Data Warehouse Modernization on the AWS Cloud

with Tableau Server, Amazon Redshift, and Amazon RDS

Quick Start Reference Deployment

June 2017

Tableau Software Team
AWS Quick Start Reference Team

Contents

amazon web services

This Quick Start deployment guide was created by Amazon Web Services (AWS) in partnership with Tableau Software.

Overview

This Quick Start deployment guide provides step-by-step instructions for deploying a modern data warehouse, based on Amazon Redshift and including the analytics and visualization capabilities of Tableau Server, on the Amazon Web Services (AWS) Cloud. Quick Starts are automated deployments that use AWS CloudFormation templates to launch, configure, and run the AWS compute, network, storage, and other services required to deploy a specific workload on AWS.

An Amazon Redshift data warehouse is an enterprise-class relational database query and management system. Amazon Redshift achieves efficient storage and optimum query performance through a combination of massively parallel processing, columnar data storage, and very efficient, targeted data compression encoding schemes.

Amazon Redshift supports client connections with many types of applications, including business intelligence (BI), reporting, data, and analytics tools. This Quick Start deploys Tableau Server for BI and data visualization capabilities. Tableau Server enables analytics teams to publish BI reports and dashboards that were authored using Tableau Desktop. With Tableau Server deployed on AWS, enterprises can take advantage of collaborative analytics and the price/performance economies and agility provided by the cloud.

The Quick Start deployment includes the following:

- A representative dataset that is loaded to an Amazon Redshift cluster to support efficient creation of aggregates at large data scale.

- A high-availability Amazon Relational Database Service (Amazon RDS) instance that demonstrates the ability to publish computed aggregates for serving dashboards at scale with low latency and high availability.
- A Tableau dashboard that employs best practices around consuming and displaying large scale, aggregated data.

This Quick Start is for users who would like to modernize their enterprise data warehouse capabilities with an AWS-validated architecture that includes Amazon Redshift for a fast, cost-efficient, fully managed data warehouse, and Tableau Server for strong data visualization. The Quick Start also provides a reference architecture for users who want to migrate or extend their existing on-premises enterprise data warehouse to the cloud.

Costs 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 templates for this Quick Start include configuration parameters that you can customize. Some of these settings, such as instance type, will affect the cost of deployment. See the pricing pages for each AWS service you will be using for cost estimates.

You can choose to bring your own Tableau Server license and use it when launching the Quick Start. If you do not have a Tableau Server license or would prefer to launch the Quick Start as a trial, the Quick Start will provide a 14-day Tableau Server evaluation license. At the end of the 14-day trial, you must enter a product key to continue using Tableau Server. For more information about Tableau Server licenses, see the Tableau website.

To develop reports and dashboards to publish to Tableau Server, you will need Tableau Desktop, which you can license separately from Tableau Server.

Architecture

Deploying this Quick Start for a new virtual private cloud (VPC) with the **default parameters** builds the following modern data warehouse environment in the AWS Cloud, including Tableau Server, Amazon Redshift, and other AWS services.

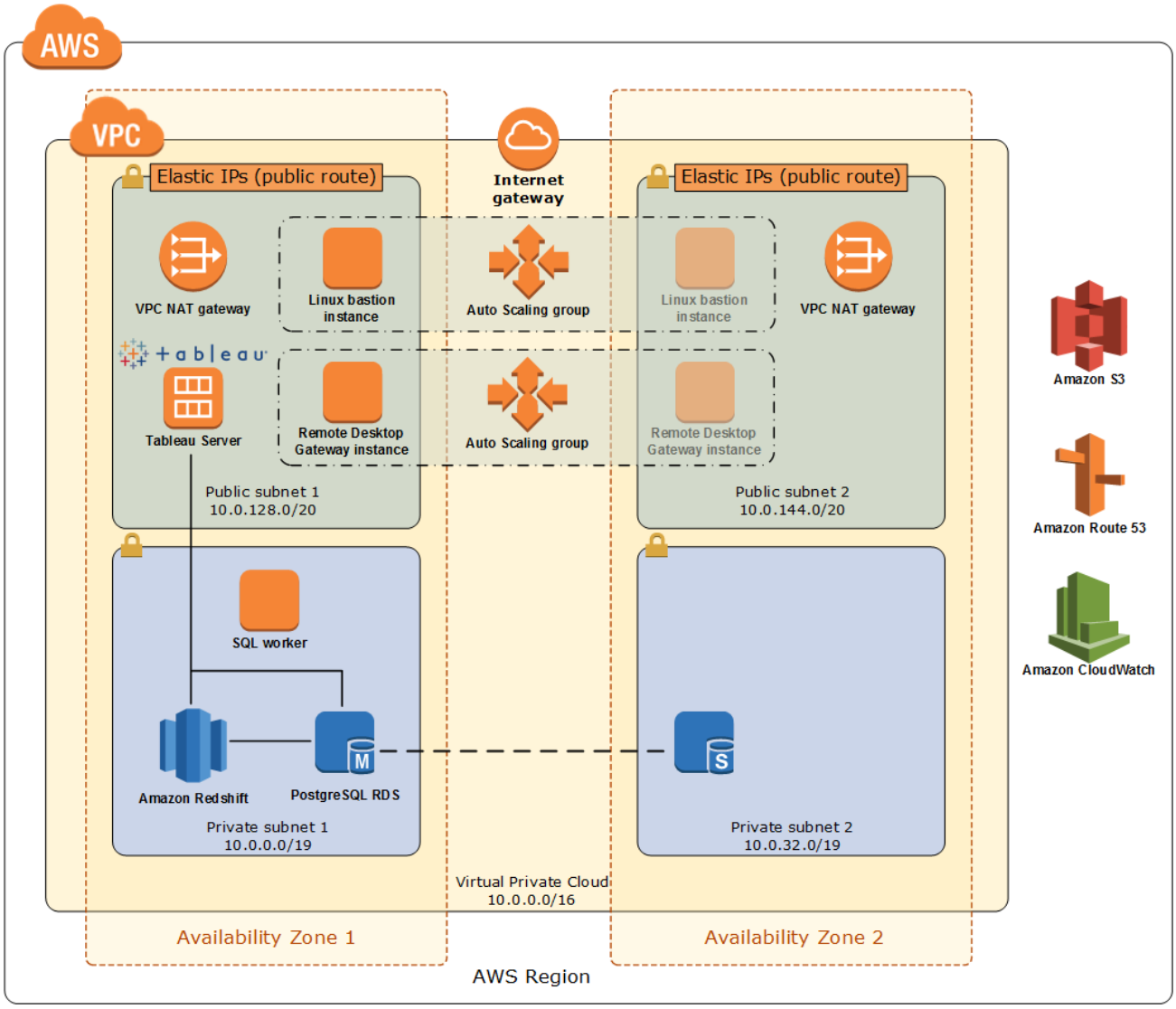


Figure 1: Quick Start architecture for a data warehouse with Tableau Server

The architecture includes the following services and components:

- **Amazon S3**. Amazon Simple Storage Service (Amazon S3) is an object store that provides artifacts necessary for the Quick Start, including datasets, dashboards, and SQL required to configure AWS database services and to compute aggregates for the sample dataset.
- **Sample dataset**. The Quick Start provides a sample dataset, available in Amazon S3, that demonstrates parallel data load to Amazon Redshift and computation of aggregates at large data scale. For more information, see the section Quick Start Dataset.

