

Matillion ETL for Amazon Redshift on the AWS Cloud

Quick Start Reference Deployment

December 2019

*Laura Malins, Tim Fletcher, Matillion
Sudhir Gupta, Shivansh Singh, Tony Bulding, AWS Quick Start team*

Visit our [GitHub repository](#) for source files and to post feedback, report bugs, or submit feature ideas for this Quick Start.

Contents

| | |
|---|---|
| Overview | 2 |
| Matillion ETL for Amazon Redshift on AWS..... | 3 |
| Cost and licenses | 3 |
| Architecture | 4 |
| Single-instance deployment..... | 4 |
| HA deployment | 6 |
| Planning the deployment | 7 |
| Specialized knowledge | 7 |
| AWS account | 7 |
| Technical requirements | 7 |
| Deployment options..... | 8 |
| Deployment steps | 9 |
| Step 1. Sign in to your AWS account..... | 9 |

| | |
|---|----|
| Step 2. Subscribe to the Matillion ETL for Amazon Redshift AMI | 9 |
| Step 3. Launch the Quick Start | 10 |
| Option 1: Single instance - Parameters for deploying Matillion ETL for Amazon Redshift into a new VPC | 11 |
| Option 2: Single instance - Parameters for deploying Matillion ETL for Amazon Redshift into an existing VPC | 16 |
| Option 3: High Availability - Parameters for deploying Matillion ETL for Amazon Redshift into a new VPC | 18 |
| Option 4: High Availability - Parameters for deploying Matillion ETL for Amazon Redshift into an existing VPC | 23 |
| Step 4. Test the deployment..... | 27 |
| Best practices for using Matillion ETL for Amazon Redshift on AWS..... | 27 |
| Security | 27 |
| Support information..... | 28 |
| FAQ..... | 28 |
| Send us feedback | 29 |
| Additional resources..... | 29 |
| Document revisions..... | 30 |

This Quick Start was created by Matillion in collaboration with Amazon Web Services (AWS).

[Quick Starts](#) are automated reference deployments that use AWS CloudFormation templates to deploy key technologies on AWS, following AWS best practices.

Overview

This Quick Start reference deployment guide provides step-by-step instructions for deploying Matillion ETL for Amazon Redshift on the AWS Cloud in either a single instance with an external Amazon Aurora PostgreSQL database or a high availability (HA) cluster.

This Quick Start is for users who want to launch Matillion ETL for Amazon Redshift without needing to manually configure related AWS services.

Please know that we may share who uses AWS Quick Starts with the AWS partner that collaborated with AWS on the content of the Quick Start.

Matillion ETL for Amazon Redshift on AWS

Matillion ETL for Amazon Redshift is an extract, transform, and load/extract, load, and transform (ETL/ELT) tool that automates data loads and transformations for greater speed, scale, and savings in the enterprise.

You can load data into Amazon Redshift from data sources such as on-premises and cloud databases, cloud and software as a service (SaaS) applications, application programming interface (API)-enabled services, files, and NoSQL databases. Once your data is available in Amazon Redshift, you can combine transformation components in Matillion ETL to build complex data transformations in a guided UI, for visualizations, business intelligence, reporting, and advanced analytics.

Launching on AWS enables your Matillion ETL instance to seamlessly use AWS services such as Amazon Simple Notification Service (Amazon SNS), Amazon Simple Queue Service (Amazon SQS), Amazon Simple Storage Service (Amazon S3), AWS Key Management Service (AWS KMS), along with change data capture, which extends the capabilities of your cloud infrastructure.

Cost and licenses

You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using the Quick Start.

The AWS CloudFormation template for this Quick Start includes configuration parameters that you can customize. Some of these settings, such as instance type, will affect the cost of deployment. For cost estimates, see the pricing pages for each AWS service you will be using. Prices are subject to change.

Tip After you deploy the Quick Start, we recommend that you enable the [AWS Cost and Usage Report](#) to track costs associated with the Quick Start. This report delivers billing metrics to an S3 bucket in your account. It provides cost estimates based on usage throughout each month and finalizes the data at the end of the month. For more information about the report, see the [AWS documentation](#).

The Quick Start requires a subscription to the Amazon Machine Image (AMI) for Matillion ETL for Amazon Redshift, which is available from [AWS Marketplace](#). Additional pricing, terms, and conditions may apply. For instructions, see [step 2](#) in the deployment section.

Architecture

You can choose a [deployment of a single instance](#) or a [high availability \(HA\) deployment](#) of a cluster.

Single-instance deployment

Deploying this Quick Start for a new virtual private cloud (VPC) with **default parameters** builds the following Matillion ETL for Amazon Redshift environment in the AWS Cloud.

