MarkLogic Server

MarkLogic Server on Microsoft® Azure® Guide

MarkLogic 9
January, 2018

Last Revised: 9.0-11, November, 2019

Copyright © 2019 MarkLogic Corporation. All rights reserved.

Table of Contents

MarkLogic Server on Microsoft® Azure® Guide

1.0 Overview of MarkLogic Server on Azure 5
- 1.1 Azure Terminology 5
- 1.2 Using MarkLogic Solution Template for Azure 6
- 1.3 Deployment Options 6
- 1.4 Typical Architecture 7

2.0 Getting Started with MarkLogic Server on Azure 9
- 2.1 System Requirements and Installation 9
- 2.2 Separate MarkLogic Converters 10
- 2.3 Set up a Simple Deployment 10
  - 2.3.1 Locate the MarkLogic Cluster Deployment Offering 10
  - 2.3.2 Configure Basic Settings 12
    - 2.3.2.1 Basic Information Values 13
  - 2.3.3 MarkLogic Configuration 14
    - 2.3.3.1 Password Policy 14
    - 2.3.3.2 MarkLogic Configuration Values 15
  - 2.3.4 Configure Cluster Resources 16
    - 2.3.4.1 Configure Cluster Resources Values 17
  - 2.3.5 Choose a VM Size 17
  - 2.3.6 Review the VM Summary 19
  - 2.3.7 Create and Purchase 20
- 2.4 Resource Configuration 21
  - 2.4.1 Virtual Network 21
  - 2.4.2 Availabilty Set 22
  - 2.4.3 Node Public IP Address 22
  - 2.4.4 Network Security Group 23
  - 2.4.5 Load Balancer 23
    - 2.4.5.1 Public IP Address Allocation 24
    - 2.4.5.2 Load Balancing Rules 24
    - 2.4.5.3 Health Probes 24
  - 2.4.6 Data Storage: Managed Disks 25
  - 2.4.7 Virtual Machines 25
- 2.5 Cluster Initialization 25
- 2.6 Limitations 25
  - 2.6.1 Availability Set 25
  - 2.6.2 Shared Disk Failover 26
- 2.7 List of Configurable Parameters 26
  - 2.7.1 Basic Information Values 26

Contains SAP BusinessObjects Text Analysis XI from SAP AG. Copyright © 1996-2011. All rights reserved.

2.7.2 Azure Configuration Values 27
2.7.3 Configure Cluster Resources Values 28

3.0 Managing MarkLogic on Azure 29
3.1 Accessing Your MarkLogic Azure VM 29
3.2 Attaching Data Disks 29
3.3 Partitioning and Mounting Disks 30
3.4 Loading Content into MarkLogic 30
3.5 Configuring MarkLogic for Azure Blob Storage 31
3.5.1 Credentials Configuration 35
3.5.2 Environment Variables 36
3.6 Adding a Node to the Existing Cluster 36
3.7 Removing a Node in the Existing Cluster 36
3.8 Resizing Virtual Machines 36
3.9 Monitoring Virtual Machine Status 37
3.10 Monitoring MarkLogic Server 38
3.11 Upgrading MarkLogic Server on Azure 38

4.0 Technical Support 39

5.0 Copyright 41
5.0 COPYRIGHT 41

1.0 Overview of MarkLogic Server on Azure

MarkLogic has partnered with Microsoft to deploy MarkLogic clusters on the Azure cloud with an officially supported MarkLogic Azure image, along with recommendations for your deployments. This guide covers the information and steps required to use MarkLogic on Azure. The solution template to configure the image can be found on Github [https://github.com/marklogic/cloud-enablement/]. In the Azure environment, you can deploy virtual machines (VMs) from the published MarkLogic Azure image. Or you can create MarkLogic clusters and resources including Load Balancers, Virtual Networks, and Virtual Machines using the Solution Template.

In addition to officially supporting MarkLogic on the Azure platform, many MarkLogic features are designed for better and easier deployment and management on cloud platforms. Features such as Encryption, Redaction, and Element Level Security enable MarkLogic to securely host your data in public cloud environments. For details about these features, see Encryption at Rest and Element Level Security in the *Security Guide*, and Redacting Content During Export or Copy Operations in the *mlcp User Guide*.

MarkLogic also includes telemetry, a support feature that aims to shorten resolution time by making system level log, metering, and configuration data accessible to MarkLogic Support. See Telemetry in the *Monitoring MarkLogic Guide* for more information.

This chapter includes the following sections:

- Azure Terminology
- Using MarkLogic Solution Template for Azure
- Deployment Options
- Typical Architecture

For more detailed information about Microsoft Azure, see the Microsoft Azure documentation.

1.1 Azure Terminology

Virtual Network is a representation of your own network in the cloud. See https://docs.microsoft.com/en-us/azure/virtual-network/virtual-networks-create-vnet-arm-cli for more information.

Availability Set is a logical grouping where virtual machines (VMs) created within an availability set are distributed across the underlying infrastructure. See https://docs.microsoft.com/en-us/azure/virtual-machines/linux/infrastructure-availability-sets-guidelines for details.

Public IP Address allows exposure of your content to the public Internet. See https://docs.microsoft.com/en-us/azure/virtual-network/virtual-network-deploy-static-pip-arm-cli for more information.

Managed Disks are the storage used to handle Azure Storage account creation and management in the background. See https://docs.microsoft.com/en-us/azure/virtual-machines/linux/convert-unmanaged-to-managed-disks for details.

Virtual Machines enable you to deploy a wide range of options for computing in an agile way. See https://docs.microsoft.com/en-us/azure/virtual-machines/linux/endorsed-distros for more information.

1.2 Using MarkLogic Solution Template for Azure

Using the Solution Template, you can customize various parameters for your MarkLogic cluster on Azure, such as Virtual Machine sizes and regions, specifying cluster names and admin credentials. The options for customization offered by the Solution Template are a subset of all of the configurable parameters for the resources associated with the cluster. The Solution Template is used to easily deploy one-node clusters or three-node clusters, using a developer's license or using your own existing MarkLogic license (known as BYOL or "bring your own license").

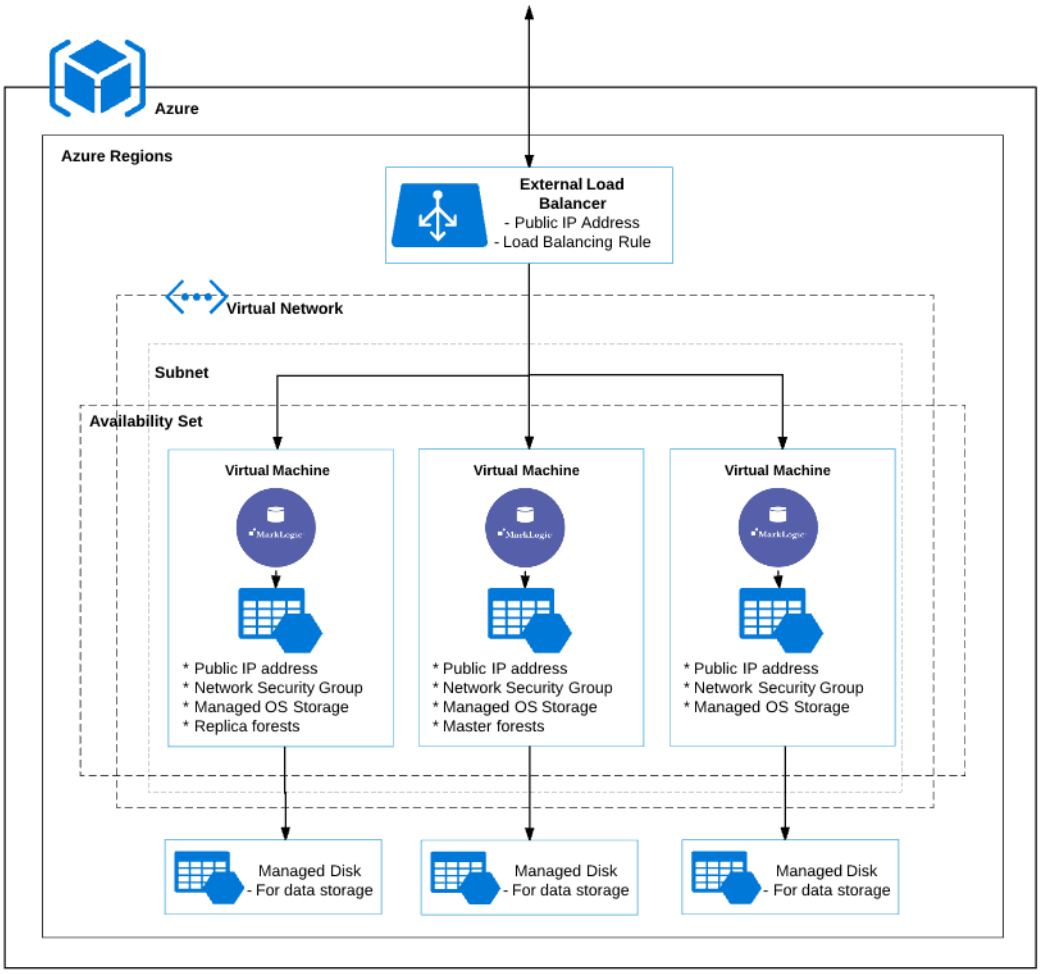
The Solution Template makes it easy to deploy MarkLogic clusters and resources with customized parameters on the Azure platform. The Azure Resource Manager can also provide additional tools to manage the many resources available in a group.

1.3 Deployment Options

In the Azure environment, you can deploy virtual machines (VMs) from the published MarkLogic Azure image. You can then add the storage, networking, and security features that you require. You can use Solution Template to setup the IT environment via infrastructure as code concept.

1.4 Typical Architecture

This diagram briefly explains the overall architecture of a cluster deployed on Azure.



2.0 Getting Started with MarkLogic Server on Azure

This chapter describes how to launch a MarkLogic Server cluster on Azure using the Solution Template to configure parameters. It includes additional information about supplying values for the template fields related to the cluster.

