Microsoft

Getting Started with Microsoft R Server on HDInsight (R Server) – Creating a cluster

Contents

Introduction

Introduction

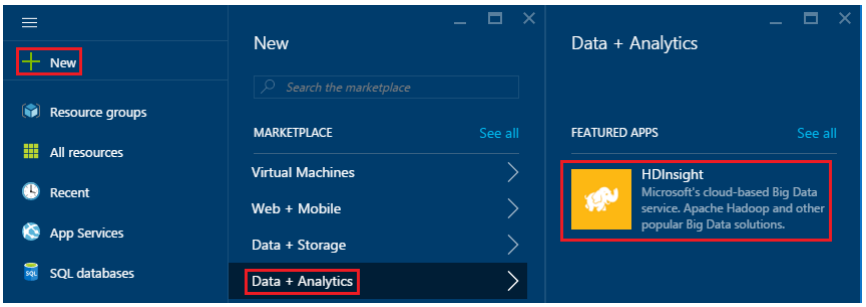
The steps in this document create an R Server on HDInsight using basic configuration information in preparation for the Microsoft R Server lab. For other cluster configuration settings (such as adding additional storage accounts, using an Azure Virtual Network, or creating a metastore for Hive,) see Create Linux-based HDInsight clusters.

The cluster creation steps should take no more than 5 minutes, however it may currently take up to 30 minutes to provision a cluster.

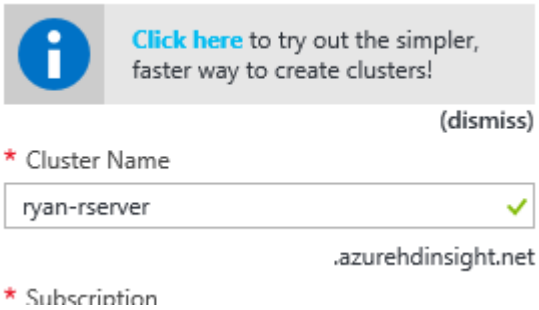
Create a Cluster

Creating a cluster in the Portal

1. Sign in to the **Azure portal**.
2. Select **NEW, Data + Analytics,** and then **HDInsight.**



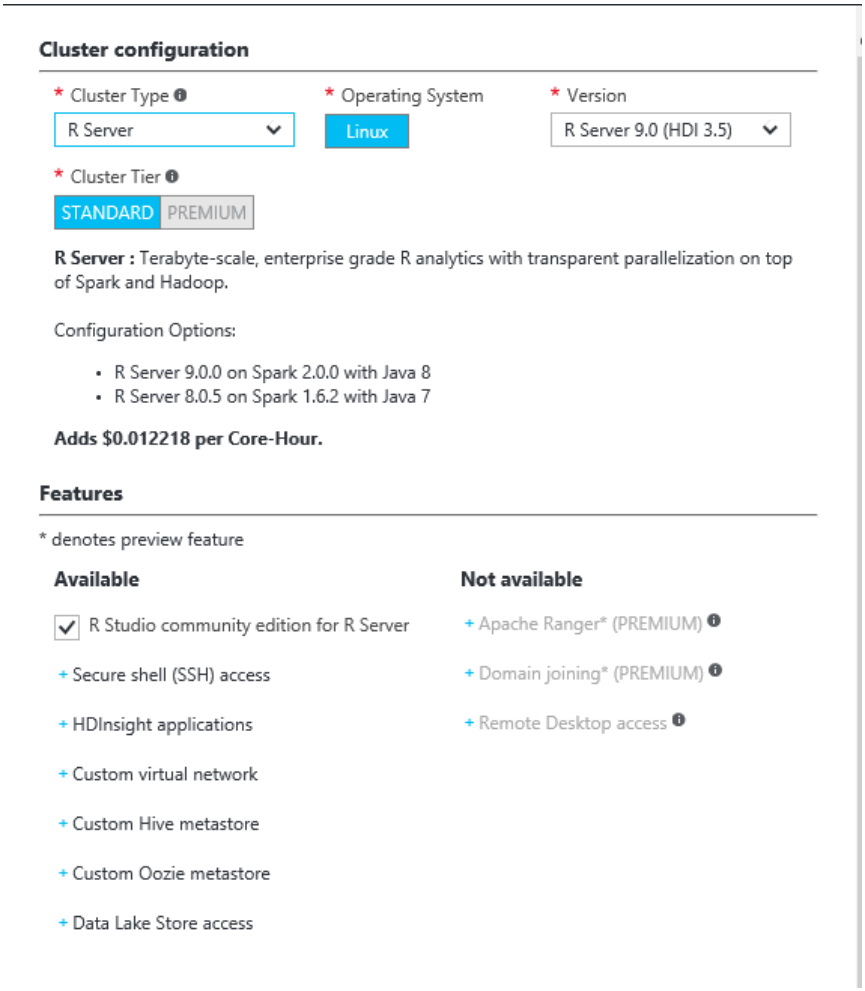
3. Enter a name for the cluster in the **Cluster Name** field. If you have multiple Azure subscriptions, use the **Subscription** entry to select the one you want to use. The name will form part of the public name in DNS <yourcluster>.azurehdinsight.net, so will need to be unique on the internet. For example, for this lab you could use your name. If presented with 'click here to try out …' please ignore it for the moment:



