
AWS
re:Invent
NOV. 29 – DEC. 3, 2021 | LAS VEGAS, NV

ANT301

Concurrency and scalability strategies with Amazon Redshift

Steven Cochrane
Director of Data & Analytics
Flutter Entertainment UK&I

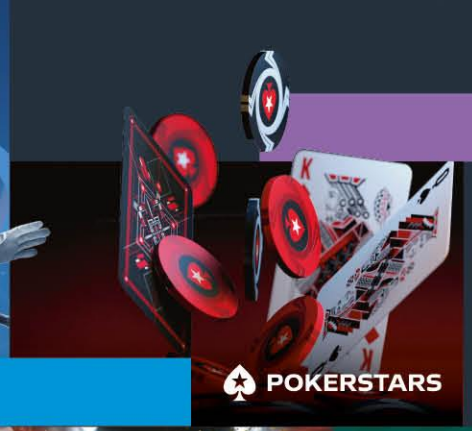
Ippokratis Pandis
Senior Principal Engineer
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift at the heart of a modern data platform

Flutter Entertainment









This is
Flutter
Entertainment









sportsbet

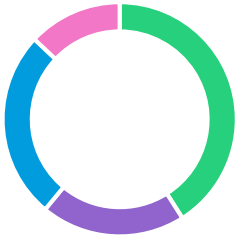




At a glance

Flutter Entertainment is the world's largest online gambling operator. Our tech-enabled business brings entertainment to life for over 14 million customers in a safe and sustainable way through some of the world's most innovative and distinctive brands operating across 100+ international markets.

Global footprint



- UK&I 41%
- Australia 20%

14 million
Active customers

£5.3 billion
Total pro forma net revenue*

100+
International markets

14,000+
Employees

UK&I	Australia	International	US
sky betting & gaming	sportsbet	POKERSTARS	FANDUEL
PADDYPOWER.		betfair	TVG
betfair		adjarabet.com	FOXBET
		JUNGLEE GAMES	POKERSTARS

Flutter

* [2020 figures]

Our UK&I division

- We operate some of the most popular sports betting and gaming brands in the UK&I market. With an estimated 28% market share, Sky Betting & Gaming, Paddy Power and Betfair offer market leading innovation to millions of customers every week.*
- Although the brands mostly operate online, this division also includes 625 Paddy Power betting shops in the UK and Ireland.
- Created following the 2020 merger with The Stars Group, the UK&I division is currently undergoing transformation to harness the combined strengths of each brand.
- Today we run a single data and analytics capability that supports the Paddy Power and Betfair brands across online and retail channels covering all sports and gaming products.

PADDYPOWER.



betfair



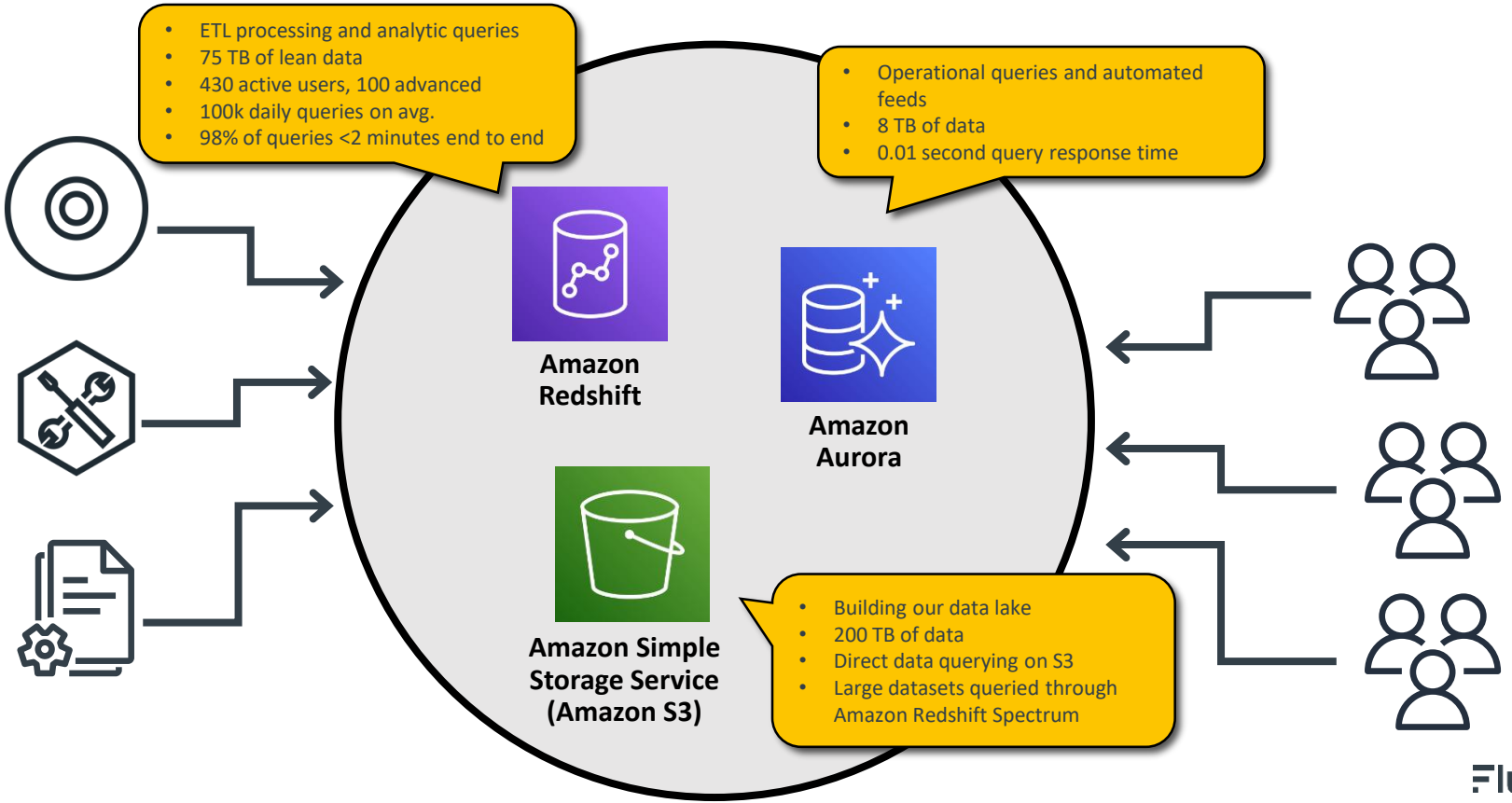
sky betting & gaming



Flutter

* [2020 figures]

Our (very simplified) data platform in 2019



Our challenges

Big events

- We had to prioritise production flows
- Restricting users and access to data during peak periods to protect mission critical processes



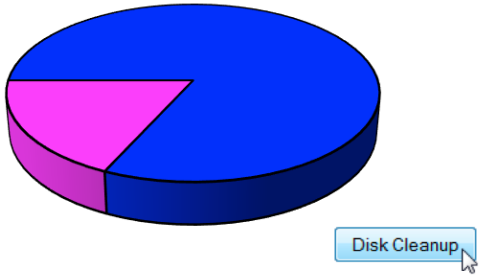
Large number of users

- We had to throttle their queries



Disk space

- We had to choose what to deliver
- We couldn't do everything



Flutter

We tried to fix it…

- We used other AWS services to supplement Amazon Redshift
- We wanted to keep close to SQL as our team is experienced there
- Amazon Aurora was used to serve lower latency requirements
- Amazon Redshift Spectrum was used to gain more storage space



We had multiple copies of the data

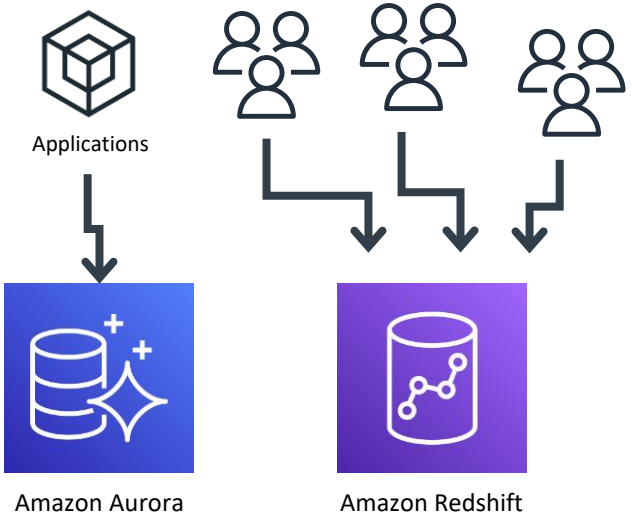
It was confusing for stakeholders and engineers to understand where a dataset was or where it should go

We still had performance, data volume, or cost challenges

Flutter

Amazon Redshift RA3 changed everything

- Amazon Redshift RA3 freed us from the constraints of data volumes as it separated storage from compute
- Amazon Redshift Concurrency Scaling allows us to automatically scale up and down
- We re-thought how we wanted our data platform to work and split out user workload from programmatic/automated load



