



Masterclass

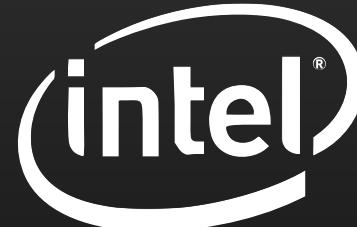
Amazon Redshift



Ian Massingham — Technical Evangelist

✉️ ianmas@amazon.com

🐦 [@IanMmmm](https://twitter.com/IanMmmm)



Masterclass

- 1 A technical deep dive that goes beyond the basics
- 2 Intended to educate you on how to get the best from AWS services
- 3 Show you how things work and how to get things done

Amazon Redshift



A fast, fully managed, petabyte-scale data warehouse
Makes it simple & cost-effective to analyse all your data

You can continue to use your existing tools

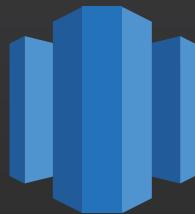
Start small & scale to a petabyte or more

Costs less than \$1,000/terabyte/year

Scalable

Optimised for
Data Warehousing

No Up-Front Costs



Amazon Redshift

Compatible

Secure

Simple

Amazon Redshift – The New AWS Data Warehouse

by Jeff Barr | on 28 NOV 2012 | [Permalink](#) [Comments](#)

You may have noticed that we announced Amazon Redshift today a data warehouse service that is a lot faster, simpler, and less expensive than alternatives available today.

Let's Talk About Data

A data warehouse is a specialized type of relational database, optimized for high-performance analysis and reporting of transactions. It collects current and historical transactional data from many disparate operational systems (manufacturing, finance, sales, shipping, etc.) and pulls it together in one place to guide analysis and decision-making. Without a data warehouse, a business lacks an integrated view of its data. Essential reports aren't available in time, visibility into changes that affect the business is reduced, and business suffers.

To date, setting up, running, and scaling a data warehouse has been complicated and expensive. Maintaining and monitoring all the moving parts (hardware, software, networking, storage) requires deep expertise across a number of disciplines. At scale, even comparatively simple things like loading or backing up data get hard.

Getting fast performance requires mastery of complicated query plans, access methods, and indexing techniques. No surprise, traditional data warehouses have been seen as gold-plated, premium-priced products with substantial upfront costs.

Enter Amazon Redshift

Amazon Redshift makes it easy for you setup, run, and scale a data warehouse of your own. AWS customers such as Netflix, JPL, and Flipboard have been testing it as part of a private beta. We are launching Amazon Redshift in a limited public beta mode today.

Amazon Redshift is a massively parallel, fully-managed data warehouse service, designed for data sets from hundreds of gigabytes to several petabytes in size, and appropriate for an organization of any size from a startup to a multi-national at a price point that will take you by surprise. Amazon Redshift is fully managed, so you no longer need to worry about provisioning hardware, installation, configuration or patching of system or database software. Because your business is dependent on your data, Amazon Redshift takes care to protect it by replicating all data within the cluster as well as in S3.

Amazon Redshift won't break the bank (or your credit card, since it is completely pay-as-you-go). We did the math and found that it would generally cost you between \$19,000 and \$25,000 per terabyte per year at list prices to build and run a good-sized data warehouse on your own. Amazon Redshift, all-in, will cost you less than \$1,000 per terabyte per year. For that price you get all of the benefits that I listed above without any of the operational headaches associated with building and running your own data warehouse.

Like every AWS service, you can create and manipulate an Amazon Redshift cluster using a set of web service

<https://aws.amazon.com/blogs/aws/category/amazon-redshift/>

User Defined Functions for Amazon Redshift

by Jeff Barr | on 11 SEP 2015 | in [Amazon Redshift](#) | [Permalink](#) | [Comments](#)

The Amazon Redshift team is on a tear. They are listening to customer feedback and rolling out new features all the time! Below you will find an announcement of another powerful and highly anticipated new feature.

— Jeff;

[Amazon Redshift](#) makes it easy to launch a petabyte-scale data warehouse. For less than \$1,000/Terabyte/year, you can focus on your analytics, while Amazon Redshift manages the infrastructure for you. Amazon Redshift's [price and performance](#) has allowed customers to unlock diverse analytical use cases to help them understand their business. As you can see from blog posts by [Yelp](#), [Amplitude](#) and [Criteo](#), our customers are constantly pushing the boundaries of what's possible with data warehousing at scale.

To extend Amazon Redshift's capabilities even further and make it easier for our customers to drive new insights, I am happy to announce that Amazon Redshift has added scalar [user-defined functions](#) (UDFs). Using PostgreSQL syntax, you can now create scalar functions in Python 2.7 custom-built for your use case, and execute them in parallel across your cluster.

Here's a template that you can use to create your own functions:

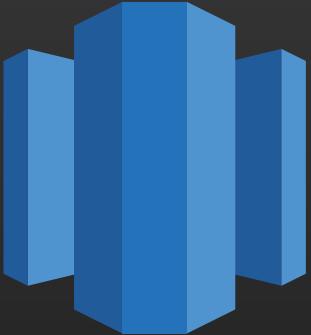
```
sql
CREATE [ OR REPLACE ] FUNCTION f_function_name
( [ argument_name arg_type, ... ] )
RETURNS data_type
{ VOLATILE | STABLE | IMMUTABLE }
AS $$  
    python_program  
$$ LANGUAGE plpythonu;
```

Scalar UDFs return a single result value for each input value, similar to built-in scalar functions such as [ROUND](#) and [SUBSTRING](#). Once defined, you can use UDFs in any SQL statement, just as you would use our built-in functions.

In addition to creating your own functions, you can take advantage of thousands of functions available through Python libraries to perform operations not easily expressed in SQL. You can even add custom libraries directly from S3 and the web. Out of the box, Amazon Redshift UDFs come integrated with the [Python Standard Library](#) and a number of other libraries, including:

- [NumPy](#) and [SciPy](#), which provide mathematical tools you can use to create multi-dimensional objects, do matrix operations, build optimization algorithms, and run statistical analyses.

<https://aws.amazon.com/blogs/aws/category/amazon-redshift/>



Agenda

Why Run Your Data Warehouse on AWS?

Getting Started

Table Design

Data Loading

Working with Data

Backup and Restoration

Upgrading & Scaling

WHY RUN YOUR DATA WAREHOUSE ON AWS?

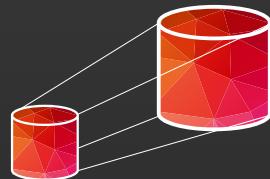
Database Management Challenges



Expensive



Difficult to administer

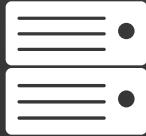


Challenging to scale

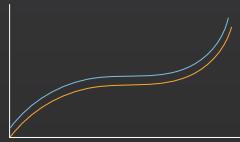


Very difficult to switch

Customers asked for data warehousing the AWS way



Easy to provision & scale up massively



No upfront costs, pay as you go



Really fast performance at a really low price



Open and flexible with support for popular BI tools

Amazon Redshift parallelises & distributes everything

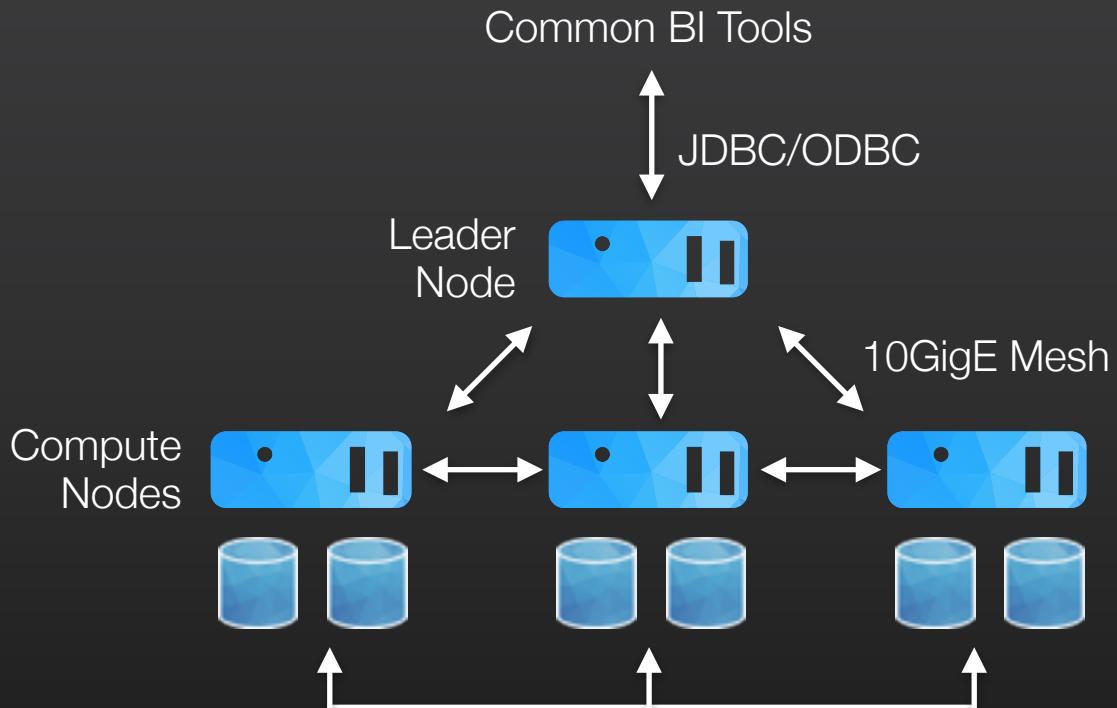
Query

Load

Backup

Restore

Resize



Amazon Redshift lets you start small & grow big

Dense Compute Nodes - dc1.large & dc1.8xlarge

dc1.large nodes

15GiB RAM

2 virtual cores, 10GigE

Single Node (160GB SSD)



Cluster 2-32 Nodes (320GB – 5.12TB SSD)



dc1.8xlarge nodes

244GiB RAM

32 virtual cores, 10GigE

Cluster 2-128 Nodes (up to 326TB SSD)



Amazon Redshift lets you start small & grow big

Dense Storage Nodes - ds1.xlarge, ds1.8xlarge, ds2.xlarge & ds2.8xlarge

ds2.8xlarge nodes

244GiB RAM

36 virtual cores, 10GigE, 16TB magnetic HDD per node

Cluster 2-128 Nodes (up to 2PB magnetic HDD)



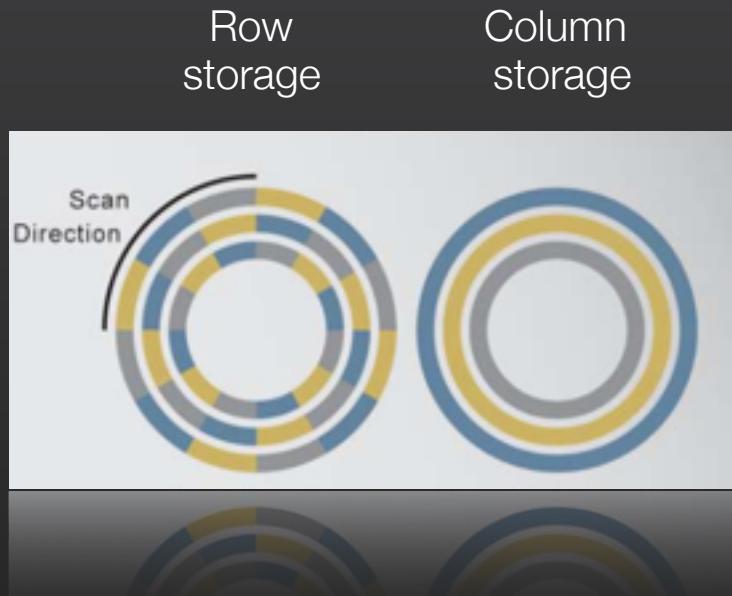
Simplify Cluster Operations with Amazon Redshift

- ▶ Built-in security in transit, at rest, when backed up
- ▶ Backup to Amazon S3 is continuous, incremental & automatic
- ▶ Streaming restores let you resume querying faster
- ▶ Disk failures are transparent & nodes recover automatically



Amazon Redshift Dramatically Reduces IO

- ▶ Column storage
- ▶ Data compression
- ▶ Zone maps
- ▶ Direct-attached storage
- ▶ Large data block sizes



Amazon Redshift in action



- Needed a way to increase speed, performance and flexibility of data analysis at a low cost
- Using AWS enabled FT to run queries 98% faster than previously—helping FT make business decisions quickly
- Easier to track and analyze trends
- Reduced infrastructure costs by 80% over traditional data center model

“When our analysts first started to do queries on Amazon Redshift, they thought it was broken because it was working so fast.

John O'Donovan
CTO, Financial Times

FINANCIAL TIMES

Financial Times is one of the world's leading business news organizations, with a combined paid print and digital circulation of 665,000.

