Informatica® Secure@Source
4.1

Installation and Configuration Guide

This product includes software licensed under the Academic Free License (http://www.opensource.org/licenses/afl-3.0.php), the Common Development and Distribution License (http://www.opensource.org/licenses/cddl1.php) the Common Public License (http://www.opensource.org/licenses/cpl1.0.php), the Sun Binary Code License Agreement Supplemental License Terms, the BSD License (http://www.opensource.org/licenses/bsd-license.php), the new BSD License (http://opensource.org/licenses/bsd-3-Clause), the MIT License (http://www.opensource.org/licenses/mit-license.php), the Artistic License (http://www.opensource.org/licenses/artistic-license-1.0) and the Initial Developer's Public License Version 1.0 (http://www.firebirdsql.org/en/initital-developer-s-public-license-version-1-0/).

This product includes software copyright © 2003-2006 Joe Walnes, 2006-2007 XStream Committers. All rights reserved. Permissions and limitations regarding this software are subject to terms available at http://xstream.codehaus.org/license.html. This product includes software developed by the Indiana University Extreme! Lab. For further information please visit http://www.extreme.indiana.edu/.

This product includes software Copyright (c) 2013 Frank Balkuff and Markus Moeller. All rights reserved. Permissions and limitations regarding this software are subject to terms of the MIT license.


DISCLAIMER: Informatica LLC provides this documentation "as is" without warranty of any kind, either express or implied, including, but not limited to, the implied warranties of merchantability, or use for a particular purpose. Informatica LLC does not warrant that this software or documentation is error free. The information provided in this software or documentation may include technical inaccuracies or typographical errors. The information in this software and documentation is subject to change at any time without notice.

NOTICES

This Informatica product (the "Software") includes certain drivers (the "DataDirect Drivers") from DataDirect Technologies, an operating company of Progress Software Corporation ("DataDirect") which are subject to the following terms and conditions:

1. THE DATADIRECT DRIVERS ARE PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESSED OR IMPLIED, INCLUDING BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NON-INFRINGEMENT.

2. IN NO EVENT WILL DATADIRECT OR ITS THIRD PARTY SUPPLIERS BE LIABLE TO THE END-USER CUSTOMER FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, CONSEQUENTIAL OR OTHER DAMAGES ARISING OUT OF THE USE OF THE ODBC DRIVERS, WHETHER OR NOT INFORMED OF THE POSSIBILITIES OF DAMAGES IN ADVANCE. THESE LIMITATIONS APPLY TO ALL CAUSES OF ACTION, INCLUDING, WITHOUT LIMITATION, BREACH OF CONTRACT, BREACH OF WARRANTY, NEGLIGENCE, STRICT LIABILITY, MISREPRESENTATION AND OTHER TORTS.

Publication Date: 2018-03-05
# Table of Contents

**Preface** ............................................................................. 8  
Informatica Resources ......................................................... 8  
  Informatica Network ......................................................... 8  
  Informatica Knowledge Base .............................................. 8  
  Informatica Documentation ............................................... 8  
  Informatica Product Availability Matrixes ............................ 9  
  Informatica Velocity .......................................................... 9  
  Informatica Marketplace ................................................... 9  
  Informatica Global Customer Support ................................. 9  

**Part I: Installation Overview** ............................................... 10  

**Chapter 1: Secure@Source** .............................................. 11  
Secure@Source Installation Overview .................................... 11  
Installation Process ............................................................... 11  
Secure@Source Services ....................................................... 12  
Secure@Source Repositories ............................................... 12  

**Part II: Before You Install Secure@Source** .......................... 14  

**Chapter 2: Plan the Domain** ............................................ 15  
Introduction to the Informatica Domain ................................. 15  
  Node and Domain .............................................................. 15  
  Nodes .................................................................................. 15  
  Service Manager ............................................................... 16  
  Application Services ......................................................... 16  
  License Key ....................................................................... 16  
  User Authentication ........................................................... 17  
  Encryption Key for Secure Data Storage ............................ 17  
  Domain Security ............................................................... 17  
  Informatica Administrator ................................................ 18  
Domain Planning Process ....................................................... 18  
Plan the Application Services ................................................. 19  
  Catalog Service ................................................................. 19  
  Content Management Service ............................................ 19  
  Data Integration Service ................................................... 20  
  Informatica Cluster Service ............................................... 21  
  Model Repository Service ................................................ 21  
Verify System Requirements ................................................ 21  
  System Requirements Considerations ................................. 22
Verify Services Installation Requirements .................................. 22
Verify Temporary Disk Space Requirements .................................. 23
Verify Port Requirements .......................................................... 23
Verify Database Requirements ..................................................... 25
Verify Application Service Hardware Requirements ..................... 25
Sizing Guidelines for Hadoop Hardware ......................................... 26
Record the Informatica Domain and Node Information ..................... 27
  Domain Object Naming Conventions .......................................... 27
  Domain .......................................................................... 28
  Nodes .......................................................................... 29
  Application Services ........................................................... 29
  Databases ................................................................. 30
  Secure Data Storage .......................................................... 32
  Domain Security .............................................................. 32
  Kerberos Authentication ......................................................... 34

Chapter 3: Prepare the Databases for the Informatica Domain .............. 35
Prepare the Databases for the Informatica Domain Overview ............... 35
Set Up Database User Accounts .................................................... 36
Domain Configuration Repository Database Requirements ................. 36
  IBM DB2 Database Requirements ........................................... 37
  Microsoft SQL Server Database Requirements ......................... 38
  Oracle Database Requirements ................................................ 38
  Sybase ASE Database Requirements ....................................... 39
Data Object Cache Database Requirements .................................. 40
  IBM DB2 Database Requirements ........................................... 40
  Microsoft SQL Server Database Requirements ......................... 40
  Oracle Database Requirements ................................................ 41
Model Repository Database Requirements ................................... 41
  IBM DB2 Database Requirements ........................................... 42
  Microsoft SQL Server Database Requirements ......................... 43
  Oracle Database Requirements ................................................ 43
Profiling Warehouse Database Requirements ................................ 44
  IBM DB2 Database Requirements ........................................... 44
  Microsoft SQL Server Database Requirements ......................... 44
  Oracle Database Requirements ................................................ 45
Operating System Parameters .................................................... 46
Reference Data Warehouse Database Requirements ......................... 46
  IBM DB2 Database Requirements ........................................... 46
  Microsoft SQL Server Database Requirements ......................... 47
  Oracle Database Requirements ................................................ 47
Secure@Source Repository Database Requirements ......................... 48
  IBM DB2 Database Requirements ........................................... 48
Chapter 4: Deployment Methods ........................................ 51
Deployment Methods Overview ........................................ 51
Internal Hadoop Cluster Deployment ................................ 51
  Prerequisites - Internal Cluster ..................................... 52
Informatica Cluster Service ............................................. 53
  Informatica Cluster Service Workflow ............................. 53
  Creating an Informatica Cluster Service ......................... 54
Preparing the Internal Hadoop Cluster Environment ............. 55
Internal Cluster Node Management ................................. 56
Troubleshooting the Internal Cluster Deployment ................ 56
External Hadoop Cluster Deployment ................................ 60
  Prerequisites for the External Hadoop Cluster .................. 60
Preparing the External Hadoop Cluster Environment .......... 61
Kerberos and SSL Setup for an External Cluster .................. 62
Troubleshooting the External Cluster Deployment ............... 63
Troubleshooting Common Cluster Deployment Issues .......... 64

Chapter 5: Prepare for Kerberos Authentication Setup .......... 70
Prepare for Kerberos Authentication Overview .................. 70
Set Up the Kerberos Configuration File ............................. 71
Generate the Service Principal and Keytab File Name Format .... 72
  Service Principal Requirements at Node Level .................. 72
  Service Principal Requirements at Process Level ............... 72
Running the Informatica Kerberos SPN Format Generator on Windows .......... 73
Running the Informatica SPN Format Generator on UNIX ........ 75
Review the SPN and Keytab Format Text File ...................... 76
Create the Service Principal Names and Keytab Files .......... 78
  Troubleshooting the SPN and Keytab Files ....................... 79

Part III: Secure@Source Installation ................................. 81

Chapter 6: Secure@Source Services Installation ................. 82
Secure@Source Services Installation Overview .................. 82
  Installing Secure@Source 4.1 ..................................... 82
  Upgrading to Secure@Source 4.1 .................................. 83
Pre-installation System Check ........................................ 83
  Pre-installation System Check Verification ...................... 83
Pre-installation System Check Results ............................ 84
Pre-installation System Check Log Files ........................... 84
Installing Secure@Source in Console Mode ....................... 84
Secure Directory for the Encryption Key and Configuration Files. 84
Installing by Creating a Domain. 85
Post-Installation Tasks. 107
Copy Elasticsearch Folder to New Nodes. 107
Set Catalog Service and Data Integration Service Properties. 108

Chapter 7: Troubleshooting the Services Installation. 111
Installation Troubleshooting Overview. 111
Installation Log Files. 111
Debug Log Files. 111
File Installation Log File. 112
Service Manager Log Files. 112
Troubleshooting Domains and Nodes. 113
Creating the Domain Configuration Repository. 113
Starting Secure@Source. 113
Pinging the Domain. 114
Adding a License. 114

Index. 115
Preface

The *Informatica Secure@Source Installation Guide* is written for the system administrator who is responsible for installing Secure@Source. This guide assumes you have knowledge of operating systems, relational database concepts, and the database engines, flat files, or mainframe systems in your environment. This guide also assumes you are familiar with the interface requirements for your supporting applications.

Informatica Resources

**Informatica Network**

Informatica Network hosts Informatica Global Customer Support, the Informatica Knowledge Base, and other product resources. To access Informatica Network, visit [https://network.informatica.com](https://network.informatica.com).

As a member, you can:

- Access all of your Informatica resources in one place.
- Search the Knowledge Base for product resources, including documentation, FAQs, and best practices.
- View product availability information.
- Review your support cases.
- Find your local Informatica User Group Network and collaborate with your peers.

**Informatica Knowledge Base**

Use the Informatica Knowledge Base to search Informatica Network for product resources such as documentation, how-to articles, best practices, and PAMs.

To access the Knowledge Base, visit [https://kb.informatica.com](https://kb.informatica.com). If you have questions, comments, or ideas about the Knowledge Base, contact the Informatica Knowledge Base team at [KB_Feedback@informatica.com](mailto:KB_Feedback@informatica.com).

**Informatica Documentation**


If you have questions, comments, or ideas about this documentation, contact the Informatica Documentation team through email at [infa_documentation@informatica.com](mailto:infa_documentation@informatica.com).
Informatica Product Availability Matrixes

Product Availability Matrixes (PAMs) indicate the versions of operating systems, databases, and other types of data sources and targets that a product release supports. If you are an Informatica Network member, you can access PAMs at https://network.informatica.com/community/informatica-network/product-availability-matrixes.

Informatica Velocity

Informatica Velocity is a collection of tips and best practices developed by Informatica Professional Services. Developed from the real-world experience of hundreds of data management projects, Informatica Velocity represents the collective knowledge of our consultants who have worked with organizations from around the world to plan, develop, deploy, and maintain successful data management solutions.

If you are an Informatica Network member, you can access Informatica Velocity resources at http://velocity.informatica.com.

If you have questions, comments, or ideas about Informatica Velocity, contact Informatica Professional Services at ips@informatica.com.

Informatica Marketplace

The Informatica Marketplace is a forum where you can find solutions that augment, extend, or enhance your Informatica implementations. By leveraging any of the hundreds of solutions from Informatica developers and partners, you can improve your productivity and speed up time to implementation on your projects. You can access Informatica Marketplace at https://marketplace.informatica.com.

Informatica Global Customer Support

You can contact a Global Support Center by telephone or through Online Support on Informatica Network.

To find your local Informatica Global Customer Support telephone number, visit the Informatica website at the following link: http://www.informatica.com/us/services-and-training/support-services/global-support-centers.

If you are an Informatica Network member, you can use Online Support at http://network.informatica.com.
Part I: Installation Overview

This part contains the following chapter:

- Secure@Source, 11
CHAPTER 1

Secure@Source

This chapter includes the following topics:

- Secure@Source Installation Overview, 11
- Installation Process, 11
- Secure@Source Services, 12
- Secure@Source Repositories, 12

Secure@Source Installation Overview

Informatica provides a single installer that installs Secure@Source and its services. Secure@Source is installed within the Informatica domain. Secure@Source uses the Catalog Service and other application services to bring together configured data assets in an enterprise and present a comprehensive view of the data assets and data asset relationships.

To use Secure@Source, you must install the Secure@Source services and create a domain. Use the Secure@Source installer to install the services. Secure@Source requires a dedicated domain to improve performance and isolate information about sensitive data. Secure@Source profiles the entire database and the domain requires specialized configuration for Secure@Source profiling.

The Secure@Source services consist of services to support the domain and application services to perform tasks and manage databases. The Informatica domain is the administrative unit for the Secure@Source environment. The domain is a collection of nodes that represent the machines on which the application services run. When you install the Secure@Source services on a machine, you install all the files for all services.

Installation Process

The installation of Secure@Source consists of multiple phases.

The installation process consists of the following phases:

1. Before you install the Secure@Source services, perform the following tasks to plan and prepare for the services installation:
   a. Plan the Informatica domain. Consider the system requirements and the type of user authentication that the domain uses.
b. Prepare the databases for the domain. Verify the database requirements, and then set up the databases required for Secure@Source.

c. Set up the machines to meet the Linux requirements to ensure that you can successfully install and run the Secure@Source services.

2. Install Secure@Source.

Use the installer to install the Secure@Source services on one or more Linux machines. The first time you run the installer, you must create the domain.

**Secure@Source Services**

Application services of Secure@Source represent server-based functionality. After you complete the installation, you can optionally create application services based on the license key generated for your organization.

When you create an application service, you designate a node to run the service process. The service process is the run-time representation of a service running on a node. The service type determines how many service processes can run at a time.

Some application services require databases to store information processed by the application service. When you plan the Informatica domain, you also need to plan the databases required by each application service.

Secure@Source uses the following application services:

- Catalog Service
- Content Management Service
- Data Integration Service
- Informatica Cluster Service
- Model Repository Service
- Secure@Source Service

**Secure@Source Repositories**

Secure@Source uses different types of repositories based on the type of data and metadata that it stores.

Secure@Source requires the following repositories:

**Domain configuration repository**

A relational database that stores domain configuration and user information.

**Model repository**

A relational database that stores metadata created by Secure@Source and application services to enable collaboration between the clients and services. Model repository also stores the resource configuration and data domain information.

**Profiling warehouse**

A relational database that stores profile results. Profile statistics form one part of the comprehensive metadata view that Secure@Source provides.
Reference data warehouse

A relational database that stores data values for the reference table objects that you define in the Model repository. When you add data to a reference table, the Content Management Service writes the data values to a table in the reference data warehouse.

Secure@Source repository database

Secure@Source stores data and metadata, such as data stores and scans, in the Secure@Source repository.
Part II: Before You Install Secure@Source

This part contains the following chapters:

- Plan the Domain, 15
- Prepare the Databases for the Informatica Domain, 35
- Deployment Methods, 51
- Prepare for Kerberos Authentication Setup, 70
CHAPTER 2

Plan the Domain

This chapter includes the following topics:

- Introduction to the Informatica Domain, 15
- Domain Planning Process, 18
- Plan the Application Services, 19
- Verify System Requirements, 21
- Record the Informatica Domain and Node Information, 27

Introduction to the Informatica Domain

An Informatica domain is a collection of nodes and services. A node is the logical representation of a machine in a domain. Services for the domain include the Service Manager that manages all domain operations and a set of application services that represent server-based functionality.

The domain requires a relational database to store configuration information and user account privileges and permissions. The first time that you install the Secure@Source services, you must create the domain configuration repository in a relational database.

Node and Domain

When you install the Secure@Source services on a machine, you create a node and a domain.

The Service Manager and all Informatica application services run on the node.

Nodes

The node in the domain runs the Service Manager that manages domain functions on that node. The Service Manager also supports the application services that run on the node.

The domain functions that a node performs and the services that a node runs depend on the following node configurations:

Node type

The node type determines whether the node can serve as a gateway or worker node and determines the domain functions that the node performs. When you install Secure@Source, you create a gateway node and the Informatica domain.

The gateway node serves as the master gateway node for the domain. The Service Manager on the master gateway node performs all domain operations on the master gateway node.
**Node role**

The node role defines the purpose of the node. A node with the service role can run application services. A node with the compute role can perform computations requested by remote application services. A node with both roles can run application services and locally perform computations for those services. By default, the gateway node has both the service and compute roles enabled.

For more information about nodes, see the *Informatica Administrator Guide*.

---

**Service Manager**

The Service Manager in the Informatica domain supports the domain and the application services.

The Service Manager supports the following areas:

**Domain**

The Service Manager performs all domain functions on the node to support the domain. Domain functions include authentication, authorization, and logging.

**Application services**

When a node has the service role, the Service Manager starts application services configured to run on that node. It starts and stops services and service processes based on requests from Informatica clients.

For more information about the Service Manager, see the *Informatica Administrator Guide*.

---

**Application Services**

Application services represent server-based functionality. Application services include services that can have multiple instances in the domain and system services that can have a single instance in the domain. System services are created for you when you create the domain. After you complete the installation, you create other application services based on the license key generated for your organization.

When you create an application service, you designate a node with the service role to run the service process. The service process is the run-time representation of a service running on a node. The service type determines how many service processes can run at a time.

Some application services require databases to store information processed by the application service. When you plan the Informatica domain, you also need to plan the databases required by each application service.

For more information about application services, see the *Informatica Application Service Guide*.

---

**License Key**

Informatica generates a license key based on the product and product options that your organization purchased. The license key controls the application services and the functionality that you can use.

When you install the Secure@Source services, you must enter the path and file name of the Informatica license key. The installer creates a license object in the domain based on the license key that you enter. When you create application services, you must assign the license object to each application service before you can run the service.
User Authentication

During installation, you can select the authentication to use for the Informatica domain.

The Informatica domain can use the following types of authentication to authenticate users in the Informatica domain:

- Native user authentication
- LDAP user authentication
- Kerberos network authentication

Native user accounts are stored in the Informatica domain and can only be used within the Informatica domain. Kerberos and LDAP user accounts are stored in an LDAP directory service and are shared by applications within the enterprise.

If you enable Kerberos authentication during installation, you must configure the Informatica domain to work with the Kerberos key distribution center (KDC). You must create the service principal names (SPN) required by the Informatica domain in the Kerberos principal database. The Kerberos principal database can be an LDAP directory service. You must also create keytab files for the SPNs and store it in the Informatica directory as required by the Informatica domain.

The installer configures the Informatica domain to use native authentication. After installation, you can set up a connection to an LDAP server and configure the Informatica domain to use LDAP authentication in addition to native authentication.

For more information about user authentication, see the Informatica Security Guide.

Encryption Key for Secure Data Storage

Informatica encrypts sensitive data, such as passwords and secure connection parameters, before it stores the data in the Secure@Source repositories. Informatica uses a keyword to create an encryption key with which to encrypt sensitive data.

When you install the Secure@Source services and create a domain, you must specify a keyword for the installer to use to generate the encryption key for the domain. Based on the keyword, the installer generates an encryption key file named siteKey and stores it in a directory you specify. If you do not specify a directory, the installer stores the siteKey file in the default directory: `<Secure@Source installation directory>/isp/config/keys`

You must specify a keyword even if you do not enable secure communication for the domain.

**Important:** You must keep the name of the domain, the keyword for the encryption key, and the encryption key file in a secure location. The encryption key is required when you change the encryption key of the domain or move a repository to another domain. If you do not have the encryption key, you must have the domain name and the keyword used to generate the encryption key.

Domain Security

When you install the Secure@Source services and create a domain, you can enable options to configure security in the domain.

You can configure the following security options for the domain:

**Secure communication for services and the Service Manager**

When you configure secure communication for the domain, you secure the connections between the Service Manager and the services in the domain. Informatica provides an SSL certificate that you can use to secure the domain. However, for better security for the domain, you can provide the SSL
certificate during the installation. Provide the keystore and truststore files that contain the SSL certificates you want to use.

Secure domain configuration repository database

When you install the Secure@Source services and create a domain, you can create the domain configuration repository in a database secured with the SSL protocol. Access to the secure database requires a truststore that contains the SSL certificates for the database. During installation, you provide the truststore file that contains the SSL certificate you want to use.

Secure connection for the Administrator tool

Informatica Administrator or the Administrator tool is the tool that you use to administer the Informatica domain. During installation, you can configure a secure HTTPS connection for the Administrator tool. You can provide the keystore file to use for the HTTPS connection.

Informatica Administrator

Informatica Administrator (the Administrator tool) is the administration tool that you use to administer the Informatica domain and security. The Administrator tool is a thin or web client application.

You use the Administrator tool to perform the following tasks:

Domain administrative tasks

- Manage logs, domain objects, and domain reports. Domain objects include application services, nodes, grids, folders, database connections, applications, and licenses.

Security administrative tasks

- Manage users, groups, roles, privileges, and permissions.

On the node where you install the Secure@Source services, the installer creates a UNIX daemon to run Informatica. When the installation completes successfully, the installer starts the Informatica daemon.

The Informatica service also runs the Administrator tool. Log in to the Administrator tool to create the user accounts for users of Informatica and to create and configure the application services in the domain.

Domain Planning Process

Before you install the Secure@Source services, you need to plan for all of the components in the Informatica domain.

When you plan the domain, you must consider the types of application services that the domain requires and the number of application services that run on each node. You must determine the database type and host name for the domain configuration repository and for the databases required by each application service.

You must choose a keyword for the installer to use to generate the encryption key for the domain. Informatica uses the encryption key to encrypt sensitive data.

If you decide to configure security for the domain, you must know the location and password for the keystore and truststore files. If you decide to use Kerberos authentication for the Informatica domain, you must work with the Kerberos administrator to set up the user and service principals required by the domain.

As part of the planning process, you must also verify that each machine and database server in the domain meets the minimum system requirements.
Plan the Application Services

When you plan the Informatica domain, you also need to plan the application services that will run in the domain. You create application services based on the license key generated for your organization.

When you plan the application services, you must account for the associated services that connect to the application service. You also must plan the relational databases that are required to create the application service.

For more information about application services, see the Informatica Application Service Guide.

Catalog Service

The Catalog Service is an application service that runs Enterprise Information Catalog in the Informatica domain. The Catalog Service manages the connections between service components and the users that have access to Enterprise Information Catalog search interface and Catalog Administrator.

The catalog represents an indexed inventory of all the configured data assets in an enterprise. You can find metadata and statistical information, such as profile statistics, data asset ratings, data domains, and data relationships, in the catalog.

**Note:** Ensure that you import the Hadoop cluster certificates to the domain trust store before you create a Catalog Service for a Hadoop cluster that uses SSL protocol.

Associated Services

The Catalog Service connects to other application services within the domain.

When you create the Catalog Service, you can associate it with the following application services:

**Model Repository Service**

The Catalog Service connects to the Model Repository Service to access resource configuration and data domain information from the Model repository. When you create the Catalog Service, you provide the name of the Model Repository Service.

**Data Integration Service**

The Catalog Service connects to the Data Integration Service to perform jobs, such as generating profile statistics for the resources. When you create the Catalog Service, you provide the name of the Data Integration Service.

**Informatica Cluster Service**

If you installed Secure@Source on the internal cluster, the Catalog Service connects to Informatica Cluster Service to administer and manage all the internal Hadoop cluster services, Apache Ambari server, and Apache Ambari agents. When you create the Catalog Service for an internal cluster deployment, you need to provide the name of the Informatica Cluster Service.

**Content Management Service**

The Catalog Service uses the Content Management Service to fetch reference data for data domains that use reference tables. When you create the Catalog Service, you can optionally provide the name of the Content Management Service.

Content Management Service

The Content Management Service is an application service that manages reference data. A reference data object contains a set of data values that Enterprise Information Catalog searches while performing data...
domain discovery on source data. Data domain discovery finds inferred business semantics based on column data. Examples include Social Security number, phone number, and credit card number.

The Content Management Service uses the Data Integration Service to run mappings to transfer data between reference tables and external data sources.

**Associated Services**

The Content Management Service connects to other application services within the domain.

