

## Before the hands-on lab

Duration: 30 minutes

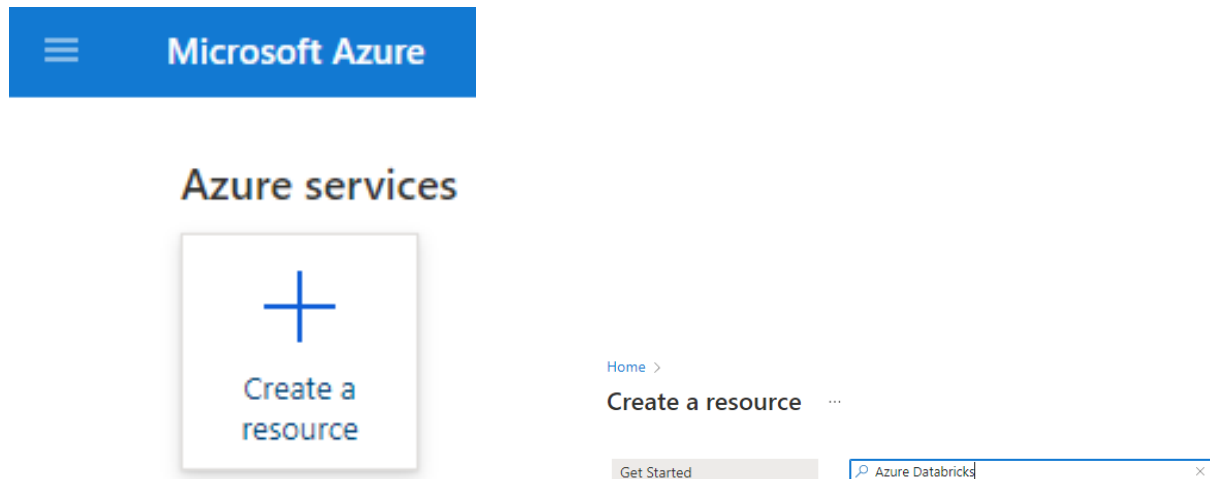
In this exercise, you will set up your environment for use in the rest of the hands-on lab. You should follow all the steps provided in the Before the Hands-on Lab section to prepare your environment *before* attending the hands-on lab.

**Recommendation:** Use an Azure VM to avoid installing any software in the laptop.  
Steps: <https://learn.microsoft.com/en-gb/azure/virtual-machines/windows/quick-create-portal>

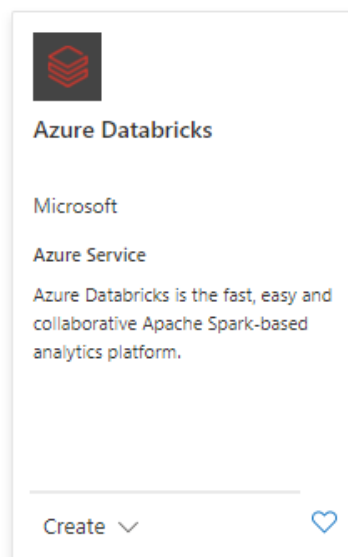
## Task 1: Provision Azure Databricks

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.

**Note:** To view the Azure portal menu, select the menu icon in the upper left-hand corner.



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



2. Select **Create**.

3. Set the following configuration on the Azure Databricks Service creation form:

- **Subscription:** Select the subscription you are using for this hands-on lab – Azure for Students
- **Resource Group:** Select **Create new** and enter a unique name, such as hands-on-lab-bigdata
- **Workspace name:** Enter a unique name, this is indicated by a green checkmark.
- **Location:** Select a region close to you. UK South
- **Pricing:** Select **Premium (+ Role-based access controls)**

## Create an Azure Databricks workspace ... ×

Basics   Networking   Advanced   Tags   Review + create

### Project Details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription * ⓘ	<input type="text" value="Azure for Students"/> <span>▼</span>
Resource group * ⓘ	<input type="text" value="hands-on-lab-bigdata"/> <span>▼</span> <a href="#">Create new</a>

