

AWS
re:Invent



ANT205

Getting the most out of Amazon Redshift automation

Paul Lappas

Principal Product Manager, Amazon Redshift
AWS

What you will learn

- Overview of Amazon Redshift's self-tuning capabilities
- How to use the Amazon Redshift Advisor
- How common use cases benefit from self-tuning features
- Demo: Automatic Table Optimization

Amazon Redshift

THE MOST WIDELY USED CLOUD DATA WAREHOUSE, WITH TENS OF THOUSANDS OF CUSTOMERS

ANALYZE ALL YOUR DATA



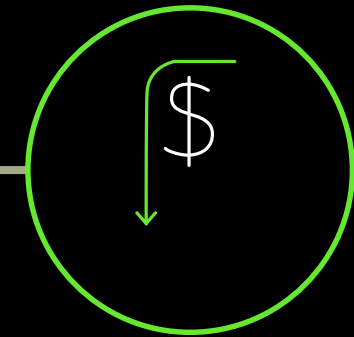
Take a **lake house approach** by analyzing all your data across your data warehouse, your Amazon S3 data lake, and operational databases with consistent security and governance policies

PERFORMANCE AT ANY SCALE



Get up to **3x better price performance** than other cloud data warehouses with a **self-tuning** system, and boost queries up to **10x with AQUA**

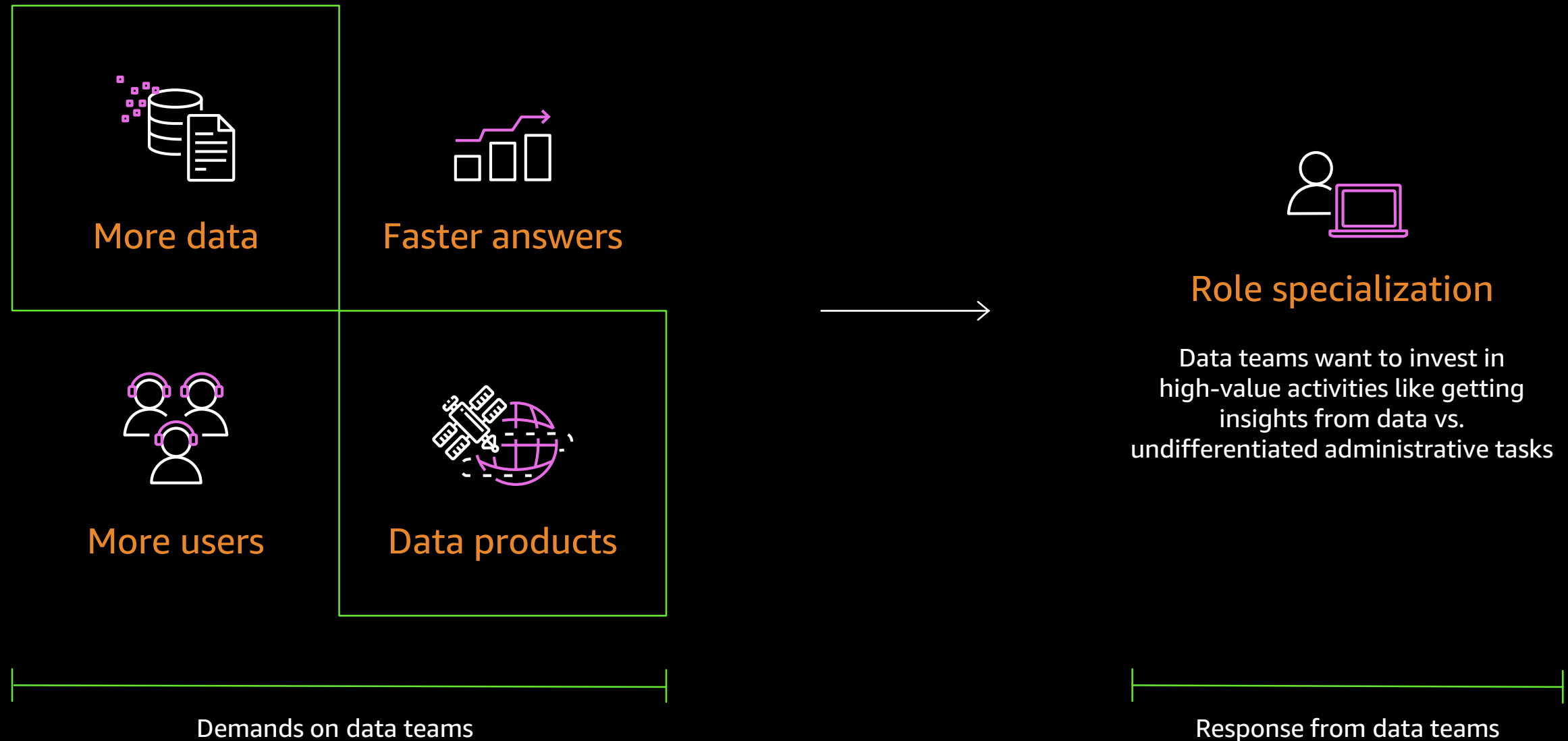
LOWER YOUR COSTS



Start small and pay only for what you use with **predictable** monthly costs; Amazon Redshift is at least **50% less expensive** than other cloud data warehouses

