Amazon Redshift Tips & Tricks:

Scaling Storage and Compute Resources

Matt Scaer: Principal Data Warehouse Specialist Solution Architect

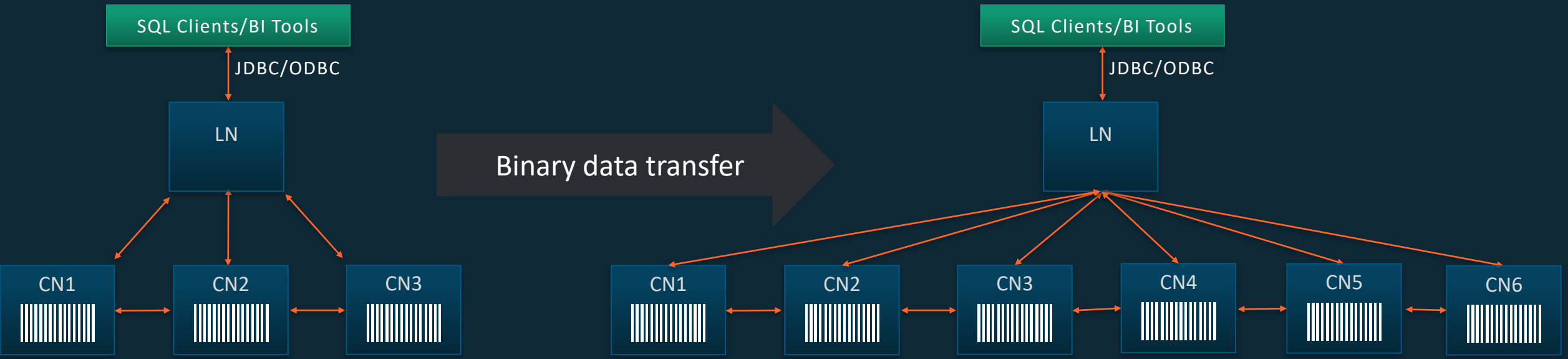
aws

Take-aways from today's session:

- How to scale compute and storage quickly on-demand:
  - Elastic and Classic Resize
- Scale compute automatically to handle thousands of concurrent users and queries:
  - Concurrency Scaling
- Understand how Redshift automatically optimizes resources and performs maintenance tasks to ensure the best performance at the lowest cost as you scale up and down.

aws

Classic resize enables scaling up and down



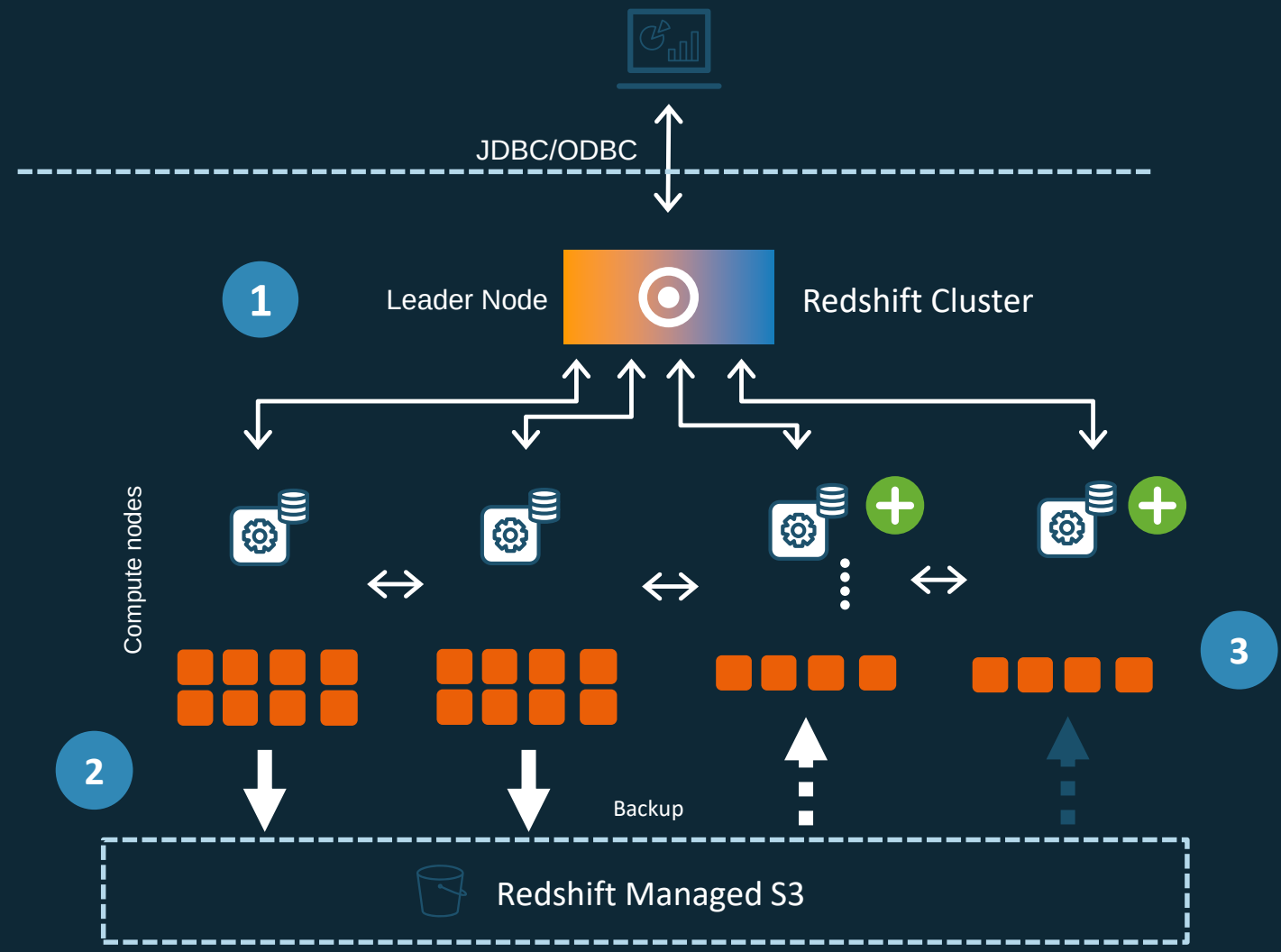
- Complete data is moved and re-distributed
- Cluster is in read-only mode during resize

