



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

General

Q: What is Amazon Redshift?

Amazon Redshift is a fast, fully managed data warehouse that makes it simple and cost-effective to analyze all your data using standard SQL and your existing Business Intelligence (BI) tools. It allows you to run complex analytic queries against petabytes of structured data, using sophisticated query optimization, columnar storage on high-performance local disks, and massively parallel query execution. Most results come back in seconds. With Redshift, you can start small for just \$0.25 per hour with no commitments and scale out to petabytes of data for \$1,000 per terabyte per year, less than a tenth the cost of traditional solutions. Amazon Redshift also includes Amazon Redshift Spectrum, allowing you to directly run SQL queries against exabytes of unstructured data in Amazon S3. No loading or transformation is required, and you can use open data formats, including Avro, CSV, Grok, Ion, JSON, ORC, Parquet, RCFile, RegexSerDe, SequenceFile, TextFile, and TSV. Redshift Spectrum automatically scales query compute capacity based on the data being retrieved, so queries against Amazon S3 run fast, regardless of data set size.

Traditional data warehouses require significant time and resource to administer, especially for large datasets. In addition, the financial cost associated with building, maintaining, and growing self-managed, on-premise data warehouses is very high. As your data grows, you have to constantly trade-off what data to load into your data warehouse and what data to archive in storage so you can manage costs, keep ETL complexity low, and deliver good performance. Amazon Redshift not



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

As with all Amazon Web Services, there are no up-front investments required, and you pay only for the resources you use. Amazon Redshift lets you pay as you go. You can even [try Amazon Redshift for free](#).

For information about Amazon Redshift regional availability, see the [AWS Region Table](#).

[Show less](#)

Q: What is Redshift Spectrum?

Redshift Spectrum is a feature of Amazon Redshift that enables you to run queries against exabytes of unstructured data in [Amazon S3](#), with no loading or ETL required. When you issue a query, it goes to the Amazon Redshift SQL endpoint, which generates and optimizes a query plan. Amazon Redshift determines what data is local and what is in Amazon S3, generates a plan to minimize the amount of Amazon S3 data that needs to be read, requests Redshift Spectrum workers out of a shared resource pool to read and process data from Amazon S3.

Redshift Spectrum scales out to thousands of instances if needed, so queries run quickly regardless of data size. And, you can use the exact same SQL for Amazon S3 data as you do for your Amazon Redshift queries today and connect to the same Amazon Redshift endpoint using your same BI tools. Redshift Spectrum lets you separate storage and compute, allowing you to scale each independently. You can setup as many Amazon Redshift clusters as you need to query your Amazon S3 data lake, providing high availability and limitless concurrency. Redshift Spectrum gives



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Amazon S3.

[Show less](#)

Q: How does the performance of Amazon Redshift compare to most traditional databases for data warehousing and analytics?

Amazon Redshift uses a variety of innovations to achieve up to ten times higher performance than traditional databases for data warehousing and analytics workloads:

- *Columnar Data Storage:* Instead of storing data as a series of rows, Amazon Redshift organizes the data by column. Unlike row-based systems, which are ideal for transaction processing, column-based systems are ideal for data warehousing and analytics, where queries often involve aggregates performed over large data sets. Since only the columns involved in the queries are processed and columnar data is stored sequentially on the storage media, column-based systems require far fewer I/Os, greatly improving query performance.
- *Advanced Compression:* Columnar data stores can be compressed much more than row-based data stores because similar data is stored sequentially on disk. Amazon Redshift employs multiple compression techniques and can often achieve significant compression relative to traditional relational data stores. In addition, Amazon Redshift doesn't require indexes or materialized views and so uses less space than traditional relational database systems. When loading data into an empty table, Amazon Redshift automatically samples your data and selects the most appropriate compression scheme.



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Q: How do I get started with Amazon Redshift?

You can sign up and get started within minutes from the Amazon Redshift detail page or via the [AWS Management Console](#). If you don't already have an AWS account, you'll be prompted to create one.

To use Redshift Spectrum, you need to first store your data in Amazon S3. You can then define the metadata about that data in your Amazon Redshift cluster or register the metadata you may already have in your Hive metastore with your cluster. You can issue a CREATE EXTERNAL SCHEMA SQL command in your Amazon Redshift cluster to define or register a database in your catalog as an external schema within Amazon Redshift. You can then issue queries against Amazon S3 using the same SQL you use for local tables and any BI tool that supports Amazon Redshift today. The external database definition you create using Amazon Redshift SQL is registered in the same data catalog that Amazon Athena uses. You can optionally manage the external database definition from the Amazon Athena Catalog as well.