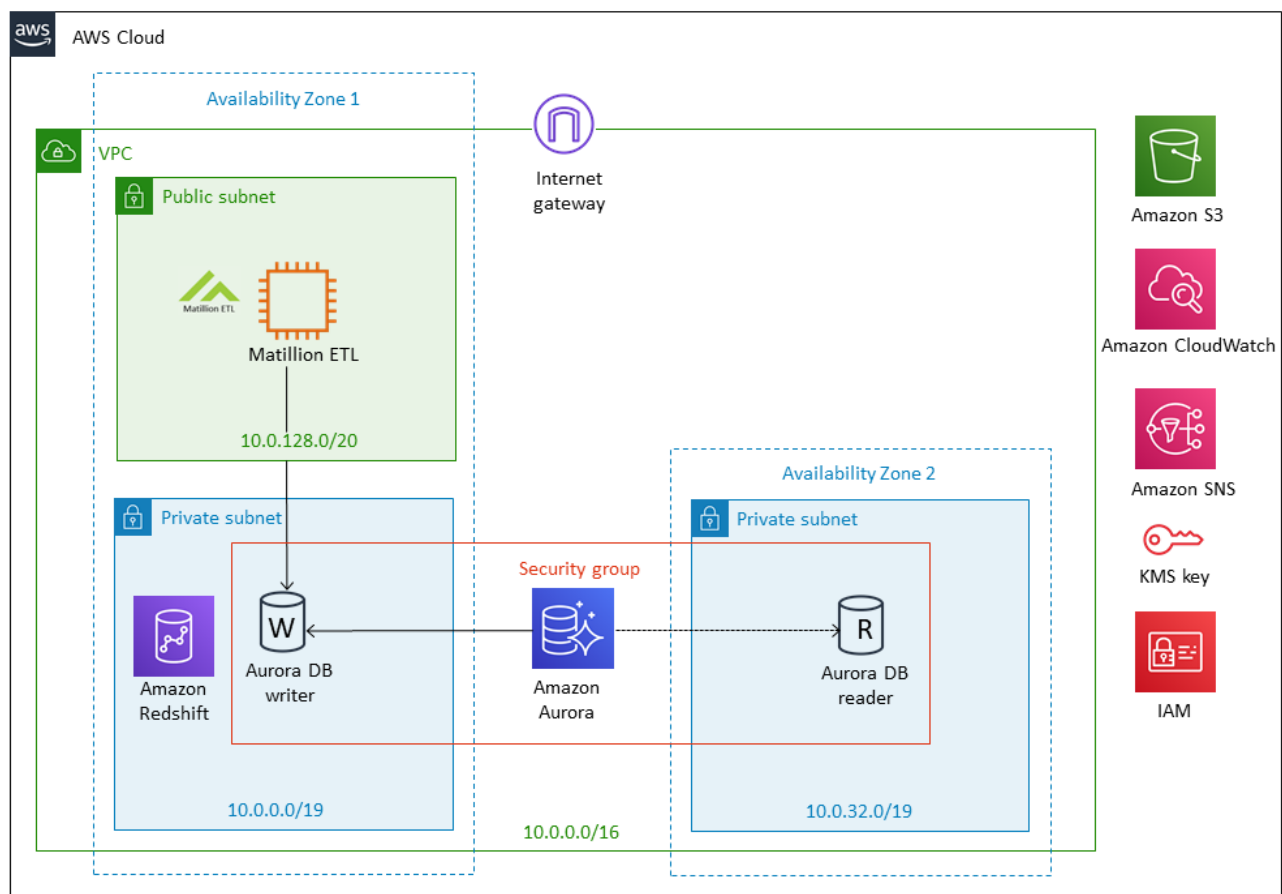


Figure 1: Quick Start architecture – Matillion ETL for Amazon Redshift on AWS

The Quick Start sets up the following:

- A highly available architecture that spans two Availability Zones.*
- A VPC configured with public and private subnets according to AWS best practices, to provide you with your own virtual network on AWS.*
- In the public subnet, a single Amazon Elastic Compute Cloud (Amazon EC2) instance running Matillion ETL.
- An AWS Identity and Access Management (IAM) role, attached to the EC2 instance.
- In the private subnets, Amazon Aurora, which is used as the Matillion ETL metadata repository.
- In a private subnet, Amazon Redshift to load data into from Matillion ETL for Amazon Redshift.*
- Amazon CloudWatch–based logging to monitor the Matillion ETL server status.
- Amazon SNS to send Amazon CloudWatch alarm and event notifications.

* The template that deploys the Quick Start into an existing VPC skips the components marked by asterisks and prompts you for your existing VPC configuration.

HA deployment

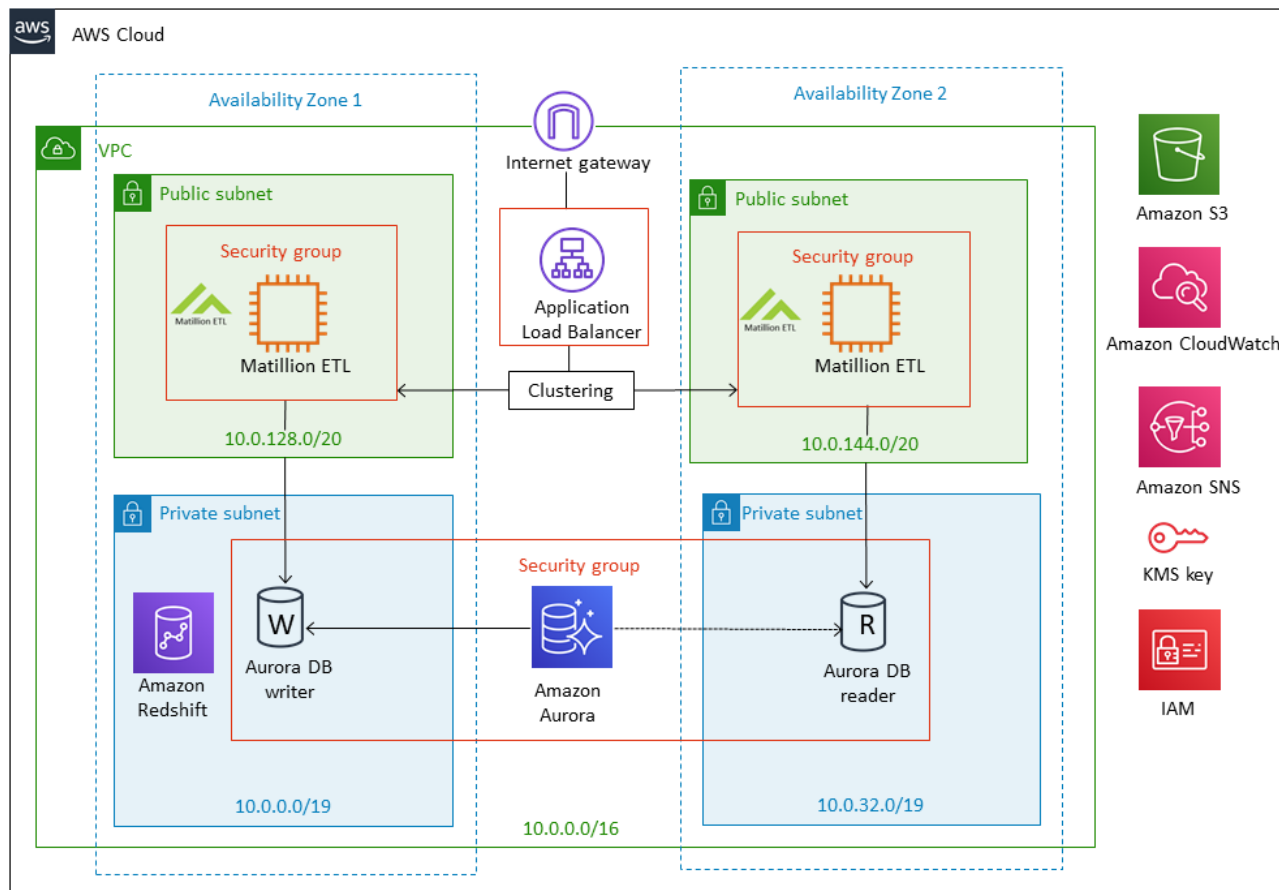


Figure 2: HA - Quick Start architecture – Matillion ETL for Amazon Redshift on AWS

The Quick Start sets up the following:

- A highly available architecture that spans two Availability Zones.*
- A VPC configured with public and private subnets according to AWS best practices, to provide you with your own virtual network on AWS.*
- In the public subnets, two Amazon EC2 instances running Matillion ETL in a cluster, deployed across two Availability Zones.
- An Application Load Balancer to direct traffic to the Matillion ETL instances.
- An IAM role, attached to the EC2 instance to specify which AWS services the Matillion ETL instance can access.
- In the private subnets, Amazon Aurora, which is used as the Matillion ETL metadata repository.

- In a private subnet, Amazon Redshift to load data into from Matillion ETL for Amazon Redshift.*
- Amazon CloudWatch–based logging to monitor the Matillion ETL server status.
- Amazon SNS to send Amazon CloudWatch alarm and event notifications.

* The template that deploys the Quick Start into an existing VPC skips the components marked by asterisks and prompts you for your existing VPC configuration.

Planning the deployment

Specialized knowledge

This Quick Start assumes familiarity with your source data and basic concepts around Amazon Redshift.

This deployment guide also requires a moderate level of familiarity with AWS services. If you're new to AWS, visit the [Getting Started Resource Center](#) and the [AWS Training and Certification website](#) for materials and programs that can help you develop the skills to design, deploy, and operate your infrastructure and applications on the AWS Cloud.