GETTING STARTED

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Instances @ [ultimo-area](#) Origins Support

Amazon Web Services

Compute

- EC2 Virtual Servers in the Cloud
- EC2 Container Service Run and Manage Docker Containers
- Elastic Beanstalk Run and Manage Web Apps
- Lambda Run Code in Response to Events

Storage & Content Delivery

- S3 Simple Storage in the Cloud
- CloudFront Global Content Delivery Network
- Elastic File System PREVIEW Fully Managed File System for S3
- Glacier Amazon Storage in the Cloud
- Import/Export Snowball Large Data Transfer
- Storage Gateway Integrates On-Premises IT Environments with Cloud Storage

Database

- RDS Managed Relational Database Service
- DynamoDB Scalable and Scalable NoSQL Data Store
- ElastiCache In-Memory Cache
- Redshift** Managed Petabyte-Scale Data Warehouse Service

Networking

- VPC Isolated Cloud Resources
- Direct Connect Dedicated Network Connection to AWS
- Route 53 Services-DNS and Domain Name Registration

Developer Tools

- CodeCommit Store Code in Private Git Repositories
- CodeDeploy Automate Code Deployments
- CodePipeline Releases Software using Continuous Delivery

Management Tools

- CloudWatch Monitor Resources and Applications
- CloudFormation Create and Manage Resources with Templates
- CloudTrail Track User Activity and API Usage
- Config Track Resource Inventory and Changes
- OpsWorks Automate Operations with Chef
- Service Catalog Create and Use Standardized Products
- Trusted Advisor Optimize Performance and Security

Security & Identity

- Identity & Access Management Manage User Access and Encryption Keys
- Directory Service Intel and Manage Active Directory
- Inspector PREVIEW Analyze Application Security
- WAF Filter Malicious Web Traffic

Analytics

- EMR Manage Big Data Framework
- Data Pipeline Orchestration for Data-Driven Workflows
- Elasticsearch Service Run and Index Elasticsearch Clusters
- Kinesis Work with Real-time Streaming Data
- Machine Learning Build Smart Applications Quickly and Easily

Internet of Things

- AWS IoT RTT Connect Devices to the Cloud

Mobile Services

- Mobile Hub RTT Build, Test, and Monitor Mobile Apps
- Cognito User Identity and App-Data Synchronization
- Device Farm Test Android, Fire OS, and iOS Apps on real devices in the Cloud

Mobile Analytics

- Mobile Analytics Collect, View and Export App Analytics

SMS Push Notification Service

Application Services

- API Gateway Build, Deploy and Manage APIs
- AppStream Low-Latency Application Streaming
- CloudSearch Managed Search Service
- Elastic Transcoder Enterprise-Scale Media Processing
- SES Email Sending Service
- SQS Message Queue Service
- SWF Workforce Service for Coordinating Application Concurrency

Enterprise Applications

- WorkSpaces Desktops in the Cloud
- WorkDocs Secure Enterprise Storage and Sharing Service
- WorkMail PREVIEW Secure Email and Calendar Service

Resource Groups

A resource group is a collection of resources that share one or more tags. Create a group for each project, application, or environment in your account.

[Create a Group](#) [Tag Editor](#)

Additional Resources

[Getting Started](#) ?
Read our documentation or view our training to learn more about AWS.

[AWS Console Mobile App](#) ?
View your resources on the go with our AWS Console mobile app, available from Amazon Appstore, Google Play, or iTunes.

[AWS Marketplace](#) ?
Find and buy software, launch with 1-Click and pay by the hour.

[AWS re:Invent Announcements](#) ?
Explore the next generation of AWS cloud capabilities. [See what's new](#)

Service Health

All services operating normally

Updated: Oct 28 2018 01:09:00 (GMT+0000)

[Service Health Dashboard](#)

Feedback English Help

© 2006 - 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use See full terms and conditions [here](#) | [Report Abuse](#) | [Feedback](#) | [Help](#) | [English](#) | [Feedback](#) | [Help](#) | [English](#)

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services SAM VPC EC2 S3 RDS CloudFront Edit Jenkins @ Jenkins-ws Oregon Support

Welcome to Amazon Redshift

You do not appear to have any clusters in the US West (Oregon) region.

Amazon Redshift is a fast and powerful, fully managed, petabyte-scale data warehouse service in the cloud. Amazon Redshift offers you fast query performance when analyzing virtually any size data set using the same SQL-based tools and business intelligence applications you use today. With a few clicks in the AWS Management Console, you can launch a Redshift cluster, starting with a few hundred gigabytes of data and scaling to a petabyte or more, for under \$1,000 per terabyte per year.

Try Amazon Redshift for free! If you've never created an Amazon Redshift cluster, you're eligible for a two-month free trial of our dc1.large node. You get 750 hours per month for free, enough hours to continuously run one dc1.large node with 160GB of compressed S3D storage. You can also build clusters with multiple nodes to test larger data sets, which will consume your free hours more quickly. Once your two month free trial expires or your usage exceeds 750 hours per month, you can shut down your cluster, avoiding any charges, or keep it running at our standard On-Demand Rate. For more information, please see the [Free Trial FAQ page](#).

[Launch Cluster](#)

Get up and running immediately

Create Cluster Manage & Configure Load & Query Data


Learn More


Learn More


Learn More

New Features

- Follower Announcements Thread
- New Dense Storage Instance
- Filter data faster with InterScan
- Custom JDBC/ODBC Drivers
- See All

Getting Started

- Free Trial
- Pricing and Specs
- Getting Started Guide

Best Practices

- Connecting to Clusters
- Loading Data
- Designing Tables
- Tuning Query Performance
- + Admin Scripts

AWS Webcast

May 2010 Eric Pernice & John Lougheed

AWS Marketplace

Informatica Cloud Advanced for Amazon Redshift (up to 1TB)
Provided by Informatica
Rating *****
From \$60.00/hour for software
[View all Business Intelligence](#)

Matillion ETL for Redshift
Provided by Matillion
Rating *****
From \$1.31/hour for software
[View all Business Intelligence](#)

Feedback English Help Feedback

© 2006 - 2010 Amazon Web Services, Inc. or its affiliates. All rights reserved.
Privacy Policy Terms of Use
Help Center Support
About Us Customer Support
Feedback
Feedback

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services SAM VPC EC2 S3 RDS CloudFront Edit Jenkins @ Jenkins-aws Drag & Drop Support

Redshift Dashboard

Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

Resources
You are using the following Amazon Redshift resources in the US West (Oregon) region (used):

Clusters (2)	Security	Parameter Groups (2)
Increase cluster limit	Subnet Groups (2)	Reserved Nodes (0)
Snapshots (2)	HEM Connections (2)	Events (0)
Manual (2)	HEM Certificates (2)	Event Subscriptions (0)
Automated (0)		

New Features
Follow Announcements Thread
New Dense Storage Instance Type
Filter data faster with Interleaved Sorting
Custom ODBC/JDBC Drivers
See All

Launch Cluster

Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse solution that makes it simple and cost-effective to efficiently analyze all your data using your existing business intelligence tools.

Getting Started
Free Trial
Pricing and Specs
Getting Started Guide

Best Practices
Connecting to Clusters
Loading Data
Designing Tables
Tuning Query Performance + Admin Scripts

AWS Webcast - Amazon Redshift Best Practices
May 2016 Drs. Francisco & John Langford

Service Health

Current Status	Details
Green Redshift (Oregon)	Service is operating normally

[View complete service health details](#)

AWS Marketplace

Informatica Cloud Advanced for Amazon Redshift (up to 1TB)
Provided by Informatica
Rating *****
From \$60.00/hour for software
[View all Business Intelligence](#)

Matillion ETL for Redshift
Provided by Matillion
Rating *****
From \$1.33/hour for software
[View all Business Intelligence](#)

Tableau Server (10 users)
Provided by Tableau
Rating *****
From \$0.63/hour for software
[View all Business Intelligence](#)

Feedback English Help

© 2006 - 2016 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use See full terms and conditions. This page is not affiliated with Amazon.com, Inc. Last updated: 2016-05-10 10:47:48 UTC Version 1.025 - 2009-SQS

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services SAM VPC EC2 S3 RDS CloudFront Edit Jenkins @ Jenkins-AWS Drag & Drop Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

CLUSTER DETAILS MORE CONFIGURATION ADDITIONAL COPY SUBMISSION REVIEW

Provide the details of your cluster. Fields marked with * are required.

Cluster Identifier* This is the unique key that identifies a cluster. This parameter is stored as a lowercase string. (e.g. my-red-instance)

Database Name* Optional: A default database named dev is created for the cluster. Optionally, specify a custom database name (e.g. mydb) to create an additional database.

Database Port* 5439 Port number on which the database accepts connections.

Master User Name* Name of master user for your cluster (e.g. dbuser).

Master User Password* Password must contain 8 to 64 printable ASCII characters excluding: !, ", ', \, and @. It must contain 1 uppercase letter, 1 lowercase letter, and 1 number.

Confirm Master User Password* Confirm Master User Password.

[Cancel](#) [Continue](#)

Feedback English Help Feedback

© 2006 - 2016 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

about this page Help Center

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services SAM VPC EC2 S3 RDS CloudFront Edit Jenkins @ Jenkins-aws Drag & Drop Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

CLUSTER DETAILS MORE CONFIGURATION ADDITIONAL COPY SUBMISSION REVIEW

Provide the details of your cluster. Fields marked with * are required.

Cluster Identifier* mycluster This is the unique key that identifies a cluster. This parameter is stored as a lowercase string. (e.g. my-clv-instance)

Database Name mydatabase Optional: A default database named dev is created for the cluster. Optionally, specify a custom database name (e.g. mydb) to create an additional database.

Database Port* 5439 Port number on which the database accepts connections.

Master User Name* master Name of master user for your cluster (e.g. dbuser).

Master User Password* ***** Password must contain 8 to 64 printable ASCII characters excluding: !, ", ', \, and @. It must contain 1 uppercase letter, 1 lowercase letter, and 1 number.

Confirm Password* ***** Confirm Master User Password.

Cancel Continue

Feedback English Help Feedback

© 2006 - 2016 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

about this page Help Center

Creating an Amazon Redshift Data Warehouse via the AWS Console

Screenshot of the AWS Redshift Cluster Configuration page, showing the creation of a new cluster.

The page title is "Create New Cluster". The top navigation bar includes "AWS Services", "IAM", "VPC", "EC2", "S3", "RDS", "CloudFront", "Edit", "Amazon S3", "Oregon", and "Support".

The left sidebar lists "Redshift Dashboard", "Clusters", "Snapshots", "Security", "Parameter Groups", "Reserved Nodes", "Events", and "Connect Client".

The main configuration area has tabs for "CLUSTER DETAILS", "NODE CONFIGURATION", "ADDITIONAL CLUSTER INFORMATION", and "SUMMARY".

Node Type: (highlighted with an orange box)

Specifies the compute, memory, storage, and I/O capacity of the cluster's nodes.

CPU: 7 EC2 Compute Units (2 virtual cores) per node

Memory: 15 GB per node

Storage: 160GB SSD storage per node

I/O Performance: Moderate

Cluster Type: (highlighted with an orange box)

Number of Compute Nodes: (highlighted with an orange box)

Single Node clusters consist of a single node which performs both leader and compute functions.

Maximum: 1

Minimum: 1

Buttons at the bottom: "Cancel", "Previous", and "Continue" (highlighted with an orange box).

Page footer: "Feedback", "English", "Feedback", "Privacy Policy", "Terms of Use", "See Us In Seattle", and "Visit AWS Help Center".

Creating an Amazon Redshift Data Warehouse via the AWS Console

The screenshot shows the 'Cluster Configuration' step of the Redshift cluster creation wizard. On the left, a sidebar lists 'Clusters', 'Parameter Groups', 'Reserved Nodes', and 'Events'. The main area is titled 'Cluster Details' and contains fields for 'Node Type' (set to 'db.t1.large'), 'CPU' (7 EC2 Compute Units), 'Memory' (15 GB per node), 'Storage' (1600GB SSD storage per node), and 'I/O Performance' (Moderate). Below this, the 'Cluster Type' is set to 'Single Node'. A note states: 'Single Node clusters consist of a single node which performs both leader and compute functions.' Under 'Number of Compute Nodes', the value is set to 1, with a range from Minimum 1 to Maximum 1. At the bottom right, there are 'Cancel', 'Previous', and 'Continue' buttons, with 'Continue' being highlighted with a red box.

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Sessions @ IaMaws-aws Oregon Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

Choose a number of nodes and Node Type below. Number of Compute Nodes is required for multi-node clusters.

Node Type: db.t1.large

CPU: 7 EC2 Compute Units (2 virtual cores) per node

Memory: 15 GB per node

Storage: 1600GB SSD storage per node

I/O Performance: Moderate

Cluster Type: Single Node

Number of Compute Nodes: 1

Maximum: 1

Minimum: 1

Single Node clusters consist of a single node which performs both leader and compute functions.

Cancel Previous Continue

Feedback English

© 2006 - 2010, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use See Us in Seattle Visit Seattle

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services IAM VPC EC2 S3 ROS CloudFront 1 Edit Home Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

CLUSTER DETAILS NODE CONFIGURATION ADDITIONAL CONFIGURATION REVIEW

Provide the optional additional configuration details below.

Cluster Parameter Group A default parameter group will be associated with this cluster.

Encrypt Database None KMS HSM Learn more about database encryption.

Configure Networking Options:

Choose a VPC Default VPC (vpc-ea6a0364) The identifier of the VPC in which you want to create your cluster.

Cluster Subnet Group default Selected Cluster Subnet Group may limit the choice of Availability Zones.

Publicly Accessible Yes Select Yes if you want the cluster to be accessible from the public Internet. Select No if you want it to be accessible only from within your private VPC network.

Choose a Public IP Address No Select Yes if you want to select your own public IP address from a list of elastic IP (EIP) addresses that are already configured for your cluster's VPC. Select No if you want Amazon Redshift to provide an EIP for you instead.

Availability Zone No Preference The EC2 Availability Zone that the cluster will be created in.

Optional: associate your cluster with one or more security groups.

VPC Security Groups default (sg-07a6a02) List of VPC Security Groups to associate with this cluster.

Optional: create a basic alarm for this cluster.

Create CloudWatch Alarms Yes No Create a CloudWatch alarm to monitor the disk usage of your cluster.

