

Deploy the Informatica® 10.5.3 Enterprise Data Catalog Solution on the Microsoft Azure Marketplace

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 10.5.3

Table of Contents

Overview	2
Intended Audience.	3
Costs and Licenses.	3
Architecture.	4
Azure Resources in the Deployment.	4
Informatica Domain.	5
Informatica Clients.	6
Before You Begin.	7
License Key Prerequisite.	7
Prerequisites.	8
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 10.5.3 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:

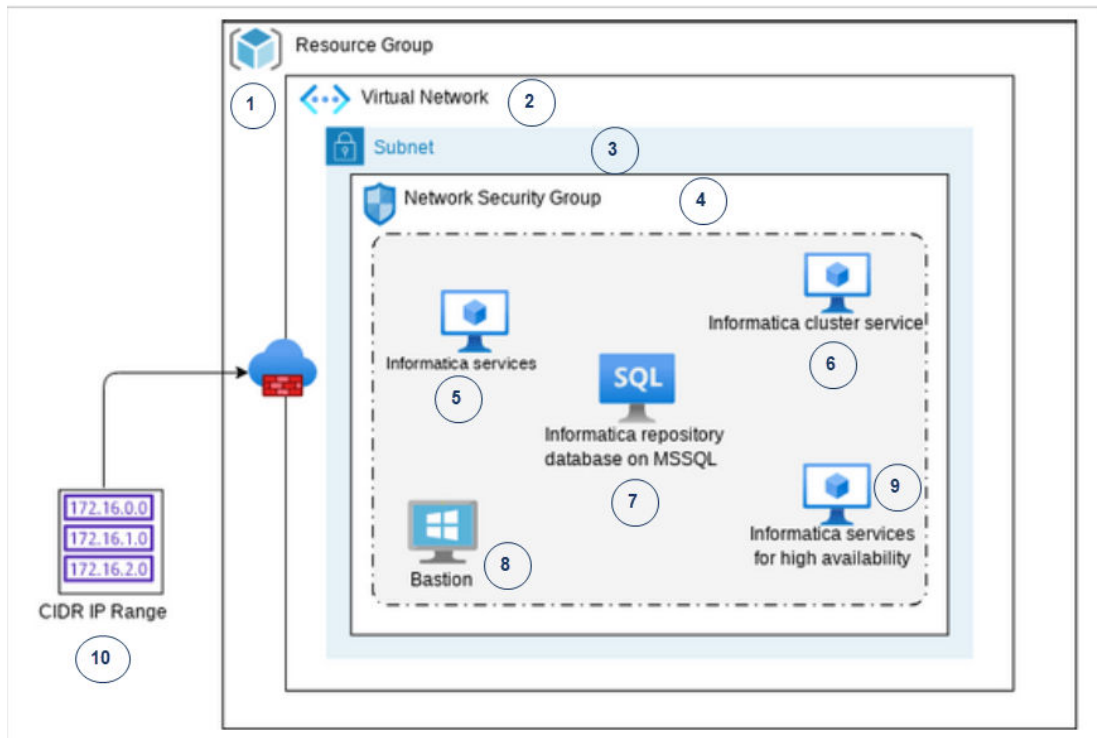
Virtual Machine	Instance Type	Cluster Size
Database	Standard_D4s_v3 / Standard_D8s_v3 / Standard_D16s_v3 / Standard_E16_v3 / Standard_D4as_v4 / Standard_D16as_v4 / Standard_D16ds_v4 / Standard_D8s_v4 / Standard_D32as_v4 / Standard_E16as_v4 / Standard_E16ds_v4 / Standard_D16ds_v5 This includes the Microsoft SQL Server 2019 on Windows Server 2019 Datacenter with pay as you go (PAYG) license model. You will be charged based on the running instances. Note: For information about changing the license mode, see the Microsoft documentation .	Small, Medium, Large
Informatica Domain	Standard_D4s_v3 / Standard_D8s_v3 / Standard_D16s_v3 / Standard_E16_v3 / Standard_D4as_v4 / Standard_D16as_v4 / Standard_D16ds_v4 / Standard_D8s_v4 / Standard_D32as_v4 / Standard_E16as_v4 / Standard_E16ds_v4 / Standard_D16ds_v5 This includes the Red Hat Enterprise Linux 8.3 with pay as you go model. You will be charged based on the running model.	Small, Medium, Large
Bastion server Optional	Standard_D4s_v3 / Standard_D8s_v3 / Standard_D16s_v3 / Standard_E16_v3 / Standard_D4as_v4 / Standard_D16as_v4 / Standard_D16ds_v4 / Standard_D8s_v4 / Standard_D32as_v4 / Standard_E16as_v4 / Standard_E16ds_v4 / Standard_D16ds_v5 This includes the Microsoft Windows Server 2019 Datacenter with pay as you go model. You will be charged based on the running instances.	-
Informatica Cluster Service	Standard_D4s_v3 / Standard_D8s_v3 / Standard_D16s_v3 / Standard_E16_v3 / Standard_D4as_v4 / Standard_D16as_v4 / Standard_D16ds_v4 / Standard_D8s_v4 / Standard_D32as_v4 / Standard_E16as_v4 / Standard_E16ds_v4 / Standard_D16ds_v5 This includes the Red Hat Enterprise Linux 8.3 with pay as you go model. You will be charged based on the running model.	Small, 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">- One for Informatica domain and high availability.- One for the Informatica cluster service.
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">- One for Informatica domain.- One for high availability.- One for bastion server.- One, three, or six for the Informatica cluster service.