> **Note:** The parameter names used here are not a one-to-one mapping with Azure resource configurations.

This chapter includes the following sections:

- System Requirements and Installation
- Separate MarkLogic Converters
- Set up a Simple Deployment
- Resource Configuration
- Limitations
- List of Configurable Parameters

2.1 System Requirements and Installation

For a MarkLogic Azure image, you will need to choose a VM that has more than 2 GB of memory.

> **Note:** For production environment, Premium storage disks in Azure are preferred. Different VM instance types in Azure may support different storage types. See Requirements and Database Compatibility in the *Installation Guide* for information regarding MarkLogic requirements for Memory, Disk Space and Swap Space, and Transparent Huge Pages.

You will need this basic information to get set up. All of these are required:

- **Subscription**: Required by Azure. It is the Azure subscription to which the cluster usage will be billed.
- **Resource Group**: Required by Azure. The resource group to which the cluster will belong.
- **Location**: Required by Azure. Where you want to deploy the cluster.
- **Deployment Name**: Required. The value to be used as a prefix for all of your resources. There is a 10 character limit for the Deployment Name. Any blank spaces will be removed.

Important: You will need to setup your Azure account before proceeding to the next steps. After setting up the account, you can find your subscription ID under “Subscriptions”.

See “Managing MarkLogic on Azure” on page 29 for more details about configuration options.

2.2 Separate MarkLogic Converters

Starting with MarkLogic 9.0-4, the MarkLogic converters/filters are offered as a package (called MarkLogic Converters package) separate from the MarkLogic Server package. For Azure, the converter installer/package is located in your default user home directory. There is a `README.txt` file in the package describing what the package is for, and pointing to the MarkLogic documentation for more information. See MarkLogic Converters Installation Changes Starting at Release 9.0-4 in the *Installation Guide* for more details.

2.3 Set up a Simple Deployment

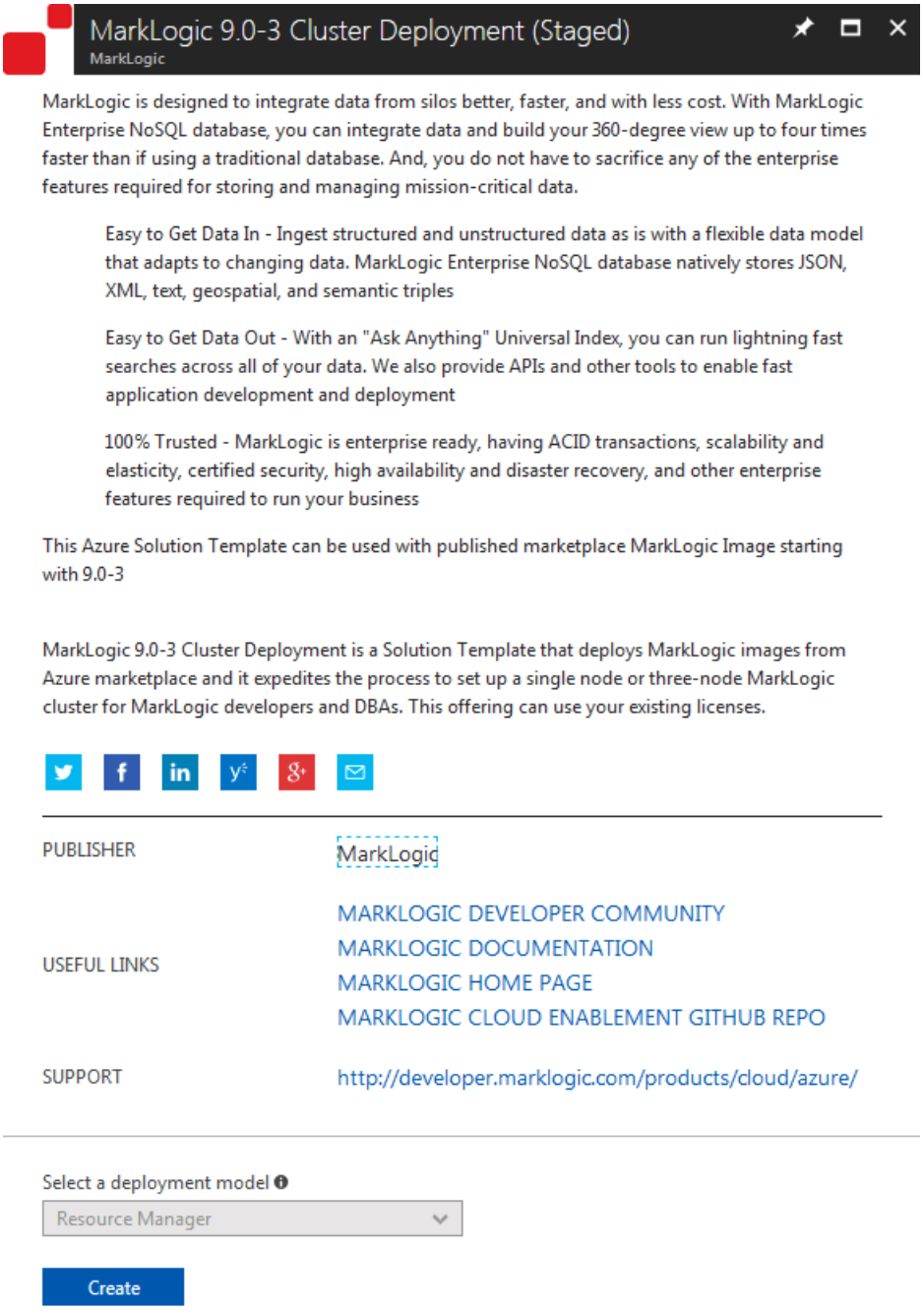
This section covers the steps to set up at simple MarkLogic deployment. It is broken into these topics:

- Locate the MarkLogic Cluster Deployment Offering
- Configure Basic Settings
- MarkLogic Configuration
- Configure Cluster Resources
- Choose a VM Size
- Review the VM Summary
- Create and Purchase

2.3.1 Locate the MarkLogic Cluster Deployment Offering

The Azure Marketplace hosts the MarkLogic Cluster Deployment Offering online. With a browser, navigate to the Azure portal at https://portal.azure.com/. Go to the Microsoft Azure Marketplace and search for “MarkLogic”.

1. Choose the cluster deployment offering for MarkLogic 9.0-x Cluster Deployment. Click on the image to see a page similar to the following one.

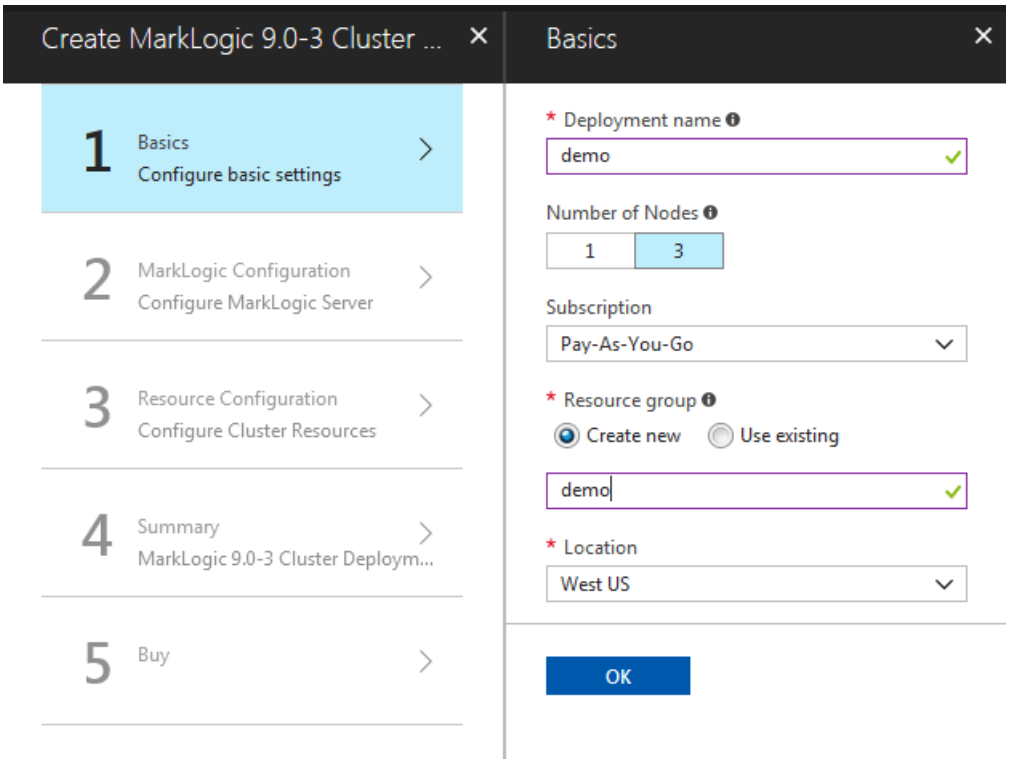


2. Click “Create” at the bottom of the page to be directed to the portal for creating the cluster.

2.3.2 Configure Basic Settings

For these steps, you will enter the appropriate information for your cluster. The highlighted fields contain the default values. See the screenshots for details.

1. Step one is to configure some basic settings like deployment name and number of nodes.



2. The default field values that are highlighted can be changed. We recommend that you leave the defaults as is for this exercise. Click OK.