4. Select **Select Cluster Type**. On the **Cluster Type** blade, select the following options:

- **Cluster Type**: R Server. Note that this will still install and configure Spark.
- **Cluster Tier**: Only standard is currently available. Premium is imminent and will enable Ranger, Remote desktop and the ability to join an Active Directory domain.

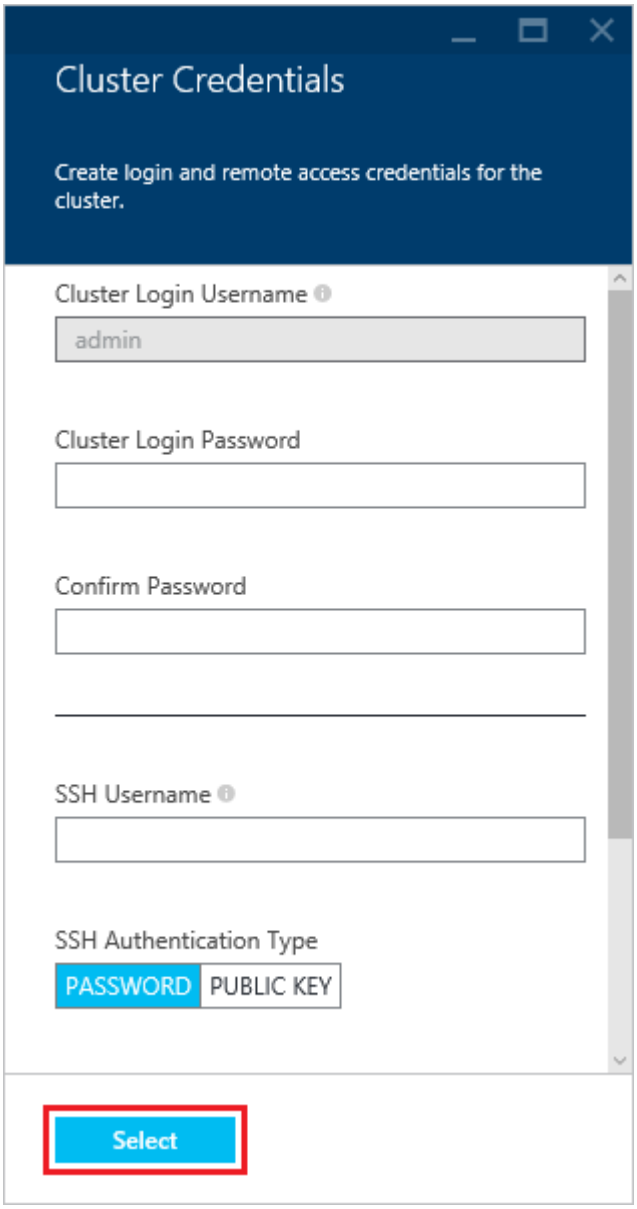
Only Linux is also currently available, and for most scenerios, Linux is the recommended build. Leave the other options at the default values, including 'R Studio Community Edition for R Server' then use the **Select** button to save the cluster type.