- **Amazon Redshift**. Amazon Redshift is a fast, fully managed, petabyte-scale data warehouse. The Quick Start uses Amazon Redshift to provide full fact tables, ad-hoc exploration and aggregation, and filtered drill-downs. Amazon Redshift is optimized for computationally intensive workloads such as computation of aggregates and complex joins, and supports analysis on both Microsoft Windows and macOS.
- **Amazon RDS with PostgreSQL**. This service makes it easy to set up, operate, and scale a relational database in the cloud. It provides the Quick Start with high-query-volume aggregate tables that feed scale-out dashboards. It is deployed in multiple Availability Zones for high availability.
- **Tableau Server on Amazon EC2**. The Quick Start provides a single-instance deployment of Tableau Server with the ability to host and serve analytics dashboards and workbooks, which is supported by the trial license. If you have a Tableau Server license, you can enter it upon deployment. For a Tableau Server standalone or cluster (multi-node) environment on AWS, see the Quick Start for Tableau Server.
- **Sample Tableau Server dashboard**. Dashboards, consistent with the sample dataset, demonstrate how to connect to multiple data sources in AWS to optimize performance.
- **SQL worker**. The Quick Start provisions a Linux instance that is used to coordinate and issue SQL commands, consistent with the sample dataset, to load data in parallel from Amazon S3 to Amazon Redshift, to set up a database link from Amazon RDS to Amazon Redshift, to define the desired aggregate computations in Amazon Redshift, and to publish computed aggregates to Amazon RDS.
- **Amazon VPC**. The Amazon Virtual Private Cloud (Amazon VPC) service lets you provision a logically isolated section of the AWS Cloud where you can launch resources in a virtual network that you define. The VPC provides a network architecture with multiple public and private subnets that span multiple Availability Zones, so that AWS resources can be deployed in highly available configurations.
- **Bastion host**. The Quick Start architecture deploys a Linux bastion host instance into a public subnet to provide readily available administrative access to the environment. This provides secure access to Linux instances located in the private and public subnets.
- **Remote Desktop Gateway**. The Quick Start deploys a Remote Desktop Gateway instance into a public subnet to provide readily available administrative access to the environment. This provides secure access to Microsoft Windows instances located in the private and public subnets. The Remote Desktop Gateway instance uses the Remote Desktop Protocol (RDP) over HTTPS to establish a secure, encrypted connection between remote users on the Internet and Windows-based EC2 instances, without

needing to configure a virtual private network (VPN) connection. This helps reduce the attack surface on your Windows-based instances and provides a remote administration solution for administrators.

- **NAT Gateway**. This service enables instances in a private subnet to connect to the Internet or to other AWS services, but prevents the Internet from initiating a connection with those instances.
- **IAM roles**. The Quick Start configures AWS Identity and Access Management (IAM) roles to provide the required access for AWS resources created through the Quick Start to access other AWS resources when required. These IAM roles enable access to data in Amazon S3, enable Amazon Redshift to copy data from the sample dataset's S3 bucket and key prefix into its tables, and enable association of the generated IAM role with the Amazon Redshift cluster.

Prerequisites

Specialized Knowledge

Before you deploy this Quick Start, we recommend that you become familiar with the following AWS services. (If you are new to AWS, see Getting Started with AWS.)

- Amazon EC2
- Amazon VPC
- Amazon S3
- Amazon Redshift
- Amazon RDS

Technical Requirements

This Quick Start includes two AWS Cloudformation templates. The first template creates a virtual private cloud (VPC) with multiple private and public subnets, which is used to house the remainder of the components in the architecture. The second template deploys those components into this new VPC or into an existing VPC that you provide.

If you'd like to use the second template to launch the Quick Start into your existing VPC, your VPC will need to be consistent with that produced by the Amazon VPC Quick Start.

Design Considerations

Quick Start Dataset

The Quick Start includes a sample dataset. The Quick Start is designed so that you can replace this dataset as needed for your use case. The dataset is stored in Amazon S3 at a

bucket and key prefix that you specify when you launch the Quick Start. The dataset includes the following data components and artifacts:

- **content.** The underlying data for the Quick Start using a folder-per-table data organization.
- **sql.** Includes the following SQL queries:

  `setup-redshift.sql` – to establish tables in Amazon Redshift consistent with the content, and to load the data for each table from Amazon S3.

  `setup-postgres.sql` – to establish a database link from Amazon RDS PostgreSQL to Amazon Redshift, and to establish any tables required in Amazon RDS PostgreSQL.

  `create-aggregates.sql` – to establish an aggregates view in Amazon Redshift.

  `publish-aggregates.sql` – to establish an aggregates table in Amazon RDS PostgreSQL from the aggregates view in Amazon Redshift.
- **dashboard.** Includes a Tableau Server dashboard to be installed on Tableau Server that references data provided by Amazon Redshift and Amazon RDS PostgreSQL data sources.

Loading the sample dataset adds about 20-25 minutes to the Quick Start deployment time. If you'd prefer not to load the dataset, set the **Load dataset** parameter to **false** during deployment. If you'd like to use another dataset, see the instructions later in this guide, and review the IAM role policies to ensure access to your own dataset.

IAM Roles

By default, the Quick Start deploys a non-sensitive, sample dataset that doesn't need to be secured. However, you can use alternative datasets with private, sensitive data. The Quick Start creates the following two IAM roles to ensure that private data is managed securely:

- **Ec2Worker.** This IAM role provides for Amazon S3 access to the referenced dataset from EC2 instances launched within the Quick Start. It also enables launched EC2 instances to associate an IAM role with the Amazon Redshift cluster that is launched within the Quick Start. Because this role is intended to be associated with EC2 instances, the Quick Start also creates an IAM instance profile that includes this IAM role.
- **RedshiftWorker.** This IAM role provides for Amazon S3 access to the referenced dataset from the Amazon Redshift cluster that is launched within the Quick Start.