Giving one place for our users to go for data

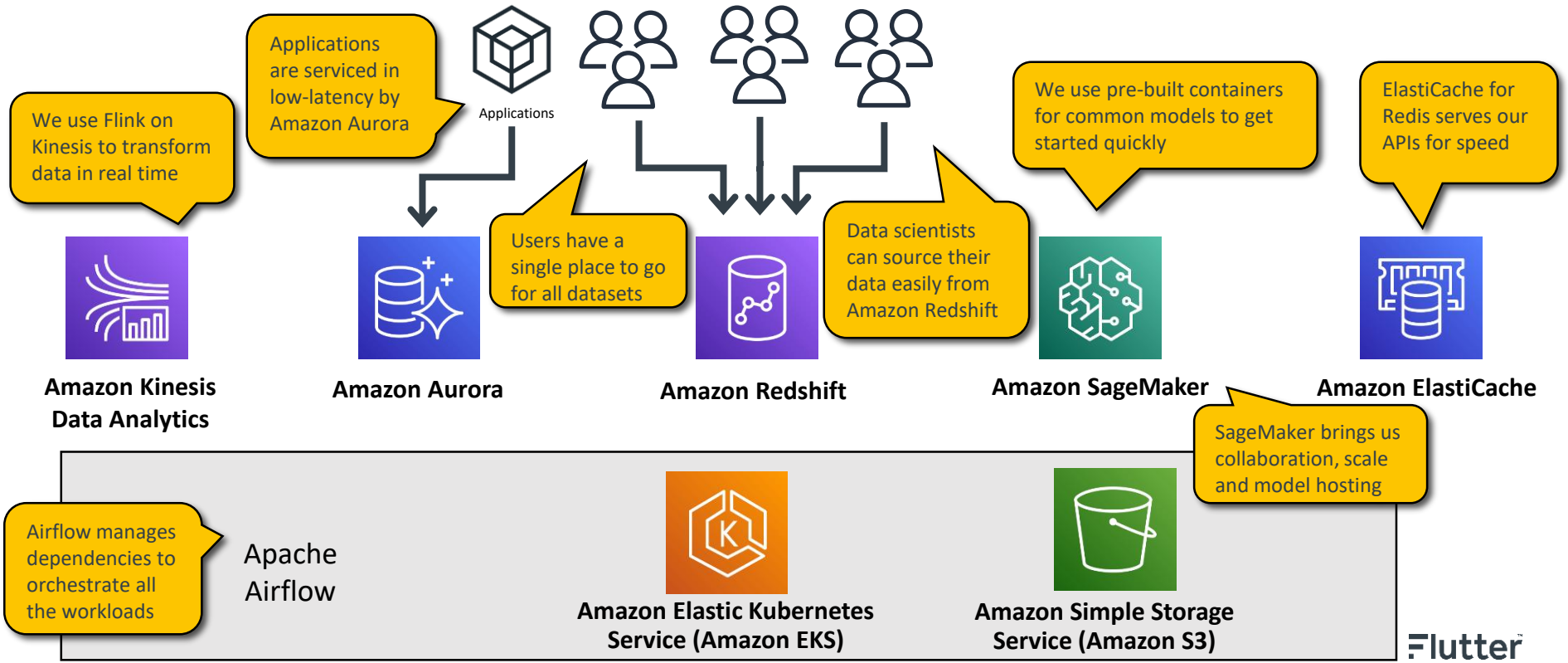
Making it easy for our engineers to understand what each technology should be used for

Giving us much more scale

Making the platform easier to maintain

Flutter

Rethinking our data platform



Flutter

The scale we now deliver at

560+
Direct access Redshift users
(430+ in 2019)

4.4 million
Queries per month
(3 million in 2019)

1 thousand
BI tool users
(1K in 2019)

910 TB
Data
(~300 TB in 2019)

35+
Data sources
(30 in 2019)

3.5+ thousand
ETL jobs
(~3K in 2019)

Flutter

Have we achieved what we set out to do?



Technical

Remove constraints of disk space and concurrency



Team

Have one place to go for each type of workload, everyone knows what each service is for



Reliability

Utilise benefits of the cloud to improve reliability with features such as automatic scaling

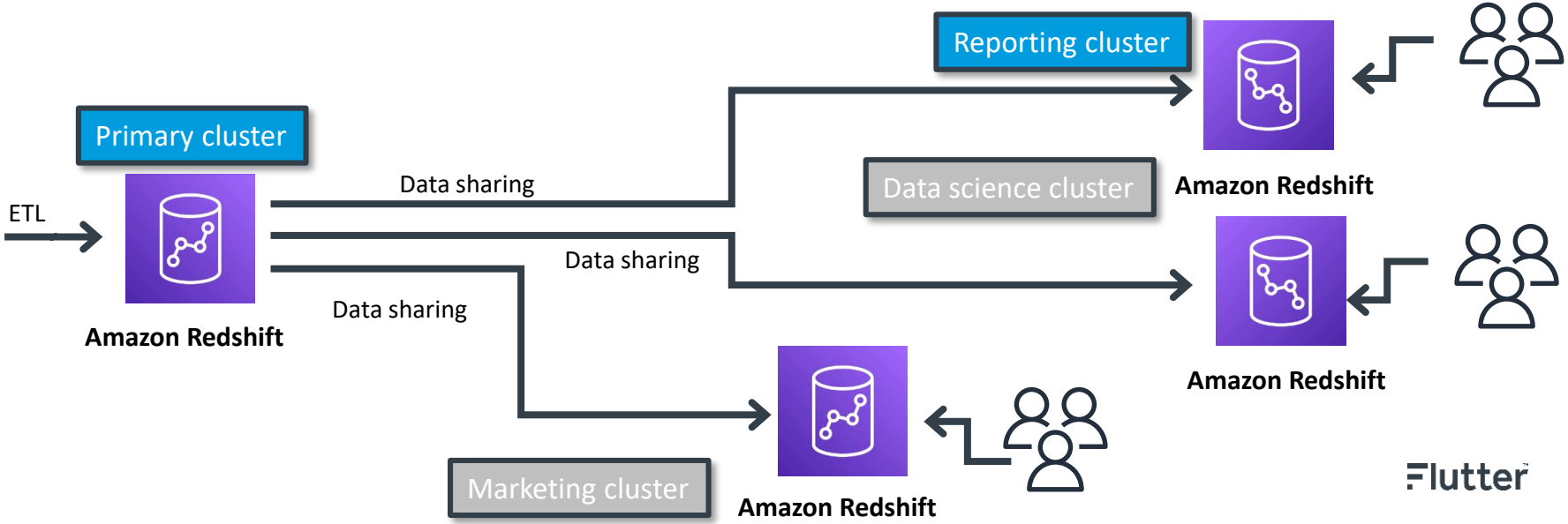


Maintenance

Reduce maintenance and tuning effort with workload isolation

What's next…

- We want to further isolate workloads to ensure scale and reliability
- Utilising Amazon Redshift data sharing to have multiple smaller clusters scaling independently for separated users and workloads
- Initial use case is separating interactive reporting/self service analytics users



Flutter

What we learned



AWS are developing their platforms to keep up with customer needs (*tell them what you need)*



Match your workloads to the appropriate AWS service, to make best use of the resources and save costs



Stick to the different best practices for each AWS service and find the gems in the docs



It’s good to understand the AWS product roadmap ahead of investing in in-house built capabilities



Early adoption of new technologies comes with a lot of benefits, and some risks, which can be mitigated by working closely with AWS

Amazon Redshift: Performance and elasticity

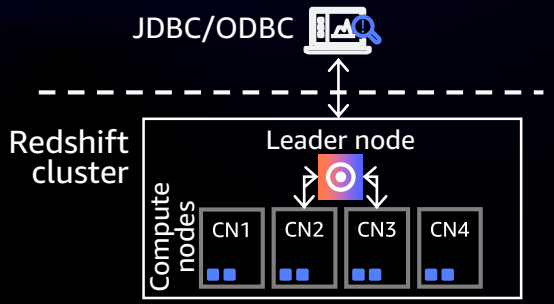
Ippokratis Pandis
Senior Principal Engineer
AWS