5. Select **Credentials**, then enter a Cluster Login Username and Cluster Login Password. This will be used for cluster admin access via the cluster dashboards.

   Enter an SSH Username and select Password, then enter the SSH Password to configure the SSH account. SSH is used to remotely connect to the cluster using a Secure Shell (SSH) client. SSH credentials will be used for R Studio Access. Note the SSH username must be different to the Cluster login, and can also be used for Secure shell.

   Make a note of both sets of credentials then use the Select button to save the credentials.



6. Select **Data Source** to select ‘Azure Storage’ for a data source for the cluster. Either select an existing storage account by selecting Select storage account and then selecting the account, or create a new account using the New link in the Select storage account section. Note that it is possible for Data Lake (our WebHDFS compliant store) to be used, as most of the labs focus on transferring data via blob store, we’ll use ‘Azure Storage’ as the primary storage method.

   If you select New, you must enter a name for the new storage account. A green check will appear if the name is accepted.

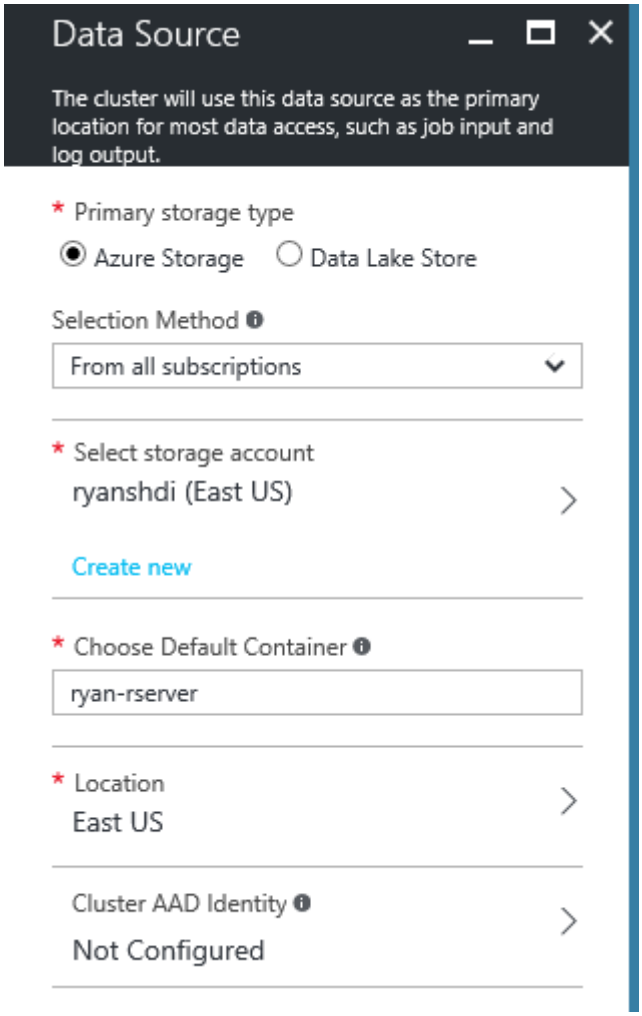
   The Default Container will default to the name of the cluster. Leave this as the value.

   Select Location to select the region to create the storage account in.