AWS account

If you don't already have an AWS account, create one at <https://aws.amazon.com> by following the on-screen instructions. Part of the sign-up process involves receiving a phone call and entering a PIN using the phone keypad.

Your AWS account is automatically signed up for all AWS services. You are charged only for the services you use.

Technical requirements

Before you launch the Quick Start, your account must be configured as specified in the following table. Otherwise, deployment might fail.

Resources

If necessary, request [service quota increases](#) for the following resources. You might need to do this if you already have an existing deployment that uses these resources, and you think you might exceed the default quotas with this deployment. For default quotas, see the [AWS documentation](#).

| Resource | This deployment uses |
|----------------------------|---|
| VPCs | 1 |
| IAM security groups | 1 |
| IAM roles | 3 |
| Application Load Balancers | 1 for HA |
| Matillion ETL instances | 1 EC2 instance – single deployment 2 EC2 instances – HA deployment |
| Databases | 2 Aurora DB instances |
| Data warehouses | 1 Amazon Redshift instance |

Regions

For a current list of supported Regions, see [Service Endpoints and Quotas](#) in the AWS documentation.

Key pair

Make sure that at least one Amazon EC2 key pair exists in your AWS account in the Region where you are planning to deploy the Quick Start. Make note of the key pair name. You'll be prompted for this information during deployment. To create a key pair, follow the [instructions in the AWS documentation](#).

If you're deploying the Quick Start for testing or proof-of-concept purposes, we recommend that you create a new key pair instead of specifying a key pair that's already being used by a production instance.

IAM permissions

To deploy the Quick Start, you must log in to the AWS Management Console with IAM permissions for the resources and actions the templates will deploy. The *AdministratorAccess* managed policy within IAM provides sufficient permissions, although your organization may choose to use a custom policy with more restrictions.

Deployment options

This Quick Start provides the following deployment options:

- **Deploy Matillion ETL for Amazon Redshift into a new VPC (end-to-end deployment).** This option builds a new AWS environment consisting of the VPC, subnets, internet gateways, security groups, and other infrastructure components, and then deploys Matillion ETL for Amazon Redshift into this new VPC.

- **Deploy Matillion ETL for Amazon Redshift into an existing VPC.** This option provisions Matillion ETL for Amazon Redshift in your existing AWS infrastructure.
- **High availability (HA): Deploy Matillion ETL for Amazon Redshift into a new VPC (end-to-end deployment).** This high-availability option builds a new AWS environment consisting of the VPC, subnets, internet gateways, security groups, and other infrastructure components, and then deploys Matillion ETL for Amazon Redshift into this new VPC.
- **High availability (HA): Deploy Matillion ETL for Amazon Redshift into an existing VPC.** This high-availability option provisions Matillion ETL for Amazon Redshift in your existing AWS infrastructure.

The Quick Start provides separate templates for these options. It also lets you configure CIDR blocks, instance types, and Matillion ETL for Amazon Redshift settings, as discussed later in this guide.

Deployment steps

Step 1. Sign in to your AWS account

1. Sign in to your AWS account at <https://aws.amazon.com> with an IAM user role that has the necessary permissions. For details, see [Planning the deployment](#) earlier in this guide.
2. Make sure that your AWS account is configured correctly, as discussed in the [Technical requirements](#) section.

Step 2. Subscribe to the Matillion ETL for Amazon Redshift AMI

This Quick Start requires a subscription to the AMI for Matillion ETL for Amazon Redshift in AWS Marketplace.

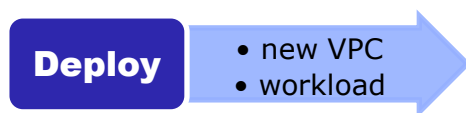
1. Sign in to your AWS account.
2. Open the page for the Matillion ETL for Amazon Redshift in [AWS Marketplace](#), and then choose **Continue to Subscribe**.
3. Review the terms and conditions for software usage, and then choose **Accept Terms**.
You will get a confirmation page, and an email confirmation will be sent to the account owner. For detailed subscription instructions, see the [AWS Marketplace documentation](#).
4. When the subscription process is complete, exit out of AWS Marketplace without further action. **Do not provision the software from AWS Marketplace—the Quick Start will deploy the AMI for you.**

Step 3. Launch the Quick Start

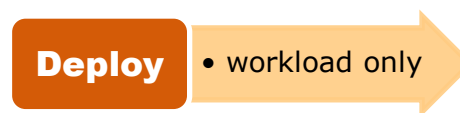
Notes The instructions in this section reflect the older version of the AWS CloudFormation console. If you're using the redesigned console, some of the user interface elements might be different.

You are responsible for the cost of the AWS services used while running this Quick Start reference deployment. There is no additional cost for using this Quick Start. For full details, see the pricing pages for each AWS service you will be using in this Quick Start. Prices are subject to change.

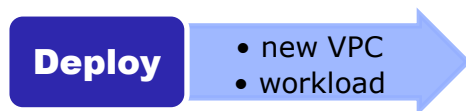
1. Sign in to your AWS account, and choose one of the following options to launch the AWS CloudFormation template. You can choose from single instance or high availability. For help choosing an option, see [deployment options](#) earlier in this guide.



[Deploy Matillion ETL for Amazon Redshift into a new VPC on AWS](#)



[Deploy Matillion ETL for Amazon Redshift into an existing VPC on AWS](#)



[HA: Deploy Matillion ETL for Amazon Redshift into a new VPC on AWS](#)



[HA: Deploy Matillion ETL for Amazon Redshift into an existing VPC on AWS](#)

Important If you're deploying Matillion ETL for Amazon Redshift into an existing VPC, make sure that your VPC has two private subnets in different Availability Zones for the Amazon Aurora databases and that the subnets aren't shared. This Quick Start doesn't support [shared subnets](#). The Amazon Redshift database also requires a private subnet. The Matillion ETL for Amazon Redshift EC2 instance requires a public subnet. You will be prompted for your VPC settings when you launch the Quick Start.

Each deployment takes about 20 minutes to complete.

2. Check the AWS Region that's displayed in the upper-right corner of the navigation bar, and change it if necessary. This is where the network infrastructure for Matillion ETL for Amazon Redshift will be built.
3. On the **Select Template** page, keep the default setting for the template URL, and then choose **Next**.
4. On the **Specify Details** page, change the stack name if needed. Review the parameters for the template. Provide values for the parameters that require input. For all other parameters, review the default settings and customize them as necessary.

In the following tables, parameters are listed by category and described separately for the various deployment options: single instance or high availability (HA):

- [Single instance: parameters for deploying Matillion ETL for Amazon Redshift into a new VPC](#)
- [Single instance: parameters for deploying Matillion ETL for Amazon Redshift into an existing VPC](#)
- [HA: Parameters for deploying Matillion ETL for Amazon Redshift into a new VPC](#)
- [HA: Parameters for deploying Matillion ETL for Amazon Redshift into an existing VPC](#)

When you finish reviewing and customizing the parameters, choose **Next**.