Amazon Route 53 Convenience DNS Names

The Quick Start launches several AWS resources that work together, and often refer to one another in configuration files and in typical usage. The Quick Start can create DNS records with convenient access names for each resource within the Quick Start, if you provide the Quick Start with the name of an Amazon Route 53 hosted zone within which it can create a record set for each resource. If you do not want the Quick Start to create DNS records, if you do not use Amazon Route 53, or if no hosted zone is available, leave the **HostedZoneName** parameter set to the default value of `<NONE>`. If you specify a hosted zone name, be sure to include the trailing period; for example, `dev.example.com.`

If you provide a hosted zone name, the Quick Start will create DNS records that will allow you to access the AWS resources launched by the Quick Start using DNS names, as shown in the following table. For clarity, the example names in the table assume a hosted zone name of `dev.example.com.` and a master AWS CloudFormation stack name of `qs`. The actual DNS names created when you deploy the Quick Start will use the hosted zone and stack names that you specify.

AWS resource	DNS record
Linux bastion host	bastion.qs.dev.example.com.
Remote Desktop Gateway host	rdgw.qs.dev.example.com.
Tableau Server host	tableau-server.qs.dev.example.com. tableau-server-private.qs.dev.example.com.
SQL worker host	sql-worker-private.qs.dev.example.com.
Amazon Redshift endpoint	redshift.qs.dev.example.com.
Amazon RDS PostgreSQL endpoint	postgres.qs.dev.example.com.

Deployment Steps

To deploy a modern data warehouse, including a strong visualization engine provided by Tableau Server, on the AWS Cloud, follow these steps.

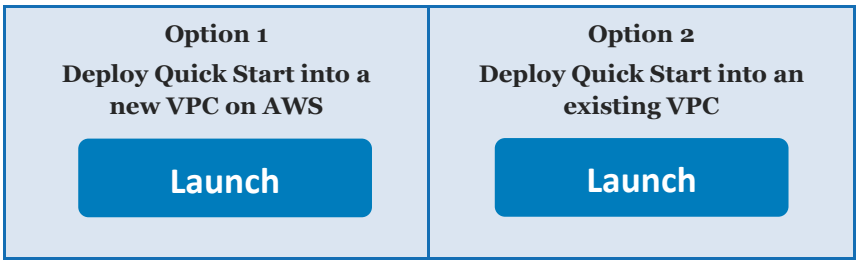
Step 1. Prepare an AWS Account

1. If you don't already have an AWS account, create one at https://aws.amazon.com by following the on-screen instructions.
2. Use the region selector in the navigation bar to choose the AWS Region where you want to deploy the data warehousing Quick Start on AWS.
3. Create a key pair in your preferred region.

4. If necessary, request a service limit increase for the Amazon EC2 **m4.4xlarge** and **t2.large** instance types. You might need to do this if you already have an existing deployment that uses these instance types, and you think you might exceed the default limits with this reference deployment.
5. If necessary, request a service limit increase for AWS CloudFormation stacks. Depending on the options that you choose while launching, this Quick Start will create up to 11 new stacks within your account.

Step 2. Launch the Quick Start

1. Deploy the AWS CloudFormation template into your AWS account by choosing one of the following links. For more information, see Technical Requirements earlier in this guide.



   Each set of stacks takes approximately 50 minutes to create. The deployment takes an additional 20-25 minutes if you choose to install the Quick Start dataset. If you choose Option 2, make sure that your existing VPC is consistent with that produced by the Amazon VPC Quick Start.

> **Note** 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. See the pricing pages for each AWS service you will be using for full details.

2. Check the region that's displayed in the upper-right corner of the navigation bar, and change it if necessary. This is where the Quick Start will be deployed. The template is launched in the US East (Ohio) Region by default.
3. On the **Select Template** page, keep the default setting for the template URL, and then choose **Next**.
4. On the **Specify Details** page, review the parameters for the template. Enter values for the parameters that require your input. For all other parameters, you can customize the default settings provided by the template.

In the following tables, parameters are listed by category and described separately for the two deployment options:

- Parameters for deploying the Quick Start into a new VPC
- Parameters for deploying the Quick Start into an existing VPC

- **Option 1: Parameters for deploying the Quick Start into a new VPC**

View template

Network Configuration:

Parameter label (name)	Default	Description
Availability Zones (AvailabilityZones)	*Requires input*	The list of Availability Zones to use for the subnets in the VPC. By default, the Quick Start preserves the logical order you specify. Be sure to include at least two Availability Zones, or as many as required by your VPC definition.
VPC Definition (VPCDefinition)	QuickstartDefault	VPC definition name from the `Mappings` section of the template. Each definition specifies a VPC configuration, including the number of Availability Zones to be used for the deployment and the CIDR blocks for the VPC, public subnets, and private subnets. You can support multiple VPC configurations by extending the map with additional definitions and choosing the appropriate name. If you do not want to change the VPC configuration, keep the default setting. For more information, see the Optional: Adding VPC Definitions section.
RDS Inbound CIDR (RDSInboundCIDR)	<VPCCIDR>	Allow inbound traffic to Amazon RDS from this CIDR block. The default setting uses the CIDR block of the new VPC created by the Quick Start.
Redshift Inbound CIDR (RedshiftInboundCIDR)	<VPCCIDR>	Allow inbound traffic to the Amazon Redshift cluster from this CIDR block. The default setting uses the CIDR block of the new VPC created by the Quick Start.
Tableau Server Inbound CIDR (TableauServerInboundCIDR)	<VPCCIDR>	CIDR block from which you may connect to the Tableau Server web interface or to the bastion host. The default setting uses the CIDR block of the new VPC created by the Quick Start.
Bastion Inbound CIDR (BastionInboundCIDR)	*Requires input*	The CIDR block that's allowed external SSH access to the bastion hosts. 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.

Parameter label (name)	Default	Description
Remote Desktop Gateway Inbound CIDR (RDGWInboundCIDR)	*Requires input*	The CIDR block that's allowed external access to the Remote Desktop gateways. Set this parameter to the string `<VPCCIDR>` to use the CIDR of the new VPC created by the Quick Start.
Hosted Zone Name (HostedZoneName)	<NONE>	The name of the hosted zone within which the Quick Start will create convenience DNS entries for AWS resources. If you don't want to create convenience DNS entries or you're not using Amazon Route 53 for DNS, keep the default setting. Otherwise, enter the hosted zone name, including the trailing period; for example, `dev.example.com.` For more information, see Amazon Route 53 Convenience DNS Names earlier in this guide.

Amazon EC2 Common Configuration:

Parameter label (name)	Default	Description
Key pair name (KeyPairName)	*Requires input*	Public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.