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

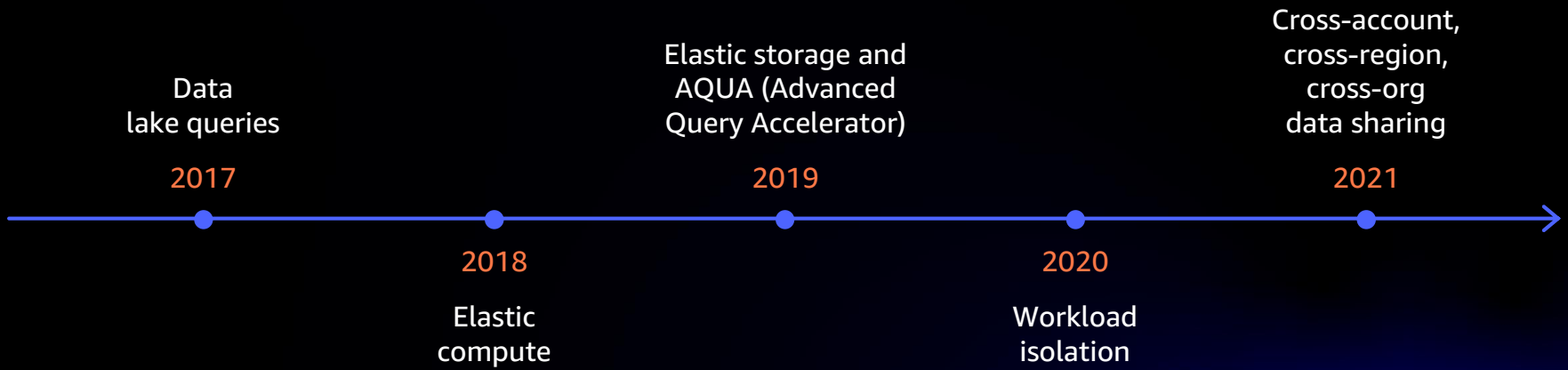
The evolution of Amazon Redshift

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift architecture: 2013

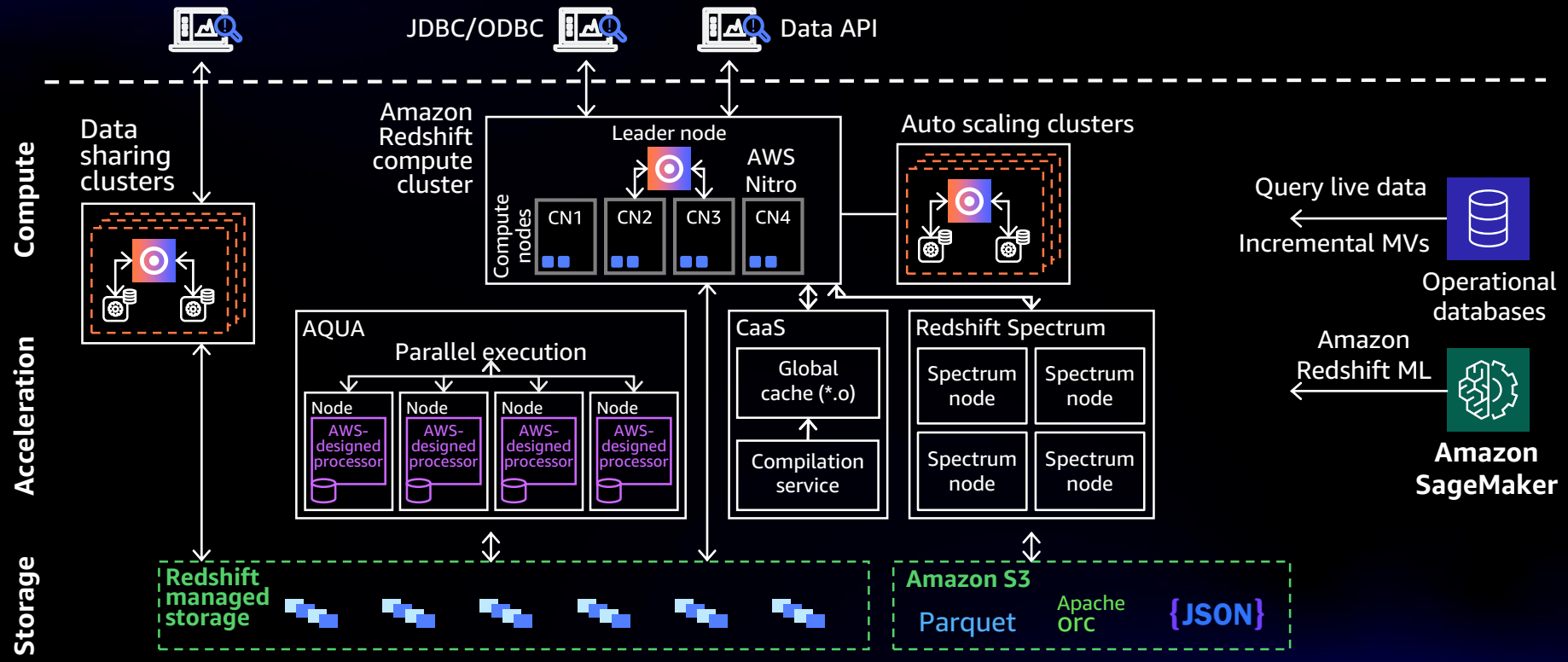


- Cloud-native data warehouse
- Best-in-class performance at scale
  - Columnar format
  - Distributed query processing
- Strong transactional consistency
- Enterprise class security
- Fully managed & self-tuning



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Amazon Redshift architecture: 2021



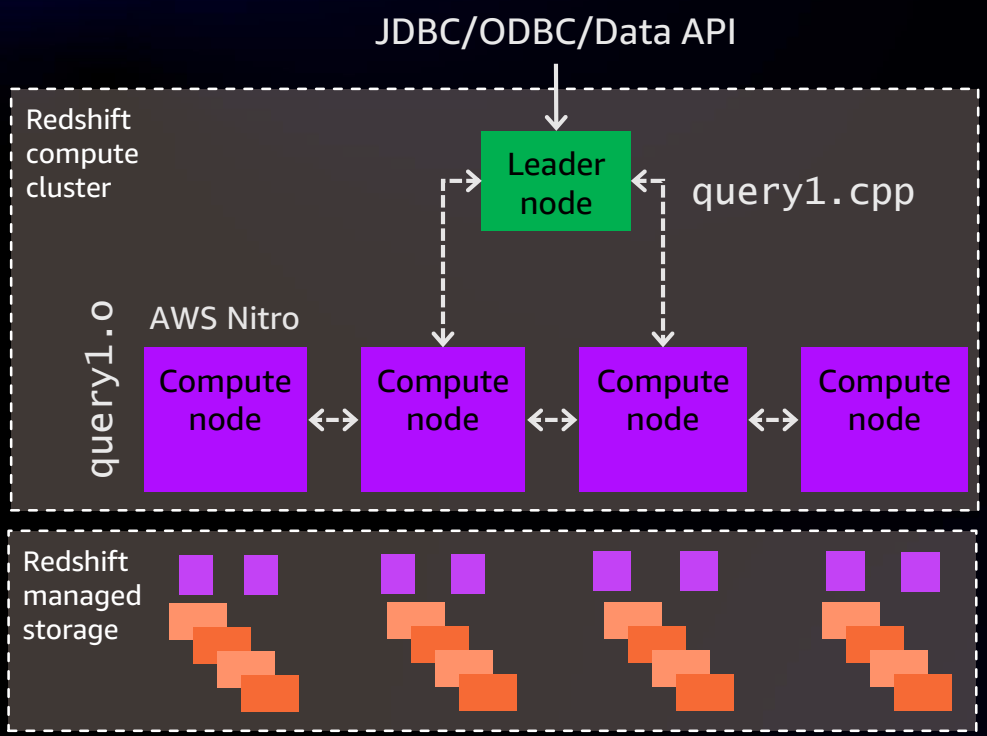
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

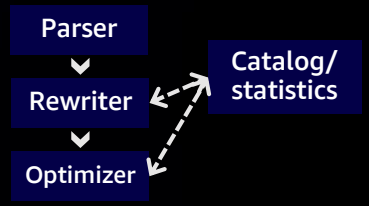
The evolution of Amazon Redshift

0. Security and availability
1. Performance & elasticity
2. Integration with the AWS environment
3. Autonomics
4. Serverless

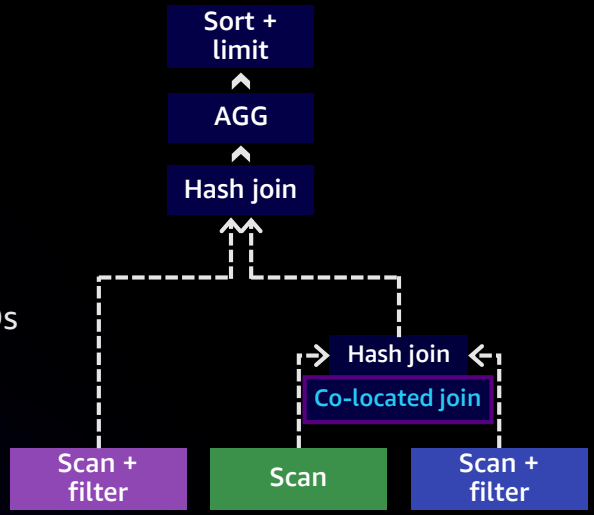
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Executing a query in Amazon Redshift





Min/max pruning
SIMD scans from local-attached SSDs
AZ64 encoding



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Code generation, pre-fetching, vectorization

```
SELECT sum(R.val) FROM R, S WHERE R.key = S.key AND R.val < 50
```

```
// Loop over the tuples of R.
while (scan_step->has_next()) {
  // Get next value for R.key.
  auto field1 = fetcher1.get_next();
  // Get next value for R.val.
  auto field2 = fetcher2.get_next();
  // Apply predicate R.val < 50.
  if (field2 < 50) {
    // Hash R.key and probe the hash table.
    size_t h1 = hash(field1) & (ht1_sz - 1);
    for (auto* p1 = ht1[h1];
         p1 != nullptr; p1 = p1->next) {
      // Evaluate the join condition R.key = S.key.
      if (field1 == p1->field1) sum1 += field2;
    }
  }
}
```