When you create the Content Management Service, you can associate it with the following application services:

**Data Integration Service**

The Content Management Service uses the Data Integration Service to transfer data between reference tables and external data sources. When you create the Content Management Service, you provide the name of the Data Integration Service. You must create the Data Integration Service and Content Management Service on the same node.

**Model Repository Service**

The Content Management Service connects to the Model Repository Service to store metadata for reference data objects in the Model repository. When you create the Content Management Service, you provide the name of the Model Repository Service.

**Required Databases**

The Content Management Service requires a reference data warehouse in a relational database. When you create the Content Management Service, you must provide connection information to the reference data warehouse.

Create the following database before you create the Content Management Service:

**Reference data warehouse**

Stores data values for the reference table objects that you define in the Model repository. When you add data to a reference table, the Content Management Service writes the data values to a table in the reference data warehouse. You need a reference data warehouse to manage reference table data in Secure@Source.

**Data Integration Service**

The Data Integration Service is an application service that runs profiles in the Informatica domain. The Data Integration Service generates profile results for resources that you have set up to fetch profile metadata and then writes the profile results to the profiling warehouse.

**Associated Services**

The Data Integration Service connects to other application services within the domain.

When you create the Data Integration Service, you can associate it with the following application service:

**Model Repository Service**

The Data Integration Service connects to the Model Repository Service to perform jobs such as running profiles. When you create the Data Integration Service, you provide the name of the Model Repository Service.
**Required Databases**

The Data Integration Service connects to other application services within the domain.

When you create the Data Integration Service, you can associate it with the following application service:

- **Model Repository Service**
  
  The Data Integration Service connects to the Model Repository Service to perform jobs such as running profiles. When you create the Data Integration Service, you provide the name of the Model Repository Service.

**Informatica Cluster Service**

The Informatica Cluster Service is an application service that runs and manages all the Hadoop services, Apache Ambari server, and Apache Ambari agents on an internal Hadoop cluster. If you choose the internal cluster deployment mode while you install Secure@Source, you need to create the Informatica Cluster Service before you create the Catalog Service. You can then specify the Informatica Cluster Service value when you create the Catalog Service.

**Model Repository Service**

The Model Repository Service is an application service that manages the Model repository. The Model repository stores metadata created by Informatica clients and application services in a relational database to enable collaboration among the clients and services.

When you access a Model repository object from Secure@Source user interface or the Data Integration Service, the client or service sends a request to the Model Repository Service. The Model Repository Service process fetches, inserts, and updates the metadata in the Model repository database tables.

**Note:** When you create the Model Repository Service, you do not associate it with other application services.

**Required Databases**

The Model Repository Service requires a Model repository in a relational database. When you create the Model Repository Service, you must provide connection information to the database.

Create the following database before you create the Model Repository Service:

- **Model repository**
  
  Stores metadata created by Informatica clients and application services in a relational database to enable collaboration among the clients and services. You need a Model repository to store the design-time and run-time objects created by Informatica clients and application services.

**Verify System Requirements**

System requirements depend on a variety of factors specific to your enterprise and goals.

The system requirements in this section are the minimum requirements necessary to install and use Secure@Source. The minimum requirements are sufficient for the following scenario:

- No user activity data.
- Data stores have a low volume of data.
Data stores will not be scanned concurrently.

For more information about product requirements and supported platforms, see the Product Availability Matrix on Informatica Network: https://network.informatica.com/community/informatica-network/product-availability-matrices

System Requirements Considerations

Consider the following factors when determining system requirements:

- The volume of data in your enterprise that you need to scan.
- The volume of user activity on the data stores.
- The time frame within which you need to complete all scans.
- Whether you want to run scans concurrently or sequentially.

Verify Services Installation Requirements

Verify that your machine meets the minimum system requirements to install the Secure@Source services.

For a production implementation, allocate at least two physical servers. Use one server for the domain services and the other for the Hadoop cluster.

Each physical server must meet the following requirements:

- 64 GB RAM
- 32 logical cores
- 500 GB disk space. All 500 GB must be allocated to one logical volume.
  To efficiently handle demanding input and output operations on a production system, Informatica recommends that you use multiple physical disks for Hadoop node. Each physical disk must be at least 15,000 RPM.

If you use multiple partitions or LVMs, Informatica recommends that each directory has the following disk space available:

<table>
<thead>
<tr>
<th>Directory</th>
<th>Recommended Disk Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>/</td>
<td>60 GB</td>
</tr>
<tr>
<td>/hadoop</td>
<td>400 GB</td>
</tr>
<tr>
<td>Note: The hadoop folder structure is a Hadoop requirement.</td>
<td></td>
</tr>
<tr>
<td>/opt</td>
<td>50 GB</td>
</tr>
<tr>
<td>/usr</td>
<td>50 GB</td>
</tr>
<tr>
<td>/var</td>
<td>40 GB</td>
</tr>
<tr>
<td>/tmp</td>
<td>8 GB</td>
</tr>
</tbody>
</table>

Verify the following operating system requirements:

- You must install Secure@Source on Red Hat Enterprise Linux 6.5. The kernel must be 2.6.32-431.el6.x86.64 or higher.
• The user that performs the installation must have passwordless sudo access for the Hadoop nodes.

• SELinux must be set to permissive or disabled on the Secure@Source node. It must be disabled on the Hadoop nodes.

• OpenSSL must be version 1.0.1e-30 or higher. The kernel-devel and kernel-headers packages must be installed.
  If multiple versions of OpenSSL are installed, remove all versions other than the one required by the Secure@Source installation from the $PATH environment variable.

• Create the Hadoop gateway user.

• Set up passwordless SSH for the Informatica user where Secure@Source is installed to the Hadoop gateway user on the Hadoop gateway node. For example, you can use the following commands:
  
  ```
  ssh-keygen
  ssh-copy-id <Hadoop gateway user>@<Hadoop gateway node hostname>
  ```

  Repeat the commands for all Hadoop node gateway users.

• Increase the soft and hard limits for the user that installs Secure@Source and the root user on all Hadoop nodes, including the gateway node. Increase nproc and nofile to 100000.

• Verify that a public Yum repository is configured. Ambari uses Yum to install the required RPMs.

• Configure the /etc/hosts file on each machine so that you have fully qualified domain names.
  Informatica recommends the following host name format in lowercase:
  
  ```
  <machine ipaddress> <fully qualified name> <alias>
  ```

  **Note:** To verify the configured host name, run the `hostname -f` command.

---

### Verify Temporary Disk Space Requirements

The installer writes temporary files to the hard disk. Verify that you have enough available disk space on the machine to support the installation. When the installation completes, the installer deletes the temporary files and releases the disk space.

The installer requires 8 GB of temporary disk space.

### Verify Port Requirements

The installer sets up the ports for components in the Informatica domain, and it designates a range of dynamic ports to use for some application services.

You can specify the port numbers to use for the components and a range of dynamic port numbers to use for the application services. Or you can use the default port numbers provided by the installer. Verify that the port numbers are available on the machines where you install the Informatica services.

The following table describes the ports used by Informatica:

<table>
<thead>
<tr>
<th>Port Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node port</td>
<td>Port number for the node created during installation. Default is 6005.</td>
</tr>
<tr>
<td>Service Manager port</td>
<td>Port number used by the Service Manager on the node. The Service Manager listens for incoming connection requests on this port. Client applications use this port to communicate with the services in the domain. This is the port that the Informatica command line programs use to communicate to the domain. This is also the port for the SQL data service JDBC/ODBC driver. Default is 6006.</td>
</tr>
<tr>
<td>Port Type</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Service Manager shutdown port</td>
<td>Port number that controls server shutdown for the domain Service Manager. The Service Manager listens for shutdown commands on this port. Default is 6007.</td>
</tr>
<tr>
<td>Informatica Administrator port</td>
<td>Port number used by Informatica Administrator. Default is 6008.</td>
</tr>
<tr>
<td>Informatica Administrator shutdown port</td>
<td>Port number that controls server shutdown for Informatica Administrator. Informatica Administrator listens for shutdown commands on this port. Default is 6009.</td>
</tr>
</tbody>
</table>
| Range of dynamic ports for application services| Range of port numbers that can be dynamically assigned to application service processes as they start up. When you start an application service that uses a dynamic port, the Service Manager dynamically assigns the first available port in this range to the service process. At a minimum, the number of ports in the range must be at least twice the number of application service processes that will run on the node. Default is 6014 to 6114. The Service Manager dynamically assigns port numbers from this range to the following application service:  
  - Model Repository Service  
  The Service Manager also assigns port numbers from this range to the system services. |
| Static ports for application services         | Static ports are ports that are assigned dedicated port numbers that do not change. When you create the application service, you can accept the default port number, or you can manually assign the port number.  
The following services use static port numbers:  
  - Content Management Service. Default is 8105 for HTTP.  
  - Data Integration Service. Default is 8095 for HTTP.  
  - Secure@Source Service. Default is 6200 for HTTP. |

**Note:** Services and nodes can fail to start if there is a port conflict. You can update the range of ports for application services after you upgrade.

### Guidelines for Port Configuration

The installer validates the port numbers that you specify to ensure that there will be no port conflicts in the domain.

Use the following guidelines to determine the port numbers:

- The port number you specify for the domain and for each component in the domain must be unique.
- The port number for the domain and domain components cannot be within the range of the port numbers that you specify for the application service processes.
- The highest number in the range of port numbers that you specify for the application service processes must be at least three numbers higher than the lowest port number. For example, if the minimum port number in the range is 6400, the maximum port number must be at least 6403.
- The port numbers that you specify cannot be lower than 1025 or higher than 65535.
Verify Database Requirements

Verify that the database server has adequate disk space for the domain configuration repository and for the other databases required by the application services.

The following table describes the database requirements for the domain configuration repository and for the other databases required by the application services:

<table>
<thead>
<tr>
<th>Database</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informatica domain configuration repository</td>
<td>The domain configuration repository supports the following database types:</td>
</tr>
<tr>
<td></td>
<td>- IBM DB2 UDB</td>
</tr>
<tr>
<td></td>
<td>- Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>- Oracle</td>
</tr>
<tr>
<td></td>
<td>- Sybase ASE</td>
</tr>
<tr>
<td></td>
<td>Allow 200 MB of disk space for the database.</td>
</tr>
<tr>
<td>Model repository</td>
<td>The Model repository supports the following database types:</td>
</tr>
<tr>
<td></td>
<td>- IBM DB2 UDB</td>
</tr>
<tr>
<td></td>
<td>- Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>- Oracle</td>
</tr>
<tr>
<td></td>
<td>Allow 3 GB of disk space for DB2. Allow 200 MB of disk space for all other</td>
</tr>
<tr>
<td></td>
<td>database types.</td>
</tr>
<tr>
<td></td>
<td>Allocate more space based on the amount of metadata you want to store.</td>
</tr>
<tr>
<td>Profiling warehouse</td>
<td>The profiling warehouse supports the following database types:</td>
</tr>
<tr>
<td></td>
<td>- IBM DB2 UDB</td>
</tr>
<tr>
<td></td>
<td>- Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>- Oracle</td>
</tr>
<tr>
<td></td>
<td>Allow 10 GB of disk space for the database.</td>
</tr>
<tr>
<td>Reference data warehouse</td>
<td>The reference data warehouse supports the following database types:</td>
</tr>
<tr>
<td></td>
<td>- IBM DB2 UDB</td>
</tr>
<tr>
<td></td>
<td>- Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>- Oracle</td>
</tr>
<tr>
<td></td>
<td>Allow 200 MB of disk space for the database.</td>
</tr>
<tr>
<td>Secure@Source repository</td>
<td>The Secure@Source repository supports the following database types:</td>
</tr>
<tr>
<td></td>
<td>- IBM DB2</td>
</tr>
<tr>
<td></td>
<td>- Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>- Oracle</td>
</tr>
<tr>
<td></td>
<td>Allow 3 GB of disk space for DB2. Allow 200 MB of disk space for all other</td>
</tr>
<tr>
<td></td>
<td>database types.</td>
</tr>
</tbody>
</table>

Verify Application Service Hardware Requirements

Verify that the nodes in the domain have adequate hardware for the Service Manager and the application services that run on the nodes.

You can create an Informatica domain with one node and run all application services on the same node. If you create an Informatica domain with multiple nodes, you can run the application services on separate nodes. When you plan the application services for the domain, consider system requirements based on the services that you run on a node.

Based on workload and concurrency requirements, you might need to optimize performance by adding cores and memory on a node.
The following table lists the minimum system requirements for a node based on some common configuration scenarios. Use this information as a guideline for other configurations in your domain.

<table>
<thead>
<tr>
<th>Services</th>
<th>Processor</th>
<th>Memory</th>
<th>Disk Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>One node runs the following services:</td>
<td>32 logical cores</td>
<td>64 GB</td>
<td>60 GB</td>
</tr>
<tr>
<td>- Data Integration Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Model Repository Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Catalog Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Content Management Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Informatica Cluster Service</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Note:</strong> Applies to internal Hadoop deployment on Hortonworks.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sizing Guidelines for Hadoop Hardware**

Based on the size of the data set, you must add additional memory and CPU cores to tune the performance of Secure@Source. You must also note the minimum number of nodes that are required to deploy supported data set sizes.

**Note:** Each node in the recommendation in the following sections requires 32 logical cores and 64 GB of available memory.

**Sizing Guidelines for a Small Data Set**

For a small data set, you require a single node. The following table lists the size of a small data set and recommendations for the number of CPU logical cores and memory settings:

<table>
<thead>
<tr>
<th>Number of Objects</th>
<th>Memory</th>
<th>Number of CPU Logical Cores</th>
<th>Minimum Number of Hadoop Nodes Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 million.</td>
<td>64 GB</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>For example, 10,000 tables x 100 columns.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Sizing Guidelines for a Medium Data Set**

For a medium data set, you require a minimum of three nodes. The following table lists the size of a medium data set and the recommended number of CPU logical cores and memory settings:

<table>
<thead>
<tr>
<th>Number of Objects</th>
<th>Memory</th>
<th>Number of CPU Logical Cores</th>
<th>Minimum Number of Hadoop Nodes Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>20 million.</td>
<td>192 GB</td>
<td>96</td>
<td>3</td>
</tr>
<tr>
<td>For example, 200,000 tables x 100 columns.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sizing Guidelines for a Large Data Set

For a large data set, you require a minimum of six nodes. The following table lists the size of a large data set and the recommended number of CPU logical cores and memory settings:

<table>
<thead>
<tr>
<th>Number of Objects</th>
<th>Memory</th>
<th>Number of CPU Logical Cores</th>
<th>Minimum Number of Hadoop Nodes Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>50 million.</td>
<td>384 GB</td>
<td>192</td>
<td>6</td>
</tr>
<tr>
<td>For example, 500,000 tables x 100 columns.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Record the Informatica Domain and Node Information

When you install the Secure@Source services, you need to know information about the domain, nodes, application services, and databases that you plan to create. If you plan to install the Informatica services on a network that uses Kerberos authentication, you also need to know information about the Kerberos authentication server.

Use the tables in this section to record the information that you need.

Domain Object Naming Conventions

Choose a naming convention to use for the domain, nodes, and application services when you plan the domain.

You cannot change domain, node, and application service names. Use names that continue to work if you migrate a node to another machine or if you add additional nodes and services to the domain. In addition, use names that convey how the domain object is used.

For more information about domain object naming conventions, see the following Informatica Velocity Best Practice article available on Informatica Network: Informatica Platform Naming Conventions.

The following table lists recommended naming conventions for domain objects:

<table>
<thead>
<tr>
<th>Object</th>
<th>Naming Convention</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>DMN, DOM, DOMAIN, <em>&lt;ORG&gt;</em>&lt;ENV&gt;</td>
<td>DOM_FIN_DEV (Finance Development)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DOMAIN_ICC_PD (Integration Competency Center Production)</td>
</tr>
<tr>
<td>Node</td>
<td>Node&lt;node##&gt;<em>&lt;ORG&gt;</em>&lt;optional distinguisher&gt;_&lt;ENV&gt;</td>
<td>Node01_ICC_DEV</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Node07_FIN_REVENUE_DV</td>
</tr>
<tr>
<td>Content Management Service</td>
<td>CMS_&lt;ORG&gt;_&lt;ENV&gt;</td>
<td>CMS_FIN_DEV</td>
</tr>
<tr>
<td>Data Integration Service</td>
<td>DIS_&lt;ORG&gt;_&lt;ENV&gt;</td>
<td>DIS_ICC_DEV</td>
</tr>
<tr>
<td>Model Repository Service</td>
<td>MRS_&lt;ORG&gt;_&lt;ENV&gt;</td>
<td>MRS_FIN_DEV</td>
</tr>
</tbody>
</table>
## Domain

The first time that you install the Informatica services, you create the master gateway node and the Informatica domain.

Use the following table to record the domain information that you need:

<table>
<thead>
<tr>
<th>Domain Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain name</td>
<td>Name of the domain that you plan to create. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; &quot; ? , &lt; &gt; \ /</td>
<td></td>
</tr>
<tr>
<td>Master gateway node host name</td>
<td>Fully qualified host name of the machine on which to create the master gateway node. If the machine has a single network name, use the default host name. The node host name cannot contain the underscore (_) character. If the machine has multiple network names, you can modify the default host name to use an alternate network name. If the machine has a single network name, use the default host name. <strong>Note:</strong> Do not use localhost. The host name must explicitly identify the machine.</td>
<td></td>
</tr>
<tr>
<td>Master gateway node name</td>
<td>Name of the master gateway node that you plan to create on this machine. The node name is not the host name for the machine.</td>
<td></td>
</tr>
</tbody>
</table>
Nodes

When you install the Informatica services, you add the installation machine to the domain as a node.

Use the following table to record the node information that you need:

<table>
<thead>
<tr>
<th>Node Information</th>
<th>Description</th>
<th>Value for Node</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node host name</td>
<td>Fully qualified host name of the machine on which to create the node. If the machine has a single network name, use the default host name. The node host name cannot contain the underscore (_) character. If the machine has multiple network names, you can modify the default host name to use an alternate network name. If the machine has a single network name, use the default host name. <strong>Note:</strong> Do not use localhost. The host name must explicitly identify the machine.</td>
<td></td>
</tr>
<tr>
<td>Node name</td>
<td>Name of the node that you plan to create on this machine. The node name is not the host name for the machine.</td>
<td></td>
</tr>
</tbody>
</table>

Application Services

The application services that you create depend on the license key generated for your organization.

**Important:** If you plan to use Kerberos authentication, you must know the application service and node name before you create the keytab files.

Use the following table to record the application services that you need in the domain and to record the nodes that will run the application services:

<table>
<thead>
<tr>
<th>Application Service</th>
<th>Service Name</th>
<th>Node Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalog Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Integration Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informatica Cluster Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model Repository Service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Secure@Source Service</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Databases

When you plan the Informatica domain, you also need to plan the required relational databases. The domain requires a database to store configuration information and user account privileges and permissions. Some application services require databases to store information processed by the application service.

Domain

Use the following table to record the database information that you need for the domain:

<table>
<thead>
<tr>
<th>Database Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain configuration database type</td>
<td>Database type for the domain configuration repository. The domain configuration repository supports IBM DB2 UDB, Microsoft SQL Server, Oracle, or Sybase ASE.</td>
<td></td>
</tr>
<tr>
<td>Domain configuration database host name</td>
<td>The name of the machine hosting the database.</td>
<td></td>
</tr>
</tbody>
</table>

Content Management Service

Use the following table to record the database information that you need for the Content Management Service:

<table>
<thead>
<tr>
<th>Database Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reference data warehouse database type</td>
<td>Database type for the reference data warehouse. The reference data warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.</td>
<td></td>
</tr>
<tr>
<td>Reference data warehouse database host name</td>
<td>The name of the machine hosting the database.</td>
<td></td>
</tr>
</tbody>
</table>

Data Integration Service

Use the following table to record the database information that you need for the Data Integration Service:

<table>
<thead>
<tr>
<th>Database Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data object cache database type</td>
<td>Database type for the data object cache database. The data object cache database supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.</td>
<td></td>
</tr>
<tr>
<td>Data object cache database host name</td>
<td>The name of the machine hosting the database.</td>
<td></td>
</tr>
<tr>
<td>Profiling warehouse database type</td>
<td>Database type for the profiling warehouse. The profiling warehouse supports IBM DB2 UDB, Microsoft SQL Server, or Oracle.</td>
<td></td>
</tr>
<tr>
<td>Profiling warehouse database host name</td>
<td>The name of the machine hosting the database.</td>
<td></td>
</tr>
</tbody>
</table>
Database Information | Description | Value |
--- | --- | --- |
Workflow database type | Database type for the workflow database. The workflow database supports IBM DB2 UDB, Microsoft SQL Server, or Oracle. |  |
Workflow database host name | The name of the machine hosting the database. |  |

**Model Repository Service**

Use the following table to record the database information that you need for the Model Repository Service:

Database Information | Description | Value |
--- | --- | --- |
Model repository database type | Database type for the Model repository. The Model repository supports IBM DB2 UDB, Microsoft SQL Server, or Oracle. |  |
Model repository database host name | The name of the machine hosting the database. |  |

**Secure@Source Service**

Use the following table to record the database information that you need for the Secure@Source Service:

Database Information | Description | Value |
--- | --- | --- |
Secure@Source repository database type | Database type for the Secure@Source repository. The Secure@Source repository supports IBM DB2, Microsoft SQL Server, or Oracle. |  |
Secure@Source repository database host name | The name of the machine hosting the database. |  |
Secure Data Storage

When you install the Informatica services, you must provide a keyword for the installer to use to generate the encryption key for the domain.

Use the following table to record the information that you need to configure secure data storage:

<table>
<thead>
<tr>
<th>Encryption Key Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
</table>
| Keyword                    | Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria:  
  - From 8 to 20 characters long  
  - Includes at least one uppercase letter  
  - Includes at least one lowercase letter  
  - Includes at least one number  
  - Does not contain spaces  
  The encryption key is created based on the keyword that you provide when you create the Informatica domain. | |
| Encryption key directory   | Directory in which to store the encryption key for the domain. The default location is the following directory: `<Informatica installation directory>/isp/config/keys`. | |

Domain Security

When you install the Secure@Source services, you can enable options in the Informatica domain to configure security for the domain.

Secure Communication for Services and the Service Manager

You can optionally configure secure communication between services and the Service Manager.

Important: If you choose to use your SSL certificates instead of the default certificates, you must provide information about the SSL certificates during the installation. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain.

Use the following table to record information about the keystore and truststore files that contain the SSL certificates you want to use:

<table>
<thead>
<tr>
<th>Security Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystore file directory</td>
<td>Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.</td>
<td></td>
</tr>
<tr>
<td>Keystore password</td>
<td>Password for the keystore infa_keystore.jks.</td>
<td></td>
</tr>
<tr>
<td>Truststore file directory</td>
<td>Directory that contains the truststore files. The directory must contain files named infa_truststore.jks and infa_truststore.pem.</td>
<td></td>
</tr>
<tr>
<td>Truststore password</td>
<td>Password for the infa_truststore.jks file.</td>
<td></td>
</tr>
</tbody>
</table>
Secure Domain Configuration Repository Database

You can optionally create the domain configuration repository in a database that is secured with the SSL protocol.

**Important**: Access to the secure database requires a truststore that contains the certificates for the database.

Use the following table to record the information about the truststore file for the secure database:

<table>
<thead>
<tr>
<th>Security Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database truststore file</td>
<td>Path and file name of the truststore file for the secure database.</td>
<td></td>
</tr>
<tr>
<td>Database truststore</td>
<td>Password for the truststore file.</td>
<td></td>
</tr>
<tr>
<td>password</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Secure Connection for the Administrator Tool

You can configure a secure HTTPS connection for the Administrator tool.

**Important**: If you choose to use a keystore file that you create instead of the default file, you must provide information about the file during installation.

Use the following table to record information about the keystore file that you want to use:

<table>
<thead>
<tr>
<th>Security Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystore password</td>
<td>A plain-text password for the keystore file.</td>
<td></td>
</tr>
<tr>
<td>Keystore file directory</td>
<td>Location of the keystore file.</td>
<td></td>
</tr>
</tbody>
</table>
Kerberos Authentication

To configure the Informatica domain to run on a network that uses Kerberos authentication, you need information about the Kerberos authentication server.