Cancel Previous Continue

Feedback English

© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Violate VSPN
Feedback
English

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Instances @ [Amazon VPC](#) Oregon Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

CLUSTER DETAILS NODE CONFIGURATION ADDITIONAL CONFIGURATION REVIEW

Provide the optional additional configuration details below.

Cluster Parameter Group A default parameter group will be associated with this cluster.

Encrypt Database None KMS HSM [Learn more about database encryption](#)

Configure Networking Options:

Choose a VPC Default VPC (vpc-e6e60364) The identifier of the VPC in which you want to create your cluster

Cluster Subnet Group default Selected Cluster Subnet Group may limit the choice of Availability Zones

Publicly Accessible Yes Select Yes if you want the cluster to be accessible from the public Internet. Select No if you want it to be accessible only from within your private VPC network.

Choose a Public IP Address No Select Yes if you want to select your own public IP address from a list of elastic IP (EIP) addresses that are already configured for your cluster's VPC. Select No if you want Amazon Redshift to provide an EIP for you instead.

Availability Zone No Preference The EC2 Availability Zone that the cluster will be created in.

Optional: associate your cluster with one or more security groups.

VPC Security Groups default-log-destination List of VPC Security Groups to associate with this cluster.

Optional: create a basic alarm for this cluster.

Create CloudWatch Alarms Yes No Create a CloudWatch alarm to monitor the disk usage of your cluster.

Disk Usage Threshold 90% Threshold at which the alarm will trigger when disk usage across all nodes reaches this percentage.

Use Existing Topic Yes Use an existing SNS topic or create a new one. SNS is a Simple Notification Service which will send email notifications to the recipients of the SNS topic when the alarm triggers.

No Name of the SNS topic that will be created.

Topic include default alarms Name of the SNS topic that will be created.

Recipients Recipients of this SNS topic. If you have multiple recipients, separate the recipients with a comma.

[Cancel](#) [Previous](#) [Continue](#)

Feedback English Help

© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Version 1.0 | Last updated 01/01/2016 | Page 14 of 24 | Page 14 of 24

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Home Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

CLUSTER DETAILS NODE CONFIGURATION ADDITIONAL CONFIGURATION REVIEW

You are about to launch a cluster with following the following specifications:

Cluster Properties

These attributes specify the name of your cluster, what type of virtual hardware it will run on, how many nodes it will contain, and the availability zone in which it will be located.

Cluster Identifier: mycluster
Node Type: ds1.large
Number of Compute Nodes: 1 (Master and compute run on a single node)
Availability Zone: No Preference

Database Configuration

These properties specify the database-name, port, and username you will use to connect to the database. The parameter group contains configuration values used by the database.

Database Name: mydatabase
Database Port: (43)
Master User Name: mroot
Cluster Parameter Group: A default parameter group will be created when the cluster is launched.

Security, Access, and Encryption

These settings control whether your cluster will be created in an existing VPC to allow for simpler integration with other AWS Services, and the security groups which define access rules to your cluster.

Virtual Private Cloud: vpc-edabit0364
Cluster Subnet Group:
Publicly Accessible: Yes
Elastic IP: Not used
VPC Security Groups: sg-b7abecb
Encrypt Database: No

CloudWatch Alarms

CloudWatch alarms are used to notify if metrics for your cluster are within a certain threshold. All recipients under the SNS topic specified for your alarm will receive notifications once an alarm is triggered.

Basic alarms will not be created for this cluster.

⚠️ Unless you are eligible for the free trial, you will start accruing charges as soon as your cluster is active.

Applicable charges:
The on-demand hourly rate for this cluster will be \$0.05, or \$0.25/node. If you have purchased reserved nodes in this region for this node type that are active, your costs will be discounted. Additional nodes will be billed at the on-demand rate.

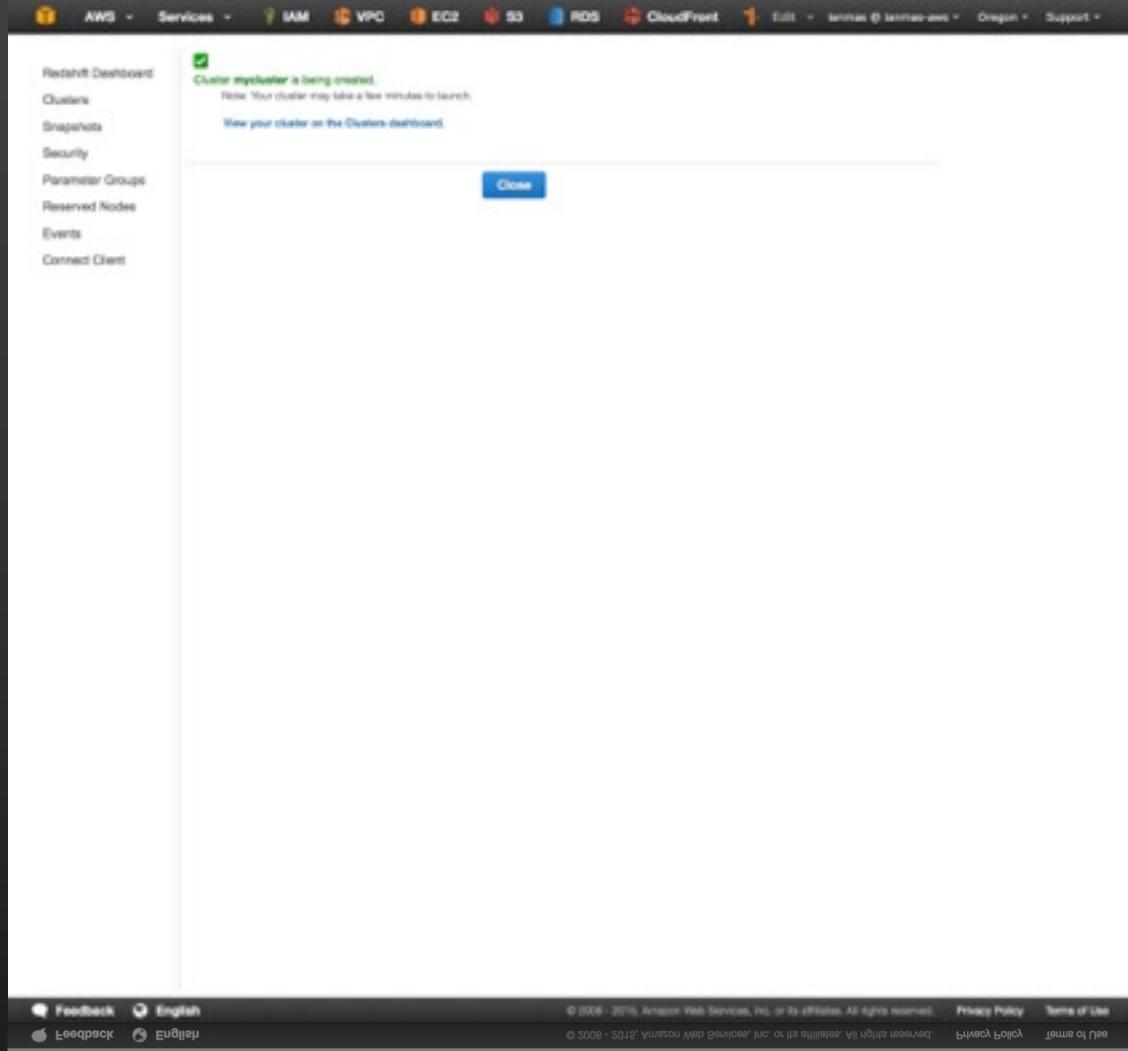
If you are eligible for a free trial, you will receive 750 hours of free usage for each month of the trial, applied across all running ds1.large nodes across all regions. Regardless of when you start your trial, you will receive two full months of free usage. Once your trial expires or your usage exceeds 750 hours/month, you can shut down your cluster, avoiding any charges, or keep it running at our standard On-Demand Rate.

For more information, see [Amazon Redshift Free Trial FAQ](#), [Amazon Redshift Pricing](#), and [Reserved Nodes Documentation](#).

Cancel Previous **Launch Cluster**

Feedback English Feedback English © 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use Help

Creating an Amazon Redshift Data Warehouse via the AWS Console



Creating an Amazon Redshift Data Warehouse via the AWS Console

The screenshot shows the AWS Redshift Clusters page. The left sidebar menu includes options like 'Clusters' (which is selected), 'Snapshots', 'Security', 'Parameter Groups', 'Reserved Nodes', 'Events', and 'Connect Client'. The main content area is titled 'Clusters' and contains a table with columns: Cluster, Cluster Status, DB Health, In Maintenance, and Recent Events. A single row is visible, showing a cluster named 'mycluster' with a status of 'creating'. The bottom of the page includes standard AWS navigation links for Feedback, Language, and Help, along with copyright information and links to Privacy Policy and Terms of Use.

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Account Oregon Support

Clusters

Launch Cluster Manage Tags

Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
mycluster	creating	unhealthy	unhealthy	0

Feedback English Feedback English

© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services IAM VPC EC2 S3 RDS CloudFront Edit Account Support

Clusters

Launch Cluster Manage Tags

Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
mycluster	creating	unknown	unknown	0

Endpoint: [aws.aaaaaaaa.cluster.redshift.amazonaws.com](#)

Cluster Properties

Cluster Name	mycluster	Cluster Status	CREATING
Node Type	dc1.large	Database Health	UNKNOWN
Nodes	1	In Maintenance Mode	UNKNOWN
Zone	us-west-2b	Parameter Group Apply Status	in-sync
Cluster Parameter Group	default:redshift-1.0 (in-sync)	Pending Modified Values	Master User Password: -----

Cluster Database Properties

Port	Backup, Audit Logging, and Maintenance
Database Name	Automated Snapshot Retention Period: 1
Master Username	Cross-Region Snapshots Enabled: No
Encrypted	Audit Logging Enabled: No
No	Maintenance Window: wed 08:00-wed 08:30
No	Allow Version Upgrades: Yes

Tags

You have not created any tags. Please add tags using the [Manage Tags](#) button above.

Feedback English FeedBack Email

© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

Powered by Apache JBoss Seam 3.0.1.Final-redshift-201607261010, on IBM WAS Dev 8.5.5.201607261010 tomcat7.0.80-redshift-201607261010

Creating an Amazon Redshift Data Warehouse via the AWS Console

The screenshot shows the AWS Redshift Clusters page. The left sidebar has a 'Clusters' section selected. The main area displays a table with one row for a cluster named 'mycluster'. The columns are: Cluster (mycluster), Cluster Status (available), DB Health (healthy), In Maintenance (no), and Recent Events (0). There are 'Launch Cluster' and 'Manage Tags' buttons at the top of the table.

Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
mycluster	available	healthy	no	0

Feedback English Feedback English
© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
About Us Help Center Customer Support
Feedback English Feedback English
© 2006 - 2016, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
About Us Help Center Customer Support

Creating an Amazon Redshift Data Warehouse via the AWS Console

AWS Services SAM VPC EC2 S3 RDS CloudFront Edit Jenkins @ Jenkins-aws Drag & Drop Support

Clusters Snapshots Security Parameter Groups Reserved Nodes Events Connect Client

Clusters

Launch Cluster Manage Tags

Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
mycluster	available	healthy	no	1

Endpoint: system-ec2-164181888.us-west-2.compute.amazonaws.com:5439 [authorized] ⓘ

Cluster Properties

Cluster Name	mycluster	Cluster Status	available
Node Type	dc1.large	Database Health	healthy
Nodes	1	In Maintenance Mode	no
Zone	us-west-2b	Parameter Group Apply Status	in sync
Cluster Parameter Group	default:redshift-1.0 (in-sync)	Pending Modified Values	None

Cluster Database Properties

Port	5439	Backup, Audit Logging, and Maintenance	
Database Name	mydatabase	Automated Snapshot Retention Period	1
Master Username	master	Cross-Region Snapshots Enabled	No
Encrypted	No	Audit Logging Enabled	No

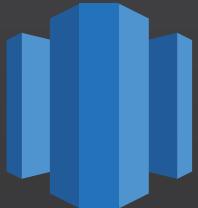
Tags ⓘ

You have not created any tags. Please add tags using the [Manage Tags](#) button above.

