[](https://github.com/Microsoft/MCW-Template-Cloud-Workshop/raw/master/Media/ms-cloud-workshop.png)

Big data and visualization

Before the hands-on lab setup guide

November 2018

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links may be provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

© 2018 Microsoft Corporation. All rights reserved.

Microsoft and the trademarks listed at <https://www.microsoft.com/en-us/legal/intellectualproperty/Trademarks/Usage/General.aspx> are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

**Contents**

* [Big data and visualization before the hands-on lab setup guide](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#big-data-and-visualization-before-the-hands-on-lab-setup-guide)
  + [Before the hands-on lab](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#before-the-hands-on-lab)
    - [Task 1: Provision Azure Databricks](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-1-provision-azure-databricks)
    - [Task 2: Create Azure Storage account](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-2-create-azure-storage-account)
    - [Task 3: Retrieve Azure Storage account information and create container](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-3-retrieve-azure-storage-account-information-and-create-container)
    - [Task 4: Provision Azure Data Factory](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-4-provision-azure-data-factory)
    - [Task 5: Create an Azure Databricks cluster](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-5-create-an-azure-databricks-cluster)

**Big data and visualization before the hands-on lab setup guide**

**Before the hands-on lab**

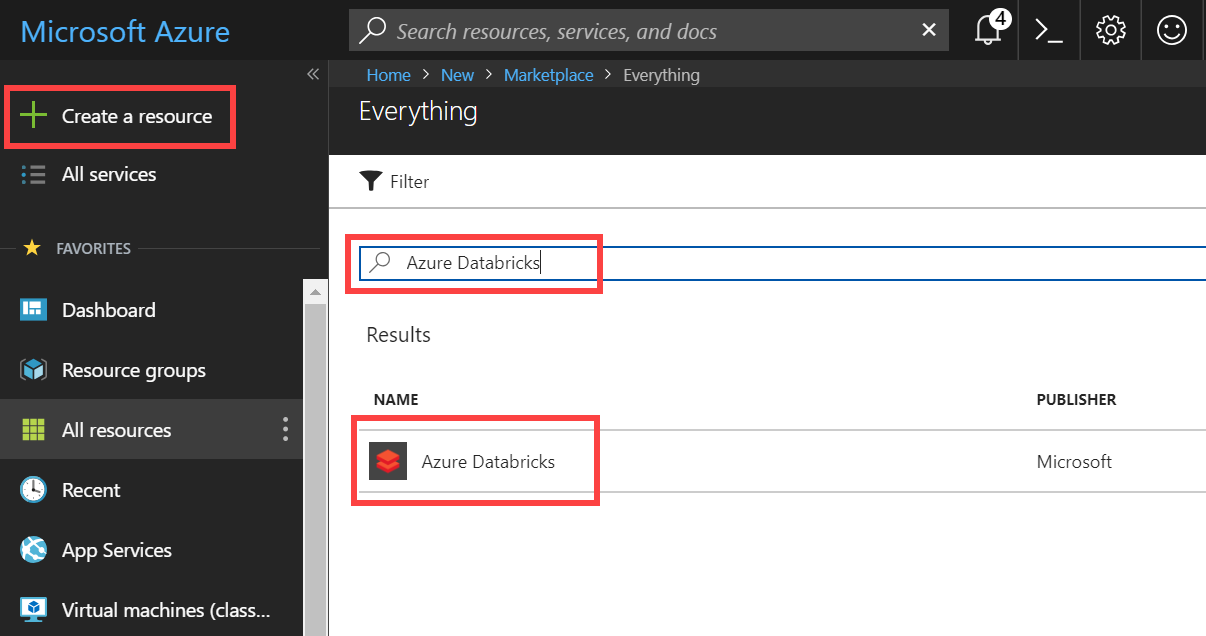
Duration: 30 minutes

In this exercise, you will set up your environment for use in the rest of the hands-on lab. You should follow all the steps provided in the Before the Hands-on Lab section to prepare your environment *before* attending the hands-on lab.

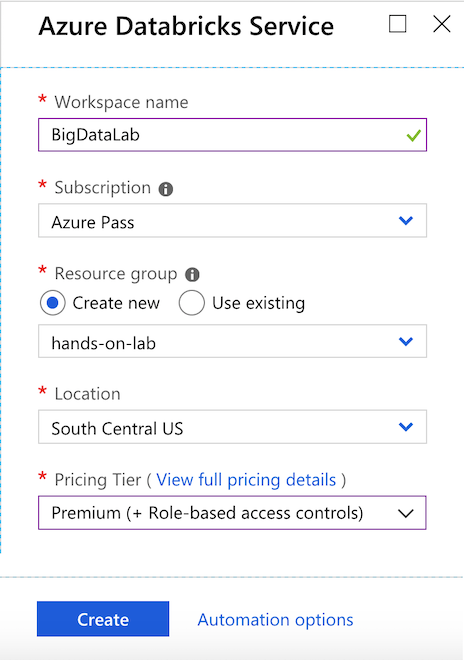
**Task 1: Provision Azure Databricks**

Azure Databricks is an Apache Spark-based analytics platform optimized for Azure. It will be used in this lab to build and train a machine learning model used to predict flight delays.

1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "Azure Databricks" into the search bar. Select Azure Databricks from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-databricks-resource.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Azure Databricks Service creation form:
   * **Name**: Enter a unique name as indicated by a green checkmark.
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource Group**: Select the same resource group you created at the beginning of this lab.
   * **Location**: Select a region close to you. ***(If you are using an Azure Pass, select South Central US.)***
   * **Pricing**: Select Premium.

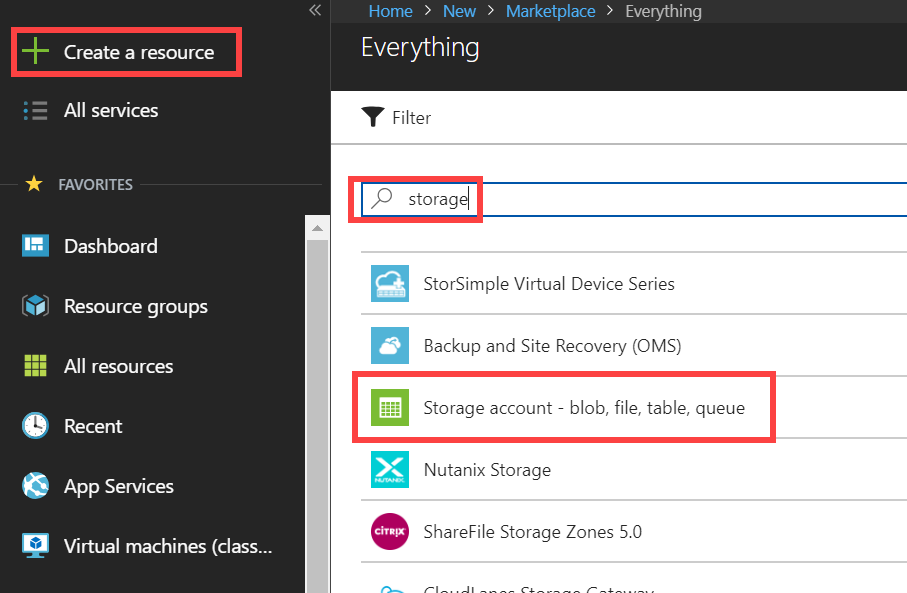
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-blade.png)

