Lecture 7
Cloud Computing II
(Cloud Computing Platforms: Microsoft Azure and AWS)

Dr. Lydia Wahid

Agenda

- Microsoft Azure
- Amazon Web Services (AWS)
- Class Assignment
- Appendix A: Definitions
- Appendix B: AutoML classification

Microsoft Azure

What is Microsoft Azure?

- Azure is a cloud computing platform that allows you to access and manage cloud services and resources provided by Microsoft.
- Azure provides the following categories of services: **AI + machine learning**, **Analytics, Compute, Databases, Developer tools, DevOps, Hybrid + multicloud, Identity, Integration, Internet of Things, Management and governance, Media, Migration, Mixed reality, Mobile, Networking, Security, Storage, Virtual desktop infrastructure, Web.**
- Each category includes a list of services that Azure provides.

Microsoft Azure Services

- **Artificial Intelligence + Machine learning:**
  - Develop applications with the latest AI and ML capabilities.
  - The following are some of the services provided in this category:
    - Customize your own state-of-the-art computer vision models
    - Create bots and connect them across channels
    - Accelerate information extraction from documents
    - Translation, speech to text, text to speech, Language understanding

Microsoft Azure Services

- **Analytics:**
  - Gather, store, process, analyze, and visualize data of any variety, volume, or velocity
  - The following are some of the services provided in this category:
    - Design AI with Apache Spark-based analytics
    - Provision cloud Hadoop, Spark, R Server, HBase, and Storm clusters (check Appendix A)
    - Predictive analytics, machine learning, and statistical modeling for big data using Microsoft R server
    - Real-time analytics on fast-moving streaming data

Microsoft Azure Services

- **Compute:**
  - Access cloud compute capacity and scale on demand—and only pay for the resources you use
  - The following are some of the services provided in this category:
    - Create, manage, operate, and optimize HPC (check Appendix A) and big compute clusters of any scale
    - Create and provision virtual machines in Windows and Linux
    - Deploy and operate always-on, scalable, distributed apps
    - Run your VMware workloads on Azure

Microsoft Azure Services

- **Containers:**
  - Develop and manage your containerized applications
  - Containers are executable units of software in which application code is packaged, along with its libraries and dependencies, so that it can be run anywhere, whether it be on the cloud or on-premise infrastructure
- **DevOps:**
  - It is a combination of the terms development and operations. DevOps is a methodology meant to improve work throughout the software development lifecycle. You can visualize a DevOps process as an infinite loop, comprising these steps: plan, code, build, test, release, deploy, operate, monitor,….

Microsoft Azure Services

- **Hybrid + multicloud:**
  - Build and run hybrid apps across cloud boundaries
- **Identity:**
  - Manage user identities and access to protect against advanced threats across devices, data, apps, and infrastructure
- **Integration:**
  - Integrate on-premises and cloud-based applications, data, and processes across your enterprise

Microsoft Azure Services

- **Migration:**
  - Migrate to the cloud with guidance, tools, and resources
- **Mixed reality:**
  - Blend your physical and digital worlds
  - Automatically align and anchor 3D content to objects in the physical world

Microsoft Azure Machine Learning

- Azure Machine Learning includes several **resources** and **assets** to enable you to perform your machine learning tasks. These resources and assets are needed to run any job.
- **Resources**: setup or infrastructural resources needed to run a machine learning workflow. Resources include:
  - Workspace
  - Datastore
  - Compute
- **Assets**: created using Azure Machine Learning commands or as part of a training/scoring (check Appendix A) run. They include:
  - Model
  - Data
  - Environment
  - Component

Microsoft Azure Machine Learning: Resources

- **Workspace**
  - The workspace is the top-level resource for Azure Machine Learning, providing a **centralized place** to work with all the artifacts you create when you use Azure Machine Learning.
  - The workspace keeps a history of all jobs, including logs, metrics, output, and a snapshot of your scripts.
- **Datastore**
  - Azure Machine Learning datastores securely keep the connection information to your data storage on Azure, so you don't have to code it in your scripts.

Microsoft Azure Machine Learning: Resources

- **Compute:** Azure Machine Learning supports the following types of compute:
  - **Compute cluster** - a cluster of CPU or GPU compute nodes in the cloud.
  - **Compute instance** - a fully configured and managed development environment in the cloud
  - **Inference cluster** - used to deploy trained machine learning models to Azure Kubernetes Service (AKS) (check Appendix A)
  - **Attached compute** - You can attach your own compute resources to your workspace and use them for training and inference.

Microsoft Azure Machine Learning: Assets

- **Model**
  - Azure machine learning models consist of the binary file(s) that represent a machine learning model and any corresponding metadata.
- **Data**
  - Azure Machine Learning allows you to work with different types of data:
    - URIs (a location in local/cloud storage)
    - Tables (a tabular data abstraction)
    - Primitives (string, Boolean, number,..)

