Informatica Secure@Source on the AWS Cloud

Quick Start Reference Deployment

November 2017
Last update: March 2018 (revisions)

Informatica Data Security Group
Santiago Cardenas and Vinod Shukla, AWS Quick Start Reference Team

Contents
Overview..................................................................................................................................................2
Costs and Licenses ..................................................................................................................................2
Architecture...............................................................................................................................................3
Prerequisites .............................................................................................................................................6
   Specialized Knowledge ......................................................................................................................6
Deployment Options ...............................................................................................................................7
Deployment Steps .....................................................................................................................................7
   Step 1. Prepare an AWS Account ........................................................................................................7
   Step 2. Obtain a License for Secure@Source ....................................................................................8
   Step 3. Launch the Quick Start ........................................................................................................8
   Step 4. Get Started with Secure@Source ........................................................................................14
FAQ.......................................................................................................................................................15
Additional Resources ............................................................................................................................17
GitHub Repository ...................................................................................................................................17
Document Revisions ..............................................................................................................................17
This Quick Start deployment guide was created by Amazon Web Services (AWS) in partnership with Informatica.

Overview

This Quick Start reference deployment guide provides step-by-step instructions for deploying Informatica Secure@Source on the AWS Cloud. Quick Starts are automated reference deployments for key technologies on the AWS Cloud, based on AWS best practices for security and high availability.

Informatica Secure@Source delivers global visibility of sensitive data assets, actionable insights into sensitive data risks, timely detection of insider and outsider threats, and accurate identification of high risk conditions; supporting data security, compliance, and governance.

Secure@Source provides visibility and control of sensitive data, allowing users to:

- **Confirm what they know about their sensitive data** – Provide global visibility of sensitive data with data classification, discovery, proliferation analysis, user access, and activity correlation and visualization for management and practitioners.
- **Monitor risk on a continuous basis** – Track sensitive data risk and remediation with multi-factor risk scoring that helps identify top risk areas based on organizational requirements.
- **Uncover the unexpected** – Detect suspicious or unauthorized data access by continuously correlating, base-lining, detecting, and alerting on high risk conditions and potential anomalous behaviors that threaten sensitive data.
- **Remediate risk** – Orchestrate data security controls to protect data at rest, prevent unauthorized access, and de-identify or anonymize sensitive data.

This Quick Start is for IT and system administrators responsible for deploying Secure@Source on the AWS Cloud in order to monitor, analyze, and prioritize sensitive data risk and obtain actionable insights.

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 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. See the pricing pages for each AWS service you will be using for cost estimates.

This Quick Start requires a license for Informatica Secure@Source, as described in step 2 of the deployment instructions. You can sign up for a trial license by registering on the Informatica website.

Architecture

Deploying this Quick Start into a new virtual private cloud (VPC) with the default parameters builds the following Secure@Source environment in the AWS Cloud.

Figure 1: Quick Start deployment architecture for Secure@Source on AWS
The architecture includes the following components:

- A VPC configured across two Availability Zones. For each Availability Zone, this Quick Start provisions one private and one public subnet. The Secure@Source deployment uses both subnets in each Availability Zone.

- Managed network address translation (NAT) gateways deployed into the public subnets and configured with an Elastic IP address for outbound internet connectivity.

- An IAM role with fine-grained permissions for access to AWS services necessary for the deployment process.

- Appropriate security groups for each instance or function to restrict access to only necessary protocols and ports.

- In the private subnets, Informatica domain and repository databases hosted on Amazon Relational Database Service (Amazon RDS) using Microsoft SQL Server, where the Informatica domain configuration repository, the Model repository, the reference data warehouse, the profiling warehouse, and the Secure@Source repository are configured.

- In the public subnets, Amazon Elastic Compute Cloud (Amazon EC2) instances for Secure@Source, including the following:
  - A single-node or multi-node embedded cluster with a configurable cluster size. A node represents an EC2 instance (virtual machine). The Informatica domain runs on a single node with all the associated services. The cluster runs on a separate node or nodes based on your selection. The application services such as the Model Repository Service and the Data Integration Service run on the Informatica domain node.
  
  - Scanners to extract metadata from all data sources supported by Secure@Source.
  
  - Applications to capture and process user activity and user behavior analytics.
  
  - **Informatica domain**, which is the fundamental administrative unit of the Informatica platform. The Informatica platform has a service-oriented architecture that provides the ability to scale services and share resources across multiple machines.
  