1. Select **Create** to finish and submit.

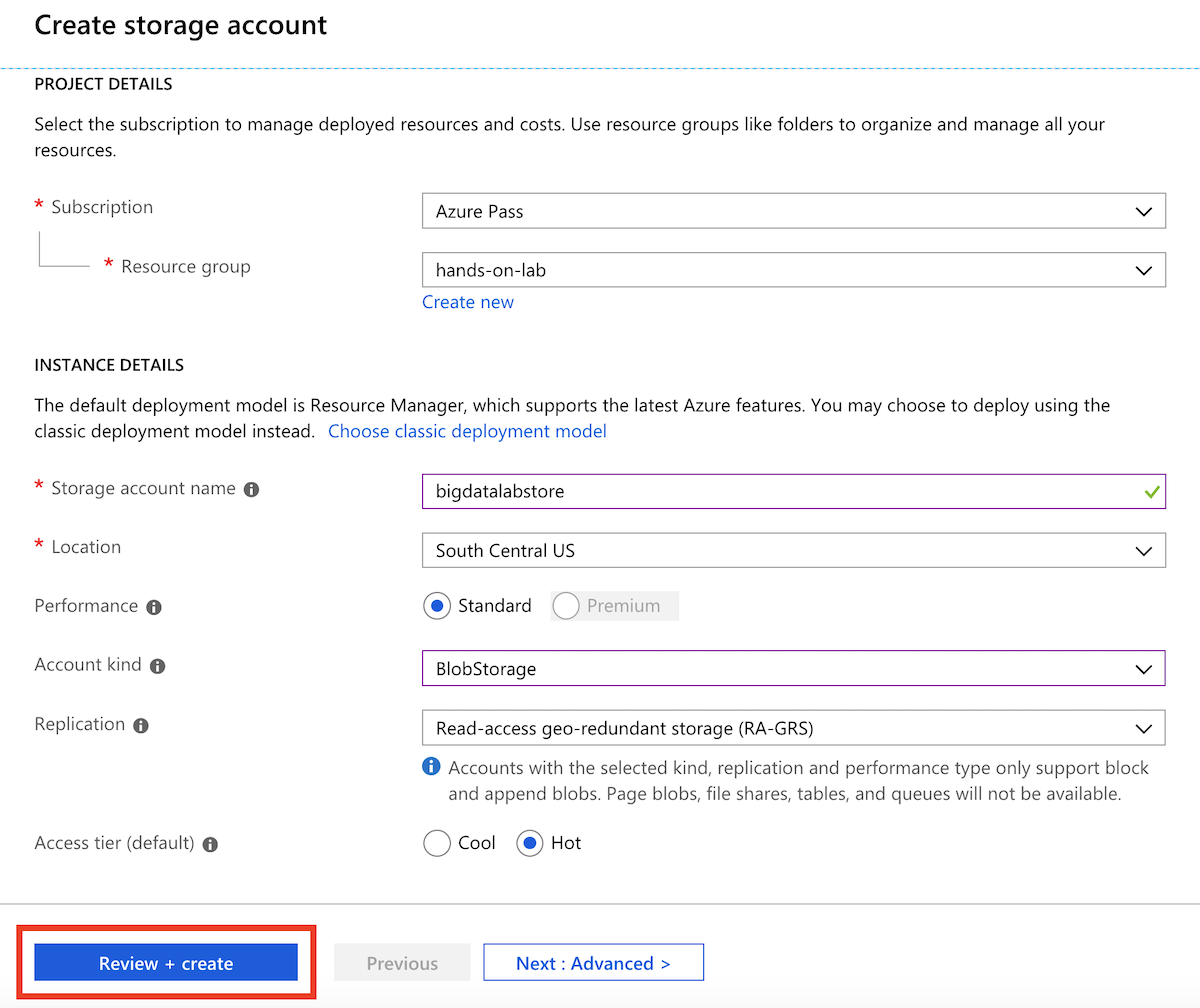
**Task 2: Create Azure Storage account**

Create a new Azure Storage account that will be used to store historic and scored flight and weather data sets for the lab.

1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "storage" into the search bar. Select **Storage account - blob, file, table, queue** from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-storage-resource.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Azure Storage account creation form:
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource group**: Select the same resource group you created at the beginning of this lab.
   * **Storage account name**: Enter a unique name as indicated by a green checkmark.
   * **Location**: Select the same region you used for Azure Databricks.
   * **Performance**: Standard
   * **Account kind**: BlobStorage
   * **Replication**: Read-access geo-redundant storage (RA-GRS)
   * **Access tier**: Hot

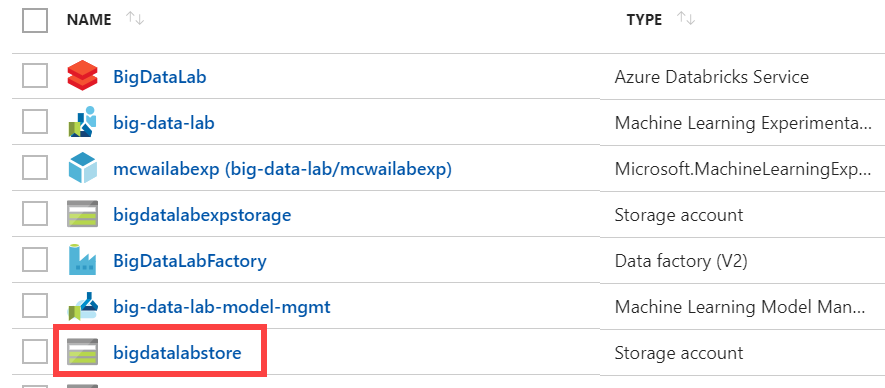
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-create-blade.png)

1. Select **Create** to finish and submit.

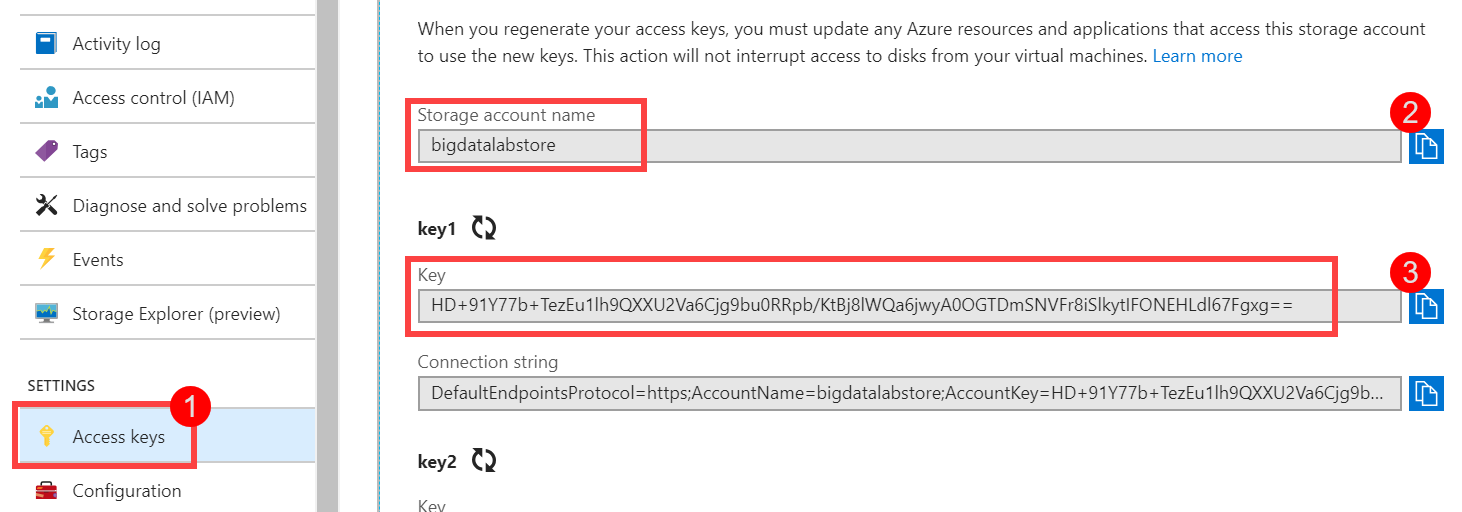
**Task 3: Retrieve Azure Storage account information and create container**

You will need to have the Azure Storage account name and access key when you create your Azure Databricks cluster during the lab. You will also need to create storage containers in which you will store your flight and weather data files.

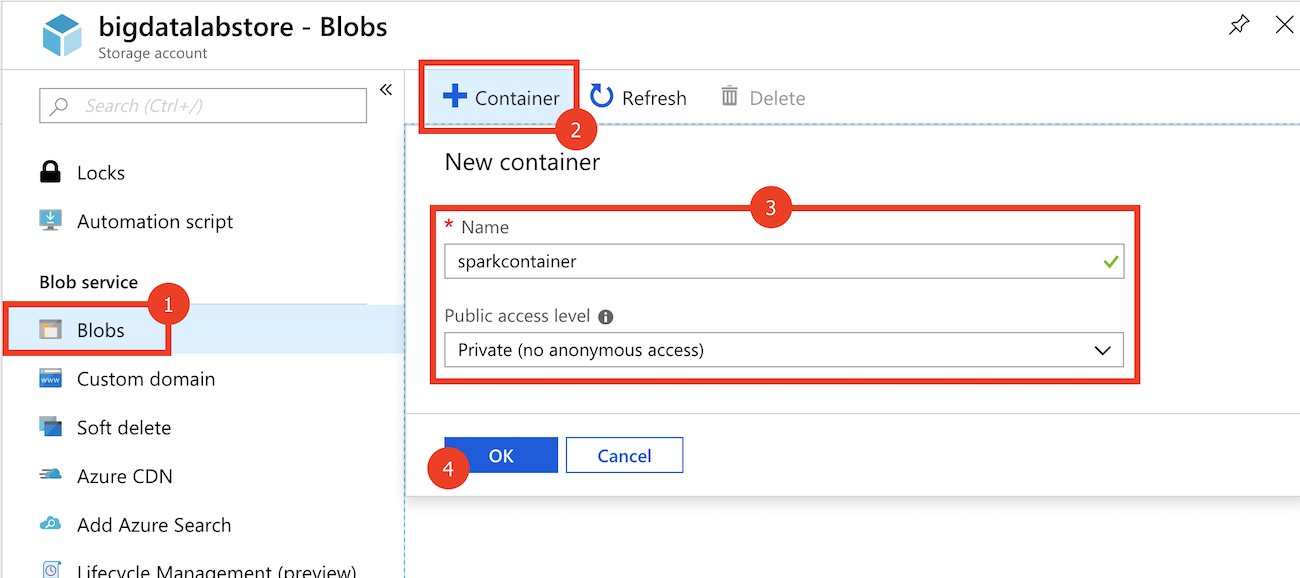
1. From the side menu in the Azure portal, choose **Resource groups**, then enter your resource group name into the filter box, and select it from the list.
2. Next, select your lab Azure Storage account from the list.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/select-azure-storage-account.png)