Feedback English Help

© 2006 - 2016 Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use See Full Terms & Conditions

Creating an Amazon Redshift Data Warehouse via the AWS Console



Getting Started with Amazon Redshift

- ▶ 2 Month free trial
- ▶ Getting Started tutorial
- ▶ Amazon Redshift System Overview
- ▶ Guides for table design, loading data & query design
- ▶ Connect using industry standard OBDC/JDBC connections
- ▶ Many BI & ETL vendors have certified Amazon Redshift

TABLE DESIGN

Choosing the optimal table design

- ▶ Compression Encodings
- ▶ Choosing the right data types
- ▶ Distributing and sorting data

Choosing a Column Compression Type

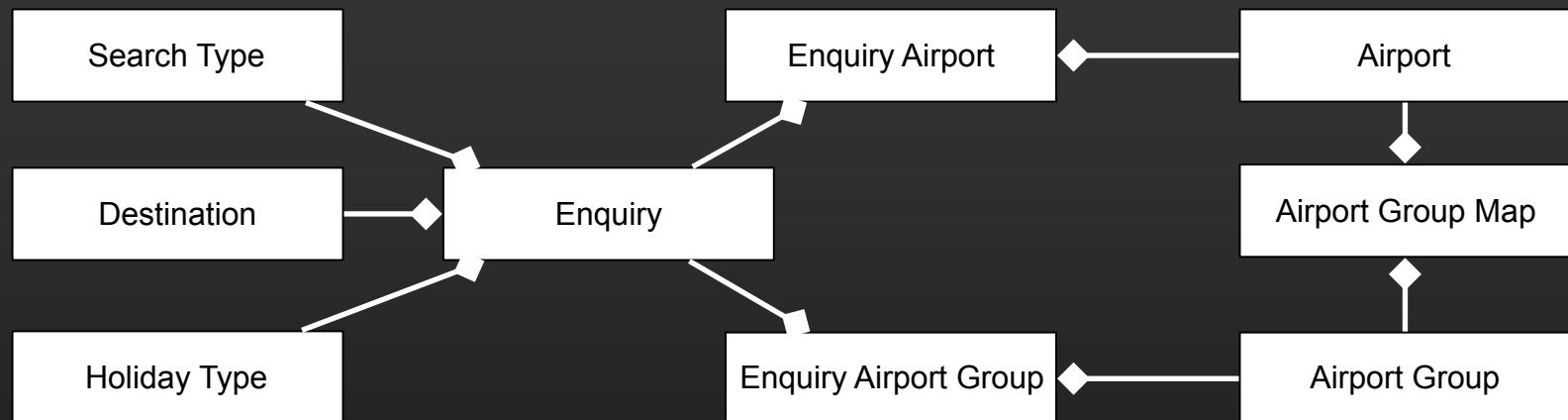
- ▶ Compression is a key tool improving performance in modern data warehouse systems to reduce the size of data that is read from storage, minimise IO & improve query performance
- ▶ The COPY command used to load data in Amazon Redshift will perform compression analysis
- ▶ We strongly recommend using the COPY command to apply automatic compression

Data Types

Data Type	Aliases	Description
SMALLINT	INT2	Signed two-byte integer
INTEGER	INT, INT4	Signed four-byte integer
BIGINT	INT8	Signed eight-byte integer
DECIMAL	NUMERIC	Exact numeric of selectable precision
REAL	FLOAT4	Single precision floating-point number
DOUBLE PRECISION	FLOAT8, FLOAT	Double precision floating-point number
BOOLEAN	BOOL	Logical Boolean (true/false)
CHAR	CHARACTER, NCHAR, BPCHAR	Fixed-length character string
VARCHAR	CHARACTER VARYING, NVARCHAR, TEXT	Variable-length character string with a user-defined limit
DATE		Calendar date (year, month, day)
TIMESTAMP	TIMESTAMP WITHOUT TIME ZONE	Date and time (without time zone)

Your Schema

► Expressed in 3rd Normal Form



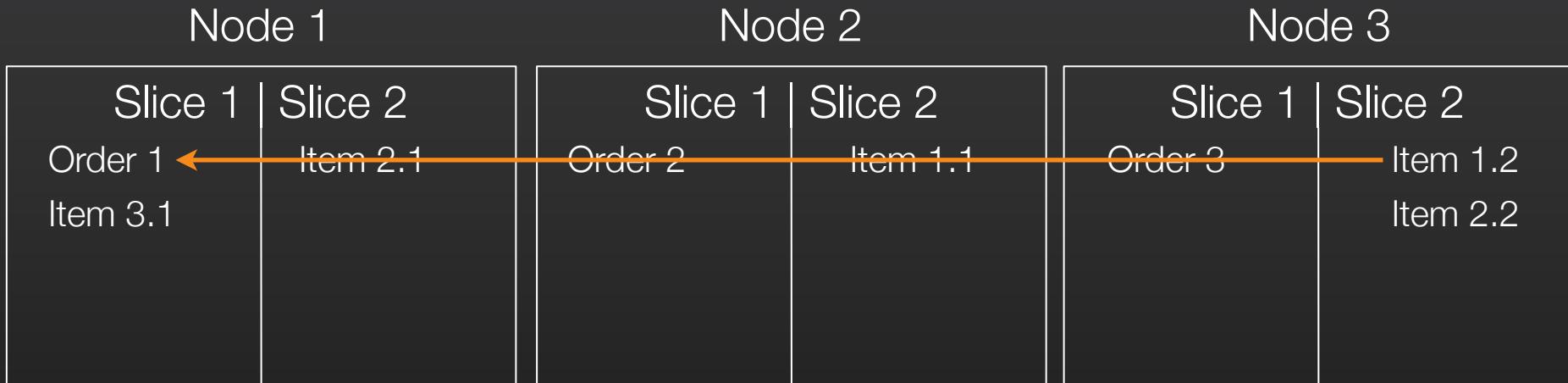
Distributing Data

- ▶ Redshift is a distributed system:
 - ▶ A cluster contains a leader node & compute nodes
 - ▶ A compute node contains slices (one per core)
 - ▶ A slice contains data
- ▶ Slices are chosen based on two types of distribution:
 - ▶ Round Robin (automated)
 - ▶ Based on a distribution key (hash of a defined column)
- ▶ Queries run on all slices in parallel: optimal query throughput can be achieved when data is evenly spread across slices

Unoptimised Distribution



Order 1: Dave Smith, Total £195
Item 1.1: Order 1, Kindle Fire HD 7", £159
Item 1.2: Order 1, Kindle Fire Case, £36



Default (No Distribution Key, Round Robin Order)

Optimised Distribution



Order 1: Dave Smith, Total £195

Item 1.1: Order 1, Kindle Fire HD 7", £159
Item 1.2: Order 1, Kindle Fire Case, £36

Node 1

Node 2

Node 3

Node 1	Node 2	Node 3
Slice 1	Slice 1	Slice 1
Slice 2	Slice 2	Slice 2
Order 1 Item 1.1 Item 1.2	Order 2 Item 2.1 Item 2.2	Order 3 Item 3.1

Customised (ORDERS.ORDER_ID DISTKEY, ITEMS.ORDER_ID DISTKEY)

Choosing a Distribution Key

- ▶ Frequently Joined
 - ▶ By most commonly run queries
 - ▶ By queries which consume the most CPU
- ▶ High Cardinality
 - ▶ Large number of discrete values
- ▶ Low Skew
 - ▶ Uniform Distribution
 - ▶ No Hotspots
- ▶ Query STV_BLOCKLIST for Skew Factor

Sorting Data

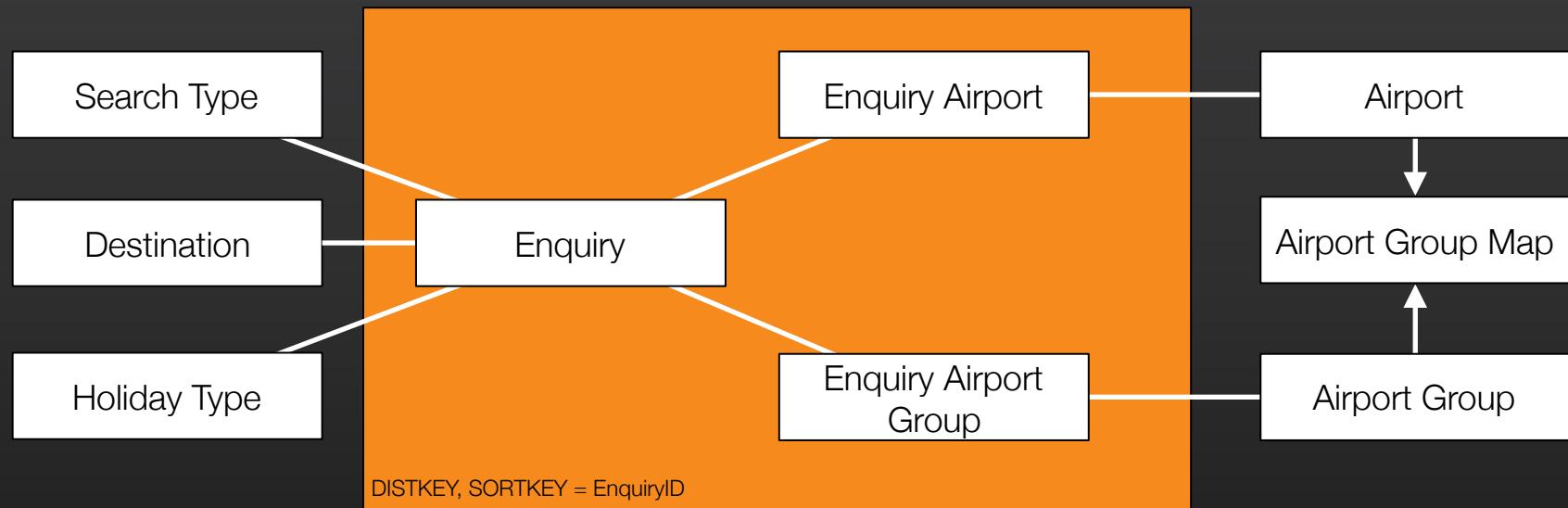
- ▶ In the slices (on disk), the data is sorted by a sort key (if none uses the insertion order)
- ▶ Choose a sort key that is frequently used in your queries
 - ▶ As a query predicate (date, identifier, ...)
 - ▶ As a join parameter (it can also be the hash key)
- ▶ The sort key allows Redshift to avoid reading entire blocks based on predicates
- ▶ E.g.: a table containing a timestamp sort key, and where only recent data is accessed, will skip blocks containing “old” data

Schema Design

- ▶ Optimizing a database for querying
- ▶ Analyse using Automatic Compression to ensure optimal IO
- ▶ Co-locate frequently joined tables: use distribution key wisely
(avoids data transfers between nodes)
- ▶ For joined tables, use sort keys on the joined columns, allowing
fast merge joins
- ▶ Compression allows you to de-normalize without penalizing
storage, simplifying queries and limiting joins

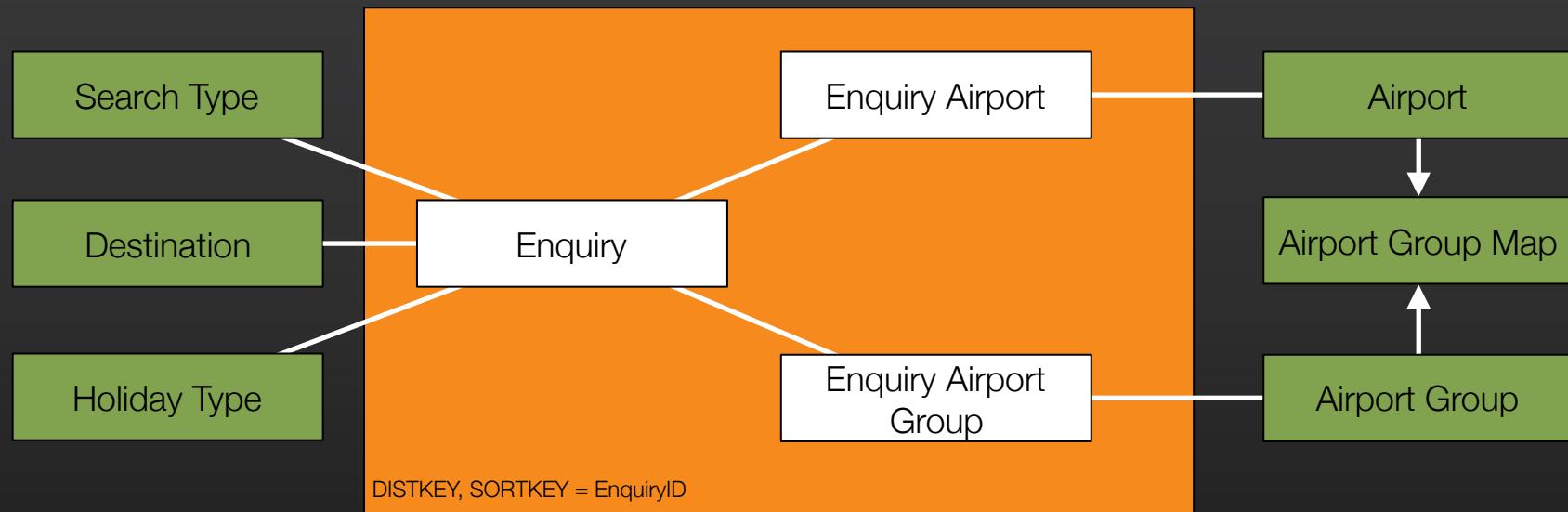
Possible Schema Optimisations

- ▶ Denormalise commonly used join attributes onto large tables



Possible Schema Optimisations

- ▶ Cache Small Tables with DISTSTYLE ALL



Vacuuming Tables

- ▶ Clean up tables after a bulk delete, a load, or a series of incremental updates with the VACUUM command, either against the entire database or against individual tables