Linux Bastion Configuration:

Parameter label (name)	Default	Description
Bastion Instance Type (BastionInstanceType)	t2.micro	The EC2 instance type for the bastion host instances.
Bastion AMI Operating System (BastionAMIOS)	Amazon-Linux-HVM	The Linux distribution for the AMI to be used for the bastion host instances. You can choose Amazon Linux, CentOS, or Ubuntu Server. If you choose CentOS, make sure that you have a subscription to the CentOS AMI in AWS Marketplace.
Bastion Banner (BastionBanner)	https://s3.amazonaws.com/quickstart-reference/linux/bastion/latest/scripts/banner_message.txt	The URL for the ASCII text file that contains the banner text to display upon login. (For more information, see the section on customizing the banner in the Quick Start for Linux bastion hosts.)

Remote Desktop Gateway Configuration:

Parameter label (name)	Default	Description
Remote Desktop Gateway Instance Type (RDGWInstanceType)	t2.large	The EC2 instance type for the Remote Desktop Gateway instances.

Parameter label (name)	Default	Description
Remote Desktop Gateway Admin User Name (RDGWAdminUser)	admin	The user name for the local administrator account for Remote Desktop Gateway.
Remote Desktop Gateway Admin Password (RDGWAdminPassword)	*Requires input*	The password for the local administrator account for Remote Desktop Gateway. The password must be at least 8 characters, and may include letters, numbers, and symbols.
Domain DNS Name (DomainDNSName)	example.com	Fully qualified domain name (FQDN) of the forest root domain.

Amazon Redshift Configuration:

Parameter label (name)	Default	Description
Redshift Database Name (RedshiftDatabaseName)	quickstart	The name of the first database to be created when the Amazon Redshift cluster is provisioned.
Redshift Number of Nodes (RedshiftNumberOfNodes)	2	The number of compute nodes in the Amazon Redshift cluster. If you specify a number larger than 1, the Quick Start will launch a multi-node cluster; otherwise, the Quick Start will launch a single-node cluster.
Redshift Node Type (RedshiftNodeType)	dc1.large	The node type to be provisioned for the Amazon Redshift cluster.
Redshift Username (RedshiftUsername)	redshift	The user name associated with the master user account for the Amazon Redshift cluster.
Redshift Password (RedshiftPassword)	*Requires input*	The password associated with the master user account for the Amazon Redshift cluster.
Redshift Database Port (RedshiftDatabasePort)	5439	The port that Amazon Redshift will listen on, which will be allowed through the security group.

Amazon RDS PostgreSQL Configuration:

Parameter label (name)	Default	Description
RDS Instance Identifier (RDSInstanceIdentifier)	quickstart-tableau-awsdata	The identifier to use for the RDS instance.
RDS Database Port (RDSDatabasePort)	5432	The port that the RDS instance will listen on, which will be allowed through the security group.
RDS Database Name (RDSDatabaseName)	quickstart	The name of the initial database created when the Amazon RDS cluster is provisioned.
RDS MultiAZ (RDSMultiAZ)	true	Set to **false** if Multi-AZ deployment isn't needed for Amazon RDS.

Parameter label (name)	Default	Description
RDS Instance Class (RDSInstanceClass)	db.t2.large	Database instance class for the provisioned RDS instance.
RDS Allocated Storage (RDSAllocatedStorage)	5	The size of the PostgreSQL database, in GiBs.
RDS Username (RDSUsername)	postgres	The user name associated with the administrator account for the provisioned PostgreSQL RDS instance.
RDS Password (RDSPassword)	*Requires input*	The password associated with the administrator account for the provisioned PostgreSQL RDS instance.

SQL Worker Configuration:

Parameter label (name)	Default	Description
SQL Worker Instance Type (SqlWorkerInstanceType)	t2.micro	EC2 instance type for the SQL Server worker instance.

Tableau Server Configuration:

Parameter label (name)	Default	Description
Content Admin User (ContentAdminUser)	admin	The user name of the initial administrator for Tableau Server.
Content Admin Password (ContentAdminPassword)	*Requires input*	The user password of the initial administrator for Tableau Server.
Reg First Name (RegFirstName)	*Requires input*	First name of the Tableau Server user.
Reg Last Name (RegLastName)	*Requires input*	Last name of the Tableau Server user.
Reg Email (RegEmail)	*Requires input*	Email address of the Tableau Server user.
Reg Company (RegCompany)	*Requires input*	Company that the Tableau Server user works for.
Reg Title (RegTitle)	*Requires input*	Job title of the Tableau Server user.
Reg Department (RegDepartment)	*Requires input*	Department of the Tableau Server user.
Reg Industry (RegIndustry)	*Requires input*	Industry that the Tableau Server user works in.

Parameter label (name)	Default	Description
Reg Phone (RegPhone)	*Requires input*	Telephone number of the Tableau Server user.
Reg City (RegCity)	*Requires input*	City that the Tableau Server user lives in.
Reg State (RegState)	*Requires input*	State that the Tableau Server user lives in.
Reg Zip (RegZip)	*Requires input*	Zip code or postal code of the Tableau Server user.
Reg Country (RegCountry)	*Requires input*	Country that the Tableau Server user lives in.
Tableau Server License Key (TableauServerLicenseKey)	*Optional*	The Tableau Server license key. If you don't specify a license key, Tableau Server will automatically be activated with a 14-day trial. At the end of the trial period, you must enter a product key to continue using Tableau Server. For more information about Tableau Server licenses, see the Tableau website.

AWS Quick Start Configuration:

Parameter label (name)	Default	Description
Quick Start S3 Bucket Name (QSS3BucketName)	quickstart-reference	S3 bucket where the Quick Start templates and scripts are installed. Use this parameter to specify the S3 bucket name you've created for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.
Quick Start S3 Key Prefix (QSS3KeyPrefix)	tableau/awsdata/latest/	S3 key prefix used to simulate a folder for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes.
Load dataset (LoadDataset)	true	By default, the Quick Start will load the sample dataset described in the Quick Start Dataset section. If you don't want to load a dataset, set this parameter to **false**. If you'd like to load your own dataset, keep the default setting of **true**, and update the next three dataset-related parameters to point to a dataset that has been prepared as described in Using Your Own Dataset.
Dataset S3 Bucket Name (DatasetS3BucketName)	aws-quickstart-datasets	S3 bucket where the sample dataset is installed. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should

Parameter label (name)	Default	Description
		not start or end with a hyphen. Use this parameter to specify the S3 bucket name for the dataset you'd like to load, if you decide to use a different dataset, or if you decide to customize or extend the Quick Start dataset. (For more information, see Using Your Own Dataset.) Keep the default setting to use the sample dataset included with the Quick Start.
Dataset S3 Key Prefix (DatasetS3KeyPrefix)	tableau-awsdata/synthetic-store-sales/v1/	S3 key prefix where the sample dataset is installed. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes. Use this parameter to specify the location for the dataset you'd like to load, if you decide to use a different dataset, or if you decide to customize or extend the Quick Start dataset. (For more information, see Using Your Own Dataset.) Keep the default setting to use the sample dataset included with the Quick Start.
Dataset S3 Region (DatasetS3Region)	us-east-1	The AWS Region where the S3 dataset is located, used for the Amazon Redshift `COPY` command. To obtain the best Amazon Redshift `COPY` performance, use a copy of the Quick Start dataset in the same region where you are launching the Quick Start. Keep the default setting to use the sample dataset included with the Quick Start.