Trends driving demand for DW automation

THE ROLE OF THE DBA IS SHIFTING TO FOCUS ON GETTING VALUE FROM DATA

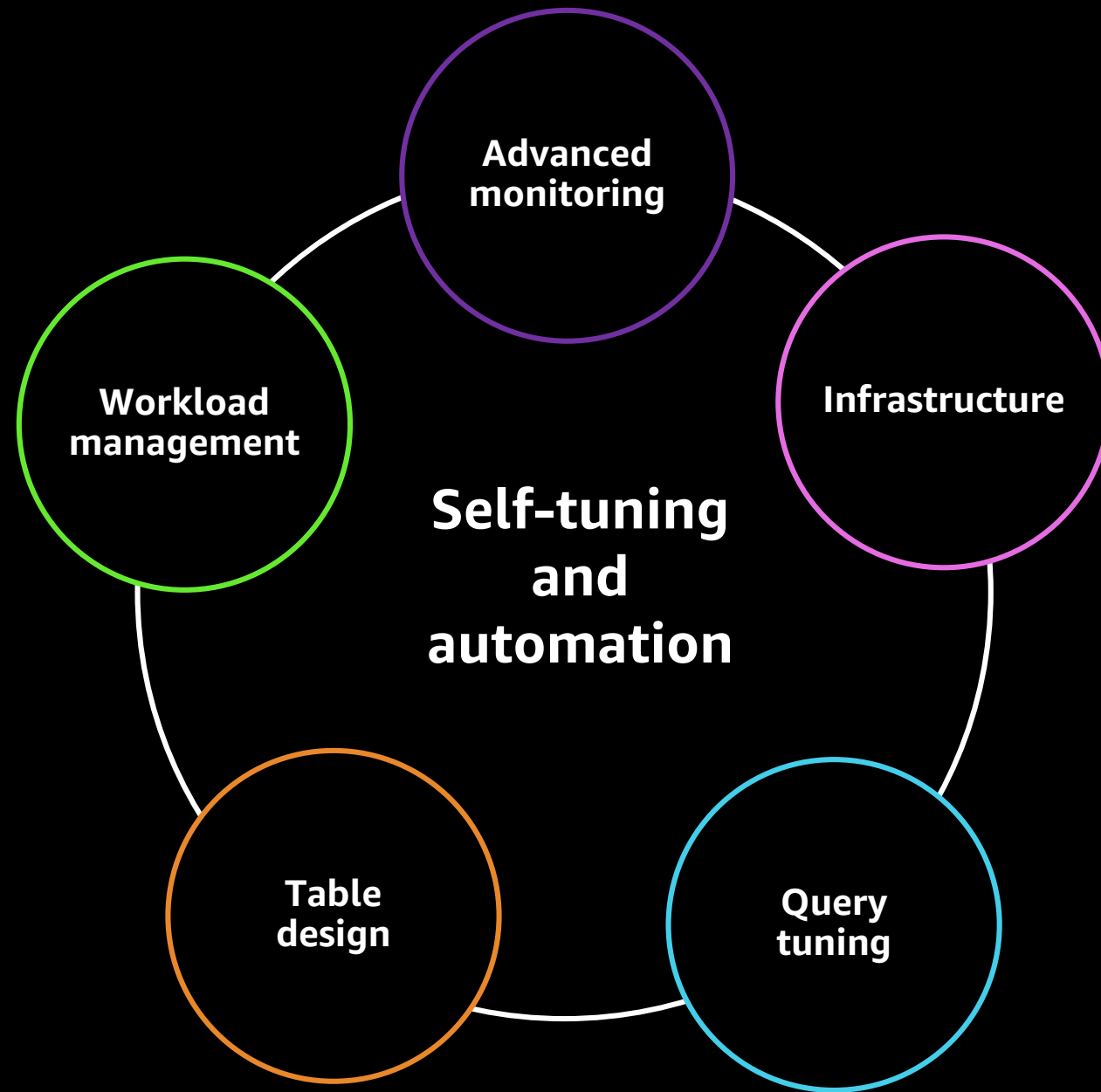


Amazon Redshift's self-tuning capabilities



Easier to get insight

WITH A FULLY MANAGED AND SELF-TUNING DATA WAREHOUSE



Amazon Redshift automates performance tuning

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

Automates physical data design and optimization

Optimizes for peak performance as data and workloads scale

Leverages machine learning to adapt to shifting workloads

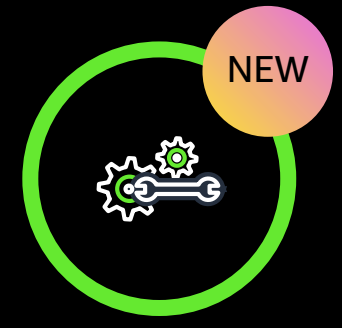
Automated performance tuning