OPTION 1: SINGLE INSTANCE - PARAMETERS FOR DEPLOYING MATILLION ETL FOR AMAZON REDSHIFT INTO A NEW VPC

[View template](#)

End User License Agreement (EULA) - Matillion ETL for Redshift:

| Parameter label (name) | Default | Description |
|---|---------|---|
| Accepted EULA from AWS Marketplace (AcceptedEULA) | Yes | PLEASE READ THE MATILLION ETL FOR REDSHIFT EULA (https://redshift-support.matillion.com/s/article/2845300) CAREFULLY BEFORE USING THE SOFTWARE. The Matillion stack can be created only if you have already accepted the EULA. To accept the EULA, see https://aws.amazon.com/marketplace/pp/B01oED5YF8 . |

VPC network configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|--|
| Availability Zones (AvailabilityZones) | <i>Requires input</i> | The list of Availability Zones to use for the subnets in the VPC. The Quick Start uses two Availability Zones from your list and preserves the logical order you specify. |
| VPC CIDR (VPCCIDR) | 10.0.0.0/16 | The CIDR block for the VPC. |
| Private subnet 1 CIDR (PrivateSubnet1CIDR) | 10.0.0.0/19 | The CIDR block for the private subnet located in Availability Zone 1. |
| Private subnet 2 CIDR (PrivateSubnet2CIDR) | 10.0.32.0/19 | The CIDR block for the private subnet located in Availability Zone 2. |
| Public subnet 1 CIDR (PublicSubnet1CIDR) | 10.0.128.0/20 | The CIDR block for the public subnet located in Availability Zone 1. |
| Public subnet 2 CIDR (PublicSubnet2CIDR) | 10.0.144.0/20 | The CIDR block for the public subnet located in Availability Zone 2. |
| Allowed external access CIDR (RemoteAccessCIDR) | <i>Requires input</i> | The CIDR IP in the format x.x.x.x/x for external SSH access to the Matillion stack. We recommend that you set this value to a trusted IP range. For example, you might want to grant only your corporate network access to the software. |

Matillion EC2 instance configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Matillion EC2 instance type (MatillionEC2InstanceType) | m4.large | The Amazon EC2 instance type for the Matillion instance. A larger instance type enables greater workload concurrency. For more information, see https://www.matillion.com/pricing/ . |
| Key pair name (KeyPairName) | <i>Requires input</i> | A public/private key pair, which allows you to connect securely to your instance after it launches. This is the key pair you created in your preferred AWS Region; see the Technical requirements section. If you do not have one in this AWS Region, create it before continuing. |

Matillion Aurora/PostgreSQL repository configuration:

| Parameter label (name) | Default | Description |
|--|-------------|--|
| Aurora/PostgreSQL instance class (PGInstanceClass) | db.r5.large | The Aurora/PostgreSQL database instance class. |

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Aurora/PostgreSQL database name (PGDBName) | matillion | The Aurora/PostgreSQL database name for the Matillion repository. |
| Aurora/PostgreSQL master username (PGMasterUsername) | matillion | The user name that is associated with the master user account for the Aurora DB that is being created. |
| Aurora/PostgreSQL master password (PGMasterUserPassword) | <i>Requires input</i> | The Aurora/PostgreSQL master user password. Minimum 8 characters, must include 1 uppercase, 1 lowercase, 1 number, and 1 (non / @ " ') symbol. |
| Aurora/PostgreSQL database port (PGDBPort) | 8201 | The port number on which the Aurora/PostgreSQL database accepts incoming connections. |
| SNS notification Email (NotificationList) | ops@company.com | The email notification list that is used to configure an SNS topic for sending Amazon CloudWatch alarm and event notifications. |

Amazon Redshift configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|---|
| Do you want to create Redshift cluster stack? (EnableRedshiftStack) | true | Enables or disables creation of an Amazon Redshift stack. If true, an Amazon Redshift stack will be created. |
| Redshift database name (RedshiftDBName) | matillion | The name of the first database to be created when the cluster is created. |
| Redshift cluster port (RedshiftDBPort) | 8200 | The port number on which the cluster accepts incoming connections. |
| Node type for Redshift cluster (RedshiftNodeType) | dc2.large | The type of node to be provisioned. |
| Number of nodes in Redshift cluster (RedshiftNumberOfNodes) | 2 | The number of compute nodes in the Amazon Redshift cluster. |
| Redshift master user name (RedshiftMasterUsername) | matillion | The user name that is associated with the master user account for the cluster that is being created. |
| Redshift master user password | <i>Requires input</i> | The password that is associated with the master user account for the cluster that is being created. It must have at least 8 characters and no more than 64 characters, and must include 1 |

| Parameter label (name) | Default | Description |
|---|---------------------|--|
| (RedshiftMasterUserPassword) | | uppercase letter, 1 lowercase letter, 1 number, and 1 symbol (excluding / @ \ " '). |
| Enable Redshift logging to S3 (RedshiftEnableLoggingToS3) | false | Enables or disables logging to an S3 bucket. To enable logging, select true. |
| Max. number of concurrent clusters (RedshiftMaxConcurrentCluster) | 1 | The maximum number of concurrency scaling Amazon Redshift clusters. |
| Encryption at rest (RedshiftEncryptionAtRest) | false | Enables or disables encryption at rest of the Amazon Redshift database. |
| KMS key ID (Redshiftkmskey) | — | The existing KMS key ID for encrypting the Amazon Redshift database at-rest. |
| Amazon S3 bucket for Redshift IAM role (RedshiftS3BucketForIAMRole) | — | The existing Amazon S3 bucket. An IAM role will be created and associated to the Amazon Redshift cluster with GET and LIST access to this bucket. |
| Maintenance window (RedshiftMaintenancewindow) | sat:05:00-sat:05:30 | The maintenance window for the Amazon Redshift cluster. |
| Redshift snapshot identifier (RedshiftSnapshotIdentifier) | — | The Amazon Redshift snapshot identifier. Leave this blank for a new cluster. Enter the snapshot identifier, only if you want to restore from a snapshot. |
| AWS account-ID of the Redshift Snapshot (RedshiftSnapshotAccountNumber) | — | The AWS account number where the Amazon Redshift snapshot was created. Leave this blank if the snapshot was created in the current AWS account. |

Tag identifiers:

| Parameter label (name) | Default | Description |
|--|--------------|---|
| Environment (TagEnvironment) | dev | Designates the environment stage of the associated AWS resource. |
| Unique friendly name (TagName) | matillion-qs | The unique friendly name as required by your company's tagging strategy document, which will be added to the tag. |

AWS Quick Start configuration:

Note We recommend that you keep the default settings for the following two parameters, unless you are customizing the Quick Start templates for your own deployment projects. Changing the settings of these parameters will automatically update code references to point to a new Quick Start location. For additional details, see the [AWS Quick Start Contributor's Guide](#).

| Parameter label (name) | Default | Description |
|---|-------------------------------|--|
| Quick Start S3 bucket name (QSS3BucketName) | aws-quickstart | The S3 bucket name for the Quick Start assets. The Quick Start bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-). It cannot start or end with a hyphen. |
| Quick Start S3 key prefix (QSS3KeyPrefix) | quickstart- matillion-etl/ | The S3 key prefix for the Quick Start assets. The Quick Start key prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slash (/). |