Microsoft Azure Machine Learning: Assets

- **Environment**
  - An encapsulation that specifies the **software packages**, **environment variables**, and **software settings** around your training and scoring scripts.
  - Azure Machine Learning supports two types of environments: curated and custom:
    - **Curated environments** Default. Intended to be used as is, they contain collections of Python packages and settings.
    - In **custom environments**, you're responsible for setting up your environment and installing packages or any other dependencies that your training or scoring script needs.

Microsoft Azure Machine Learning: Assets

- **Component**
  - An Azure Machine Learning component is a self-contained **piece of code** that does one step in a machine learning pipeline.
  - Components can do tasks such as **data processing**, **model training**, **model scoring**, and so on.
  - A component is analogous to a function - it has a name, parameters, expects input, and returns output.

Microsoft Azure AutoML

- Automated machine learning, also referred to as automated ML or AutoML, is the process of automating the iterative tasks of machine learning model development.

- Automated ML in Azure Machine Learning is based on a breakthrough from our Microsoft Research division.

Microsoft Azure AutoML

- **How does AutoML work?**
  - During training, Azure Machine Learning creates a number of pipelines in parallel that try different algorithms and parameters for you.
  - The service iterates through ML algorithms paired with feature selections, where each iteration produces a model with a training score.
  - The better the score for the metric you want to optimize for, the better the model is considered to "fit" your data. It will stop once it hits the exit criteria defined in the experiment.

Microsoft Azure AutoML

- Using **Azure Machine Learning**, you can design and run your automated ML training experiments with these steps:
  1. **Identify the ML problem** to be solved.
  2. **Choose whether you want a code-first experience or a no-code studio web experience**:
     - Code-first experience: you can use the Azure Machine Learning SDKv2 or the Azure Machine Learning CLIv2.
     - Limited/no-code experience: you can use the web interface in Azure Machine Learning studio at https://ml.azure.com (this link will take you to your account if signed in)

Microsoft Azure AutoML

- Using **Azure Machine Learning**, you can design and run your automated ML training experiments with these steps:
  3. **Specify the source of the labeled training data**: You can bring your data to Azure Machine Learning either from your local machine or an existing cloud-based storage.
  4. **Configure the automated machine learning parameters** that determine how many iterations over different models, hyperparameter settings, advanced preprocessing/featurization, and what metrics to look at when determining the best model.
  5. **Submit the training job**
  6. **Review the results**

Microsoft Azure AutoML: Demo Example

- **Task:** Train a classification model with AutoML in the Azure Machine Learning studio.
- Check Appendix B

Microsoft Azure Cognitive Service

- Azure Cognitive Service for Language is a cloud-based service that provides Natural Language Processing (NLP) features for understanding and analyzing text such as:
  - Named Entity Recognition (NER), Personally Identifying Information (PII)
  - Text analytics for health
  - Key phrase extraction
  - Language detection
  - Sentiment analysis and opinion mining
  - Question answering
  - Summarization
  - Custom Named Entity Recognition (Custom NER)
  - Custom text classification
  - Conversational language understanding

Microsoft Azure Cognitive Service

- You can build applications using the web-based Language Studio, REST APIs, and client libraries.
- You can access the language studio from: language.cognitive.azure.com
- Note that you first create a **language resource** in azure portal.

Amazon Web Services (AWS)

Amazon Web Services (AWS)

- Amazon Web Services (AWS) provide on-demand cloud computing platforms and APIs to individuals, companies, and governments, on a metered, pay-as-you-go basis.

- AWS categories of services include **computing**, **storage**, **networking**, **database**, **analytics**, **application services**, **deployment**, **management**, **machine learning**, **mobile**, **developer tools**, and tools for the **Internet of Things**.

Amazon Web Services (AWS)

- **Amazon EC2:** Amazon Elastic Compute Cloud
  - Allows users to create and rent virtual computers to run their own applications.
  - Allows users to configure their own virtual machine (which is called an "instance") such as processor, storage, networking, operating system.
- **Amazon S3:** Amazon Simple Storage Service
  - Provides object storage
  - Manages data as objects (an object includes: data itself, metadata, and a globally unique identifier)

Amazon Web Services (AWS)

- **Amazon VPC:** Amazon Virtual Private Cloud
  - Provides provisioning to a logically isolated section of AWS Cloud.
  - Customers access the EC2 over an IPsec (Check Appendix A) based virtual private network.
- **Amazon EMR:** Amazon Elastic MapReduce
  - It is cluster platform for running big data frameworks, such as Apache Hadoop and Apache Spark, on AWS to process and analyze vast amounts of data.