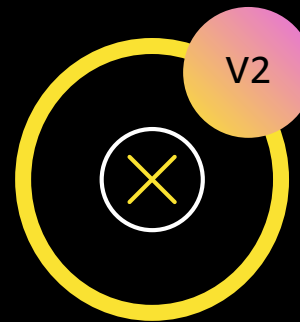
Automatic vacuum



Automatic distribution keys



Automatic sort keys



Workload Manager



Automatic table sort

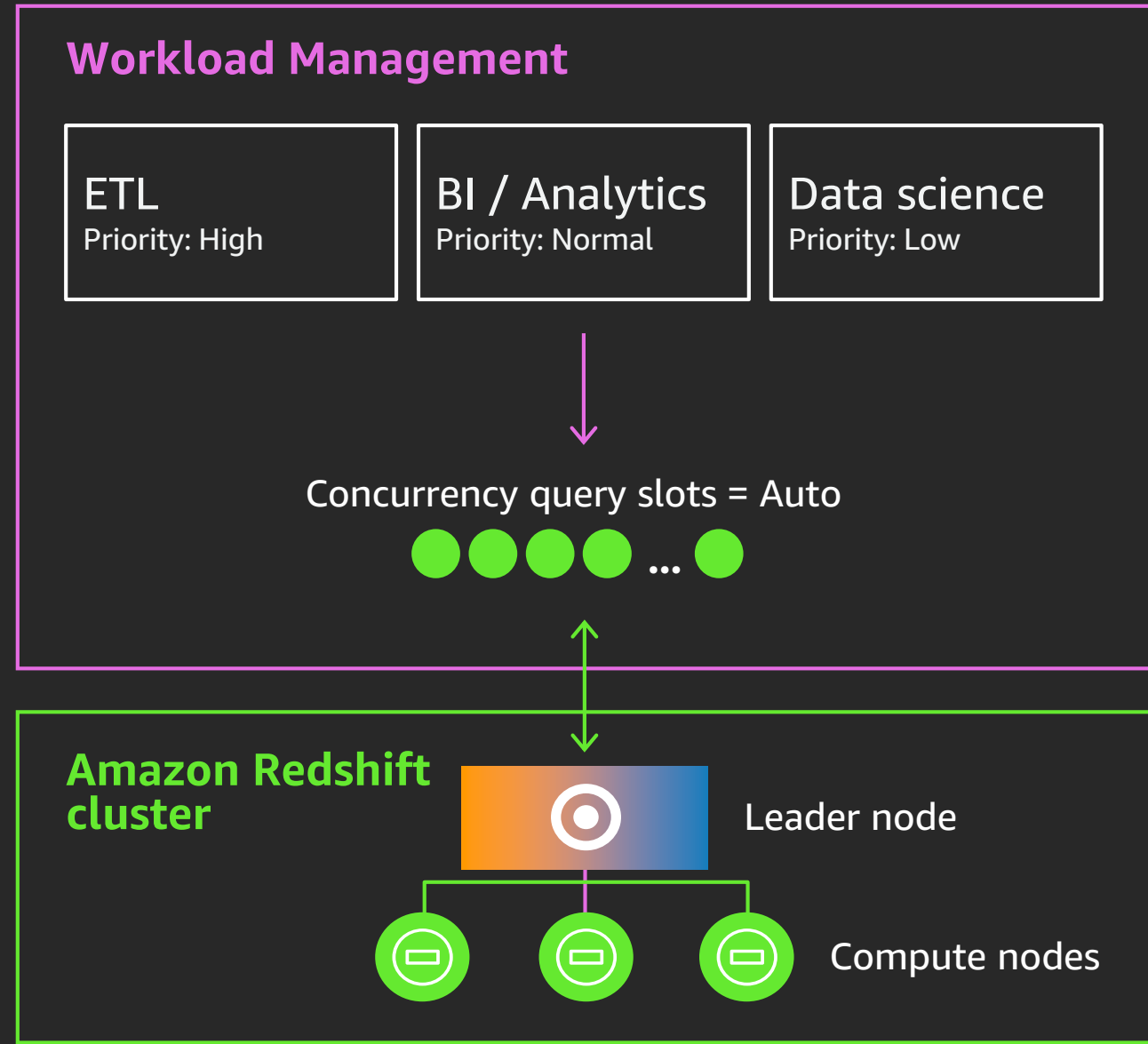


MV auto-refresh and rewrite

Automatic WLM and query priorities

OPTIMIZE RESOURCES FOR MAXIMUM QUERY THROUGHPUT AND CONSISTENT PERFORMANCE

- ✓ Ensure that high-priority workloads get preferential treatment
- ✓ Prevent expensive queries from consuming system resources
- ✓ Uses machine learning to optimize query throughput
- ✓ Ensures highly efficient resource utilization for maximum ROI
- ✓ Fair sharing of resources