OPTION 2: SINGLE INSTANCE - PARAMETERS FOR DEPLOYING MATILLION ETL FOR AMAZON REDSHIFT INTO AN EXISTING VPC

[View template](#)

End User License Agreement (EULA) - Matillion ETL for Redshift:

| Parameter label (name) | Default | Description |
|---|---------|---|
| Accepted EULA from AWS Marketplace (AcceptedEULA) | Yes | PLEASE READ THE MATILLION ETL FOR REDSHIFT EULA (https://redshift-support.matillion.com/s/article/2845300) CAREFULLY BEFORE USING THE SOFTWARE. The Matillion stack can be created only if you have already accepted the EULA. To accept the EULA, see https://aws.amazon.com/marketplace/pp/B01oED5YF8 . |

VPC network configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|---|
| VPC CIDR (VPCID) | <i>Requires input</i> | The ID of the existing VPC that contains the subnets. |
| Private subnet 1 CIDR (PrivateSubnet1) | <i>Requires input</i> | An existing private subnet 1 to launch secondary resources (e.g., PostgreSQL database). |
| Private subnet 2 CIDR (PrivateSubnet2) | <i>Requires input</i> | An existing private subnet 2 to launch secondary resources (e.g., PostgreSQL database). |
| Public subnet 1 CIDR (PublicSubnet1) | <i>Requires input</i> | An existing public subnet 1 to launch the Matillion EC2 instances into. |
| Allowed external access CIDR (RemoteAccessCIDR) | <i>Requires input</i> | The inbound IPv4 CIDR range for the Application Load Balancer. |

Matillion EC2 instance configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Matillion EC2 instance type (MatillionEC2InstanceType) | m4.large | The Amazon EC2 instance type for the Matillion instance. A larger instance type enables greater workload concurrency. For more information, see https://www.matillion.com/pricing/ . |
| Key pair name (KeyPairName) | <i>Requires input</i> | A public/private key pair, which allows you to connect securely to your instance after it launches. This is the key pair you created in your preferred AWS Region; see the Technical requirements section. If you do not have one in this AWS Region, create it before continuing. |

Matillion Aurora/PostgreSQL repository configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|---|
| Aurora/PostgreSQL instance class (PGInstanceClass) | db.r5.large | The Aurora/PostgreSQL database instance class. |
| Aurora/PostgreSQL database name (PGDBName) | matillion | The Aurora/PostgreSQL database name for the Matillion repository. |
| Aurora/PostgreSQL database port (PGDBPort) | 8201 | The port number on which the Aurora/PostgreSQL database accepts incoming connections. |
| Aurora/PostgreSQL master username (PGMasterUsername) | matillion | The user name that is associated with the master user account for the Aurora DB that is being created. |
| Aurora/PostgreSQL master password (PGMasterUserPassword) | <i>Requires input</i> | The Aurora/PostgreSQL master user password. Minimum 8 chars, must include 1 uppercase, 1 lowercase, 1 number, and 1 (non / @ " ') symbol. |
| SNS notification email (NotificationList) | ops@company.com | The email notification that is used to configure an SNS topic for sending an Amazon CloudWatch alarm and Amazon RDS event notifications. |

Tag identifiers:

| Parameter label (name) | Default | Description |
|--|---------|--|
| Environment (TagEnvironment) | dev | Designates the environment stage of the associated AWS resource. |

AWS Quick Start configuration:

Note We recommend that you keep the default settings for the following two parameters, unless you are customizing the Quick Start templates for your own deployment projects. Changing the settings of these parameters will automatically update code references to point to a new Quick Start location. For additional details, see the [AWS Quick Start Contributor's Guide](#).

| Parameter label (name) | Default | Description |
|---|---------------------------|--|
| Quick Start S3 bucket name (QSS3BucketName) | aws-quickstart | The S3 bucket name for the Quick Start assets. The Quick Start bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-). It cannot start or end with a hyphen. |
| Quick Start S3 key prefix (QSS3KeyPrefix) | quickstart-matillion-etl/ | The S3 key prefix for the Quick Start assets. The Quick Start key prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slash (/). |

OPTION 3: HIGH AVAILABILITY - PARAMETERS FOR DEPLOYING MATILLION ETL FOR AMAZON REDSHIFT INTO A NEW VPC

[View template](#)

End User License Agreement (EULA) - Matillion ETL for Redshift:

| Parameter label (name) | Default | Description |
|---|---------|---|
| Accepted EULA from AWS Marketplace (AcceptedEULA) | Yes | PLEASE READ THE MATILLION ETL FOR REDSHIFT EULA (https://redshift-support.matillion.com/s/article/2845300) CAREFULLY BEFORE USING THE SOFTWARE. The Matillion stack can be created only if you have already accepted the EULA. To accept the EULA, see https://aws.amazon.com/marketplace/pp/B01oED5YF8 . |

VPC network configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|---|
| Availability Zones (AvailabilityZones) | <i>Requires input</i> | The list of Availability Zones to use for the subnets in the VPC. The Quick Start uses two Availability Zones from your list and preserves the logical order you specify. |
| VPC CIDR (VPCCIDR) | 10.0.0.0/16 | The CIDR block for the VPC. |
| Private subnet 1 CIDR (PrivateSubnet1CIDR) | 10.0.0.0/19 | The CIDR block for the private subnet located in Availability Zone 1. |
| Private subnet 2 CIDR (PrivateSubnet2CIDR) | 10.0.32.0/19 | The CIDR block for the private subnet located in Availability Zone 2. |
| Public subnet 1 CIDR (PublicSubnet1CIDR) | 10.0.128.0/20 | The CIDR block for the public subnet located in Availability Zone 1. |
| Public subnet 2 CIDR (PublicSubnet2CIDR) | 10.0.144.0/20 | The CIDR block for the public subnet located in Availability Zone 2. |

| Parameter label (name) | Default | Description |
|---|-----------------------|---|
| Allowed external access CIDR (RemoteAccessCIDR) | <i>Requires input</i> | The CIDR IP in the format x.x.x.x/x for external access to the Application Load Balancer and for SSH access to the Matillion instance. We recommend that you set this value to a trusted IP range. For example, you might want to grant only your corporate network access to the software. |

Matillion EC2 instance and Application Load Balancer configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|--|
| Matillion EC2 instance type (MatillionEC2Instance Type) | m4.large | The Amazon EC2 instance type for the Matillion instance. A larger instance type enables greater workload concurrency. For more information, see https://www.matillion.com/pricing/ . |
| Key pair name (KeyPairName) | <i>Requires input</i> | A public/private key pair, which allows you to connect securely to your instance after it launches. This is the key pair you created in your preferred AWS Region; see the Technical requirements section. If you do not have one in this AWS Region, create it before continuing. |
| Matillion ALB DNS prefix (LBDnsName) | matillion | The prefix for the Load Balancer DNS name. Example: [matillion]-1731869672.eu-west-1.elb.amazonaws.com |