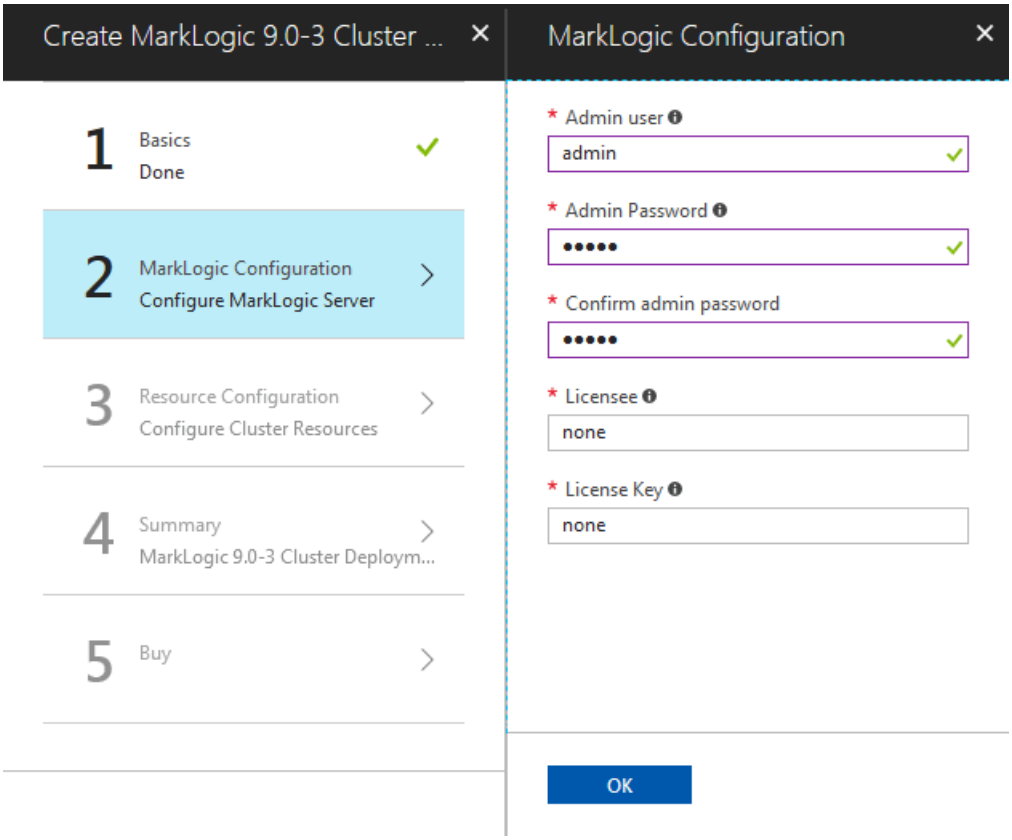
Note: Each time you click OK, a validation program runs in the background to validate the entries in the fields. You cannot move to the next screen until the values have be validated. Validation is indicated by a green checkmark next to the step.

2.3.2.1 Basic Information Values

Field	Value
Deployment Name	*Required*. The value to be used as a prefix for all of your resources. There is a 10 character limit for the Deployment Name. Any blank spaces will be removed.
Number of Nodes	1 or 3. Default is 3.
Subscription	*Required by Azure*. It is the Azure subscription to which the cluster usage will be billed.
Resource Group	*Required by Azure*. The name of the resource group to which the cluster will belong. Options are Create new or Use existing. A name for the group is required.
Location	*Required* . Choose a physical location for your resource group.

2.3.3 MarkLogic Configuration

1. Fill in the fields for the MarkLogic Admin user and password, licensee, and license key. See the section on “Password Policy” on page 14 for details on creating a password.



2. Click OK.

2.3.3.1 Password Policy

Azure requires that when you use the MarkLogic Solution template to deploy MarkLogic on Azure, the admin password policy must be stronger than the default MarkLogic policy. Because of this, MarkLogic enforces a different admin password policy on Azure.

The regular expression to validate the policy is:

```
^(?=.*[A-Z])(?=.*[.!@#$%^&()-_=+])(?=.*[0-9])(?=.*[a-z]).{12,40}$
```

This means that your password must be 12-40 characters long and contain at least one uppercase letter, a digit, and a special character - one of `.!@#$%^&()-_=+`.

You can change your password later after the initial set up. When you change your password, the MarkLogic password policy will be used to validate the new password.

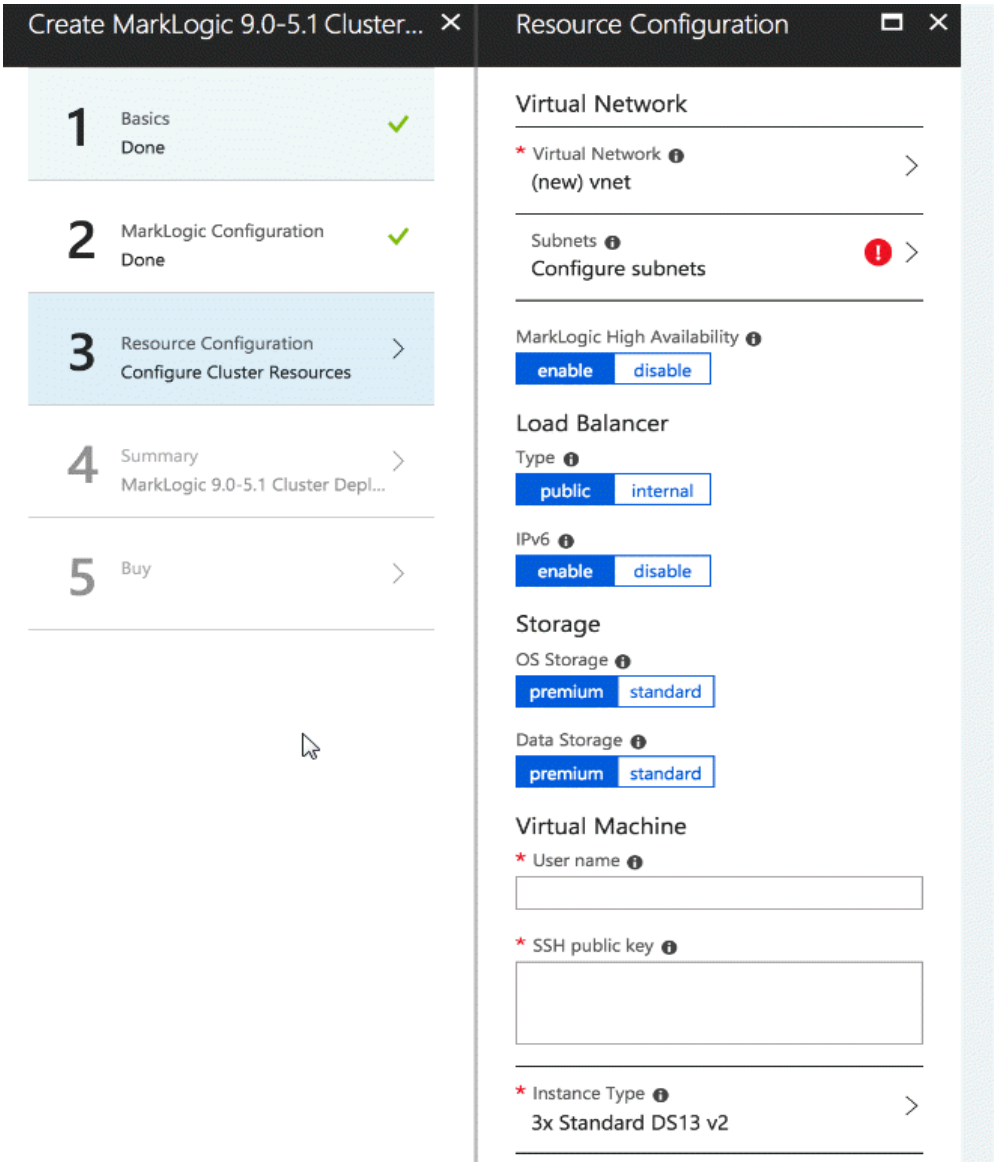
2.3.3.2 MarkLogic Configuration Values

Field	Value
Admin User	*Required*. The MarkLogic administrator username.
Admin Password	*Required*. The MarkLogic administrator password.
Licensee	*Optional*. The MarkLogic licensee. Use “none” for no license.
License Key	*Optional*. The MarkLogic license key. Use “none” for no license key.

Note: When you provide a license key and a licensee, the template will deploy clusters using BYOL image.

2.3.4 Configure Cluster Resources

1. Configure your cluster resources. The default fields are highlighted.



2. Click OK to continue. See “Configure Cluster Resources Values” on page 17 for additional information.

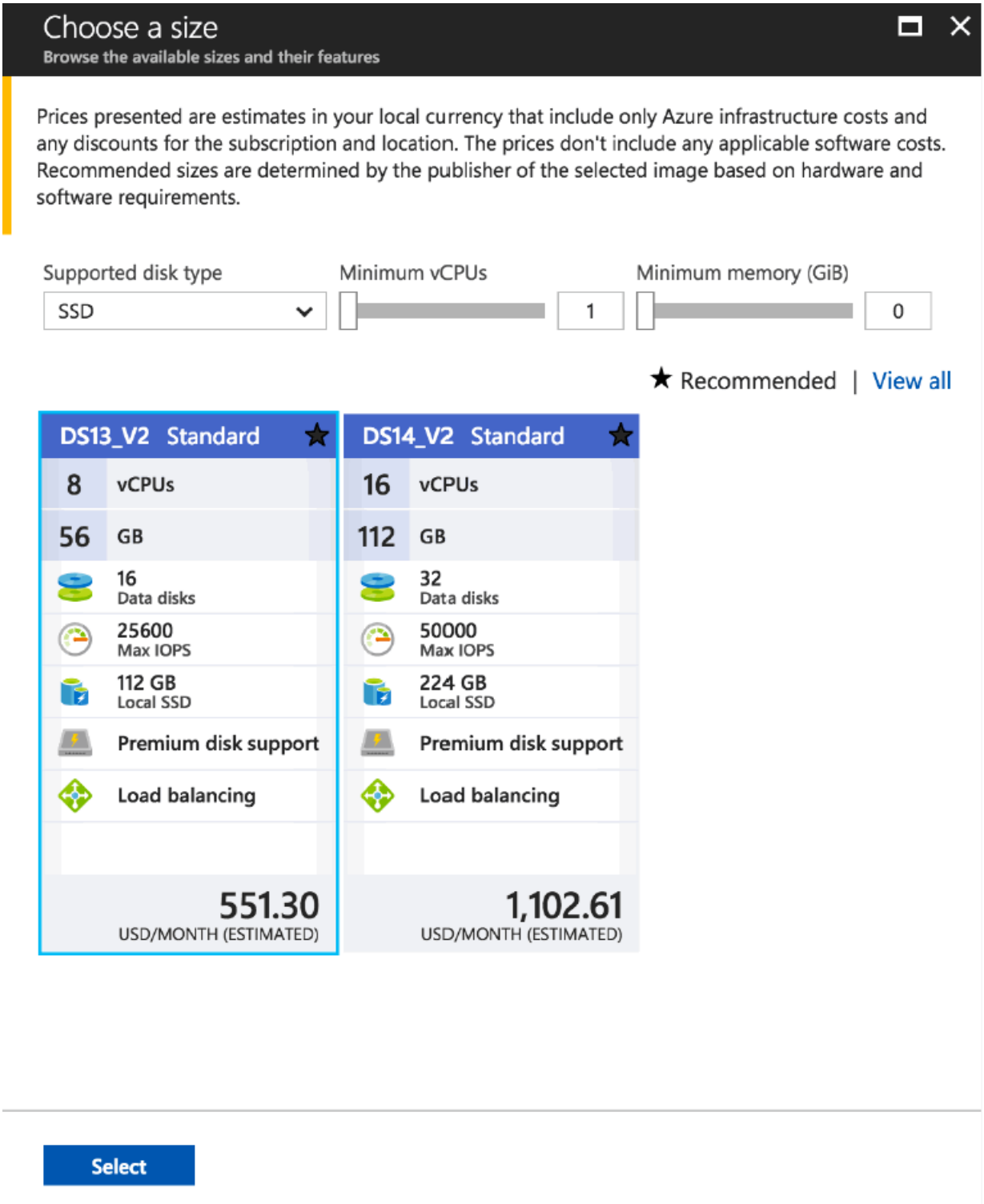
2.3.4.1 Configure Cluster Resources Values