Materialized views auto-refresh and query rewrite

SPEED UP QUERY PERFORMANCE BY ORDERS OF MAGNITUDE WITH PRECOMPUTED RESULTS

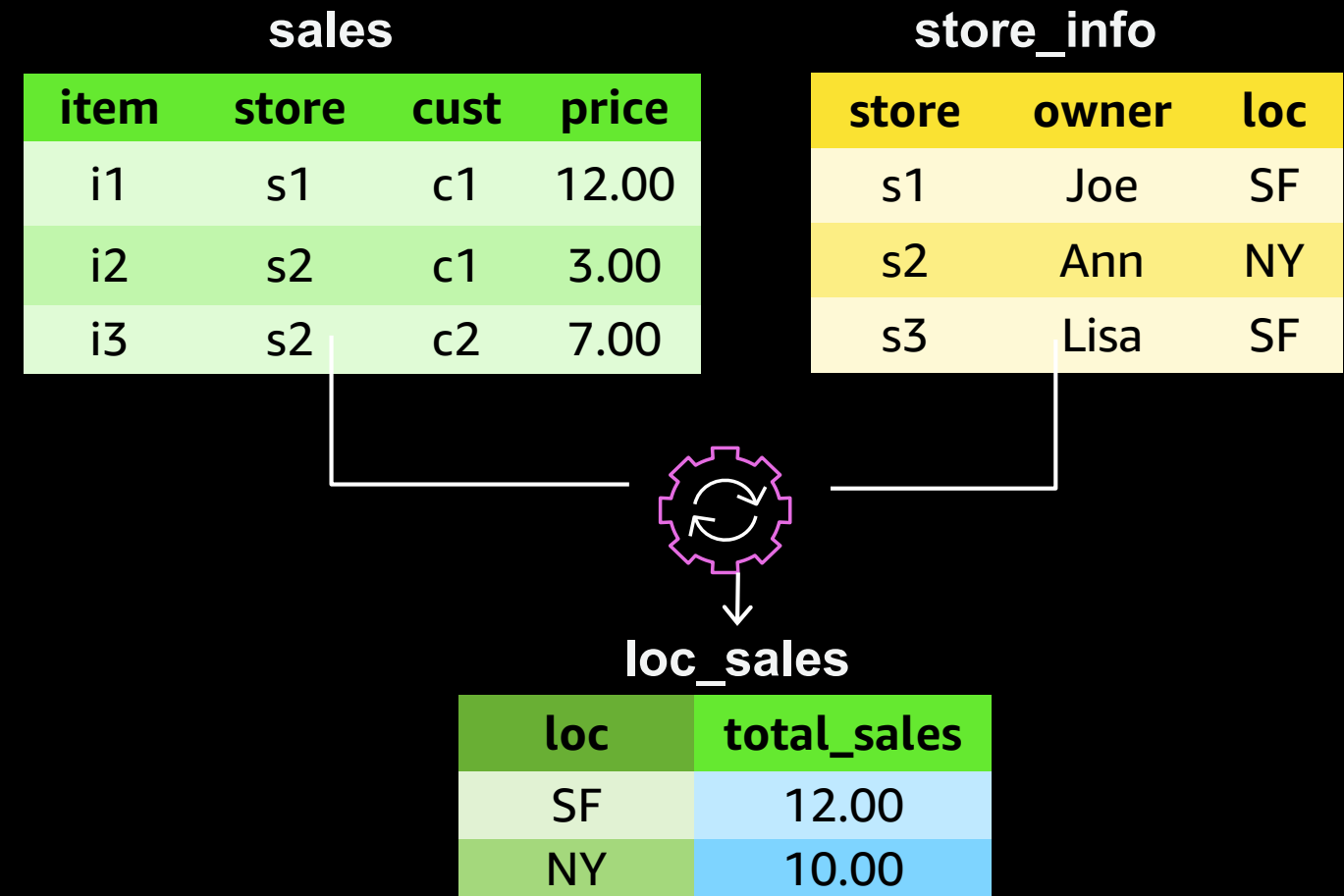
Simplify and accelerate iterative and predictable workloads, such as ETL, BI/dashboarding queries

MVs can be based on one or more Amazon Redshift tables or external tables (Spectrum, Federated)

Efficient incremental maintenance

Scheduled, automatic, or manually timed refresh

Amazon Redshift auto query rewrite optimizes queries by replacing native tables with materialized views



"The Amazon Redshift materialized view auto query rewrite feature reduced dashboard load times from 8 minutes to just 500 ms. The best part is that this is completely transparent for Tableau and the business user."