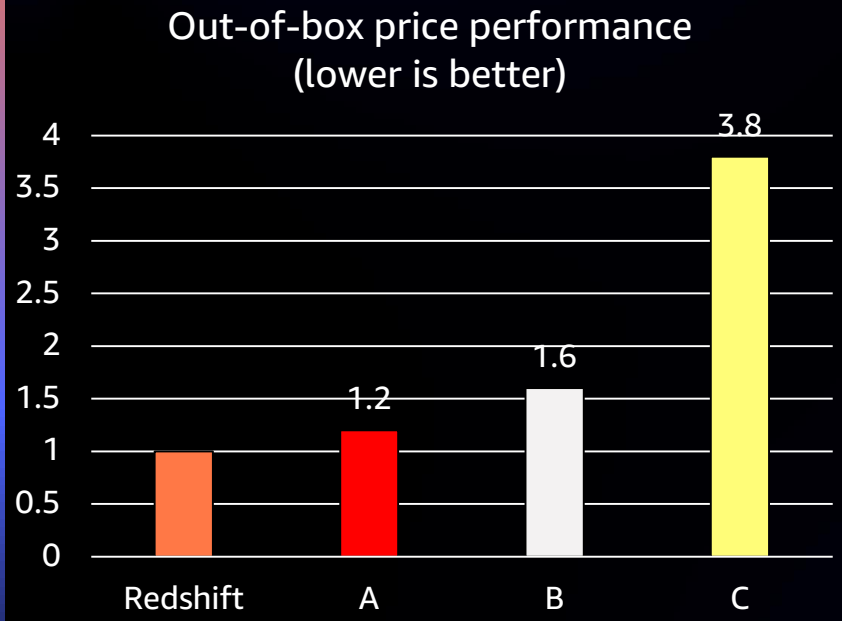
```
// Hash R.key and probe the hash table.
size_t h1 = hash(field1) & (ht1_sz - 1);
// Prefetch the current tuple.
prefetch(ht1[h1]);
// Push the current tuple in the buffer.
prefetchbuffer1.push(field1, field2);
// Pop a prefetched earlier tuple.
h1, field1, field2 =
    prefetchbuffer1.pop();
// Probe the earlier tuple.
for (auto* p1 = ht1[h1];
     p1 != nullptr; p1 = p1->next) {
   ...
}
```

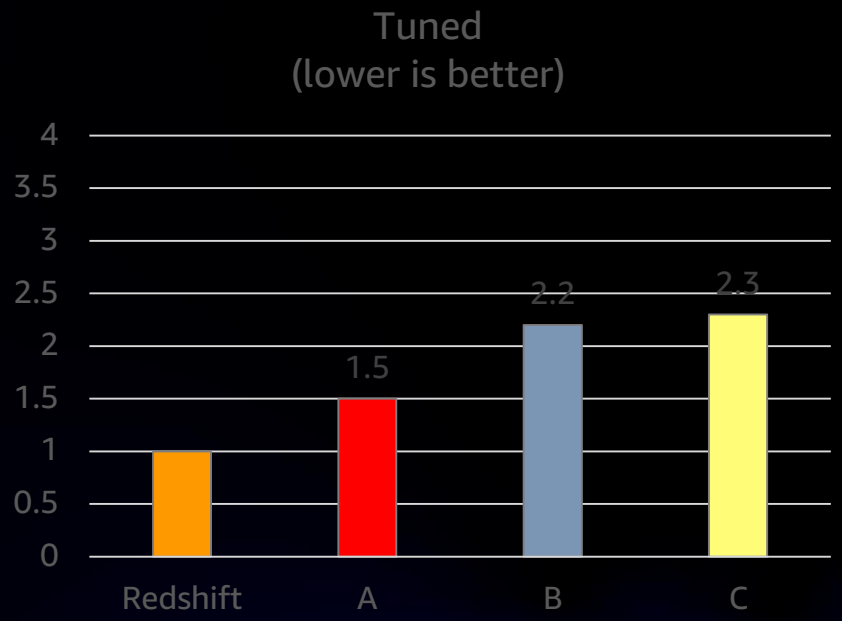
+ AVX2/AVX512 vectorization
+ Min/max pruning, late materialization

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Leading price/performance

Cloud DW 3TB benchmark*





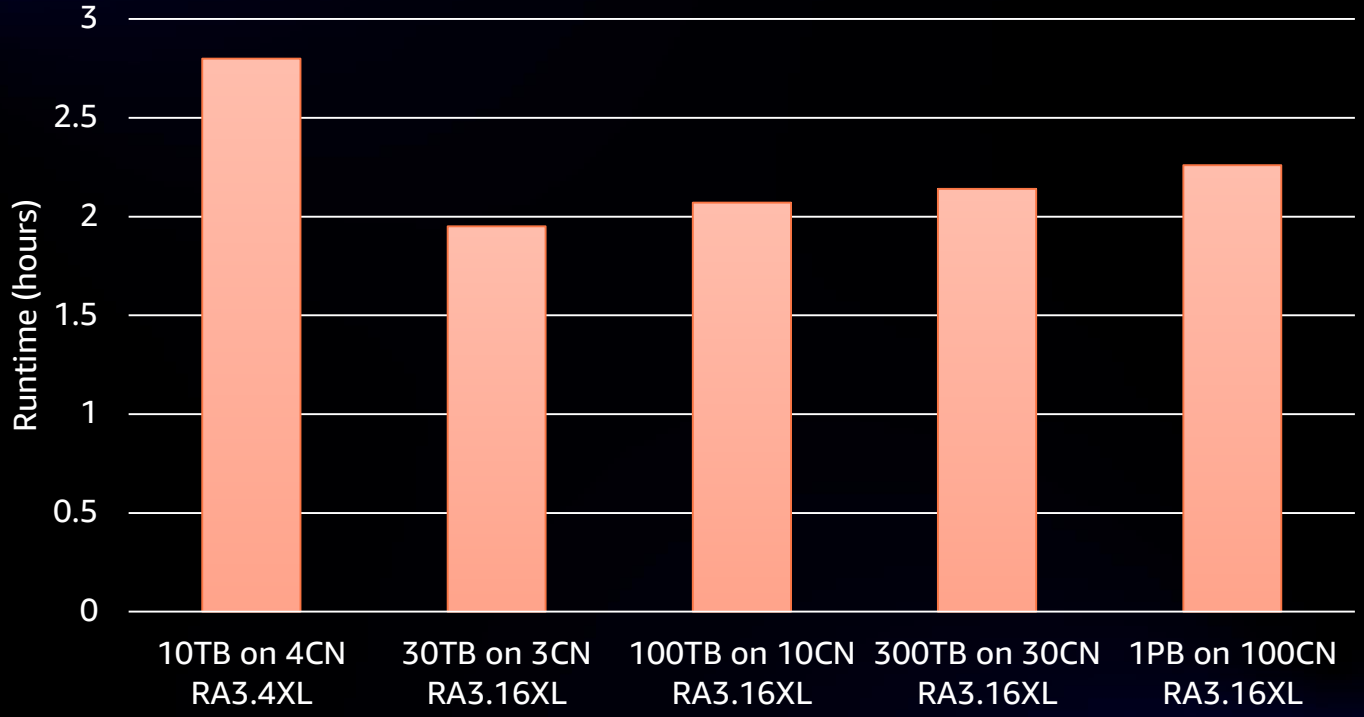
* Cloud DW benchmark is based on TPC-DS (v2.13) with no query modifications done

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Linear scalability to the PB scale