Redshift Elastic Resize (GA)

Scale your Redshift clusters up and down in minutes to get the optimal performance.

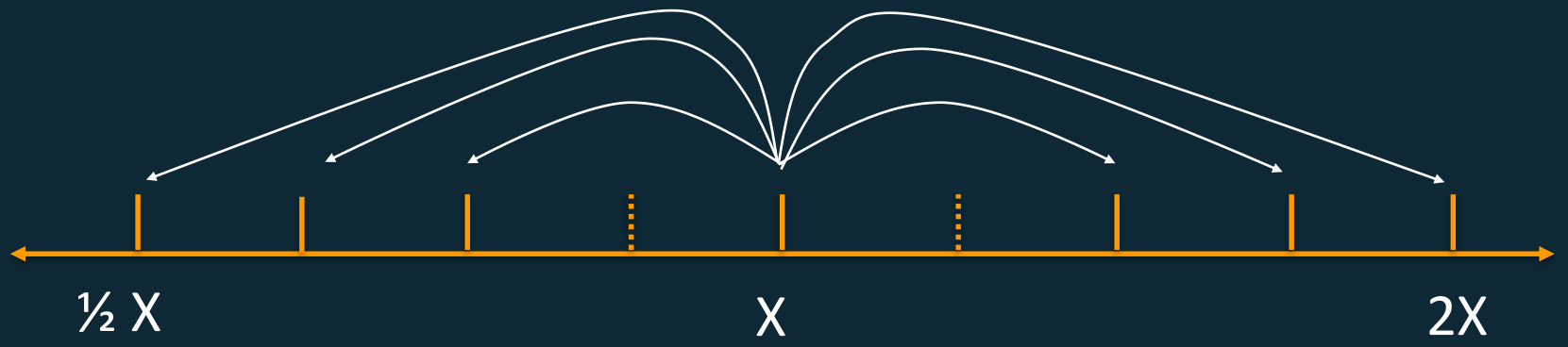
How it works:

1. Redshift updates the snapshot on S3 with the most recent data.
2. New nodes are added (for scaling up) or removed (for scaling down) during this period.
3. The cluster is fully available for read and write queries. Queries that were being held are queued for execution automatically.