- **Option 2: Parameters for deploying the Quick Start into an existing VPC**

View template

Network Configuration:

Parameter label (name)	Default	Description
Existing VPC ID (VPCID)	*Requires input*	ID of your existing VPC (e.g., vpc-0343606e).
Existing VPC CIDR (VPCCIDR)	10.0.0.0/16	CIDR block for your existing VPC.
Existing VPC Private Subnet 1 ID (PrivateSubnet1ID)	*Requires input*	ID of the private subnet in Availability Zone 1 (e.g., subnet-a0246dcd).
Existing VPC Private Subnet 2 ID (PrivateSubnet2ID)	*Requires input*	ID of the private subnet in Availability Zone 2 (e.g., subnet-b58c3d67).
Existing VPC Public Subnet1 ID (PublicSubnet1ID)	*Requires input*	ID of the public subnet in Availability Zone 1 (e.g., subnet-9bc642ac).

Parameter label (name)	Default	Description
Existing VPC Public Subnet2 ID (PublicSubnet2ID)	*Requires input*	ID of the public subnet in Availability Zone 2 (e.g., subnet-e3246d8e).
RDS Inbound CIDR (RDSInboundCIDR)	<VPCCIDR>	Allow inbound traffic to Amazon RDS from this CIDR block. The default setting uses the CIDR block you specified in the **VPCCIDR** parameter.
Redshift Inbound CIDR (RedshiftInboundCIDR)	<VPCCIDR>	Allow inbound traffic to the Amazon Redshift cluster from this CIDR block. The default setting uses the CIDR block you specified in the **VPCCIDR** parameter.
Tableau Server Inbound CIDR (TableauServerInbound CIDR)	<VPCCIDR>	CIDR block from which you may connect to web interface or bastion host. The default setting uses the CIDR block you specified in the **VPCCIDR** parameter.
Hosted Zone Name (HostedZoneName)	<NONE>	The name of the hosted zone within which the Quick Start will create convenience DNS entries for AWS resources. If you don't want to create convenience DNS entries or you're not using Amazon Route 53 for DNS, keep the default setting. Otherwise, enter the hosted zone name, including the trailing period; for example, `dev.example.com.` For more information, see Amazon Route 53 Convenience DNS Names earlier in this guide.
Bastion Elastic IP (BastionEIP)	169.254.169.254	Public IP of the bastion host to use to create convenience DNS entries for AWS resources. If you haven't specified a hosted zone name parameter, this parameter will not be used and can remain set to its default value.
RDGW Elastic IP (RDGWElasticIP)	169.254.169.254	Public IP of the Remote Desktop Gateway instance to use to create convenience DNS entries for AWS resources. If you haven't specified a hosted zone name, this parameter will not be used and can remain set to its default value.

Amazon EC2 Common Configuration:

Parameter label (name)	Default	Description
Key pair name (KeyPairName)	*Requires input*	Public/private key pair, which allows you to connect securely to your instance after it launches. When you created an AWS account, this is the key pair you created in your preferred region.

Amazon Redshift Configuration:

Parameter label (name)	Default	Description
Redshift Database Name (RedshiftDatabaseName)	quickstart	The name of the first database to be created when the Amazon Redshift cluster is provisioned.
Redshift Number of Nodes (RedshiftNumberOfNodes)	2	The number of compute nodes in the Amazon Redshift cluster. If you specify a number larger than 1, the Quick Start will launch a multi-node cluster; otherwise, the Quick Start will launch a single-node cluster.
Redshift Node Type (RedshiftNodeType)	dc1.large	The node type to be provisioned for the Amazon Redshift cluster.
Redshift Username (RedshiftUsername)	redshift	The user name associated with the master user account for the Amazon Redshift cluster.
Redshift Password (RedshiftPassword)	*Requires input*	The password associated with the master user account for the Amazon Redshift cluster.
Redshift Database Port (RedshiftDatabasePort)	5439	The port that Amazon Redshift will listen on, which will be allowed through the security group.

Amazon RDS PostgreSQL Configuration:

Parameter label (name)	Default	Description
RDS Instance Identifier (RDSInstanceIdentifier)	quickstart-tableau-awsdata	The identifier to use for the RDS instance.
RDS Database Port (RDSDatabasePort)	5432	The port that the RDS instance will listen on, which will be allowed through the security group.
RDS Database Name (RDSDatabaseName)	quickstart	The name of the initial database created when the Amazon RDS cluster is provisioned.
RDS MultiAZ (RDSMultiAZ)	true	Set to **false** if Multi-AZ deployment isn't needed for Amazon RDS.
RDS Instance Class (RDSInstanceClass)	db.t2.large	Database instance class for the provisioned RDS instance.
RDS Allocated Storage (RDSAllocatedStorage)	5	The size of the PostgreSQL database, in GiBs.
RDS Username (RDSUsername)	postgres	The user name associated with the administrator account for the provisioned PostgreSQL RDS instance.
RDS Password (RDSPassword)	*Requires input*	The password associated with the administrator account for the provisioned PostgreSQL RDS instance.

SQL Worker Configuration:

Parameter label (name)	Default	Description
SQL Worker Instance Type (SqlWorkerInstanceType)	t2.micro	EC2 instance type for the SQL Server worker instance.

Tableau Server Configuration:

Parameter label (name)	Default	Description
Content Admin User (ContentAdminUser)	admin	The user name of the initial administrator for Tableau Server.
Content Admin Password (ContentAdminPassword)	*Requires input*	The user password of the initial administrator for Tableau Server.
Reg First Name (RegFirstName)	*Requires input*	First name of the Tableau Server user.
Reg Last Name (RegLastName)	*Requires input*	Last name of the Tableau Server user.
Reg Email (RegEmail)	*Requires input*	Email address of the Tableau Server user.
Reg Company (RegCompany)	*Requires input*	Company that the Tableau Server user works for.
Reg Title (RegTitle)	*Requires input*	Job title of the Tableau Server user.
Reg Department (RegDepartment)	*Requires input*	Department of the Tableau Server user.
Reg Industry (RegIndustry)	*Requires input*	Industry that the Tableau Server user works in.
Reg Phone (RegPhone)	*Requires input*	Telephone number of the Tableau Server user.
Reg City (RegCity)	*Requires input*	City that the Tableau Server user lives in.
Reg State (RegState)	*Requires input*	State that the Tableau Server user lives in.
Reg Zip (RegZip)	*Requires input*	Zip code or postal code of the Tableau Server user.
Reg Country (RegCountry)	*Requires input*	Country that the Tableau Server user lives in.