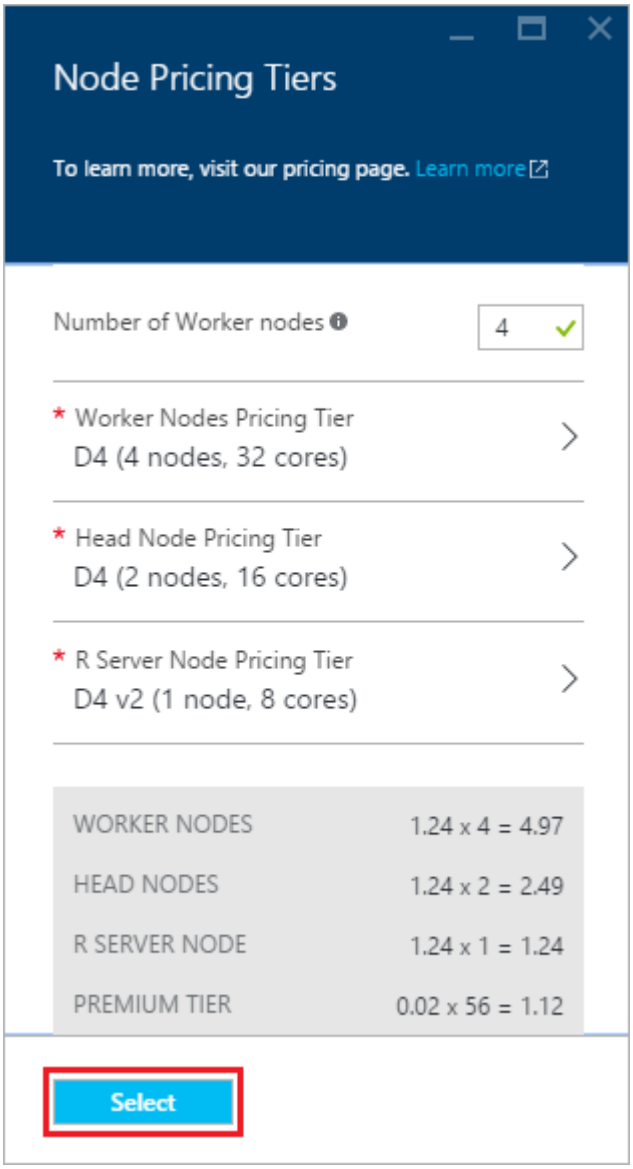
Important: Selecting the location for the default data source will also set the location of the HDInsight cluster. The cluster and default data source must be located in the same region.

A Cluster AAD identity is used to set POSIX permissions for the Cluster access to Azure Data Lake, and is not required for this lab.

Use the **Select** button to save the data source configuration.



7. Select **Node Pricing Tiers** to display information about the nodes that will be created for this cluster. Unless you know that you'll need a larger cluster, **leave the number of worker nodes at the default of 4** The estimated cost of the cluster will be shown within the blade.

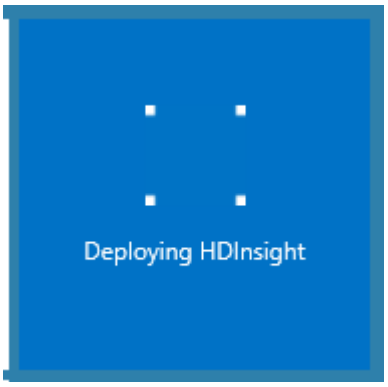


8. Resource Groups allow you to logically group separate resources. Anything within the resource group is dropped when the resource group is dropped. On the Resource Group blade, if you've created a resource group for the lab, 'select existing' alternatively create a new resource group by giving it a name – this is not public so can be any meaningful name, without a space.

9. On the **New HDInsight Cluster** blade, make sure that **Pin to Startboard** is selected, and then select **Create**. This will create the cluster and add a tile for it to the Startboard of your Azure Portal. The icon will indicate that the cluster is creating, and will change to display the HDInsight icon once creation has completed. Note, that because HDInsight clusters cannot be paused, only deleted (the data remains), you may wish to select 'Automation Options' to save, and then deploy the Azure Resource Manager

template which will save you from manually going through this wizard for the next deployment.