1. Select **Access keys** (1) from the menu. Copy the **storage account name** (2) and the **key1** key (3) and copy the values to a text editor such as Notepad for later.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-access-keys.png)

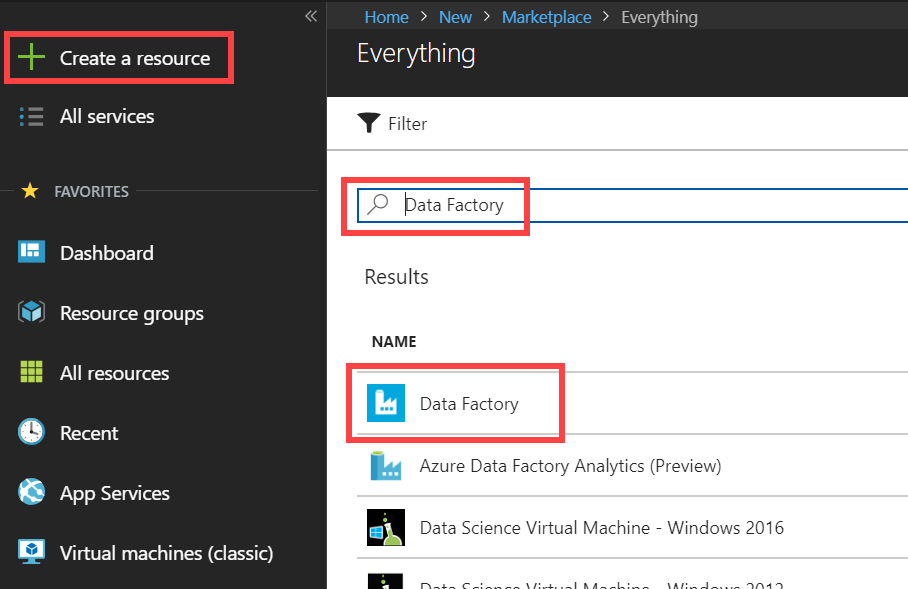
1. Select **Blobs** (1) from the menu. Select **+ Container** (2) on the Blobs blade, enter **sparkcontainer** for the name (3), leaving the public access level set to Private. Select **OK** (4) to create the container.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-create-container.png)

**Task 4: Provision Azure Data Factory**

Create a new Azure Data Factory instance that will be used to orchestrate data transfers for analysis.

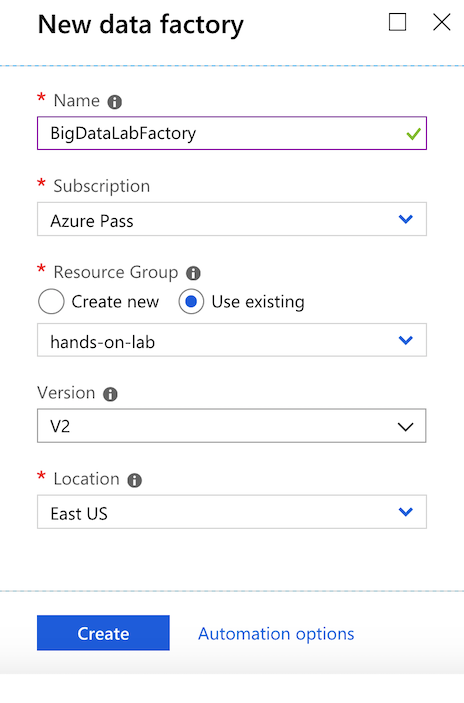
1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "Data Factory" into the search bar. Select **Data Factory** from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-data-factory.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Data Factory creation form:
   * **Name**: Enter a unique name as indicated by a green checkmark.
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource Group**: Select the same resource group you created at the beginning of this lab.
   * **Version**: V2
   * **Location**: Select any region close to you.

***Understanding Data Factory Location:*** The Data Factory location is where the metadata of the data factory is stored and where the triggering of the pipeline is initiated from. Meanwhile, a data factory can access data stores and compute services in other Azure regions to move data between data stores or process data using compute services. This behavior is realized through the [globally available IR](https://azure.microsoft.com/en-us/global-infrastructure/services/?products=data-factory) to ensure data compliance, efficiency, and reduced network egress costs.

The IR Location defines the location of its back-end compute, and essentially the location where the data movement, activity dispatching, and SSIS package execution are performed. The IR location can be different from the location of the data factory it belongs to.

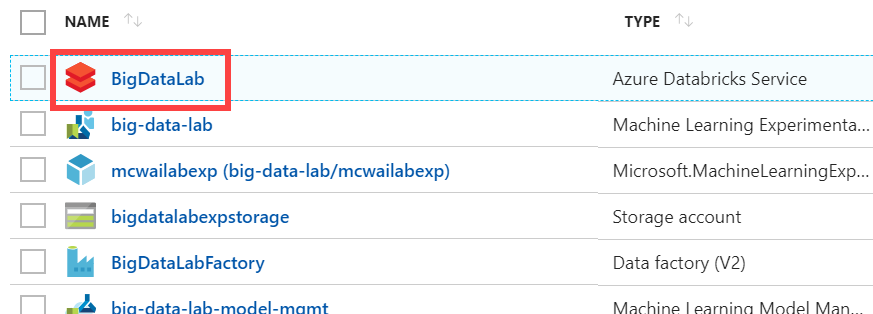
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-data-factory-create-blade.png)

1. Select **Create** to finish and submit.

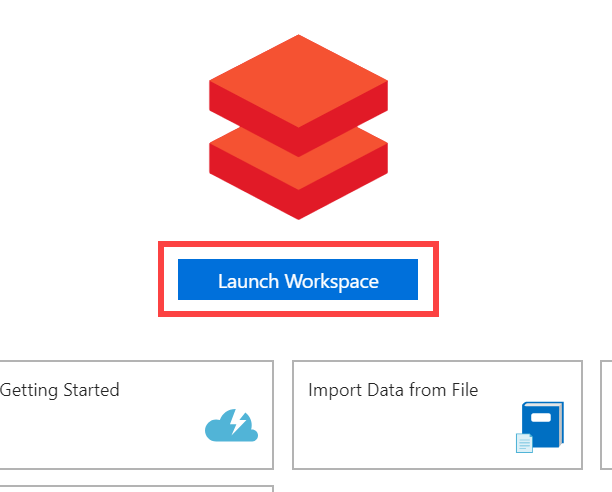
**Task 5: Create an Azure Databricks cluster**

You have provisioned an Azure Databricks workspace, and now you need to create a new cluster within the workspace. Part of the cluster configuration includes setting up an account access key to your Azure Storage account, using the Spark Config within the new cluster form. This will allow your cluster to access the lab files.

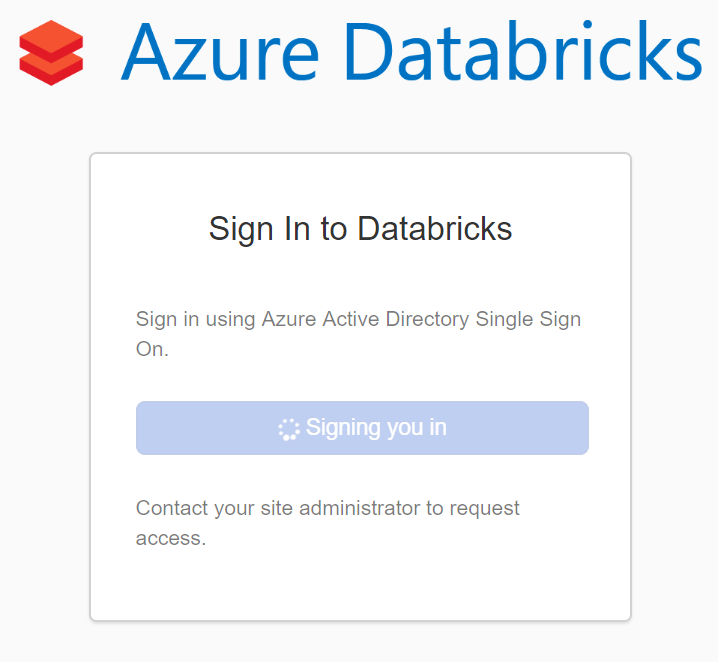
1. From the side menu in the Azure portal, select **Resource groups**, then enter your resource group name into the filter box, and select it from the list.
2. Next, select your Azure Databricks service from the list.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/select-azure-databricks-service.png)

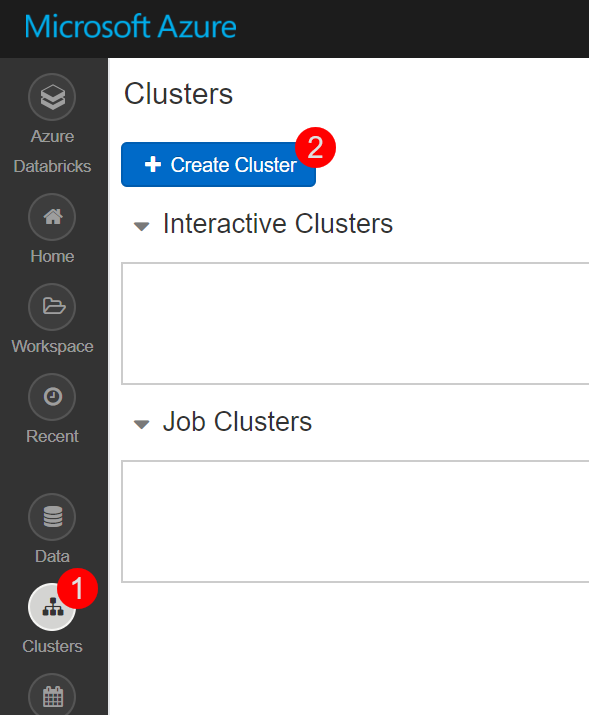
1. In the Overview pane of the Azure Databricks service, select **Launch Workspace**.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-launch-workspace.png)

Azure Databricks will automatically log you in using Azure Active Directory Single Sign On.