Parameter label (name)	Default	Description
Tableau Server License Key (TableauServerLicenseKey)	*Optional*	The Tableau Server license key. If you don't specify a license key, Tableau Server will automatically be activated with a 14-day trial. At the end of the trial period, you must enter a product key to continue using Tableau Server. For more information about Tableau Server licenses, see the Tableau website.

AWS Quick Start Configuration:

Parameter label (name)	Default	Description
Stack Name (StackName)	<NONE>	Name of top-level stack. This string will be used for DNS entries. For more information, see Amazon Route 53 Convenience DNS Names earlier in this guide.
Quick Start S3 Bucket Name (QSS3BucketName)	quickstart-reference	S3 bucket where the Quick Start templates and scripts are installed. Use this parameter to specify the S3 bucket name you've created for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen.
Quick Start S3 Key Prefix (QSS3KeyPrefix)	tableau/awsdata/latest/	S3 key prefix used to simulate a folder for your copy of Quick Start assets, if you decide to customize or extend the Quick Start for your own use. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes.
Load dataset (LoadDataset)	true	By default, the Quick Start will load the sample dataset described in the Quick Start Dataset section. If you don't want to load a dataset, set this parameter to **false**. If you'd like to load your own dataset, keep the default setting of **true**, and update the next three dataset-related parameters to point to your dataset that has been prepared as described in Using Your Own Dataset..
Dataset S3 Bucket Name (DatasetS3BucketName)	aws-quickstart-datasets	S3 bucket where the Quick Start dataset is installed. The bucket name can include numbers, lowercase letters, uppercase letters, and hyphens, but should not start or end with a hyphen. Use this parameter to specify the S3 bucket name for the dataset you'd like to load, if you decide to use a different dataset, or if you decide to customize or extend the Quick Start dataset. (For more information, see Using Your Own Dataset.) Keep the default setting to use the sample dataset included with the Quick Start.

Parameter label (name)	Default	Description
Dataset S3 Key Prefix (DatasetS3KeyPrefix)	tableau-awsdata/synthetic-store-sales/v1/	S3 key prefix where the Quick Start dataset is installed. This prefix can include numbers, lowercase letters, uppercase letters, hyphens, and forward slashes. Use this parameter to specify the location for the dataset you'd like to load, if you decide to use a different dataset, or if you decide to customize or extend the Quick Start dataset. (For more information, see Using Your Own Dataset.) Keep the default setting to use the sample dataset included with the Quick Start.
Dataset S3 Region (DatasetS3Region)	us-east-1	The AWS Region where the S3 dataset is located, used for the Amazon Redshift `COPY` command. To obtain the best Amazon Redshift `COPY` performance, use a copy of the Quick Start dataset in the same region where you are launching the Quick Start. Keep the default setting to use the sample dataset included with the Quick Start.

5. 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**.
6. On the **Review** page, review and confirm the template settings. Under **Capabilities**, select the check box to acknowledge that the template will create IAM resources.

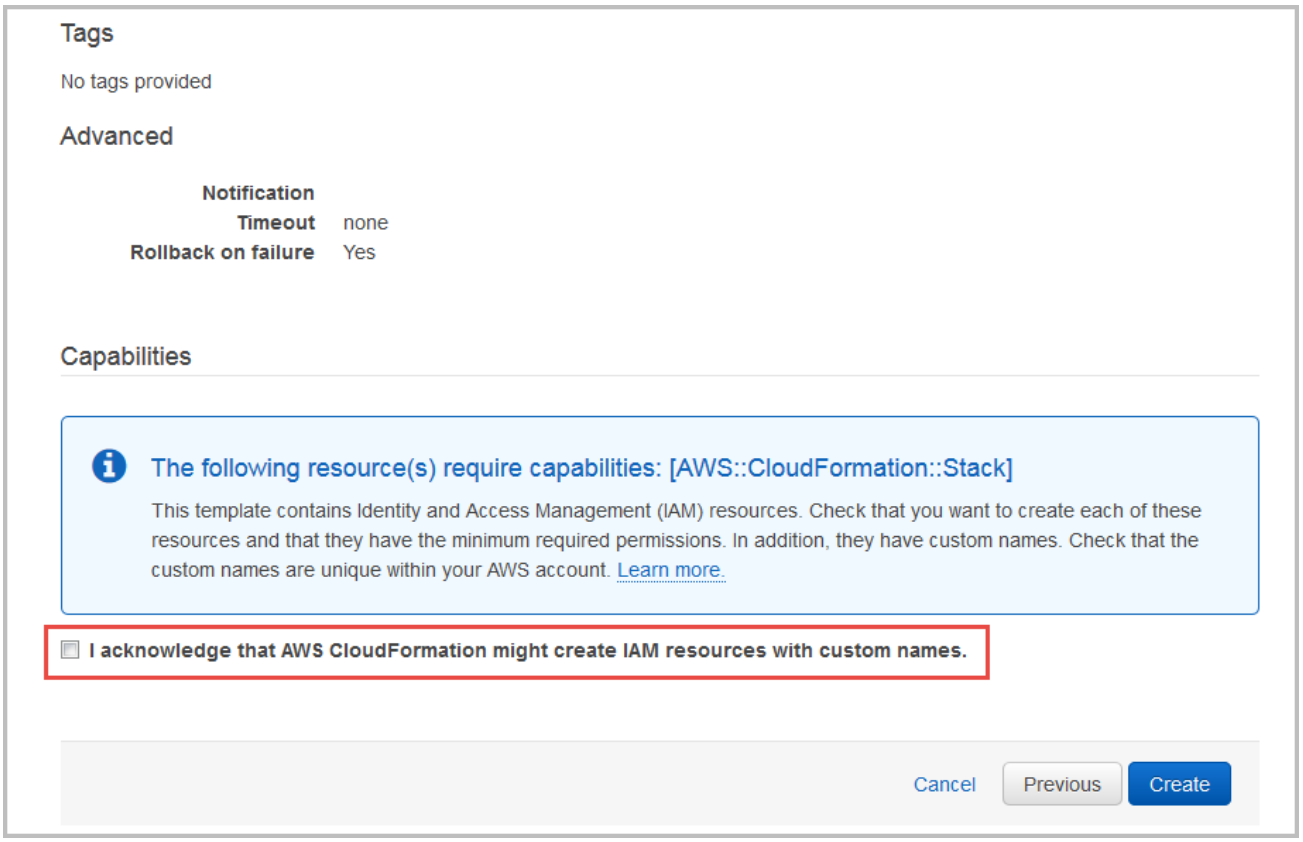


Figure 2: Check box for IAM resources

7. Choose **Create** to deploy the stack.
8. Monitor the status of the stack. When the status is **CREATE_COMPLETE**, the deployment is complete.