Field	Value
MarkLogic High Availability	*enable* or *disable*. Default is *enable*. This option is only applicable to a multi-node cluster. For a single node cluster, high availability will not be configured. When this option is set to *enable*, local disk failover will be configured for all database forests initialized with MarkLogic. Master forests will be configured on first node coming up in the clusters. The second node will be configured with replica forests. See “Typical Architecture” on page 7 for an example of a forest topology.
Virtual Network	Virtual Network for MarkLogic cluster
Subnets	Subnets for MarkLogic.
Load Balancer: Type	*public* or *internal*. Default is Public Load Balancer
Load Balancer: IPv6	*enable* or *disable*. Default is enable. IPv6 address on the load balancer. Only applicable if load balancer is public.
Storage: OS Storage	*premium* or *standard*. Default is *premium*. The storage type for the operating system of the virtual machines.
Storage: Data Storage	*premium* or *standard*. Default is *premium*. Storage type for data directory of virtual machines.
Virtual Machine: User name	*Required*. Operating system username for virtual machines.
Virtual Machine: SSH public key	*Required*. Public SSH key for the virtual machine listed.
Instance Type	*Required*. Type of virtual machine instance to launch. The list only includes instance types that meet the minimum standard requirements for MarkLogic Server.

2.3.5 Choose a VM Size

All of the instance types displayed meet the minimum MarkLogic requirements. Other options have been filtered out for you. The choices will be displayed based on your prior selections.

1. Choose a VM size. You can click View all to see all of the available options. Choose an instance type by clicking on it.

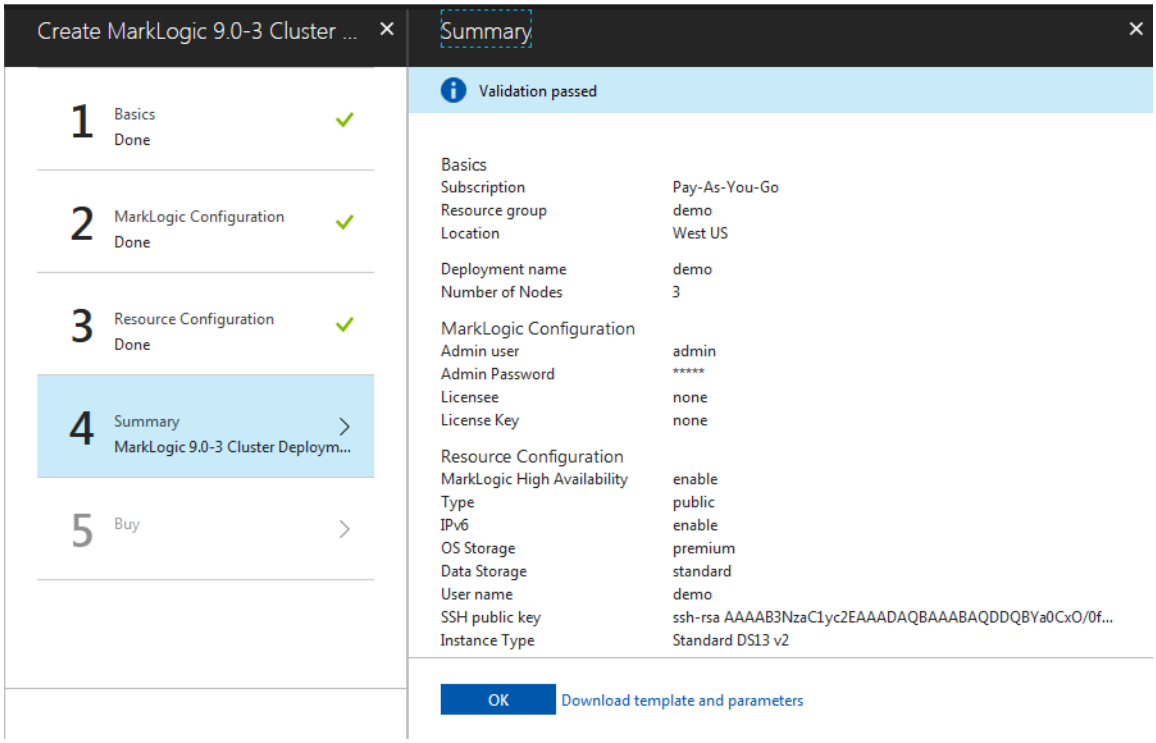


Note: Choosing premium storage will limit the set of options to select from.

2. Click Select to contine.

2.3.6 Review the VM Summary

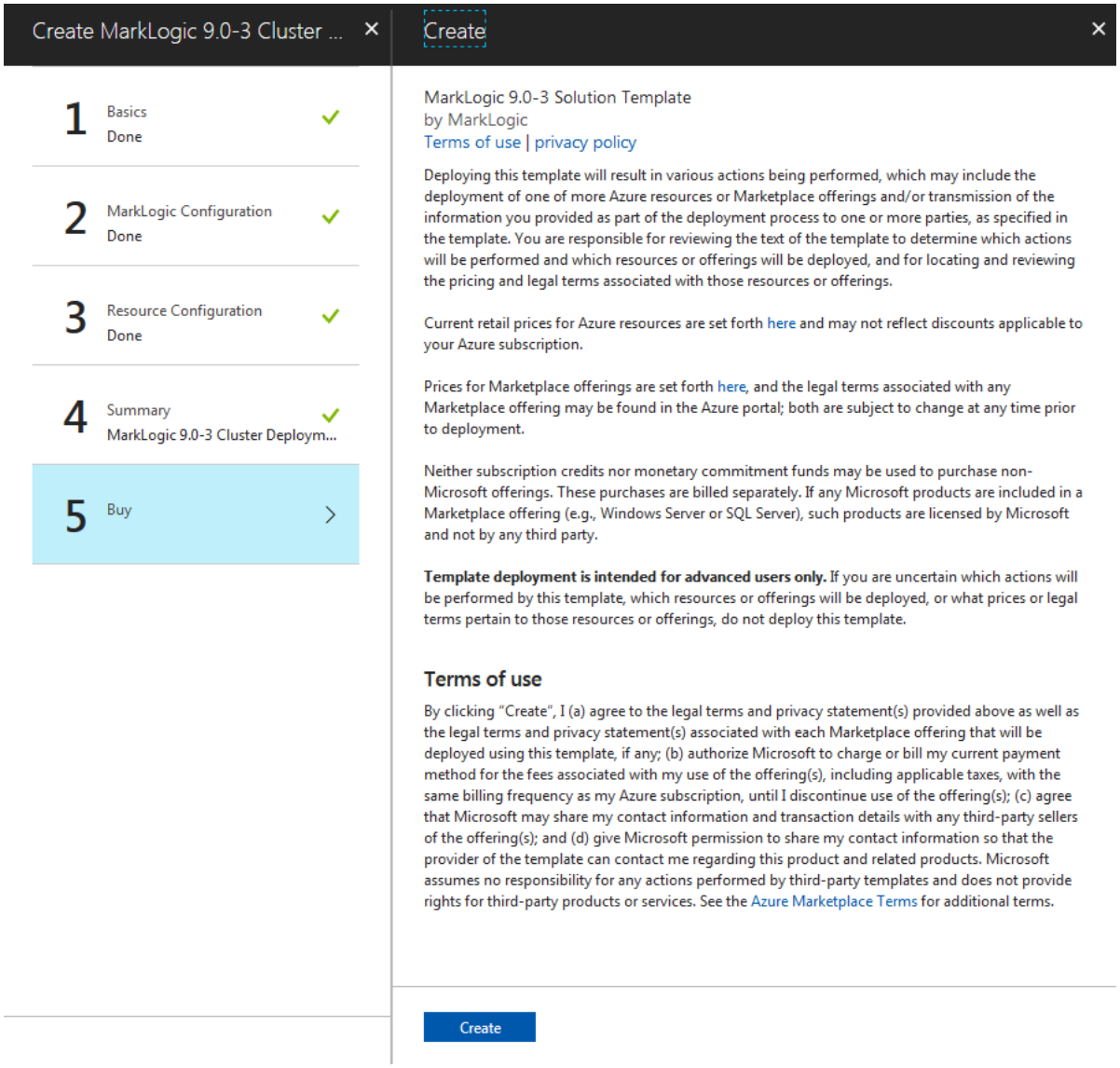
1. Edit details if you need to modify anything, before proceeding to the next step. If you need to make changes, click on the tabs at the left to go back and edit your choices.