Learning more about Table Design

- ▶ Table Design Documentation

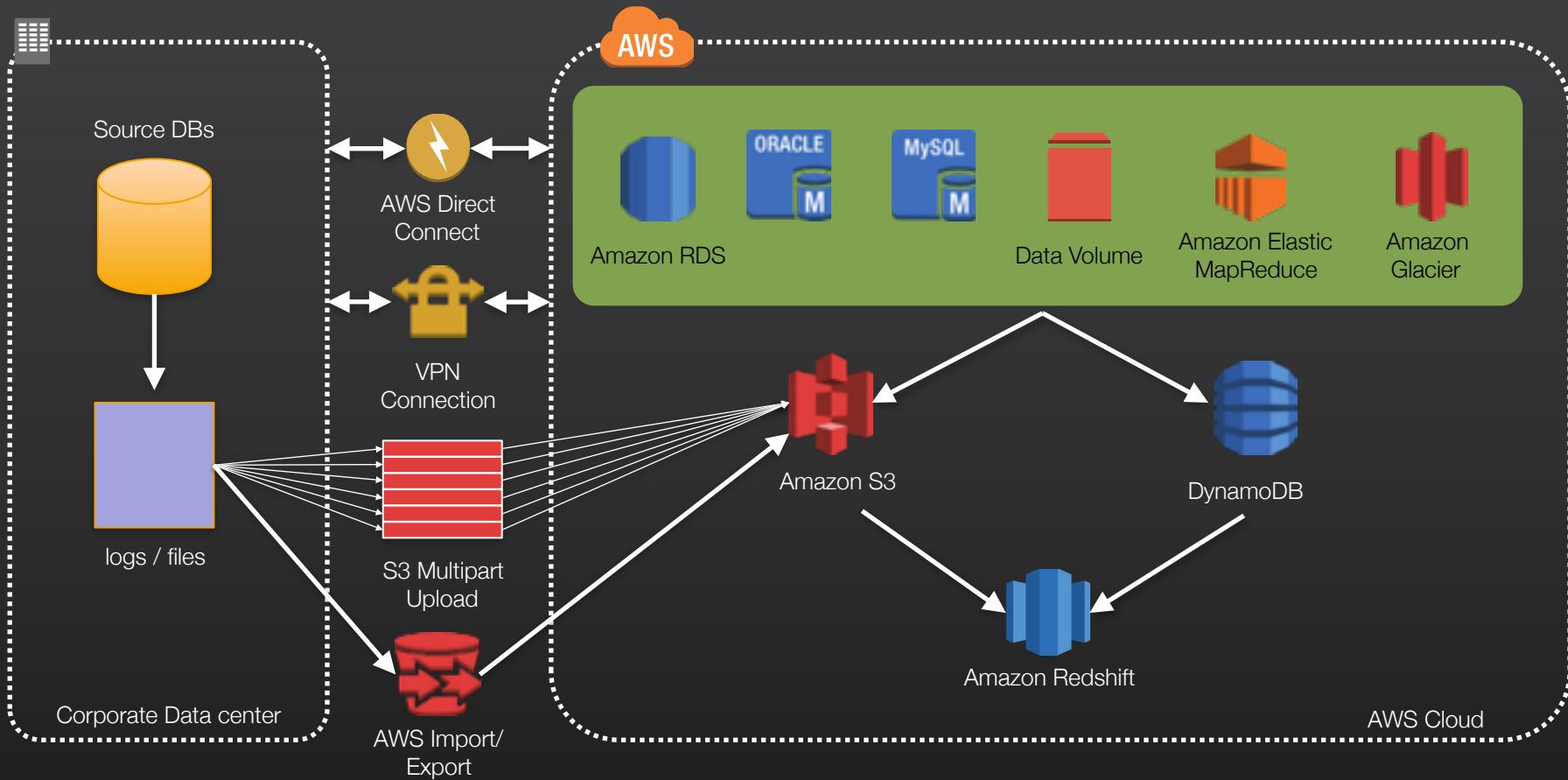
http://docs.aws.amazon.com/redshift/latest/dg/t_Creating_tables.html

- ▶ Table Design Tutorial

<http://docs.aws.amazon.com/redshift/latest/dg/tutorial-tuning-tables.html>

DATA LOADING

Amazon Redshift Loading Data Overview

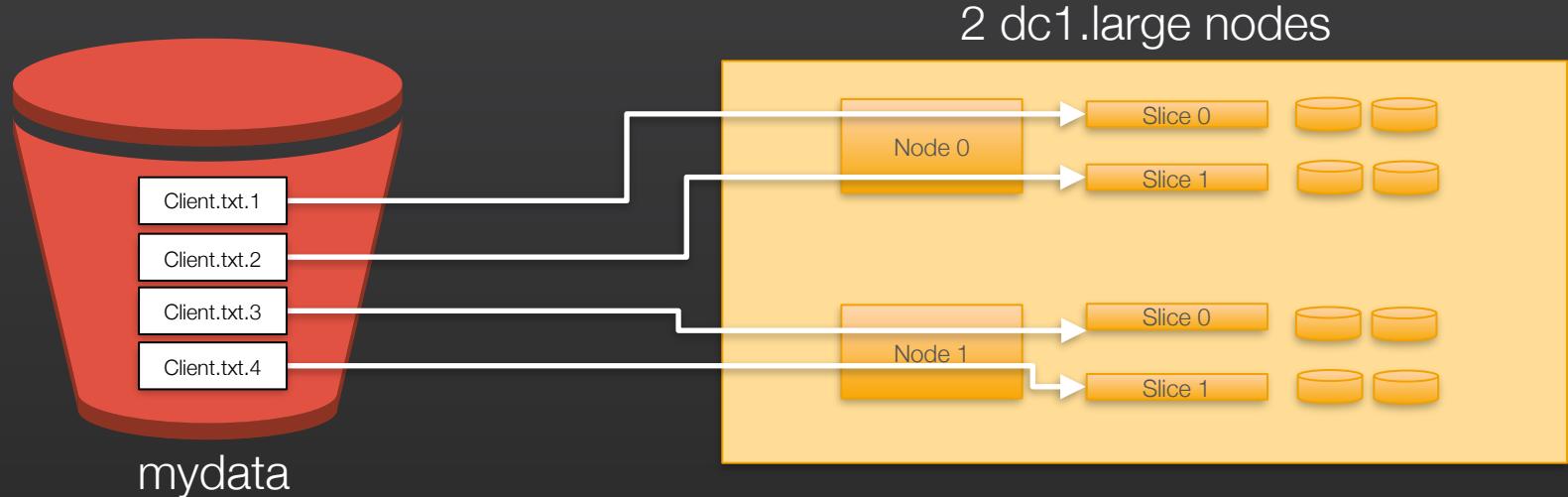


Sharing Your Data

- ▶ Data Dumped from DB
- ▶ CSV Files Encrypted
 - ▶ Review Encryption Code
- ▶ Uploaded to S3
 - ▶ Bucket <bucket>
- ▶ Data shared with AWS

```
{  
  "Id": "Policy1234567890",  
  "Statement": [  
    {  
      "Action": [  
        "s3:Get*",  
        "s3>List*"  
      ],  
      "Effect": "Allow",  
      "Principal": {  
        "AWS": ["arn:aws:iam::887210671223:root"]  
      },  
      "Resource": [  
        "arn:aws:s3::::<bucket-name>",  
        "arn:aws:s3::::<bucket-name>/*"  
      ],  
      "Sid": "Stmt0987654321"  
    }  
  ]  
}
```

Splitting Data Files



```
copy customer
from 's3://mydata/client.txt'
credentials 'aws_access_key_id=<your-access-key>; aws_secret_access_key=<your_secret_key>'
delimiter '|';
```

Typical Load Issues

- ▶ Mismatch between data types in table and values in input data fields
- ▶ Mismatch between number of columns in table and number of fields in input data
- ▶ Mismatched quotes
 - ▶ Redshift supports both single and double quotes; however, these quotes must be balanced appropriately
- ▶ Incorrect format for date/time data in input files
 - ▶ Use DATEFORMAT and TIMEFORMAT to control
- ▶ Out-of-range values in input files (for numeric columns)
- ▶ Number of distinct values for a column exceeds the limitation for its compression encoding

Direct SQL

- ▶ Redshift supports standard DML commands
 - ▶ INSERT, UPDATE, DELETE
- ▶ Redshift does not support single-command merge (upsert) statement
 - ▶ Load data into a staging table
 - ▶ Joining the staging table with the target table
 - ▶ UPDATE data where row exists
 - ▶ INSERT where no row exists
- ▶ All Direct SQL Commands go via Leader Node

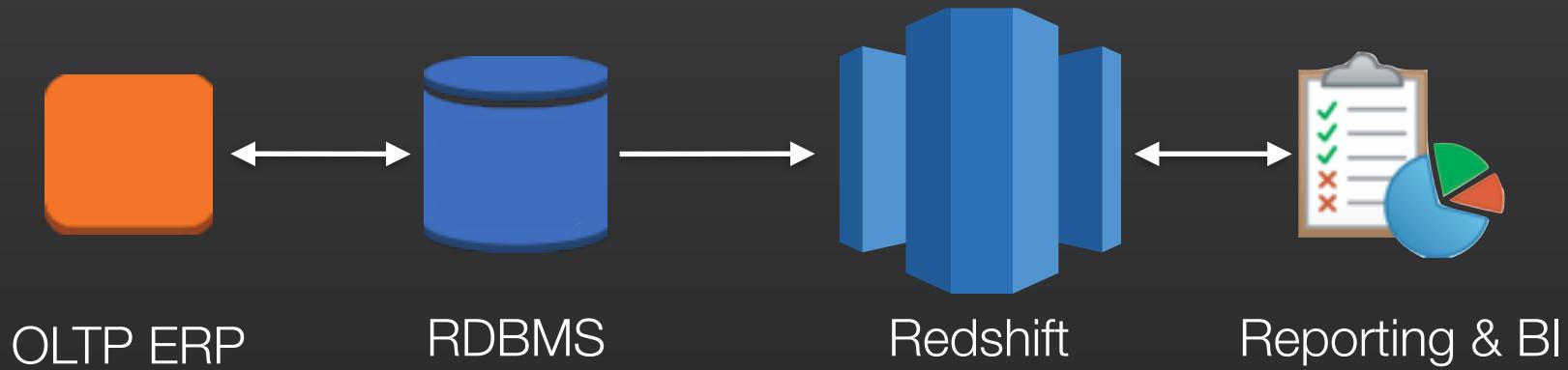
Data Loading Best Practices

- ▶ Use a COPY Command to load data
- ▶ Use a single COPY command
- ▶ Split your data into multiple files
- ▶ Compress your data files with GZIP
- ▶ Use multi-row inserts if COPY is not possible
- ▶ Bulk insert operations (INSERT INTO...SELECT and CREATE TABLE AS) provide high performance data insertion

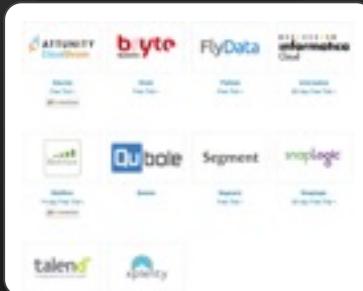
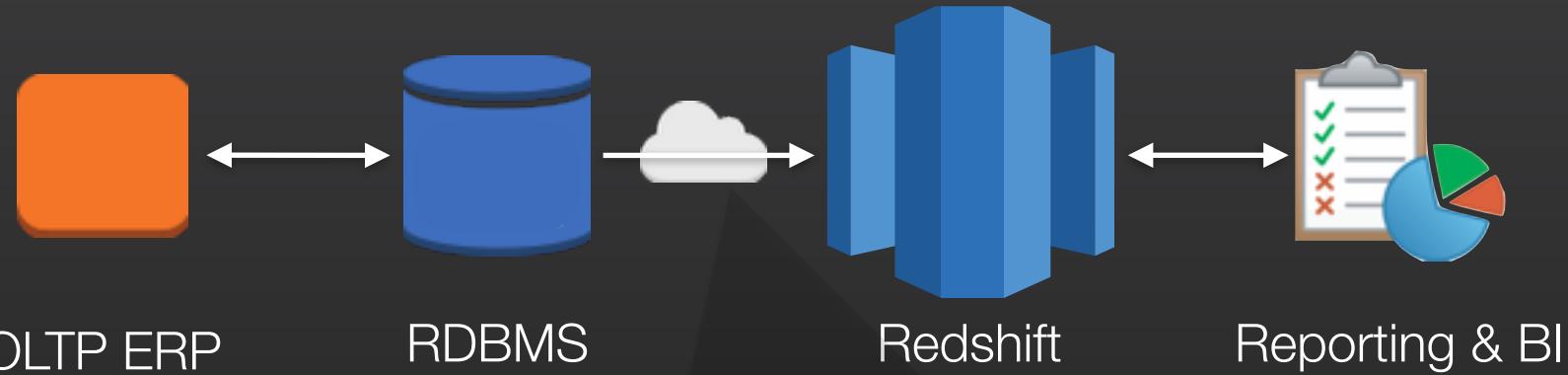
Data Loading Best Practices

- ▶ Load your data in sort key order to avoid needing to vacuum
- ▶ Organize your data as a sequence of time-series tables, where each table is identical but contains data for different time ranges
- ▶ Use staging tables to perform an upsert
- ▶ Run the VACUUM command whenever you add, delete, or modify a large number of rows, unless you load your data in sort key order
- ▶ Increase the memory available to a COPY or VACUUM by increasing wlm_query_slot_count
- ▶ Run the ANALYZE command whenever you've made a non-trivial number of changes to your data to ensure your table statistics are current