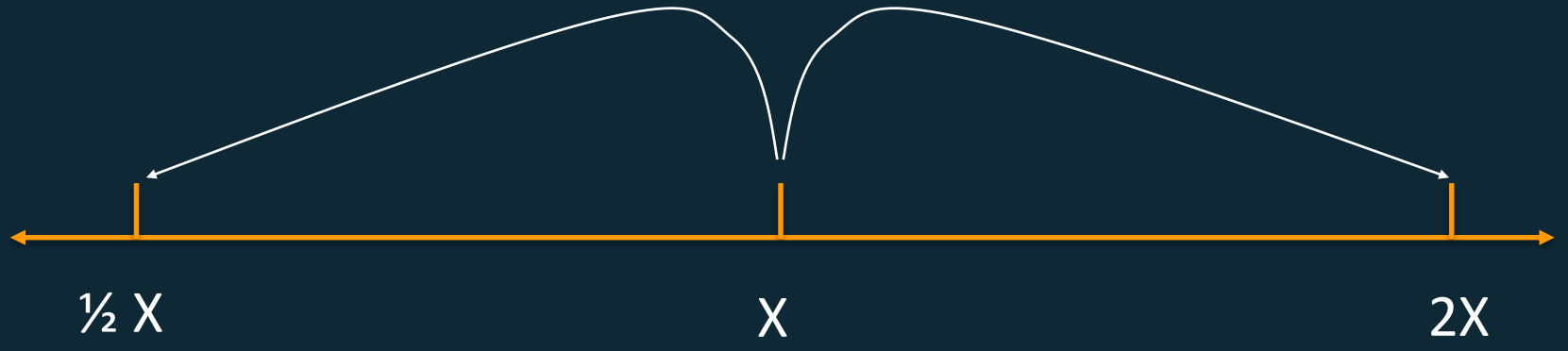


Elastic resize options depend on the node type

DS2.8XL
DC2.8XL

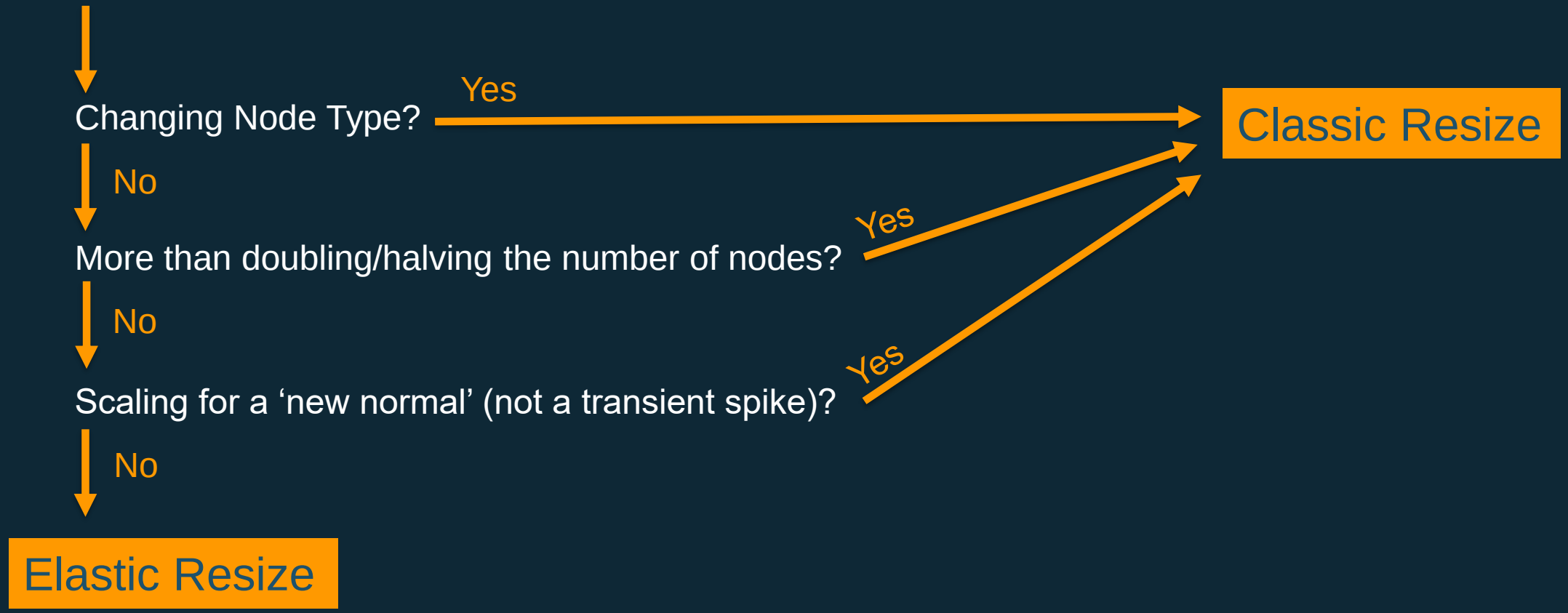


DS2.XL
DC2.L



aws

When to use Elastic vs. Classic resize



Wrap-up: Elastic Resize

With elastic resize you can now add or subtract nodes in minutes

- ✓ Scale compute and storage on-demand
- ✓ Faster query processing
- ✓ Finish large ETL jobs faster
- ✓ Save on-demand cost during off-peak hours

aws

Redshift Elastic Resize FAQs

When to pick Elastic Resize and Classic?