2. Review the summary details and click OK.

2.3.7 Create and Purchase

The next screen explains the terms of use and privacy policy. After reading the information and agreeing to these terms, click Create to create your MarkLogic cluster on Azure.



After you click create, the Azure VM will start up. The start up process can take a few minutes.

2.4 Resource Configuration

This section describes the configuration of cluster resources pre-defined in the template. These are fields that you fill in when you use the Solution Template to set up a cluster. These fields contain a subset of all of the configurable parameters for the resources associated with the cluster. The field names used in the Solution Template are not a one-to-one mapping of the Azure Resource configuration. In addition, for simplicity some of the configuration options are not included here (such as name of the VM instances, load balancer, availability set, and so on).

These topics are covered in this section:

- Virtual Network
- Availabilty Set
- Node Public IP Address
- Network Security Group
- Load Balancer
- Data Storage: Managed Disks
- Virtual Machines
- Cluster Initialization

2.4.1 Virtual Network

For the virtual network, specify the following information:

Field	Value
Address Space	10.0.0.0/16
Subnet Address Range	10.0.1.0/24

2.4.2 Availabilty Set

Your VMs are placed into a logical grouping called an *availability set*. When VMs are created in an availability set, Azure distributes the placement of the VMs across the infrastructure. Availability Sets ensure that at least one VM remains running during planned or unplanned events.

Field	Value	Description
Fault Domains	3	Defines the group of virtual machines that share a common power source and network switch.
Update Domains	20	Indicate groups of virtual machines and underlying physical hardware that can be updated at the same time.
Use Managed Disk	Yes. This is the recommended setting and is not configurable.	Handles storage for you.

See also “Availability Set” on page 25 for about the limitations of Availability Sets.

2.4.3 Node Public IP Address

By default, Public IP addresses are dynamic, so they can change when the VM is deleted. To guarantee that a VM always uses the same public IP address, create a static Public IP.

Field	Value
Public IPv4 Allocation Method	Static
Public IPv6 Allocation Method	Dynamic
Idle Timeout (minutes)	4 - Default value set by Azure

2.4.4 Network Security Group

These are the applicable security rules for the cluster for allowed access, inbound.

Usage/Name	Protocol	Source Port Range	Source Address Prefix	Destination Port Range
SSH	tcp	*	*	22
Admin	tcp	*	*	8000-8010
Health-Check	tcp	*	*	7997
Communication	tcp	*	*	7778-7999

2.4.5 Load Balancer

It is a good practice to use a network load balancer between the client applications and a MarkLogic deployment. Depending on your deployment topology, you may use either an internet-facing load balancer or an internal load balancer.

An internet-facing load balancer should be used when a client application needs to access a MarkLogic deployment using public IP addresses. You should also consider network security in Azure for this type of deployment.

An internal load balancer should be used when the client application accesses a MarkLogic deployment using internal IP addresses, or the client application runs on premises and a secure VPN connection is established between the two networks.

The load balancer detects proper running of MarkLogic via the HealthCheck App Server on port 7997 and will only direct traffic to that node if it has verified that the MarkLogic instance is up. Therefore, for HTTP, the Load Balancer Probe in Azure for MarkLogic is on port 7997.

Field	Value	Description
Load Balancer	Internal or Public	Default is Public Load Balancer
Load Balancer IPv6	Enable or Disable	Default is Enable. IPv6 address on the load balancer. Only applicable if the load balancer is public

2.4.5.1 Public IP Address Allocation

Field	Value
Public IPv4 Address Allocation Method	*Static*. Only applicable to public load balancer.
Public IPv6 Address Allocation Method	*Dynamic*. Only applicable to public load balancer.

2.4.5.2 Load Balancing Rules

Protocol	Frontend Port	Backend Port	Idle Timeout (in minutes)
tcp	8000	8000	5
tcp	8001	8001	5
tcp	8002	8002	5
tcp	8003	8003	5
tcp	8004	8004	5
tcp	8005	8005	5
tcp	8006	8006	5
tcp	8007	8007	5
tcp	8008	8008	5

2.4.5.3 Health Probes

Field	Value
Port	7997
Interval (in seconds)	5
Number of Probes	2

2.4.6 Data Storage: Managed Disks

Azure recommends using Managed Disks for your virtual machine data. Managed Disks handle storage for you behind the scenes, while providing better reliability for Availability Sets.

Field	Value
Create Option	Empty
Size (GiB)	1023

2.4.7 Virtual Machines

Azure virtual machines enable you to deploy virtually any workload and any language on nearly any operating system.

Field	Value
Boot diagnostics	Enabled
Guest OS diagnostics	Disabled

2.5 Cluster Initialization

The template will initialize all the cluster nodes with information provided by the user, and then start the cluster.

> **Note:** MarkLogic will mount the VM disk device `/dev/sdc` to the MarkLogic data directory.

On a multi-node cluster, if High Availability is enabled, the template will configure local-disk failover for forests initialized with MarkLogic Server in the cluster. The replica forests for App-Services, Documents, Extensions, Fab, Last-Login, Meters, Modules, Schemas, Security, and Triggers will be configured on a node other than the bootstrap node (the first node initialized in the cluster). The third node in a three node cluster, will have no forests at the start.

2.6 Limitations

This section includes limitations in Azure.

2.6.1 Availability Set

Azure Availability Set is a logical group for virtual machines. Microsoft Azure SLA guarantees that at least one of the (two or more) nodes in Availability Set is available 99.95% of the time. The Availability Set parameter is only applicable to a multi-node cluster. For a one-node cluster, the Availability Set parameter will be disabled.

See the Microsoft Azure SLA for your type of deployment for more information.

The configuration of an Availability Set includes the number of update domain and number of fault domain. The combination of the two domains can only guarantee that one of the nodes in Availability Set is available most of the time. You cannot configure an Availability Set that guarantees two of three nodes of a cluster will be available most of the time. For more information on Availability Set configuration, see the Azure documentation (https://docs.microsoft.com/en-us/azure/virtual-machines/linux/manage-availability)

2.6.2 Shared Disk Failover

Users have the option to have High Availability (HA) configured for database forests initialized with MarkLogic. However only local disk failover is possible because Azure does not support mounting a managed disk for multiple machines. An alternative for managed disk is Azure File Storage - a shared storage service. The performance of the File Storage is not comparable to managed disk for mounting to virtual machines.

2.7 List of Configurable Parameters

The following tables contain fields from the set up example, for the configurable parameters in the MarkLogic Solutions template:

These topics are covered in this section:

- Basic Information Values
- Azure Configuration Values
- Configure Cluster Resources Values

2.7.1 Basic Information Values

Field	Value
Deployment Name	*Required*. The value to be used as a prefix for all of your resources. There is a 10 character limit for the Deployment Name. Any blank spaces will be removed.
Number of Nodes	1 or 3. Default is 3.
Subscription	*Required by Azure*. It is the Azure subscription to which the cluster usage will be billed.
Resource Group	*Required by Azure*. The name of the resource group to which the cluster will belong. Options are Create new or Use existing. A name for the group is required.
Location	*Required.* Choose a physical location for your resource group.

2.7.2 Azure Configuration Values

Field	Value
MARKLOGIC_LICENSE_KEY	A license key to use for this MarkLogic instance. This license key is only valid for a Bring Your Own License (BYOL) Image or a user-created Image.
MARKLOGIC_LICENSEE	The Licensee corresponding to MARKLOGIC_LICENSE_KEY.
MARKLOGIC_NODE_NAME	A distinct name of a node within a cluster.
MARKLOGIC_ADMIN_USERNAME	The MarkLogic Administrator username used for initial installations.
MARKLOGIC_ADMIN_PASSWORD	The MarkLogic Administrator password used for initial installations.
MARKLOGIC_AZURE_DISK	The LUN numbers of the disks to use for the MarkLogic data directories. The disks will be mounted, and a file system will be created on each one, if needed. A single disk is acceptable, in which case, just enter the LUN number for that disk. When multiple disks are to be mounted, enter a list of integers, delimited by comma: Example: 4,5,6 If left undefined, the default is: 4.
MARKLOGIC_AZURE_STORAGE_PROXY	The URL of the proxy server used by the group to access Azure blob storage. If MARKLOGIC_AZURE_STORAGE_PROXY is set and azure storage proxy is not set in the Admin Interface group configuration, the value of MARKLOGIC_AZURE_STORAGE_PROXY is used . If MARKLOGIC_AZURE_STORAGE_PROXY is set and azure storage proxy is also set in the Admin Interface group configuration, the Admin Interface azure storage proxy setting is used.

2.7.3 Configure Cluster Resources Values

Field	Value
MarkLogic High Availability	*enable* or *disable*. Default is *enable*. This option is only applicable to a multi-node cluster. For a single node cluster, high availability will not be configured. When this option is set to *enable*, local disk failover will be configured for all database forests initialized with MarkLogic. Master forests will be configured on first node coming up in the clusters. The second node will be configured with replica forests. See “Typical Architecture” on page 7 for an example of a forest topology.
Load Balancer: Type	*public* or *internal*. Default is Public Load Balancer
Load Balancer: IPv6	*enable* or *disable*. Default is *enable*. IPv6 address on the load balancer. Only applicable if load balancer is public.
Storage: OS Storage	*premium* or *standard*. Default is *premium*. The storage type for the operating system of the virtual machines.
Storage: Data Storage	*premium* or *standard*. Default is *premium*. Storage type for data directory of virtual machines.
Virtual Machine: User name	*Required*. Operating system username for virtual machines.
Virtual Machine: SSH public key	*Required*. Public SSH key for the virtual machine listed.
Instance Type	*Required*. Type of virtual machine instance to launch. The list only includes instance types that meet the minimum standard requirements for MarkLogic Server.