Matillion Aurora/PostgreSQL repository configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Aurora/PostgreSQL instance class (PGInstanceClass) | db.r5.large | The Aurora/PostgreSQL database instance class. |
| Aurora/PostgreSQL database name (PGDBName) | matillion | The Aurora/PostgreSQL database name for the Matillion repository. |
| Aurora/PostgreSQL database port (PGDBPort) | 8201 | The port number on which the Aurora database accepts incoming connections. |
| Aurora/PostgreSQL master username (PGMasterUsername) | matillion | The user name that is associated with the master user account for the Aurora DB that is being created. |
| Aurora/PostgreSQL master password (PGMasterUserPassword) | <i>Requires input</i> | The Aurora/PostgreSQL master user password. Minimum 8 characters, must include 1 uppercase, 1 lowercase, 1 number, and 1 (non / @ " ") symbol. |

| Parameter label (name) | Default | Description |
|---|-----------------|---|
| SNS notification Email (NotificationList) | ops@company.com | The email notification list that is used to configure an SNS topic for sending Amazon CloudWatch alarm and event notifications. |

Amazon Redshift configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|--|
| Do you want to create Redshift cluster stack? (EnableRedshiftStack) | true | Enables or disables creation of an Amazon Redshift stack. If true, an Amazon Redshift stack will be created. |
| Redshift database name (RedshiftDBName) | matillion | The name of the first database to be created when the cluster is created. |
| Redshift cluster port (RedshiftDBPort) | 8200 | The port number on which the Amazon Redshift cluster accepts incoming connections. |
| Node type for Redshift cluster (RedshiftNodeType) | dc2.large | The type of node to be provisioned. |
| Number of nodes in Redshift cluster (RedshiftNumberOfNodes) | 2 | The number of compute nodes in the Amazon Redshift cluster. |
| Redshift master user name (RedshiftMasterUsername) | matillion | The user name that is associated with the master user account for the cluster that is being created. |
| Redshift master user password (RedshiftMasterUserPassword) | <i>Requires input</i> | The password that is associated with the master user account for the cluster that is being created. It must have at least 8 characters and no more than 64 characters, and must include 1 uppercase letter, 1 lowercase letter, 1 number, and 1 symbol (excluding / @ \ " '). |
| Enable Redshift logging to S3 (RedshiftEnableLoggingToS3) | false | Enables or disables logging to an S3 bucket. To enable logging, select true. |
| Max. number of concurrent clusters (RedshiftMaxConcurrentCluster) | 1 | The maximum number of concurrency scaling Amazon Redshift clusters. |
| Encryption at rest (RedshiftEncryptionAtRest) | false | Enables or disables encryption at rest of the Amazon Redshift database. |

| Parameter label (name) | Default | Description |
|---|---------------------|---|
| KMS key ID (Redshiftkmskey) | — | The existing KMS key ID for encrypting the Amazon Redshift database at-rest. |
| Amazon S3 bucket for Redshift IAM role (RedshiftS3BucketForIAMRole) | — | The existing Amazon S3 bucket. An IAM role will be created and associated to the Amazon Redshift cluster with GET and LIST access to this bucket. |
| Maintenance window (RedshiftMaintenancewindow) | sat:05:00-sat:05:30 | The maintenance window for the Amazon Redshift cluster. |
| Redshift snapshot identifier (RedshiftSnapshotIdentifier) | — | The Amazon Redshift snapshot identifier. Leave this blank for a new cluster. Enter the snapshot identifier only if you want to restore from a snapshot. |
| AWS account-ID of the Redshift Snapshot (RedshiftSnapshotAccountNumber) | — | The AWS account number where the Amazon Redshift snapshot was created. Leave this blank if the snapshot was created in the current AWS account. |

Matillion ETL realm configuration:

| Parameter label (name) | Default | Description |
|---|---------|---|
| Username (MatillionRealmConName) | — | The connection user name (e.g., administrator@INTERNAL.DOMAIN.COM). |
| Connection password (MatillionRealmConPass) | — | The password for the connection user name for the initial bind. |
| URL (MatillionRealmConURL) | — | The URL to your directory server (e.g., ldap://10.10.10.254:389) |
| User Base (MatillionRealmUserBase) | — | The subtree below which users are stored in the directory tree (e.g., cn=Users,dc=INTERNAL,dc=domain,dc=com). |
| User Search (MatillionRealmUserSearch) | — | The LDAP attribute to use for identifying users (e.g., sAMAccountName={0}). |
| Role Base (MatillionRealmRoleBase) | — | The subtree below which groups are stored in the directory tree (e.g., cn=Groups,dc=INTERNAL,dc=domain,dc=com). |
| Role Name (MatillionRealmRoleName) | — | The LDAP attribute used to identify a group or role (e.g., cn). |

| Parameter label (name) | Default | Description |
|--|---------|--|
| Role Search (MatillionRealmRoleSearch) | — | The LDAP attribute to use to identify groups or roles (e.g., member={o}). |
| User Subtree (MatillionRealmUserSubtree) | false | Sets the scope of the search. Select true if you wish to search the entire subtree, rooted at the 'User Base' entry. Selecting false (default) requests a lone top-level search. |
| Login Role (MatillionRealmMETLRole) | — | The name of an existing group in the directory server whose users will be allowed to log in. Role names are case-sensitive. |
| Admin Role (MatillionRealmMETLAdminRole) | — | The name of an existing group in the directory server whose users will be allowed to administer Matillion. Role names are case-sensitive. |
| API Role (MatillionRealmMETLAPIRole) | — | The name of an existing group in the directory server whose users will be allowed to administer Matillion. Role names are case-sensitive. |

Tag identifiers:

| Parameter label (name) | Default | Description |
|--|--------------|---|
| Environment (TagEnvironment) | dev | Designates the environment stage of the associated AWS resource. |
| Unique friendly name (TagName) | matillion-qs | The unique friendly name as required by your company's tagging strategy document, which will be added to the tag. |

AWS Quick Start configuration:

Note We recommend that you keep the default settings for the following two parameters, unless you are customizing the Quick Start templates for your own deployment projects. Changing the settings of these parameters will automatically update code references to point to a new Quick Start location. For additional details, see the [AWS Quick Start Contributor's Guide](#).

| Parameter label (name) | Default | Description |
|---|----------------|--|
| Quick Start S3 bucket name (QSS3BucketName) | aws-quickstart | The S3 bucket name for the Quick Start assets. The Quick Start bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-). It cannot start or end with a hyphen. |

| Parameter label (name) | Default | Description |
|---|---------------------------|---|
| Quick Start S3 key prefix (QSS3KeyPrefix) | quickstart-matillion-etl/ | The S3 key prefix for the Quick Start assets. The Quick Start key prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slash (/). |

OPTION 4: HIGH AVAILABILITY - PARAMETERS FOR DEPLOYING MATILLION ETL FOR AMAZON REDSHIFT INTO AN EXISTING VPC

[View template](#)

End User License Agreement (EULA) - Matillion ETL for Redshift:

| Parameter label (name) | Default | Description |
|---|---------|--|
| Accepted EULA from AWS Marketplace (AcceptedEULA) | Yes | PLEASE READ THE MATILLION ETL FOR REDSHIFT EULA (https://redshift-support.matillion.com/s/article/2845300) CAREFULLY, BEFORE USING THE SOFTWARE. The Matillion stack can be created only if you have already accepted the EULA. To accept the EULA, see https://aws.amazon.com/marketplace/pp/B01oED5YF8 . |