Step 3. Test the Deployment

1. Navigate to Tableau Server by pointing a web browser to the URL for **TableauServerURL** in the **Outputs** tab. You must access this from an endpoint that's allowed by the **Tableau Server Inbound CIDR** parameter of the stack.
2. View the data sources within Tableau Server. The Quick Start configures distinct data sources for selected tables in Amazon Redshift, and to published aggregates in Amazon RDS. In this architecture, Amazon Redshift can compute aggregates on a periodic basis. Amazon RDS effectively functions as a highly available, scalable cache for aggregate data that can be consumed by a large number of seats on a very frequent basis.
3. View the dashboards that have been published to Tableau Server from the sample dataset. The dashboards are designed to combine AWS data sources and use the

architecture in powerful ways while retaining good performance as the volume of data, number of users, and scale of queries increase over time.

Optional: Using Your Own Dataset

The Quick Start uses a sample dataset that you can replace with your own data, along with corresponding SQL and Tableau dashboards, if needed for your use case. This section provides additional details about how to structure your own dataset and where to store its components so that it will work with the Quick Start.

Dataset Location and Structure

The dataset used by the Quick Start is stored in the location determined by the following parameters when you launch the Quick Start: **DatasetS3BucketName**, **DatasetS3KeyPrefix**, and **DatasetS3Region**. The dataset has the following structure:

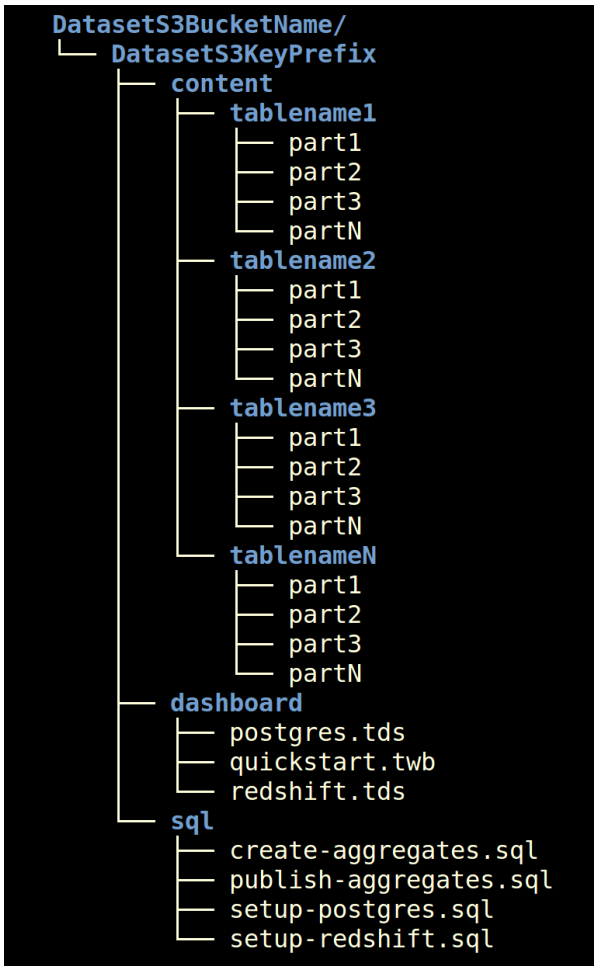


Figure 3: Dataset structure

Dataset Content

The dataset content is the data itself, typically the contents of each fact table and dimension table in your dataset. You can include as few or as many tables as you'd like. The `setup-redshift.sql` that you provide elsewhere in the dataset should contain the DDL corresponding to each table within your dataset's `content/` directory.

Because datasets can be quite large, Amazon Redshift supports loading tables through a parallel copy operation from Amazon S3 into each node in the Amazon Redshift cluster. Beneath each table, you can divide your data into multiple parts so that these parallel load operations complete more quickly. You can store your data in any format supported by the Amazon Redshift `COPY` command. The SQL `COPY` commands in your `setup-redshift.sql` should be consistent with the format that you have chosen for your data.

If you divide the data for a table into multiple parts, you can either provide a manifest file that lists all the parts that Amazon Redshift should load, or you can specify a key prefix, and Amazon Redshift will load data from any objects that match that prefix. If you decide to use a key prefix instead of a manifest file, Amazon Redshift will require the `S3:ListBucket` permission so that it can iterate through the objects that match the key prefix. The Quick Start provides this permission in the IAM role that it creates and associates with Amazon Redshift.

Some organizations that host S3 buckets that provide public datasets prevent users from accessing the `S3:ListBucket` action as a matter of policy. If you would like your dataset to reside in such an S3 bucket, you should specify a manifest file in your SQL `COPY` commands instead of using a key prefix.

Dataset SQL

The Quick Start creates two database resources that hold different subsets of your data:

- An Amazon Redshift cluster to act as a data warehouse that can hold all fact and dimension content and that can efficiently manufacture aggregates at large data scale.
- An Amazon RDS PostgreSQL instance to store computed aggregates with high availability and scalable read throughput suitable for large numbers of users and concurrent queries.

After provisioning these database resources, the Quick Start executes the SQL that you provide in each of the following four files, in the following order, for the stated purpose:

1. `setup-redshift.sql` – Executed on Amazon Redshift. Creates tables with the specified DDL, and copies data from Amazon S3 into each created table. You could also choose to create any required schemas and users here.

2. `setup-postgres.sql` – Executed on Amazon RDS PostgreSQL. The sample dataset creates a DB link that allows Amazon RDS to use Amazon Redshift as a foreign data source. You may also wish to include DDL for small dimension tables and copy dimension data into Amazon RDS at the outset. You could also choose to create any specified schemas and users here.
3. `create-aggregates.sql` – Executed on Amazon Redshift. This SQL computes aggregates that can be offloaded to the RDS DB instance to service very large query volumes, or alternatively specifies views such that aggregate computation occurs when the view is queried by the PostgreSQL DB link to Amazon Redshift.
4. `publish-aggregates.sql` – Executed on PostgreSQL RDS. Inserts aggregates queried from Amazon Redshift into aggregate tables on Amazon RDS PostgreSQL so that they can be served with high availability and scalable read throughput with high query volumes.