Use the following table to verify and record information about the Kerberos authentication server:

<table>
<thead>
<tr>
<th>Domain Information</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Service realm name</td>
<td>Name of the Kerberos realm to which the Informatica domain services belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.</td>
<td></td>
</tr>
<tr>
<td>User realm name</td>
<td>Name of the Kerberos realm to which the Informatica domain users belong. The realm name must be in uppercase. The service realm name and the user realm name must be the same.</td>
<td></td>
</tr>
<tr>
<td>Location of the Kerberos configuration file</td>
<td>Directory where the Kerberos configuration file named krb5.conf is stored. Informatica requires specific properties to be set in the configuration file. If you do not have permission to copy or update the Kerberos configuration file, you might have to request the Kerberos administrator to update the file.</td>
<td></td>
</tr>
</tbody>
</table>
Prepare the Databases for the Informatica Domain

This chapter includes the following topics:

- Prepare the Databases for the Informatica Domain Overview, 35
- Set Up Database User Accounts, 36
- Domain Configuration Repository Database Requirements, 36
- Data Object Cache Database Requirements, 40
- Model Repository Database Requirements, 41
- Profiling Warehouse Database Requirements, 44
- Reference Data Warehouse Database Requirements, 46
- Secure@Source Repository Database Requirements, 48

Prepare the Databases for the Informatica Domain Overview

Informatica stores data and metadata in repositories in the domain. Before you create the domain and the application services, set up the databases and database user accounts for the repositories.

Set up a database and user account for the following repositories:

- Domain configuration repository
- Model repository
- Profiling warehouse
- Reference data warehouse
- Secure@Source repository

To prepare the databases, verify the database requirements and set up the database. The database requirements depend on the application services that you create in the domain and the number of data integration objects that you build and store in the repositories.
Set Up Database User Accounts

Set up a database and user account for the domain configuration repository and for the repository databases associated with the applications services.

Use the following rules and guidelines when you set up the user accounts:

• The database user account must have permissions to create and drop tables, indexes, and views, and to select, insert, update, and delete data from tables. The users must have sufficient tablespace sizes.
• Use 7-bit ASCII to create the password for the account.
• To prevent database errors in one repository from affecting any other repository, create each repository in a separate database schema with a different database user account. Do not create a repository in the same database schema as the domain configuration repository or any other repository in the domain.
• If you create more than one domain, each domain configuration repository must have a separate user account.

You must create the following database user accounts:

• Secure@Source repository
• Model repository
• Domain configuration repository
• Content Management
• Profiling warehouse

Domain Configuration Repository Database Requirements

Informatica components store metadata in relational database repositories. The domain stores configuration and user information in a domain configuration repository.

You must set up a database and user account for the domain configuration repository before you run the installation. The database must be accessible to all gateway nodes in the Informatica domain.

When you install Secure@Source, you provide the database and user account information for the domain configuration repository. The Secure@Source installer uses JDBC to communicate with the domain configuration repository.

The domain configuration repository supports the following database types:

• IBM DB2 UDB
• Microsoft SQL Server
• Oracle
• Sybase ASE

Allow 200 MB of disk space for the disk space.
IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- If the repository is in an IBM DB2 9.7 database, verify that IBM DB2 Version 9.7 Fix Pack 7 or a later fix pack is installed.
- On the IBM DB2 instance where you create the database, set the following parameters to ON:
  - DB2_SKIPINSERTED
  - DB2_EVALUNCOMMITTED
  - DB2_SKIPDELETED
  - AUTO_RUNSTATS
- On the database, set the configuration parameters.
  The following table lists the configuration parameters that you must set:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
<tr>
<td>applheapsz</td>
<td>8192</td>
</tr>
<tr>
<td>appl_ctl_heap_sz</td>
<td>8192</td>
</tr>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
</tbody>
</table>

- Set the tablespace pageSize parameter to 32768 bytes.
  In a single-partition database, specify a tablespace that meets the pageSize requirements. If you do not specify a tablespace, the default tablespace must meet the pageSize requirements.
  In a multi-partition database, specify a tablespace that meets the pageSize requirements. Define the tablespace in the catalog partition of the database.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Verify that the database user has CREATETAB, CONNECT, and BINDADD privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
• In the DataDirect Connect for JDBC utility, update the DynamicSections parameter to 3000.

The default value for DynamicSections is too low for the Informatica repositories. Informatica requires a larger DB2 package than the default. When you set up the DB2 database for the domain configuration repository or a Model repository, you must set the DynamicSections parameter to at least 3000. If the DynamicSections parameter is set to a lower number, you can encounter problems when you install or run Informatica services.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository on Microsoft SQL Server:

• Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention.

To set the isolation level for the database, run the following command:

ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON

To verify that the isolation level for the database is correct, run the following command:

SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName

• Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON.

To set the parameter to ON, run the following command:

ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON

To verify that the parameter is set correctly, run the following command:

SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName

• Set the AUTO_UPDATE_STATISTICSASYNC parameter to ON.

To set the parameter to ON, run the following command:

ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICSASYNC ON

To verify that the parameter is set correctly, run the following command:

SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName

• Set the AUTO_UPDATE_STATISTICS parameter to ON.

To set the parameter to ON, run the following command:

ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON

To verify that the parameter is set correctly, run the following command:

SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName

• The database user account must have the CONNECT, CREATE TABLE, and CREATE VIEW privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

• Set the open_cursors parameter to 4000 or higher.

• Set the permissions on the view $parameter for the database user.

• Set the privileges for the database user to run show parameter open_cursors in the Oracle database.

When you run the pre-installation (i10Pi) system check tool, i10Pi runs the command against the database to identify the OPEN_CURSORS parameter with the domain database user credentials.

You can run the following query to determine the open cursors setting for the domain database user account:

SELECT VALUE OPEN_CURSORS FROM V$PARAMETER WHERE UPPER(NAME)=UPPER('OPEN_CURSORS')

• Verify that the database user has the following privileges:
CREATE SEQUENCE
CREATE SESSION
CREATE SYNONYM
CREATE TABLE
CREATE VIEW

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

**Sybase ASE Database Requirements**

Use the following guidelines when you set up the repository on Sybase ASE:

- Set the database server page size to 16K or higher. You must set the page size to 16K as this is a one-time configuration and cannot be changed afterwards.

- Set the database locking configuration to use row-level locking.
  The following table describes the database locking configuration that you must set:

<table>
<thead>
<tr>
<th>Database Configuration</th>
<th>Sybase System Procedure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lock scheme</td>
<td>sp_configure &quot;lock scheme&quot;</td>
<td>0, datarows</td>
</tr>
</tbody>
</table>

- Set the Sybase database option "ddl in tran" to TRUE.
- Set "allow nulls by default" to TRUE.
- Turn ON the Sybase database option select into/bulkcopy/pllsort.
- Enable the "select" privilege for the sysobjects system table.
- Create the following login script to disable the default VARCHAR truncation:
  
  ```
  create procedure dbo.sp_string_rtrunc_proc as set string_rtruncation on
  sp_modifylogin "user_name", "login script", sp_string_rtrunc_proc
  ```

  The login script is executed every time the user logs into the Sybase instance. The stored procedure sets the parameter at the session level. The sp_modifylogin system procedure updates "user_name" with the stored procedure as its "login script". The user must have permission to invoke the stored procedure.

- Verify that the database user has CREATE DEFAULT, CREATE PROCEDURE, CREATE RULE, CREATE TABLE, and CREATE VIEW privileges.

- Set the database configurations to the recommended baseline values.
  The following table lists the database memory configuration parameters that you must set:

<table>
<thead>
<tr>
<th>Database Configuration</th>
<th>Sybase System Procedure</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum amount of total physical memory</td>
<td>sp_configure &quot;max memory&quot;</td>
<td>2097151</td>
</tr>
<tr>
<td>Procedure cache size</td>
<td>sp_configure &quot;procedure cache size&quot;</td>
<td>500000</td>
</tr>
<tr>
<td>Number of open objects</td>
<td>sp_configure &quot;number of open objects&quot;</td>
<td>5000</td>
</tr>
<tr>
<td>Number of open indexes</td>
<td>sp_configure &quot;number of open indexes&quot;</td>
<td>5000</td>
</tr>
<tr>
<td>Database Configuration</td>
<td>Sybase System Procedure</td>
<td>Value</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>---------------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>Number of open partitions</td>
<td>sp_configure &quot;number of open partitions&quot;</td>
<td>5000</td>
</tr>
<tr>
<td>Heap memory per user</td>
<td>sp_configure &quot;heap memory per user&quot;</td>
<td>49152</td>
</tr>
<tr>
<td>Number of locks</td>
<td>sp_configure &quot;number of locks&quot;</td>
<td>100000</td>
</tr>
</tbody>
</table>

# Data Object Cache Database Requirements

The data object cache database stores cached logical data objects and virtual tables for the Data Integration Service. You specify the data object cache database connection when you create the Data Integration Service.

The data object cache database supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Oracle

Allow 200 MB of disk space for the database.

**Note:** Ensure that you install the database client on the machine on which you want to run the Data Integration Service.

## IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Verify that the database user account has CREATETAB and CONNECT privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.

## Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository on Microsoft SQL Server:

- Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention. To set the isolation level for the database, run the following command:
  
  ```sql
  ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
  ```

  To verify that the isolation level for the database is correct, run the following command:
  
  ```sql
  SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
  ```

- Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON. To set the parameter to ON, run the following command:
  
  ```sql
  ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
  ```
To verify that the parameter is set correctly, run the following command:

```
SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName
```

- Set the AUTO_UPDATE_STATISTICS_ASYNC parameter to ON.
  To set the parameter to ON, run the following command:

```
ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS_ASYNC ON
```

To verify that the parameter is set correctly, run the following command:

```
SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName
```

- Set the AUTO_UPDATE_STATISTICS parameter to ON.
  To set the parameter to ON, run the following command:

```
ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON
```

To verify that the parameter is set correctly, run the following command:

```
SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName
```

- The database user account must have the CONNECT and CREATE TABLE privileges.

### Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:
  
  ```
  CREATE INDEX
  CREATE SESSION
  CREATE SYNONYM
  CREATE TABLE
  CREATE VIEW
  DROP TABLE
  INSERT INTO TABLE
  UPDATE TABLE
  ```

- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

### Model Repository Database Requirements

Secure@Source services and Informatica clients store data and metadata in the Model repository. Before you create the Model Repository Service, set up a database and database user account for the Model repository.

The Model repository supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Oracle

Allow 3 GB of disk space for DB2. Allow 200 MB of disk space for all other database types.

For more information about configuring the database, see the documentation for your database system.
IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- If the repository is in an IBM DB2 9.7 database, verify that IBM DB2 Version 9.7 Fix Pack 7 or a later fix pack is installed.
- On the IBM DB2 instance where you create the database, set the following parameters to ON:
  - DB2_SKIPINSERTED
  - DB2_EVALUNCOMMITTED
  - DB2_SKIPDELETED
  - AUTO_RUNSTATS
- On the database, set the configuration parameters.
  The following table lists the configuration parameters that you must set:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>applheapsz</td>
<td>8192</td>
</tr>
<tr>
<td>appl_ctl_heap_sz</td>
<td>8192</td>
</tr>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
</tbody>
</table>

- Set the tablespace pageSize parameter to 32768 bytes.
  In a single-partition database, specify a tablespace that meets the pageSize requirements. If you do not specify a tablespace, the default tablespace must meet the pageSize requirements.
  In a multi-partition database, specify a tablespace that meets the pageSize requirements. Define the tablespace in the catalog partition of the database.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Verify that the database user has CREATETAB, CONNECT, and BINDADD privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
• In the DataDirect Connect for JDBC utility, update the DynamicSections parameter to 3000.
  The default value for DynamicSections is too low for the Informatica repositories. Informatica requires a
  larger DB2 package than the default. When you set up the DB2 database for the domain configuration
  repository or a Model repository, you must set the DynamicSections parameter to at least 3000. If the
  DynamicSections parameter is set to a lower number, you can encounter problems when you install or run
  Informatica services.

**Microsoft SQL Server Database Requirements**

Use the following guidelines when you set up the repository on Microsoft SQL Server:

• Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention.  
  To set the isolation level for the database, run the following command:

  ```sql
  ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
  ```

  To verify that the isolation level for the database is correct, run the following command:

  ```sql
  SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
  ```

• Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON.  
  To set the parameter to ON, run the following command:

  ```sql
  ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
  ```

  To verify that the parameter is set correctly, run the following command:

  ```sql
  SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName
  ```

• Set the AUTO_UPDATE_STATISTICS_ASYNC parameter to ON.  
  To set the parameter to ON, run the following command:

  ```sql
  ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS_ASYNC ON
  ```

  To verify that the parameter is set correctly, run the following command:

  ```sql
  SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName
  ```

• Set the AUTO_UPDATE_STATISTICS parameter to ON.  
  To set the parameter to ON, run the following command:

  ```sql
  ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON
  ```

  To verify that the parameter is set correctly, run the following command:

  ```sql
  SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName
  ```

• The database user account must have the CONNECT, CREATE TABLE, and CREATE VIEW privileges.

**Oracle Database Requirements**

Use the following guidelines when you set up the repository on Oracle:

• Set the open_cursors parameter to 4000 or higher.

• Verify that the database user has the CONNECT, RESOURCE, and CREATE VIEW privileges.

• Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms
  have not been created for any tables in the database.
Profiling Warehouse Database Requirements

The profiling warehouse database stores profiling and scorecard results. You specify the profiling warehouse connection when you create the Data Integration Service.

The profiling warehouse supports the following database types:

- IBM DB2 UDB
- Microsoft SQL Server
- Oracle

Allow 10 GB of disk space for the database.

**Note:** Ensure that you install the database client on the machine on which you want to run the Data Integration Service. You can specify a JDBC connection or Hive connection as a profiling warehouse connection for IBM DB2 UDB, Microsoft SQL Server, and Oracle database types. You can create column profiles, rule profiles, data domain discovery profiles, and scorecards with a JDBC connection as a profiling warehouse connection.

For more information about configuring the database, see the documentation for your database system.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- The database user account must have the CREATETAB, CONNECT, CREATE VIEW, and CREATE FUNCTION privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.

**Note:** Informatica does not support the partitioned database environment for IBM DB2 databases when you use a JDBC connection as the profiling warehouse connection.

Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository on Microsoft SQL Server:

- Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention. To set the isolation level for the database, run the following command:

```
ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
```

To verify that the isolation level for the database is correct, run the following command:

```
SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
```

- Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON. To set the parameter to ON, run the following command:

```
ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
```

To verify that the parameter is set correctly, run the following command:

```
SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName
```

- Set the AUTO_UPDATE_STATISTICS_ASYNC parameter to ON. To set the parameter to ON, run the following command:
ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS_ASYNC ON
To verify that the parameter is set correctly, run the following command:
SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName

• Set the AUTO_UPDATE_STATISTICS parameter to ON.
To set the parameter to ON, run the following command:
ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON
To verify that the parameter is set correctly, run the following command:
SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName

• The database user account must have the CONNECT, CREATE TABLE, CREATE VIEW, and CREATE FUNCTION privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

• Verify that the database user has the following privileges:
  ALTER TABLE
  CREATE ANY INDEX
  CREATE PROCEDURE
  CREATE SESSION
  CREATE TABLE
  CREATE VIEW
  DROP TABLE
  UPDATE TABLE

• Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.

• Set the tablespace parameter. Use the following formula to determine the value: 2 MB x (number of tables in each scan x number of concurrent scans)
  For example, you have 1,000 tables in each scan and you plan to run 10 scans concurrently. Calculate the tablespace parameter value as follows: 2 MB x (100 x 10) = 20 GB
  Note: Tablespace must be distributed across multiple disks.

• Set the following parameters to the Informatica recommended values:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Recommended Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>open_cursors</td>
<td>3000</td>
</tr>
<tr>
<td>Sessions</td>
<td>1000</td>
</tr>
<tr>
<td>Processes</td>
<td>1000</td>
</tr>
</tbody>
</table>
Operating System Parameters

Set the following parameters to the Informatica recommended values:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Recommended Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>descriptors</td>
<td>65535</td>
</tr>
<tr>
<td>maxproc</td>
<td>65535</td>
</tr>
<tr>
<td>stacksize</td>
<td>50000 KB</td>
</tr>
</tbody>
</table>

Reference Data Warehouse Database Requirements

The reference data warehouse stores the data values for reference table objects that you define in a Model repository. You configure a Content Management Service to identify the reference data warehouse and the Model repository.

You associate a reference data warehouse with a single Model repository. You can select a common reference data warehouse on multiple Content Management Services if the Content Management Services identify a common Model repository. The reference data warehouse must support mixed-case column names.

The reference data warehouse supports the following database types:
- IBM DB2 UDB
- Microsoft SQL Server
- Oracle

Allow 200 MB of disk space for the database.

**Note:** Ensure that you install the database client on the machine on which you want to run the Content Management Service.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- Verify that the database user account has CREATETAB and CONNECT privileges.
- Verify that the database user has SELECT privileges on the SYSCAT.DBAUTH and SYSCAT.DBTABAUTH tables.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- Set the tablespace pageSize parameter to 32768 bytes.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository on Microsoft SQL Server:

- Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention. To set the isolation level for the database, run the following command:

  `ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON`

  To verify that the isolation level for the database is correct, run the following command:

  `SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName`

- Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON. To set the parameter to ON, run the following command:

  `ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON`

  To verify that the parameter is set correctly, run the following command:

  `SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName`

- Set the AUTO_UPDATE_STATISTICS_ASYNC parameter to ON. To set the parameter to ON, run the following command:

  `ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS_ASYNC ON`

  To verify that the parameter is set correctly, run the following command:

  `SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName`

- Set the AUTO_UPDATE_STATISTICS parameter to ON. To set the parameter to ON, run the following command:

  `ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON`

  To verify that the parameter is set correctly, run the following command:

  `SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName`

- The database user account must have the CONNECT and CREATE TABLE privileges.

Oracle Database Requirements

Use the following guidelines when you set up the repository on Oracle:

- Verify that the database user has the following privileges:

  `ALTER SEQUENCE`
  `ALTER TABLE`
  `CREATE SEQUENCE`
  `CREATE SESSION`
  `CREATE TABLE`
  `CREATE VIEW`
  `DROP SEQUENCE`
  `DROP TABLE`

  Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.
Secure@Source Repository Database Requirements

Secure@Source stores data and metadata, such as data stores and scans, in the Secure@Source repository. Before you create the Secure@Source Service, set up a database and database user account for the Secure@Source repository.

The Secure@Source repository supports the following database types:

- IBM DB@ UDB
- Microsoft SQL Server
- Oracle

Allow 3 GB of disk space for DB2. Allow 200 MB of disk space for all other database types.

For more information about configuring the database, see the documentation for your database system.

IBM DB2 Database Requirements

Use the following guidelines when you set up the repository on IBM DB2:

- If the repository is in an IBM DB2 9.7 database, verify that IBM DB2 Version 9.7 Fix Pack 7 or a later fix pack is installed.
- On the IBM DB2 instance where you create the database, set the following parameters to ON:
  - DB2_SKIPINSERTED
  - DB2_EVALUNCOMMITTED
  - DB2_SKIPDELETED
  - AUTO_RUNSTATS
- On the database, set the configuration parameters.
  The following table lists the configuration parameters that you must set:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>applheapsz</td>
<td>8192</td>
</tr>
<tr>
<td>appl_ctl_heap_sz</td>
<td>8192</td>
</tr>
<tr>
<td>logfilsiz</td>
<td>8000</td>
</tr>
<tr>
<td>maxlocks</td>
<td>98</td>
</tr>
</tbody>
</table>
Microsoft SQL Server Database Requirements

Use the following guidelines when you set up the repository on Microsoft SQL Server:

- Set the read committed isolation level to READ_COMMITTED_SNAPSHOT to minimize locking contention.
  To set the isolation level for the database, run the following command:
  ```sql
  ALTER DATABASE DatabaseName SET READ_COMMITTED_SNAPSHOT ON
  ```
  To verify that the isolation level for the database is correct, run the following command:
  ```sql
  SELECT is_read_committed_snapshot_on FROM sys.databases WHERE name = DatabaseName
  ```
- Set the ALLOW_SNAPSHOT_ISOLATION parameter to ON.
  To set the parameter to ON, run the following command:
  ```sql
  ALTER DATABASE DatabaseName SET ALLOW_SNAPSHOT_ISOLATION ON
  ```
  To verify that the parameter is set correctly, run the following command:
  ```sql
  SELECT is_allow_snapshot_isolation_on FROM sys.databases WHERE name = DatabaseName
  ```
- Set the AUTO_UPDATE_STATISTICS_ASYNC parameter to ON.
  To set the parameter to ON, run the following command:
  ```sql
  ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS_ASYNC ON
  ```
  To verify that the parameter is set correctly, run the following command:
  ```sql
  SELECT is_auto_update_statistics_async_on FROM sys.databases WHERE name = DatabaseName
  ```
- Set the AUTO_UPDATE_STATISTICS parameter to ON.
  To set the parameter to ON, run the following command:
  ```sql
  ALTER DATABASE DatabaseName SET AUTO_UPDATE_STATISTICS ON
  ```
  To verify that the parameter is set correctly, run the following command:
  ```sql
  SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = DatabaseName
  ```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>locklist</td>
<td>50000</td>
</tr>
<tr>
<td>auto_stmt_stats</td>
<td>ON</td>
</tr>
</tbody>
</table>

- Set the tablespace pageSize parameter to 32768 bytes.
  In a single-partition database, specify a tablespace that meets the pageSize requirements. If you do not specify a tablespace, the default tablespace must meet the pageSize requirements.
  In a multi-partition database, specify a tablespace that meets the pageSize requirements. Define the tablespace in the catalog partition of the database.
- Set the NPAGES parameter to at least 5000. The NPAGES parameter determines the number of pages in the tablespace.
- Verify that the database user has CREATETAB, CONNECT, and BINDADD privileges.
- Informatica does not support IBM DB2 table aliases for repository tables. Verify that table aliases have not been created for any tables in the database.
- In the DataDirect Connect for JDBC utility, update the DynamicSections parameter to 3000.
  The default value for DynamicSections is too low for the Informatica repositories. Informatica requires a larger DB2 package than the default. When you set up the DB2 database for the domain configuration repository or a Model repository, you must set the DynamicSections parameter to at least 3000. If the DynamicSections parameter is set to a lower number, you can encounter problems when you install or run Informatica services.
SELECT is_auto_update_statistics_on FROM sys.databases WHERE name = 'DatabaseName'

- The database user account must have the CONNECT, CREATE TABLE, and CREATE VIEW privileges.

**Oracle Database Requirements**

Use the following guidelines when you set up the repository on Oracle:

- Set the SESSIONS parameter to 1500.
- Set the PROCESSES parameter to 2272.
- Set the OPEN_CURSORS parameter to 2000 or higher.
- Verify that the database user has the CONNECT, RESOURCE, and CREATE VIEW privileges.
- Informatica does not support Oracle public synonyms for repository tables. Verify that public synonyms have not been created for any tables in the database.
Deployment Methods Overview

You can deploy Secure@Source either in an internal Hadoop distribution on Hortonworks or external Hadoop distribution on Cloudera version 5.5 or Hortonworks version 2.3. If you choose the external Hadoop cluster option to deploy Secure@Source, set up the Cloudera version 5.5 or Hortonworks version 2.3 cluster environment in your enterprise before you install Secure@Source. If you choose the internal Hadoop cluster option, you perform the Hortonworks cluster configuration when you install Secure@Source.

Internal Hadoop Cluster Deployment

The Secure@Source installer creates an Informatica Cluster Service as an ISP service if you choose the internal Hadoop distribution. Secure@Source uses Apache Ambari to manage and monitor the internal Hadoop cluster. The internal Hadoop cluster for Secure@Source supports the high availability option.

The following components of the Secure@Source internal Hadoop cluster environments support the high availability option:

- HDFS
• HBase
• Yarn
• Solr

Prerequisites - Internal Cluster

Before you install Secure@Source on an internal Hadoop cluster, you must verify that the system environment meets the prerequisites required to deploy Secure@Source.

