

## Deploy the Informatica® Enterprise Data Catalog Solution on the Microsoft Azure Marketplace (10.4.1)

## Abstract

The automated marketplace solution uses Azure Resource Manager to launch, configure, and run the Azure virtual machine, virtual network, and other services required to deploy a specific workload on Azure. This deployment reference provides step-by-step instructions for deploying Informatica Enterprise Data Catalog on the Microsoft Azure Marketplace.

## Supported Versions

- Enterprise Data Catalog 10.4.1

## Table of Contents

|  |    |
|--|----|
| Overview. . . . .  | 2  |
| Intended Audience. . . . .   | 3  |
| Costs and Licenses. . . . .  | 3  |
| Architecture. . . . .  | 3  |
| Azure Resources in the Deployment. . . . .                             | 4  |
| Informatica Domain. . . . .  | 5  |
| Informatica Clients. . . . .   | 6  |
| Before You Begin. . . . .  | 7  |
| License Key Prerequisite. . . . .                                      | 7  |
| Prerequisites. . . . .   | 7  |
| Deploying Enterprise Data Catalog on the Azure Marketplace. . . . .    | 8  |
| Step 1. Begin Provisioning. . . . .                                    | 8  |
| Step 2. Deploy a Domain and Configure Azure Resources. . . . .         | 9  |
| Monitoring Instance Provision and Informatica Domain Creation. . . . . | 10 |
| Logs. . . . .  | 11 |
| Database Account and User Details. . . . .                             | 12 |
| Enterprise Data Catalog FAQ. . . . .                                   | 12 |

## Overview

Customers of Microsoft Azure and Informatica can execute a Enterprise Data Catalog deployment from the Azure marketplace to create an Informatica domain in the Azure and explore Enterprise Data Catalog functionality.

This deployment reference provides step-by-step instructions for deploying Enterprise Data Catalog on Microsoft Azure. Automated reference deployments use Azure Resource Manager to launch, configure, and run the Azure virtual machine, virtual network, and other services required to deploy a specific workload on Microsoft Azure.

## Intended Audience

This guide is for users who are responsible for deploying the marketplace solution of Enterprise Data Catalog 10.4.1 on Microsoft Azure.

As a user with administrator privileges to deploy applications on Microsoft Azure, you must be familiar with Azure platform elements such as Azure Resource Manager, Virtual Machine, Virtual Network, Azure Databricks, Azure Functions, Azure Active Directory, Azure database, and Azure Blob storage. See the Microsoft Azure documentation.

To find Enterprise Data Catalog documentation, see the [Informatica documentation portal](#).

## Costs and Licenses

You are responsible for the cost of the Azure services used while running this deployment. There is no additional cost for using this marketplace deployment.

The Azure resource manager template for this deployment includes configuration parameters that you can customize. Some of these settings, such as instance type, will affect the cost of deployment. See the pricing pages for each Azure service that you plan to use for cost estimates.

This deployment requires a license for Informatica Enterprise Data Catalog. To sign up for a license, contact your organization's Informatica sales contact or [Informatica Global Customer Support](#).

**Note:** You supply the license key value in the Informatica Enterprise Data Catalog License Key parameter when you configure the deployment.

The following table lists the instance types that you can choose based on sizing requirements:

| Virtual Machine             | Instance Type  | Cluster Size         |
|-----------------------------|--|----------------------|
| Database                    | Standard_D3_v2 / Standard_DS3_v2 / Standard_D3<br>This includes the SQL Server 2014 SP2 on Windows Server 2012 R2 Datacenter with pay as you go (PAYG) license model. You will be charged based on the running instances.<br><b>Note:</b> For information about changing the license mode, see the <a href="#">Microsoft documentation</a> . | Small, Medium, Large |
| Informatica Domain          | Standard_D5_v2 / Standard_DS5_v2 / Standard_E16_v3 / Standard_D14_v2   | Small, Medium, Large |
| Bastion server              | Standard_D2_v2 / Standard_A2_v2 / Standard_DS2_v2 / Standard_B2s   | Small, Medium, Large |
| Informatica Cluster Service | Standard_D5_v2 / Standard_DS5_v2 / Standard_E16_v3 / Standard_D14_v2   | Small                |
| Informatica Cluster Service | Standard_D4_v2 / Standard_DS4_v2 / Standard_E16_v3 / Standard_D14_v2 / Standard_DS5_v2 / Standard_D5_v2  | Medium, Large        |