VPC network configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|---|
| VPC CIDR (VPCID) | <i>Requires input</i> | The ID of the existing VPC. This must be the VPC that contains the subnets. |
| Private subnet 1 CIDR (PrivateSubnet1) | <i>Requires input</i> | An existing private subnet to launch secondary resources, e.g. PostgreSQL database. |
| Private subnet 2 CIDR (PrivateSubnet2) | <i>Requires input</i> | An existing private subnet to launch secondary resources, e.g. PostgreSQL database. |
| Public subnet 1 CIDR (PublicSubnet1) | <i>Requires input</i> | An existing public subnet to launch the Matillion EC2 instance(s) into. |
| Public subnet 2 CIDR (PublicSubnet2) | <i>Requires input</i> | An existing public subnet to launch the Matillion EC2 instance(s) into. |
| Allowed external access CIDR (RemoteAccessCIDR) | <i>Requires input</i> | Inbound IPv4 CIDR range for the Application Load Balancer. |

Matillion EC2 instance and Application Load Balancer configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Matillion EC2 instance type (MatillionEC2InstanceType) | m4.large | The Amazon EC2 instance type for the Matillion instance. A larger instance type enables greater workload concurrency. For more information, see https://www.matillion.com/pricing/ . |
| Key pair name (KeyPairName) | <i>Requires input</i> | A public/private key pair, which allows you to connect securely to your instance after it launches. This is the key pair you created in your preferred AWS Region; see the Technical requirements section. If you do not have one in this AWS Region, create it before continuing. |
| Matillion ALB DNS prefix (LBDnsNAME) | matillion | The Application Load Balancer DNS name prefix (e.g., [matillion]-1731869672.eu-west-1.elb.amazonaws.com). |

Matillion Aurora/PostgreSQL repository configuration:

| Parameter label (name) | Default | Description |
|--|-----------------------|--|
| Aurora/PostgreSQL instance class (PGInstanceClass) | db.r5.large | The Aurora/PostgreSQL database instance class. |
| Aurora/PostgreSQL database name (PGDBName) | matillion | The Aurora/PostgreSQL database name for the Matillion repository. |
| Aurora/PostgreSQL database port (PGDBPort) | 8201 | The port number on which the Aurora/PostgreSQL database accepts incoming connections. |
| Aurora/PostgreSQL master username (PGMasterUsername) | matillion | The user name that is associated with the master user account for the Aurora DB that is being created. |
| Aurora/PostgreSQL master password (PGMasterUserPassword) | <i>Requires input</i> | The Aurora/PostgreSQL master user password. Minimum 8 chars, must include 1 uppercase, 1 lowercase, 1 number, and 1 (non / @ " ") symbol |
| SNS notification email (NotificationList) | db-ops@ domain.com | The email notification that is used to configure an SNS topic for sending an Amazon CloudWatch alarm and Amazon RDS event notifications. |

Matillion ETL realm configuration:

| Parameter label (name) | Default | Description |
|---|-----------------------|--|
| Username (MatillionRealmConName) | <i>Requires input</i> | The connection user name (e.g., administrator@INTERNAL.DOMAIN.COM). |
| Connection password (MatillionRealmConPass) | <i>Requires input</i> | The password for the connection user name for the initial bind. |
| URL (MatillionRealmConURL) | <i>Requires input</i> | The URL to your directory server (e.g., ldap://10.10.10.254:389). |
| User Base (MatillionRealmUserBase) | <i>Requires input</i> | The subtree below which users are stored in the directory tree (e.g., cn=Users,dc=INTERNAL,dc=domain,dc=com). |
| User Search (MatillionRealmUserSearch) | <i>Requires input</i> | The LDAP attribute to use for identifying users (e.g., sAMAccountName={0}). |
| Role Base (MatillionRealmRoleBase) | <i>Requires input</i> | The subtree below which groups are stored in the directory tree (e.g., cn=Groups,dc=INTERNAL,dc=domain,dc=com). |
| Role Name (MatillionRealmRoleName) | <i>Requires input</i> | The LDAP attribute used to identify a group or role (e.g., cn). |
| Role Search (MatillionRealmRoleSearch) | <i>Requires input</i> | The LDAP attribute to use to identify groups or roles (e.g., member={0}). |
| User Subtree (MatillionRealmUserSubtree) | false | Sets the scope of the search. Select true if you want to search the entire subtree, rooted at the 'User Base' entry. Selecting false (default) requests a lone top-level search. |
| Login Role (MatillionRealmMETLRole) | <i>Requires input</i> | The name of an existing group in the directory server whose users will be allowed to log in. Role names are case-sensitive. |
| Admin Role (MatillionRealmMETLAdminRole) | <i>Requires input</i> | The name of an existing group in the directory server whose users will be allowed to administer Matillion. Role names are case-sensitive. |
| API Role (MatillionRealmMETLAPIRole) | <i>Requires input</i> | The name of an existing group in the directory server whose users will be allowed to administer Matillion. Role names are case-sensitive. |

Tag identifiers:

| Parameter label (name) | Default | Description |
|--|---------|--|
| Environment (TagEnvironment) | dev | Designates the environment stage of the associated AWS resource. |

AWS Quick Start configuration:

Note We recommend that you keep the default settings for the following two parameters, unless you are customizing the Quick Start templates for your own deployment projects. Changing the settings of these parameters will automatically update code references to point to a new Quick Start location. For additional details, see the [AWS Quick Start Contributor's Guide](#).

| Parameter label (name) | Default | Description |
|---|---------------------------|--|
| Quick Start S3 bucket name (QSS3BucketName) | aws-quickstart | The S3 bucket name for the Quick Start assets. The Quick Start bucket name can include numbers, lowercase letters, uppercase letters, and hyphens (-). It cannot start or end with a hyphen. |
| Quick Start S3 key prefix (QSS3KeyPrefix) | quickstart-matillion-etl/ | The S3 key prefix for the Quick Start assets. The Quick Start key prefix can include numbers, lowercase letters, uppercase letters, hyphens (-), and forward slash (/). |

- On the **Options** page, you can [specify tags](#) (key-value pairs) for resources in your stack and [set advanced options](#). When you're done, choose **Next**.
- On the **Review** page, review and confirm the template settings. Under **Capabilities**, select the two checkboxes to acknowledge that the template will create IAM resources and that it might require the capability to auto-expand macros.
- Choose **Create** to deploy the stack.
- Monitor the status of the stack. When the status is **CREATE_COMPLETE**, the Matillion ETL for Amazon Redshift cluster is ready.
- Use the URLs displayed in the **Outputs** tab for the stack, to view the resources that were created.