Arman Nasrollahi, Home24

Automatic Table Optimization

INTRODUCING AN AUTO-PROPERTY FOR TABLES

Continuously scans workload patterns

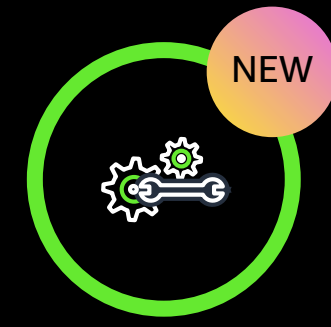
Automatically adjusts sort and distribution styles to account for changes in workload

Can be enabled or disabled per table

Audit trail of changes



Automatic
distribution keys



Automatic
sort keys

“When we tested ATO in our development environment, the performance of our queries was 25% faster than our production workload not using ATO, without requiring any additional effort by our administrators.”

Nishesh Aggarwal, Enterprise Architecture Manager, ZS Associates

Distribution style optimizations

SPEED UP QUERY PERFORMANCE WHEN JOINING TABLES

Tables start off as EVEN or ALL based on size

Optimizes data distribution to minimize network communication among nodes

Redistribution operations are done in the background with minimal user impact

Join multi-graph

Query workload

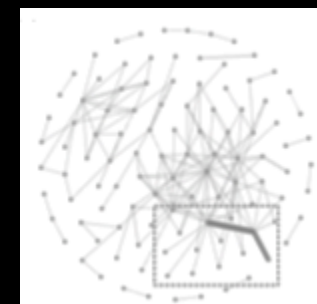
...C join E on C.c = E.c...

...B join D on B.b = D.d...

...B join D on B.b₁ = D.d₁...

...B join D on B.b₁ = D.d₁...

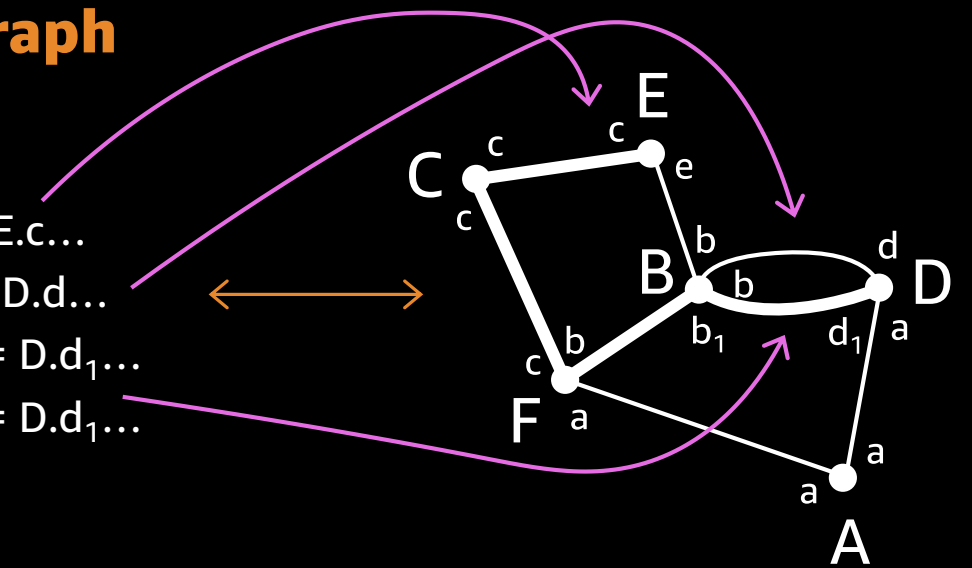
— weight = 2
— weight = 1



Join multi-graph of a real Amazon Redshift customer



Most beneficial columns across two tables



Sort key optimizations

SPEED UP QUERY PERFORMANCE WHEN ORDERING AND FILTERING COLUMNS

Tables start off with no sort key

Monitors the workload for the most selective column that is used frequently as a filter or join predicate

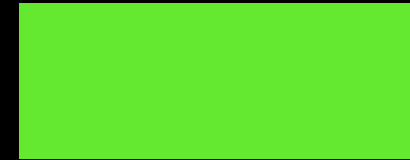
After sort key is applied, table sorting is handled by automatic table sort

Sorted by ???