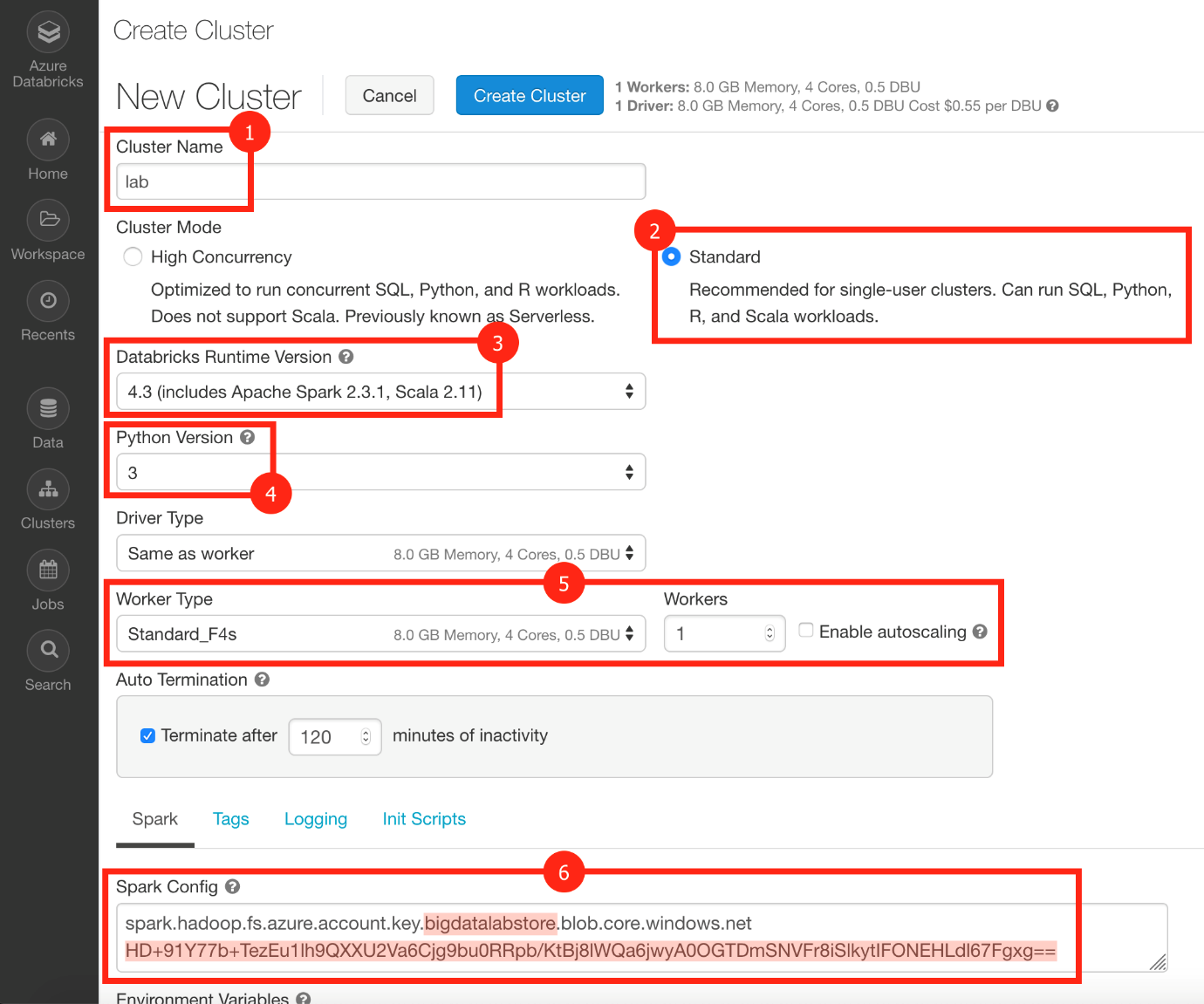
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-aad.png)

1. Select **Clusters** (1) from the menu, then select **Create Cluster** (2).

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-cluster-button.png)

1. On the Create New Cluster form, provide the following:
   * **Cluster Name**: lab
   * **Cluster Type**: Standard
   * **Databricks Runtime Version**: 4.3 (includes Apache Spark 2.3.1, Scala 2.11)
   * **Python Version**: 3
   * **Driver Type**: Same as worker
   * **Worker Type**: Standard\_F4s
   * **Workers**: 1
   * **Enable Autoscaling**: Uncheck this option.
   * **Auto Termination**: Check the box and enter 120.
   * **Spark Config**: Edit the Spark Config by entering the connection information for your Azure Storage account that you copied earlier in Task 5. This will allow your cluster to access the lab files. Enter the following:

spark.hadoop.fs.azure.account.key.<STORAGE\_ACCOUNT\_NAME>.blob.core.windows.net <ACCESS\_KEY>, where <STORAGE\_ACCOUNT\_NAME> is your Azure Storage account name, and <ACCESS\_KEY> is your storage access key.

1. **Example:** spark.hadoop.fs.azure.account.key.bigdatalabstore.blob.core.windows.net HD+91Y77b+TezEu1lh9QXXU2Va6Cjg9bu0RRpb/KtBj8lWQa6jwyA0OGTDmSNVFr8iSlkytIFONEHLdl67Fgxg==
2. [](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-cluster-form.png)
3. Select **Create Cluster**.

You should follow all these steps provided *before* attending the Hands-on lab.

* © 2019 GitHub, Inc.
* [Terms](https://github.com/site/terms)
* [Privacy](https://github.com/site/privacy)
* [Security](https://github.com/security)
* [Status](https://githubstatus.com/)
* [Help](https://help.github.com/)
* [Contact GitHub](https://github.com/contact)
* [Pricing](https://github.com/pricing)
* [API](https://developer.github.com/)
* [Training](https://training.github.com/)
* [Blog](https://github.blog/)
* [About](https://github.com/about)

[Skip to content](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#start-of-content)

* Why GitHub?
* [Enterprise](https://github.com/enterprise)
* Explore
* [Marketplace](https://github.com/marketplace)
* Pricing

Top of Form



Bottom of Form

[Sign in](https://github.com/login?return_to=%2Fmicrosoft%2FMCW-Big-data-and-visualization)[Sign up](https://github.com/join)

* [**Watch**](https://github.com/login?return_to=%2Fmicrosoft%2FMCW-Big-data-and-visualization)[**67**](https://github.com/microsoft/MCW-Big-data-and-visualization/watchers)
* [**Star**](https://github.com/login?return_to=%2Fmicrosoft%2FMCW-Big-data-and-visualization)[**33**](https://github.com/microsoft/MCW-Big-data-and-visualization/stargazers)
* [**Fork**](https://github.com/login?return_to=%2Fmicrosoft%2FMCW-Big-data-and-visualization)[**61**](https://github.com/microsoft/MCW-Big-data-and-visualization/network/members)

[microsoft](https://github.com/microsoft)/[**MCW-Big-data-and-visualization**](https://github.com/microsoft/MCW-Big-data-and-visualization)

[Code](https://github.com/microsoft/MCW-Big-data-and-visualization)[Issues **5**](https://github.com/microsoft/MCW-Big-data-and-visualization/issues)[Pull requests **0**](https://github.com/microsoft/MCW-Big-data-and-visualization/pulls)[Projects **0**](https://github.com/microsoft/MCW-Big-data-and-visualization/projects)[Insights](https://github.com/microsoft/MCW-Big-data-and-visualization/pulse)

Top of Form

Dismiss

Bottom of Form

**Join GitHub today**

GitHub is home to over 36 million developers working together to host and review code, manage projects, and build software together.