  - **MRS** or **Model Repository Service**, which manages a relational database that stores all the metadata for projects created using Informatica client tools. The Model repository also stores run-time and configuration information for applications that are deployed to a Data Integration Service and the administrative module of Enterprise Information Catalog.
- **DIS** or **Data Integration Service**, which is a compute component within the Informatica domain that manages requests to submit big data integration, big data quality, and profiling jobs to the Hadoop cluster for processing.

- **Catalog Service**, which runs Enterprise Information Catalog and manages connections between service components and external applications.

- **Informatica Cluster Service**, which runs and manages all the Hadoop services, Apache Ambari server, and Apache Ambari agents on an internal Hadoop cluster.

- **CMS** or **Content Management Service**, which manages reference data. It provides reference data information to the Data Integration Service and Informatica Developer.

- **Profiling**, which discovers the functional meaning of data in a data source. A profile is a repository object that performs data domain discovery on the source metadata, data, or both. The Secure@Source Service uses the profiling results to identify and classify sensitive data in a data source.

- **Metadata and Catalog**, which include the metadata persistence store, search index, and graph database in an embedded Hortonworks cluster. The catalog represents an indexed inventory of all the data assets in the enterprise that you configure in Enterprise Information Catalog. Enterprise Information Catalog organizes all the enterprise metadata in the catalog and enables the users of external applications to discover and understand the data.

- **Secure@Source Service**, which runs the Secure@Source web application and manages communication with the Secure@Source repository.

- **Data Sources** representing the source databases or metadata sources that Secure@Source scans to identify sensitive data. You can also use Secure@Source on AWS to collect metadata from resources deployed on premises, as shown in Figure 2.
You can collect metadata both from resources deployed on a VPC and from resources deployed on premises. For on-premises resources, Secure@Source collects metadata through a virtual private network (VPN), a VPC, or by using AWS Direct Connect.

After deploying Informatica Secure@Source, you must configure the required resources to extract metadata.

If you decide to deploy Secure@Source into your existing VPC (see Deployment Options later in this guide), the Quick Start assumes that the infrastructure components already exist, and deploys Secure@Source into the environment you specify during deployment.

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 the Getting Started Resource Center.)

- Amazon Virtual Private Cloud (Amazon VPC)
- Amazon EC2
- Amazon RDS
- Amazon Simple Storage Service (Amazon S3)
- Elastic IP Addresses
Deployment Options
This Quick Start provides two deployment options:

- **Deploy Secure@Source into a new VPC** (end-to-end deployment). This option builds a new AWS environment consisting of the VPC, subnets, NAT gateways, security groups, and other infrastructure components, and then deploys Secure@Source into this new VPC.

- **Deploy Secure@Source into an existing VPC**. This option provisions Secure@Source in your existing AWS infrastructure.

The Quick Start also lets you configure additional settings such as CIDR blocks, instance types, and Secure@Source settings, as discussed later in this guide.

Deployment Steps

**Step 1. Prepare an AWS Account**

1. If you don’t already have an AWS account, create one at [http://aws.amazon.com](http://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 Secure@Source.

3. Create a key pair in your preferred region.

   When you log in to an EC2 instance, you use a password file for authentication. The file is called a private key file and has a file name extension of .pem.

   If you do not have an existing .pem key to use, follow the instructions in the [AWS documentation](http://aws.amazon.com) to create a key pair.

   **Note** Your administrator might ask you to use an existing key pair.

   When you create a key pair, you save the .pem file to your desktop system. Simultaneously, AWS saves the key pair to your account. Make a note of the key pair that you want to use for the Secure@Source instance, so that you can provide the key pair name during network configuration.

4. If necessary, request a service limit increase for the EC2 instance type you’ve decided to use (c4.4xlarge or c4.8xlarge) for the Informatica domain. You might need to do this if you already have an existing deployment that uses that instance type, and you think you might exceed the default limit with this reference deployment.
**Step 2. Obtain a License for Secure@Source**

This Quick Start requires a license for Informatica Secure@Source. To obtain a license key for Secure@Source, do the following:

1. To sign up for a trial license, please register on the Informatica website. After you register, an Informatica contact will follow up to help you procure a license. You may skip this step if you already have a valid Secure@Source license key.

2. In your AWS account, create an S3 bucket.

3. (Optional) Create a directory under the bucket.

4. Place the Secure@Source license key file for the software in the S3 bucket or directory. You will be prompted for its location in step 3.

**Step 3. Launch the 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.

1. Choose one of the following options to launch the AWS CloudFormation template into your AWS account. For help choosing an option, see deployment options earlier in this guide.