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">- One for Informatica domain and high availability.- One for the Informatica cluster service.
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">- One for Informatica domain.- One for high availability.- One for bastion server.- One, three, or six for the Informatica cluster service.

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

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

Performs data integration tasks for the Developer tool and for external clients.

Metadata Access Service

Allows the Developer tool to access cluster 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

Manages reference data and is responsible for compiling rule specifications into maplets. 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

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 the nodes and services associated with Enterprise Data Catalog.

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 can upload the license key file from your local system to the container in the Microsoft Azure storage account. When you configure the Enterprise Data Catalog deployment, you then need to add the license key from your Microsoft Azure storage account.

Microsoft Azure Storage Account

You can use your existing Microsoft Azure storage account or create a new account.

To create a new Microsoft Azure storage account, complete the following the steps:

1. Log in to the Microsoft Azure portal.
2. Click **Create a resource**.
3. Search for the storage account.
4. Click **Create** and enter the required parameters.
5. Click **Review + create**.
6. Navigate to **Storage resource > Containers > + Container**.
7. Enter a container name.

8. Select `Private` (no anonymous access level) for the public access level.
9. Click to open the created container.

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 10.5.3**.
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 10.5.3 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. Redirects you to the list of storage account under your subscription. Select the container that has the license file for upload.
Informatica High Availability	Indicates whether you want to enable high availability for the Enterprise Data Catalog deployment. Default is Disabled. For information about high availability for the Informatica domain, see the High Availability documentation.
Informatica Server	Required. Indicates the virtual machine size of the Informatica domain. Default is Standard_D16ds_v4.
Database Server	Required. Indicates the virtual machine size of the database server. Default is Standard_D8s_v3.
Informatica Cluster Service Deployment Type	Select the size of the Informatica cluster service required from the following options: <ul style="list-style-type: none">- Small- Medium- Large
Cluster Service Virtual Machine Size	Size of the virtual machine that hosts the Informatica cluster service. Default is Standard_D16ds_v4.
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. Default is No.
Bastion server size	Virtual machine size. Applicable only when you choose to deploy the bastion server. Default is Standard_D4s_v3.

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. Default is Yes.
Virtual Network	Required. The identifier for the Azure virtual network where you want to deploy Enterprise Data Catalog. Note: 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 cluster service

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

The Advanced Scanners tool is available with Enterprise Data Catalog. For high availability, Advanced Scanners runs only on the primary node.

The URL for accessing the Advanced Scanners tool has the following format:

https://<Public IP address_or_DNS name_or_Private IP address>:48090

User name: admin

Password: admin

Note: You must change the default password after you log in to the Advanced Scanners tool.

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
daadb	daauser	Data Asset Analytics

Enterprise Data Catalog FAQ

Q. 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?

A. 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.

Q. The virtual machine displays the following error message while running the extension script:

```
Enable failed: processing file downloads failed: failed to download file[0]: failed to download the file: http request failed: Get [REDACTED] dial tcp 13.107.246.10:443: i/o timeout
```

The deployment status of Enterprise Data Catalog is shown as unsuccessful. How do I solve this issue?

A. To solve this issue, whitelist the following outbound firewall security rules:

Ports	IP Address/Sites	Description
80 and 443	catalogartifact.azureedge.net	Required to download the extension script from an artefact location.
All	VirtualNetwork	Required to communicate with the Informatica Cluster Service nodes and database to connect to the Domain.
443	Storage	Required to download the product license.
123	0.rhel.pool.ntp.org	Required for the NTP servers such as Informatica Cluster Service services. Note: The IP address changes dynamically as the NTP servers are connected in the group.
123	1.rhel.pool.ntp.org	
123	2.rhel.pool.ntp.org	
123	3.rhel.pool.ntp.org	
1023	*.pool.ntp.org	Required to communicate with the Informatica Cluster Service nodes and the Catalog Service. Replace * with the name of the NTP Pool server closest to your location. Note: The IP address changes dynamically as the NTP servers are connected in the group.
443	13.91.47.76/32	Required for the Azure RHUI content delivery servers. For more information, see: https://docs.microsoft.com/en-us/azure/virtual-machines/
443	40.85.190.91/32	
443	52.187.75.218/32	
443	52.174.163.213/32	
443	52.237.203.198/32	

Q. A "No license available" message appears in the Advanced Scanner user interface.

A. Verify that you set the EDC_USER, EDC_PASSWORD, and EDC_URL global variables under **Administration**. Also verify that the Catalog Service is enabled.

Author

Ashwin G