Cloud DW benchmark*

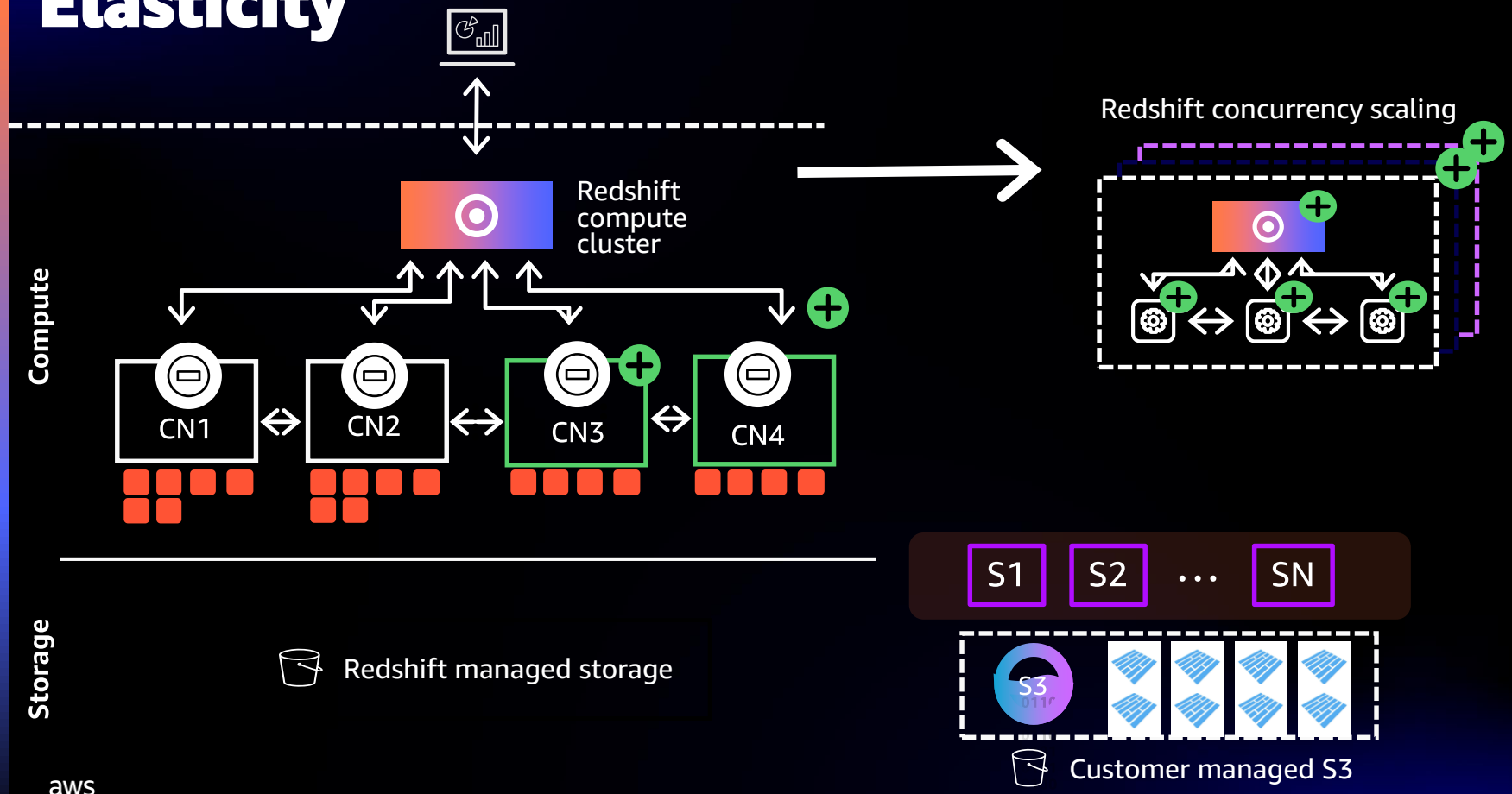


* Cloud DW benchmark is based on TPC-DS (v2.13) with no query modifications done

aws

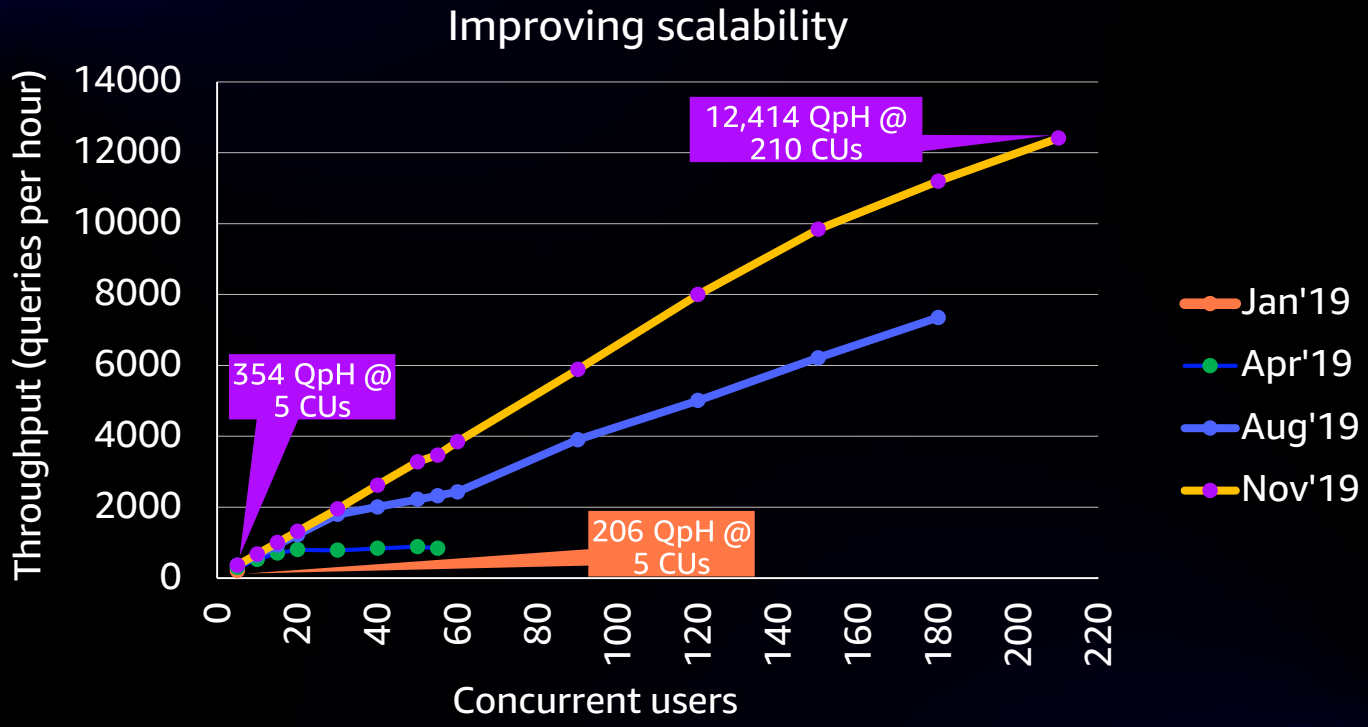
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Elasticity



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Compute elasticity improvements



