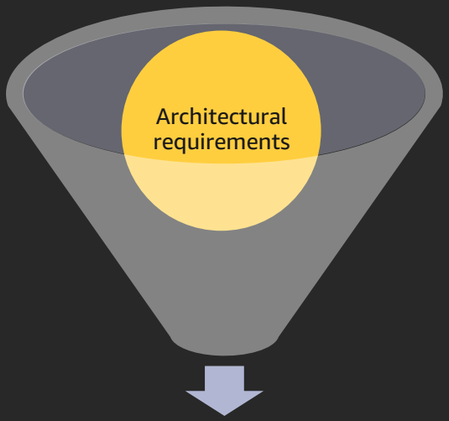
Common taxonomy across all tenants

How did Amazon Redshift meet the design goals of Workday DaaS?

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

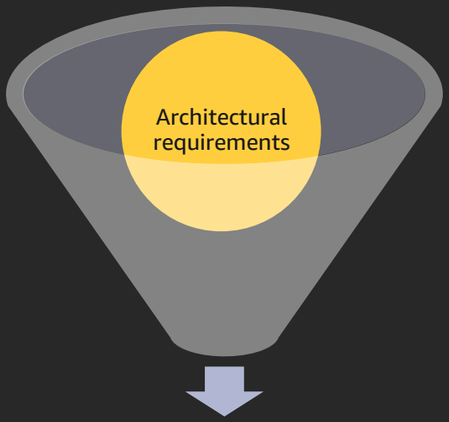
Why Amazon Redshift

Workday DaaS design goal	Architectural capability	Possible AWS solutions
Scalable, secure global storage	**TLS, and encryption at rest. 99.99% availability as per Workday SLA.**	**AWS KMS Amazon S3**
Ability to programmatically evolve	DB transactions with read after write consistency	Amazon Redshift Amazon RDS
Single master and federated	Programmatic backup and restore	Amazon Redshift Amazon RDS
Customers can opt in at any time	Ability to process large number of contributions	Amazon Redshift Massively Parallel Processing
Advance analytical queries on large dimensional datasets	Window slicing, dynamic partitioning	Amazon Redshift



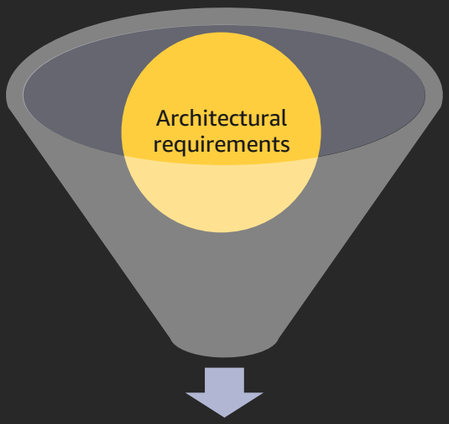
Why Amazon Redshift

Workday DaaS design goal	**Architectural capability**	**Possible AWS solutions**
Scalable, secure global storage	TLS, and encryption at rest. 99.99% availability as per Workday SLA.	AWS KMS AWS S3
Ability to programmatically evolve	**DB transactions with read after write consistency**	**Amazon Redshift Amazon RDS**
Single master and federated	Programmatic backup and restore	Amazon Redshift Amazon RDS
Customers can opt in at any time	Ability to process large number of contributions	Amazon Redshift Massively Parallel Processing
Advance analytical queries on large dimensional datasets	Window slicing, dynamic partitioning	Amazon Redshift