### Instance Details

Workspace name *	<input type="text" value="BigDataLab"/> <span>✓</span>
Region *	<input type="text" value="UK South"/> <span>▼</span>
Pricing Tier * ⓘ	<input type="text" value="Premium (+ Role-based access controls)"/> <span>▼</span>

**i** We selected the recommended pricing tier for your workspace. You can change the tier based on your needs. ×

**Review + create**

[< Previous](#)

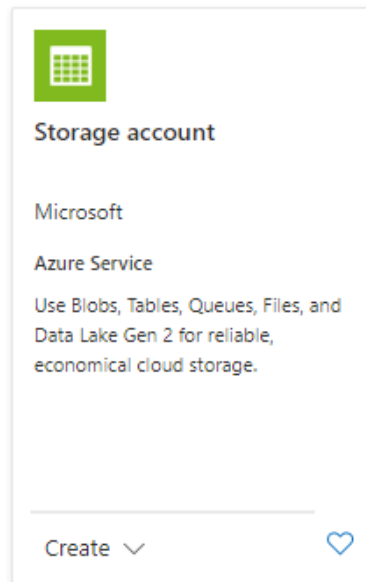
[Next : Networking >](#)

4. Select **Review + Create**.
5. Wait for validation to pass, then select **Create**.

## 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>), select + **Create a resource**, then type "storage" into the search bar. Select **Storage account** from the results.



2. Select **Create**.
3. Set the following configuration on the Azure Storage account creation form:
  - **Subscription:** Select the subscription you are using for this hands-on lab. Azure for Students
  - **Resource group:** Select the same resource group you created at the beginning of this lab. hands-on-lab-bigdata
  - **Storage account name:** Enter a unique name, this is indicated by a green checkmark.
  - **Location:** Select the same region you used for Azure Databricks.

- **Performance: Standard**
- **Replication: Geo-redundant storage (GRS)**
- **Made read access to data available in the event of regional unavailability.**

## Create a storage account ...



Basics   Advanced   Networking   Data protection   Encryption   Tags   Review

Azure Storage is a Microsoft-managed service providing cloud storage that is highly available, secure, durable, scalable, and redundant. Azure Storage includes Azure Blobs (objects), Azure Data Lake Storage Gen2, Azure Files, Azure Queues, and Azure Tables. The cost of your storage account depends on the usage and the options you choose below. [Learn more about Azure storage accounts](#)

### Project details

Select the subscription in which to create the new storage account. Choose a new or existing resource group to organize and manage your storage account together with other resources.

Subscription \* Azure for Students ▼

Resource group \* hands-on-lab-bigdata ▼

[Create new](#)

### Instance details

If you need to create a legacy storage account type, please click [here](#).

Storage account name ⓘ \* bigdatastore

Region ⓘ \* (Europe) UK South ▼

[Deploy to an edge zone](#)

Performance ⓘ \*   
☒ **Standard:** Recommended for most scenarios (general-purpose v2 account)   
☐ **Premium:** Recommended for scenarios that require low latency.

Redundancy ⓘ \* Geo-redundant storage (GRS) ▼

☒ Make read access to data available in the event of regional unavailability.

[Review](#)

[< Previous](#)

[Next : Advanced >](#)

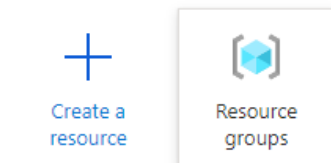
4. Select **Review + create**.
5. Wait for validation to pass, then select **Create**.

## Task 3: Create storage container

In this task, you will create a storage container 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.

### Azure services



### Resource groups

University of Reading (livereadingac.onmicrosoft.com)

+ Create ⚙️ Manage view ↻ Refresh ⬇️ Export to CSV 🔗 Open query 🏷️ Assign tags

Filter for any field...