3.0 Managing MarkLogic on Azure

This chapter describes how to manage, scale, and monitor MarkLogic Servers in the Azure environment. This section contains the following topics related to managing and scaling MarkLogic on Azure:

- Accessing Your MarkLogic Azure VM
- Attaching Data Disks
- Partitioning and Mounting Disks
- Loading Content into MarkLogic
- Configuring MarkLogic for Azure Blob Storage
- Adding a Node to the Existing Cluster
- Removing a Node in the Existing Cluster
- Resizing Virtual Machines
- Monitoring Virtual Machine Status
- Monitoring MarkLogic Server
- Upgrading MarkLogic Server on Azure

3.1 Accessing Your MarkLogic Azure VM

After the Azure VM is up and available, navigate to the “Virtual Machines” tab on your Azure portal and click on the VM you just created. You can find the IP address there for accessing the MarkLogic Admin Console at port 8001.

> **Note:** In the MarkLogic Azure image, the following ports are open by default: 8000, 8001, 8002, 7997.

3.2 Attaching Data Disks

To preserve your data, you must attach a persistent data disk to your Azure VM, in case of VM re-provisioning. Microsoft recommends using managed disks over unmanaged disks for new VMs.

The steps for attaching data disks to your VM from the Azure Portal can be found here:

> https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-linux-attach-disk-portal/

After you go through the steps on the portal, follow the instructions under the section “Initialize a new data disk in Linux” from the following Azure tutorial:

> https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-linux-classic-attach-disk/#how-to-initialize-a-new-data-disk-in-linux

Microsoft Azure Storage is the back-end to store VM disks and has two main SKUs, with different performance characteristics:

- The Premium storage SKUs provide high-performance, low-latency disk support for I/O intensive workloads.
- The Standard storage based SKUs offer a better price/GB ratio.

Not all VMs support attaching Premium Storage. In addition, Microsoft Azure provides unmanaged and managed disks in each of the two performance tiers. Azure recommends using managed disks for new VMs and converting your previous unmanaged disks to managed disks, to take advantage of the many features available in Managed Disks.

The MarkLogic Tiered Storage feature enables you to take advantage of many Azure Storage types. For more information, see Tiered Storage in the *Administrator's Guide*.

Place your most frequently used data on Premium Storage data disks for faster access. Place your less frequently used data on Standard Storage data disks at a lower cost. If you are considering leveraging the MarkLogic Tiered Storage capabilities with different Azure Storage types, please read the following discussion on scalability and performance targets:

https://docs.microsoft.com/en-us/azure/storage/common/storage-scalability-targets

3.3 Partitioning and Mounting Disks

To use data disks after they have been attached, you must also mount the data disk. The steps for partitioning and mounting data disks can be found here:

https://docs.microsoft.com/en-us/azure/virtual-machines/linux/add-disk

3.4 Loading Content into MarkLogic

To load contents into your MarkLogic server in Azure, refer to the *Loading Content Into MarkLogic Server Guide*.

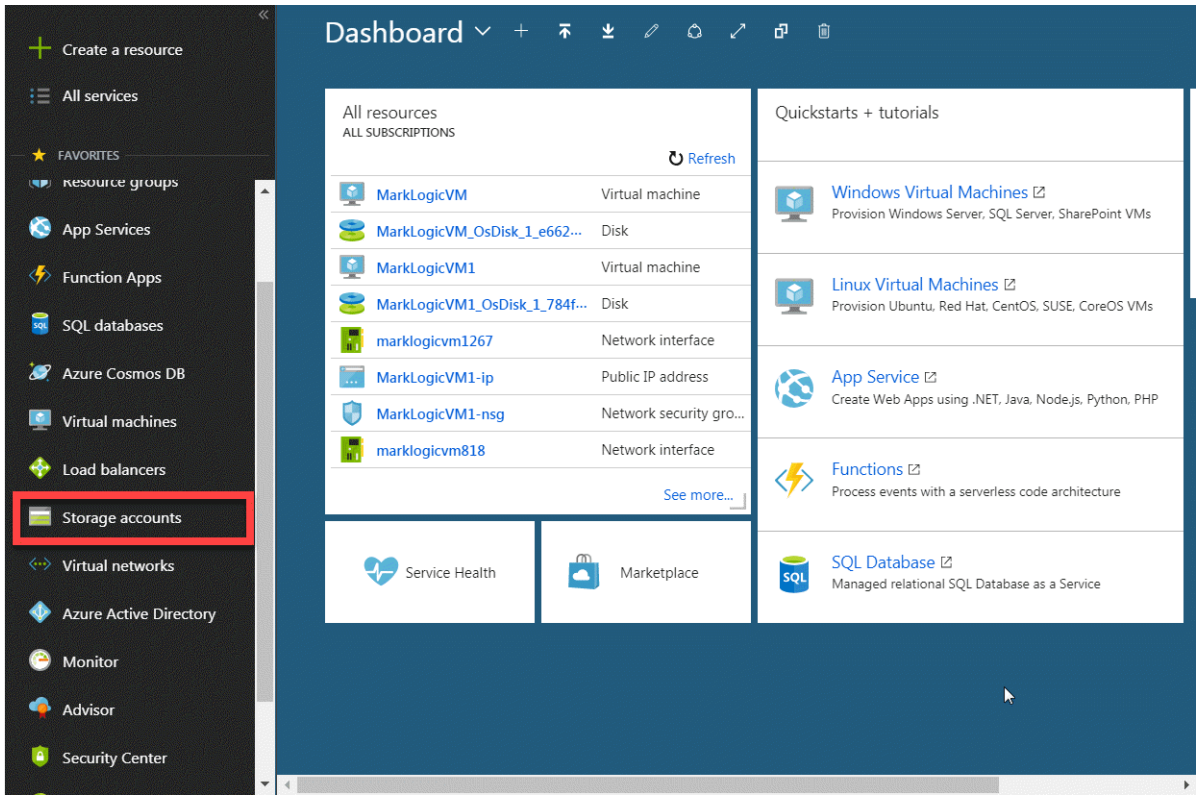
3.5 Configuring MarkLogic for Azure Blob Storage

Azure Blob Storage is similar to Amazon S3, even better in some ways, because it doesn't have S3's read-after-write eventual consistency limitation and because it also supports the ability append, so journal files work on it.

In MarkLogic, Azure pathnames are of the form: azure://mycontainer/mydirectory/myfile. They are similar to MarkLogic S3 pathnames of the form: S3://mybucket/mydirectory/myfile. What are called "buckets" on S3 are called "containers" on Azure.

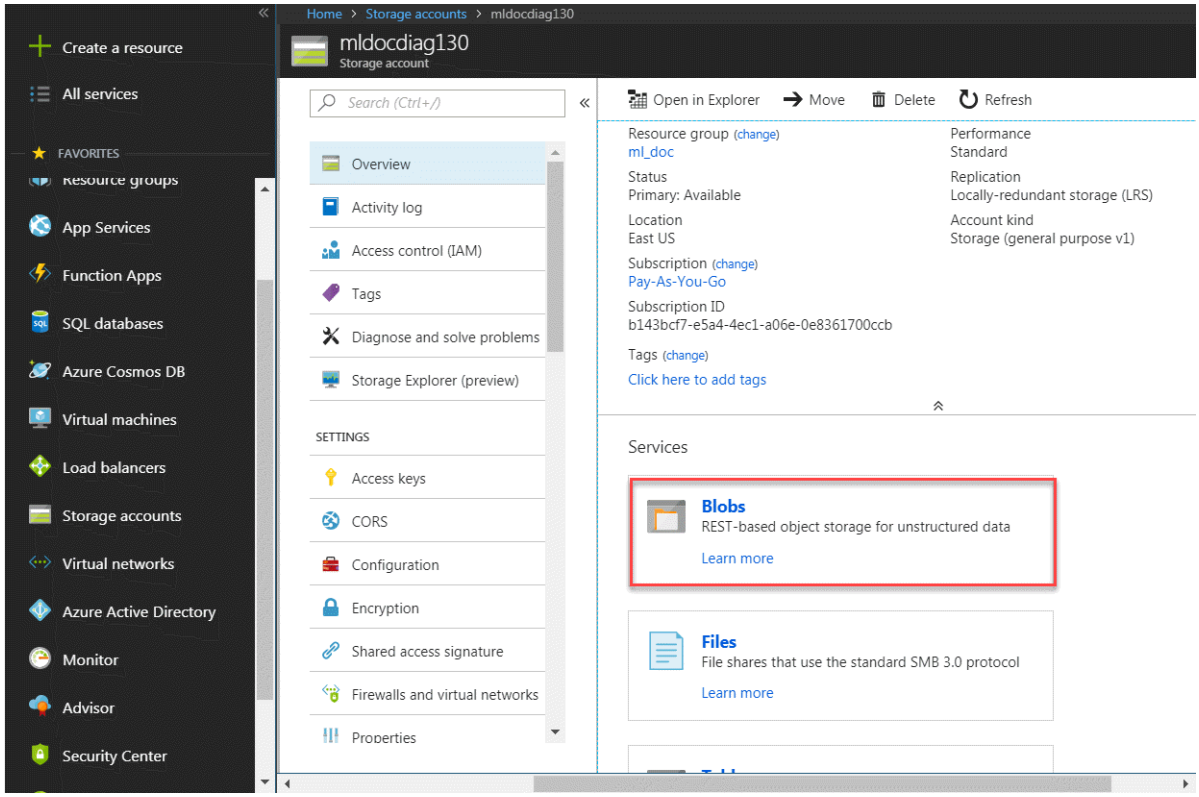
To use Azure Blob Storage, you need an Azure storage account, and you need a container for your blobs. You can create and manage Azure storage accounts on the Azure portal: https://portal.azure.com/

After logging into the portal, click on "Storage accounts" on the left navigation area. From there you can manage an existing storage account or add a new one. Storage account names are global, so Microsoft provides guidance how to name them: https://docs.microsoft.com/en-us/azure/architecture/best-practices/naming-conventions#naming-rules-and-restrictions