Visit our [Getting Started](#) page to see how to try Amazon Redshift for free.

[Show less](#)

Q: How do I create and access an Amazon Redshift data warehouse cluster?

You can easily create an Amazon Redshift data warehouse cluster by using the [AWS Management Console](#) or the [Amazon Redshift APIs](#). You can start with a single node, 160GB data warehouse and scale all the way to a petabyte or more with a few clicks in the AWS Console or a single API call.



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

connection string from the [AWS Management Console](#) or by using the [Redshift APIs](#). You can then use this connection string with your favorite database tool, programming language, or Business Intelligence (BI) tool. You will need to authorize network requests to your running data warehouse cluster. For a detailed explanation please refer to our [Getting Started Guide](#).

[Show less](#)

Q: What does a leader node do? What does a compute node do?

A leader node receives queries from client applications, parses the queries and develops execution plans, which are an ordered set of steps to process these queries. The leader node then coordinates the parallel execution of these plans with the compute nodes, aggregates the intermediate results from these nodes and finally returns the results back to the client applications.

Compute nodes execute the steps specified in the execution plans and transmit data among themselves to serve these queries. The intermediate results are sent back to the leader node for aggregation before being sent back to the client applications.

[Show less](#)

Q: What is the maximum storage capacity per compute node? What is the recommended amount of data per compute node for optimal performance?



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

storage, 2 Intel Xeon E5-2670v2 (Ivy Bridge) virtual cores and 15GiB of RAM. The Eight Extra Large is sixteen times bigger with 2.56TB of SSD storage, 32 Intel Xeon E5-2670v2 virtual cores and 244GiB of RAM. You can get started with a single DC2.Large node for \$0.25 per hour and scale all the way up to 128 8XL nodes with 326TB of SSD storage, 3,200 virtual cores and 24TiB of RAM.

Amazon Redshift's MPP architecture means you can increase your performance by increasing the number of nodes in your data warehouse cluster. The optimal amount of data per compute node depends on your application characteristics and your query performance needs. An Amazon Redshift data warehouse cluster can contain from 1-128 compute nodes, depending on the node type. For details please see our [documentation](#).

[Show less](#)

Q: When would I use Amazon Redshift vs. Amazon RDS?

Both Amazon Redshift and [Amazon RDS](#) enable you to run traditional relational databases in the cloud while offloading database administration. Customers use Amazon RDS databases both for online-transaction processing (OLTP) and for reporting and analysis. Amazon Redshift harnesses the scale and resources of multiple nodes and uses a variety of optimizations to provide order of magnitude improvements over traditional databases for analytic and reporting workloads against very large data sets. Amazon Redshift provides an excellent scale-out option as your data and query complexity grows or if you want to prevent your reporting and analytic processing from interfering with the performance of your OLTP workload.

[Show less](#)



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

While Redshift Spectrum is great for running queries against data in Amazon Redshift and S3, it really isn't a fit for the types of use cases that enterprises typically ask from processing frameworks like Amazon EMR. Amazon EMR goes far beyond just running SQL queries. Amazon EMR is a managed service that lets you process and analyze extremely large data sets using the latest versions of popular big data processing frameworks, such as Spark, Hadoop, and Presto, on fully customizable clusters. With Amazon EMR you can run a wide variety of scale-out data processing tasks for applications such as machine learning, graph analytics, data transformation, streaming data, and virtually anything you can code.

You can use Redshift Spectrum together with EMR. Redshift Spectrum uses the same approach to store table definitions as Amazon EMR. Redshift Spectrum can support the same Apache Hive Metastore used by Amazon EMR to locate data and table definitions. If you're using Amazon EMR and have a Hive Metastore already, you just have to configure your Amazon Redshift cluster to use it. You can then start querying that data right away along with your Amazon EMR jobs. So, if you're already using EMR to process a large data store, you can use Redshift Spectrum to query that data right at the same time without interfering with your Amazon EMR jobs.

Query services, data warehouses, and complex data processing frameworks all have their place, and they are used for different things. You just need to choose the right tool for the job.

[Show less](#)

Q: When should I use Amazon Athena vs. Redshift Spectrum?



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Q. Why should I use Amazon Redshift instead of running my own MPP data warehouse cluster on Amazon EC2?