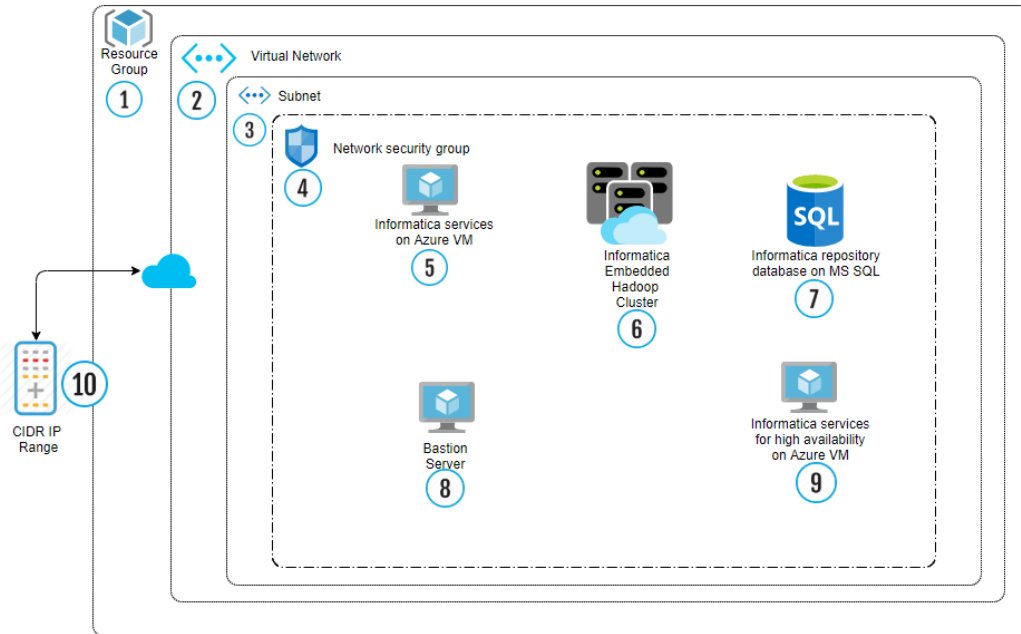
## Architecture

The Microsoft Azure marketplace solution, when you deploy on a virtual network, creates and connects the following resources in the network:

- Informatica domain server on a virtual machine, with one additional high availability virtual machine.
- Informatica clients on a bastion server.

- Microsoft SQL Server for the repositories in the Informatica domain.
- Informatica embedded Hadoop cluster on a virtual machine.

The following image shows the architecture of the Enterprise Data Catalog on Microsoft Azure:



The numbers in the architecture diagram correspond to items in the following list:

1. A resource group on the Azure platform.
2. A virtual network that includes a subnet.
3. A subnet to contain specific elements of the deployment.
4. A network security group that includes the Enterprise Data Catalog deployment.
5. The Informatica services on Azure Virtual Machine.
6. Informatica embedded Hadoop cluster.
7. Microsoft SQL Server database instance for the Informatica domain configuration repository.
8. Bastion server, if you had selected the option to deploy a bastion server.
9. Informatica services for high availability on Azure Virtual Machine.
10. The CIDR IP address range that you use to access the Informatica services URL and virtual machines.

## *Azure Resources in the Deployment*

The deployment process creates or includes the components listed in this section.

## Components in a Deployment in a New Virtual Network

The following components are created when you deploy the software:

| Component              | Number of Components Created   |
|------------------------|--|
| Virtual Network        | One virtual network.   |
| Network security group | Two network security groups: <ul style="list-style-type: none"><li>- One for Informatica domain and high availability.</li><li>- One for the Informatica embedded Hadoop cluster.</li></ul>  |
| Subnet                 | One subnet.  |
| Microsoft SQL Server   | One Microsoft SQL Server instance on virtual machine.  |
| Virtual machines       | Up to nine virtual machines with the following assignments: <ul style="list-style-type: none"><li>- One for Informatica domain.</li><li>- One for high availability.</li><li>- One for bastion server.</li><li>- One, three, or six for the Informatica embedded Hadoop cluster.</li></ul> |