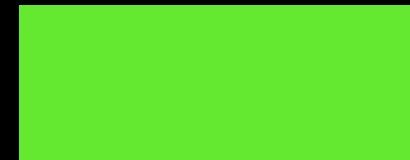
MIN: 01-MAR-2019

MAX: 31-MAR-2019



MIN: 01-APRIL-2019

MAX: 30-APRIL-2019



MIN: 01-MAY-2019

MAX: 31-MAY-2019

Most beneficial column



Automatic table sort

MAINTAIN SORT ORDER OF THE DATA IN AMAZON REDSHIFT TABLES TO CONTINUOUSLY OPTIMIZE QUERY PERFORMANCE

Eliminates DBA-managed sorting

Automates sort order maintenance

Maximizes efficient use of compute

Leverages machine learning

No impact on performance or accessibility

Sorted by date



MIN: 01-MAR-2019

MAX: 31-MAR-2019



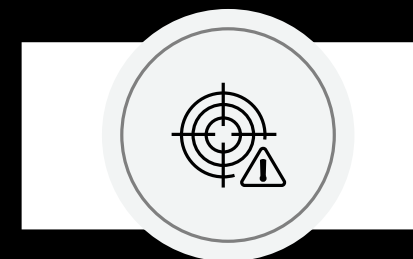
MIN: 01-APRIL-2019

MAX: 30-APRIL-2019



MIN: 01-MAY-2019

MAX: 31-MAY-2019



MIN: 01-JUNE-2019

MAX: 30-JUNE-2019

Most accessed partition

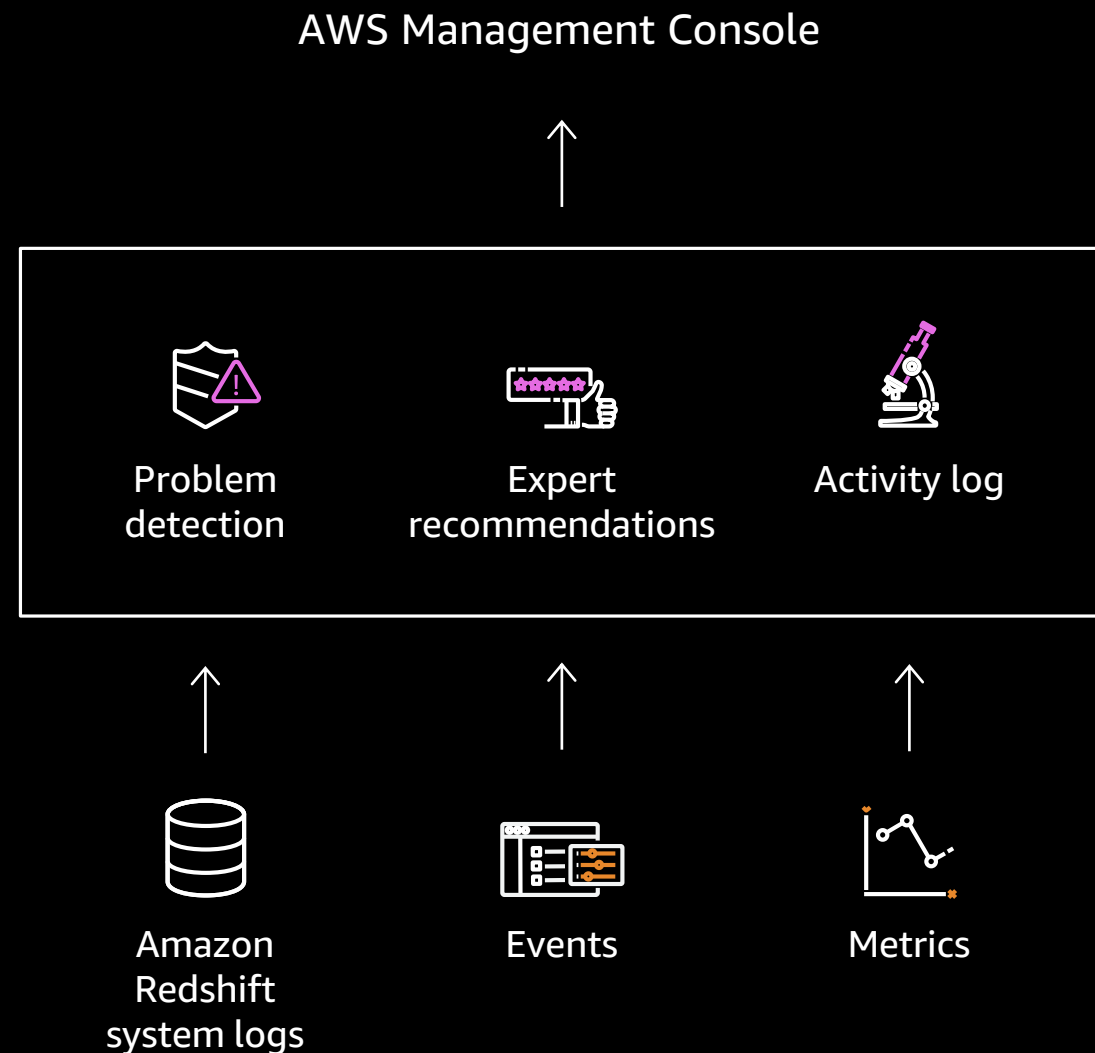


Amazon Redshift Advisor



Amazon Redshift Advisor

EXPERT RECOMMENDATIONS PERSONALIZED FOR YOUR CLUSTER



Continuously monitor millions of data points to detect, analyze, and surface issues before they impact your users

Expert health checks

Actionable recommendations to optimize cost and performance

Machine learning powered

Types of Advisor recommendations

FUTURE AUTONOMOUS CAPABILITIES ARE FIRST SURFACED AS RECOMMENDATIONS

Data ingestion



Amazon Simple Storage Service (Amazon S3) compression
Skip compression analysis
Split objects