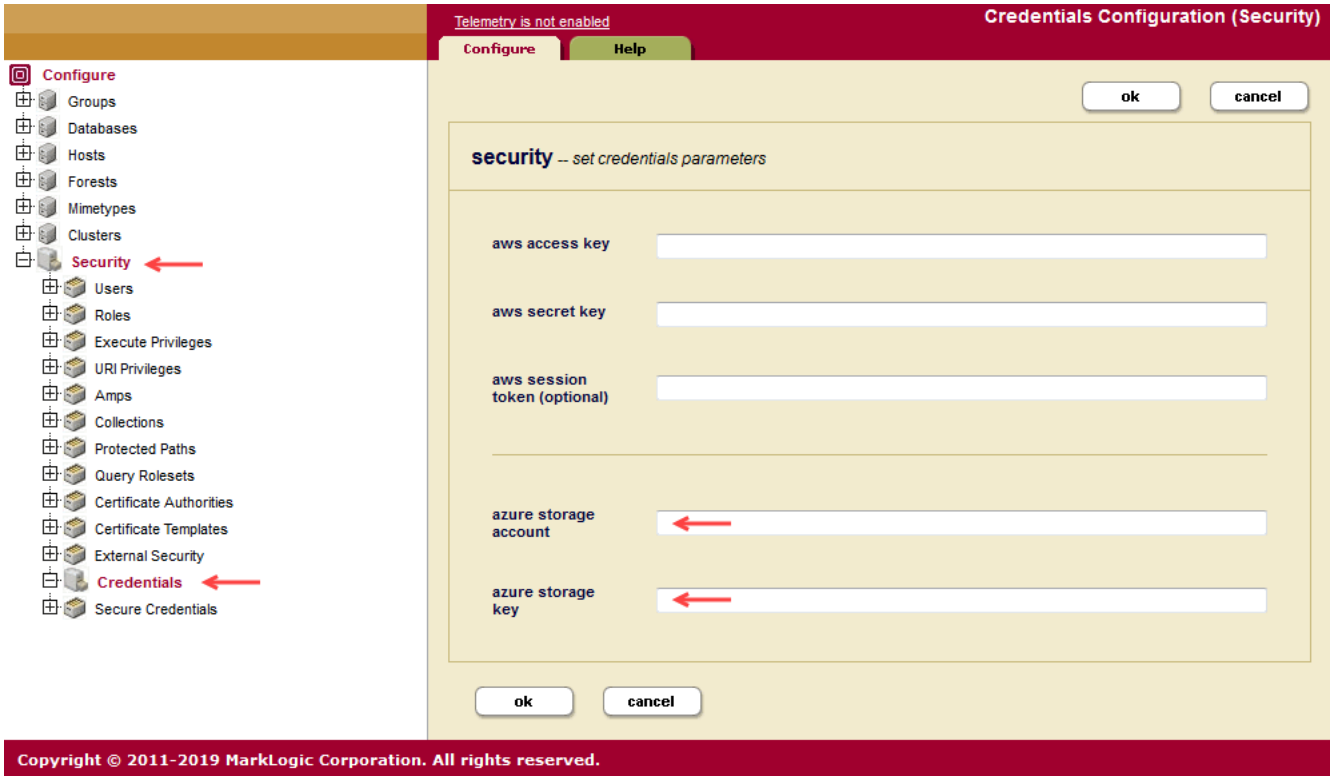
Within your storage account you also need to have at least one container for your blobs. From there your can manage an existing accound or add a new onw. On the storage account management page, click "Blobs" to get to the containers page. From there you can manage an existing container or add a new one.

Also on the Storage Account Management page is an “Access keys” link. In a bit, you will need to click this link, so keep the browser open to your Storage Account Management page.

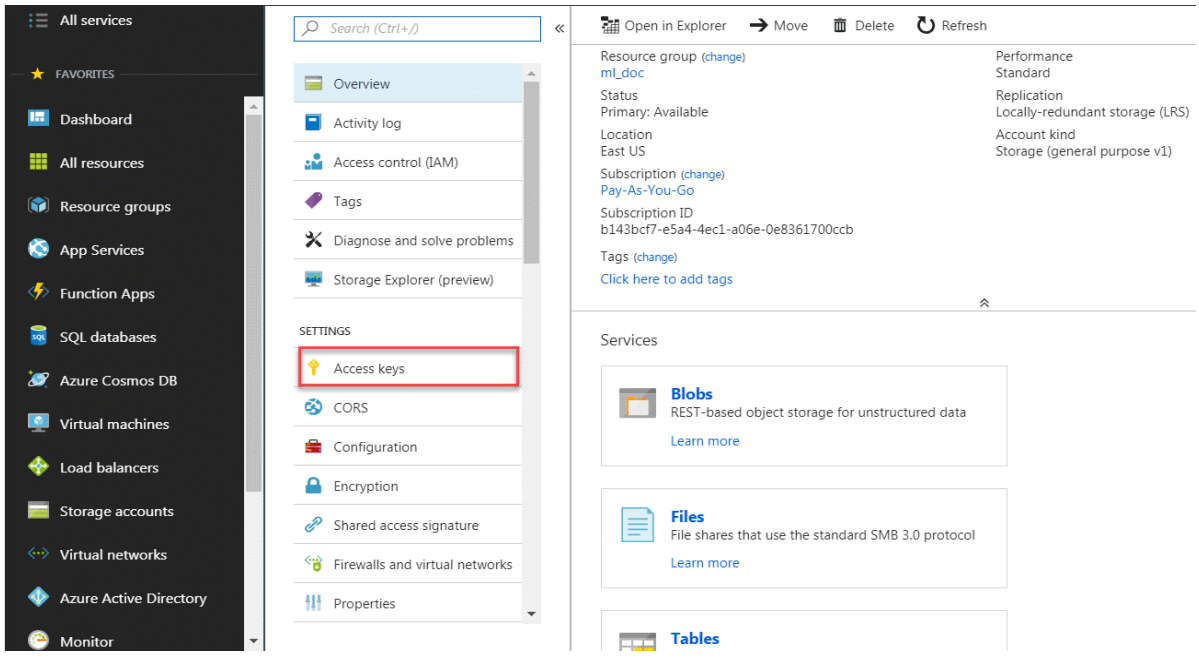


MarkLogic needs Azure storage credentials to access Azure Blob Storage. To configure Azure storage credentials in MarkLogic, in the Admin GUI, navigate to Security > Credentials. On that page, right below where you would enter AWS credentials for S3, you can enter your Azure storage credentials.

Note: You may also use your Azure VM identity to access Azure Blob Storage. Please see this web tutorial to learn how to create a storage account and a VM with identity on Azure.



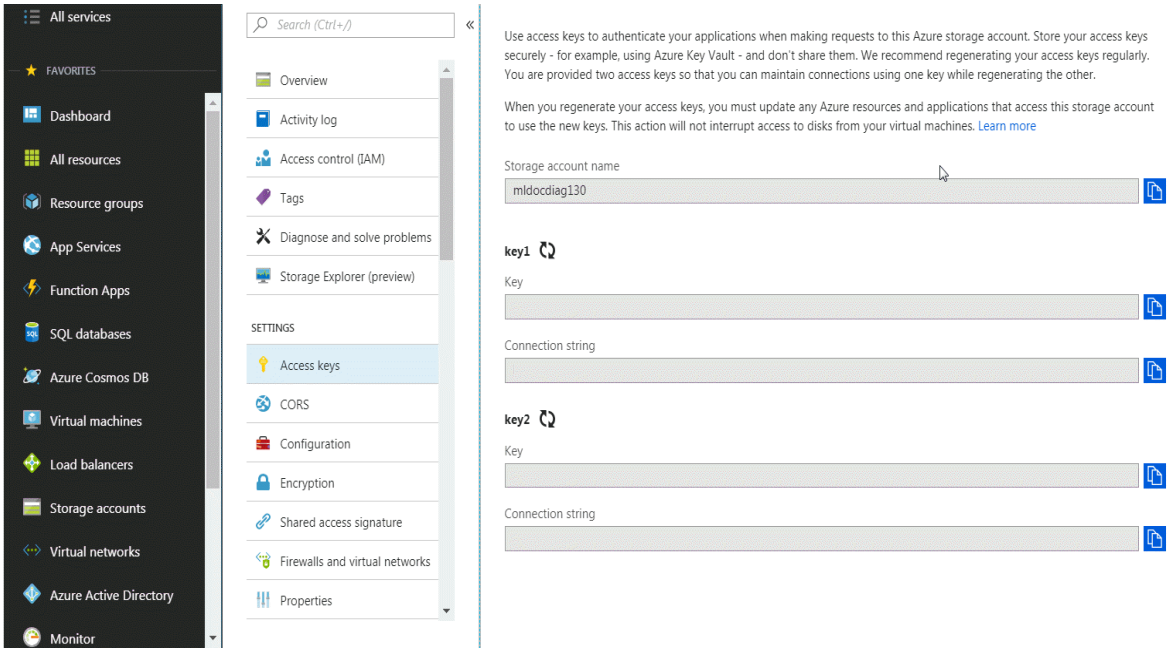
Enter your Azure storage account and your Azure storage key. Your Azure storage account is the name of your storage account. Your Azure storage key is an access key automatically generated by Microsoft and provided to you. You can get your Azure storage key from the Storage Account Management page of the Azure portal.



On the Storage Account Management page of Azure the portal, click on the “Azure keys” link. You an use either the “Key” value for key1 or the “Key” value for key2; it doesn’t matter which you use.

The easiest way to do this is to have two open browser windows, one open to the MarkLogic Admin GUI, and one open to the Azure Portal. Then you can just copy the value from one browser window and paste it into the other.

Azure storage access keys are like passwords, so they should be treated as passwords. MarkLogic stores the storage access key in encrypted form in its Security database.



Once you have configured your Azure storage credentials in MarkLogic, you can access your containers as file systems. The filesystem’s built-in functions work with Azure pathnames, which are in the the form `azure://mycontainer/mydirectory/myfile`. External binary references can have Azure pathnames and MarkLogic will transparently access them.