Amazon Redshift automatically handles many of the time-consuming tasks associated with managing your own data warehouse including:

- *Setup:* With Amazon Redshift, you simply create a data warehouse cluster, define your schema, and begin loading and querying your data. Provisioning, configuration and patching are all managed for you.
- *Data Durability:* Amazon Redshift replicates your data within your data warehouse cluster and continuously backs up your data to Amazon S3, which is designed for eleven nines of durability. Amazon Redshift mirrors each drive's data to other nodes within your cluster. If a drive fails, your queries will continue with a slight latency increase while Redshift rebuilds your drive from replicas. In case of node failure(s), Amazon Redshift automatically provisions new node(s) and begins restoring data from other drives within the cluster or from Amazon S3. It prioritizes restoring your most frequently queried data so your most frequently executed queries will become performant quickly.
- *Scaling:* You can add or remove nodes from your Amazon Redshift data warehouse cluster with a single API call or via a few clicks in the [AWS Management Console](#) as your capacity and performance needs change.
- *Automatic Updates and Patching:* Amazon Redshift automatically applies upgrades and patches your data warehouse so you can focus on your application and not on its administration.



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

data warehouse cluster as soon as the data warehouse cluster is available. Billing continues until the data warehouse cluster terminates, which would occur upon deletion or in the event of instance failure. You are billed based on:

- *Compute node hours:* Compute node hours are the total number of hours you run across all your compute nodes for the billing period. Node usage hours are billed for each hour your data warehouse cluster is running in an available state. If you no longer wish to be charged for your data warehouse cluster, you must terminate it to avoid being billed for additional node hours. Partial node hours consumed are billed as full hours. You are billed for 1 unit per node per hour, so a 3-node data warehouse cluster running persistently for an entire month would incur 2,160 instance hours. You will not be charged for leader node hours; only compute nodes will incur charges.
- *Backup Storage:* Backup storage is the storage associated with your automated and manual snapshots for your data warehouse. Increasing your backup retention period or taking additional snapshots increases the backup storage consumed by your data warehouse. There is no additional charge for backup storage up to 100% of your provisioned storage for an active data warehouse cluster. For example, if you have an active Single Node XL data warehouse cluster with 2TB of local instance storage, we will provide up to 2TB-Month of backup storage at no additional charge. Backup storage beyond the provisioned storage size and backups stored after your cluster is terminated are billed at [standard Amazon S3 rates](#).
- *Data transfer:* There is no data transfer charge for data transferred to or from Amazon Redshift and Amazon S3 within the same AWS Region. For all other data transfers into and out of Amazon Redshift, you will be billed at standard AWS data transfer rates.



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Data Integration and Loading

Q: How do I load data into my Amazon Redshift data warehouse?

You can load data into Amazon Redshift from a range of data sources including [Amazon S3](#), [Amazon DynamoDB](#), [Amazon EMR](#), [AWS Glue](#), [AWS Data Pipeline](#) and or any SSH-enabled host on Amazon EC2 or on-premises. Amazon Redshift attempts to load your data in parallel into each compute node to maximize the rate at which you can ingest data into your data warehouse cluster. For more details on loading data into Amazon Redshift please view our [Getting Started Guide](#).

Yes, clients can connect to Amazon Redshift using ODBC or JDBC and issue 'insert' SQL commands to insert the data. Please note this is slower than using S3 or DynamoDB since those methods load data in parallel to each compute node while SQL insert statements load via the single leader node.

[Show less](#)

Q: How do I load data from my existing Amazon RDS, Amazon EMR, Amazon DynamoDB, and Amazon EC2 data sources to Amazon Redshift?

You can use our [COPY command](#) to load data in parallel directly to Amazon Redshift from Amazon EMR, Amazon DynamoDB, or any SSH-enabled host. Redshift Spectrum also enables you to load



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Q: I have a lot of data for initial loading into Amazon Redshift. Transferring via the Internet would take a long time. How do I load this data?

You can use [AWS Import/Export](#) to transfer the data to Amazon S3 using portable storage devices. In addition, you can use [AWS Direct Connect](#) to establish a private network connection between your network or datacenter and AWS. You can choose 1Gbit/sec or 10Gbit/sec connection ports to transfer your data.

[Show less](#)

Security

Q: How does Amazon Redshift keep my data secure?

Amazon Redshift encrypts and keeps your data secure in transit and at rest using industry-standard encryption techniques. To keep data secure in transit, Amazon Redshift supports SSL-enabled connections between your client application and your Redshift data warehouse cluster. To keep your data secure at rest, Amazon Redshift encrypts each block using hardware-accelerated AES-256 as it is written to disk. This takes place at a low level in the I/O subsystem, which encrypts everything written to disk, including intermediate query results. The blocks are backed up as is, which means that backups are encrypted as well. By default, Amazon Redshift takes care of key



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

You can use Redshift Spectrum with an Amazon Redshift cluster that is part of your VPC. Note that Redshift Spectrum does not currently support [Enhanced VPC Routing](#).