Verify that the internal Hadoop distribution meets the following prerequisites:
• Operating system is 64-bit Red Hat Enterprise Linux version 6.5 or later.
• Verify that the cluster nodes meet the following requirements:

<table>
<thead>
<tr>
<th>Node Type</th>
<th>Minimum Requirements</th>
</tr>
</thead>
</table>
| Master node | - The number of CPUs is 4.  
- Unused memory available for use is 16 GB.  
- Disk space is 50 GB. |
| Slave node | - The number of CPUs is 4.  
- Unused memory available for use is 16 GB.  
- Disk space is 50 GB. |

• Ensure that the Linux base repositories are configured.
• Verify that you have the write permission on the /home directory.
• On each host machine, verify that you have the following tools and applications available:
  - YUM and RPM (RHEL/CentOS/Oracle Linux)
  - Zypper and php_curl (SLES)
  - apt (Ubuntu)
  - scp, curl, unzip, tar, and wget
  - awk
  - OpenSSL version 1.0.1e-30.el6_6.5.x86_64 or above
  - Python version 2.6

  **Note:** If you install SUSE Linux Enterprise 11, update all the hosts to Python version 2.6.8-0.15.1.

  - If any version of Java Development Kit (JDK) is installed, uninstall it.

  **Note:** Secure@Source installs JDK version 1.8 and PostgreSQL version 8.4 as part of Apache Ambari installation. The location of the JDK package is /var/lib/ambari-server/resources/jdk-8u60-linux-x64.tar.gz.

• Apache Ambari requires certain ports that are open and available during the installation to communicate with the hosts that Apache Ambari deploys and manages. You need to temporarily disable the tables to meet this requirement.
• Verify that you meet the memory and package requirements for Apache Ambari. For more information, refer to the Hortonworks documentation.
• When you install Apache Ambari Metrics, verify that you install kernel headers and kernel devel package as well.
Informatica Cluster Service

If the operating system is Red Hat Enterprise Linux version 6.x, rename `ambaribinariesRH6.tar.gz` to `ambaribinaries.tar.gz`.

If the operating system is Red Hat Enterprise Linux version 7.x, rename `ambaribinariesRH7.tar.gz` to `ambaribinaries.tar.gz`.

Copy the Apache Ambari binaries file to the installation directory: `Installer/sats/SecureAtSource_Installer_linux-x64_4.0/Server/LDM/source`

Verify that the maximum number of open file descriptors is 10,000 or more. Use the `ulimit` command to verify the current value and change the value if required.

If you plan to use a custom SSL setting and if the keystores for the Domain and Adminconsole are different, you must export the certificate from the Adminconsole keystore and add it to Domain keystore. Use the following commands to export and import the certificate:

```bash
keytool -export -keystore <ac_keystore_file> -storepass <ac_keystore_password> -file <file>.cert -alias <alias_of_ac_cert>

keytool -import -keystore <domain_keystore_file> -storepass <domain_keystore_password> -file <file>.cert -alias <alias_name>
```

For each Hadoop node, create the following folder: `/opt/SecureAtSource`

Verify that the Hadoop custom directory meets the following requirements:
- No spaces in name.
- A minimum of 50 GB disk space.
- Has file level read, write, and execute permissions.
- The EXEC flag is set.

Informatica Cluster Service

The Informatica Cluster Service is an application service that runs and manages all the Hadoop services, Apache Ambari server, and Apache Ambari agents on an internal Hadoop cluster. If you choose the internal cluster deployment mode, you need to create the Informatica Cluster Service before you create the Catalog Service. Then, you can pass the Informatica Cluster Service value to when you create the Catalog Service.

Informatica Cluster Service distributes the Hortonworks binaries and launches the required Hadoop services on the hosts where the internal cluster runs.

Informatica Cluster Service Workflow

The Informatica Cluster Service is an ISP service that manages the internal Hadoop cluster in Secure@Source.

After Informatica Cluster Service is created, it performs the following actions:

1. Launches the Apache Ambari server and associated agents.
2. Creates Hadoop services and monitoring systems on Apache Ambari including HDFS, Apache Zookeeper, Yarn, and related monitoring services.
3. Starts the Hadoop services.
4. When you shut down Secure@Source, the Informatica Cluster Service stops all the Hadoop services and stops the Apache Ambari server and its agents.
Creating an Informatica Cluster Service

You can choose to generate the Informatica Cluster Service when you install Secure@Source or create the application service manually using Informatica Administrator.

1. In the Administrator tool, select a domain, and click the Services and Nodes tab.
2. On the Actions menu, click New > Informatica Cluster Service.
   The New Informatica Cluster Service: Step 1 of 3 dialog box appears.
3. Configure the general properties in the dialog box.
   The following table describes the properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Name of the service. The name is not case-sensitive and must be unique within the domain. The name cannot exceed 128 characters or begin with @. The name cannot contain character spaces. The characters in the name must be compatible with the code page of the Model repository that you associate with the Catalog Service. The name cannot contain the following special characters: `~ % * + = { } \ ; : ' &quot; / ? . , &lt; &gt;</td>
</tr>
<tr>
<td>Description</td>
<td>Description of the service. The description cannot exceed 765 characters.</td>
</tr>
<tr>
<td>Location</td>
<td>Domain in which the application service runs.</td>
</tr>
<tr>
<td>License</td>
<td>License to assign to the Informatica Cluster Service. Select the license that you installed with Secure@Source.</td>
</tr>
<tr>
<td>Assign</td>
<td>Node configuration type. Specify whether the node is in a single node, high availability, or grid environment.</td>
</tr>
<tr>
<td>Node</td>
<td>Node in the Informatica domain on which the Informatica Cluster Service runs. If you change the node, you must recycle the Informatica Cluster Service.</td>
</tr>
</tbody>
</table>

4. Click Next.
   The New Informatica Cluster Service - Step 2 of 3 dialog box appears.
5. Configure the security properties in the dialog box.
   The following table describes the properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTTP Port</td>
<td>A unique HTTP port number used for each Data Integration Service process. The defaults is 8085.</td>
</tr>
<tr>
<td>Enable Transport Layer Security (TLS)</td>
<td>Indicates that the Informatica Cluster Service must use HTTPS. If you did not configure the Data Integration Service to use HTTPS, the Informatica Cluster Service does not start.</td>
</tr>
<tr>
<td>HTTPS Port</td>
<td>Port number for the HTTPS connection. Required if you select Enable Transport layer Security.</td>
</tr>
</tbody>
</table>
6. Click Next.

   The New Informatica Cluster Service - Step 3 of 3 dialog box appears.

7. Configure the Hadoop cluster properties in the dialog box.

   The following table describes the properties:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystore File</td>
<td>Path and file name of the keystore file. The keystore file contains the keys and certificates required if you use the SSL security protocol with Catalog Administrator. Required if you select Enable Transport layer Security.</td>
</tr>
<tr>
<td>Keystore Password</td>
<td>Password for the keystore file. Required if you select Enable Transport layer Security.</td>
</tr>
<tr>
<td>SSL Protocol</td>
<td>Secure Sockets Layer protocol to use.</td>
</tr>
</tbody>
</table>

   8. Click Finish.

Preparing the Internal Hadoop Cluster Environment

You need to perform multiple validation checks before you can before you install Secure@Source on an internal Hadoop cluster.

Perform the following steps before you install Secure@Source on an internal Hadoop cluster environment:

- Configure the /etc/hosts file on each machine so that you have fully qualified domain names. Informatica recommends the following host name format in lowercase:
  
  `<machine ipaddress> <fully qualified name> <alias>

  Note: To verify the configured host name, run the `hostname -f` command.

- Set up passwordless Secure Shell (SSH) connections between the following components:
  
  - From Informatica Cluster Service to Hadoop Gateway
  - From the Hadoop Gateway to Apache Hadoop nodes
  - For a multi-node cluster, set up a passwordless SSH connection from the Informatica domain to the Hadoop gateway host and Hadoop nodes.
Internal Cluster Node Management

A Hadoop cluster has a set of machines that is configured to run Hadoop applications and services. A typical Hadoop cluster includes a master node and multiple slave or worker nodes. The master node runs the master daemons JobTracker and NameNode. A slave node runs the DataNode and TaskTracker daemons. In small clusters, the master node might also run the slave daemons.

Cluster with High Availability

You can use the highly availability option for the HDFS, HBase, Yarn, and Solr components of the internal Hadoop cluster environment. If you set up Informatica Cluster Service on a multi-node and highly available cluster, you need a minimum of three nodes for Secure@Source to function successfully. If you have already set up Informatica Cluster Service on a single node, you cannot make the cluster highly available by adding more nodes to the cluster.

If the internal cluster contains only three nodes, Secure@Source distributes all master and slave services on all the three nodes. if the internal cluster contains more than three nodes, Secure@Source automatically chooses top three nodes with the highest system configuration as master nodes. The remaining nodes serve as slave nodes. When you add nodes to the internal cluster, the newly added nodes serve as slave nodes. The nodes that you add to the cluster must meet the minimum configuration requirements for slave nodes.

Cluster without High Availability

You can set up Informatica Cluster Service on a single node that is not highly available. In such cases, the master and worker nodes remain on the same node. You cannot bring up Informatica Cluster Service if you add a single node to an existing single-node cluster or try to set up Informatica Cluster Service with two nodes.

Delete Nodes

You can delete nodes from the internal cluster provided they meet the following conditions:

- You cannot delete a master node.
- You cannot delete a node if the number of live data nodes in the cluster becomes less than three on deleting the node.

Troubleshooting the Internal Cluster Deployment

I see failures with high availability in the internal cluster.

High availability is possible if you install Secure@Source the first time on more than two nodes. Verify that the count of cluster nodes for Secure@Source is not less than three. If you use a single node for Secure@Source during installation or add more nodes after the installation one after the other, you cannot implement high availability. If any one of the highly available hosts shuts down or is unreachable, Secure@Source might continue running but the cluster might not be highly available. You must then bring up the node that was shut down or unreachable to make the cluster highly available.

When Oracle and the Informatica Hadoop Service run on the same machine, the Informatica Hadoop Service fails to start.

Verify if Oracle is running on port 8080. Port 8080 is reserved for Apache Ambari. Make sure that you run Oracle on a different port if you run Oracle and Informatica Hadoop Service on the same machine.
The Apache Ambari tool fails to operate after I install Secure@Source on an internal Hadoop cluster.

- The Ambari host name that you specify during install is case-sensitive. Verify that the host name meets the requirements. Informatica recommends using lowercase for hostnames.

Run the following command to verify that the host name is successfully set up:

```
#hostname -f
```

The command returns the fully qualified domain name that you configured.

- If you installed Secure@Source on Red Hat Enterprise Linux (RHEL) version 6.5 or later, verify that all the base operating system repositories are available. The Secure@Source installer pulls many packages from the base operating system repositories. For example, the hosts must be able to access the Red Hat Enterprise Linux `rhel-6-serveroptional-rpms` repository for the installation to complete successfully. If you do not have the complete set of base operating system repositories, you might run into installation issues.

- Verify the version of Python installed. Secure@Source supports Python version 2.6.8-0.15.1 or earlier. Python version 2.7.9 or later is not supported.

- Before you install Secure@Source, verify that the ports 8080, 8440, and 8411 are unused. Apache Ambari uses these ports during the installation.

Informatica Cluster Service fails to respond.

Verify that you do not have `777` permissions on the `/var` directory. However, you must have the write permission on the `/var` directory.

The Ambari Server on the default PostgreSQL database shuts down or you want to move the Ambari server from one host to another.

You can perform the following steps to set up the Ambari Server on a new host:

1. Disable the Informatica Cluster Service using Informatica Administrator.
2. Restart the PostgreSQL instance on the Ambari Server host. Back up the PostgreSQL databases `ambarirca` and `ambari`.

   **Note:** To know more about backing up PostgreSQL databases, refer the *Moving the Ambari Server* section in Hortonworks documentation.

3. Update the Ambari Server host name value for Informatica Cluster Service with the new Ambari Server host name. To update the host name in Informatica Administrator, you can select the Catalog Service and then in the **Informatica Cluster Service** field under **Hadoop Cluster Options**, enter the host name.

4. Enable Informatica Cluster Service. This action installs Ambari Server in the new Hadoop Gateway host that you provided and updates the Ambari agent with the new Ambari Server host.

5. Disable Informatica Cluster Service.

6. Drop both `ambarirca` and `ambari` databases from the PostgreSQL instance on the new Hadoop gateway host.

7. Restore both `ambarirca` and `ambari` databases from the backup of the previous Hadoop gateway host.

   **Note:** To know more about restoring the PostgreSQL databases from a backup, refer the *Moving the Ambari Server* section in Hortonworks documentation.

8. Enable Informatica Cluster Service.

The Apache Ambari server fails to start up with the following error message in the Ambari Server log file:

```
ERROR: Exiting with exit code 1.REASON: Unable to start PostgreSQL server.
```

Add `127.0.0.1 localhost localhost.localdomain` to the `/etc/hosts` file.
I see that the cluster is not highly available.

Verify that the all the nodes that host high-availability components are up and running. If one of the nodes that hosts high-availability components shuts down, the cluster ceases to be highly available. Make sure that you bring up the node that had shut down.

When one of the slave nodes in the cluster is unreachable, I cannot enable Informatica Cluster Service.

If one of the slave nodes is unreachable due to unexpected failure, you must either remove the host from the list of Apache Ambari agents in Informatica Administrator or bring up the slave node again before you recycle Informatica Cluster Service.

I deleted an existing Catalog Service and enabled a new one. However, I am unable to access and use the data in the catalog.

If you want to point a new Catalog Service to the existing data in catalog, make sure you specify the same service cluster name used for the deleted Catalog Service for the new Catalog Service. You specify the service cluster name in Informatica Administrator under the Hadoop Cluster Options section on the Properties tab.

After I add a few nodes to an existing internal cluster, the Apache Ambari Metrics System fails to launch.

This failure can occur due to some machines, especially virtual machines, not having the required system packages. Make sure that the new hosts that you add meet the pre-requirements for the internal cluster deployment of Secure@Source. Refer to the Prerequisites and Preparing the Internal Hadoop Cluster Environment sections in the Deployment Methods chapter of this guide.


One of the prerequisites for Apache Ambari Agent to work successfully is an OpenSSL version 16 or later. You can perform the following steps:

1. Verify the OpenSSL library version installed on your hosts using the following command:

   ```
   rpm -qa | grep openssl
   ```

2. If the output of the previous command reads openssl-1.0.1e-15.x86_64 (1.0.1 build 15), you must upgrade the OpenSSL library. To upgrade the OpenSSL library, run the following command:

   ```
   yum upgrade openssl
   ```

3. Verify that the newer version of OpenSSL is installed using the `rpm` command.

4. Restart Apache Ambari Agents and click Retry > Failed in the wizard user interface.

When you install Secure@Source on an internal cluster, the installer does not create Informatica Cluster Service due to permission issues.

If the root user does not have the required permission to add users under the user home directory, the task log file displays the following error message:

```_execution of 'useradd -m -G hadoop -g hadoop mapred' returned 12. useradd: cannot create directory /home/mapred
```

You can then change the user home directory for the virtual machine to the directory that has the required permission for root user to create a directory in it:

```
sudo vim /etc/default/useradd
```

Change home directory to `HOME=/export/home`. 

---

58 Chapter 4: Deployment Methods
The Apache Ambari Application Timeline Server fails with the following error:

```
ps -p 'hadoop-yarn/yarn/yarn-timelineserver.pid' failed: https://issues.apache.org/jira/browse/AMBARI-4825
```

This issue can occur due to more than one active terminal. Use the w, whoami, and kill commands to verify and terminate all terminals except the one that is relevant. You can then restart the YARN cluster.

The Apache Ambari server installation fails on some of the host machines and the following error shows up in the Ambari Server log files: Ambari-server status Traceback (most recent call last) File "/usr/sbin/ambari-server.py", line 26, in <module> from ambari_commons.exceptions import FatalException, NonFatalException ImportError: No module named ambari_commons.exceptions.

This issue occurs when you try to install Ambari Server on a single host that has an Ambari Agent from a previous installation. The link /usr/lib/python2.6/site-packages/ambari_commons needs to point to /usr/lib/ambari-server/lib/ambari_commons instead of /usr/lib/ambari-agent/lib/ambari_commons.

When you add only two nodes to the cluster, the cluster setup fails with the following error code 00030: Cannot create a cluster with 2 hosts. Minimum 3 hosts are required for creating the High Availability cluster.

You cannot create a cluster with only two nodes. You need to provide at least three nodes for a highly available cluster and one host for a cluster that is not highly available. You can add nodes to an existing single-node cluster that is already set up. However, you cannot configure the cluster as a highly available cluster.

When you create a cluster, the verification of minimum system configuration fails with one of the following error codes: 00026, 00027, or 00028.

If you have a single-node cluster, the host machine must meet the minimum configuration criteria for master nodes. If you have a highly available cluster, at least three of the host machines must meet the minimum configuration criteria for master nodes. The remaining host machines must meet the minimum configuration criteria for slave nodes.

You are unable to delete a node from the cluster and you see one of the following error codes in the log file: 00035 or 00036.

Error code 00035 indicates that on deleting the node the number of live data nodes in the cluster reduces to less than three. The minimum number of live nodes required in the cluster is three. Error code 0036 indicates that there was an attempt to remove a node that hosts master services. You cannot remove a node that hosts master services.

HDFS data node shows up errors frequently indicating unknown operations with the following errors in the data node log files:
```
DataXceiver error processing unknown operation src: /127.0.0.1:33349 dst: /127.0.0.1:50010 java.io.EOFException at java.io.DataInputStream.readShort
DataInputStream.java:315) at org.apache.hadoop.hdfs.protocol.datatransfer.Receiver.readOp
Receiver.java:58)
```

Apache Ambari opens a socket connection to HDFS data node every minute to monitor it. You can ignore these errors as they do not affect the cluster operations.

I have Informatica Domain, Data Integration Service, and Content Management Service secured with the SSL protocol. If I change the Catalog Service settings to enable the SSL mode, the service fails to start. How do I resolve the issue?

After you configure Informatica Domain, Data Integration Service, and Content Management Service to secure them with SSL protocol using the Secure@Source installer, perform the following steps:

1. Export the Catalog Service keystore certificate.
2. Import the keystore certificate into the Informatica truststore file.
3. Place Informatica truststore file on all Hadoop nodes. Make sure that you follow the same directory structure for the truststore file on all host machines.
4. In Informatica Administrator, specify the common Informatica truststore file location in the **Domain TrustStore File Location** field of the **Advance Options** section for the Informatica Cluster Service.

5. In the **Edit Security Properties** dialog box on the **Processes** tab of the Informatica Cluster Service, select **Enable Transport Layer Security (TLS)** and specify the path to the keystore file mentioned in Step 1.

6. Enable Informatica Cluster Service. If the service already enabled, disable the service in the **Complete** mode and then enable the service again.

7. Navigate to the Security Properties section of the Catalog Service and specify the path to the keystore file mentioned in Step 1.

8. Enable the Catalog Service.

---

**External Hadoop Cluster Deployment**

You can deploy Secure@Source on a Hadoop cluster that you have set up on Cloudera packages and parcel version 5.10 or Hortonworks version 2.5. If you have enabled Kerberos authentication in your enterprise to authenticate users and services on a network, you can configure the Informatica domain to use Kerberos network authentication.

You must configure Zookeeper, HDFS, and Yarn specifications when you install Secure@Source on an external Hadoop cluster in your enterprise. Catalog Service uses these specifications and launches the following services and components on the Hadoop cluster as Yarn application:

- Solr version 6.5.0
- HBase version 1.1.2
- Scanner components

**Prerequisites for the External Hadoop Cluster**

Before you install Secure@Source on a Cloudera package or parcel version 5.10 or Hortonworks version 2.5 environment, you must verify that the system environment meets the prerequisites required to deploy Secure@Source.

Verify that the external Hadoop distribution meets the following prerequisites:

- OpenSSL version on the cluster nodes is openssl-1.0.1e-30.el6_6.5.x86_64. If you have a previous version of OpenSSL, upgrade to this version before you install Secure@Source on Cloudera.
- You have Read, Write, and Execute permissions for owners, groups, and others on HDFS directories.
- If you plan to use a custom SSL setting and if Cloudera or Hortonworks is SSL enabled, then perform the following actions:
  - Export the certificate from the Domain keystore and import it to the Cloudera or Hortonworks truststore.
  - Export the certificate from the Cloudera or Hortonworks keystore and import it to the Domain truststore.

Use the following command to export the certificate:

```
keytool -export -keystore <ac_keystore_file> -storepass <ac_keystore_password> -file <file>.cert -alias <alias_of_ac_cert>
```

Use the following command to import the certificate:

```
keytool -import -keystore <domain_keystore/truststore file> -storepass <domain_keystore/truststore password> -file <file>.cert -alias <alias_name>
```
• Java 8 is installed on the external Hadoop cluster.
• The JAVA_HOME environmental variable is set to the Java 8 location. Alternatively, the /usr/bin/java link points to the Java 8 location.

Preparing the External Hadoop Cluster Environment

You must perform multiple validation checks before you install Secure@Source on an external Hadoop cluster.

Perform the following steps before you install Secure@Source on an external Cloudera cluster:
• If you enabled Kerberos on the cluster in your enterprise, create the required keytab files.
• Create the following directories in HDFS before you create the Catalog Service:
  - /Informatica/LDM/<ServiceClusterName>
  - /Informatica/LDM/<ServiceClusterName>/service-logs
  - /user/<username>

Where <ServiceClusterName> is the name of the service cluster that you need to enter when you create the Catalog Service and <username> is the user name of the Informatica domain user.
• Make the Informatica domain <username> the owner of the /Informatica/LDM/<ServiceClusterName> and /user/<username> directories.
• Make Hadoop a group of /Informatica/LDM/<ServiceClusterName>/service-logs.
• When you create the Catalog Service that connects to an SSL-enabled Kerberos external Cloudera or Hortonworks cluster, verify that you configure the following properties:
  - A keytab file that contains all the users in LDAP.
  - Kerberos domain name.
  - HDFS namenode and YARN Resource Manager service principals
  - Path to Solr keystore file and password.
• Create the folder <SecureAtSource_ServiceName>_elasticsearch in the directory /var/lib/hadoop-yarn.
  The owner of this folder must be the same as the Principal User and the group must be 1000.
  The ServiceClusterName must be same as the Principal User.
  The folder name you enter in <SecureAtSource_ServiceName>_elasticsearch must be in lower case.

Example:

    sudo mkdir /var/lib/hadoop_yarn/sats_elasticsearch
    sudo chown infa:1000 /var/lib/hadoop_yarn/sats_elasticsearch
Kerberos and SSL Setup for an External Cluster

You can install Secure@Source on an external cluster that uses Kerberos network authentication to authenticate users and services on a network. Secure@Source also supports SSL authentication for secure communication in the cluster.

Kerberos is a network authentication protocol which uses tickets to authenticate access to services and nodes in a network. Kerberos uses a Key Distribution Center (KDC) to validate the identities of users and services and to grant tickets to authenticated user and service accounts. In the Kerberos protocol, users and services are known as principals. The KDC has a database of principals and their associated secret keys that are used as proof of identity. Kerberos can use an LDAP directory service as a principal database.

Informatica does not support cross or multi-realm Kerberos authentication. The server host, client machines, and Kerberos authentication server must be in the same realm.

The Informatica domain requires keytab files to authenticate nodes and services in the domain without transmitting passwords over the network. The keytab files contain the service principal names (SPN) and associated encrypted keys. Create the keytab files before you create nodes and services in the Informatica domain.

Prerequisites for SSL Authentication

Verify that the external cluster meets the following requirements before you can enable SSL authentication in the cluster:

- Informatica domain is configured in the SSL mode.
- The cluster and YARN REST endpoints are Kerberos-enabled.
- Create a keystore file for the Apache Solr application on all nodes in the cluster. Import public certificates of Apache Solr keystore files on all the hosts into all the truststore files configured for HDFS and YARN. This step is required for Apache Spark and scanner jobs to connect to the Apache Solr application.
- Import the public certificates of Apache Solr and YARN applications into the truststore file of the Informatica domain. This step is required for Catalog Service to connect to YARN and Solr applications.

Prerequisites for Kerberos Authentication

Perform the following steps before you enable the Kerberos authentication for the external cluster:

- Create the following users in the LDAP security domain where <username> is the service cluster name.
  - <username>@KERBEROSDOMAIN.COM
  - <username>/<hostname>@KERBEROSDOMAIN.COM
    
    **Note:** Create the user ID for all the hosts in the cluster.
  - HTTP/<hostname>@KERBEROSDOMAIN.COM
    
    **Note:** Create the user ID for all the hosts in the cluster.
  - Create a keytab file with credentials for all these users created in LDAP. You can create keytab files for each one of the users in KDC server and merge them with the hdfs.keytab file using the ktutil command to create a single keytab file.
    