Query tuning



Reallocate memory*
Enable short query acceleration*

Table design



Sort key recommendation*
Distribution style recommendation*
Vacuum tables*
Analyze tables*
Column encoding

Cost



Delete unused clusters
Compress table data

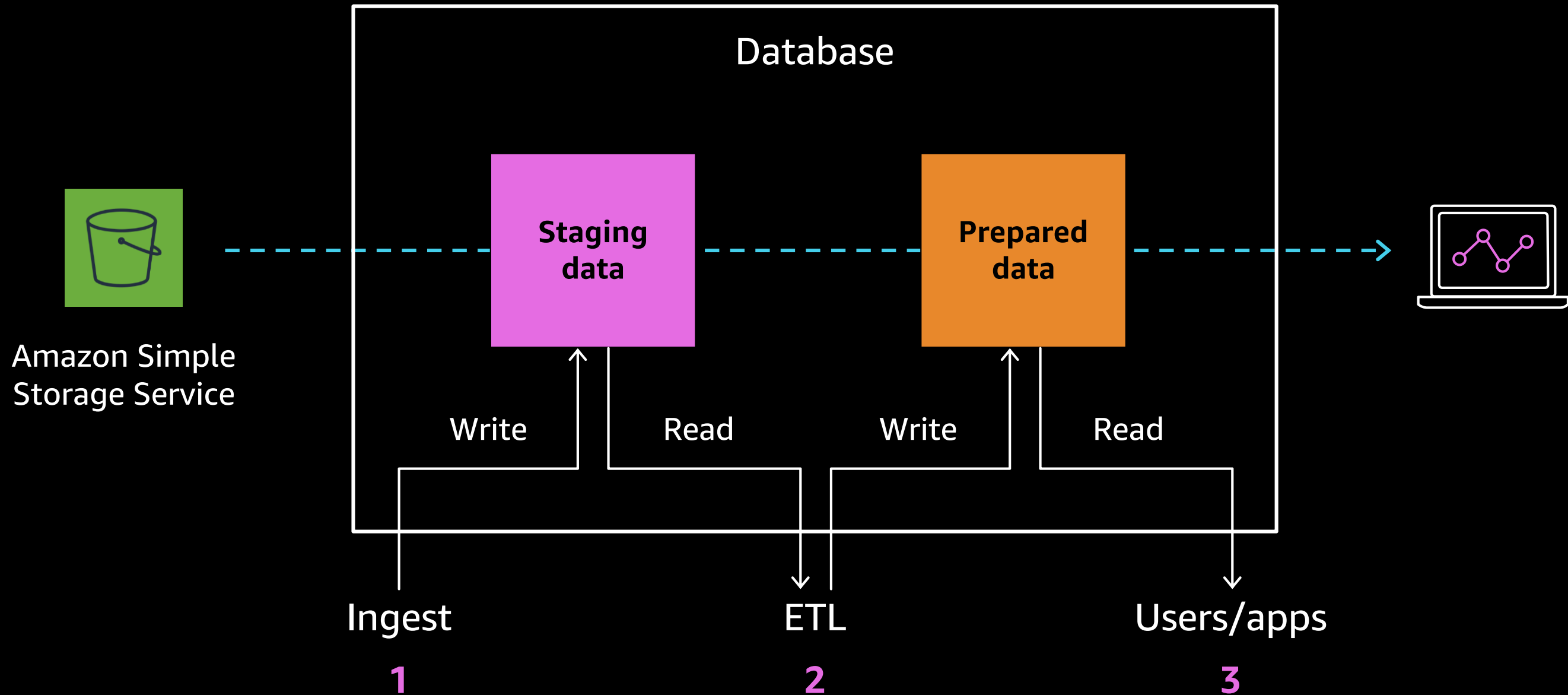
*Existing self-tuning capability

Use cases



Reference architecture

COMMON DATA ARCHITECTURE FOR AMAZON REDSHIFT



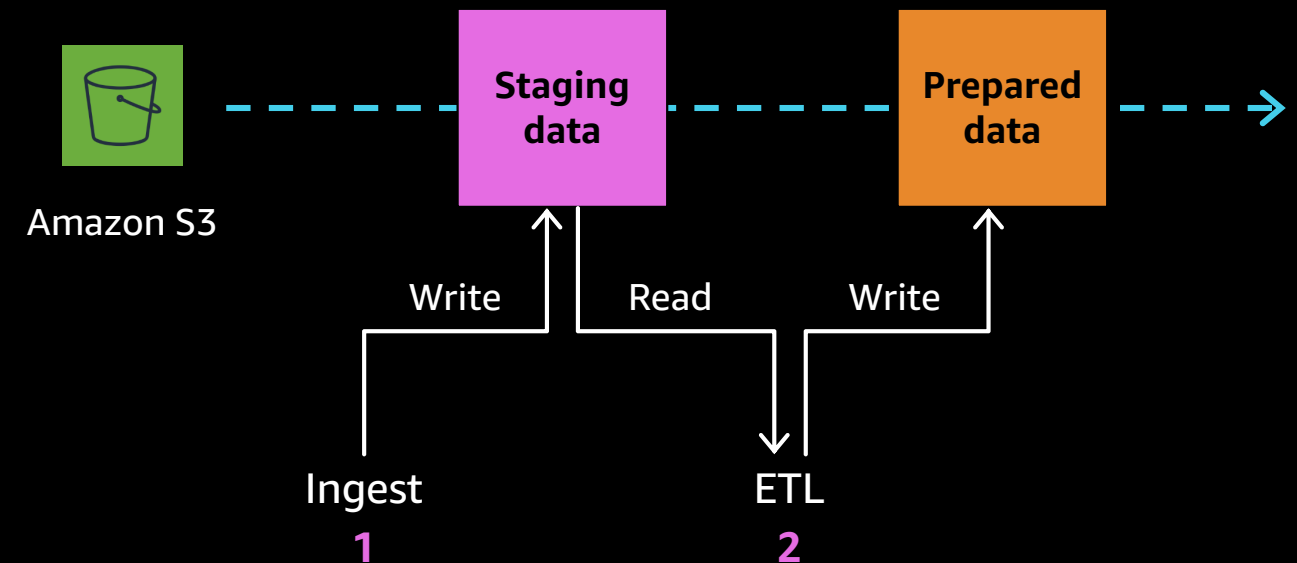
Use case: Operational analytics

HOW AUTOMATION MAKES IT EASIER AND FASTER

Automatic Table Optimization will set the best sort and dist keys

Set "medium" WLM query priorities for ETL queries

Automated sort and vacuum maintains table data as they are modified



Use case: Business intelligence and reporting

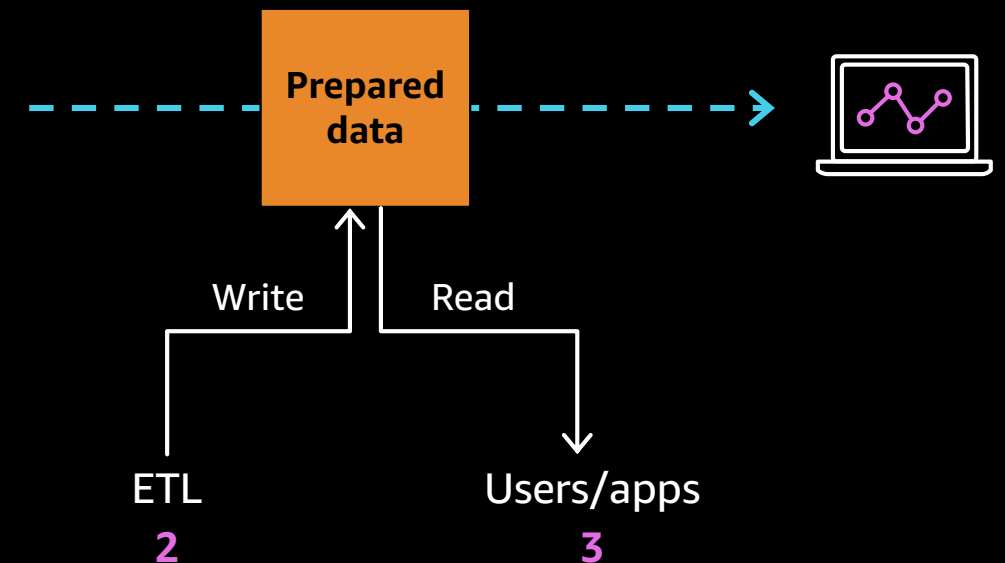
HOW AUTOMATION MAKES IT EASIER AND FASTER

Automatic Table Optimization will set the best sort and dist keys

Configure WLM query monitoring rules (QMR) to abort poorly written queries

Set "high" WLM query priority for user queries

Create materialized views to speed up common queries



Demo: Automatic Table Optimization



Learn more

https://docs.aws.amazon.com/redshift/latest/dg/t_Creating_tables.html

<https://docs.aws.amazon.com/redshift/latest/dg/advisor.html>

<https://docs.aws.amazon.com/redshift/latest/dg/automatic-wlm.html>

Next steps

Already using Amazon Redshift?

Ask your account team for a free Amazon Redshift optimization session

Learn more about what's new

<https://aws.amazon.com/redshift/whats-new/>

Getting started with Amazon Redshift?

Ask your account team for a 10-minute demo

Learn more about starting a proof of concept

<https://docs.aws.amazon.com/redshift/proof-of-concept-playbook.html>

Or request help from AWS

<https://pages.awscloud.com/redshift-proof-of-concept-request>

Thank you!





Please complete
the session survey