## Components in a Deployment on an Existing Virtual Network

The following components are created when you deploy the software:

| Component              | Number of Components Created   |
|------------------------|--|
| Network security group | Two network security groups: <ul style="list-style-type: none"><li>- One for Informatica domain and high availability.</li><li>- One for the Informatica embedded Hadoop cluster.</li></ul>  |
| Microsoft SQL Server   | One Microsoft SQL Server instance on virtual machine.  |
| Virtual machines       | Up to nine virtual machines with the following assignments: <ul style="list-style-type: none"><li>- One for Informatica domain.</li><li>- One for high availability.</li><li>- One for bastion server.</li><li>- One, three, or six for the Informatica embedded Hadoop cluster.</li></ul> |

## Bastion Server

You can optionally deploy a bastion server as an element in the solution. The bastion server acts as a firewall between the internet and the cloud platform network where the solution deploys. It can also act as a remote server which you can log in to run Informatica clients.

## Informatica Domain

The Informatica domain is a server component that hosts application services, such as the Model Repository Service, cluster configuration for an existing configuration, cloud provisioning configuration for an autodeployed configuration, and Data Integration Service. These services, together with domain clients, enable you to create and run mappings and other objects to extract, transform, and write data.

## Application Services

The Informatica domain includes the following application services:

### **Model Repository Service**

The Model Repository Service manages the Model repository. The Model repository stores metadata created by Informatica products in a relational database to enable collaboration among the products. Informatica Developer, the Data Integration Service, and the Administrator tool store metadata in the Model repository.

### **Data Integration Service**

The Data Integration Service is an application service in the Informatica domain that performs data integration tasks for the Developer tool and for external clients.

### **Metadata Access Service**

The Metadata Access Service is an application service that allows the Developer tool to access Hadoop connection information to import and preview metadata. The Metadata Access Service contains information about the Service Principal Name (SPN) and keytab information if the Hadoop cluster uses Kerberos authentication.

### **Content Management Service**

The Content Management Service is a component within the Informatica domain that manages reference data and is responsible for compiling rule specifications into mapplets. The Content Management Service provides reference data information to the Data Integration Service and to the Developer tool and Analyst tool. The Content Management Service stores reference data in a database that you specify.

### **Analyst Service**

The Analyst Service manages the connection between the service components and the users who log in to Analyst tool. You can perform column and rule profiling, manage scorecards, and manage bad records and duplicate records in the Analyst tool. The Analyst Service stores profiling, scorecarding, and bad and duplicate record data in databases that you specify.

### **Informatica Cluster Service**

Runs and manages all the Hadoop services, Apache Ambari server, and Apache Ambari agents on an embedded Hadoop cluster.

### **Catalog Service**

Runs Enterprise Data Catalog and manages connections between service components and external applications.

### **Domain Repositories**

Informatica repositories, hosted on Microsoft SQL Server databases, store metadata about domain objects. The Informatica domain includes the following repositories:

#### **Domain configuration repository**

The domain configuration repository stores configuration metadata about the Informatica domain. It also stores user privileges and permissions.

### **Application Service Databases**

The Informatica domain and the application services use a series of databases to store information. You must set up the databases with the database names and user names that the domain and the services expect.

To read a list of the database names and user names that you must apply to the databases, see [“Database Account and User Details” on page 12](#).

### **Informatica Clients**

You can use several different clients with Enterprise Data Catalog:

**Administrator tool**

The Administrator tool enables you to create and administer services, connections, and other domain objects.

**Developer tool**

The Developer tool enables you to create and run mappings and other objects that enable you to access, transform, and write data to targets.

**Command line interface**

The command line interface offers hundreds of commands to assist in administering the Informatica domain, creating and running repository objects, administering security features, and maintaining domain repositories.

**Catalog Administrator**

The Enterprise Data Catalog administration tool that you can use to manage and monitor the resources, schedules, attributes, and connections.

**Enterprise Data Catalog**

The tool that you can use to perform the following tasks:

- Search for assets in the catalog.
- Verify the quality of data such as profiling information.
- View the lineage and relationship of assets in the catalog.
- Enrich assets with attributes for ease of search.

**Bastion server**

You can optionally deploy a bastion server as an element in the solution. The bastion server is a Windows instance installed with the Developer tool and the command line interface clients. The bastion server acts as a firewall for access to the network. It can also act as a remote Windows server that you can log in to run Informatica clients.

## Before You Begin

Before you launch the automated deployment on Microsoft Azure, verify the prerequisites and make the choices described in this section.