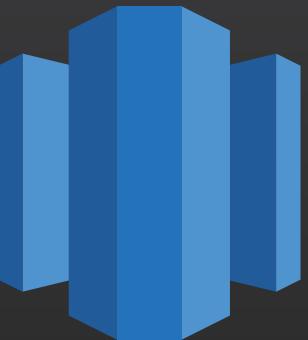
WORKING WITH DATA IN AMAZON REDSHIFT



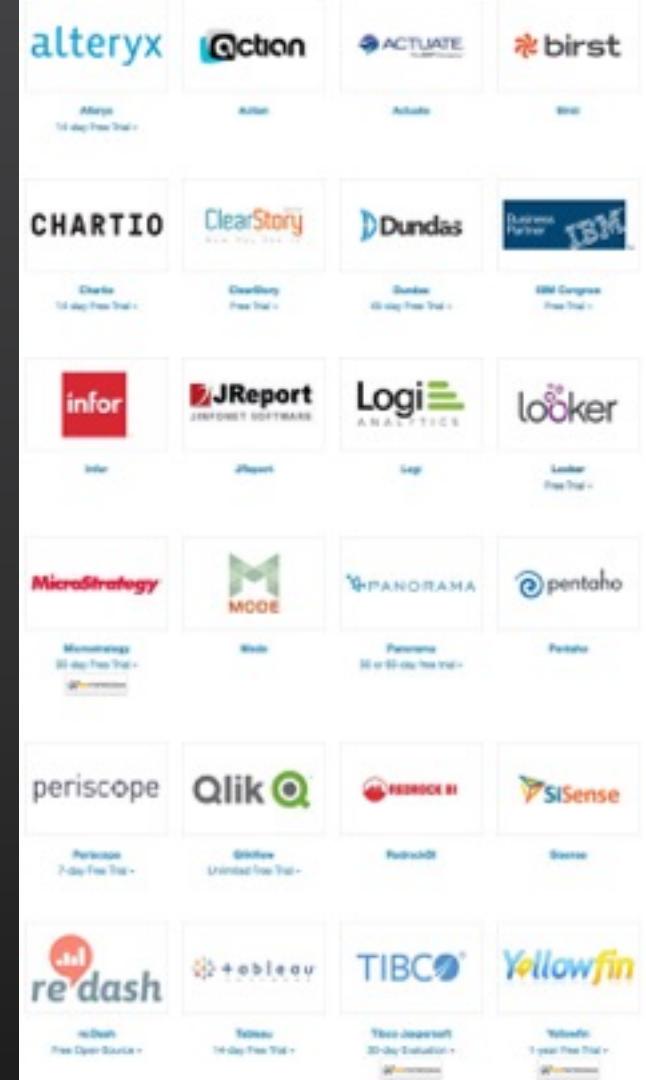
Data Integration



Business Intelligence & Visualisation



JDBC/ODBC



<https://aws.amazon.com/marketplace/redshift>

The screenshot shows the AWS Marketplace interface for the Redshift category. At the top, there's a search bar and a navigation bar with links like 'Your Recently Viewed', 'View All Products', and 'View Software'. Below the search bar, there are two tabs: 'Sort by Categories' and 'Search AWS Marketplace'. The main content area displays several product cards:

- Amazon Redshift**: A fast and powerful, fully managed, petabyte-scale data warehousing service in the cloud. Amazon Redshift offers you fast query performance while analyzing virtually any sort of data you bring to the service. SQL-based tools and business intelligence applications can use today's Redshift tables, starting with a few hundred gigabytes of data and scaling to tens of terabytes. [Read more](#).
- Featured Products**:
 - TIBCO Jaspersoft**: Bring your data closer. TIBCO Jaspersoft offers you fast query performance while analyzing virtually any sort of data you bring to the service. SQL-based tools and business intelligence applications can use today's Redshift tables, starting with a few hundred gigabytes of data and scaling to tens of terabytes. [Read more](#).
 - MicroStrategy**: [View Product Details](#)
 - ATTUNIITY CloudBeam**: [View Product Details](#)
- TIBCO Jaspersoft Reporting and Analytics for AWS**: [View Product Details](#)
- TIBCO Jaspersoft for AWS with Multi-Retention (Hourly)**: [View Product Details](#)
- TIBCO Jaspersoft for AWS Reporting and Analytics for AWS**: [View Product Details](#)
- Yellowfin**: [View Product Details](#)
- Matillion ETL for Redshift**: [View Product Details](#)
- LoadRBM**: [View Product Details](#)
- TIBCO Jaspersoft for AWS with Multi-Retention (Hourly)**: [View Product Details](#)
- TIBCO Jaspersoft Reporting and Analytics for AWS (BYOL)**: [View Product Details](#)
- JackDB**: [View Product Details](#)

Looker Analytics Platform - 10 Users, Multi-node Redshift (plus RDS)

Find it on AWS Marketplace | View Details | See product page

Looker Analytics Platform is your business's source of truth and the insights in your Redshift Data Warehouse. By looking directly at your Redshift data, Looker gives you access to highly granular and detailed reporting and visualization tools. Looker is built for business intelligence, data science, and data engineering teams. Looker provides a full suite of data engineering services from raw data ingestion analysis, and migration your data warehouse, your data team can quickly support your data engineering needs.

Customer Rating: 4.6 out of 5 stars based on 19 reviews | **Review**

Pricing Details: \$1,000/mo. **Monthly Subscription**

Amazon Redshift System Requirements:

- Compute: Amazon Lambda Image (1GB) or more required.
- Storage: 1TB Amazon S3 bucket.
- Request: AWS Lambda function.
- Additional Requirements: Amazon RDS, Amazon VPC.

Detailed Product Description:

Looker for AWS allows anyone in your business to quickly analyze and find insights in your Redshift data warehouse. Looker is built for business intelligence, data science, and data engineering teams. Looker gives you access to highly granular and detailed reporting and visualization tools. Looker provides a full suite of data engineering services from raw data ingestion analysis, and migration your data warehouse, your data team can quickly support your data engineering needs.

Product Details:

Version: 1.0.0 Available on AWS Marketplace Since: 03/02/2014 Help Amazon continue operating system software for your needs.

Resources:

Looker for AWS allows anyone in your business to quickly analyze and find insights in your Redshift data warehouse. Looker is built for business intelligence, data science, and data engineering teams. Looker provides a full suite of data engineering services from raw data ingestion analysis, and migration your data warehouse, your data team can quickly support your data engineering needs.

Usage Instructions:

Once the "Install Step" of your Lambda needs to "Running", please click your Looker URL. This will take you to the Looker dashboard where you can start exploring your data.

Support Details:

Looker Analytics Platform - 10 users, Multi-node Redshift (plus RDS)

Support for Looker can be obtained through the Looker Support site or via email.

Looker for AWS enables the Looker implementation process in a plug-and-play manner through direct integration with Amazon Redshift. Looker also provides a full suite of data engineering services from raw data ingestion analysis, and migration your data warehouse, your data team can quickly support your data engineering needs.

Tableau Server (10 users)

Find it on AWS Marketplace | View Details | See product page

Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process. Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process.

Customer Rating: 4.6 out of 5 stars based on 19 reviews | **Review**

Pricing Details: \$1,000/mo. **Monthly Subscription**

Amazon Redshift System Requirements:

- Compute: Amazon Lambda Image (1GB) or more required.
- Storage: 1TB Amazon S3 bucket.
- Request: AWS Lambda function.
- Additional Requirements: Amazon RDS, Amazon VPC.

Detailed Product Description:

Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process. Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process.

Product Details:

Version: 1.0.0 Available on AWS Marketplace Since: 03/02/2014 Help Amazon continue operating system software for your needs.

Resources:

Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process.

Usage Instructions:

Once the "Install Step" of your Lambda needs to "Running", please click your Tableau URL. This will take you to the Tableau dashboard where you can start exploring your data.

Support Details:

Tableau Server for AWS is a secure and multi-tenant visual analysis engine you can use. Pull in interactive dashboards with Tableau Studio and share them throughout your organization or across the globe. Deploying Tableau Server for AWS is a simple step-by-step process.

TIBCO Jaspersoft Reporting and Analytics for AWS

Find it on AWS Marketplace | View Details | See product page

TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards. TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards.

Customer Rating: 4.6 out of 5 stars based on 19 reviews | **Review**

Pricing Details: \$1,000/mo. **Annual Subscription**

Amazon Redshift System Requirements:

- Compute: Amazon Lambda Image (1GB) or more required.
- Storage: 1TB Amazon S3 bucket.
- Request: AWS Lambda function.
- Additional Requirements: Amazon RDS, Amazon VPC.

Detailed Product Description:

TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards. TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards.

Product Details:

Version: 1.0.0 Available on AWS Marketplace Since: 03/02/2014 Help Amazon continue operating system software for your needs.

Resources:

TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards.

Usage Instructions:

Once the "Install Step" of your Lambda needs to "Running", please click your TIBCO Jaspersoft Reporting and Analytics for AWS URL. This will take you to the TIBCO Jaspersoft Reporting and Analytics for AWS dashboard where you can start exploring your data.

Support Details:

TIBCO Jaspersoft Reporting and Analytics for AWS is a comprehensive data source reporting and analysis solution for AWS that can be deployed in the cloud. It includes a reporting and analysis interface that allows users to interactively query and analyze data from various sources, such as Amazon Redshift, Amazon S3, and Amazon Kinesis. The reporting and analysis interface provides a drag-and-drop interface for creating reports and dashboards.

AMAZON QUICKSIGHT[®]

FAST, EASY TO USE, CLOUD POWERED BUSINESS INTELLIGENCE

AVAILABLE IN PREVIEW NOW

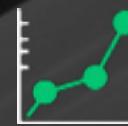
BRINGING BUSINESS INTELLIGENCE TO ALL, WITH AMAZON QUICKSIGHT



FIRST ANALYSIS IN
LESS THAN 60 SECONDS



BLAZING FAST QUERIES
WITH A NEW IN-MEMORY
QUERY ENGINE



DYNAMIC, BEAUTIFUL
DATA VISUALIZATIONS



SHARE LIVE AND
SNAPSHOT ANALYSES
WITH EVERYONE



INTEGRATE WITH
DATA SOURCES
ON AWS



1/10TH THE COST
OF OLD-GUARD BI
TOOLS

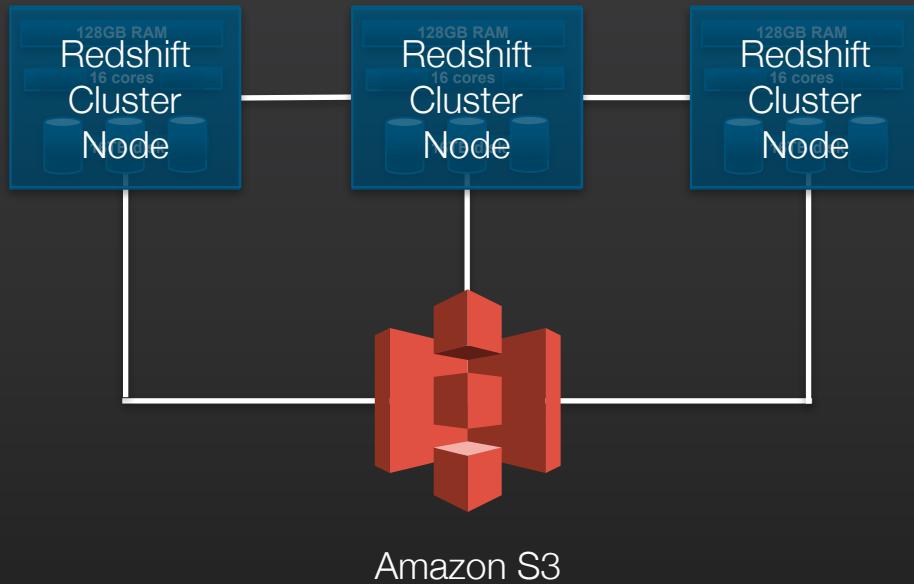
Amazon QuickSight Launch & Demo



<https://youtu.be/Tj0gW4XI6vU>

BACKUP AND RESTORATION

Backup & Restoration



Backups to Amazon S3 are automatic, continuous & incremental

Configurable system snapshot retention period

User snapshots on-demand

Streaming restores enable you to resume querying faster

AWS Services: Lambda, MPC, EC2, S3, RDS, CloudFront, Route 53, Amazon Machine Images, Amazon CloudWatch, Amazon CloudFront, Support

Clusters

Launch Cluster Manage Tags

Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
mycluster	available	healthy	no	1

Feedback English
Logout English

© 2006 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
© 2006 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use
all rights reserved. Legal Notices Customer Support

SNS S3 MPC EBS RDS CloudFront Rekognition Services Support

Clusters

Launch Cluster Manage Tags

Cluster Cluster Status DB Health In Maintenance Recent Events

AWS Services SNS MPC EBS RDS CloudFront Rekognition Services Support

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Notes Events Connect Client

Redshift Dashboard Clusters Snapshots Security Parameter Groups Reserved Notes Events Connect Client

Create Snapshot Actions

Filter: Last 7 Weeks At Cluster Filter

Meeting 1 of 1 Snapshots

Snapshot Identifier	Cluster Identifier	Status	Type	VPC	Zone	Create Time
is-mycluster-2015-10-28-09-30-12	mycluster	available	automated	us-east-1	us-west-2a	October 28, 2015 at 09:30:12 AM UTC

Snapshot Properties

Snapshot Identifier	is-mycluster-2015-10-28-09-30-12	Cluster Identifier	mycluster
Snapshot Type	automated	Cluster Version	1.0
Snapshot Status	available	Node Type	db.t1.small
Snapshot Create Time	October 28, 2015 at 09:30:12 AM UTC	Number of Nodes	1
Encrypted	No	Availability Zone	us-west-2b
		Master Username	master
		Cluster Create Time	October 28, 2015 at 09:29:38 AM UTC
		VPC	us-east-1

Access Configuration

Demand Account ID	123456789012 (This account)	Backup Details	
Accounts With Resource Access	None	Total Backup Size	12.00 MB
		Incremental Backup Size	12.00 MB
		Progress	Complete
		Backup Rate	>10.00 MB/s
		Time taken	<1m

Tags

You have not created any tags. Please add tags using the Manage Tags button above.