<table>
<thead>
<tr>
<th>Option 1</th>
<th>Option 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Deploy Secure@Source into a new VPC on AWS</strong></td>
<td><strong>Deploy Secure@Source into an existing VPC on AWS</strong></td>
</tr>
<tr>
<td><img src="#" alt="Launch" /></td>
<td><img src="#" alt="Launch" /></td>
</tr>
</tbody>
</table>

**Important** If you’re deploying Secure@Source into an existing VPC, make sure that your VPC has two public and two private subnets in different Availability Zones for the Informatica domain and database instances. These subnets require NAT gateways in their route tables, to allow the instances to download packages and software without exposing them to the internet. You will also need the domain name option configured in the DHCP options, as explained in the Amazon VPC documentation. You will be prompted for your VPC settings when you launch the Quick Start.

Each deployment takes approximately three hours to complete.
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 network infrastructure for Secure@Source will be built. The template is launched in the US West (Oregon) 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, 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. When you finish reviewing and customizing the parameters, choose **Next**.

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

- **Parameters for deploying Secure@Source into a new VPC**
- **Parameters for deploying Secure@Source into an existing VPC**

**Option 1: Parameters for deployment into a new VPC**

**View template**

**Network Configuration:**

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Availability Zones</td>
<td>Requires input</td>
<td>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.</td>
</tr>
<tr>
<td>VPC CIDR</td>
<td>10.0.0.0/16</td>
<td>CIDR block for the VPC.</td>
</tr>
<tr>
<td>Private Subnet 1 CIDR</td>
<td>10.0.0.0/19</td>
<td>CIDR block for the private subnet located in Availability Zone 1.</td>
</tr>
<tr>
<td>Private Subnet 2 CIDR</td>
<td>10.0.32.0/19</td>
<td>CIDR block for the private subnet located in Availability Zone 2.</td>
</tr>
<tr>
<td>Public Subnet 1 CIDR</td>
<td>10.0.128.0/20</td>
<td>CIDR block for the public (DMZ) subnet located in Availability Zone 1.</td>
</tr>
<tr>
<td>Public Subnet 2 CIDR</td>
<td>10.0.144.0/20</td>
<td>CIDR block for the public (DMZ) subnet located in Availability Zone 2.</td>
</tr>
<tr>
<td>IP Address Range</td>
<td>Requires input</td>
<td>The CIDR IP range that is permitted to access the Informatica domain and the Informatica embedded cluster. We recommend that you use a constrained CIDR range to reduce the potential of inbound attacks from unknown IP addresses. For example, to specify the</td>
</tr>
</tbody>
</table>
### Amazon EC2 Configuration:

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Pair Name (KeyPairName)</td>
<td>Requires input</td>
<td>Name of an existing Amazon EC2 key pair from step 1. You must specify this option to enable Secure Shell (SSH) access to the Informatica domain and the cluster nodes.</td>
</tr>
<tr>
<td>Informatica Domain Instance Type</td>
<td>c4.4xlarge</td>
<td>EC2 instance type for the instance that hosts the Informatica domain. The two options are c4.4xlarge and c4.8xlarge.</td>
</tr>
<tr>
<td>Informatica Embedded Cluster Size</td>
<td>Small</td>
<td>The number of nodes in the cluster. The instance type for each node is c4.8xlarge. Options are:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Small: single node</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Medium: three nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>- Large: six nodes</td>
</tr>
</tbody>
</table>

### Amazon RDS Configuration:

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica Database Instance Username (DBUser)</td>
<td>Requires input</td>
<td>User name (8-18 character string) for the RDS database instance associated with the Informatica domain and services (such as Model Repository Service, Data Integration Service, Content Management Service, Secure@Source Service).</td>
</tr>
<tr>
<td>Informatica Database Instance Password (DBPassword)</td>
<td>Requires input</td>
<td>Password (8-18 character string) for the RDS database instance associated with the Informatica domain and services (such as Model Repository Service, Data Integration Service, Content Management Service, Secure@Source Service).</td>
</tr>
</tbody>
</table>

### Informatica Secure@Source Configuration:

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica Administrator Username (InformaticaAdminUsername)</td>
<td>Requires input</td>
<td>User name for accessing Informatica Administrator.</td>
</tr>
<tr>
<td>Informatica Administrator Password (InformaticaAdminPassword)</td>
<td>Requires input</td>
<td>Password for accessing Informatica Administrator.</td>
</tr>
</tbody>
</table>
Parameter label (name) | Default | Description
--- | --- | ---
Secure@Source License Key Location (InformaticaSATSKeyS3Bucket) | Requires input | Name of the S3 bucket in your account that contains the Informatica Secure@Source license key, from step 2.
Secure@Source License Key Name (InformaticaSATSKeyName) | Requires input | The Informatica Secure@Source license key name, from step 2. If you’ve placed the license file in a folder in your S3 bucket, specify the folder name as well; for example, license-key or bucket-folder/license-key.
Import Sample Content (ImportSampleData) | Yes | Keep the default setting Yes to import the sample catalog data. You can use the sample data to get started with the product. Make sure that you do not run resources using the sample data.