### *License Key Prerequisite*

Verify that you have a license to deploy Enterprise Data Catalog.

You supply the license key value in the Informatica Enterprise Data Catalog License Key parameter when you configure the deployment.

### *Prerequisites*

Before you deploy Enterprise Data Catalog on Microsoft Azure, verify the prerequisites.

- You must have a Microsoft Azure subscription with owner role.
- You must have access and permissions to create the following resources on the Azure platform:
  - Virtual network
  - Network security group

- Virtual machines
- You have a Contributor or higher role.
- You have sufficient number of CPU cores based on the instance types in the region where you plan to deploy the Enterprise Data Catalog solution.

**Note:** Not all Azure resources are supported in all regions. See the Azure documentation to verify that the resources for your solution are supported in your desired region.

In addition to geographical regions, the solution supports government cloud regions. Contact Informatica Global Customer Support to check support for your desired region.

## Deploying Enterprise Data Catalog on the Azure Marketplace

The automated deployment of Enterprise Data Catalog on the Azure marketplace uses the Azure Resource Manager template to guide your choices and launch the solution deployment.

When you provision the Enterprise Data Catalog solution on the Azure marketplace, launch the wizard and configure the basic properties. Later, configure the solution.

### Step 1. Begin Provisioning

Use the Azure Marketplace website to provision Azure cluster resources including a Enterprise Data Catalog deployment.

When you implement the Enterprise Data Catalog solution on Azure marketplace, you launch the wizard, configure basic properties.

1. Search for and select the Enterprise Data Catalog solution.
  - a. Log in to the [Azure marketplace](#) website. Use the search bar to search for Informatica Enterprise Data Catalog.
  - b. Select **Informatica Enterprise Data Catalog 10.4.1**.  
Click **Get it now** to launch the solution wizard.
  - c. Read the details of the terms of use and click **Continue**.  
The wizard redirects the browser window to the Enterprise Data Catalog 10.4.1 solution on the Azure portal.
  - d. Click **Create**.

A series of panels opens to enable you to configure the solution on the Azure platform.

2. Enter the information in the Basics panel, and click **OK**.



## Step 2. Deploy a Domain and Configure Azure Resources

Create an Informatica domain and configure new or existing Azure resources to use with it.

### Basics

Enter values for the following parameters:

| Parameter      | Description   |
|----------------|---|
| Subscription   | Required. Azure subscription you use to manage the deployment.  |
| Resource Group | Required. The Azure resource group containing the Virtual Network where you deploy Enterprise Data Catalog. |
| Region         | Required. Azure location where you deploy Enterprise Data Catalog.  |

### Informatica Enterprise Data Catalog

Enter values for the following parameters:

| Parameter   | Description  |
|---|--|
| Informatica License Key                             | Required. Indicates the Enterprise Data Catalog license key value to upload.   |
| Informatica High Availability                       | Indicates whether you want to enable high availability for the Enterprise Data Catalog deployment.<br>Default is Disabled.<br>For information about high availability for the Informatica domain, see the <a href="#">High Availability</a> documentation. |
| Informatica Server                                  | Required. Indicates the virtual machine size of the Informatica server. Default is Standard_D5_v2.   |
| Database Server                                     | Required. Indicates the virtual machine size of the database server.   |
| Informatica Embedded Hadoop Cluster Deployment Type | Select the size of the Informatica embedded Hadoop cluster required from the following options:<br>- Small<br>- Medium<br>- Large  |
| Embedded Hadoop Cluster Virtual Machine Size        | Size of the virtual machine that hosts the embedded Hadoop cluster.  |
| Password  | Indicates the password for SSH, RDP, database, and database users.   |
| Confirm Password                                    | Confirms the password that you entered.  |

## Bastion Server

Enter values for the following parameters:

| Parameter             | Description   |
|-----------------------|---|
| Deploy Bastion Server | Deploys a bastion server to access other resources in the virtual network.<br>Default is No.                      |
| Bastion server size   | Virtual machine size. Applicable only when you choose to deploy the bastion server.<br>Default is Standard_D2_v2. |

## Network Configuration

Enter values for the following parameters:

| Parameter             | Description   |
|-----------------------|---|
| CIDR IP Address Range | Required. The CIDR public IP range of clients that are permitted to access the Informatica Enterprise Data Catalog. Format is x.x.x.x/x.  |
| Assign Public IP      | Assigns a public IP address to the network interface that is attached to the virtual machine.<br>Default is Yes.  |
| Virtual Network       | Required. The identifier for the Azure virtual network where you want to deploy Enterprise Data Catalog.<br><b>Note:</b> The deployment supports new and existing virtual networks. The Azure location must be same for the virtual network resource group and the deployment resource group. |
| Subnet                | Required. The identifier for the subnet within the virtual network where Enterprise Data Catalog is deployed.   |

After you configure the parameters, verify the choices in **Review + create**, read the terms of use, and click **Create**.