What are the factors to estimate the resize runtime?

How can customers test Elastic Resize offline?

How to track the progress of an Elastic Resize?

What are the current constraints on Elastic Resize?

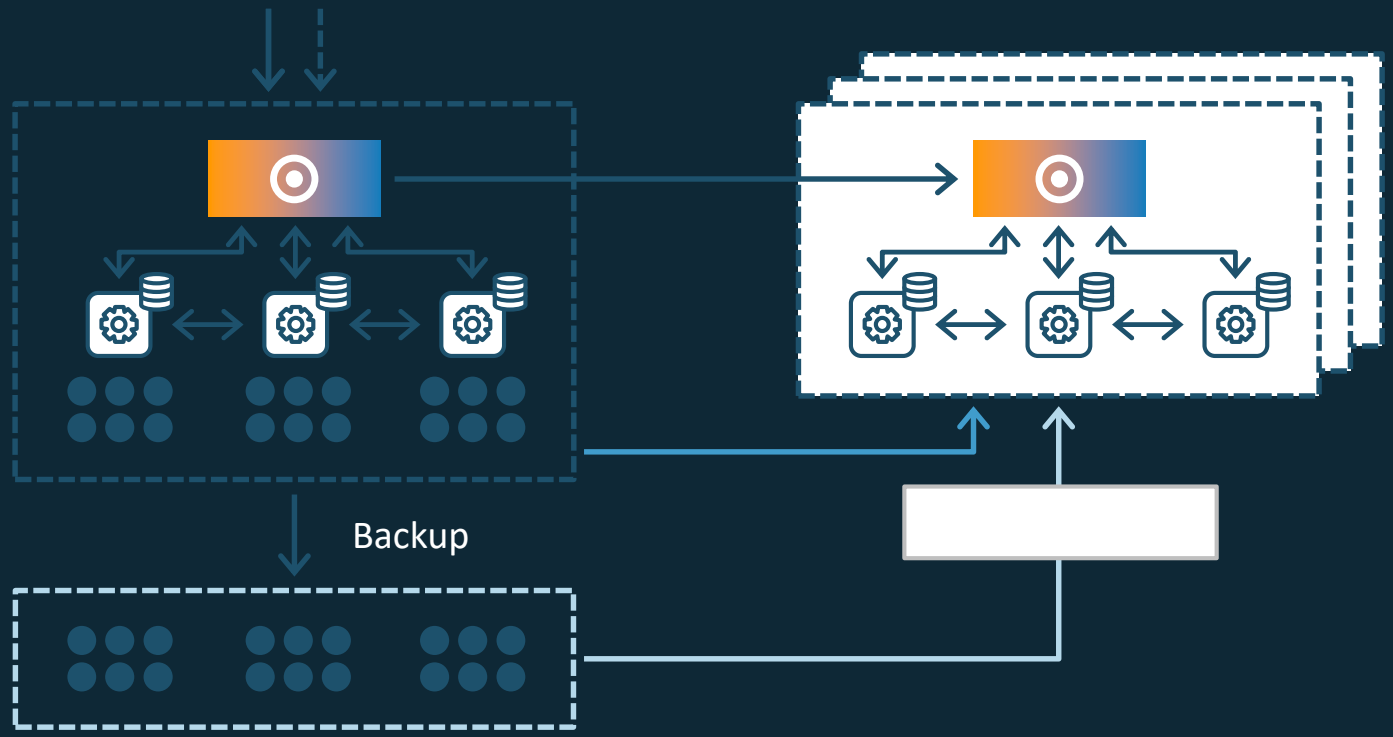
aws

Introducing Concurrency Scaling

Concurrency Scaling

Preview

Redshift automatically adds transient clusters, in seconds, to serve sudden spike in concurrent requests with consistently fast performance.