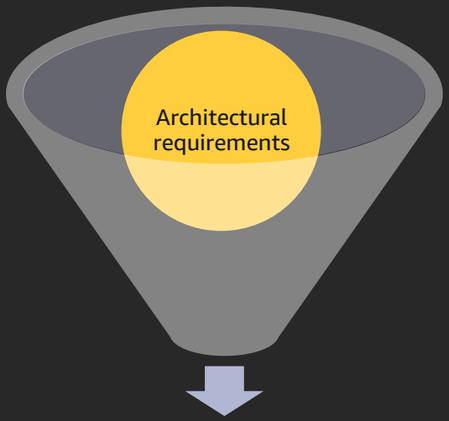
Why Amazon Redshift

Workday DaaS design goal	**Architectural capability**	**Possible AWS solutions**
Scalable, secure global storage	TLS, and encryption at rest. 99.99% availability as per Workday SLA.	AWS KMS Amazon S3
Ability to programmatically evolve	DB transactions with read after write consistency	Amazon Redshift Amazon RDS
Single master and federated	**Programmatic backup and restore**	**Amazon Redshift Amazon RDS**
Customers can opt in at any time	Ability to process large number of contributions	Amazon Redshift Massively Parallel Processing
Advance analytical queries on large dimensional datasets	Window slicing, dynamic partitioning	Amazon Redshift



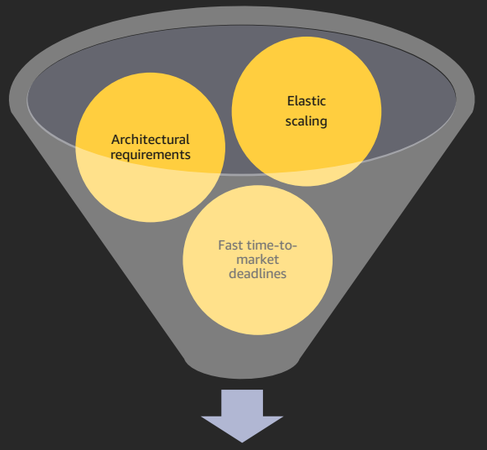
Why Amazon Redshift

Workday DaaS design goal	Architectural capability	Possible AWS solutions
Scalable, secure global storage	TLS, and encryption at rest. 99.99% availability as per Workday SLA.	AWS KMS Amazon S3
Ability to programmatically evolve	DB transactions with read after write consistency	Amazon Redshift Amazon RDS
Single master and federated	Programmatic backup and restore	Amazon Redshift Amazon RDS
Customers can opt in at any time	**Ability to process large number of contributions**	**Amazon Redshift Massively Parallel Processing**
Advance analytical queries on large dimensional datasets	Window slicing, dynamic partitioning	Amazon Redshift



Why Amazon Redshift

Workday DaaS design goal	Architectural capability	Possible AWS solutions
Scalable, secure global storage	TLS, and encryption at rest. 99.99% availability as per Workday SLA.	AWS KMS Amazon S3
Ability to programmatically evolve	DB transactions with read after write consistency	Amazon Redshift Amazon RDS
Single master and federated	Programmatic backup and restore	Amazon Redshift Amazon RDS
Customers can opt in at any time	Ability to process large number of contributions	Amazon Redshift Massively Parallel Processing
Advance analytical queries on large dimensional datasets	**Window slicing, dynamic partitioning**	**Amazon Redshift**