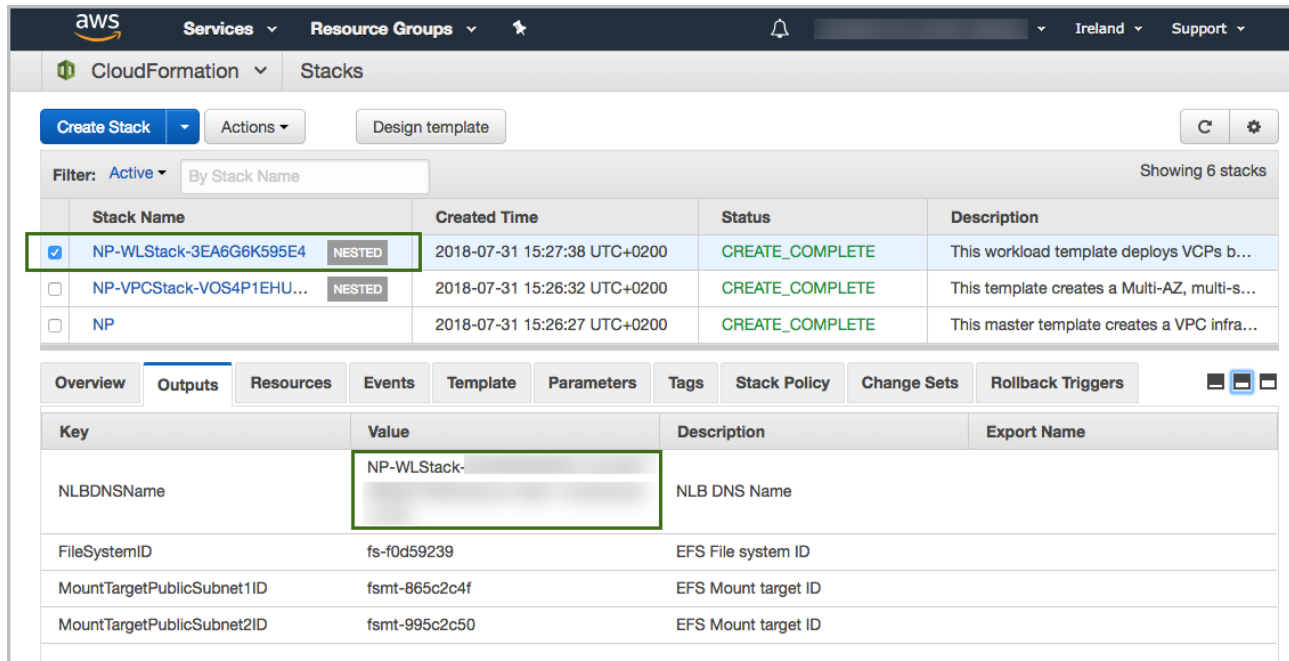


Figure 3: Matillion ETL for Amazon Redshift outputs after successful deployment

Step 4. Test the deployment

Once the stack has been successfully created, you can access Matillion ETL for Amazon Redshift via the IP address from the outputs with the user name and password that are provided in the outputs.

Best practices for using Matillion ETL for Amazon Redshift on AWS

For information about getting started and best practices related to Matillion ETL for Amazon Redshift, see <https://redshift-support.matillion.com/s/article/1975061>.

Security

Matillion is accessed via the web browser. This can be limited to HTTPS, and you should restrict the IP range in the EC2 security group to just the IP addresses that you will use to access Matillion. For more information about configuring Matillion ETL for Secure Sockets Layer (SSL), see <https://redshift-support.matillion.com/s/article/2824459#header4>.

Support information

Reach out to Matillion Support for further assistance at support@matillion.com or via <https://redshift-support.matillion.com/>.

FAQ

Q. I encountered a `CREATE_FAILED` error when I launched the Quick Start.

A. If AWS CloudFormation fails to create the stack, we recommend that you relaunch the template with **Rollback on failure** set to **No**. (This setting is under **Advanced** in the AWS CloudFormation console, **Options** page.) With this setting, the stack's state will be retained and the instance will be left running, so you can troubleshoot the issue. (The log file is located in `/var/log/tomcat8/`.)

Important When you set **Rollback on failure** to **No**, you will continue to incur AWS charges for this stack. Please make sure to delete the stack when you finish troubleshooting.

For additional information, see [Troubleshooting AWS CloudFormation](#) on the AWS website.

Q. I encountered a size limitation error when I deployed the AWS CloudFormation templates.

A. We recommend that you launch the Quick Start templates from the links in this guide or from another S3 bucket. If you deploy the templates from a local copy on your computer or from a location other than an S3 bucket, you might encounter template size limitations when you create the stack. For more information about AWS CloudFormation quotas, see the [AWS documentation](#).

Q. I can't access the Matillion ETL instance.

A. Check that the instance is running in the EC2 instances, and ensure that the security group allows you to access it. Matillion is accessed via the web browser.

Q. The product loads, but then immediately says "Connection lost." Why?

A. The "connection" that is being lost refers to a websocket connection. Once the site loads, all further communication is done over a websocket in order to broadcast changes to other users. Ensure that the client (browser) has a strong network connection.

Q. Matillion ETL can't connect to Amazon Redshift, but I don't know why. Any ideas?

A. If you see the error "Connection attempt timed out." while editing or creating a project or environment, this is usually related to security groups. Often, the security group assigned to the Amazon Redshift cluster and the security group assigned to the Matillion ETL instance are different groups. To resolve this issue, add the Matillion ETL security group as an inbound rule to the VPC Security Group section in the Amazon Redshift Cluster configuration.

Q. When using Matillion ETL over SSL, my browser warns me the site isn't secure. Why?

A. Matillion ETL comes with a self-signed SSL certificate, and so the browser can't validate it. You can upload your own certificate provided by AWS or another provider. For more information, see <https://redshift-support.matillion.com/s/article/2824459#header4>.

Q. How do I get further Matillion support?

A. Reach out to Matillion Support at support@matillion.com or via <https://redshift-support.matillion.com/s/>.

Send us feedback

To post feedback, submit feature ideas, or report bugs, use the **Issues** section of the [GitHub repository](#) for this Quick Start. If you'd like to submit code, please review the [Quick Start Contributor's Guide](#).

Additional resources

AWS resources

- [Getting Started Resource Center](#)
- [AWS General Reference](#)
- [AWS Glossary](#)

AWS services

- [AWS CloudFormation](#)
- [Amazon EC2](#)
- [IAM](#)
- [Amazon Route 53](#)

- [Amazon S3](#)
- [Amazon SNS](#)
- [Amazon VPC](#)

Matillion ETL for Amazon Redshift documentation

- [Matillion Support](#)

Other Quick Start reference deployments

- [AWS Quick Start home page](#)

Document revisions

| Date | Change | In sections |
|---------------|---------------------|-------------|
| December 2019 | Initial publication | |

© 2020, Amazon Web Services, Inc. or its affiliates, and Matillion. All rights reserved.

Notices

This document is provided for informational purposes only. It represents AWS's current product offerings and practices as of the date of issue of this document, which are subject to change without notice. Customers are responsible for making their own independent assessment of the information in this document and any use of AWS's products or services, each of which is provided "as is" without warranty of any kind, whether express or implied. This document does not create any warranties, representations, contractual commitments, conditions or assurances from AWS, its affiliates, suppliers or licensors. The responsibilities and liabilities of AWS to its customers are controlled by AWS agreements, and this document is not part of, nor does it modify, any agreement between AWS and its customers.

The software included with this paper is licensed under the Apache License, Version 2.0 (the "License"). You may not use this file except in compliance with the License. A copy of the License is located at <http://aws.amazon.com/apache2.0/> or in the "license" file accompanying this file. This code is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.