[**Sign up**](https://github.com/join?source=prompt-blob-show)

Branch: master

[**MCW-Big-data-and-visualization**](https://github.com/microsoft/MCW-Big-data-and-visualization)**/**[**Hands-on lab**](https://github.com/microsoft/MCW-Big-data-and-visualization/tree/master/Hands-on%20lab)**/Before the HOL - Big data and visualization.md**

[**Find file**](https://github.com/microsoft/MCW-Big-data-and-visualization/find/master)Copy path

[@kylebunting](https://github.com/kylebunting) [kylebunting](https://github.com/kylebunting) [Fix for TOC links in before the HOL MD file.578e9be](https://github.com/microsoft/MCW-Big-data-and-visualization/commit/578e9be5b41f49162511eb918b2f2b355545dfe3) on Mar 13

**6** contributors[@DawnmarieDesJardins](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=DawnmarieDesJardins)[@nansravn](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=nansravn)[@hopero929](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=hopero929)[@kylebunting](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=kylebunting)[@Abhishekpathania01](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=Abhishekpathania01)[@joelhulen](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md?author=joelhulen)

206 lines (110 sloc)  12.1 KB

[**Raw**](https://github.com/microsoft/MCW-Big-data-and-visualization/raw/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md)[**Blame**](https://github.com/microsoft/MCW-Big-data-and-visualization/blame/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md)[**History**](https://github.com/microsoft/MCW-Big-data-and-visualization/commits/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md)

[](https://github.com/Microsoft/MCW-Template-Cloud-Workshop/raw/master/Media/ms-cloud-workshop.png)

Big data and visualization

Before the hands-on lab setup guide

November 2018

Information in this document, including URL and other Internet Web site references, is subject to change without notice. Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places, and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, e-mail address, logo, person, place or event is intended or should be inferred. Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this document may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this document. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this document does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

The names of manufacturers, products, or URLs are provided for informational purposes only and Microsoft makes no representations and warranties, either expressed, implied, or statutory, regarding these manufacturers or the use of the products with any Microsoft technologies. The inclusion of a manufacturer or product does not imply endorsement of Microsoft of the manufacturer or product. Links may be provided to third party sites. Such sites are not under the control of Microsoft and Microsoft is not responsible for the contents of any linked site or any link contained in a linked site, or any changes or updates to such sites. Microsoft is not responsible for webcasting or any other form of transmission received from any linked site. Microsoft is providing these links to you only as a convenience, and the inclusion of any link does not imply endorsement of Microsoft of the site or the products contained therein.

© 2018 Microsoft Corporation. All rights reserved.

Microsoft and the trademarks listed at <https://www.microsoft.com/en-us/legal/intellectualproperty/Trademarks/Usage/General.aspx> are trademarks of the Microsoft group of companies. All other trademarks are property of their respective owners.

**Contents**

* [Big data and visualization before the hands-on lab setup guide](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#big-data-and-visualization-before-the-hands-on-lab-setup-guide)
  + [Before the hands-on lab](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#before-the-hands-on-lab)
    - [Task 1: Provision Azure Databricks](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-1-provision-azure-databricks)
    - [Task 2: Create Azure Storage account](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-2-create-azure-storage-account)
    - [Task 3: Retrieve Azure Storage account information and create container](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-3-retrieve-azure-storage-account-information-and-create-container)
    - [Task 4: Provision Azure Data Factory](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-4-provision-azure-data-factory)
    - [Task 5: Create an Azure Databricks cluster](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/Before%20the%20HOL%20-%20Big%20data%20and%20visualization.md#task-5-create-an-azure-databricks-cluster)

**Big data and visualization before the hands-on lab setup guide**

**Before the hands-on lab**

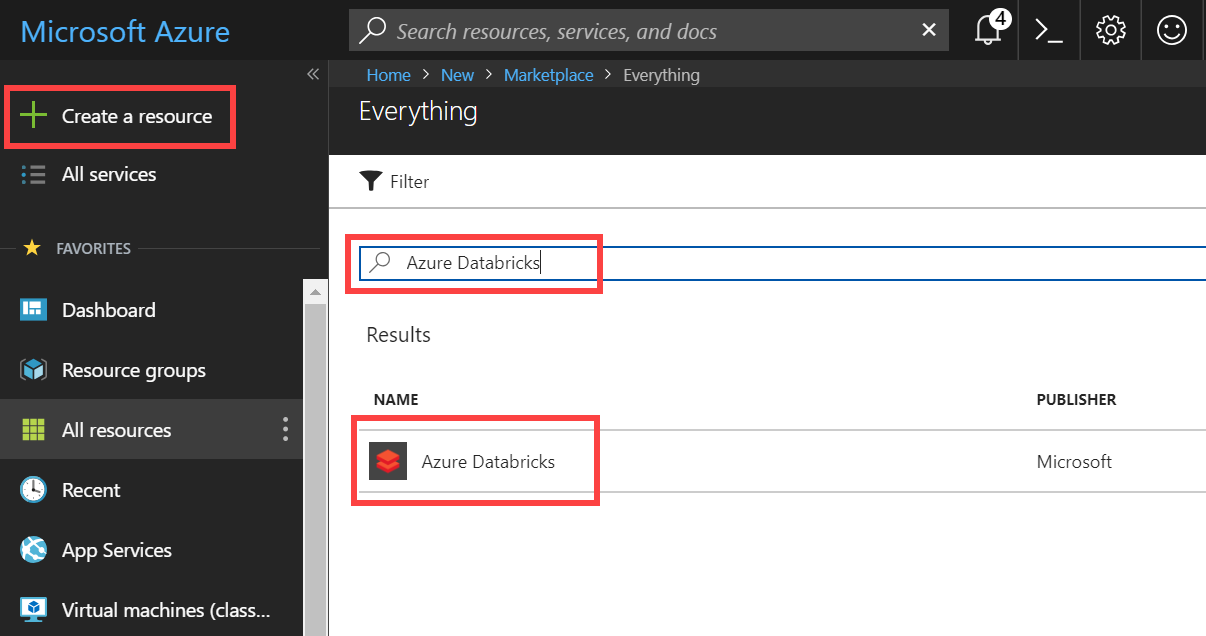
Duration: 30 minutes

In this exercise, you will set up your environment for use in the rest of the hands-on lab. You should follow all the steps provided in the Before the Hands-on Lab section to prepare your environment *before* attending the hands-on lab.

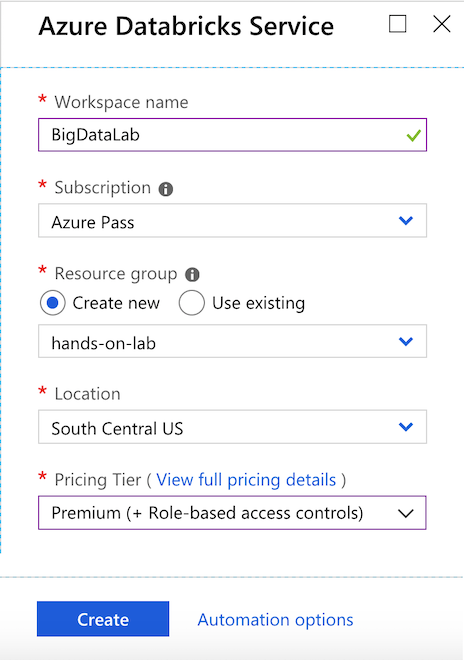
**Task 1: Provision Azure Databricks**

Azure Databricks is an Apache Spark-based analytics platform optimized for Azure. It will be used in this lab to build and train a machine learning model used to predict flight delays.

1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "Azure Databricks" into the search bar. Select Azure Databricks from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-databricks-resource.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Azure Databricks Service creation form:
   * **Name**: Enter a unique name as indicated by a green checkmark.
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource Group**: Select the same resource group you created at the beginning of this lab.
   * **Location**: Select a region close to you. ***(If you are using an Azure Pass, select South Central US.)***
   * **Pricing**: Select Premium.

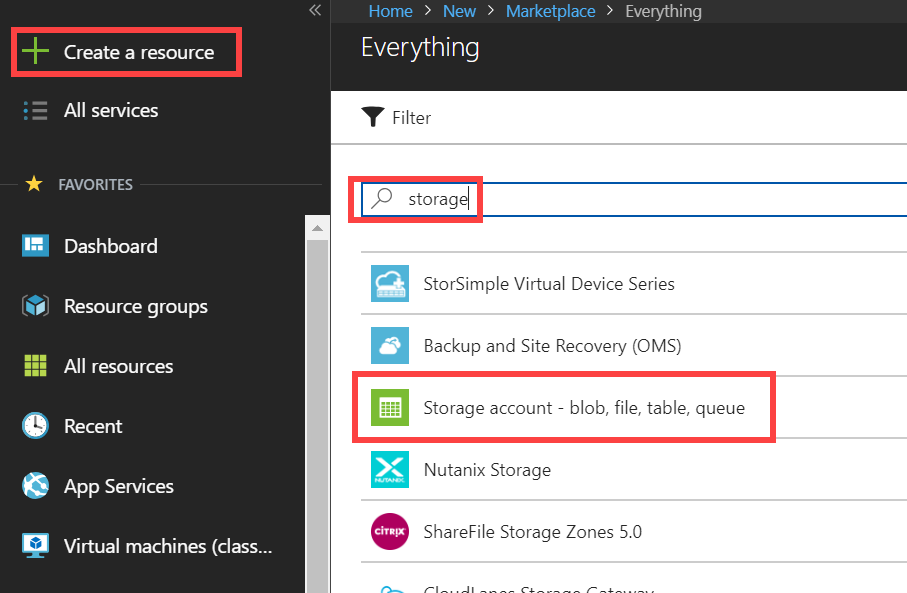
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-blade.png)