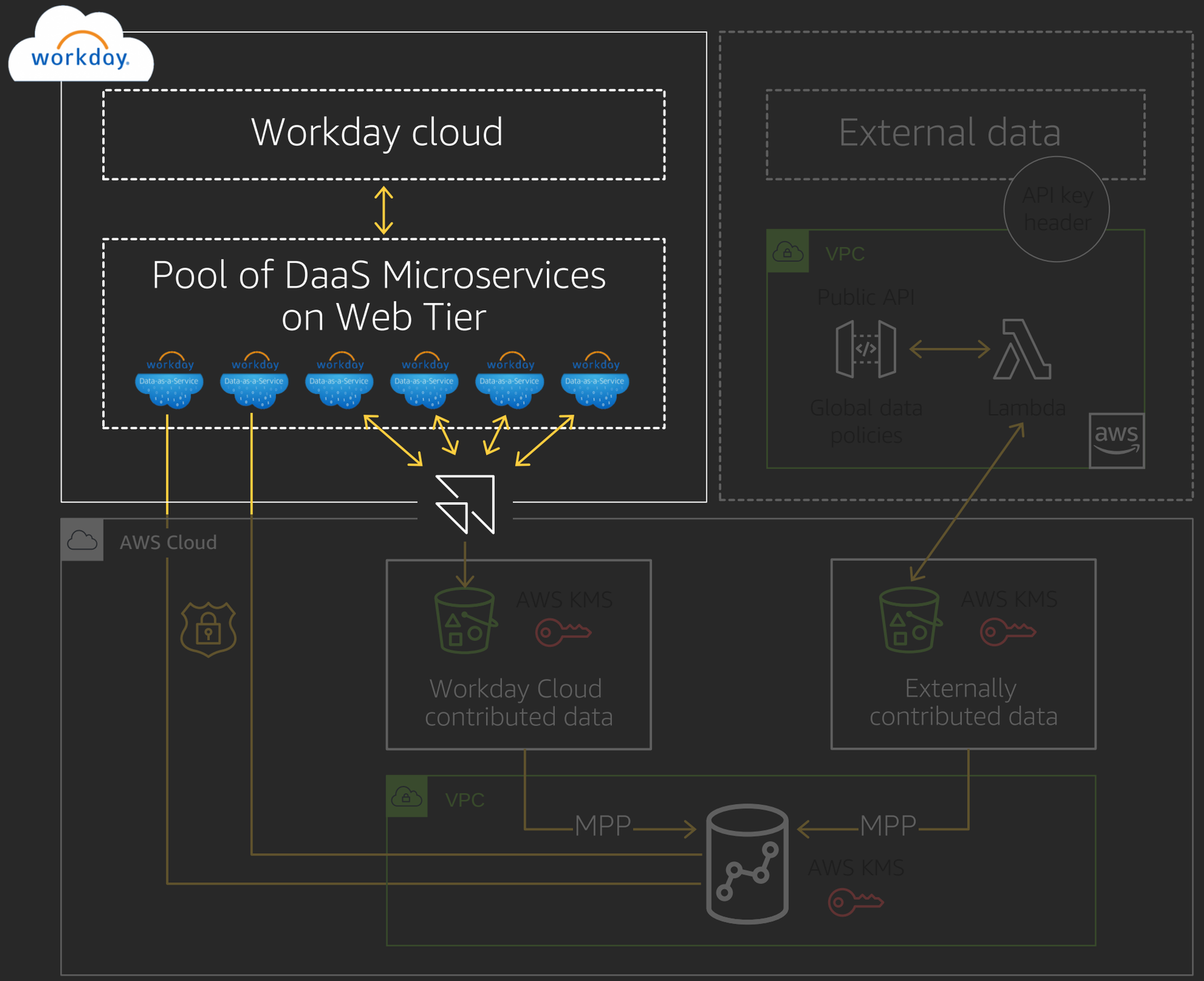


Amazon Redshift

Napkin architecture

High-level diagram to illustrate data flow

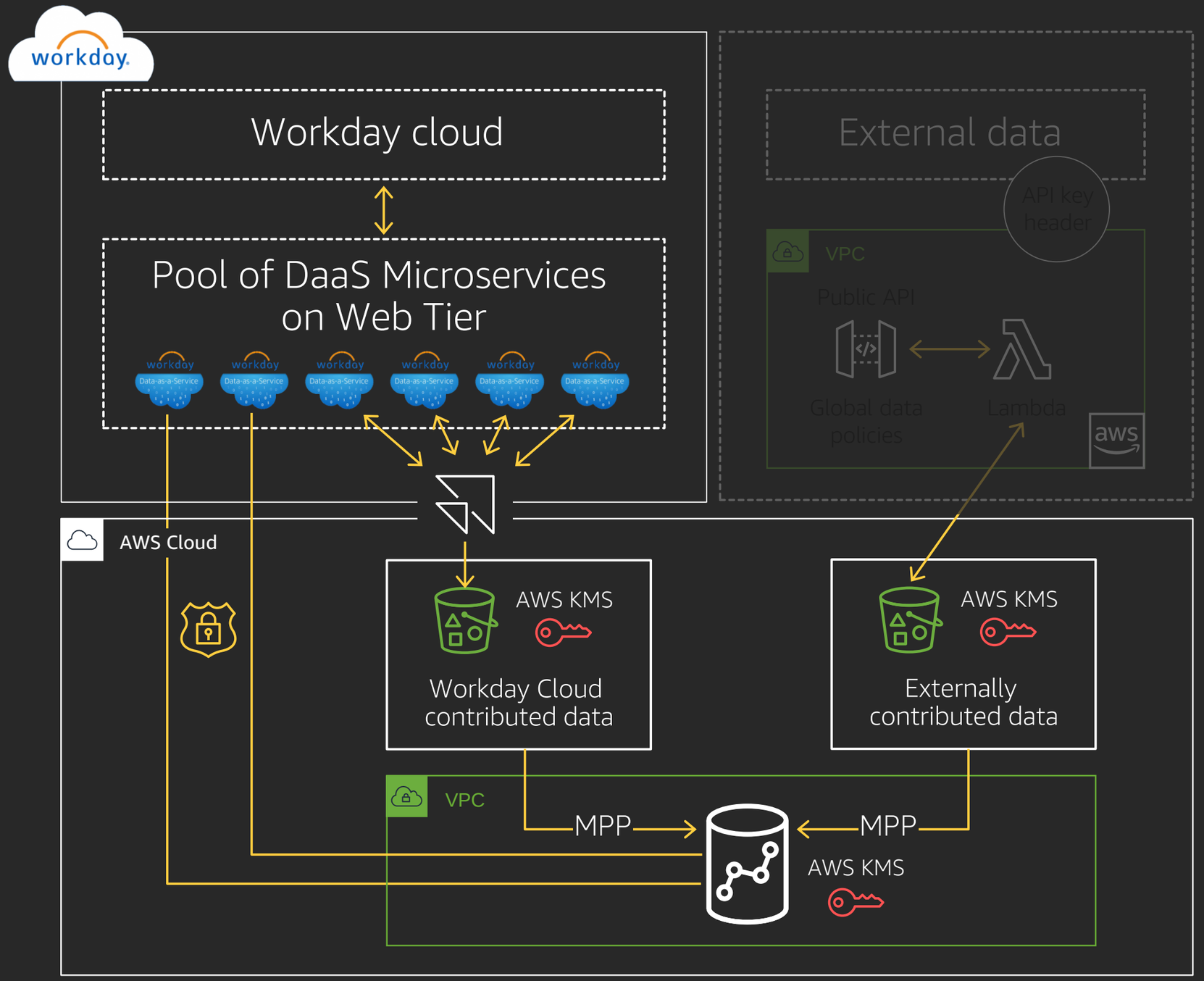
MPP: Massively parallel processing



Napkin architecture

High-level diagram to illustrate data flow

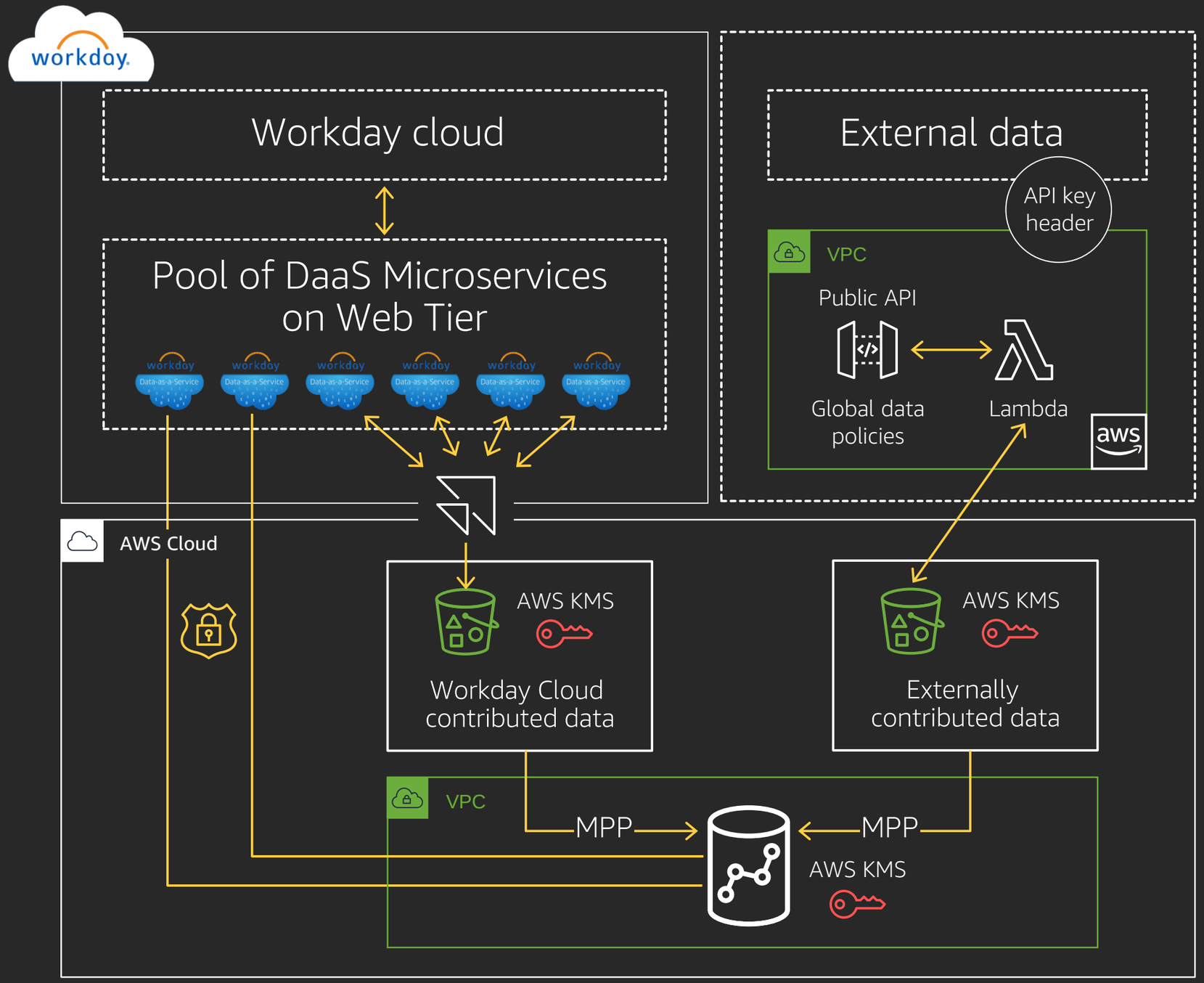
MPP: Massively parallel processing



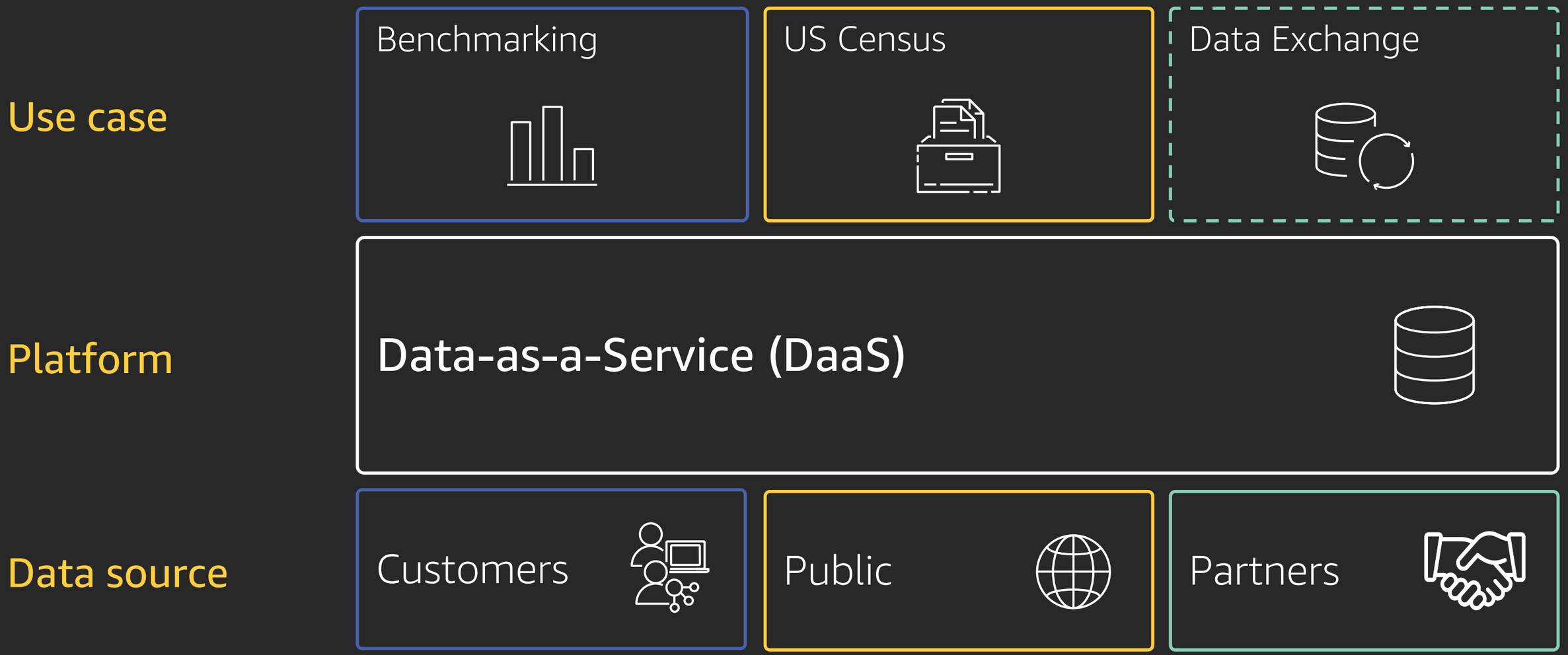
Napkin architecture

High-level diagram to illustrate data flow