It will take some time for the cluster to be created, usually around 15 minutes, however allow up to 30. Use the tile on the Startboard, or the Notifications entry on the left of the page to check on the creation process. From the Dashboard, if you select ‘pin to Dashboard’ you should see the following during the deployment:



Conclusion

Completion

Once these steps are completed, you will have a HDInsight cluster and an edge Server with Microsoft R Server and R Studio Community Edition in preparation for the Microsoft R Server Lab.

Terms of Use

By using this Hands-on Lab, you agree to the following terms:

The technology/functionality described in this Hands-on Lab is provided by Microsoft Corporation in a "sandbox" testing environment for purposes of obtaining your feedback and to provide you with a learning experience. You may only use the Hands-on Lab to evaluate such technology features and functionality and provide feedback to Microsoft. You may not use it for any other purpose. You may not modify copy, distribute, transmit, display, perform, reproduce, publish, license, create derivative works from, transfer, or sell this Hands-on Lab or any portion thereof.

COPYING OR REPRODUCTION OF THE HANDS-ON LAB (OR ANY PORTION OF IT) TO ANY OTHER SERVER OR LOCATION FOR FURTHER REPRODUCTION OR REDISTRIBUTION IS EXPRESSLY PROHIBITED.

THIS HANDS-ON LAB PROVIDES CERTAIN SOFTWARE TECHNOLOGY/PRODUCT FEATURES AND FUNCTIONALITY, INCLUDING POTENTIAL NEW FEATURES AND CONCEPTS, IN A SIMULATED ENVIRONMENT WITHOUT COMPLEX SET-UP OR INSTALLATION FOR THE PURPOSE DESCRIBED ABOVE. THE TECHNOLOGY/CONCEPTS REPRESENTED IN THIS HANDS-ON LAB MAY NOT REPRESENT FULL FEATURE FUNCTIONALITY AND MAY NOT WORK THE WAY A FINAL VERSION MAY WORK. WE ALSO MAY NOT RELEASE A FINAL VERSION OF SUCH FEATURES OR CONCEPTS. YOUR EXPERIENCE WITH USING SUCH FEATURES AND FUNCITONALITY IN A PHYSICAL ENVIRONMENT MAY ALSO BE DIFFERENT.

FEEDBACK. If you give feedback about the technology features, functionality and/or concepts described in this Hands-on Lab to Microsoft, you give to Microsoft, without charge, the right to use, share and commercialize your feedback in any way and for any purpose. You also give to third parties, without charge, any patent rights needed for their products, technologies and services to use or interface with any specific parts of a Microsoft software or service that includes the feedback. You will not give feedback that is subject to a license that requires Microsoft to license its software or documentation to third parties because we include your feedback in them. These rights survive this agreement.

MICROSOFT CORPORATION HEREBY DISCLAIMS ALL WARRANTIES AND CONDITIONS WITH REGARD TO THE HANDS-ON LAB, INCLUDING ALL WARRANTIES AND CONDITIONS OF MERCHANTABILITY, WHETHER EXPRESS, IMPLIED OR STATUTORY, FITNESS FOR A PARTICULAR PURPOSE, TITLE AND NON-INFRINGEMENT. MICROSOFT DOES NOT MAKE ANY ASSURANCES OR REPRESENTATIONS WITH REGARD TO THE ACCURACY OF THE RESULTS, OUTPUT THAT DERIVES FROM USE OF THE VIRTUAL LAB, OR SUITABILITY OF THE INFORMATION CONTAINED IN THE VIRTUAL LAB FOR ANY PURPOSE.

DISCLAIMER

This lab contains only a portion of the features and enhancements in Microsoft Azure. Some of the features might change in future releases of the product.