1. Select **Create** to finish and submit.

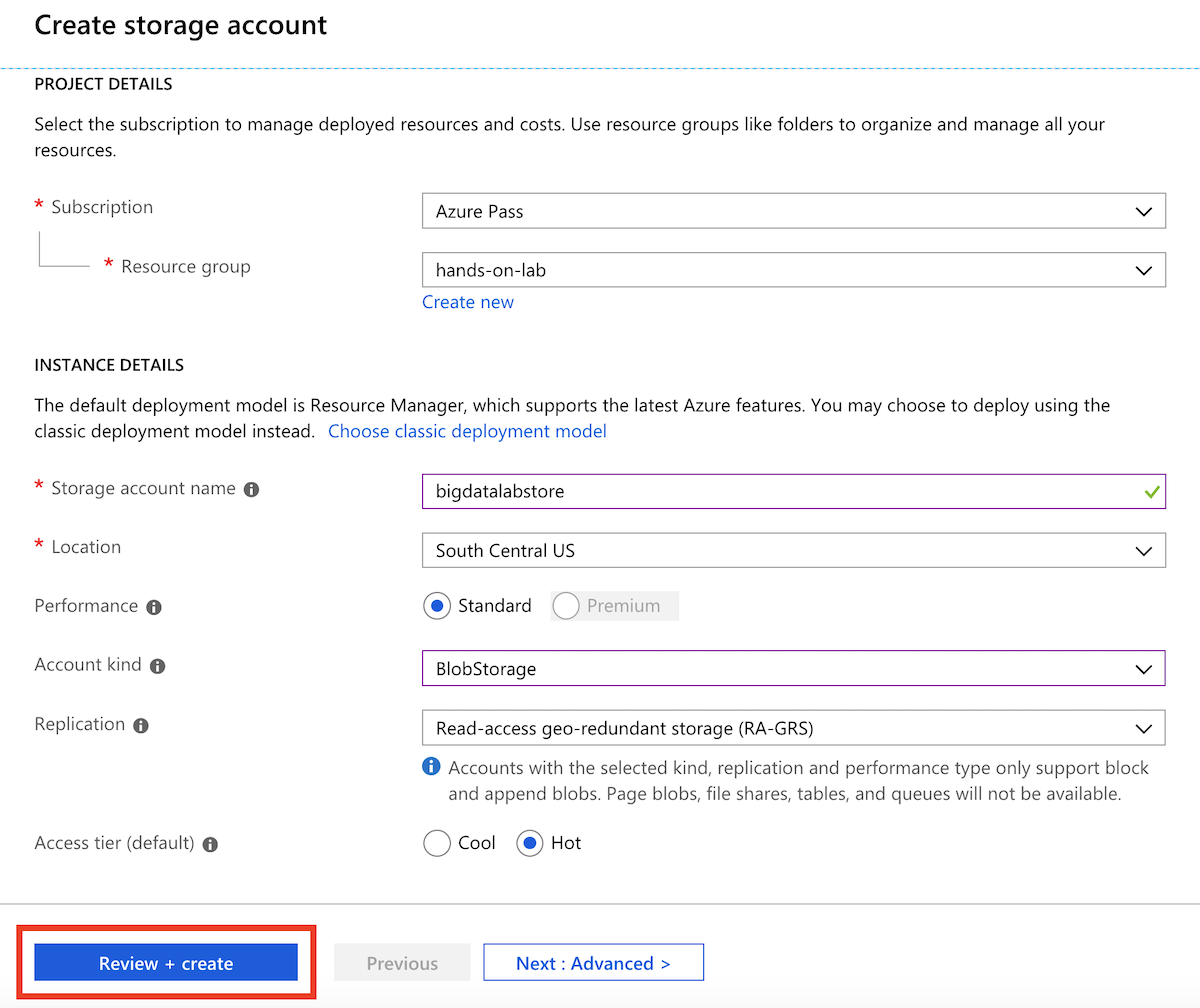
**Task 2: Create Azure Storage account**

Create a new Azure Storage account that will be used to store historic and scored flight and weather data sets for the lab.

1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "storage" into the search bar. Select **Storage account - blob, file, table, queue** from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-storage-resource.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Azure Storage account creation form:
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource group**: Select the same resource group you created at the beginning of this lab.
   * **Storage account name**: Enter a unique name as indicated by a green checkmark.
   * **Location**: Select the same region you used for Azure Databricks.
   * **Performance**: Standard
   * **Account kind**: BlobStorage
   * **Replication**: Read-access geo-redundant storage (RA-GRS)
   * **Access tier**: Hot

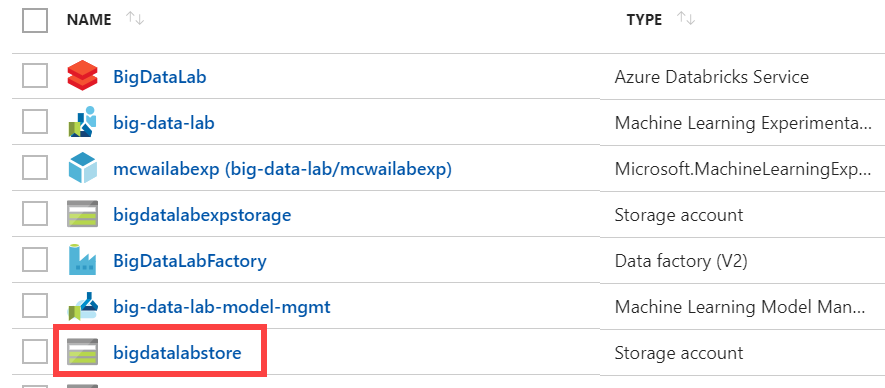
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-create-blade.png)

1. Select **Create** to finish and submit.

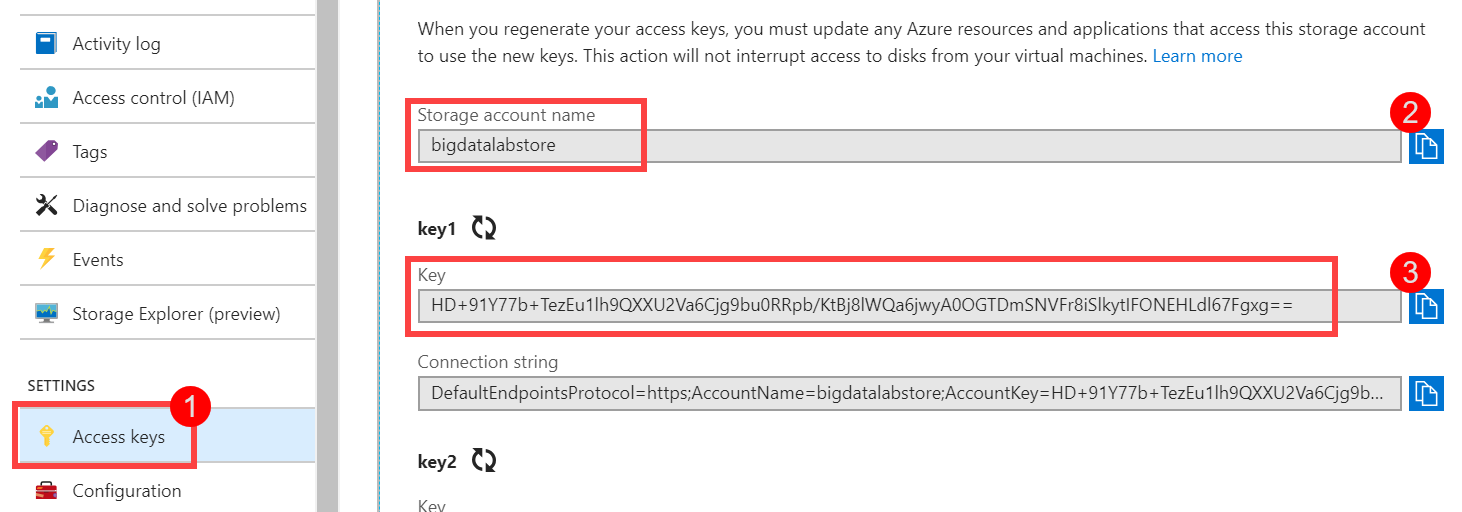
**Task 3: Retrieve Azure Storage account information and create container**

You will need to have the Azure Storage account name and access key when you create your Azure Databricks cluster during the lab. You will also need to create storage containers in which you will store your flight and weather data files.