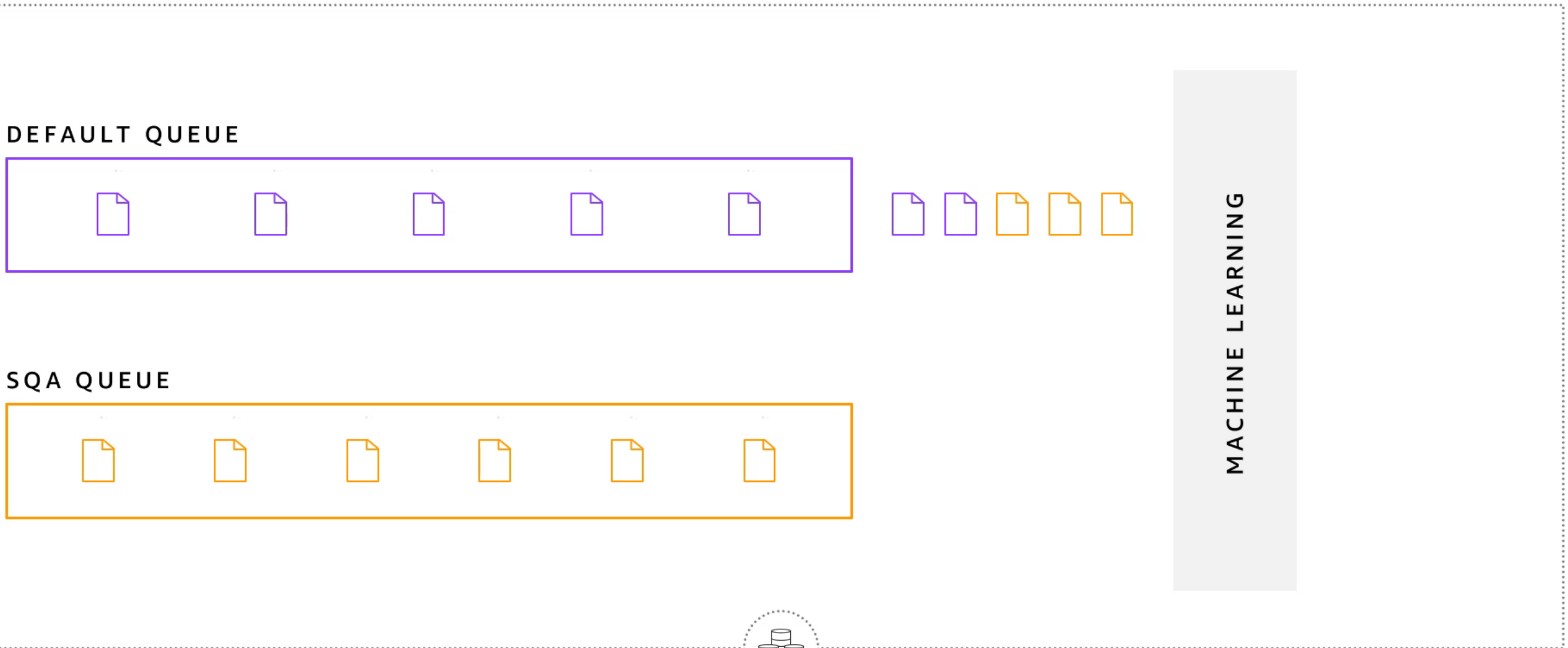
For every 24 hours that your main cluster is in use, you accrue a one-hour credit for Concurrency Scaling. This means that Concurrency Scaling is free for > 97% of customers.

How it works:

1. All queries go to the leader node, user only sees less wait for queries.
2. When queries in designated WLM queue begin queuing, Redshift automatically routes them to the new clusters, enabling Concurrency Scaling automatically.
3. Redshift automatically spins up a new cluster, processes waiting queries and automatically shuts down the Concurrency Scaling cluster.