    The hdfs.keytab file is in the directory /var/run/cloudera-scm-agent/process/</LatestNumber>-hdfs-DATANODE on the Cloudera server machine.
  - Create the following folders in HDFS that Enterprise Information Catalog uses as data directories for the Catalog Service: /Informatica/LDM/<username> and /user/<username>.
  - Change the owner of these two folders to <username>.

- Set up the udp_preference_limit parameter in the krb5.conf Kerberos configuration file to 1. This parameter determines the protocol that Kerberos uses when it sends a message to the KDC. Set
udp_preference_limit = 1 to always use TCP. The Informatica domain supports only the TCP protocol. If the udp_preference_limit parameter is set to any other value, the Informatica domain might shut down unexpectedly.

- Open the ports on the Secure@Source host:
  - 6005 to 6009
  - 6013
  - 6113
  - 52,652
  - 18,295
  - 18,105
  - 9,553
  - 9,443
  - 9,085
  - 8,443
  - 18,095
  - 9090
- Verify all system requirements. For more information, see "Verify System Requirements" on page 21.

## Troubleshooting the External Cluster Deployment

The ingestion of metadata into the catalog does not happen successfully.

Verify that the script files in the Secure@Source installer that launch the Solr and Ingestion jobs are copied to the target host of the external cluster.

**Can I implement Kerberos security if I install Secure@Source on an external cluster?**

Yes. Secure@Source supports Kerberos network authentication on an external cluster.

**I see that the Catalog Service has unexpectedly shut down and the error message in the log file reads as follows: "GSSEException: No valid credentials provided (Mechanism level: Server not found in Kerberos database". How do I resolve the issue?**

Verify that all the cluster nodes in the /etc/hosts domain have fully qualified host names and fix the incorrect host names.

**I have a Cloudera version 4 cluster environment and am not able to install Secure@Source on the cluster.**

Secure@Source supports Cloudera version 5.4 or later for external Hadoop cluster deployment. Upgrade the Cloudera version to 5.4 or later.

**I added a few hosts that contain high availability services to the cluster using CDH Manager. However, I do not see the nodes enabled.**

If you added nodes to an existing cluster, verify that you added them to the Hadoop nodes list in Informatica Administrator and restart the Catalog Service.

**I encounter issues with Kerberos Distribution Center (KDC) availability and see messages similar to the following one: "(java.security.PrivilegedActionException: javax.security.sasl.SaslException: GSS initiate failed [Caused by GSSEException: No valid credentials provided (Mechanism level: Connection reset)]) occurred when evaluating**
Zookeeper Quorum Member's received SASL token. Zookeeper Client will go to AUTH_FAILED state. How do I resolve the issues?

A Kerberos-enabled cluster requires a highly available KDC. Verify that you have enabled high availability for KDC.

The ingestion service is terminated by Apache Yarn due to memory overhead issues with errors similar to the following one: "Container killed by YARN for exceeding memory limits. 10.0 GB of 10 GB physical memory used. Consider boosting spark.yarn.executor.memoryOverhead." How do I resolve the errors?

The issue occurs due to Apache YARN memory checks. It is recommended that the following two properties are set to False:

- yarn.nodemanager.pmem-check-enabled
- yarn.nodemanager.vmem-check-enabled

After you abruptly shut down a domain and Catalog Service, you see that YARN application continue to run.

If you do not shut down the domain gracefully, the YARN applications for HBase, Solr, and Spark might continue to run. You must manually shut down these YARN applications before you bring up the domain and application services again.

Troubleshooting Common Cluster Deployment Issues

The Apache Zookeeper client connection count is low and you see the following error in the Zookeeper log file: "Too many connections from <ipaddress>- max is 60." You might also encounter Ingestion Service failures with the following error message in the log files: "Unexpected error, closing socket connection and attempting reconnect java.io.IOException: Connection reset by peer."

Apache Zookeeper is a shared application and requires multiple open and configured connections.

Change the value of the maxclientCnxns parameter to the recommended value based on the cluster load and start the entire cluster again.

The Catalog Service fails to start automatically when you recycle the Informatica Hadoop Service.

You must restart the Catalog Service manually when you recycle the Informatica Hadoop Service.

One of the cluster components fails to start up with the following error in the log file: "Caused by: java.langNumberFormatException: For input string: "0LdmCustomOptions.HbaseMasterProperties."

This issue might occur due to incorrect custom property settings for the Catalog Service that were set up in Informatica Administrator. Verify that each custom property that you need to update is defined as a separate LdmCustomOptions parameter in Informatica Administrator. You can then start the Catalog Service again to bring up the cluster.

Catalog Service fails to respond with the following error message in the log file: "Connection timed out for connection string () and timeout () / elapsed () org.apache.curator.CuratorConnectionLossException: KeeperErrorCode = ConnectionLoss at org.apache.curator.ConnectionState.checkTimeouts(ConnectionState.java:197). The Ingestion client or HBase log file might contain the following error message: "Possibly transient ZooKeeper, quorum=…, exception=org.apache.zookeeper.KeeperException ConnectionLossException: KeeperErrorCode = ConnectionLoss for / hbase/meta-region-server."

This issue might occur due to the inability of the Catalog Service to reach some of the core cluster components, such as Apache Zookeeper. The Zookeeper issues can occur due to temporary disk issues. Fix the disk issues and verify that Apache Zookeeper is up and running.
Apache Zookeeper log file displays the following error message due to high disk latency: "fsync-ing the write ahead log in SyncThread:3 took 25115ms which will adversely affect operation latency."

It is recommended that you allocate a dedicated disk, not a disk partition, to Apache Zookeeper because it has to provide high-consistency guarantees to its client. Verify that you have allocated the recommended number of hard disks for the size of the workload. You also need to point the Zookeeper data directory to its dedicated disk.

Some of the quorum members of the Apache Zookeeper ensemble are not reachable and the log file displays warnings similar to the following message: "[QuorumPeer[myid=3]/0:0:0:0:0:0:0:0:2181:QuorumCnxManager@383] - Cannot open channel to 2 at election address 10.65.144.18:3888 java.net.ConnectException: Connection refused at java.net.PlainSocketImpl.socketConnect(Native Method)"

Verify that the Zookeeper hosts are reachable over the network. It is important that Zookeeper quorum members do not run out of memory. Verify that the nodes meet the recommended memory requirements. Verify that only processes related to the Catalog Service run on the same host.

Multiple Apache Zookeeper quorum members display session termination messages at the client, such as the following message: "Caught end of stream exception EndOfStreamException: Unable to read additional data from client sessionid 0x0, likely client has closed socket, Processed session termination for sessionid."

Monitor the heap memory usage for Zookeeper quorum members. You can consider increasing the heap memory for Zookeeper and start the entire cluster again.

Ingestion Service fails to start up and the log file displays the following error message: "Initial job has not accepted any resources; check your cluster UI to ensure that workers are registered and have sufficient memory".

This error indicates insufficient amount of memory or CPU cores in the entire cluster. Verify that the cluster has sufficient resources to launch new applications.

The Catalog Service fails to start after you change the LdmCustomOptions.loadType custom property in Informatica Administrator and the log file contains the following error: "Caused by: org.springframework.beans.factory.BeanCreationException: Error creating bean with name 'hbaseGraphFactory': Invocation of init method failed; nested exception is com.thinkaurelius.titan.core.TitanConfigurationException: Local settings present for one or more globally managed options: [cluster.max-partitions]. These options are controlled through the ManagementSystem interface; local settings have no effect."

You must back up all data before you change the load type custom property, change the load type setting, start the cluster again, and then load the backed up data.

The Catalog Service or Ingestion Service fails due to unavailability of some HBase database tables with the following error in the log files: "Caused by: com.thinkaurelius.titan.diskstorage.TemporaryBackendException: Temporary failure in storage backend at com.thinkaurelius.titan.diskstorage.HBaseStoreManager.ensureTableExists (HBaseStoreManager.java:754) Caused by: org.apache.hadoop.hbase.TableNotFoundException: ldmns:exDocStore at org.apache.hadoop.hbase.master.procedure.CreateTableProcedure.prepareCreate."

This error occurs due to improper cleanup of data related to the Catalog Service or Ingestion Service. If you changed the load type using the LdmCustomOptions.loadType custom property in Informatica Administrator, verify that you backed up all the service-related data, removed the data completely, and then reloaded it.

The Catalog Service or Ingestion Service fails due to the presence of some HBase database tables with the following error in the log files: "Caused by: org.apache.hadoop.hbase.ipc.RemoteWithExtrasException (org.apache.hadoop.hbase.TableExistsException): org.apache.hadoop.hbase.TableExistsException: ldmns:exDocStore at org.apache.hadoop.hbase.master.procedure.CreateTableProcedure.prepareCreate."

This error occurs due to improper cleanup of data related to the Catalog Service or Ingestion Service. If you changed the load type using the LdmCustomOptions.loadType custom property in Informatica Administrator, verify that you backed up all the service-related data, removed the data completely, and then reloaded it.

The Catalog Service or Ingestion Service fails due to some disabled HBase database tables with the following error in the log files: "Caused by: org.apache.hadoop.hbase.TableNotEnabledException: ldmns:DataDomain_stage is disabled at
This error occurs due to improper cleanup of data related to the Catalog Service or Ingestion Service. If you changed the load type using the `LdmCustomOptions.loadType` custom property in Informatica Administrator, verify that you backed up all the service-related data, removed the data completely, and then reloaded it.

The Catalog Service or Ingestion Service fails with one of the following errors in the log files: HBase log file contains the error message "Caused by: com.thinkaurelius.titan.diskstorage.TemporaryBackendException: Temporary failure in storage backend Caused by: org.apache.hadoop.hbase.client.RetriesExhaustedException: Failed after attempts=4, exceptions: failed on local exception: java.io.IOException: Connection reset by peer This server is in the failed servers list." The Ingestion Service log file might contain the error "Caused by: org.apache.spark.SparkException: Job aborted due to stage failure: Task 0 in stage 9468.0 failed 4 times, most recent failure: Lost task 0.3 in stage 9468.0 (TID 12018): org.apache.hadoop.hbase.client.RetriesExhaustedException: Failed after attempts=4, exceptions: This server is in the failed servers list."

The error might occur when HBase server is not reachable due to factors, such as network partitioning, HBase region server unavailability due to workload, or its internal housekeeping activities such as data splitting and compactions. You can try restarting the Catalog Service with increased memory for HBase.

HBase servers might fail if Apache Zookeeper or HDFS is not reachable. The Catalog Service tries to automatically start HBase instances up to the configured number of attempts unless the error is fatal. In such cases, you might have to manually restart the Catalog Service.

Apache YARN application shuts down intermittently and cluster nodes are not used to submit cluster applications. The YARN Resource Manager log file contains the following error message: "Node irl66dsq04.xxx.com:8041 reported UNHEALTHY with details: 1/1 log-dirs are bad: /var/log/hadoop-yarn/container, Node Transitioned from RUNNING to UNHEALTHY, Container Transitioned from RUNNING to KILLED, Removed node irl66dsq04.xxx.com:8041 cluster capacity: <memory:184320, vCores:96>"

Verify the disk space for `/partition` on hard disk using commands, such as `df`. Apache YARN considers a node unhealthy and terminates the node applications if disk space use is greater than 80%. Delete unnecessary data from `/partition`. If you have multiple disks, then point `/partition` to an uncontended disk.

The HBase region server shuts down with an error message similar to the following one in the log file: "Sleeper: Slept 15559ms instead of 3000ms, this is likely due to a long garbage collecting pause and it’s usually bad. HeapMemoryManager: heapOccupancyPercent 0.9935025 is above heap occupancy alarm watermark (0.95). JvmPauseMonitor: Detected pause in JVM or host machine (eg GC): pause of approximately 3733ms. GC pool ‘ParNew’ had collection(s): count=1 time=4075ms."

This error occurs due to HBase memory issues. Verify that you configured the correct workload or data set type for your Enterprise Information Catalog workload using the `LdmCustomOptions.loadType` custom property in Informatica Administrator. Sometimes, You might need to manually increase the HBase heap settings using Informatica Administrator and then restart the Catalog Service.

The Ingestion Service fails with memory issues and the Ingestion Service log file contains error messages similar to the following message: "TaskSetManager: Lost task 1.0 in stage 18.0 (TID 39, INVRGX65CMD03.informatica.com): org.apache.spark.util.TaskCompletionListenerException: GC overhead limit exceeded at org.apache.spark.TaskContextImpl.markTaskCompleted(TaskContextImpl.scala:83) at org.apache.spark.scheduler.Task.run(Task.scala:72)"

This error occurs due to reduced HBase memory. Verify that you configured the correct workload or data set type for your Enterprise Information Catalog workload using the `LdmCustomOptions.loadType` custom property in Informatica Administrator. In some cases, You might need to manually increase the HBase heap settings using Informatica Administrator and then restart the Catalog Service.

This error indicates that the Hadoop file system is not reachable. Use the Apache Ambari user interface to verify that HDFS is up and running.

An HDFS data node fails with an error message similar to the following message: "BlockStateChange: BLOCK NameSystem.addToCorruptReplicasMap: blk_1073876841 added as corrupt on 10.65.145.216:50010 by irlcmg07.informatica.com/10.65.145.216 because reported RWR replica with genstamp 136273 does not match COMPLETE block's genstamp in block map 138353."

This error usually occurs in a single-node deployment because the data is not replicated. The issue might occur due to data corruption in some HDFS data blocks. The data corruption might happen due to either a corrupt disk volume or disk that has its memory full.

If you have more than one disk directory configured for HDFS, you can try changing the value for dfs.datanode.failed.volumes.tolerated from 0. A value of 0 results in data node shutdown even with a minimum of one disk volume that has corrupt data.

Where can I find all the log files related to Informatica Cluster Service and how do I troubleshoot using the log files?

You can find out the details of issues related to the Informatica Cluster Service by performing the following steps:

1. Open and verify the Informatica Cluster Service log file at the following location: $<Install Directory>/logs/<Informatica Cluster Service Node Name>/services/InfraHadoopService/<Informatica Cluster Service Name>.
2. Open and verify the log file for the Apache Ambari server on Ambari server host at the following location: /var/log/ambari-server
3. Open and verify the log file for the Apache Ambari agent on Ambari agent host at the following location: /var/log/ambari-agent.
4. If you cannot troubleshoot the issue by following the previous steps, perform the following steps:
   a. Launch the Apache Ambari application using the http://<ambari-server host>:8080/ URL.
   b. Click ops at the top of the application to verify the failed requests:

![Ambari Interface](image)

   c. Record the name of the host on which the request failed and the Hadoop component related to the failed request.
   d. Log in to the host where the request failed.
   e. Verify the log file for the specific Hadoop component related to the failed request at the following locations:

<table>
<thead>
<tr>
<th>Name of the Component</th>
<th>Log File Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>NameNode</td>
<td>/var/log/hadoop/hdfs or /var/log/hadoop-hdfs</td>
</tr>
<tr>
<td>SecondaryNameNode</td>
<td>/var/log/hadoop/hdfs or /var/log/hadoop-hdfs</td>
</tr>
<tr>
<td>JournalNode</td>
<td>/var/log/hadoop/hdfs or /var/log/hadoop-hdfs</td>
</tr>
<tr>
<td>ZKFC</td>
<td>/var/log/hadoop/hdfs or /var/log/hadoop-hdfs</td>
</tr>
</tbody>
</table>
### Name of the Component | Log File Location
---|---
DataNode | /var/log/hadoop/hdfs or /var/log/hadoop-hdfs
HistoryServer | /var/log/hadoop/mapreduce or /var/log/hadoop-mapreduce
MetricsCollector | /var/log/ambari-metrics-collector
MetricsMonitor | /var/log/ambari-metrics-monitor
AppTimelineServer | /var/log/hadoop-yarn or /var/log/hadoop/yarn
ResourceManager | /var/log/hadoop-yarn or /var/log/hadoop/yarn
NodeManager | /var/log/hadoop-yarn or /var/log/hadoop/yarn
ZookeeperServer | /var/log/zookeeper

**Note:** If there are some Hadoop services running that were not stopped when you shut down the Informatica Cluster Service previously, you might see the text `java.net.BindException: Address already in use` in the log file. You must then shut down such processes on the ports mentioned in the exception.

**Where can I find all the log files for Apache YARN applications, such as Solr, HBase, and Ingestion Service?**

You can perform the following steps to view the log files:

1. Log in to the Apache Ambari user interface, and click the **Service** tab at the top of the page to open the following page:

![Ambari Service Tab](image)

2. Click **Quick Links > ResourceManager UI** to open the following page that lists all the applications:

![ResourceManager UI](image)

You can view the different applications launched by that Catalog Service on Apache YARN. The **State** column indicates the current status of applications.

3. Click the link under the **ID** column to open the following page:

![Application Details](image)

4. To view the log file, click **logs**.
The Ingestion Service fails with the following error message:

"java.io.IOException: Connection reset by peer." when running jobs in a sequence.

This issue occurs due to limited Zookeeper client connections allowed for Secure@Source. You can change the Zookeeper client connection value to 0, which indicates unlimited connections.
Prepare for Kerberos Authentication Setup

This chapter includes the following topics:

- **Prepare for Kerberos Authentication Overview, 70**
- **Set Up the Kerberos Configuration File, 71**
- **Generate the Service Principal and Keytab File Name Format, 72**
- **Review the SPN and Keytab Format Text File, 76**
- **Create the Service Principal Names and Keytab Files, 78**

## Prepare for Kerberos Authentication Overview

You can configure the Informatica domain to use Kerberos network authentication to authenticate users, services, and nodes.

Kerberos is a network authentication protocol which uses tickets to authenticate access to services and nodes in a network. Kerberos uses a Key Distribution Center (KDC) to validate the identities of users and services and to grant tickets to authenticated user and service accounts. In the Kerberos protocol, users and services are known as principals. The KDC has a database of principals and their associated secret keys that are used as proof of identity. Kerberos can use an LDAP directory service as a principal database.

To use Kerberos authentication, you must install and run the Informatica domain on a network that uses Kerberos network authentication. Informatica can run on a network that uses Kerberos authentication with Microsoft Active Directory service as the principal database.

The Informatica domain requires keytab files to authenticate nodes and services in the domain without transmitting passwords over the network. The keytab files contain the service principal names (SPN) and associated encrypted keys. Create the keytab files before you create nodes and services in the Informatica domain.

Before you configure Kerberos authentication for the domain, perform the following tasks:

- Set up the Kerberos configuration file.
- Generate the service principal and keytab file names in the Informatica format.
- Review the SPN and keytab format text file.
- Request the Kerberos administrator to add the SPN to the Kerberos principal database and create the keytab files.
Set Up the Kerberos Configuration File

Kerberos stores configuration information in a file named krb5.conf. Informatica requires specific properties set in the Kerberos configuration file so that the Informatica domain can use Kerberos authentication correctly. You must set the properties in the krb5.conf configuration file.

The configuration file contains the information about the Kerberos server, including the Kerberos realm and the address of the KDC. You can request the Kerberos administrator to set the properties in the configuration file and send you a copy of the file.

1. Back up the krb5.conf file before you make any changes.
2. Edit the krb5.conf file.
3. In the libdefaults section, set or add the properties required by Informatica.

The following table lists the values to which you must set properties in the libdefaults section:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>default_realm</td>
<td>Name of the service realm for the Informatica domain.</td>
</tr>
<tr>
<td>forwardable</td>
<td>Allows a service to delegate client user credentials to another service. Set this parameter to True. The Informatica domain requires application services to authenticate the client user credentials with other services.</td>
</tr>
<tr>
<td>default_tkt_enctypes</td>
<td>Encryption types for the session key in ticket-granting tickets (TGT). Set this parameter only if session keys must use specific encryption types.</td>
</tr>
<tr>
<td>udp_preference_limit</td>
<td>Determines the protocol that Kerberos uses when it sends a message to the KDC. Set udp_preference_limit = 1 to always use TCP. The Informatica domain supports only the TCP protocol. If the udp_preference_limit is set to any other value, the Informatica domain can shut down unexpectedly.</td>
</tr>
</tbody>
</table>

4. In the realms section, include the port number in the address of the KDC separated by a colon.

For example, if the KDC address is kerberos.example.com and the port number is 88, set the kdc parameter to the following:

```
kdc = kerberos.example.com:88
```

5. Save the krb5.conf file.
6. Store the krb5.conf file in a directory that is accessible to the machine where you plan to install the Informatica services.

The following example shows the content of a krb5.conf with the required properties:

```
[libdefaults]
default_realm = AFNIKRBFNIDEV.COM
forwardable = true
udp_preference_limit = 1

[realms]
AFNIKRBFNIDEV.COM = {
    admin_server = SMPLKERDC01.AFINKRBFNIDEV.COM
    kdc = SMPLKERDC01.AFINKRBFNIDEV.COM:88
}

[domain_realms]
afnikrbafnidev.com = AFNIKRBFNIDEV.COM
.afnikrbafnidev.com = AFNIKRBFNIDEV.COM
```

For more information about the Kerberos configuration file, see the Kerberos network authentication documentation.
Generate the Service Principal and Keytab File Name Format

If you run the Informatica domain with Kerberos authentication, you must associate Kerberos service principal names (SPN) and keytab files with the nodes and processes in the Informatica domain. Informatica requires keytab files to authenticate services without requests for passwords.

Based on the security requirements for the domain, you can set the service principal level to one of the following levels:

**Node Level**

If the domain is used for testing or development and does not require a high level of security, you can set the service principal at the node level. You can use one SPN and keytab file for the node and all the service processes on the node. You must also set up a separate SPN and keytab file for the HTTP processes on node.

**Process Level**

If the domain is used for production and requires a high level of security, you can set the service principal at the process level. Create a unique SPN and keytab file for each node and each process on the node. You must also set up a separate SPN and keytab file for the HTTP processes on node.

The Informatica domain requires the service principal and keytab file names to follow a specific format. To ensure that you follow the correct format for the service principal and keytab file names, use the Informatica Kerberos SPN Format Generator to generate a list of the service principal and keytab file names in the format required by the Informatica domain.

The Informatica Kerberos SPN Format Generator is shipped with the Informatica services installer.

**Service Principal Requirements at Node Level**

If the Informatica domain does not require a high level of security, the node and service processes can share the same SPNs and keytab files. The domain does not require a separate SPN for each service process in a node.

The Informatica domain requires SPNs and keytab files for the following components at node level:

- **Principal distinguished name (DN) for the LDAP directory service**
  
  Principal name for the bind user DN that is used to search the LDAP directory service. The name of the keytab file must be `infaldapuser.keytab`.

- **Node process**
  
  Principal name for the Informatica node that initiates or accepts authentication calls. The same principal name is used to authenticate the services in the node. Each gateway node in the domain requires a separate principal name.

- **HTTP processes in the domain**
  
  Principal name for all web application services in the Informatica domain, including Informatica Administrator. The browser uses this principal name to authenticate with all HTTP processes in the domain. The name of the keytab file must be `webapp_http.keytab`.

**Service Principal Requirements at Process Level**

If the Informatica domain requires a high level of security, create a separate SPN and keytab file for each node and each service in the node.

The Informatica domain requires SPNs and keytab files for the following components at process level:
Principal distinguished name (DN) for the LDAP directory service

Principal name for the bind user DN that is used to search the LDAP directory service. The name of the keytab file must be infa_ldapuser.keytab.

Node process

Principal name for the Informatica node that initiates or accepts authentication calls.

Informatica Administrator service

Principal name for the Informatica Administrator service that authenticates the service with other services in the Informatica domain. The name of the keytab file must be AdminConsole.keytab.

HTTP processes in the domain

Principal name for all web application services in the Informatica domain, including Informatica Administrator. The browser uses this principal name to authenticate with all HTTP processes in the domain. The name of the keytab file must be webapp_http.keytab.

Service process

Principal name for the service that runs on a node in the Informatica domain. Each service requires a unique service principal and keytab file name.

You do not need to create the SPNs and keytab files for the services before you run the installer. You can create the SPN and keytab file for a service when you create the service in the domain. The SPN and keytab file for a service must be available when you enable the service.

Running the Informatica Kerberos SPN Format Generator on Windows

You can run the Informatica Kerberos SPN Format Generator to generate a file that shows the correct format for the SPNs and keytab file names required in the Informatica domain.