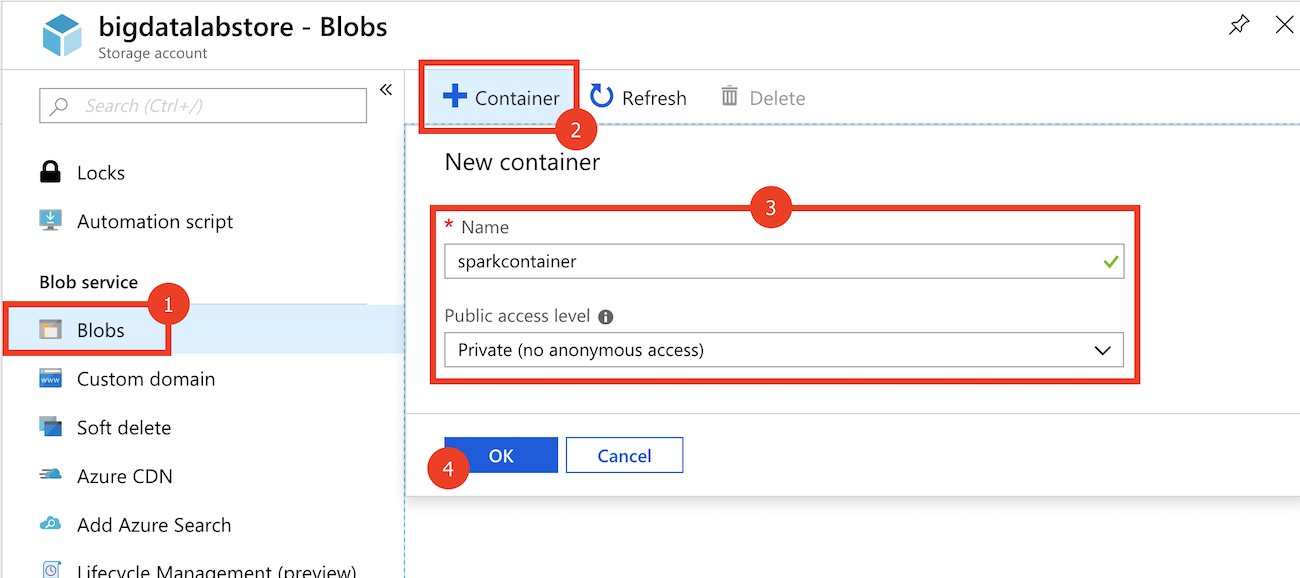
1. From the side menu in the Azure portal, choose **Resource groups**, then enter your resource group name into the filter box, and select it from the list.
2. Next, select your lab Azure Storage account from the list.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/select-azure-storage-account.png)

1. Select **Access keys** (1) from the menu. Copy the **storage account name** (2) and the **key1** key (3) and copy the values to a text editor such as Notepad for later.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-access-keys.png)

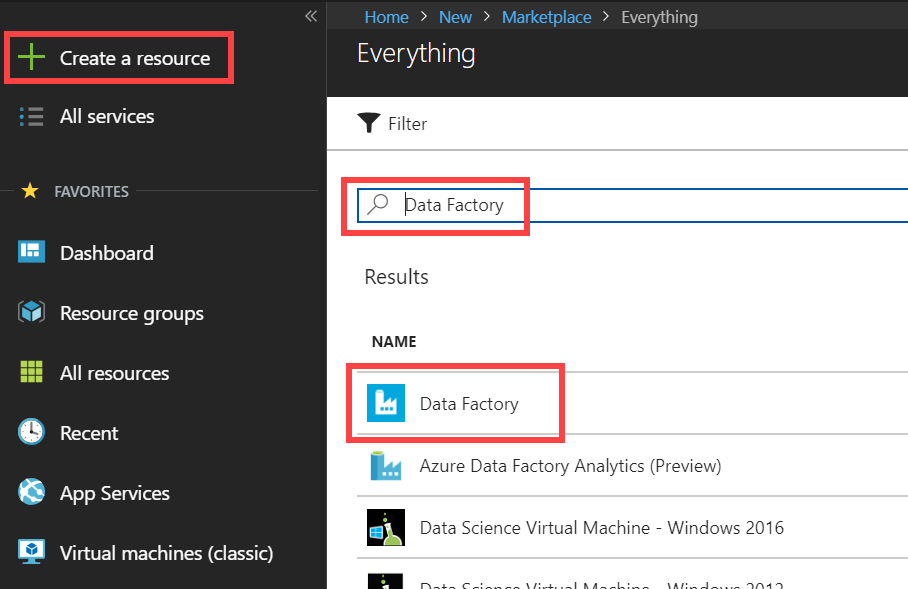
1. Select **Blobs** (1) from the menu. Select **+ Container** (2) on the Blobs blade, enter **sparkcontainer** for the name (3), leaving the public access level set to Private. Select **OK** (4) to create the container.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-storage-create-container.png)

**Task 4: Provision Azure Data Factory**

Create a new Azure Data Factory instance that will be used to orchestrate data transfers for analysis.

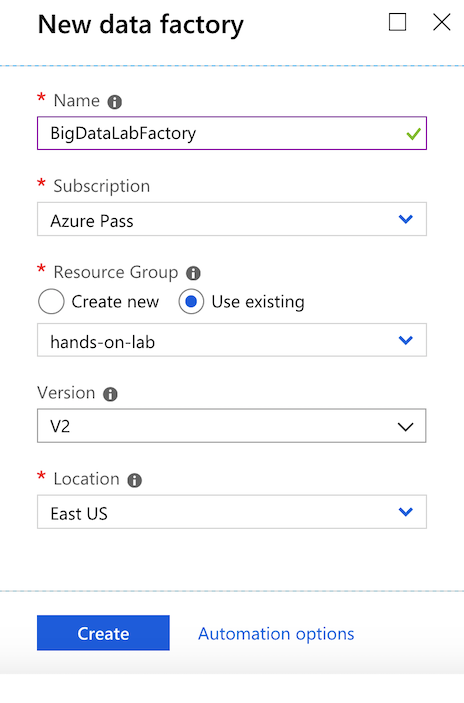
1. In the [Azure Portal](https://portal.azure.com/) ([https://portal.azure.com](https://portal.azure.com/)), select **+ Create a resource**, then type "Data Factory" into the search bar. Select **Data Factory** from the results.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/create-azure-data-factory.png)

1. Select Create on the bottom of the blade that follows.
2. Set the following configuration on the Data Factory creation form:
   * **Name**: Enter a unique name as indicated by a green checkmark.
   * **Subscription**: Select the subscription you are using for this hands-on lab.
   * **Resource Group**: Select the same resource group you created at the beginning of this lab.
   * **Version**: V2
   * **Location**: Select any region close to you.

***Understanding Data Factory Location:*** The Data Factory location is where the metadata of the data factory is stored and where the triggering of the pipeline is initiated from. Meanwhile, a data factory can access data stores and compute services in other Azure regions to move data between data stores or process data using compute services. This behavior is realized through the [globally available IR](https://azure.microsoft.com/en-us/global-infrastructure/services/?products=data-factory) to ensure data compliance, efficiency, and reduced network egress costs.

The IR Location defines the location of its back-end compute, and essentially the location where the data movement, activity dispatching, and SSIS package execution are performed. The IR location can be different from the location of the data factory it belongs to.

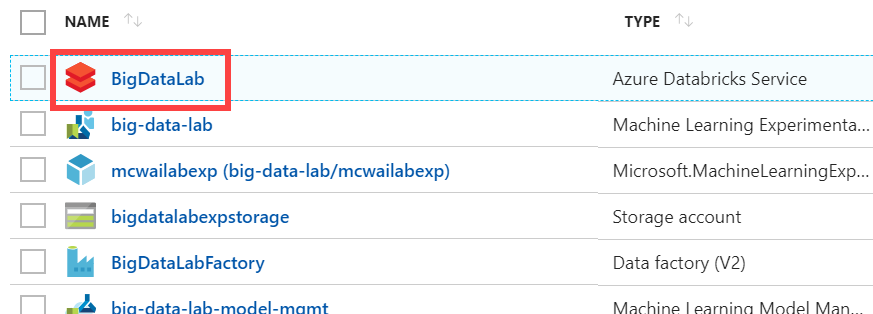
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-data-factory-create-blade.png)

1. Select **Create** to finish and submit.

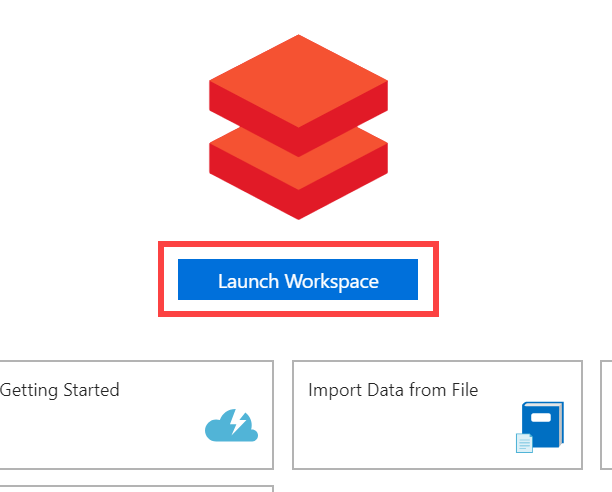
**Task 5: Create an Azure Databricks cluster**

You have provisioned an Azure Databricks workspace, and now you need to create a new cluster within the workspace. Part of the cluster configuration includes setting up an account access key to your Azure Storage account, using the Spark Config within the new cluster form. This will allow your cluster to access the lab files.