**AWS Quick Start Configuration:**

Parameter label (name) | Default | Description
--- | --- | ---
Quick Start S3 Bucket Name (QSS3BucketName) | aws-quickstart | 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) | quickstart-informatica-secureatsource/ | The S3 key name 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.

- **Option 2: Parameters for deployment into an existing VPC**

  View template

**Network Configuration:**

Parameter label (name) | Default | Description
--- | --- | ---
VPC (VPCID) | Requires input | ID of your existing VPC where you want to deploy Secure@Source (for example, vpc-0343606e). The VPC must meet the following requirements:
  - It must be set up with public access through the internet via an attached internet gateway.
  - The **DNS Resolution** property of the VPC must be set to **Yes**.
  - The **Edit DNS Hostnames** property of the VPC must be set to **Yes**.
<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica Domain Subnet (InformaticaServerSubnetID)</td>
<td>Requires input</td>
<td>Publicly accessible subnet ID for the Informatica domain.</td>
</tr>
<tr>
<td>Informatica Database Subnets (DBSubnetIDs)</td>
<td>Requires input</td>
<td>IDs of two private subnets in the selected VPC. These must be in different Availability Zones in the selected VPC (for example, us-west-1b, us-west-1c).</td>
</tr>
<tr>
<td>IP Address Range (IPAddressRange)</td>
<td>Requires input</td>
<td>The CIDR IP range that is permitted to access the Informatica domain and the Informatica embedded cluster. We recommend that you use a constrained CIDR range to reduce the potential of inbound attacks from unknown IP addresses. For example, to specify the range of 10.20.30.40 to 10.20.30.49, enter 10.20.30.40/49.</td>
</tr>
</tbody>
</table>

**Amazon EC2 Configuration:**

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Pair Name (KeyName)</td>
<td>Requires input</td>
<td>Name of an existing Amazon EC2 key pair from step 1. You must specify this option to enable Secure Shell (SSH) access to the Informatica domain.</td>
</tr>
<tr>
<td>Informatica Domain Instance Type (InformaticaServerInstanceType)</td>
<td>c4.4xlarge</td>
<td>The EC2 instance type for the instance that hosts the Informatica domain. The two options are c4.4xlarge and c4.8xlarge.</td>
</tr>
<tr>
<td>Informatica Embedded Cluster Size (ICSClusterSize)</td>
<td>Small</td>
<td>The number of nodes in the cluster. The instance type for each node is c4.8xlarge. Options are:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Small: single node</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Medium: three nodes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Large: six nodes</td>
</tr>
</tbody>
</table>

**Amazon RDS Configuration:**

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica Database Instance Username (DBUsername)</td>
<td>Requires input</td>
<td>User name (8-18 character string) for the database instance associated with the Informatica domain and services (such as Model Repository Service, Data Integration Service, Content Management Service, Secure@Source Service).</td>
</tr>
<tr>
<td>Informatica Database Instance Password (DBPassword)</td>
<td>Requires input</td>
<td>Password (8-18 character string) for the database instance associated with the Informatica domain and services (such as Model Repository Service, Data Integration Service, Content Management Service, Secure@Source Service).</td>
</tr>
</tbody>
</table>
### Informatica Secure@Source Configuration:

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica Administrator Username (InformaticaAdminUsername)</td>
<td>Requires input</td>
<td>User name for accessing Informatica Administrator.</td>
</tr>
<tr>
<td>Informatica Administrator Password (InformaticaAdminPassword)</td>
<td>Requires input</td>
<td>Password for accessing Informatica Administrator.</td>
</tr>
<tr>
<td>Secure@Source License Key Location (InformaticaSATSKeyS3Bucket)</td>
<td>Requires input</td>
<td>Name of the S3 bucket in your account that contains the Informatica Secure@Source license key, from step 2.</td>
</tr>
<tr>
<td>Secure@Source License Key Name (InformaticaSATSKeyName)</td>
<td>Requires input</td>
<td>The Informatica Secure@Source license key name, from step 2. If you've placed the license file in a folder in your S3 bucket, specify the folder name as well; for example, license-key or bucket-folder/license-key.</td>
</tr>
<tr>
<td>Import Sample Content (ImportSampleData)</td>
<td>Yes</td>
<td>Keep the default setting Yes to import the sample catalog data. The resources do not get imported. You can use the sample data to get started with the product. Make sure that you do not run resources using the sample data.</td>
</tr>
</tbody>
</table>