You can run the SPN Format Generator from the command line or from the Informatica installer. The SPN Format Generator generates a file with the names of the service principal and keytab files based on the parameters you provide.

Note: Verify that the information you provide is correct. The SPN Format Generator does not validate the values you enter.

To run the SPN Format Generator, perform the following steps:

1. Log in to the machine with a system user account.
2. Close all other applications.
3. On the machine where you extracted the installation files, go to the following directory: <Informatica installation files directory>/Server/Kerberos
4. Run the SPNFormatGenerator.bat file.
   The Informatica Kerberos SPN Format Generator Welcome page appears.
5. Click Next.
   The Service Principal Level page appears.
6. Select the level at which to set the Kerberos service principals for the domain.
The following table describes the levels you can select:

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process Level</td>
<td>Configures the domain to use a unique service principal name (SPN) and keytab file for each node and each application service on a node. The number of SPNs and keytab files required for each node depends on the number of application service processes that run on the node. Use the process level option for domains that require a high level of security, such as production domains.</td>
</tr>
<tr>
<td>Node Level</td>
<td>Configures the domain to share SPNs and keytab files on a node. This option requires one SPN and keytab file for the node and all application services that run on the node. It also requires a separate SPN and keytab file for all HTTP processes on the node. Use the node level option for domains that do not require a high level of security, such as test and development domains.</td>
</tr>
</tbody>
</table>

7. Click **Next**.

   The **Authentication Parameters - Kerberos Authentication** page appears.

8. Enter the domain and node parameters to generate the SPN format.

   The following table describes the parameters you must specify:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Name</td>
<td>Name of the domain. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; &quot; ? , &lt; &gt; /</td>
</tr>
<tr>
<td>Node name</td>
<td>Name of the Informatica node.</td>
</tr>
<tr>
<td>Node host name</td>
<td>Fully qualified host name or the IP address of the machine on which to create the node. The node host name cannot contain the underscore (_) character. <strong>Note:</strong> Do not use <code>localhost</code>. The host name must explicitly identify the machine.</td>
</tr>
<tr>
<td>Service Realm Name</td>
<td>Name of the Kerberos realm for the Informatica domain services. The realm name must be in uppercase.</td>
</tr>
</tbody>
</table>

   If you set the service principal at node level, the utility displays the **+Node** button. If you set the service principal at process level, the utility displays the **+Node** and **+Service** buttons.

9. To generate the SPN format for an additional node, click **+Node** and specify the node name and host name.

10. To generate the SPN format for a service, click **+Service** and specify the service name in the **Service On Node** field.

    The **Service On Node** field displays only if you set the service principal at process level and you click **+Service**. You can enter multiple services for a node. The services appear immediately below the node that they run on.

11. To remove a node from the list, click **-Node**.

    The Informatica SPN Format Generator deletes the node. If you have added services to the node, the services are deleted with the node.
12. To remove a service from a node, clear the service name field.

13. Click **Next**.

The SPN Format Generator displays the path and file name of the file that contains the list of service principal and keytab file names.

14. Click **Done** to exit the SPN Format Generator.

The SPN Format Generator generates a text file that contains the SPN and keytab file names in the format required for the Informatica domain.

### Running the Informatica SPN Format Generator on UNIX

You can run the Informatica Kerberos SPN Format Generator to generate a file that shows the correct format for the SPNs and keytab file names required in the Informatica domain.

You can run the SPN Format Generator from the command line or from the Informatica installer. The SPN Format Generator generates a file with the names of the service principal and keytab files based on the parameters you provide.

**Note:** Verify that the information you provide is correct. The SPN Format Generator does not validate the values you enter.

1. On the machine where you extracted the installation files, go to the following directory: `<Informatica installation files directory>/Server/Kerberos`

2. On a shell command line, run the `SPNFormatGenerator.sh` file.

3. Press **Enter** to continue.

4. In the **Service Principal Level** section, select the level at which to set the Kerberos service principals for the domain.

   The following table describes the levels you can select:

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-&gt;Process Level</td>
<td>Configures the domain to use a unique service principal name (SPN) and keytab file for each node and each application service on a node. The number of SPNs and keytab files required for each node depends on the number of application service processes that run on the node. Use the process level option for domains that require a high level of security, such as productions domains.</td>
</tr>
<tr>
<td>2-&gt;Node Level</td>
<td>Configures the domain to share SPNs and keytab files on a node. This option requires one SPN and keytab file for the node and all application services that run on the node. It also requires a separate SPN and keytab file for all HTTP processes on the node. Use the node level option for domains that do not require a high level of security, such as test and development domains.</td>
</tr>
</tbody>
</table>

5. Enter the domain and node parameters required to generate the SPN format.
The following table describes the parameters you must specify:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain Name</td>
<td>Name of the domain. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; &quot; ? , &lt; &gt; \ /</td>
</tr>
<tr>
<td>Node name</td>
<td>Name of the Informatica node.</td>
</tr>
</tbody>
</table>
| Node host name    | Fully qualified host name or the IP address of the machine on which to create the node. The node host name cannot contain the underscore (_) character.  
  **Note:** Do not use `localhost`. The host name must explicitly identify the machine. |
| Service Realm Name| Name of the Kerberos realm for the Informatica domain services. The realm name must be in uppercase.                                    |

If you set the service principal at node level, the prompt **Add Node?** appears. If you set the service principal at process level, the prompt **Add Service?** appears.

6. At the **Add Node?** prompt, enter 1 to generate the SPN format for an additional node. Then enter the node name and node host name.
   
   To generate the SPN formats for multiple nodes, enter 1 at each **Add Node?** prompt and enter a node name and node host name.

7. At the **Add Service?** prompt, enter 1 to generate the SPN format for a service that will run on the preceding node. Then enter the service name.
   
   To generate the SPN formats for multiple services, enter 1 at each **Add Service?** prompt and enter a service name.

8. Enter 2 to end the **Add Service?** or **Add Node?** prompts.
   
   The SPN Format Generator displays the path and file name of the file that contains the list of service principal and keytab file names.

9. Press Enter to exit the SPN Format Generator.
   
   The SPN Format Generator generates a text file that contains the SPN and keytab file names in the format required for the Informatica domain.

### Review the SPN and Keytab Format Text File

The Kerberos SPN Format Generator generates a text file named `SPNKeytabFormat.txt` that lists the format for the service principal and keytab file names required by the Informatica domain. The list includes the SPN and keytab file names based on the service principal level you select.

Review the text file and verify that there are no error messages.

The text file contains the following information:

**Entity Name**

Identifies the node or service associated with the process.
**SPN**

Format for the SPN in the Kerberos principal database. The SPN is case sensitive. Each type of SPN has a different format.

An SPN can have one of the following formats:

<table>
<thead>
<tr>
<th>Keytab type</th>
<th>SPN Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>NODE_SPN</td>
<td>isp/&lt;NodeName&gt;/&lt;DomainName&gt;@&lt;REALMNAME&gt;</td>
</tr>
<tr>
<td>NODE_AC_SPN</td>
<td>_AdminConsole/&lt;NodeName&gt;/&lt;DomainName&gt;@&lt;REALMNAME&gt;</td>
</tr>
<tr>
<td>NODE_HTTP_SPN</td>
<td>HTTP/&lt;NodeHostName&gt;@&lt;REALMNAME&gt;</td>
</tr>
<tr>
<td>SERVICE_PROCESS_SPN</td>
<td>&lt;ServiceName&gt;/&lt;NodeName&gt;/&lt;DomainName&gt;@&lt;REALMNAME&gt;</td>
</tr>
</tbody>
</table>

**Note:** The Kerberos SPN Format Generator validates the node host name. If the node host name is not valid, the utility does not generate an SPN. Instead, it displays the following message: Unable to resolve host name.

**Keytab File Name**

Format for the name of the keytab file to be created for the associated SPN in the Kerberos principal database. The keytab file name is case sensitive.

The keytab file names use the following formats:

<table>
<thead>
<tr>
<th>Keytab type</th>
<th>Keytab File Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>NODE_SPN</td>
<td>&lt;NodeName&gt;.keytab</td>
</tr>
<tr>
<td>NODE_AC_SPN</td>
<td>_AdminConsole.keytab</td>
</tr>
<tr>
<td>NODE_HTTP_SPN</td>
<td>webapp_http.keytab</td>
</tr>
<tr>
<td>SERVICE_PROCESS_SPN</td>
<td>&lt;ServiceName&gt;.keytab</td>
</tr>
</tbody>
</table>

**Keytab Type**

Type of the keytab. The keytab type can be one of the following types:

- NODE_SPN. Keytab file for a node process.
- NODE_AC_SPN. Keytab file for the Informatica Administrator service process.
- NODE_HTTP_SPN. Keytab file for HTTP processes in a node.
- SERVICE_PROCESS_SPN. Keytab file for a service process.

**Service Principals at Node Level**

The following example shows the contents of the SPNKeytabFormat.txt file generated for service principals at the node level:

<table>
<thead>
<tr>
<th>ENTITY_NAME</th>
<th>SPN</th>
<th>KEY_TAB_NAME</th>
</tr>
</thead>
<tbody>
<tr>
<td>Node01</td>
<td>isp/Node01/InfDomain@MY.SVCREALM.COM</td>
<td>Node01.keytab</td>
</tr>
<tr>
<td>NODE_SPN</td>
<td>HTTP/NodeHost01.enterprise.com@MY.SVCREALM.COM</td>
<td>webapp_http.keytab</td>
</tr>
<tr>
<td>NODE_HTTP_SPN</td>
<td>isp/Node02/InfDomain@MY.SVCREALM.COM</td>
<td>Node02.keytab</td>
</tr>
</tbody>
</table>
Create the Service Principal Names and Keytab Files

After you generate the list of SPN and keytab file names in Informatica format, send a request to the Kerberos administrator to add the SPNs to the Kerberos principal database and create the keytab files.

Use the following guidelines when you create the SPN and keytab files:

**The user principal name (UPN) must be the same as the SPN.**

When you create a user account for the service principal, you must set the UPN with the same name as the SPN. The application services in the Informatica domain can act as a service or a client depending on the operation. You must configure the service principal to be identifiable by the same UPN and SPN.

A user account must be associated with only one SPN. Do not set multiple SPNs for one user account.

**Enable delegation in Microsoft Active Directory.**

You must enable delegation for all user accounts with service principals used in the Informatica domain. In the Microsoft Active Directory Service, set the Trust this user for delegation to any service (Kerberos only) option for each user account that you set an SPN.

Delegated authentication happens when a user is authenticated with one service and that service uses the credentials of the authenticated user to connect to another service. Because services in the Informatica domain need to connect to other services to complete an operation, the Informatica domain requires the delegation option to be enabled in Microsoft Active Directory.

For example, when a PowerCenter client connects to the PowerCenter Repository Service, the client user account is authenticated with the PowerCenter Repository Service principal. When the PowerCenter
Troubleshooting the SPN and Keytab Files

You can use Kerberos utilities to verify that the service principal and keytab file names created by the Kerberos administrator match the service principal and keytab file names that you requested. You can also use the utilities to determine the status of the Kerberos key distribution center (KDC).

Use the ktpass utility to create the service principal keytab files.

Microsoft Active Directory supplies the ktpass utility to create keytab files. Informatica supports Kerberos authentication only on Microsoft Active Directory and has certified only keytab files that are created with ktpass.

The keytab files for a node must be available on the machine that hosts the node. By default, the keytab files are stored in the following directory: {Informatica installation directory}/isp/config/keys. During installation, you can specify a directory on the node to store the keytab files.

When you receive the keytab files from the Kerberos administrator, copy the keytab files to a directory that is accessible to the machine where you plan to install the Informatica services. When you run the Informatica installer, specify the location of the keytab files. The Informatica installer copies the keytab files to the directory for keytab files on the Informatica node.

Use the following utilities to verify the SPNs and keytab files:

klist

You can use klist to list the Kerberos principals and keys in a keytab file. To list the keys in the keytab file and the time stamp for the keytab entry, run the following command:

```
  klist -k -t <keytab_file>
```

The following output example shows the principals in a keytab file:

```
Keytab name: FILE:int_svc01.keytab
KVNO Timestamp Principal
----- --------------- -------------------------------
 3 12/31/69 19:00:00 int_svc01/node01_vMPE/Domain96_vMPE@REALM
 3 12/31/69 19:00:00 int_svc01/node01_vMPE/Domain96_vMPE@REALM
 3 12/31/69 19:00:00 int_svc01/node01_vMPE/Domain96_vMPE@REALM
 3 12/31/69 19:00:00 int_svc01/node01_vMPE/Domain96_vMPE@REALM
 3 12/31/69 19:00:00 int_svc01/node01_vMPE/Domain96_vMPE@REALM
```

kinit

You can use kinit to request a ticket-granting ticket for a user account to verify that the KDC is running and can grant tickets. To request a ticket-granting ticket for a user account, run the following command:

```
kinit <user_account>
```
You can also use *kinit* to request a ticket-granting ticket and verify that the keytab file can be used to establish a Kerberos connection. To request a ticket-granting ticket for an SPN, run the following command:

```bash
kinit -V -k -t <keytab_file> <SPN>
```

The following output example shows the ticket-granting ticket created in the default cache for a specified keytab file and SPN:

```
Using default cache: /tmp/krb5cc_10000073
Using principal: int_srvc01/node01_vMPE/Domn96_vMPE@REALM
Using keytab: int_srvc01.keytab
Authenticated to Kerberos v5
```

**setspn**

You can use *setspn* to view, modify, or delete the SPN of an Active Directory service account. On the machine that hosts the Active Directory service, open a command line window and run the command.

To view the SPNs that are associated with a user account, run the following command:

```bash
setspn -L <user_account>
```

The following output example shows the SPN associated with the user account *is96svc*:

```
Registered ServicePrincipalNames for CN=is96svc,OU=AllSvcAccts,OU=People, DC=ds,DC=intrac0rp,DC=ze0rp:
 int_srvc01/node02_vMPE/Domn96_vMPE
```

To view the user accounts associated with an SPN, run the following command:

```bash
setspn -O <SPN>
```

The following output example shows the user account associated with the SPN *int_srvc01/node02_vMPE/Domn96_vMPE*:

```
Checking domain DC=ds,DC=intrac0rp,DC=ze0rp
CN=is96svc,OU=AllSvcAccts,OU=People,DC=ds,DC=intrac0rp,DC=ze0rp
 int_srvc01/node02_vMPE/Domn96_vMPE

Existing SPN found!
```

To search for duplicate SPNs, run the following command:

```bash
setspn -X
```

The following output example shows multiple user accounts associated with one SPN:

```
Checking domain DC=ds,DC=intrac0rp,DC=ze0rp
Processing entry 1125
HOST/mtb01.REALM is registered on these accounts:
 CN=Team1svc,OU=AllSvcAccts,OU=People,DC=ds,DC=intrac0rp,DC=ze0rp
 CN=MTB1svc,OU=11S,OU=WPC960K3,OU=WINServers,DC=ds,DC=intrac0rp,DC=ze0rp
```

**Note:** Searching for duplicate SPNs can take a long time and a large amount of memory.

**kdestroy**

You can use *kdestroy* to delete the active Kerberos authorization tickets and the user credentials cache that contains them. If you run *kdestroy* without parameters, you delete the default credentials cache.
Part III: Secure@Source
Installation

This part contains the following chapters:

- Secure@Source Services Installation, 82
- Troubleshooting the Services Installation, 111
CHAPTER 6

Secure@Source Services Installation

This chapter includes the following topics:

- Secure@Source Services Installation Overview, 82
- Pre-installation System Check, 83
- Installing Secure@Source in Console Mode, 84
- Post-installation Tasks, 107

Secure@Source Services Installation Overview

You can install the Secure@Source services on a Linux machine and run the installer in console mode. Complete the pre-installation tasks to prepare for the installation. The installation process includes a system check to verify whether the machine that you are installing on meets the system requirements for installation.

The installation process also creates a service named Informatica that runs as a daemon on Linux. When you start the Informatica service, it starts the Service Manager, which manages all domain operations.

After the installation, use Informatica Administrator to log in to the domain and configure the application services.

Installing Secure@Source 4.1

Complete the following steps to install Secure@Source 4.1:

1. Run the Secure@Source 4.0 install.sh file to install Secure@Source 4.0.
2. Bring down the Informatica domain and services.
3. Set the INFA_HOME variable.
4. Run the Secure@Source 4.1 install.sh file.
5. Start the Informatica domain and services.
6. From the Administrator tool, disable the Secure@Source Service and upgrade the Secure@Source repository contents.
7. Enable the Secure@Source Service.
Upgrading to Secure@Source 4.1

Complete the following steps to upgrade to Secure@Source 4.1 from Secure@Source 4.0:

1. From the Administrator tool, disable the Secure@Source Service.
2. Download and extract the Secure@Source 4.1 installer.
3. Run the Secure@Source 4.1 install.sh file.
4. From the Administrator tool, upgrade the Secure@Source repository contents.
5. Enable the Secure@Source Service.

Pre-installation System Check

The Secure@Source installer includes the Pre-installation System Check utility. After you enter all the required information, the utility runs a series of checks to verify whether the Informatica domain host, Hadoop Gateway host, and Hadoop nodes meet the requirements for the installation.

Pre-installation System Check Verification

The Pre-installation System Check utility verifies that the system meets the following requirements:

- The time on the machine is maintained by the Network Time Protocol (NTP) daemon. This ensures that the time is in sync across all the nodes involved in the Secure@Source installation.
- A passwordless SSH exists between the Hadoop gateway host and Hadoop node.
- The root disk space is larger than 50 GB.
- JDK version 1.8 is installed.
- A minimum of 32 GB of RAM is available for the Informatica domain.
- A minimum of 64 GB of RAM is available for the Hadoop Gateway host and Hadoop nodes.
- A minimum of 25 GB disk space is available for the Informatica domain.
- A minimum of 50 GB disk space is available for the Hadoop Gateway host and Hadoop nodes root partition.
- A minimum of 8 cores processor for the Informatica domain.
- A minimum of 16 cores processor for the Hadoop Gateway host and Hadoop nodes.
- In Hadoop nodes, the /opt folder contains the SecureAtSource folder.
- The current user has write permission to the /opt directory.
- The kernel-headers.x86_64 package is in the list of installed packages.
- The kernel-devel.x86_64 package is in the list of installed packages.
- OpenSSL version 1.0.1e-30.el6_6.5.x86_64.rpm or later is available.
- The ports that you specified in the installer are available.
Pre-installation System Check Results

After the Pre-installation System Check utility completes the evaluation, the utility displays any failures on the installer screen. The log file for the utility shows the complete results. For each requirement, the log file shows one of the following result statuses:

- [Pass] - The requirement meets the criteria for the installation.
- [Fail] - The requirement does not meet the criteria for the installation. You can proceed with the installation. However, Informatica recommends that you first resolve the issue and then continue with the installation. To resolve the issue before installing Secure@Source, stop the installer, resolve the issue, and then run the installer again.
- [Information] - Verify the information and perform any additional tasks as outlined in the details.

Pre-installation System Check Log Files

You can view the results of the Pre-installation System Check on the console or in the following log files:

- %110Pi_server_summary.log<timestamp>. Contains the results for the Informatica domain host. The log file is located in the folder: %<Installer extracted location>/logs
- %110Pi_results_<Hadoop host/node name>.log<timestamp>. Contains the results for the Hadoop Gateway host and nodes. The log file is located in the folder: %<Installer extracted location>/logs/I10Pi_NodeResults

Installing Secure@Source in Console Mode

You can install the Secure@Source services in console mode on Linux.

When you run the installer in console mode, the words Quit and Back are reserved words. Do not use them as input text.

Secure Directory for the Encryption Key and Configuration Files

When you install or upgrade Informatica, the installer creates directories to store Informatica files that require restricted access, such as the domain encryption key file and the nodemeta.xml. On Linux, the installer assigns different permissions for the directories and the files in the directories.

By default, the installer creates the following directories within the Informatica installation directory:

- %Informatica installation directory%/isp/config
  Contains the nodemeta.xml file. Also contains the /keys directory where the encryption key file is stored. You can specify a different directory in which to store the files. The installer assigns the same permissions to the specified directory as the default directory.

- %Informatica installation directory%/services/shared/security
  This directory is not used by Secure@Source.

The installer assigns the following permissions to the directories and the files in the directories:
Directory Permissions

The owner of the directory has -wx permissions to the directory but no r permission. The owner of the directory is the user account used to run the installer. The group to which the owner belongs also has -wx permissions to the directory but no r permission.

For example, the user account ediqa owns the directory and belongs to the infaadmin group. The ediqa user account and the infaadmin group have the following permissions: -wx-wx---

The ediqa user account and the infaadmin group can write to and run files in the directory. They cannot display the list of files in directory but they can list a specific file by name.

If you know the name of a file in the directory, you can copy the file from the directory to another location. If you do not know the name of the file, you must change the permission for the directory to include the read permission before you can copy the file. You can use the command chmod 730 to give read permission to the owner of the directory and subdirectories.

For example, you need to copy the encryption key file named siteKey to a temporary directory to make it accessible to another node in the domain. Run the command chmod 730 on the <Informatica installation directory>/isp/config directory to assign the following permissions: rwx-wx---. You can then copy the encryption key file from the /keys subdirectory to another directory.

After you complete copying the files, change the permissions for the directory back to write and execute permissions. You can use the command chmod 330 to remove the read permission.

Note: Do not use the -R option to recursively change the permissions for the directory and files. The directory and the files in the directory have different permissions.

File Permissions

The owner of the files in the directory has rwx permissions to the files. The owner of the files in the directory is the user account used to run the installer. The group to which the owner belongs also has rwx permissions to the files in the directory.

The owner and group have full access to the file and can display or edit the file in the directory.

Note: You must know the name of the file to be able to list or edit the file.

Installing by Creating a Domain

Run the installer to create the domain and configure the services.

1. Log in to the machine with the system user account.
2. Close all other applications.
3. On a shell command line, run the install.sh file from the installation directory.
The Welcome to the Informatica Secure@Source 4.0 Installer panel appears.

Before you continue, read the following documents:
* Informatica Secure@Source 4.0 Installation and Configuration Guide
* Informatica Secure@Source 4.0 Release Notes

Configure the LANG and LC_ALL variables to generate appropriate code pages and to create and connect to repositories and Repository Services.
Do you want to continue? (Y/N)Y

Current operating system meets minimum requirements.

Checking for existing 4.0 product installation.
Preparing to install...
Extracting the JRE from the installer archive...
Unpacking the JRE...
Extracting the installation resources from the installer archive...
Configuring the installer for this system's environment...

4. If the environment variables are not set, press N to exit the installer and set them as required.
   If the environment variables are set, press Y to continue.

The Installation Type panel appears.

5. Read the terms and conditions for Java SE Development Kit Software. Press 2 to accept the agreement.

6. Read the Informatica End User License Agreement. Press 2 to accept the agreement.

7. Press 1 to install Secure@Source.

The Installation Prerequisites - Step 2 of 11 panel appears.

8. Review the list of prerequisites and press Enter to continue.

The License and Installation Directory - Step 3 of 11 panel appears.

9. Type the full path and file name of the Informatica license key and press Enter.

   Default is /export/ilmsatsqe/license.key.

10. Type the full path for the installation directory.
    The directory names in the path must not contain spaces or the following special characters: ! @ # $ % * () [] {} , ; '

    Default is /export/ilmsatsqe/Informatica/SATS4.0.

    Note: Informatica recommends using alphanumeric characters in the installation directory path. If you use a special character such as â or €, unexpected results might occur at run time.

The Domain Selection - Step 4A of 11 panel appears.

11. Press 1 to create a domain.
    When you create a domain, the node that you create becomes a gateway node in the domain. The gateway node contains a Service Manager that manages all domain operations.