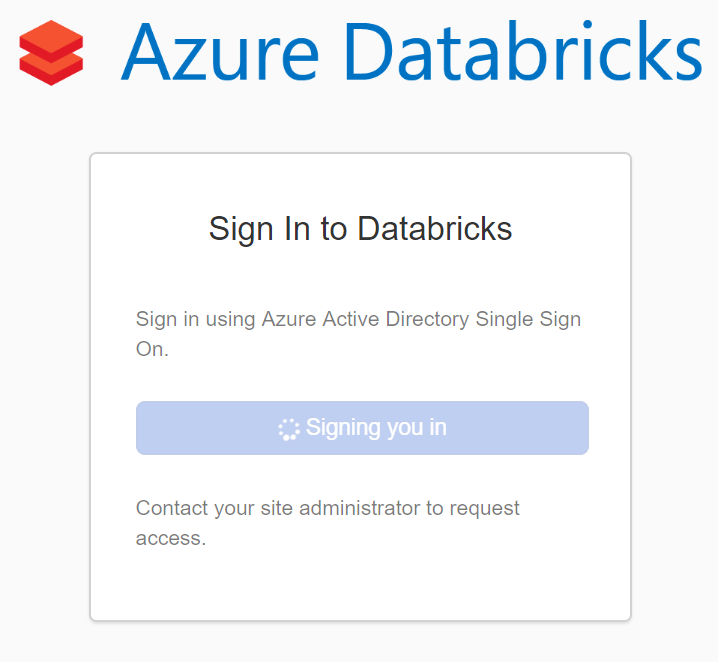
1. From the side menu in the Azure portal, select **Resource groups**, then enter your resource group name into the filter box, and select it from the list.
2. Next, select your Azure Databricks service from the list.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/select-azure-databricks-service.png)

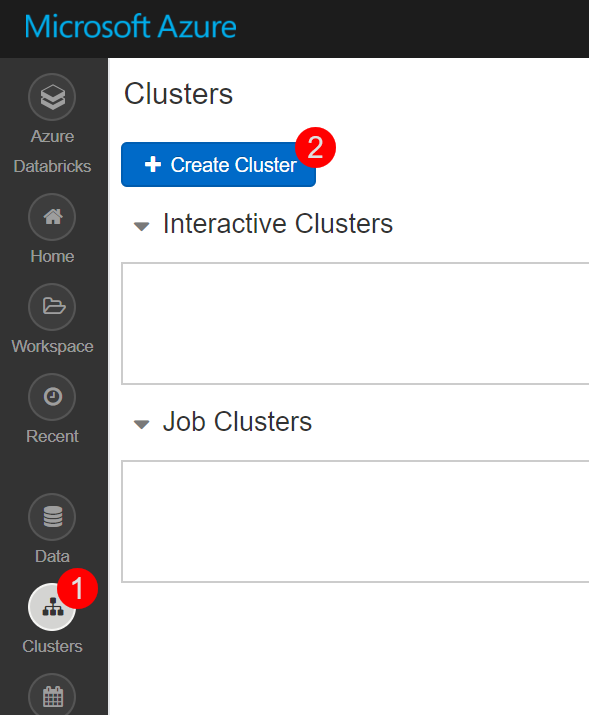
1. In the Overview pane of the Azure Databricks service, select **Launch Workspace**.

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-launch-workspace.png)

Azure Databricks will automatically log you in using Azure Active Directory Single Sign On.

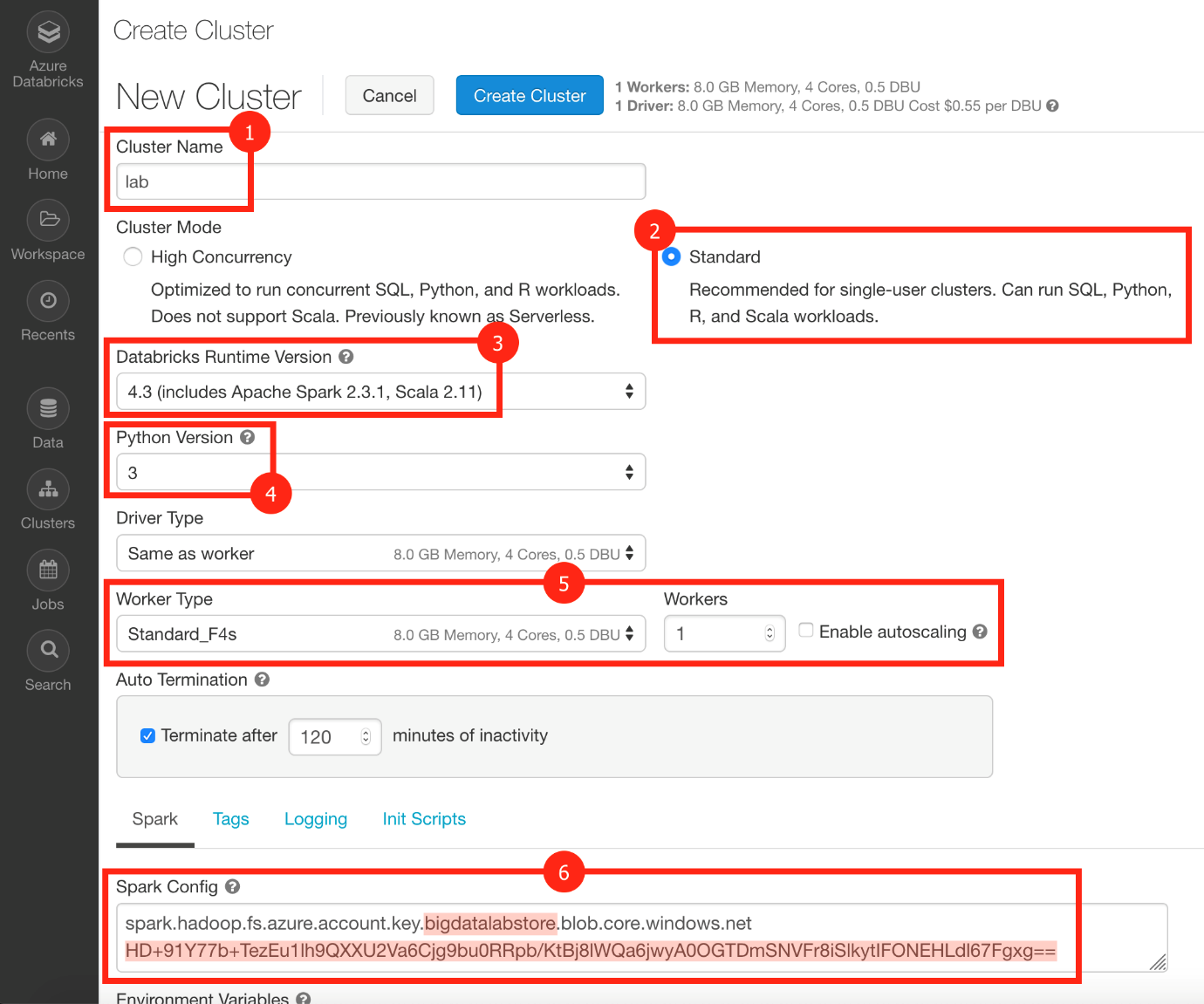
[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-aad.png)

1. Select **Clusters** (1) from the menu, then select **Create Cluster** (2).

[](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-cluster-button.png)

1. On the Create New Cluster form, provide the following:
   * **Cluster Name**: lab
   * **Cluster Type**: Standard
   * **Databricks Runtime Version**: 4.3 (includes Apache Spark 2.3.1, Scala 2.11)
   * **Python Version**: 3
   * **Driver Type**: Same as worker
   * **Worker Type**: Standard\_F4s
   * **Workers**: 1
   * **Enable Autoscaling**: Uncheck this option.
   * **Auto Termination**: Check the box and enter 120.
   * **Spark Config**: Edit the Spark Config by entering the connection information for your Azure Storage account that you copied earlier in Task 5. This will allow your cluster to access the lab files. Enter the following:

spark.hadoop.fs.azure.account.key.<STORAGE\_ACCOUNT\_NAME>.blob.core.windows.net <ACCESS\_KEY>, where <STORAGE\_ACCOUNT\_NAME> is your Azure Storage account name, and <ACCESS\_KEY> is your storage access key.

1. **Example:** spark.hadoop.fs.azure.account.key.bigdatalabstore.blob.core.windows.net HD+91Y77b+TezEu1lh9QXXU2Va6Cjg9bu0RRpb/KtBj8lWQa6jwyA0OGTDmSNVFr8iSlkytIFONEHLdl67Fgxg==
2. [](https://github.com/microsoft/MCW-Big-data-and-visualization/blob/master/Hands-on%20lab/media/azure-databricks-create-cluster-form.png)
3. Select **Create Cluster**.

You should follow all these steps provided *before* attending the Hands-on lab.