MPP: Massively parallel processing



DaaS data sources

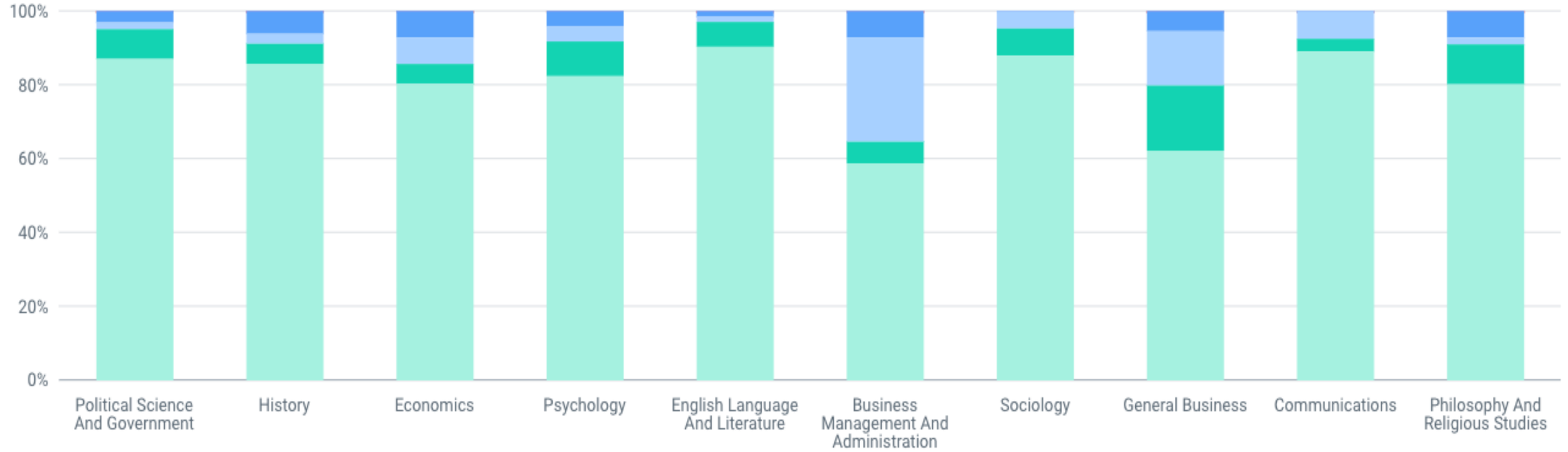


US Labor Market - Degree+Levels

Actions

Year 2017 **Occupation** Lawyers, And Judges, Magistrates, And Other Judicial Workers

State California



Professional degree beyond a bachelor's degree
Doctorate degree
Bachelor's degree
Master's degree
1 or more years of college credit, no degree
Regular high school diploma
GED or alternative credential
Associate's degree
Some college, but less than 1 year