12. Enter the port number for the Informatica Administrator to use.

13. Select whether to use a keystore file generated by the installer or a keystore file that you create. You can use a keystore file with a self-signed certificate or a certificate signed by a certification authority.
   a. Press 1 to use a keystore file generated by the installer.
If you select to use a keystore file generated by the installer, the installer creates a self-signed keystore file named `default.keystore` in the following location:

```<Informatica installation directory>/tomcat/conf/```

b. Press 2 to specify a keystore file and password.

If you specify the keystore, enter the password and location of the keystore file.

14. Specify if you want to enable SAML-based single sign-on.
   - Press 1 if you do not want to enable SAML-based single sign-on.
   - Press 2 if you want to enable SAML-based single sign-on.

The **Domain Security - Secure Communication - Step 4B of 11** panel appears.

15. In the Domain Security - Secure Communication section, specify whether to use the default Informatica SSL certificates or to use your SSL certificates to secure domain communication.
   a. Select the type of SSL certificates to use.

   The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - Use the default Informatica SSL certificate files</td>
<td>Use the default SSL certificates provided by Informatica. <strong>Note:</strong> If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files.</td>
</tr>
<tr>
<td>2 - Specify the location of the SSL certificate files</td>
<td>Use SSL certificates that you provide. You must specify the location of the keystore and truststore files. You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain.</td>
</tr>
</tbody>
</table>

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.

   The following table describes the parameters that you must enter for the SSL certificate files:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystore file directory</td>
<td>Directory that contains the keystore files. The directory must contain files named infa_keystore.jks and infa_keystore.pem.</td>
</tr>
<tr>
<td>Keystore password</td>
<td>Password for the keystore infa_keystore.jks.</td>
</tr>
</tbody>
</table>
The Domain Configuration Repository section appears.

16. Select the database to use for the domain configuration repository.

The following table lists the databases you can use for the domain configuration repository:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database type</td>
<td>Type of database for the domain configuration repository. Select from the following options: 1 - Oracle 2 - Microsoft SQL Server 3 - IBM DB2 4 - Sybase ASE</td>
</tr>
</tbody>
</table>

The Informatica domain configuration repository stores metadata for domain operations and user authentication. The domain configuration repository must be accessible to all gateway nodes in the domain.

17. Enter the properties for the database user account.

The following table lists the properties for the database user account:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the domain configuration database user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the domain configuration database user account.</td>
</tr>
</tbody>
</table>

18. Enter the parameters for the database.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.
The following table describes the properties that you must configure for the IBM DB2 database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
</table>
| Configure tablespace | Select whether to specify a tablespace:  
|                 | 1 - No  
|                 | 2 - Yes  
|                 | In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes. |
| Tablespace      | Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes.  
|                 | In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables.  
|                 | In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database. |

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema name</td>
<td>Name of the schema that will contain domain configuration tables. If this parameter is blank, the installer creates the tables in the default schema.</td>
</tr>
</tbody>
</table>

19. Choose whether to create a secure domain configuration repository.

You can create a domain configuration repository in a database secured with the SSL protocol. To create a domain configuration repository in a secure database, press 1.

To create a domain configuration repository in an unsecure database, press 2.

20. Specify the JDBC connection.

a. To enter the JDBC connection information using the JDBC URL information, press 1. To enter the JDBC connection information using a custom JDBC connection string, press 2.

b. Enter the JDBC connection information.

   • To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
The following table describes the JDBC connection information:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: <code>&lt;host name&gt;:&lt;port number&gt;</code> For example: <code>abc123@company.com:50000</code></td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft SQL Server and Sybase ASE.</td>
</tr>
<tr>
<td>Configure JDBC Parameters</td>
<td>Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.</td>
</tr>
</tbody>
</table>

- To enter the connection information using a custom JDBC connection string, type the connection string.
  Use the following syntax for the JDBC connection string for the databases:
  **IBM DB2**
  ```
  jdbc:Informatica:db2://host_name:port_no;DatabaseName=
  ```
  **Oracle**
  ```
  jdbc:Informatica:oracle://host_name:port_no;ServiceName=
  ```
  **Microsoft SQL Server**
  ```
  jdbc:Informatica:sqiserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
  ```
  **Sybase**
  ```
  jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
  ```
  Verify that the connection string contains all the connection parameters required by your database system.

21. If you create a secure domain configuration repository, enter the parameters for the secure database.
    If you create the domain configuration repository on a secure database, you must provide the truststore information for the database. You must also provide a JDBC connection string that includes the security parameters for the database.
The following table describes the options available to create a secure domain configuration repository database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database truststore file</td>
<td>Path and file name of the truststore file for the secure database.</td>
</tr>
<tr>
<td>Database truststore password</td>
<td>Password for the truststore file.</td>
</tr>
<tr>
<td>Custom JDBC Connection String</td>
<td>Complete JDBC connection for the secure database, including the host name and port number and the secure database parameters.</td>
</tr>
</tbody>
</table>

In addition to the host name and port number for the database server, you must include the following secure database parameters:

**EncryptionMethod**

- Required. Indicates whether data is encrypted when transmitted over the network. This parameter must be set to SSL.

**ValidateServerCertificate**

- Optional. Indicates whether Informatica validates the certificate that the database server sends.
  - If this parameter is set to True, Informatica validates the certificate that the database server sends.
  - If you specify the HostNameInCertificate parameter, Informatica also validates the host name in the certificate.
  - If this parameter is set to False, Informatica does not validate the certificate that the database server sends. Informatica ignores any truststore information that you specify.
  - Default is True.

**HostNameInCertificate**

- Optional. Host name of the machine that hosts the secure database. If you specify a host name, Informatica validates the host name included in the connection string against the host name in the SSL certificate.

**cryptoProtocolVersion**

- Required. Specifies the cryptographic protocol to use to connect to a secure database. You can set the parameter to `cryptoProtocolVersion=TLSv1.1` or `cryptoProtocolVersion=TLSv1.2` based on the cryptographic protocol used by the database server.

You can use the following syntax for the connection strings:

- **Oracle**: `jdbc:Informatica:oracle://host_name:port_no;ServiceName=service_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

- **IBM DB2**: `jdbc:Informatica:db2://host_name:port_no;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

- **Microsoft SQL Server**: `jdbc:Informatica:sqlserver://host_name:port_no;SelectMethod=cursor;DatabaseName=database_name;EncryptionMethod=SSL;HostNameInCertificate=DB_host_name;ValidateServerCertificate=true_or_false`

**Note**: The installer does not validate the connection string. Verify that the connection string contains all the connection parameters and security parameters required by your database.
22. If the database contains a domain configuration repository for a previous domain, choose to overwrite the data or set up another database.

The following table describes the options of overwriting the data or setting up another database when you create a domain configuration repository for a previous domain:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 - OK</td>
<td>Enter the connection information for a new database.</td>
</tr>
<tr>
<td>2 - Continue</td>
<td>The installer overwrites the data in the database with new domain configuration.</td>
</tr>
</tbody>
</table>

The Model Repository Service Database - Step 5A of 11 panel appears.

23. Select the database to use for the Model Repository Service database.

The following table lists the databases that you can use for the Model Repository Service database:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database type</td>
<td>Type of database for the Model Repository Service database. Select from the following options:</td>
</tr>
<tr>
<td>1 - Oracle</td>
<td>2 - Microsoft SQL Server</td>
</tr>
<tr>
<td>3 - IBM DB2</td>
<td></td>
</tr>
</tbody>
</table>

24. Enter the properties for the database user account.

The following table lists the properties for the database user account:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the Model Repository Service database user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the Model Repository Service database user account.</td>
</tr>
</tbody>
</table>

25. Specify the tablespace properties.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.
The following table describes the properties that you must configure for the IBM DB2 database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configure tablespace</td>
<td>Select whether to specify a tablespace:</td>
</tr>
<tr>
<td></td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td></td>
<td>In a single-partition database, if you select No, the installer creates the tables</td>
</tr>
<tr>
<td></td>
<td>in the default tablespace. In a multi-partition database, you must select Yes.</td>
</tr>
<tr>
<td>Tablespace</td>
<td>Name of the tablespace in which to create the tables. Specify a tablespace that meets the page size requirement of 32768 bytes.</td>
</tr>
<tr>
<td></td>
<td>In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables.</td>
</tr>
<tr>
<td></td>
<td>In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.</td>
</tr>
</tbody>
</table>

b. If you select Microsoft SQL Server, enter the schema name for the database.

The following table describes the properties that you must configure for the Microsoft SQL Server database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema name</td>
<td>Name of the schema that will contain Model Repository Service database tables. If this parameter is blank, the installer creates the tables in the default schema.</td>
</tr>
</tbody>
</table>

c. Choose whether to create a secure domain configuration repository. You can create a domain configuration repository in a database secured with the SSL protocol. To create a domain configuration repository in a secure database, press 1. To create a domain configuration repository in an unsecure database, press 2.

d. To enter the JDBC connection information using the JDBC URL information, press 1. To enter the JDBC connection information using a custom JDBC connection string, press 2.

e. Enter the JDBC connection information.
   - To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
The following table describes the database connection information:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: &lt;host name&gt;:&lt;port number&gt;</td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft SQL Server and Sybase ASE.</td>
</tr>
</tbody>
</table>
| Configure JDBC Parameters     | Select whether to add additional JDBC parameters to the connection string: 1 · Yes 2 · No  
  If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters. |

- To enter the connection information using a custom JDBC connection string, type the connection string.

Use the following syntax for the JDBC connection string for the databases:

**IBM DB2**

```
jdbc:Informatica:db2://host_name:port_no;DatabaseName=
```

**Oracle**

```
jdbc:Informatica:oracle://host_name:port_no;ServiceName=
```

**Microsoft SQL Server**

```
jdbc:Informatica:sqlserver://
  host_name:port_no;SelectMethod=cursor;DatabaseName=
```

**Sybase**

```
jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
```

Verify that the connection string contains all the connection parameters required by your database system.

The Profiling Warehouse Connection Database panel appears.

26. Select one of the following database types to use for the Profiling Warehouse Connection database:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
</table>
| Database type| Type of database for the Profiling Warehouse Connection database. Select from the following options:  
  1 · Oracle  
  2 · Microsoft SQL Server  
  3 · IBM DB2 |

27. Enter the following properties for the database user account:
28. Specify the tablespace properties.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.
      The following table describes the properties that you must configure for the IBM DB2 database:

      | Property          | Description                                                                                                                                 |
      |-------------------|-------------------------------------------------------------------------------------------------------------------------------------------|
      | Configure tablespace | Select whether to specify a tablespace: 1 - No 2 - Yes  
In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes. |
      | Tablespace         | Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes.  
In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables.  
In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database. |

   b. If you select Microsoft SQL Server, enter the schema name for the database.
      The following table describes the properties that you must configure for the Microsoft SQL Server database:

      | Property | Description                                                                                                                                 |
      |----------|-------------------------------------------------------------------------------------------------------------------------------------------|
      | Schema name | Name of the schema that will contain Profiling Warehouse Connection database tables. If this parameter is blank, the installer creates the tables in the default schema. |

   c. To enter the JDBC connection information using the JDBC URL information, press 1. To enter the JDBC connection information using a custom JDBC connection string, press 2.
   d. Enter the JDBC connection information.
      • To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
The following table describes the database connection information:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: &lt;host name&gt;:&lt;port number&gt;</td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft SQL Server and Sybase ASE.</td>
</tr>
<tr>
<td>Configure JDBC Parameters</td>
<td>Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.</td>
</tr>
</tbody>
</table>

- To enter the connection information using a custom JDBC connection string, type the connection string.

Use the following syntax for the JDBC connection string for the databases:

**IBM DB2**

```
jdbc:Informatica:db2://host_name:port_no;DatabaseName=...
```

**Oracle**

```
jdbc:Informatica:oracle://host_name:port_no;ServiceName=...
```

**Microsoft SQL Server**