Feedback English

© 2006 - 2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use

about:blank To enable this feature, go to Services > Redshift > Cluster Settings > Advanced Options > Allow connections from this VPC. You must also have a VPC endpoint for your cluster.

curl to merge VPC endpoint

curl -X GET "http://127.0.0.1:5000/v1/clusters?cluster_id=1&vpc_id=1®ion=us-east-1&status=available"

curl -X GET "http://127.0.0.1:5000/v1/clusters?cluster_id=1&vpc_id=1®ion=us-east-1&status=available"

The screenshot shows the AWS Redshift Cluster Snapshot creation interface. At the top, there are tabs for Launch Cluster, Manage Tags, and Create Snapshot. The main area displays a table with one row, showing a cluster named 'hypothetical'. A context menu is open over this row, with the 'Actions' option expanded. The menu includes options like 'Review Previous Snapshot', 'Copy Automated Snapshot', 'Delete Manual Snapshot', 'Manage Access', and 'Manage Tags'. Below the table, there are sections for Snapshot Properties, Snapshot Identifier, Snapshot Type, Snapshot Status, Snapshot Create Time, and Encrypted. To the right, there are sections for Owner Configuration, Backup Details, and Progress. The progress bar indicates the backup is complete at 10.00 MB/s. At the bottom, there is a note about tags and a 'Get Started' button.

Clusters

Launch Cluster Manage Tags

Create Snapshot Actions

Review Previous Snapshot Copy Automated Snapshot Delete Manual Snapshot Manage Access Manage Tags

Snapshots

Cluster Cluster Status DB Health In Maintenance Recent Events

AWS Services VPC EC2 S3 RDS CloudFront Rep. Logos Services & Services Oregon Support

Filter: Last 2 Weeks All Clusters Filter

Viewing 1 of 1 Snapshots

Cluster Identifier: hypothetical Cluster Version: 1.0 Node Type: db.t2.small Number of Nodes: 1 Availability Zone: us-west-2b Master Username: master Cluster Create Time: October 26, 2015 at 0:30:12 AM UTC Encrypted: No

Owner Configuration

Owner Account ID: 034925704707 | View account Accounts With Restore Access: None

Backup Details

Total Backup Size: 10.00 MB Incremental Backup Size: 10.00 MB Progress: Complete Backup Rate: 10.00 MB/s Throughput: 1/1

You have not created any tags. Please add tags using the Manage Tags button above.

Get Started

The screenshot shows the AWS CloudFormation console interface for creating a new cluster from a snapshot. The main navigation bar at the top includes services like Lambda, VPC, EC2, S3, RDS, CloudFront, and Route 53.

The left sidebar navigation pane is collapsed, showing options like Clusters, Snapshots, Security, Parameter Groups, Reserved Notes, Events, and Connect Client.

The main content area displays three nested modal windows:

- Outermost Modal:** "Create Cluster" window. It shows a summary of the cluster creation process:
 - Cluster Identifier: mycluster-2019-12-28-20-12
 - Node Type: t2.small (On demand price: \$0.25 per cluster or \$0.08 per node)
- Middle Modal:** "Restore Cluster From Snapshot" window. It contains fields for specifying the cluster identifier and other configuration parameters:
 - Cluster Identifier: mycluster (Cluster already exists)
 - Port: 8000
 - Allow Version Upgrade: Yes
 - Cluster Subnet Group: default-sg-012345678901234567
 - Publicly Accessible: Yes
 - Choose an Public IP Address: No
 - Availability Zone: us-east-1a
 - Cluster Parameter Group: default.r5.2xlarge
 - VPC Security Groups: default (sg-012345678901234567)
- Innermost Modal:** "Create Cluster" window. This is a confirmation step for the cluster creation process.

At the bottom of the interface, there are footer links for Feedback, English, Logout, and Help.

Screenshot of the AWS Redshift Cluster creation process.

The main navigation bar shows services like Lambda, VPC, EC2, S3, RDS, CloudFront, Route 53, and others.

The left sidebar navigation includes:

- Clusters
- Snapshots
- Security
- Parameter Groups
- Reserved Notes
- Events
- Connect Client

The main content area shows the "Create Cluster" wizard steps:

- Create Cluster**: Step 1 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:
- Configure Cluster Options**: Step 2 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:
- Review Cluster Configuration**: Step 3 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:
- Finalize Cluster Creation**: Step 4 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:
- Review Cluster Configuration**: Step 5 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:
- Finalize Cluster Creation**: Step 6 of 6. Shows the "Cluster Identifier" field with "redshift-123" and the "Node Type" dropdown set to "M4.4xlarge".
 - Snapshot Identifier**:
 - Node Type**: M4.4xlarge On demand price: \$0.25 per cluster or \$0.25 per node
 - Cluster Identifier**:
 - Port**:
 - Allow Version Upgrade**:
 - Cluster Subnet Group**:
 - Publicly Accessible**:
 - Choose an Public IP Address**:
 - Availability Zone**:
 - Cluster Parameter Group**:
 - VPC Security Groups**:

Bottom navigation bar includes: Feedback, English, Logoff, and Help.

The image shows a composite view of the AWS Redshift console interface, illustrating the process of creating a new cluster from a snapshot.

Left Column: A sidebar navigation menu for the Redshift service, listing options like Clusters, Snapshots, Security, Parameter Groups, Reserved Notes, Events, and Connect Client.

Top Bar: The standard AWS navigation bar with links for AWS Services, Support, and Documentation.

Central Area: The main content area displays the "Clusters" page, which includes a table with columns: Cluster, Cluster Status, DB Health, In Maintenance, and Recent Events. Two clusters are listed: "mycluster" (status: available, healthy) and "second mycluster" (status: creating, healthy).

Modals and Popups:

- Create Snapshot:** A modal window showing the "Create Snapshot" button and a dropdown menu with "Recent First Snapshot" and "Copy Automated Snapshot".
- Restore Cluster From Snapshot:** A modal window titled "Restore Cluster From Snapshot" with fields for "Snapshot Identifier" (set to "mycluster") and "Node Type" (set to "Standard"). It also lists "Cluster Identifier" and "Port" (both set to their current values).
- Restore Cluster From Snapshot (Details):** A detailed modal window showing the progress of restoring the cluster. It indicates "You are creating a new cluster from a snapshot" and provides status information: "Snapshot Identifier: mycluster", "Node Type: Standard", "Cluster Identifier: mycluster", "Port: 5439", and "Publicly Accessible: No". It also lists "Availability Zone: us-east-1", "Cluster Parameter Group: mycluster", and "VPC Security Groups: mycluster".

Bottom Navigation: The standard AWS footer with links for Feedback, English, Logoff, and Help.

CLUSTER DATA REPLICATION

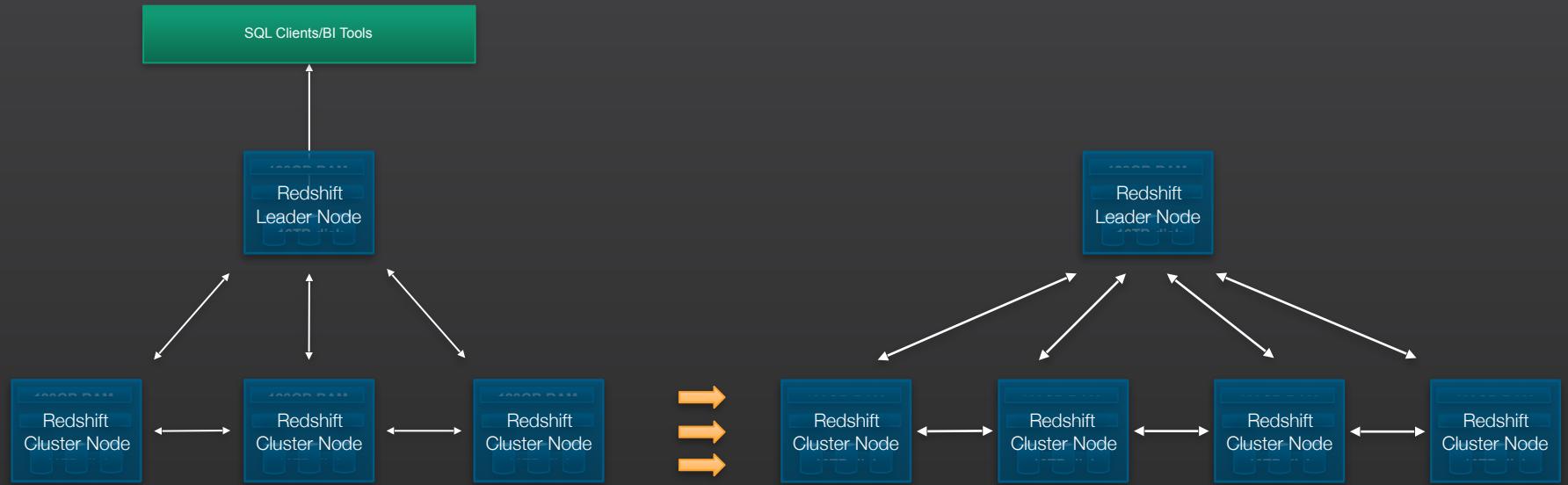
+

**AUTOMATED BACKUPS ON
AMAZON S3**

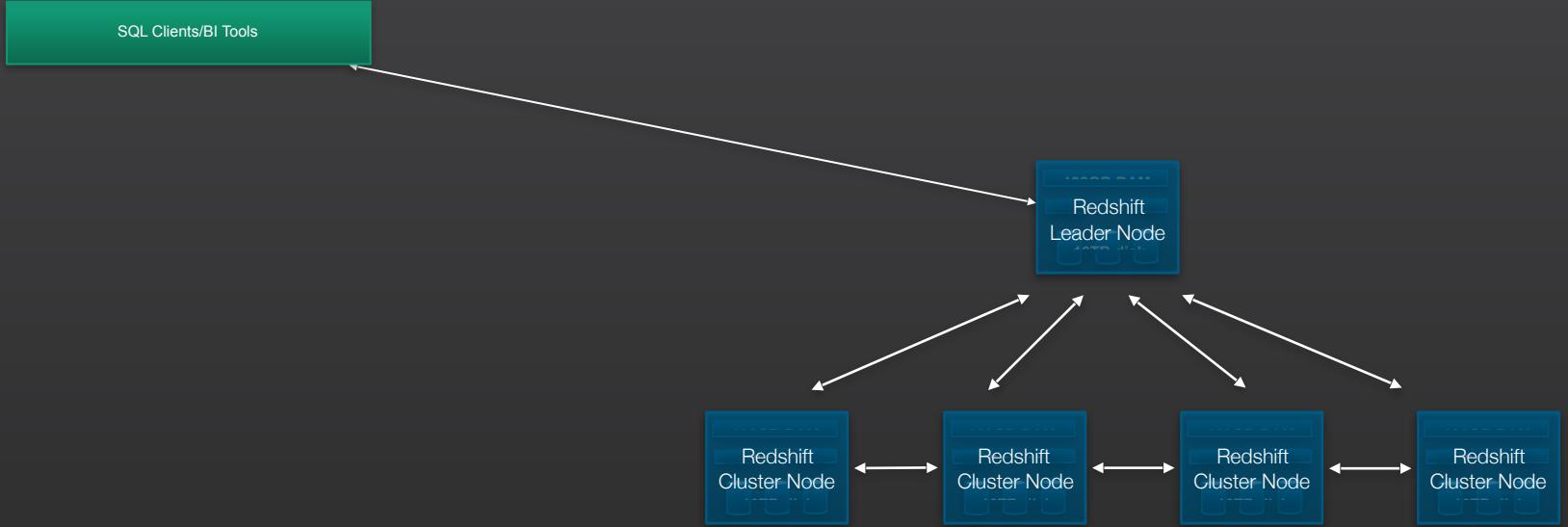
+

NODE MONITORING

UPGRADING & SCALING



Resize while remaining online
Provision a new cluster in the background
Copy data in parallel from node to node
Only charged for source cluster



Automatic SQL endpoint switchover via DNS
Decommission the source cluster

Resizing a Redshift Data Warehouse via the AWS Console

The screenshot shows the AWS Redshift console interface. The top navigation bar includes 'AWS Services', 'IAM', 'VPC', 'EC2', 'S3', 'ROS', 'CloudFront', 'File', 'Edit', 'James @ James-AWS', 'Oregon', and 'Support'. The main left sidebar has links for 'Clusters', 'Snapshots', 'Security', 'Parameter Groups', 'Reserved Nodes', 'Events', and 'Connect Client'. The main content area is titled 'Cluster: mycluster' with tabs for 'Configuration', 'Status', 'Performance', 'Queries', and 'Logs'. The 'Configuration' tab is selected. It displays the following information:

Cluster Properties		Cluster Status	
Cluster Name	mycluster	Cluster Status	available
Cluster Type	Single Node	Database Health	healthy
Node Type	dc2.large	In Maintenance Mode	No
Nodes	1	Parameter Group Apply Status	In sync
Zone	us-west-2b	Pending Modified Values	None
Created Time	October 26, 2015 at 8:39:30 AM UTC		
Cluster Version	1.0.1033		
VPC ID	vpc-0d60038f (View VPCs)		
Cluster Subnet Group	default		
VPC Security Groups	default (sg-b7ebed5) (active)		
Cluster Parameter Group	default.redshift-1.0 (in sync)		

Below this, there are sections for 'Cluster Database Properties' and 'Backup, Audit Logging, and Maintenance'.

Cluster Database Properties

Port	5439
Publicly Accessible	No
Database Name	mydatabase
Master Username	master
Encrypted	No
JDBC URL	jdbc:redshift://mycluster.us-west-2.amazonaws.com:5439/mydatabase

Backup, Audit Logging, and Maintenance

Automated Snapshot Retention Period	1
Cross-Region Snapshots Enabled	No
Audit Logging Enabled	No
Maintenance Window	wed 09:00-wed 09:30

At the bottom, there are links for 'Feedback', 'English', 'Feedback', 'Help', 'Privacy Policy', 'Terms of Use', and copyright information: © 2006–2015, Amazon Web Services, Inc. or its affiliates. All rights reserved.