Total for Count of Laborers 159,725

US Labor Market Statistics - Occupation

Actions

State California

Year 2017

County San Francisco County (Central)--Inner Mission & Castro
San Francisco County (Central)--South of Market & Potrero
San Francisco County (Central)--Sunset District (North)
San Francisco County (North & East)--North Beach & Chinatown
San Francisco County (North & West)--Richmond District
More (2)

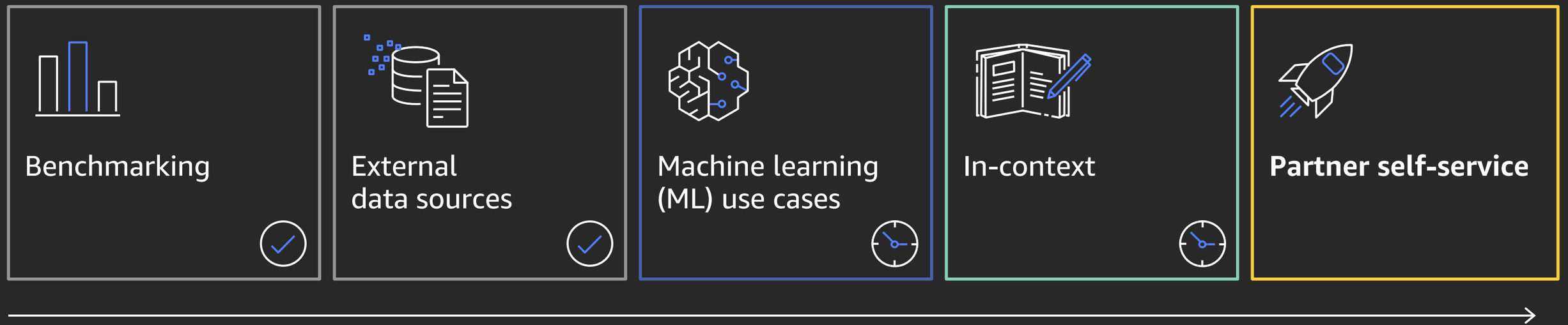
338 items

Occupation	Year	Size of Workforce
Miscellaneous Managers, Including Funeral Service Managers And Postmasters And Mail Superintendents	2017	26,485
Software Developers, Applications And Systems Software	2017	22,985
Lawyers, And Judges, Magistrates, And Other Judicial Workers	2017	14,017
Designers	2017	12,888
Accountants And Auditors	2017	12,743
Marketing And Sales Managers	2017	11,354
Retail Salespersons	2017	10,792
Registered Nurses	2017	8,550
Chief Executives And Legislators	2017	8,538

So much data out there!



What’s next for Workday DaaS



Related sessions

- 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!

Thank you!

Michalis Petropoulos
Director, Engineering
Amazon Redshift

Erol Guney
Architect, Data Platform
Workday Inc.

AWS re:Invent

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

aws



Please complete the session survey in the mobile app.

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