The Quick Start runs each of these SQL fragments exactly once, in the order listed. However, you should consider authoring each SQL fragment so that they are idempotent; that is, so they can be executed multiple times without undesired side effects, to ensure predictable results. For example, you may want to consider including `DROP TABLE IF EXISTS` prior to your `CREATE TABLE` statements so that setup SQL can be run multiple times with predictable results if the underlying data is updated. Similarly, `publish-aggregates` could be run on a schedule to ensure that aggregates remain current within a 10 to 15-minute window. This will increase the flexibility for maintaining your dataset incrementally after the Quick Start is deployed.

Dataset Dashboard

The dataset's `dashboard/` directory contains Tableau data source files for Amazon Redshift and Amazon RDS PostgreSQL. The Quick Start publishes these files to Tableau Server as Data Server data sources after the database resources become available and are ready to serve. You should not need to change either of these data sources. The Quick Start updates the data sources before publishing them to Tableau Server so that their endpoints correspond to those of the actual database resources provisioned within the Quick Start.

The `dashboard/` directory also contains a workbook that is published to Tableau Server. You can modify this workbook to describe your dataset. This workbook should be named workbook.twb and should include the following:

- Each live Amazon Redshift and PostgreSQL data source. The Quick Start has no capability to automatically refresh extracts.
- Any dashboards that are relevant and valuable in the context of your dataset.

Optional: Adding VPC Definitions

When you launch the Quick Start in the mode where a new VPC is created, the Quick Start uses VPC parameters that are defined in a mapping within the Quick Start templates. If you choose to download the templates from the GitHub repository, you can add new named VPC definitions to the mapping, and choose one of the named VPC definitions that you have defined when you launch the Quick Start.

The following table shows the parameters defined within each VPC definition. You can define as many VPC definitions as you need within your environments. When you deploy the Quick Start, use the **VPCDefinition** parameter to specify the configuration you want to use.

Parameter	Default	Description
CreateAdditional PrivateSubnets	false	Set to **true** to create a network ACL-protected subnet in each Availability Zone. If **false**, the CIDR parameters for those subnets will be ignored.
NATInstanceType	t2.small	EC2 instance type for the NAT instances. This parameter is used only if the AWS Region doesn't support NAT gateways.
NumberOfAZs	2	Number of Availability Zones to use in the VPC.
PrivateSubnet1A CIDR	10.0.0.0/19	CIDR block for private subnet 1A located in Availability Zone 1.
PrivateSubnet1B CIDR	10.0.192.0/21	CIDR block for private subnet 1B with dedicated network ACL located in Availability Zone 1.
PrivateSubnet2A CIDR	10.0.32.0/19	CIDR block for private subnet 2A located in Availability Zone 2.
PrivateSubnet2B CIDR	10.0.200.0/21	CIDR block for private subnet 2B with dedicated network ACL located in Availability Zone 2.
PrivateSubnet3A CIDR	10.0.64.0/19	CIDR block for private subnet 3A located in Availability Zone 3.
PrivateSubnet3B CIDR	10.0.208.0/21	CIDR block for private subnet 3B with dedicated network ACL located in Availability Zone 3.
PrivateSubnet4A CIDR	10.0.96.0/19	CIDR block for private subnet 4A located in Availability Zone 4.
PrivateSubnet4B CIDR	10.0.216.0/21	CIDR block for private subnet 4B with dedicated network ACL located in Availability Zone 4.
PublicSubnet1 CIDR	10.0.128.0/20	CIDR block for the public (DMZ) subnet 1 located in Availability Zone 1.
PublicSubnet2 CIDR	10.0.144.0/20	CIDR block for the public (DMZ) subnet 2 located in Availability Zone 2.

Parameter	Default	Description
PublicSubnet3 CIDR	10.0.160.0/20	CIDR block for the public (DMZ) subnet 3 located in Availability Zone 3.
PublicSubnet4 CIDR	10.0.176.0/20	CIDR block for the public (DMZ) subnet 4 located in Availability Zone 4.
VPCCIDR	10.0.0.0/16	CIDR block for the VPC.

Troubleshooting and FAQ

Q. I encountered a CREATE_FAILED error when I launched the Quick Start. What should I do?

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. (You'll want to look at the log files in `%ProgramFiles%\Amazon\EC2ConfigService` and `C:\cfn\log`.)

> **Important** When you set **Rollback on failure** to **No**, you'll continue to incur AWS charges for this stack. Please make sure to delete the stack when you've finished troubleshooting.

For additional information, see Troubleshooting AWS CloudFormation on the AWS website or contact us on the AWS Quick Start Discussion Forum.

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 location we've provided or from another S3 bucket. If you deploy the templates from a local copy on your computer or from a non-S3 location, you might encounter template size limitations when you create the stack. For more information about AWS CloudFormation limits, see the AWS documentation.

Q. Can I use the QuickStart with my own data?

A. Of course! The structure within the dataset (per-table content; purpose-specific SQL for database and table setup, aggregation and publishing; and consistent workbooks, dashboards and data sources) makes it easy to extend or replace the sample dataset that is included with the Quick Start.

Q. Does my data need to be public for the Quick Start to work?

A. No, it doesn't. You can set the dataset location parameters (**DatasetS3BucketName**, **DatasetS3KeyPrefix**, and **DatasetS3Region**) to refer to a private location in Amazon S3. The Quick Start will create IAM roles and apply them to resources provisioned within the Quick Start so that Amazon Redshift and Tableau Server will be able to access the required data.

Q. Can I use this Quick Start to set up a Tableau Server production environment?

A. This Quick Start uses a Tableau Server trial license, which supports a single Tableau Server instance. If you want to deploy Tableau Server in a production environment, we recommend that you sign up for a standard Tableau Server license, and customize the Quick Start templates to deploy a Tableau Server cluster.

Additional Resources

AWS services

- AWS CloudFormation
  http://aws.amazon.com/documentation/cloudformation/
- Amazon EC2
  http://docs.aws.amazon.com/AWSEC2/latest/WindowsGuide/
- Amazon Redshift
  https://aws.amazon.com/documentation/redshift/
- Amazon RDS
  https://aws.amazon.com/documentation/rds/
- Amazon S3
  https://aws.amazon.com/documentation/s3/
- Amazon VPC
  http://aws.amazon.com/documentation/vpc/

Tableau

- Tableau Desktop
  http://www.tableau.com/products/desktop
- Tableau Server
  http://www.tableau.com/products/server

- Designing Efficient Workbooks whitepaper
  https://www.tableau.com/learn/whitepapers/designing-efficient-workbooks

Quick Start reference deployments

- AWS Quick Start home page
  https://aws.amazon.com/quickstart/

Send Us Feedback

You can visit our GitHub repository to download the templates and scripts for this Quick Start, to post your comments, and to share your customizations with others.

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.