aws

Amazon Redshift Burst

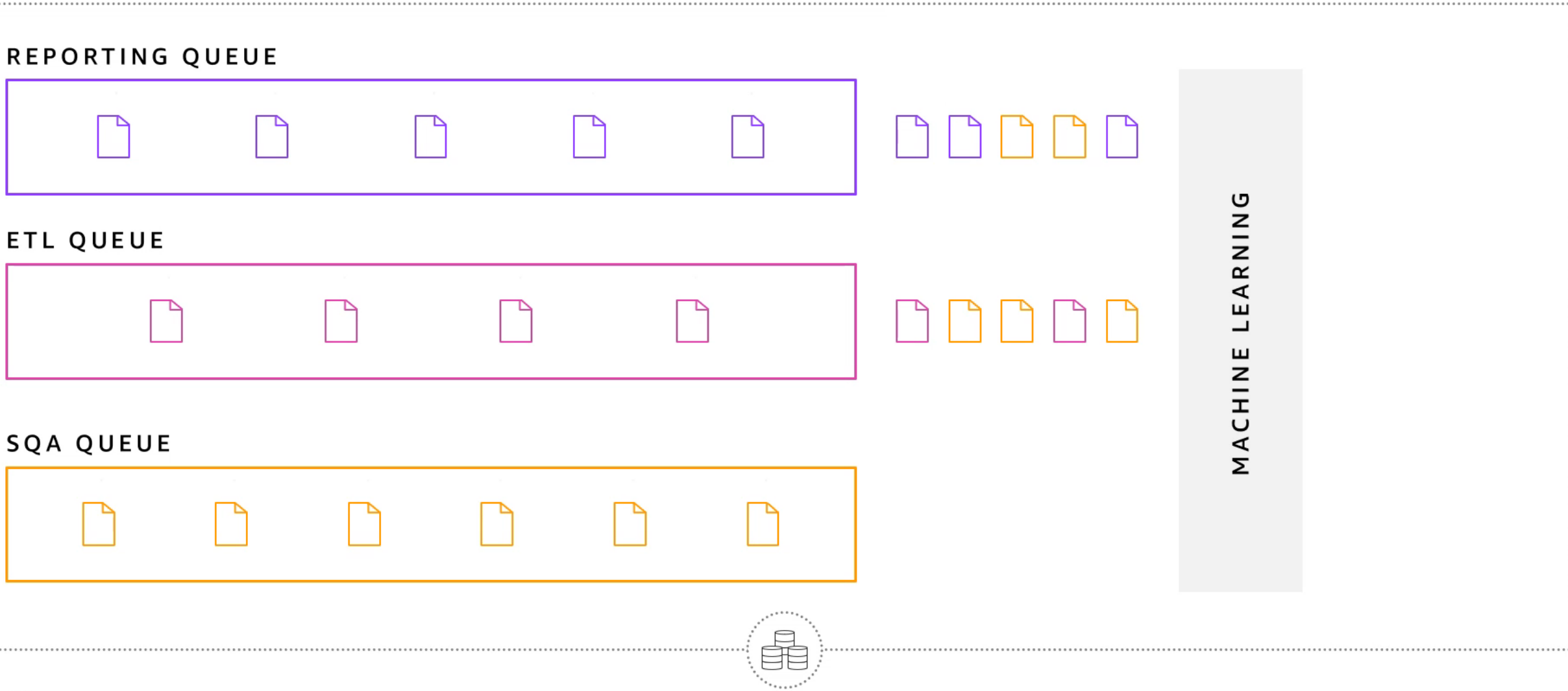


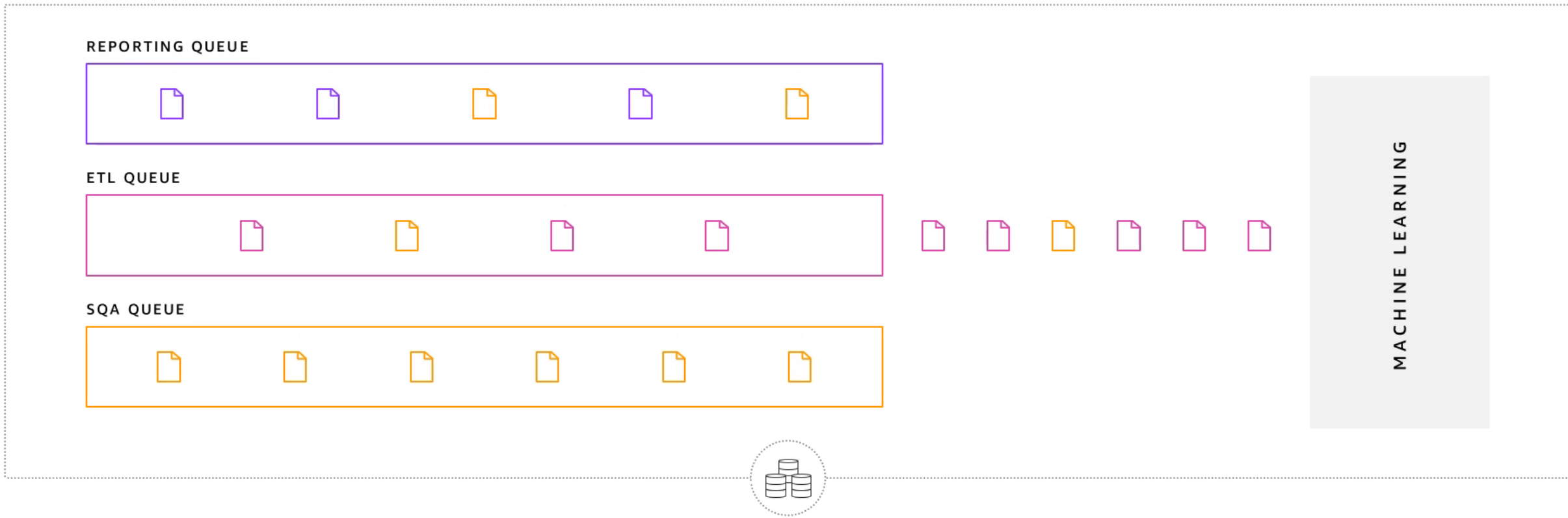
DEFAULT QUEUE

SQA QUEUE

MACHINE LEARNING

Amazon Redshift Burst





Concurrency Scaling FAQs

- Are there any constraints around what type of workload can be run on new clusters? Does the customer need to control this?
- How does a customer configure concurrency scaling (what are the options)?
- Will I see a big cost impact during the periods when concurrency and transient clusters are deployed?

aws

Intelligent Administration and Maintenance

Availability of intelligent administration and maintenance features

In Preview	Distribution Key		Recommendation for Distribution Key
Q2 2019	Sort Key		Recommendation for Sort Key
In Preview	Concurrency Setting		Automation for concurrent setting, making it dynamic