Resizing a Redshift Data Warehouse via the AWS Console

The screenshot shows the AWS Redshift Cluster configuration page for a cluster named 'mycluster'. The left sidebar lists 'Clusters', 'Schemas', 'Security', 'Parameter Groups', 'Reserved Nodes', 'Events', and 'Connect Client'. The main content area displays cluster details, including:

- Cluster Name:** mycluster
- Cluster Type:** Single Node
- Node Type:** dc2.large
- Nodes:** 1
- Zone:** us-west-2b
- Created Time:** October 26, 2015 at 8:39:30 AM UTC
- Cluster Version:** 1.0.1032
- VPC ID:** vpc-0d60038f (View VPCs)
- Cluster Subnet Group:** default
- VPC Security Groups:** default (sg-b7ebed5) (active)
- Cluster Parameter Group:** default.redshift-1.0 (in sync)

Cluster Status:

- Cluster Status: available
- Database Health: healthy
- In Maintenance Mode: no
- Parameter Group Apply Status: in sync
- Pending Modified Values: none

Cluster Database Properties:

- Port: 5439
- Publicly Accessible: No
- Database Name: mydatabase
- Master Username: master
- Encrypted: No
- JDBC URL: jdbc:redshift://mycluster.cjyv1t.us-west-2.amazonaws.com:5439/

Backup, Audit Logging, and Maintenance:

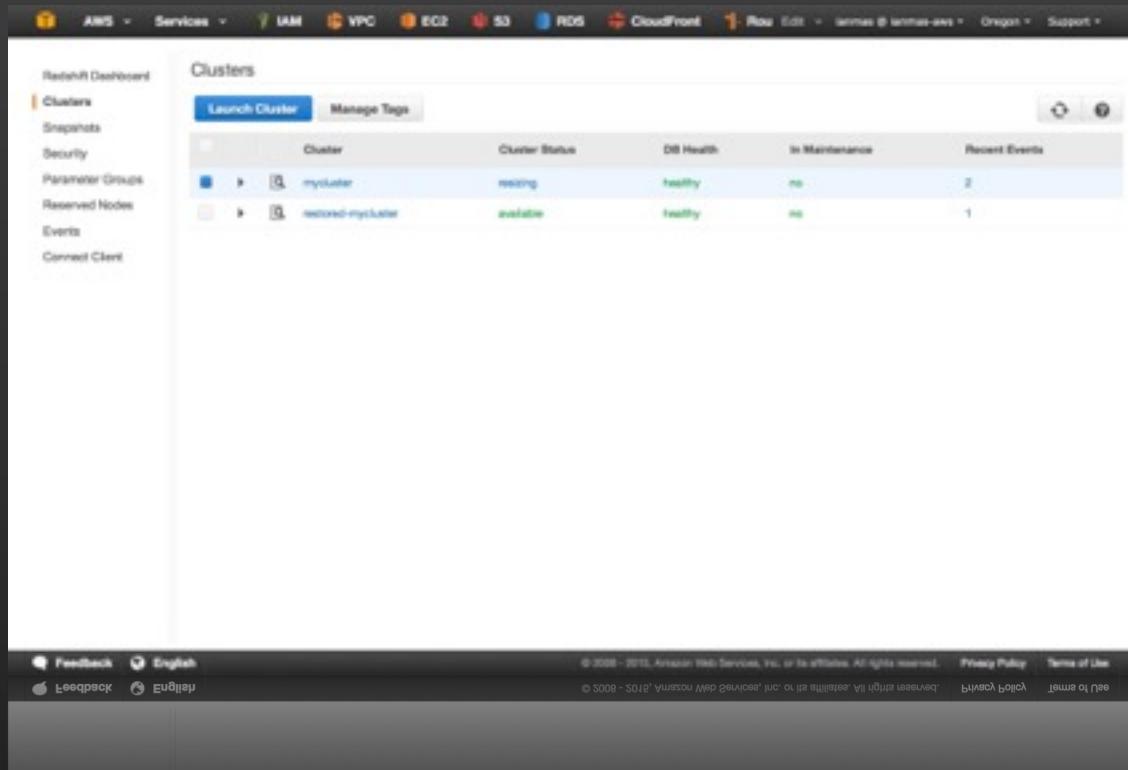
- Automated Snapshot Retention Period: 1
- Cross-Region Snapshots Enabled: No
- Audit Logging Enabled: No
- Maintenance Window: wed 09:00-wed 09:30

At the bottom, there are links for Feedback, English, Help, and Support, along with copyright information: © 2006–2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy Terms of Use.

Resizing a Redshift Data Warehouse via the AWS Console

The screenshot shows the AWS Redshift console interface. On the left, there's a sidebar with navigation links like 'Clusters', 'Schemas', 'Security', 'Parameter Groups', 'Reserved Nodes', 'Events', and 'Connect Client'. The main area displays cluster details for 'Cluster: mycluster'. A modal window titled 'Resize Cluster' is open in the center. Inside the modal, it says: 'Choose the number of nodes and optionally a new node type for the resize operation. Note that the available node type and cluster type options may be limited by the cluster's current availability zone.' Below this, there are dropdown menus for 'Node Type' (set to 'dc2.large') and 'Cluster Type' (set to 'Multi Node'), and a 'Number Of Nodes' input field with a value of '8'. A warning message in a box states: 'Warning: Resizing the cluster will cause it to be restarted into read-only mode for the duration of the resize operation. All currently executing queries and database connections on the cluster will be terminated when the resize operation begins and again when it is complete. For more information, see [Resizing a Cluster](#).' At the bottom right of the modal are 'Cancel' and 'Resize' buttons. To the right of the modal, there's a summary section with cluster status ('available'), database health ('healthy'), instance mode ('no'), step apply status ('in sync'), and modified values ('None'). At the very bottom, there's a footer with links for 'Feedback', 'English', 'Logout', and 'Help/FAQ'.

Resizing a Redshift Data Warehouse via the AWS Console



The screenshot shows the AWS Redshift Dashboard. On the left, there's a sidebar with options: Redshift Dashboard, Clusters (which is selected and highlighted in orange), Snapshots, Security, Parameter Groups, Reserved Nodes, Events, and Connect Client. The main area is titled "Clusters" and contains a table with the following data:

	Cluster	Cluster Status	DB Health	In Maintenance	Recent Events
	mycluster	missing	healthy	no	2
	restored-mycluster	available	healthy	no	1

At the bottom of the page, there are links for Feedback, English, Feedback (again), and Help. The footer contains copyright information: © 2008–2015, Amazon Web Services, Inc. or its affiliates. All rights reserved. Privacy Policy, Terms of Use, and a link to the AWS User Agreement.

<http://docs.aws.amazon.com/redshift/latest/mgmt/working-with-clusters.html#cluster-resize-intro>

RECENT FEATURES

Scalar User Defined Functions

You can write UDFs using Python 2.7

Syntax is largely identical to PostgreSQL UDF syntax
System & network calls within UDFs are prohibited

Comes with Pandas, NumPy, and SciPy pre-installed

You'll also be able import your own libraries for even more flexibility

<https://aws.amazon.com/blogs/aws/category/amazon-redshift/>

AWS Official Blog

User Defined Functions for Amazon Redshift

by Jeff Barr | on 11 SEP 2016 | in Amazon Redshift | [Permalink](#) | [Comments](#)

The Amazon Redshift team is on a tear. They are listening to customer feedback and rolling out new features all the time! Below you will find an announcement of another powerful and highly anticipated new feature.

-- Jeff;

Amazon Redshift makes it easy to launch a petabyte-scale data warehouse. For less than \$1,000/Terabyte/year, you can focus on your analytics, while Amazon Redshift manages the infrastructure for you. Amazon Redshift's price and performance has allowed customers to unlock diverse analytical use cases to help them understand their business. As you can see from blog posts by [Yelp](#), [Amplitude](#) and [Criteo](#), our customers are constantly pushing the boundaries of what's possible with data warehousing at scale.

To extend Amazon Redshift's capabilities even further and make it easier for our customers to drive new insights, I am happy to announce that Amazon Redshift has added scalar [user-defined functions](#) (UDFs). Using PostgreSQL syntax, you can now create scalar functions in Python 2.7 custom-built for your use case, and execute them in parallel across your cluster.

Here's a template that you can use to create your own functions:

sql

```
CREATE [ OR REPLACE ] FUNCTION f_function_name
( [ argument_name arg_type, ... ] )
RETURNS data_type
{ VOLATILE | STABLE | IMMUTABLE }
AS $$
    python_program
$$ LANGUAGE plpythonu;
```

Scalar UDFs return a single result value for each input value, similar to built-in scalar functions such as `ROUND` and `SUBSTRING`. Once defined, you can use UDFs in any SQL statement, just as you would use our built-in functions.

In addition to creating your own functions, you can take advantage of thousands of functions available through Python libraries to perform operations not easily expressed in SQL. You can even add custom libraries directly from S3 and the web. Out of the box, Amazon Redshift UDFs come integrated with the [Python Standard Library](#) and a number of other libraries, including:

- [NumPy](#) and [SciPy](#), which provide mathematical tools you can use to create multi-dimensional objects, do matrix operations, build optimization algorithms, and run statistical analyses.
- [Pandas](#), which offers high level data manipulation tools built on top of NumPy and SciPy, and that enables you to perform data analysis or an end-to-end modeling workflow.
- [Dateutil](#) and [Pytz](#), which make it easy to manipulate dates and time zones (such as figuring out how many months are left before the

New DS2 Instances

DS2 instances have twice the memory and compute power of their Dense Storage predecessor, DS1 (previously called DW1), but the same storage.

DS2 also supports Enhanced Networking and provides 50% more disk throughput than DS1.

AWS Official Blog

Amazon Redshift – Now Faster and More Cost-Effective than Ever

by Jeff Barr | on 09 JUN 2015 | in Amazon Redshift | Permalink

My colleague Tina Adams sent me a guest post to share news of a new instance type and new Reserved Instance offerings for [Amazon Redshift](#).

— Jeff;

Amazon Redshift makes analyzing petabyte-scale data fast, cheap, and simple. It delivers advanced technology capabilities, including parallel execution, compressed columnar storage, and end-to-end encryption, as a fully managed service, letting you focus on your data not your database. All for less than \$1,000/TB/YR. When launching a cluster, you can choose between our Dense Compute (SSD) and Dense Storage (HDD) instance families.

Today, we are making our Dense Storage family even faster and more cost effective with a second-generation instance type, DS2. Moreover, you can now reserve all dense storage and dense compute instances types for one year with No Upfront payment, and receive a 20% discount over On-Demand rates. For steeper discounts, you can pay for your entire reserved instance term with one All Upfront payment.

New DS2 Instances

DS2 instances have twice the memory and compute power of their Dense Storage predecessor, DS1 (previously called DW1), but the same storage. DS2 also supports Enhanced Networking and provides 50% more disk throughput than DS1. On average, DS2 provides 50% better performance than DS1, but is priced exactly the same.

Instance	vCPU	Memory (GiB)	Network	Storage	I/O	Price/TB/Year (On Demand)	Price/TB/Year (3 Year RI)
Dense Storage – Current Generation							
ds2.xlarge	4	31	Enhanced	2 TB HDD	0.50 Gbps	\$3,330	\$999
ds2.8xlarge	36	244	Enhanced - 10 Gbps	16 TB HDD	4.00 Gbps	\$3,330	\$999
Dense Storage – Previous Generation (formerly DW1)							
ds1.xlarge	2	15	Moderate	2 TB HDD	0.30 Gbps	\$3,330	\$999
ds1.8xlarge	16	120	10 Gbps	16 TB HDD	2.40 Gbps	\$3,330	\$999

**RESOURCES YOU CAN USE
TO LEARN MORE**

aws.amazon.com/redshift

Getting Started with Amazon Redshift

aws.amazon.com/redshift/getting-started/

Customers running Analytics in the AWS Cloud

<http://aws.amazon.com/solutions/case-studies/analytics/>

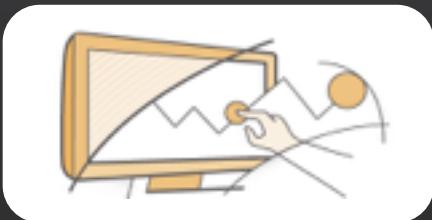
AWS re:Invent 2015 | Amazon Redshift Deep Dive: Tuning & Best Practices

<https://youtu.be/fmy3jCxUliM?list=PLhr1KZpdzukdsbIOEVXrCYtvUsDakzYJl>



AWS Training & Certification

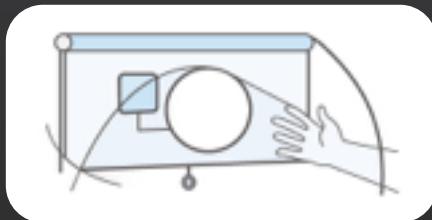
Self-Paced Labs



Try products, gain new skills, and get hands-on practice working with AWS technologies

[aws.amazon.com/training/
self-paced-labs](https://aws.amazon.com/training/self-paced-labs)

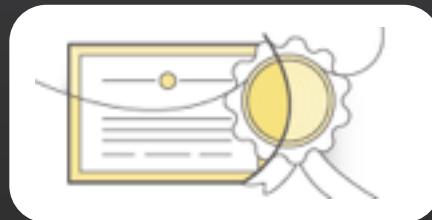
Training



Build technical expertise to design and operate scalable, efficient applications on AWS

aws.amazon.com/training

Certification



Validate your proven skills and expertise with the AWS platform

aws.amazon.com/certification

Follow us for more
events & webinars



Ian Massingham — Technical Evangelist

 @ianMmmm



@AWS_UK1 for local AWS events & news



@AWScloud for Global AWS News & Announcements