```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=...
```

**Sybase**

```
jdbc:Informatica:sybase://host_name:port_no;DatabaseName=...
```

Verify that the connection string contains all the connection parameters required by your database system.

29. Enter the value for the **Data Access Connect String** property. The value cannot exceed 32 characters. You can find the value in the following location for each database:
- IBM DB2. See the Catalog entry.
- Microsoft SQL Server. See the odbc.ini file.
- Oracle. See the tnsnames.ora file.

The **Content Management Service Parameters and Database** panel appears.

30. Enter the following Content Management Service configuration parameters:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Management Service name</td>
<td>Name of the Content Management Service.</td>
</tr>
<tr>
<td>HTTPS Port</td>
<td>HTTPS port number to use for the Content Management Service.</td>
</tr>
</tbody>
</table>
31. Select one of the following database types to use for the Content Management Service staging database:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database type</td>
<td>Type of database for the Content Management Service staging database. Select from the following options:</td>
</tr>
<tr>
<td></td>
<td>1 - Oracle</td>
</tr>
<tr>
<td></td>
<td>2 - Microsoft SQL Server</td>
</tr>
<tr>
<td></td>
<td>3 - IBM DB2</td>
</tr>
</tbody>
</table>

32. Enter the following properties for the database user account:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the Content Management Service staging database user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the Content Management Service staging database user account.</td>
</tr>
</tbody>
</table>

33. Specify the tablespace properties.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.
      The following table describes the properties that you must configure for the IBM DB2 database:

      | Property       | Description                                                                 |
      |----------------|-----------------------------------------------------------------------------|
      | Configure tablespace | Select whether to specify a tablespace:                                       |
      |                 | 1 - No                                                                      |
      |                 | 2 - Yes                                                                     |
      |                 | In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes. |
      | Tablespace      | Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes. |
      |                 | In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables. |
      |                 | In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database. |

   b. If you select Microsoft SQL Server, enter the schema name for the database.
      The following table describes the properties that you must configure for the Microsoft SQL Server database:

      | Property       | Description                                                                 |
      |----------------|-----------------------------------------------------------------------------|
      | Schema name    | Name of the schema that will contain Content Management Service staging database tables. If this parameter is blank, the installer creates the tables in the default schema. |
c. To enter the JDBC connection information using the JDBC URL information, press 1. To enter the JDBC connection information using a custom JDBC connection string, press 2.

d. Enter the JDBC connection information.
   • To enter the connection information using the JDBC URL information, specify the JDBC URL properties.
   The following table describes the database connection information:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: &lt;host name&gt;:&lt;port number&gt;</td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft SQL Server and Sybase ASE.</td>
</tr>
<tr>
<td>Configure JDBC Parameters</td>
<td>Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.</td>
</tr>
</tbody>
</table>

   • To enter the connection information using a custom JDBC connection string, type the connection string.
   Use the following syntax for the JDBC connection string for the databases:
   
   IBM DB2
   
   jdbc:Informatica:db2://host_name:port_no;DatabaseName=
   
   Oracle
   
   jdbc:Informatica:oracle://host_name:port_no;ServiceName=
   
   Microsoft SQL Server
   
   jdbc:Informatica:sqlserver://
   host_name:port_no;SelectMethod=cursor;DatabaseName=
   
   Sybase
   
   jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
   
   Verify that the connection string contains all the connection parameters required by your database system.

34. Enter the value for the Data Access Connect String property. The value cannot exceed 32 characters. You can find the value in the following location for each database:
   • IBM DB2. See the Catalog entry.
   • Microsoft SQL Server. See the odbc.ini file.
   • Oracle. See the tnsnames.ora file.

   The Secure@Source Parameters and Database panel appears.
35. Enter the following Secure@Source Service configuration parameters:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secure@Source Service name</td>
<td>Name of the Secure@Source Service.</td>
</tr>
<tr>
<td>HTTPS Port</td>
<td>HTTPS port number to use for the Secure@Source Service.</td>
</tr>
</tbody>
</table>

36. Select one of the following database types to use for the Secure@Source repository:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
</table>
| Database type           | Type of database for the Secure@Source repository. Select from the following options:
|                         | 1 - Oracle                                                                  |
|                         | 2 - Microsoft SQL Server                                                     |
|                         | 3 - IBM DB2                                                                 |

37. Enter the following properties for the database user account:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the Secure@Source repository user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the Secure@Source repository user account.</td>
</tr>
</tbody>
</table>

38. If you do not want to create a secure Secure@Source repository, enter the parameters for the database.
   a. If you select IBM DB2, select whether to configure a tablespace and enter the tablespace name.
      The following table describes the properties that you must configure for the IBM DB2 database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configure tablespace</td>
<td>Select whether to specify a tablespace:</td>
</tr>
<tr>
<td></td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td></td>
<td>In a single-partition database, if you select No, the installer creates the tables in the default tablespace. In a multi-partition database, you must select Yes.</td>
</tr>
<tr>
<td>Tablespace</td>
<td>Name of the tablespace in which to create the tables. Specify a tablespace that meets the pageSize requirement of 32768 bytes.</td>
</tr>
<tr>
<td></td>
<td>In a single-partition database, if you select Yes to configure the tablespace, enter the name of the tablespace in which to create the tables.</td>
</tr>
<tr>
<td></td>
<td>In a multi-partition database, specify the name of the tablespace that resides in the catalog partition of the database.</td>
</tr>
</tbody>
</table>

   b. If you select Microsoft SQL Server, enter the schema name for the database.
The following table describes the properties that you must configure for the Microsoft SQL Server database:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schema name</td>
<td>Name of the schema that will contain Secure@Source repository tables. If this parameter is blank, the installer creates the tables in the default schema.</td>
</tr>
</tbody>
</table>

c. Choose whether to create a secure Secure@Source repository.

You can create a Secure@Source repository in a database secured with the SSL protocol. To create a Secure@Source repository in a secure database, press 1.

To create a Secure@Source repository in an unsecure database, press 2.

d. To enter the JDBC connection information using the JDBC URL information, press 1. To enter the JDBC connection information using a custom JDBC connection string, press 2.

e. Enter the JDBC connection information.

- To enter the connection information using the JDBC URL information, specify the JDBC URL properties.

The following table describes the database connection information:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: &lt;host name&gt;:&lt;port number&gt;</td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft Microsoft SQL Server and Sybase ASE.</td>
</tr>
<tr>
<td>Configure JDBC Parameters</td>
<td>Select whether to add additional JDBC parameters to the connection string: 1 - Yes 2 - No If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.</td>
</tr>
</tbody>
</table>
To enter the connection information using a custom JDBC connection string, type the connection string. Use the following syntax for the JDBC connection string for the databases:

**IBM DB2**
```
jdbc:Informatica:db2://host_name:port_no;DatabaseName=
```

**Oracle**
```
jdbc:Informatica:oracle://host_name:port_no;ServiceName=
```

**Microsoft SQL Server**
```
jdbc:Informatica:sqlserver://
host_name:port_no;SelectMethod=cursor;DatabaseName=
```

**Sybase**
```
jdbc:Informatica:sybase://host_name:port_no;DatabaseName=
```

Verify that the connection string contains all the connection parameters required by your database system.

39. Specify if you want to enable user activity.
   - Press 1 to enable user activity.
   - Press 2 if you do not want to enable user activity.

40. Specify if you want to install and enable Persistent Data Masking.
   - Press 1 to install and enable Persistent Data Masking.
   - Press 2 if you do not want to install and enable Persistent Data Masking.

   The PowerCenter Repository Service Parameters and Database panel appears.

41. Enter the following PowerCenter configuration parameters:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>PowerCenter Repository Service Name</td>
<td>Name of the PowerCenter Repository Service.</td>
</tr>
<tr>
<td>PowerCenter Integration Service Name</td>
<td>Name of the PowerCenter Integration Service.</td>
</tr>
</tbody>
</table>

42. Select one of the following database types to use for the PowerCenter Repository Service:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database type</td>
<td>Type of database for the PowerCenter Repository Service database. Select from the following options:</td>
</tr>
<tr>
<td>1 - Oracle</td>
<td>2 - Microsoft SQL Server</td>
</tr>
<tr>
<td>2 - Microsoft SQL Server</td>
<td>3 - IBM DB2</td>
</tr>
</tbody>
</table>

43. Enter the following properties for the database user account:
<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the PowerCenter Repository Service database user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the PowerCenter Repository Service database user account.</td>
</tr>
</tbody>
</table>

44. Enter the value for the **Data Access Connect String** property. The value cannot exceed 32 characters. You can find the value in the following location for each database:
- IBM DB2. See the Catalog entry.
- Microsoft SQL Server. See the `odbc.ini` file.
- Oracle. See the `tnsnames.ora` file.

The **Persistent Data Masking Parameters and Database** panel appears.

45. Enter the following Persistent Data Masking Service configuration parameters:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent Data Masking Service Name</td>
<td>Name of the Persistent Data Masking Service.</td>
</tr>
<tr>
<td>HTTPS Port</td>
<td>HTTPS port number to use for the Persistent Data Masking Service.</td>
</tr>
</tbody>
</table>

46. Select one of the following database types to use for the Persistent Data Masking Service database:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
</table>
| Database type | Type of database for the Persistent Data Masking Service database. Select from the following options:
- 1 - Oracle
- 2 - Microsoft SQL Server
- 3 - IBM DB2 |

47. Enter the following properties for the database user account:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database user ID</td>
<td>Name for the Persistent Data Masking Service database user account.</td>
</tr>
<tr>
<td>User password</td>
<td>Password for the Persistent Data Masking Service database user account.</td>
</tr>
</tbody>
</table>

48. Choose whether to create a secure Persistent Data Masking Service database.
- Press 1 to create a secure Persistent Data Masking Service database.
- Press 2 to create an unsecure Persistent Data Masking Service database.

49. Specify how you want to enter the JDBC connection information.
- To enter the JDBC connection information using the JDBC URL, press 1.
- To enter the JDBC connection information using the custom JDBC connection string, press 2.
50. If you enter the JDBC connection information using the JDBC URL, you must specify the following parameters:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database address</td>
<td>Host name and port number of the database in the following format: &lt;host name&gt;:&lt;port number&gt;</td>
</tr>
<tr>
<td>Database service name</td>
<td>Service name for Oracle and IBM DB2 databases or database name for Microsoft Microsoft SQL Server and Sybase ASE.</td>
</tr>
<tr>
<td>Configure JDBC Parameters</td>
<td>Select whether to add additional JDBC parameters to the connection string:</td>
</tr>
<tr>
<td></td>
<td>1: Yes</td>
</tr>
<tr>
<td></td>
<td>2: No</td>
</tr>
<tr>
<td></td>
<td>If you select Yes, enter the parameters or press Enter to accept the default. If you select No, the installer creates the JDBC connection string without parameters.</td>
</tr>
</tbody>
</table>

51. Enter the value for the **Data Access Connect String** property. The value cannot exceed 32 characters. You can find the value in the following location for each database:

- IBM DB2. See the Catalog entry.
- Microsoft SQL Server. See the `odbc.ini` file.
- Oracle. See the `tnsnames.ora` file.

The **Domain Security - Encryption Key** panel appears.

52. Enter the keyword and encryption key directory for the Informatica domain.

The following table describes the encryption key parameters that you must specify:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyword</td>
<td>Keyword to use to create a custom encryption key to secure sensitive data in the domain. The keyword must meet the following criteria:</td>
</tr>
<tr>
<td></td>
<td>- From 8 to 20 characters long</td>
</tr>
<tr>
<td></td>
<td>- Includes at least one uppercase letter</td>
</tr>
<tr>
<td></td>
<td>- Includes at least one lowercase letter</td>
</tr>
<tr>
<td></td>
<td>- Includes at least one number</td>
</tr>
<tr>
<td></td>
<td>- Does not contain spaces</td>
</tr>
<tr>
<td></td>
<td>The encryption key is created based on the keyword that you provide when you create the Informatica domain.</td>
</tr>
<tr>
<td>Encryption key directory</td>
<td>Directory in which to store the encryption key for the domain. The default location is the following directory: <code>&lt;Secure@Source installation directory&gt;/data/MD-SatS/sats/isp/config/keys</code>.</td>
</tr>
</tbody>
</table>

The installer sets different permissions to the directory and the files in the directory. Save the name of the domain, the keyword for the encryption key, and the encryption key file in a secure location. You must specify the domain name, keyword, and encryption key when you change the encryption key for the domain or move a repository to another domain.

53. Press 1.

The **Domain and Node Configuration** panel appears.

54. Enter the information for the domain and the node that you want to create.
The following table describes the properties that you set for the domain and gateway node.

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain name</td>
<td>Name of the domain to create. The default domain name is Domain_&lt;MachineName&gt;. The name must not exceed 128 characters and must be 7-bit ASCII only. It cannot contain a space or any of the following characters: ` % * + ; &quot; ? , &lt; &gt; /</td>
</tr>
<tr>
<td>Node host name</td>
<td>Host name of the machine on which to create the node. The node host name cannot contain the underscore (_) character. If the machine has a single network name, use the default host name. If the a machine has multiple network names, you can modify the default host name to use an alternate network name. Optionally, you can use the IP address. <strong>Note:</strong> Do not use localhost. The host name must explicitly identify the machine.</td>
</tr>
<tr>
<td>Node name</td>
<td>Name of the node to create on this machine. The node name is not the host name for the machine.</td>
</tr>
<tr>
<td>Node port number</td>
<td>Port number for the node. The default port number for the node is 6005. If the port number is not available on the machine, the installer displays the next available port number.</td>
</tr>
<tr>
<td>Domain user name</td>
<td>User name for the domain administrator. You can use this user name to initially log in to Informatica Administrator. Use the following guidelines:</td>
</tr>
<tr>
<td></td>
<td>- The name is not case sensitive and cannot exceed 128 characters.</td>
</tr>
<tr>
<td></td>
<td>- The name cannot include a tab, newline character, or the following special characters: % * + / ? ; &lt; &gt;</td>
</tr>
<tr>
<td></td>
<td>- The name can include an ASCII space character except for the first and last character. Other space characters are not allowed.</td>
</tr>
<tr>
<td>Domain password</td>
<td>Password for the domain administrator. The password must be more than 2 characters and must not exceed 16 characters.</td>
</tr>
<tr>
<td>Confirm password</td>
<td>Enter the password again to confirm.</td>
</tr>
</tbody>
</table>

55. Select whether to display the default ports for the domain and node components assigned by the installer.

The following table describes the advanced port configuration page:

<table>
<thead>
<tr>
<th>Prompt</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display advanced port configuration page</td>
<td>Select whether to display the port numbers for the domain and node components assigned by the installer:</td>
</tr>
<tr>
<td></td>
<td>1 - No</td>
</tr>
<tr>
<td></td>
<td>2 - Yes</td>
</tr>
<tr>
<td></td>
<td>If you select Yes, the installer displays the default port numbers assigned to the domain components. You can specify the port numbers to use for the domain and node components. You can also specify a range of port numbers to use for the service process that will run on the node. You can use the default port numbers or specify new port numbers. Verify that the port numbers you enter are not used by other applications.</td>
</tr>
</tbody>
</table>

56. If you display the port configuration page, enter new port numbers at the prompt or press **Enter** to use the default port numbers.
The following table describes the ports that you can set:

<table>
<thead>
<tr>
<th>Port</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum port number</td>
<td>Lowest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6014.</td>
</tr>
<tr>
<td>Maximum port number</td>
<td>Highest port number in the range of dynamic port numbers that can be assigned to the application service processes that run on this node. Default is 6114.</td>
</tr>
</tbody>
</table>

The **Service Parameters** panel appears.

57. Enter the Model Repository Service name.
58. Enter the Data Integration Service name.
59. Enter the port number for the Data Integration Service if you do not want to use the default value.
60. Specify whether to use the default Informatica SSL certificates or to use your SSL certificates to enable secure communication for the Data Integration Service.
   a. Select the type of SSL certificates to use.
      The following table describes the options for the SSL certificates that you can use to secure the Informatica domain:

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| 1 - Use the default Informatica SSL certificate files | Use the default SSL certificates provided by Informatica.  
   **Note:** If you do not provide an SSL certificate, Informatica uses the same default private key for all Informatica installations. If you use the default Informatica keystore and truststore files, the security of your domain could be compromised. To ensure a high level of security for the domain, select the option to specify the location of the SSL certificate files. |
| 2 - Specify the location of the SSL certificate files | Use SSL certificates that you provide. You must specify the location of the keystore and truststore files.  
   You can provide a self-signed certificate or a certificate issued by a certificate authority (CA). You must provide SSL certificates in PEM format and in Java Keystore (JKS) files. Informatica requires specific names for the SSL certificate files for the Informatica domain. You must use the same SSL certificates for all nodes in the domain. Store the truststore and keystore files in a directory accessible to all the nodes in the domain and specify the same keystore file directory and truststore file directory for all nodes in the same domain. |

b. If you provide the SSL certificate, specify the location and passwords of the keystore and truststore files.
The following table describes the parameters that you must enter for the SSL certificate files:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keystore file directory</td>
<td>Directory that contains the keystore files. The directory must contain files named \texttt{infa_keystore.jks} and \texttt{infa_keystore.pem}.</td>
</tr>
<tr>
<td>Keystore password</td>
<td>Password for the keystore \texttt{infa_keystore.jks}.</td>
</tr>
<tr>
<td>Truststore file directory</td>
<td>Directory that contains the truststore files. The directory must contain files named \texttt{infa_truststore.jks} and \texttt{infa_truststore.pem}.</td>
</tr>
<tr>
<td>Truststore password</td>
<td>Password for the \texttt{infa_truststore.jks} file.</td>
</tr>
</tbody>
</table>

The \textbf{Cluster Type Selection} panel appears.

61. Choose the Hadoop cluster type for Secure@Source.

   • Press 1 to deploy Secure@Source on an internal Hadoop distribution on Hortonworks using Ambari tool.
   • Press 2 to deploy Secure@Source on an external Hadoop distribution on Cloudera CDH 5.10 or Hortonworks version 2.5.

   Depending on the settings that you specify, Secure@Source creates an INFA Hadoop Service for internal Hadoop distribution.

62. Specify whether the cluster is Kerberos-enabled.

   • Press 1 to indicate that the cluster is not Kerberos-enabled.
   • Press 2 to indicate that the cluster is Kerberos-enabled.

   The \textbf{Informatica Cluster Service Creation} panel appears.

63. Enter values for the following parameters:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gateway User</td>
<td>Username for the Apache Ambari server.</td>
</tr>
<tr>
<td>Informatica Cluster Service Name</td>
<td>Name of the Hadoop service for the internal cluster.</td>
</tr>
<tr>
<td>Informatica Cluster Service Port</td>
<td>Port number for the Hadoop service.</td>
</tr>
<tr>
<td>Informatica Hadoop Cluster Gateway Host</td>
<td>Host where the Apache Ambari server runs.</td>
</tr>
<tr>
<td>Informatica Hadoop Cluster Nodes</td>
<td>Hosts where the Apache Ambari agents run.</td>
</tr>
<tr>
<td>Informatica Hadoop Cluster Gateway Port</td>
<td>Web port for the Apache Ambari server.</td>
</tr>
<tr>
<td>Informatica Hadoop Service HTTPS Port</td>
<td>HTTPS port number for the Hadoop Service.</td>
</tr>
<tr>
<td>Catalog Service Name</td>
<td>Name of the catalog service.</td>
</tr>
<tr>
<td>Catalog Service Port</td>
<td>Port number of the catalog service.</td>
</tr>
<tr>
<td>Hadoop Trust Store File</td>
<td>Path of the Hadoop truststore file.</td>
</tr>
</tbody>
</table>
64. Choose whether you want to override the Ambari password.
   • Press 1 to indicate that you do not want to override the Ambari password.
   • Press 2 to indicate that you want to override the Ambari password.

65. Choose whether you want to provide the file path of the HDFS, Yarn, and Zookeeper logs and data directories.
   • Press 1 to indicate No.
   • Press 2 to indicate Yes.
   The Pre-Installation Checks panel appears.

66. Press Enter.
   The Pre-installation System Check tool (i10Pi) runs an evaluation to verify whether the machine meets the system requirements for installation. When the evaluation completes, the Pre-Installation Summary panel appears with the results.

67. Review the results of the system check. You can also access the results of the system check from the log files in the following locations:
   • <Installer extracted location>/logs/I10Pi_server_summary.log-<timestamp>
   • <Installer extracted location>/logs/I10Pi_NodeResults/I10Pi_results_<Hadoop host/node name>.log-<timestamp>

68. Select one of the following options:
   • To continue with the installation, press Enter.
     You can proceed with the installation even if the Pre-installation System Check tool finds failures. However, Informatica highly recommends that you resolve the failed requirements before you proceed.
   • To stop the installation, press Ctrl + C.
   • To go back and run the Pre-installation System Check tool again, type back.

After you complete the installation, the Post-installation Summary panel appears and indicates whether the installation completed successfully. You can view the installation log files to get more information about the tasks performed by the installer and to view configuration properties for the installed components.

Set the properties for the Data Integration Service, the Catalog Service, and the Secure@Source Service.

For information on setting service properties, see the Secure@Source Administrator and User Guide.

Post-Installation Tasks

After you install Secure@Source, perform the post-installation tasks.

Copy Elasticsearch Folder to New Nodes

If you add an extra node to an existing Hadoop cluster, perform the following steps:

1. Copy the Elasticsearch folder from the existing Hadoop node to the new node.
   The Elasticsearch folder on the existing Hadoop node is located in the following directory: /var/lib/hadoop-yarn

2. Ensure that the permissions for the new node are the same as the permissions for the existing node.
Set Catalog Service and Data Integration Service Properties

Set the properties for the Catalog Service and the Data Integration Service from Informatica Administrator (the Administrator tool).

1. From Informatica Administrator, click the **Manage** tab.
2. Click the **Services and Nodes** view.
3. Click the Data Integration Service in the **Domain Navigator** pane.
4. Ensure that the **Properties** view is selected in the service details pane.
5. Configure the Data Integration Service parameters. Informatica recommends the following values based on the number of rows in the data stores:

<table>
<thead>
<tr>
<th>Data Integration Service Property</th>
<th>Property Group Name</th>
<th>Less than 100,000 rows</th>
<th>100,000 rows or More</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Execution Pool Size</td>
<td>Execution Options</td>
<td>&lt;Number of logical cores * 0.5&gt; For example, if you have 32 logical cores, enter 16.</td>
<td>&lt;Number of logical cores * 0.5&gt; For example, if you have 96 logical cores, enter 48.</td>
</tr>
<tr>
<td>Maximum Profile Execution Pool Size</td>
<td>Profiling Warehouse Database Properties</td>
<td>&lt;Number of logical cores * 0.5&gt;</td>
<td>&lt;Number of logical cores * 0.5&gt;</td>
</tr>
<tr>
<td>Maximum Concurrent Profile Jobs</td>
<td>Advanced Profiling Properties</td>
<td>&lt;Number of logical cores * 0.5&gt;</td>
<td>&lt;Number of logical cores * 0.5&gt;</td>
</tr>
<tr>
<td>Maximum Concurrent Columns</td>
<td>Advanced Profiling Properties</td>
<td>100</td>
<td>5</td>
</tr>
<tr>
<td>Maximum Column Heap Size</td>
<td>Advanced Profiling Properties</td>
<td>512</td>
<td>512</td>
</tr>
<tr>
<td>AdvancedProfilingServiceOptions.DomainMappingColumnThreshold</td>
<td>Custom Properties</td>
<td>100</td>
<td>Note: Do not add this parameter.</td>
</tr>
<tr>
<td>ExecutionContextOptions.now_youre_thinking_with_portals</td>
<td>Custom Properties</td>
<td>false</td>
<td>Note: Do not add this parameter.</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Data Integration Service Property</th>
<th>Property Group Name</th>
<th>Less than 100,000 rows</th>
<th>100,000 rows or More</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExecutionOptions.MaxProcessLifeTime</td>
<td>Custom Properties</td>
<td>3600000</td>
<td>3600000</td>
</tr>
</tbody>
</table>
| Maximum Heap Size                                 | Advanced Properties     | Use the following formula to calculate the value: \(((200/1024) \times (\text{number of cores} \times 0.5)) + 1.2 \times (\text{number of cores} \times 0.5)\) Where:  
- \(200 = 200 \text{ MB for each mapping.}\)  
- \(1024 = \text{number of MB in one GB.}\)  
- \(1.2 = 1.2 \text{ GB for each PMDTM.}\) | Use the following formula to calculate the value: \(((200/1024) \times (\text{number of cores} \times 0.5)) + 1.2 \times (\text{number of cores} \times 0.5)\) Where:  
- \(200 = 200 \text{ MB for each mapping.}\)  
- \(1024 = \text{number of MB in one GB.}\)  
- \(1.2 = 1.2 \text{ GB for each PMDTM.}\) |

6. Click the **Recycle** icon to recycle the service and apply the updates.
7. Click the Catalog Service in the **Domain Navigator** pane.
8. Ensure that the **Properties** view is selected in the service details pane.
9. Configure the Catalog Service parameters. Informatica recommends the following values based on the number of Hadoop nodes:

<table>
<thead>
<tr>
<th>Catalog Service Property</th>
<th>Property Group Name</th>
<th>One Hadoop Node</th>
<th>Three Hadoop Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>LdmCustomOptions.loadType</td>
<td>Custom Properties</td>
<td>low</td>
<td>medium</td>
</tr>
<tr>
<td>LdmCustomOptions.ingest.enable.propagation</td>
<td>Custom Properties</td>
<td>false</td>
<td>false</td>
</tr>
</tbody>
</table>

10. Ensure that the **Process** view is selected in the service details pane.
11. Configure the following Catalog Service parameter. Informatica recommends the following values based on the number of Hadoop nodes:

<table>
<thead>
<tr>
<th>Catalog Service Property</th>
<th>Property Group Name</th>
<th>One Hadoop Node</th>
<th>Three Hadoop Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum Heap Size</td>
<td>Advanced Properties</td>
<td>8 GB</td>
<td>12 GB</td>
</tr>
</tbody>
</table>

12. Click the **Recycle** icon to recycle the service and apply the updates.
13. Click the Secure@Source Service in the **Domain Navigator** pane.
14. Ensure that the **Process** view is selected in the service details pane.
15. Configure the Secure@Source Service parameters. Informatica recommends the following values based on the number of Hadoop nodes:

<table>
<thead>
<tr>
<th>Secure@Source Service Property</th>
<th>Property Group Name</th>
<th>One Hadoop Node</th>
<th>Three Hadoop Nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>-DmaxProfilingPoolConnections</td>
<td>Advanced Process Configuration &gt; Additional JVM Options</td>
<td>Set to half or one-third of the value specified in the Maximum Profile Execution Pool Size property of the Data Integration Service.</td>
<td>Set to half or one-third of the value specified in the Maximum Profile Execution Pool Size property of the Data Integration Service.</td>
</tr>
<tr>
<td>Maximum Heap Size</td>
<td>Advanced Process Configuration</td>
<td>8 GB</td>
<td>4 GB</td>
</tr>
</tbody>
</table>

16. Click the **Recycle** icon to recycle the service and apply the updates.
Chapter 7

Troubleshooting the Services Installation

This chapter includes the following topics:

- Installation Troubleshooting Overview, 111
- Installation Log Files, 111
- Troubleshooting Domains and Nodes, 113

Installation Troubleshooting Overview

This chapter provides information about the Secure@Source installation process and the cause and resolution of errors that occur during installation. The examples included in this chapter describe general troubleshooting strategies and are not a comprehensive list of possible causes of installation issues.

Installation Log Files

You can use the following log files to troubleshoot an Informatica installation:

Installation log files

The installer produces log files during and after the installation. You can use these logs to get more information about the tasks completed by the installer and errors that occurred during installation. The installation log files include the following logs:

- Debug logs
- File installation logs

Service Manager log files

Log files generated when the Service Manager starts on a node.

Debug Log Files

The installer writes actions and errors to the debug log file. The name of the log file depends on the Informatica component you install.
The following table describes the properties of the debug log files:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log File Name</td>
<td>- Informatica_&lt;Version&gt;_Services.log</td>
</tr>
<tr>
<td></td>
<td>- Informatica_&lt;Version&gt;_Client.log</td>
</tr>
<tr>
<td></td>
<td>- Informatica_&lt;Version&gt;_Services_Upgrade.log</td>
</tr>
<tr>
<td></td>
<td>- Informatica_&lt;Version&gt;_Client_Upgrade.log</td>
</tr>
<tr>
<td>Location</td>
<td>Installation directory.</td>
</tr>
<tr>
<td>Usage</td>
<td>Get more information about the actions performed by the installer and get more information about installation errors. The installer writes information to this file during the installation. If the installer generates an error, you can use this log to troubleshoot the error.</td>
</tr>
<tr>
<td>Contents</td>
<td>Detailed summary of each action performed by the installer, the information you entered in the installer, each command line command used by the installer, and the error code returned by the command.</td>
</tr>
</tbody>
</table>

The debug log contains output from the infacmd and infasetup commands used to create the domain, node, and application services. It also contains information about starting the application services.

**File Installation Log File**

The file installation log file contains information about the installed files.

The following table describes the properties of the installation log file:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log File Name</td>
<td>- Informatica_&lt;Version&gt;_Services_InstallLog.log</td>
</tr>
<tr>
<td></td>
<td>- Informatica_&lt;Version&gt;_Client_InstallLog.log</td>
</tr>
<tr>
<td>Location</td>
<td>Installation directory.</td>
</tr>
<tr>
<td>Usage</td>
<td>Get information about the files installed and registry entries created.</td>
</tr>
<tr>
<td>Contents</td>
<td>Directories created, names of the files installed and commands run, and status for each installed file.</td>
</tr>
</tbody>
</table>

**Service Manager Log Files**

The installer starts the Informatica service. The Informatica service starts the Service Manager for the node. The Service Manager generates log files that indicate the startup status of a node. Use these files to troubleshoot issues when the Informatica service fails to start and you cannot log in to Informatica Administrator. The Service Manager log files are created on each node.
The following table describes the files generated by the Service Manager:

<table>
<thead>
<tr>
<th>Property</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>catalina.out</td>
<td>Log events from the Java Virtual Machine (JVM) that runs the Service Manager. For example, a port is available during installation, but is in use when the Service Manager starts. Use this log to get more information about which port was unavailable during startup of the Service Manager. The catalina.out file is in the following directory: <code>&lt;Informatica installation directory&gt;/logs/&lt;node name&gt;/catalina.out</code></td>
</tr>
<tr>
<td>node.log</td>
<td>Log events generated during the startup of the Service Manager on a node. You can use this log to get more information about why the Service Manager for a node failed to start. For example, if the Service Manager cannot connect to the domain configuration database after 30 seconds, the Service Manager fails to start. The node.log file is in the /tomcat/logs directory.</td>
</tr>
</tbody>
</table>

**Note:** The Service Manager also uses node.log to record events when the Log Manager is unavailable. For example, if the machine where the Service Manager runs does not have enough available disk space to write log event files, the Log Manager is unavailable.

---

**Troubleshooting Domains and Nodes**

The installer can generate errors when creating and configuring domains and nodes during the Informatica installation.

You can encounter errors with the following installer tasks:

- Adding the domain configuration database
- Starting Informatica
- Pinging the domain
- Adding a license

---

**Creating the Domain Configuration Repository**

If you create a domain, the installer creates a domain configuration repository to store domain metadata. The installer uses the options you enter during installation to add configuration metadata to the domain configuration repository. The installer uses JDBC to communicate with the database. You do not need to configure ODBC or native connectivity on the machine where you install the Secure@Source services.

The installer creates and drops a table in the domain configuration repository database to verify the connection information. The user account for the database must have create privileges on the database. Each domain must have a separate domain configuration repository.

---

**Starting Secure@Source**

The installer runs infaservice to start the Informatica service. To troubleshoot issues when Secure@Source fails to start, use the information in the installation debug log and the node.log and catalina.out Service Manager log files to identify the cause of the error.
After you create a domain, log in to Informatica Administrator after the Informatica service starts to verify that the domain is available.

Secure@Source can fail to start for the following reasons:

- **The Service Manager is out of system memory.** The Java Runtime Environment (JRE) that starts Informatica and runs the Service Manager might not have enough system memory to start. Set the INFA_JAVA_OPTS environment variable to configure the amount of system memory used by Secure@Source. On Linux, you can set the memory configuration when you start Informatica.

- **The domain configuration database is not available.** Secure@Source fails to start on a node if the Service Manager on a gateway node cannot connect to the domain configuration database within 30 seconds. Verify that the domain configuration repository is available.

- **Some of the folders in the Informatica installation directory do not have the appropriate execute permissions.** Grant execute permission on the Informatica installation directory.

- **The localhost does not resolve successfully.** if you use an internal cluster and the localhost does not resolve successfully, Informatica Cluster Service might fail. You need to verify that the localhost resolves successfully.

### Pinging the Domain

The installer runs the `infacmd` Ping command to verify that the domain is available before it continues the installation. The domain must be available so that license objects can be added to the domain. If the Ping command fails, start Secure@Source on the gateway node.

### Adding a License

The installer runs the `infacmd` AddLicense command to read the Informatica license key file and create a license object in the domain. To run the application services in Informatica Administrator, a valid license object must exist in the domain.

You can get more information about the contents of the license key file used for installation, including serial number, version, expiration date, operating systems, and connectivity options in the installation debug log.

You can get more information about existing licenses for the domain in Informatica Administrator.
INDEX

A
AddLicense (infacmd)
  troubleshooting 114
Administrator tool
  overview 18
application services
  installation requirements 25
  naming conventions 27
  overview 16
  ports 23
authentication
  Kerberos 62

C
catalina.out
  troubleshooting installation 112
Catalog Service
  associated services 19
  overview 19
cluster management
  overview 56
compute role
  nodes 15
configuration
  Kerberos files 71
console mode
  installing Secure@Source 84
Content Management Service
  associated services 20
  overview 20
  required databases 20

data integration service
  associated services 20, 21
  overview 20
data object cache
  database requirements 40
  IBM DB2 database requirements 40
  Microsoft SQL Server database requirements 40
  Oracle database requirements 41
data base preparations
  repositories 35
database requirements
  data object cache 40
  installation requirements 25
Model repository 41
  profiling warehouse 44
reference data warehouse 46
Secure@Source repository 48
database user accounts
  guidelines for setup 36
deo path logs
  troubleshooting the installation 111
disk space requirements
  installation requirements 23
domain configuration
  Microsoft SQL Server database requirements 38
  domain configuration repository
IBM DB2 database requirements 48
  Oracle database requirements 38, 43
  preparing databases 36
  requirements 25
  Sybase ASE database requirements 39
  troubleshooting 113
domain objects
  naming conventions 27
domain security
  overview 17
domains
  application services 16
  naming conventions 27
  nodes 15
  overview 15
  planning 18
  ports 23
  security 17
  Service Manager 16
  user authentication 17

E
encryption key
  overview 17
  secure data storage 17
external Hadoop cluster
  preparing 61
  prerequisites 60

G
gateway nodes
  creating during installation 15

I
IBM DB2 database requirements
  data object cache 40
  domain repository 48
  profiling warehouse 44
  reference data warehouse 46
  Secure@Source repository database 48
infacmd
pinging objects 114
InfraHadoop Service
overview 21
Informatica Administrator
overview 18
Informatica Cluster Service
creating 54
overview 53
workflow 53
Informatica services
troubleshooting 113
installation
process 11
installation logs
descriptions 112
installation requirements
application service requirements 25
database requirements 25
disk space 23
minimum system requirements 22
port requirements 23
internal Hadoop cluster
preparing 55
prerequisites 52
K
Kerberos authentication
configuration files 71
creating keytab files 78
creating service principal names 78
external cluster 62
generating keytab file name formats 72
generating SPN formats 72
planning 34
Kerberos SPN Format Generator
Windows 73
keytab files
Kerberos authentication 72, 78
L
license keys
overview 16
licenses
adding 114
overview 16
Linux
installing Secure@Source in console mode 84
log files
catalina.out 112
debug logs 111
installation 111
installation logs 112
node.log 112
types 111
M
Microsoft SQL Server database requirements
data object cache 40
domain configuration 38
Model Repository 43
Profiling Warehouse 44
Microsoft SQL Server database requirements (continued)
Reference Data Warehouse 47
Secure@Source 49
minimum system requirements
nodes 25
Model repository
database requirements 41
Model Repository
Microsoft SQL Server database requirements 43
Model Repository Service
overview 21
required databases 21
N
node
installation 15
node.log
troubleshooting installation 112
nodes
application services 16
compute role 15
gateways 15
naming conventions 27
overview 15
roles 15
Service Manager 16
service role 15
O
Oracle database requirements
data object cache 41
domain configuration repository 38, 43
profiling warehouse 45
reference data warehouse 47
Secure@Source repository 50
P
Ping (infacmd)
troubleshooting 114
port requirements
installation requirements 23
ports
application services 23
domains 23
requirements 23
profiling warehouse
database requirements 44
IBM DB2 database requirements 44
Oracle database requirements 45
Profiling Warehouse
Microsoft SQL Server database requirements 44
R
reference data warehouse
database requirements 46
IBM DB2 database requirements 46
Oracle database requirements 47
Reference Data Warehouse
Microsoft SQL Server database requirements 47