GA	Vacuum		Auto Vacuum in the background
GA	Analyze		Auto Analyze in the background

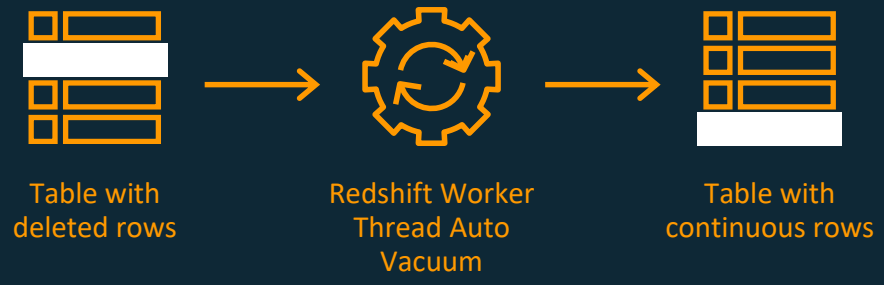
aws

Ease of use: Vacuum Delete

New!

When you delete or update rows, Redshift marks them for delete and Vacuum Delete removes the deleted rows to improve performance and free up space.

Redshift now automatically reclaims space from deleted rows, improving performance and space utilization.



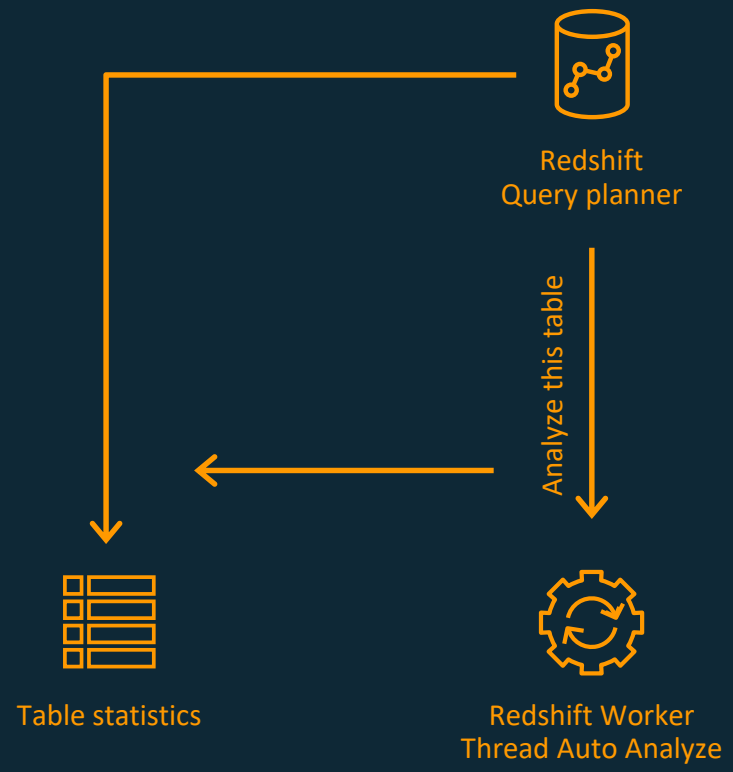
Redshift now prioritizes tables that need to be vacuumed and when the cluster usage is low, Redshift **automatically and incrementally runs vacuum in the background**.

This feature is available in select regions in Dec'18 and all regions in Jan'19.

Ease of use: Auto Analyze

New!

Redshift now automatically collects table statistics to deliver enhanced query performance.