You can do easily do backups to an Azure pathname, and you can also do journal archiving to Azure. You can specify an Azure directory as a forest data directory. Forests on Azure blob storage can have journals. Forests on Azure blob storage can be transactionally and robustly read/write, instead of just read-only as they are on S3.

There are security API functions to script the configuration of Azure storage credentials: `sec:credentials-get-azure` and `sec:credentials-set-azure`. The REST management API endpoint `/manage/v2/credentials/properties` will support credentials configuration.

If there are no Azure credentials in the Security database, MarkLogic will check environment variables. See “Environment Variables” on page 36 for more information.

If you want to configure MarkLogic Server to access Azure blob storage through a proxy server, you can specify the URL to the proxy server by setting the `MARKLOGIC_AZURE_STORAGE_PROXY` in the MarkLogic Azure Solution Template, or use the Admin Interface to configure MarkLogic Server to access S3 Storage through a proxy server, as follows:

1. Log into the Admin Interface.
2. Click the Groups icon on the left tree menu.
3. Click the Configure tab at the top right.
4. Locate the group for which you want to view settings.
5. Click the icon for this group.
6. Enter the URL to the Azure Blob Storage proxy server in the `azure storage proxy` field toward the bottom of the configuration page. The proxy URL should start with `https://` (for example, `https://proxy.marklogic.com:8080`). If you don't specify the port number, MarkLogic assumes the proxy server is listening on port `8080`.



Note: If the `MARKLOGIC_AZURE_STORAGE_PROXY` variable is set and `azure storage proxy` is not set in the Admin Interface group configuration, the value of `MARKLOGIC_AZURE_STORAGE_PROXY` is used . If `MARKLOGIC_AZURE_STORAGE_PROXY` is set and `azure storage proxy` is also set in the Admin Interface group configuration, the Admin Interface `azure storage proxy` setting is used.

3.5.1 Credentials Configuration

There are security API functions to script the configuration of Azure storage credentials, specifically the functions `sec:credentials-get-azure` and `sec:credentials-set-azure`, which are a part of `security.xqy`. The REST management API endpoint `/manage/v2/credentials/properties` will support credentials configuration.

3.5.2 Environment Variables

If there are no Azure credentials in the Security database, MarkLogic will check environment variables. You can specify your Azure storage account name with the environment variable `MARKLOGIC_AZURE_STORAGE_ACCOUNT`. You can specify your Azure storage access key with the environment variable `MARKLOGIC_AZURE_STORAGE_KEY`. In Linux, you can set these MarkLogic environment variables with export commands in `/etc/marklogic.conf`.

> **Note:** Since Azure storage access keys are like passwords, it's better to have MarkLogic keep them encrypted in its Security database, but this alternate mechanism is available if you need it.

3.6 Adding a Node to the Existing Cluster

To add more VMs as additional hosts for MarkLogic Clusters on Azure platform, set up the number of VM instances as needed from your Azure portal. For more details, see Adding a Host to a Cluster in the *Administrator's Guide*.

Since the IP address for each node changes as you start and stop an instance on Azure, make sure a public DNS name is assigned to every node before stopping an instance so that this node can join the cluster again.

MarkLogic recommends assigning a public DNS name to your VMs. For instructions on how to do this, see the Microsoft Azure documentation.

> https://azure.microsoft.com/en-us/documentation/articles/virtual-machines-linux-portal-create-fqdn

3.7 Removing a Node in the Existing Cluster

For details on removing nodes, see Leaving the Cluster in the *Administrator's Guide*.

3.8 Resizing Virtual Machines

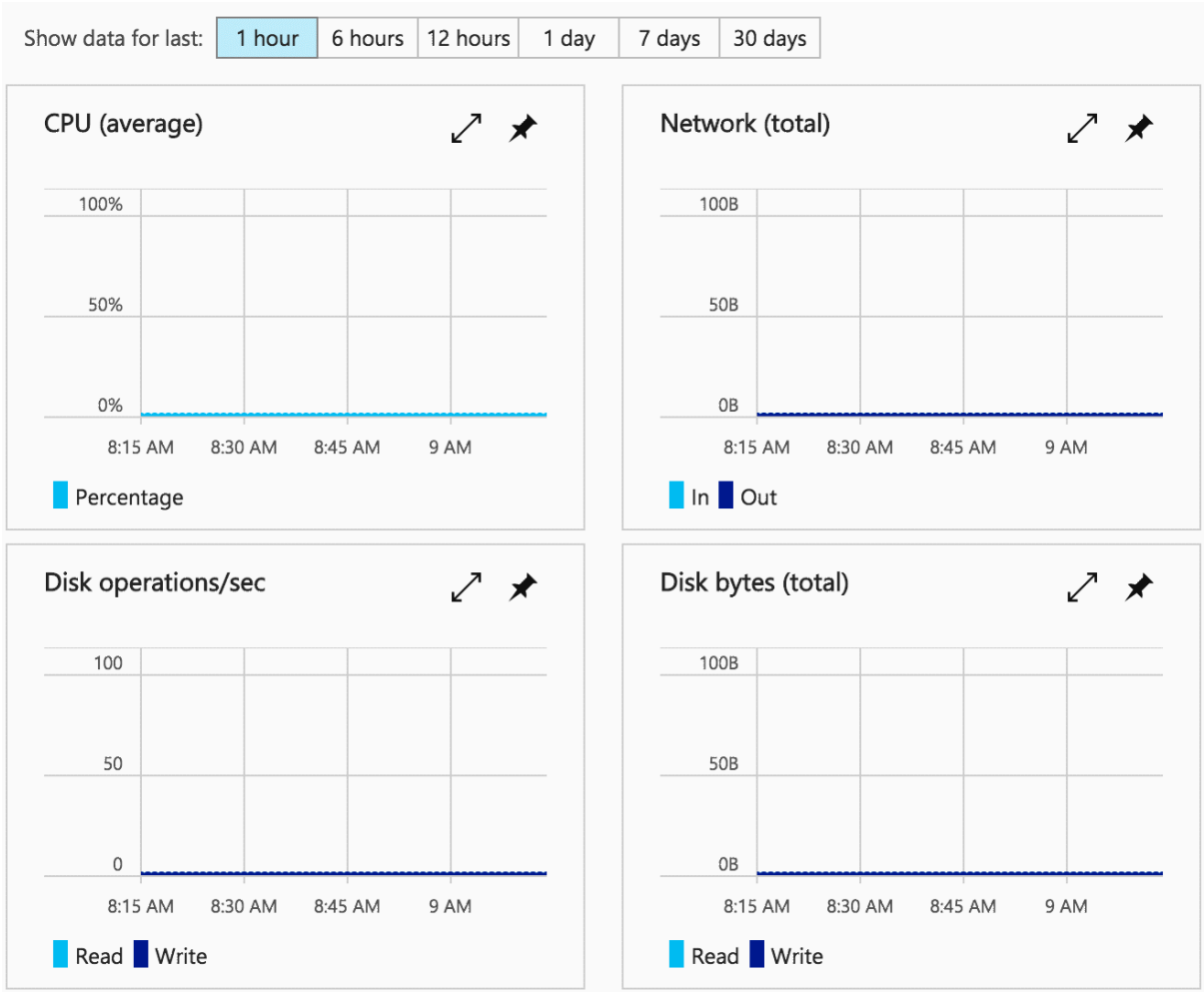
See the following Microsoft Azure blog for recommendations on changing VM instance sizes:

> https://azure.microsoft.com/en-us/blog/resize-virtual-machines/

You should expect some downtime when resizing a VM. In addition, data stored on temporary OS disk will be lost during resizing.

3.9 Monitoring Virtual Machine Status

Azure has built-in single VM monitoring, and you will be able to see the VM status. You can select which performance metrics you want to see. The VM status display is similar to the following:



3.10 Monitoring MarkLogic Server

MarkLogic provides a monitoring dashboard and REST Management APIs that you can use to integrate with other monitoring tools. For details, see Monitoring MarkLogic Server in the *Monitoring MarkLogic Guide*.

Plugins for New Relic and App Dynamics are also available. The Open Source, community supported New Relic plugs in can be found in the tools section on www.developer.marklogic.com.

3.11 Upgrading MarkLogic Server on Azure

The Azure Solution Template does not support updating nodes. MarkLogic Server on Azure must be upgraded manually.

To upgrade a MarkLogic cluster on Azure:

1. Detach the data disks:
   https://docs.microsoft.com/en-us/azure/virtual-machines/windows/detach-disk.

2. Create new VMs:
   https://docs.microsoft.com/en-us/azure/virtual-machines/windows/quick-create-portal.

3. Attach and mount the data disks:
   https://docs.microsoft.com/en-us/azure/virtual-machines/linux/attach-disk-portal.

4. Join the upgraded MarkLogic node to the cluster, as described in Adding Another Node to a Cluster in the *Scalability, Availability, and Failover Guide*.

4.0 Technical Support

MarkLogic provides technical support according to the terms detailed in your Software License Agreement or End User License Agreement.

We invite you to visit our support website at http://help.marklogic.com to access information on known and fixed issues, knowledge base articles, and more. For licensed customers with an active maintenance contract, see the Support Handbook for instructions on registering support contacts and on working with the MarkLogic Technical Support team.

Complete product documentation, the latest product release downloads, and other useful information is available for all developers at http://developer.marklogic.com. For technical questions, we encourage you to ask your question on Stack Overflow.

5.0 Copyright

MarkLogic Server 9.0 and supporting products.
Last updated: November 18, 2019

COPYRIGHT

© 2019 MarkLogic Corporation. All rights reserved.
This technology is protected by U.S. Patent No. 7,127,469B2, U.S. Patent No. 7,171,404B2, U.S. Patent No. 7,756,858 B2, and U.S. Patent No 7,962,474 B2, US 8,892,599, and US 8,935,267.

The MarkLogic software is protected by United States and international copyright laws, and incorporates certain third party libraries and components which are subject to the attributions, terms, conditions and disclaimers set forth below.

For all copyright notices, including third-party copyright notices, see the Combined Product Notices for your version of MarkLogic.