Cloud DW derived from TPC-DS 3TB

4-node DC2.8XL

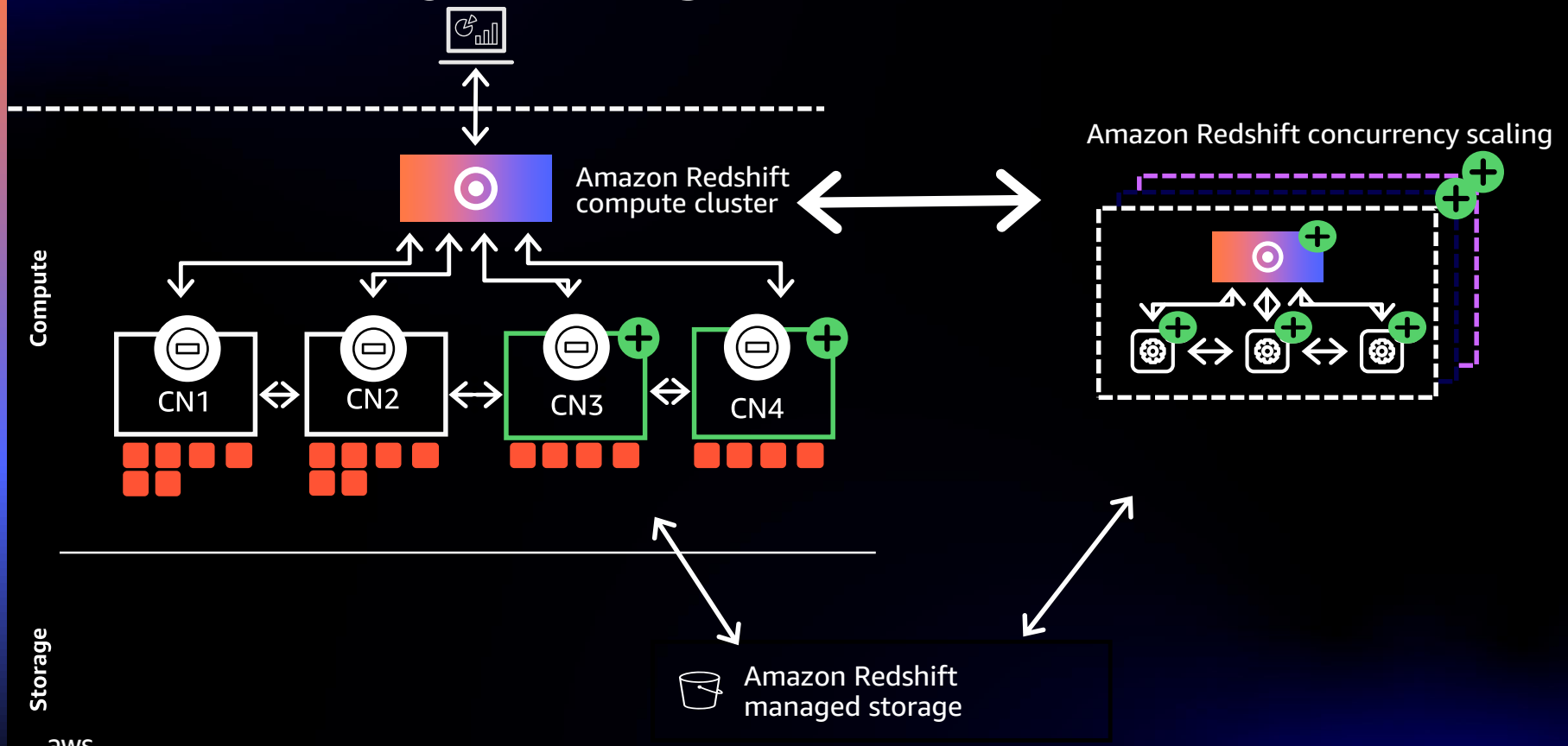
>60x improvement in concurrency

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

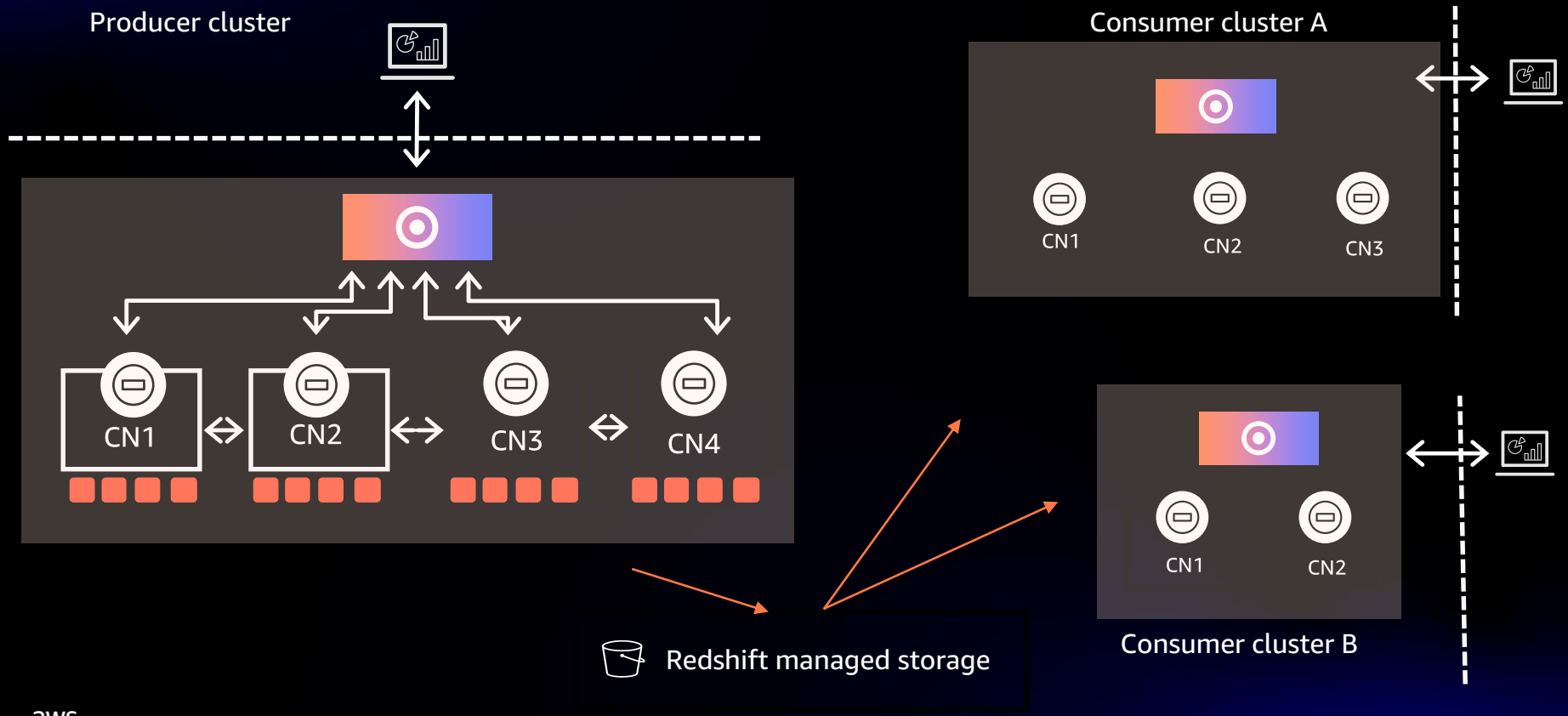
Concurrency scaling writes

Introducing!



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

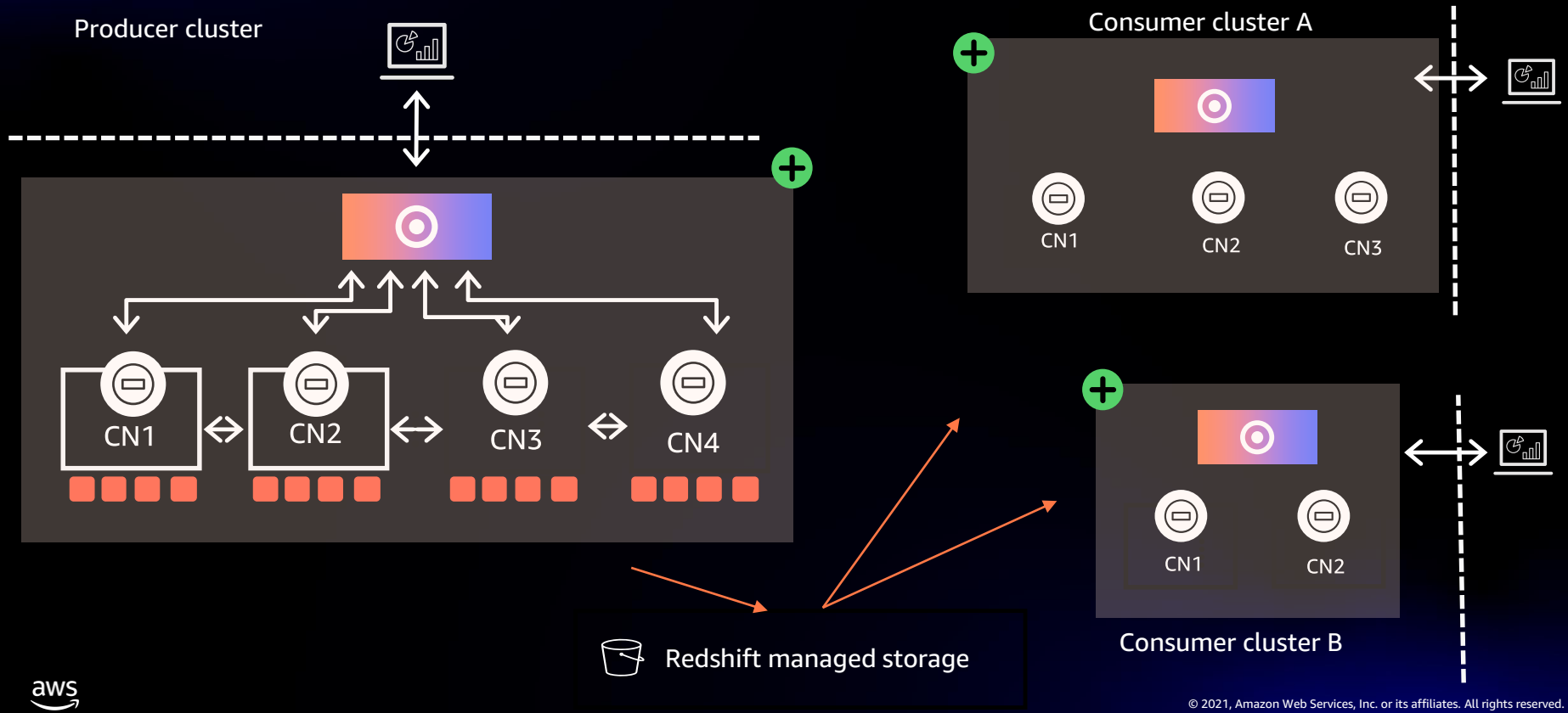
Data sharing



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Data sharing and concurrency scaling

Introducing!