When you click **Create**, Azure deploys the Enterprise Data Catalog and creates resources in the environment that you configured.

## Monitoring Instance Provision and Informatica Domain Creation

You can use cloud platform dashboards, logs, or other artifacts to see whether cluster creation succeeded and how to locate and identify the Informatica domain on the cloud platform.

### During Deployment

After you finish configuring the solution and start the deployment process, the Azure dashboard indicates deployment status in the top right corner.

To view the detailed status of the deployment job, including resources, click **Deployment in progress...**

### When Deployment is Complete

The automated deployment includes the following resources:

- Virtual network
- Network security group

- Microsoft SQL Server database
- Informatica domain
- Informatica embedded Hadoop cluster

Perform the following steps to use your Azure dashboard to verify the status of resource deployment:

1. Use the dashboard search bar to search for the resource group that contains the Enterprise Data Catalog deployment.  
The dashboard displays the **Overview** view of the resource group, with resource deployment status as a clickable link in the upper right corner.
2. Click the resource deployment status link.  
When you click the deployment status link, a detail window opens listing the failed and successful deployments.
3. Click **Error details** for information about failed resource deployments.
4. Click **Overview** to see a list of the resources in a resource group.
5. You can click column headings in the display to sort by name, type, or location of the resource.

When the deployment is complete, you can open the Informatica Administrator tool in a browser. The Administrator tool URL has the following format:

*https://<Public IP address\_or\_DNS name\_or\_Private IP address>:8443*

*User name: infauser*

You can read the values from the properties of the Informatica services virtual machine in your resource group.

## Logs

After the completion of the Enterprise Data Catalog deployment, consult logs to see the success or failure of solution element creation.

You can access the following logs on the virtual machine that hosts the Informatica domain:

### Azure extension operation logs

Records the installation of Azure resources and services.

You can find the file in the following location:

`/var/log/azure/custom-script/handler.log`

**Note:** The directory path `/var/lib/waagent/custom-script/download/0` contains the `stdout` and `stderr` logs. The directory also contains the file `silentlaunch.sh`, which contains the script that was executed to install Azure resources and services.

### Command execution log

This log records the following events:

- Creation of Informatica connections and services.
- Population of the domain and its repositories.

You can find the file in the following location:

`/opt/Oneclicksolution.log`

### Informatica domain and services configuration log

At the top of the log file is a summary section that lists automated tasks and their status. You can view the details about each task under the summary section. If any of the tasks failed complete successfully, you can look at the detailed section for the task to troubleshoot the task.

## Database Account and User Details

The Informatica domain and the application services use a series of databases to store information. Verify that database names and user account details on each database match the names and details that the databases expect.

The following table describes the database and user account information:

| Database Name | User        | Applicable For   |
|---------------|-------------|--|
| domaindb      | domainuser  | Informatica Domain   |
| mrsdb         | mrsuser     | Model Repository Service                                   |
| cmsdb         | cmsuser     | Content Management Service                                 |
| pwhdb         | pwduser     | Data Integration Service as profiling warehouse connection |
| monitordb     | monitoruser | Monitoring Model repository                                |
| wfhdb         | wfhuser     | Data Integration Service as workflow connection            |
| analystdb     | analystuser | Analyst Service  |

## Enterprise Data Catalog FAQ

The `Extension Script time out` error message appears and the deployment status of Enterprise Data Catalog is shown as unsuccessful. The Informatica Administrator displays the status of the Enterprise Data Catalog application services as up and running. Is my deployment successful?

You can ignore the error message and unsuccessful deployment status. Enterprise Data Catalog is successfully deployed and you can start to use Enterprise Data Catalog. Microsoft Azure Marketplace displays the deployment status as unsuccessful if the deployment time exceeds a specific time limit.

## Author

Suraj Jayan