[Show less](#)

Q: Can I access my Amazon Redshift compute nodes directly?

No. Your Amazon Redshift compute nodes are in a private network space and can only be accessed from your data warehouse cluster's leader node. This provides an additional layer of security for your data.

[Show less](#)

Availability and Durability

Q: What happens to my data warehouse cluster availability and data durability if a drive on one of my nodes fails?

Your Amazon Redshift data warehouse cluster will remain available in the event of a drive failure however you may see a slight decline in performance for certain queries. In the event of a drive failure, Amazon Redshift will transparently use a replica of the data on that drive which is stored on



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

the event of a drive failure you will need to restore the cluster from snapshot on S3. We recommend using at least two nodes for production.

[Show less](#)

Q: What happens to my data warehouse cluster availability and data durability if my data warehouse cluster's Availability Zone (AZ) has an outage?

If your Amazon Redshift data warehouse cluster's Availability Zone becomes unavailable, you will not be able to use your cluster until power and network access to the AZ are restored. Your data warehouse cluster's data is preserved so you can start using your Amazon Redshift data warehouse as soon as the AZ becomes available again. In addition, you can also choose to restore any existing snapshots to a new AZ in the same Region. Amazon Redshift will restore your most frequently accessed data first so you can resume queries as quickly as possible.

[Show less](#)

Q: Does Amazon Redshift support Multi-AZ Deployments?

Currently, Amazon Redshift only supports Single-AZ deployments. You can run data warehouse clusters in multiple AZ's by loading data into two Amazon Redshift data warehouse clusters in separate AZs from the same set of Amazon S3 input files. With Redshift Spectrum, you can spin up multiple clusters across AZs and access data in Amazon S3 without having to load it into your



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

35). Redshift can also asynchronously replicate your snapshots to S3 in another region for disaster recovery.

By default, Amazon Redshift enables automated backups of your data warehouse cluster with a 1-day retention period. You can configure this to be as long as 35 days.

Free backup storage is limited to the total size of storage on the nodes in the data warehouse cluster and only applies to active data warehouse clusters. For example, if you have total data warehouse storage of 8TB, we will provide at most 8TB of backup storage at no additional charge. If you would like to extend your backup retention period beyond one day, you can do so using the [AWS Management Console](#) or the [Amazon Redshift APIs](#). For more information on automated snapshots, please refer to the Amazon Redshift Management Guide. Amazon Redshift only backs up data that has changed so most snapshots only use up a small amount of your free backup storage.

When you need to restore a backup, you have access to all the automated backups within your backup retention window. Once you choose a backup from which to restore, we will provision a new data warehouse cluster and restore your data to it.

[Show less](#)

Q: How do I manage the retention of my automated backups and snapshots?

You can use the [AWS Management Console](#) or [ModifyCluster](#) API to manage the period of time your automated backups are retained by modifying the RetentionPeriod parameter. If you desire to



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Scalability

Q: How do I scale the size and performance of my Amazon Redshift data warehouse cluster?

If you would like to increase query performance or respond to CPU, memory or I/O over-utilization, you can increase the number of nodes within your data warehouse cluster via the [AWS Management Console](#) or the [ModifyCluster](#) API. When you modify your data warehouse cluster, your requested changes will be applied immediately. Metrics for compute utilization, storage utilization, and read/write traffic to your Amazon Redshift data warehouse cluster are available free of charge via the [AWS Management Console](#) or Amazon CloudWatch APIs. You can also add additional, user-defined metrics via [Amazon Cloudwatch](#) custom metric functionality.

With Redshift Spectrum, you can run multiple Amazon Redshift clusters accessing the same data in Amazon S3. You can use different clusters for different use cases. For example, you can use one cluster for standard reporting and another for data science queries. Your marketing team can use their own clusters different from your operations team. Depending on the type and number of nodes in your local cluster, and the number of files need to be processed for your query, Redshift Spectrum automatically distributes the execution of your query to several Redshift Spectrum workers out of a shared resource pool to read and process data from Amazon S3, and pulls results back into your Amazon Redshift cluster for any remaining processing.

[Show less](#)



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Querying and Analytics

Q: Are Amazon Redshift and Redshift Spectrum compatible with my preferred business intelligence software package and ETL tools?