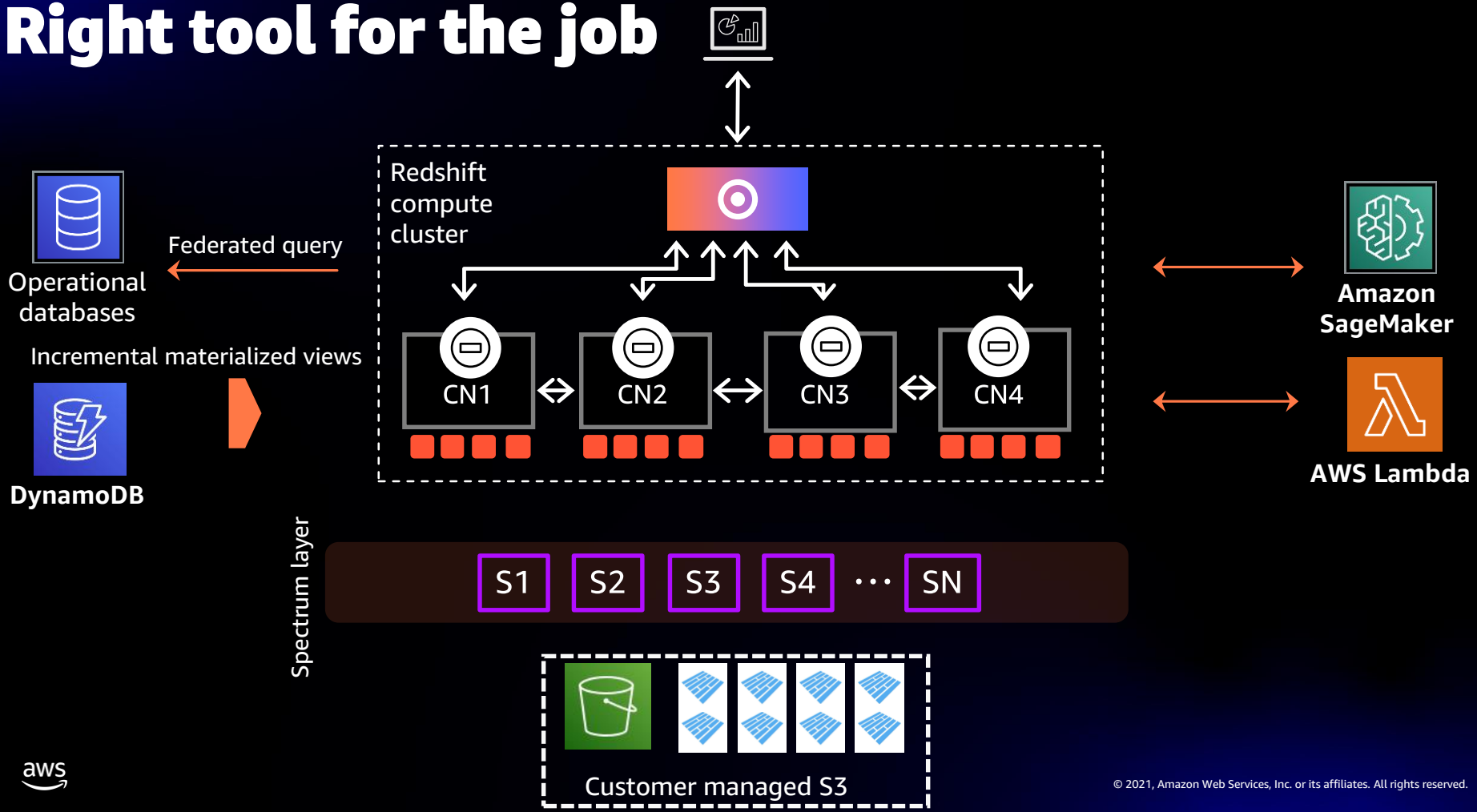
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The evolution of Amazon Redshift

0. Security and availability
1. Performance & elasticity
2. Integration with the AWS environment
3. Autonomics
4. Serverless

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Right tool for the job



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The evolution of Amazon Redshift

0. Security and availability
1. Performance & elasticity
2. Integration with the AWS environment
3. Autonomics
4. Serverless

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automated performance tuning

UPDATED

ML-BASED OPTIMIZATIONS TO GET STARTED EASILY AND GET THE FASTEST PERFORMANCE QUICKLY



Automatic vacuum delete



ATO: Automatic distribution keys



ATO: Automatic sort keys



Auto workload manager



Automatic table sort



ATO: Automatic column compression encoding



Auto Analyze



Auto refresh & rewrite Materialized Views

Automates physical data design and optimization

Optimizes for peak performance as data and workloads scale

Leverages machine learning to adapt to shifting workloads

aws

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automated materialized views

NEW PREVIEW

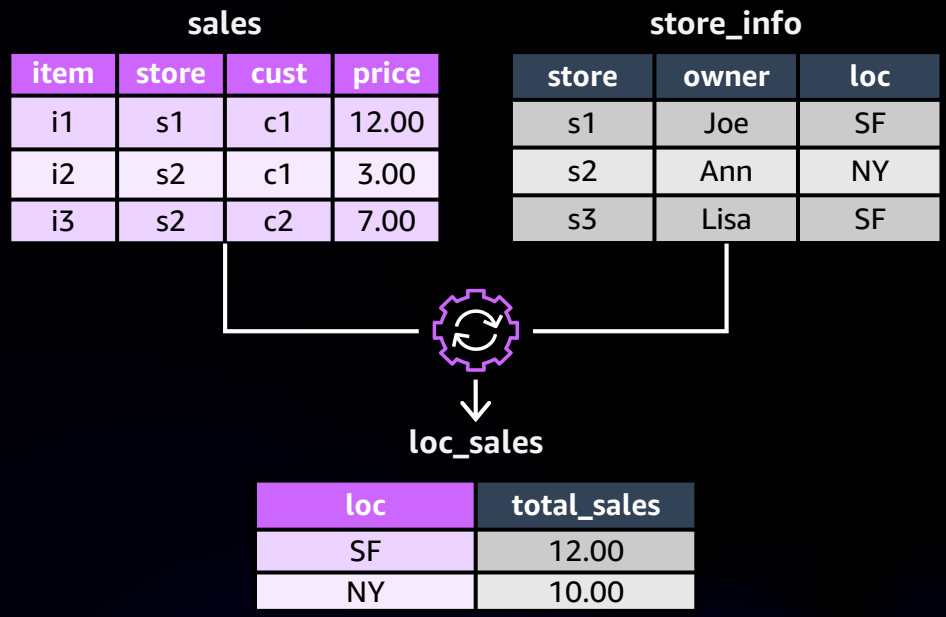
Automated creation and deletion of materialized views, with incremental refresh

Up to orders of magnitude query performance improvement

Continual monitoring to identify queries that will benefit from having a MV and automatically create

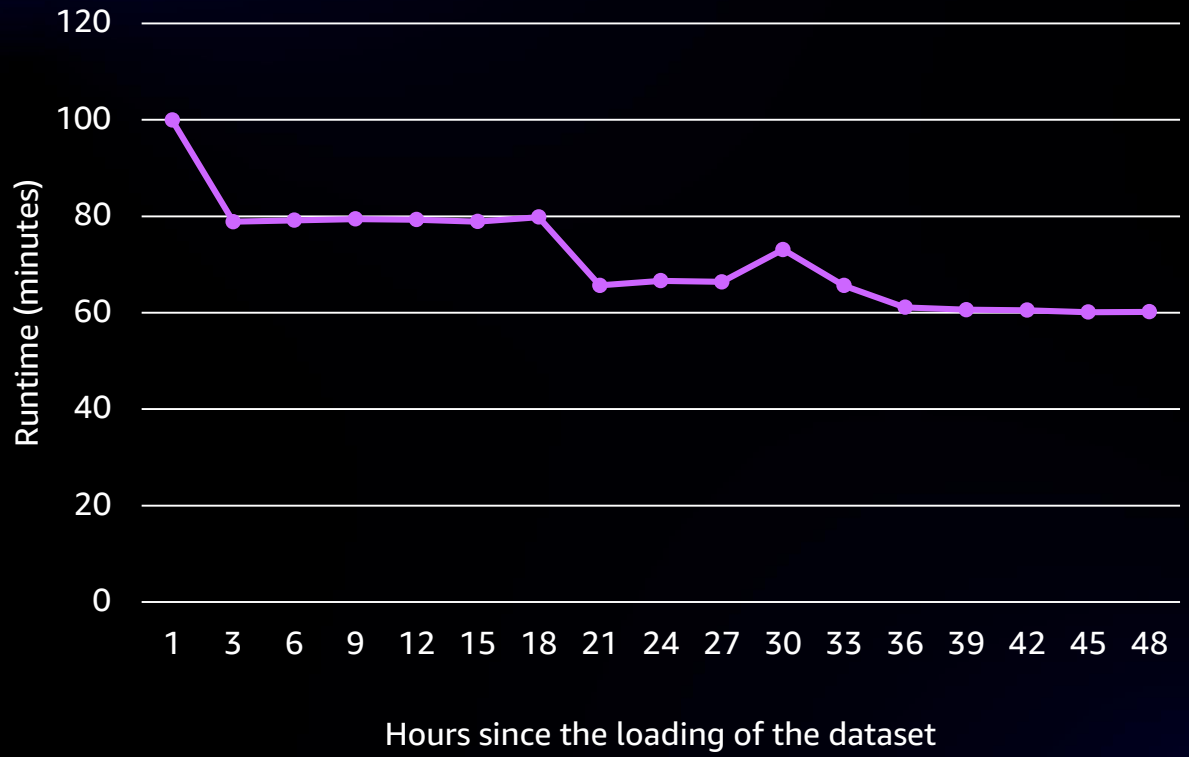
Automated query rewrite to leverage MV

Automatically deletes MVs that are no longer useful



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Automatic table optimization