### AWS Quick Start Configuration:

<table>
<thead>
<tr>
<th>Parameter label (name)</th>
<th>Default</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quick Start S3 Bucket Name (QSS3BucketName)</td>
<td>aws-quickstart</td>
<td>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.</td>
</tr>
<tr>
<td>Quick Start S3 Key Prefix (QSS3KeyPrefix)</td>
<td>quickstart-informatica-secureatsource/</td>
<td>The S3 key name 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.</td>
</tr>
</tbody>
</table>

5. On the **Options** page, you can specify tags (key-value pairs) for resources in your stack and set advanced options. When you are 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.
7. Choose **Create** to deploy the stack.

8. Monitor the status of the stack. When the status is CREATE_COMPLETE, the deployment is complete.

A sample URL for Secure@Source deployed on AWS is shown in Figure 3.

![CloudFormation stack deployment](image)

**Figure 3: URL to access Secure@Source on AWS**

**Step 4. Get Started with Secure@Source**

1. You can use the URL displayed in the **Outputs** tab to launch Secure@Source on AWS. See the [Informatica Secure@Source User Guide](#) for information about logging in to Secure@Source.

2. Perform the following high-level steps to identify sensitive data. See the [Informatica Secure@Source User Guide](#) for detailed instructions.
   a. Define locations to identify the geographic areas where your data sources exist.
b. Define data store groups to logically group similar data stores.

c. Define data stores to access data from data sources such as Informatica PowerCenter and relational databases.

d. Define data domains to include the rules that Secure@Source uses to identify sensitive columns in a data store.

e. Define classification policies to classify sensitive data based on industry or organization data security standards.

f. Define scans to determine how and when Secure@Source discovers and classifies sensitive data for a data store.

g. Monitor the status of the scan job to verify that the job completed successfully.

h. View the scan results to analyze the discovery and classification of sensitive data.

i. Import user and user access metadata to display which users have access and activity on sensitive data.

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. (You will want to look at the log files in /var/log/cfn-init-cmd.log.)

**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 have finished troubleshooting.

For additional information, see Troubleshooting AWS CloudFormation on the AWS website.

Q. I encountered errors while installing the Informatica domain and services. Where can I find more information?

A. View the /installation.log log file to see more information about the errors that you encountered.
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. I need to configure connectivity to data sources to be able to scan them. Where can I find more information?

A. The Secure@Source documentation describes the steps you need to follow to connect and scan various data sources supported by Secure@Source. For a detailed explanation of these steps, see the Secure@Source Administrator Guide.

Q. The Load Catalog step of my scan fails because the service fails to connect to the Model Repository Service.

A. This happens when there is no host entry corresponding to the domain instance on Hadoop nodes. To resolve this issue, add a host entry corresponding to the domain instance in /etc/hosts on all Hadoop nodes.

Q. Scans fail when the Oracle data source I am trying to scan has a RAC (Real Application Clusters) architecture.

A. This happens when there is no host entry corresponding to the RAC on Hadoop and domain nodes. To resolve this issue, add a host entry corresponding to the RAC in /etc/hosts on the domain node and on all Hadoop nodes.

Q. The profiling job fails on data stores that have a large number of tables (typically in the order of thousands).

A. This can happen when Enterprise Information Catalog runs out of memory when trying to process large amounts of data. Increase the Enterprise Information Catalog heap size to 8 GB or above, and set the scanner memory option on the data store to medium or large.
Additional Resources

AWS services

- AWS CloudFormation

- Amazon EBS

- Amazon EC2

- Amazon VPC

Informatica Secure@Source

- Informatica Secure@Source Overview

- Informatica Secure@Source user Guide

Quick Start reference deployments

- AWS Quick Start home page
  [https://aws.amazon.com/quickstart/](https://aws.amazon.com/quickstart/)

GitHub Repository

You can visit our [GitHub repository](https://github.com) to download the templates and scripts for this Quick Start, to post your comments, and to share your customizations with others.

Document Revisions

<table>
<thead>
<tr>
<th>Date</th>
<th>Change</th>
<th>In sections</th>
</tr>
</thead>
<tbody>
<tr>
<td>March 2018</td>
<td>Updated text and links to reflect version 4.1 of Secure@Source. Added new FAQs.</td>
<td>Deployment Steps, FAQ</td>
</tr>
<tr>
<td>November 2017</td>
<td>Initial publication</td>
<td>—</td>
</tr>
</tbody>
</table>