Redshift detects changes in the table and when the cluster usage is low, it **automatically updates table statistics** in the background, so you don't have to run ANALYZE manually.

Updated table statistics are important inputs to Redshift query optimizer for optimal query plans.

This feature is available in select regions in Dec'18 and all regions in Jan'19.

aws

Intelligent Administration and Maintenance FAQs

Do I still need to run vacuum and analyze?

How can I track the auto vacuum's historical work?

How to know what's next for Redshift management improvements?

Redshift Advisor: Your DBA's best friend

- Redshift expert system available in AWS Console
- Identifies undesirable end user behaviors for resolution by providing high-impact recommendations to improve performance and reduce cost
- >96% of clusters have tailored feedback
- Actionable WLM, COPY, storage, and system maintenance feedback
- Analyses have doubled since launch (July '18); will double again by EOY

Recap of today's session:

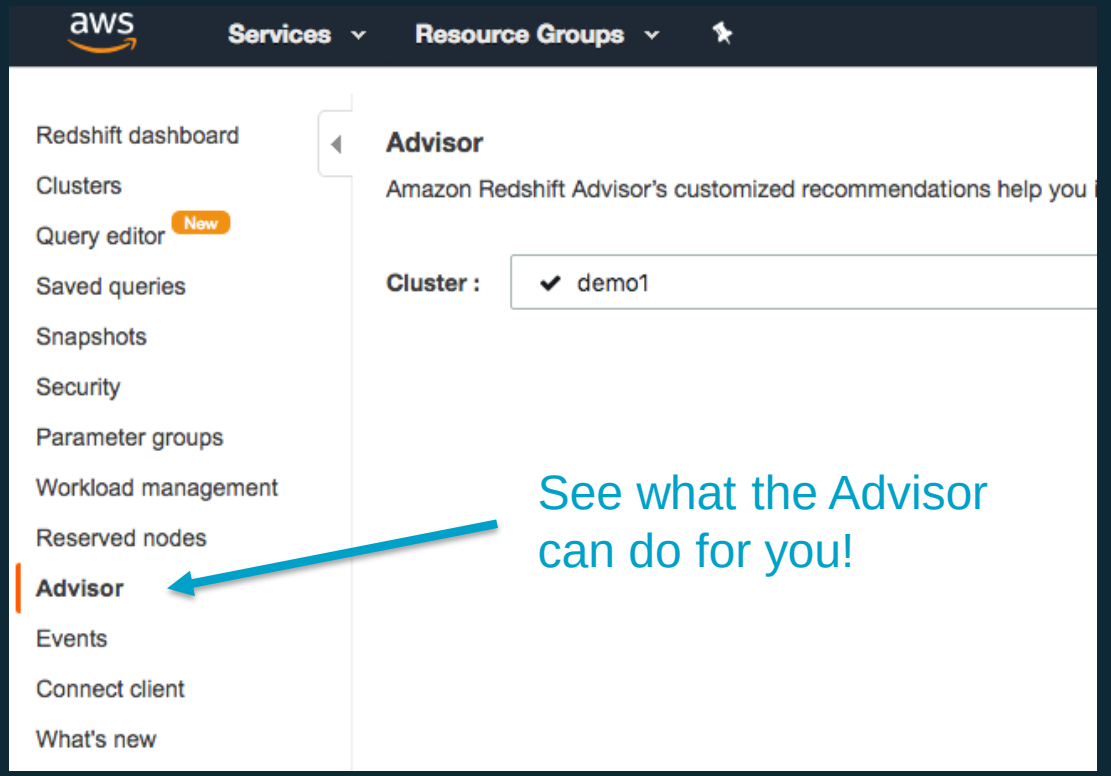
- How to scale compute and storage quickly on-demand:
  - Elastic and Classic Resize
- Scale compute automatically to handle thousands of concurrent users and queries:
  - Concurrency Scaling
- Understand how Redshift automatically optimizes resources and performs maintenance tasks to ensure the best performance at the lowest cost as you scale up and down.

aws

What's next?

Existing Redshift Users:

- Check the Redshift Advisor on your cluster(s).
- See how you can increase your clusters power and decrease costs.
- Look at your Workload Management with an eye towards Concurrency Scaling.
- Watch for maintenance window reductions/eliminations.
- Share your feedback/ideas with us. You really do drive the Redshift roadmap.



New to Redshift?

- Start with https://aws.amazon.com/redshift/getting-started/
- Try Amazon Redshift with a free trial.
- Share your feedback/ideas as well!

Tweet with #AmazonRedshift

Thank you!