Amazon Redshift uses industry-standard SQL and is accessed using standard JDBC and ODBC drivers. You can download Amazon Redshift custom JDBC and ODBC drivers from the Connect Client tab of the [Redshift Console](#). We have validated integrations with popular [BI and ETL vendors](#), a number of which are offering [free trials](#) to help you get started loading and analyzing your data. You can also go to the [AWS Marketplace](#) to deploy and configure solutions designed to work with Amazon Redshift in minutes.

Redshift Spectrum supports all Amazon Redshift client tools. The client tools can continue to connect to the Amazon Redshift cluster endpoint using ODBC or JDBC connections. No changes are required.

You use exactly the same query syntax and have the same query capabilities to access tables in Redshift Spectrum as you have for tables in the local storage of your Redshift cluster. External tables are referenced using the schema name defined in the CREATE EXTERNAL SCHEMA command where they were registered.

[Show less](#)



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Q: I use a Hive Metastore to store metadata about my S3 data lake. Can I use Redshift Spectrum?

Yes. The CREATE EXTERNAL SCHEMA command supports Hive Metastores. We do not currently support DDL against the Hive Metastore.

[Show less](#)

Q: How do I get a list of all external database tables created in my cluster?

You can query the system table SVV_EXTERNAL_TABLES to get that information.

[Show less](#)

Monitoring

Q: How do I monitor the performance of my Amazon Redshift data warehouse cluster?

Metrics for compute utilization, storage utilization, and read/write traffic to your Amazon Redshift data warehouse cluster are available free of charge via the [AWS Management Console](#) or [Amazon](#)



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

a lot of Redshift Spectrum workers and complete them quickly.

[Show less](#)

Maintenance

Q: What is a maintenance window? Will my data warehouse cluster be available during software maintenance?

Amazon Redshift periodically performs maintenance to apply fixes, enhancements and new features to your cluster. You can change the scheduled maintenance windows by modifying the cluster, either programmatically or by using the [Redshift Console](#). During these maintenance windows, your Amazon Redshift cluster is not available for normal operations. For more information about maintenance windows and schedules by region, see [Maintenance Windows](#) in the Amazon Redshift Management Guide.

[Show less](#)



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

understanding of the AWS Cloud



WHAT'S NEW WITH AWS

Learn about the latest products, services, and more



Sign In to the Console

[Twitter](#) [Facebook](#) [Podcast](#) [Twitch](#) [AWS Blog](#) [RSS News Feed](#)
 [Email Updates](#)

AWS & Cloud Computing

[What is Cloud Computing?](#)

[What is Caching?](#)

[What is NoSQL?](#)

[What is DevOps?](#)

[What is Docker?](#)

[Products & Services](#)

[Customer Success](#)

[Economics Center](#)

[Architecture Center](#)

[Security Center](#)

Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

Serverless Computing

Big Data

High Performance Computing

Mobile Services

Digital Marketing

Game Development

Digital Media

Government & Education

Health

Financial Services

Windows on AWS

Retail

Power & Utilities

Oil & Gas

Automotive

Blockchain

Manufacturing

Resources & Training

Developers

Java on AWS

JavaScript on AWS

Mobile on AWS

PHP on AWS

Python on AWS

Ruby on AWS

.NET on AWS

SDKs & Tools



Amazon Redshift ▾

Overview

Features

Pricing

Getting Started ▾

Resources ▾

FAQs

Migrations

[Security Credentials](#)

[Request Service Limit Increases](#)

[Contact Us](#)

Amazon Web Services is Hiring.

Amazon Web Services (AWS) is a dynamic, growing business unit within Amazon.com. We are currently hiring Software Development Engineers, Product Managers, Account Managers, Solutions Architects, Support Engineers, System Engineers, Designers and more. Visit our [careers](#) page to learn more.

Amazon.com is an Equal Opportunity-Affirmative Action Employer – Minority / Female / Disability / Veteran / Gender Identity / Sexual Orientation.

Language [Bahasa Indonesia](#) | [Deutsch](#) | [English](#) | [Español](#) | [Français](#) | [Italiano](#) | [Português](#) | [Tiếng Việt](#) | [Türkçe](#)
[Русский](#) | [ไทย](#) | [日本語](#) | [한국어](#) | [中文 \(简体\)](#) | [中文 \(繁體\)](#)

[Site Terms](#) | [Privacy](#)

© 2018, Amazon Web Services, Inc. or its affiliates. All rights reserved.