Cloud DW derived from TPC-H 30TB

5-node RA3.16XL

aws

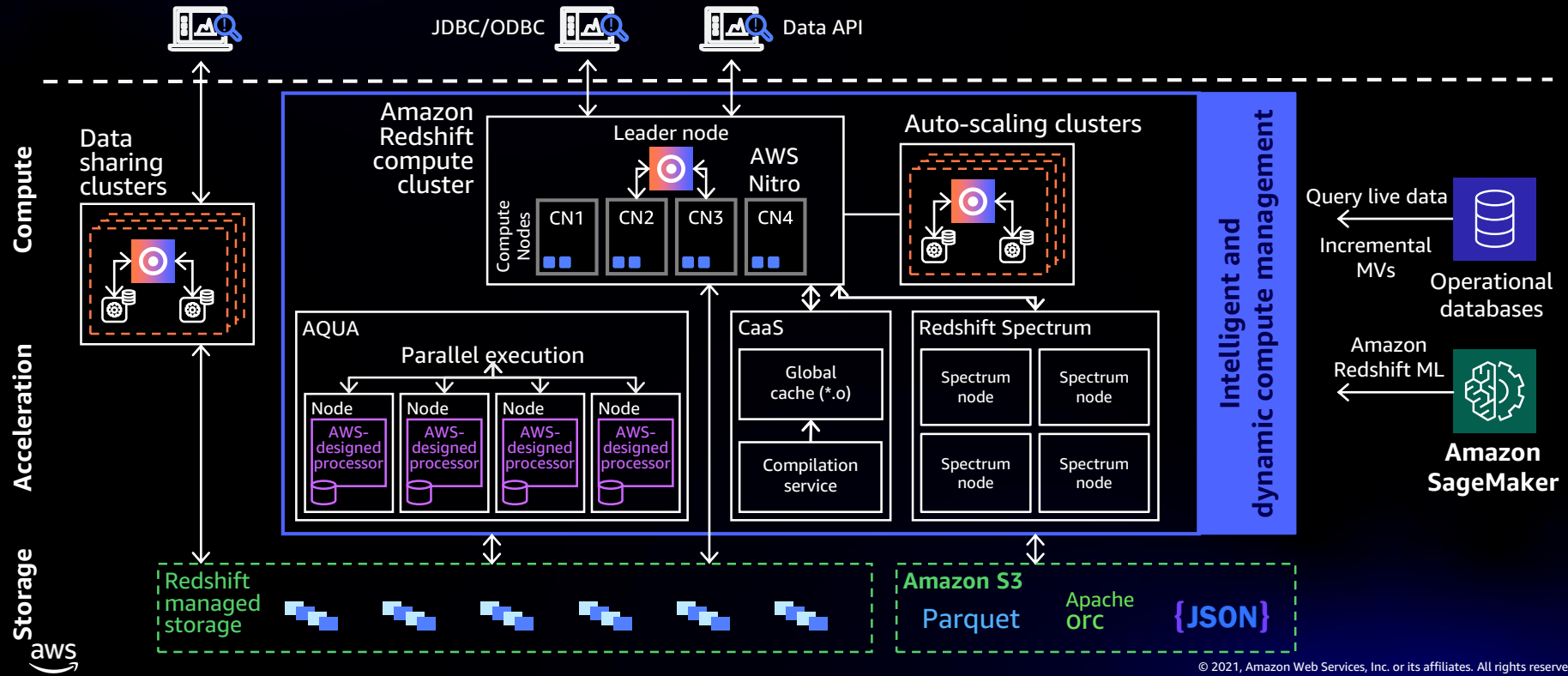
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The evolution of Amazon Redshift

0. Security and availability
1. Performance & elasticity
2. Integration with the AWS environment
3. Autonomics
4. Serverless

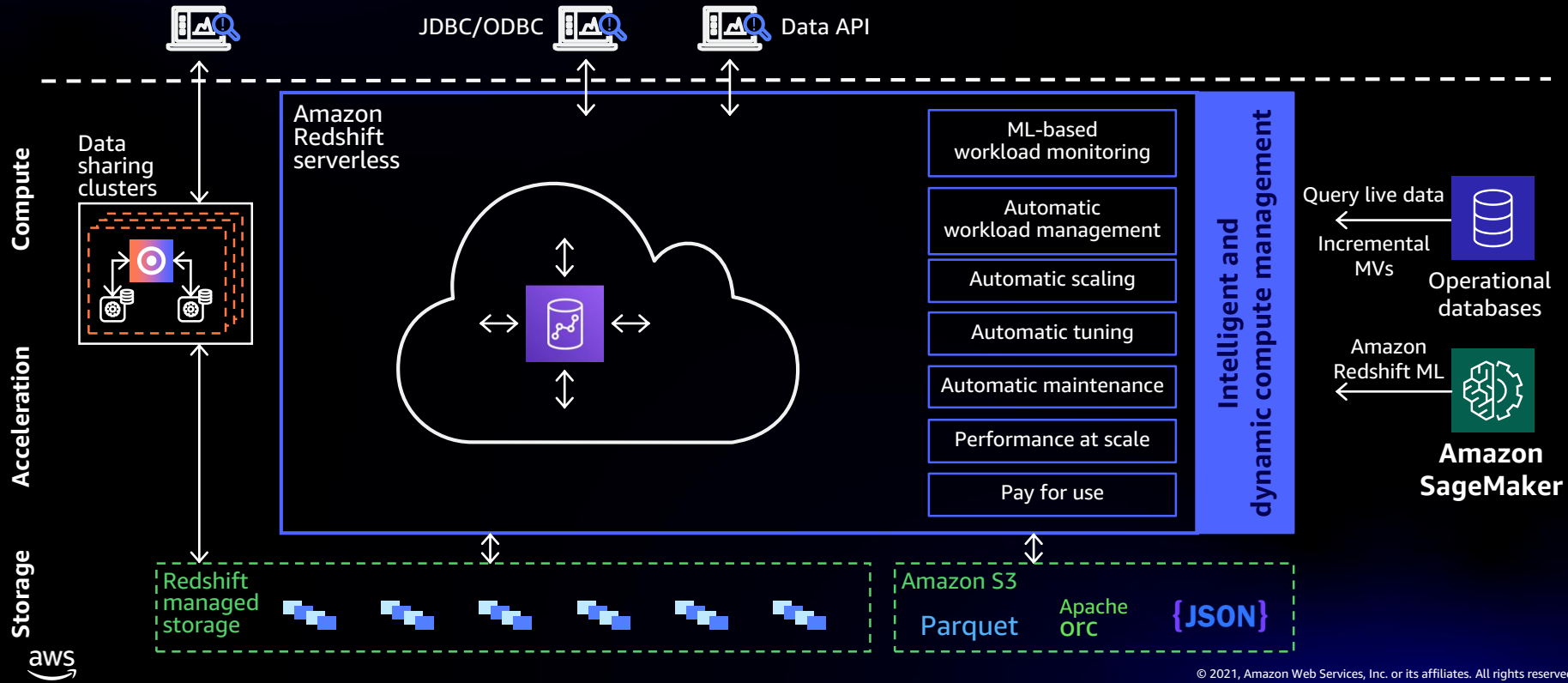
© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Intelligent and dynamic compute management evolution



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

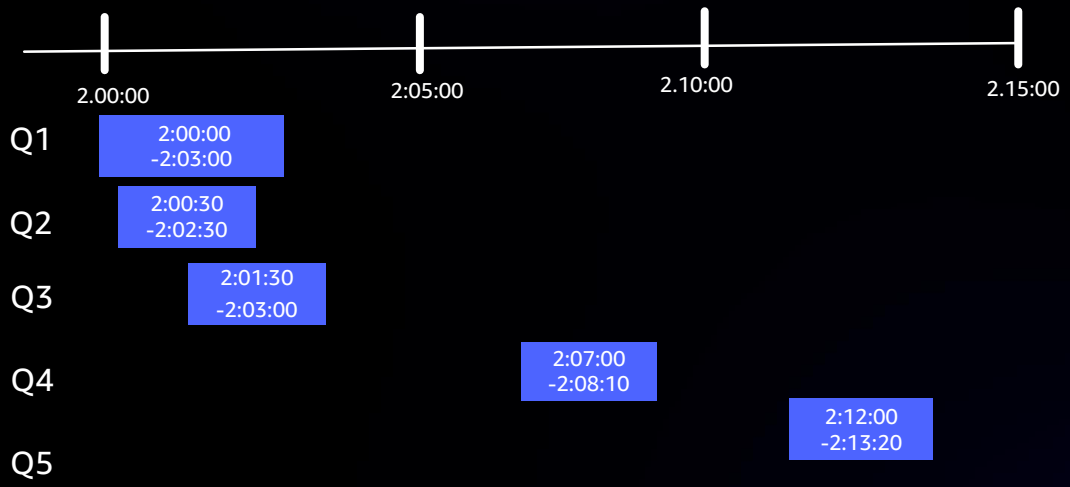
Intelligent and dynamic compute management evolution



© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Pay for use

Pay for the compute capacity only for the workload duration (metered on a per-second basis)



Billing duration	Query execution time
@2:03	3 minutes (for Q1, Q2, Q3)
@2:09	1 minute and 10 seconds (for Q4)
@2:14	1 minute and 20 seconds (for Q1)
Total charges	5 minutes and 30 seconds

No charges for idleness

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

The evolution of Amazon Redshift



- Tens of thousands of customers process exabytes of data daily with Amazon Redshift
- Continuous focus on performance and scalability
- Ability to elastically support tens of PBs of data and 1,000s of users
- Tight integration with the broad AWS environment
- Autonomics that make Amazon Redshift easy to use
- Serverless experience

And it is still Day 1

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Thank you!

Steven Cochrane
steven.cochrane@ppb.com

Ippokratis Pandis
ippo@amazon.com

© 2021, Amazon Web Services, Inc. or its affiliates. All rights reserved.