Subscription equals all

Location equals all X

+ Add filter

🛡️ 0 Unsecure resources

🔗 0 Recommendations

No grouping

List view

<input type="checkbox"/> Name ↑↓	Subscription ↑↓	Location ↑↓	
<input type="checkbox"/> databricks-rg-BigDataLab-3mwkua7dv2khw	Azure for Students	UK South	...
<input type="checkbox"/> hands-on-lab-bigdata	Azure for Students	UK South	...
<input type="checkbox"/> NetworkWatcherRG	Azure for Students	UK South	...

2. Next, select your lab Azure Storage account from the list.

### hands-on-lab-bigdata

Resource group

+ Create ⚙️ Manage view 🗑️ Delete resource group ↻ Refresh ⬇️ Export to CSV 🔗 Open query 🏷️ Assign tags → Move ...

#### Essentials

JSON View

Subscription (move) : [Azure for Students](#)

Deployments : [3 Succeeded](#)

Subscription ID : 76068279-40db-4b9d-8f16-fa90826c76c1

Location : UK South

Tags (edit) : [Click here to add tags](#)

#### Resources

Recommendations

Filter for any field...

Type equals all X

Location equals all X

+ Add filter

Showing 1 to 3 of 3 records.

☐ Show hidden types ⓘ

No grouping

List view

<input type="checkbox"/> Name ↑↓	Type ↑↓	Location ↑↓	
<input type="checkbox"/> BigDataLab	Azure Databricks Service	UK South	...
<input type="checkbox"/> storebigdatalab	Storage account	UK South	...

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

The screenshot shows the Azure portal interface for a storage account named 'storebigdatalab'. The left sidebar contains a navigation menu with 'Containers' highlighted. The main area shows the 'Containers' blade with a table of containers. A '+ Container' button is visible. A 'New container' dialog box is open, showing the 'Name' field with 'sparkcontainer' and the 'Public access level' dropdown set to 'Private (no anonymous access)'.

Home > hands-on-lab-bigdata > storebigdatalab

storebigdatalab | Containers

Storage account

Search

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Data storage

Containers

+ Container

Change access level

Restore containers

Refresh

Search containers by prefix

Name	Last modified	Public access level
<input type="checkbox"/> \$logs	1/13/2023, 12:15:15 ...	Private
<input type="checkbox"/> sparkcontainer	1/13/2023, 12:17:25 ...	Private

New container

Name \*

sparkcontainer

Public access level ⓘ

Private (no anonymous access)

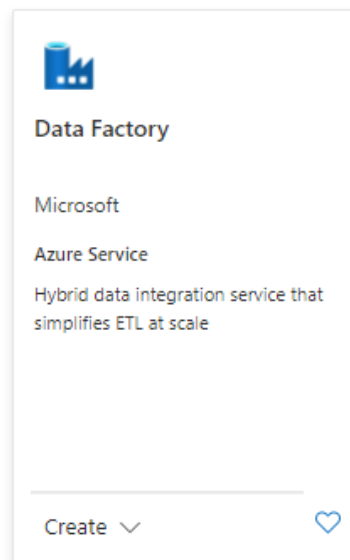
Advanced



## Task 4: Provision Azure Data Factory

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

In the [Azure Portal](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.



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

**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](#) 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.

## Create Data Factory ...



**Basics**   Git configuration   Networking   Advanced   Tags   Review + create

One-click to create data factory with sample pipeline and datasets. [Try it](#)

### Project details

Select the subscription to manage deployed resources and costs. Use resource groups like folders to organize and manage all your resources.

Subscription \* ⓘ

Azure for Students



Resource group \* ⓘ

hands-on-lab-bigdata



[Create new](#)

### Instance details

Name \* ⓘ

BigDataLabUpdateFactory



Region \* ⓘ

UK South



Version \* ⓘ

V2



**Review + create**

< Previous

Next : Git configuration >

[Give feedback](#)

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

## Task 5: Download and install Power BI Desktop

Power BI desktop is required to make a connection to your Azure Databricks environment when creating the Power BI dashboard.

Download and install [Power BI Desktop](#).

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