Amazon Web Services (AWS): EMR

- The central component of Amazon EMR is the **cluster**.
- A cluster is a collection of **Amazon EC2 instances**. Each instance in the cluster is called a *node*.
- Each node has a role within the cluster, referred to as the *node type*.
  - **Primary node:** A node that manages the cluster by running software components to coordinate the distribution of data and tasks among other nodes for processing.
  - **Core node:** A node with software components that run tasks and store data in the HDFS on your cluster.
  - **Task node:** A node with software components that only runs tasks and does not store data in HDFS. Task nodes are optional.

Amazon Web Services (AWS): EMR

- When you launch your cluster, you choose the frameworks and applications to install for your data processing needs such as: Hive, Hadoop, Spark, and so on.
- Amazon EMR service architecture consists of several layers, each of which provides certain capabilities and functionality to the cluster:
  - Storage
  - Cluster resource management
  - Data processing frameworks
  - Applications and programs

Amazon Web Services (AWS): EMR

- Amazon EMR service architecture layers:
  - **Storage:** There are several different types of storage:
    - **HDFS:** it is reclaimed when you terminate a cluster
    - **EMR File System (EMRFS):** directly access data stored in Amazon S3 as if it were a file system like HDFS
    - **Local file system:** refers to a locally connected disk

Note: When you create a Hadoop cluster, each node is created from an Amazon EC2 instance. Data on instance store volumes persists only during the lifecycle of its Amazon EC2 instance.)

Amazon Web Services (AWS): EMR

- Amazon EMR service architecture layers:
  - **Cluster resource management:**
    - responsible for managing cluster resources and scheduling the jobs for processing data.
    - By default, Amazon EMR uses YARN.
  - **Data processing frameworks:**
    - It is the engine used to process and analyze data.
    - The main processing frameworks available for Amazon EMR are Hadoop MapReduce and Spark.

Amazon Web Services (AWS): EMR

- Amazon EMR service architecture layers:
  - **Applications and programs:**
    - Amazon EMR supports many applications, such as Hive, Pig, and the Spark Streaming library.
    - You can use libraries and languages to interact with the applications that you run in Amazon EMR. For example, you can use Java, Hive, or Pig with MapReduce or Spark Streaming, Spark SQL, MLlib, and GraphX with Spark.

References

- azure.microsoft.com/en-us/products/
- learn.microsoft.com/en-us/azure/machine-learning/concept-azure-machine-learning-v2?tabs=sdk
- learn.microsoft.com/en-us/azure/machine-learning/concept-automated-ml
- learn.microsoft.com/en-us/azure/cognitive-services/language-service/
- aws.amazon.com/products
- docs.aws.amazon.com/emr/latest/ManagementGuide/emr-what-is-emr.html

Class Assignment

Class Assignment

- You are required to use Microsoft Azure and apply it on a problem of your choice.
- The problem can fall in the following categories: Classification, Regression, Time series forecasting, Natural language processing and Computer vision.
- Bonus grades will be given for the following:
  - Choosing either Time series forecasting or NLP or Computer vision problem.
  - OR using Azure Machine Learning SDKv2
  - OR using Azure Machine Learning CLIv2
- Delivery date: During the lecture of week12
- Assignment grade: 5 marks (without the bonus)

Appendix A: Definitions

Appendix A: Definitions

- **Storm:** Apache Storm is a distributed stream processing computation framework
- **Microsoft R Server:** is now Microsoft Machine Learning Server which is a flexible enterprise platform for analyzing data at scale, building intelligent apps, and discovering valuable insights across a business now with full support for Python and R.
- **HPC:** High-performance computing (HPC), also called "big compute", uses a large number of CPU or GPU-based computers to solve complex mathematical tasks.

Appendix A: Definitions

- **Model scoring:** In machine learning, model scoring is the process of assessing the accuracy of a model.
- **Azure Kubernetes Service:** Azure Kubernetes Service (AKS) is a cloud service that lets you run Kubernetes clusters without managing the underlying infrastructure. Kubernetes is an open-source system for automating the deployment, scaling, and management of containerized applications across multiple nodes.
- **Deploy:** Software deployment is to make a software system available to its intended users.

Appendix A: Definitions

- **IPsec:** Internet Protocol Security (IPsec) is a secure network protocol suite that authenticates and encrypts packets of data to provide secure encrypted communication between two computers over an Internet Protocol network. It is used in virtual private networks (VPNs).

Appendix B: AutoML classification

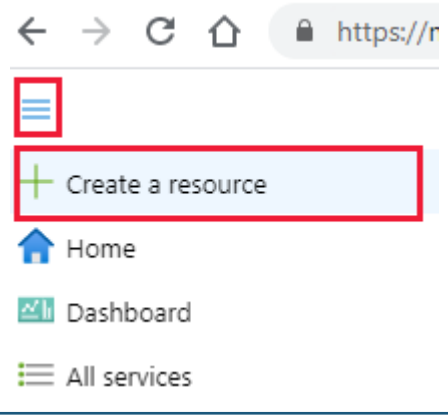
Microsoft Azure AutoML: Demo Example

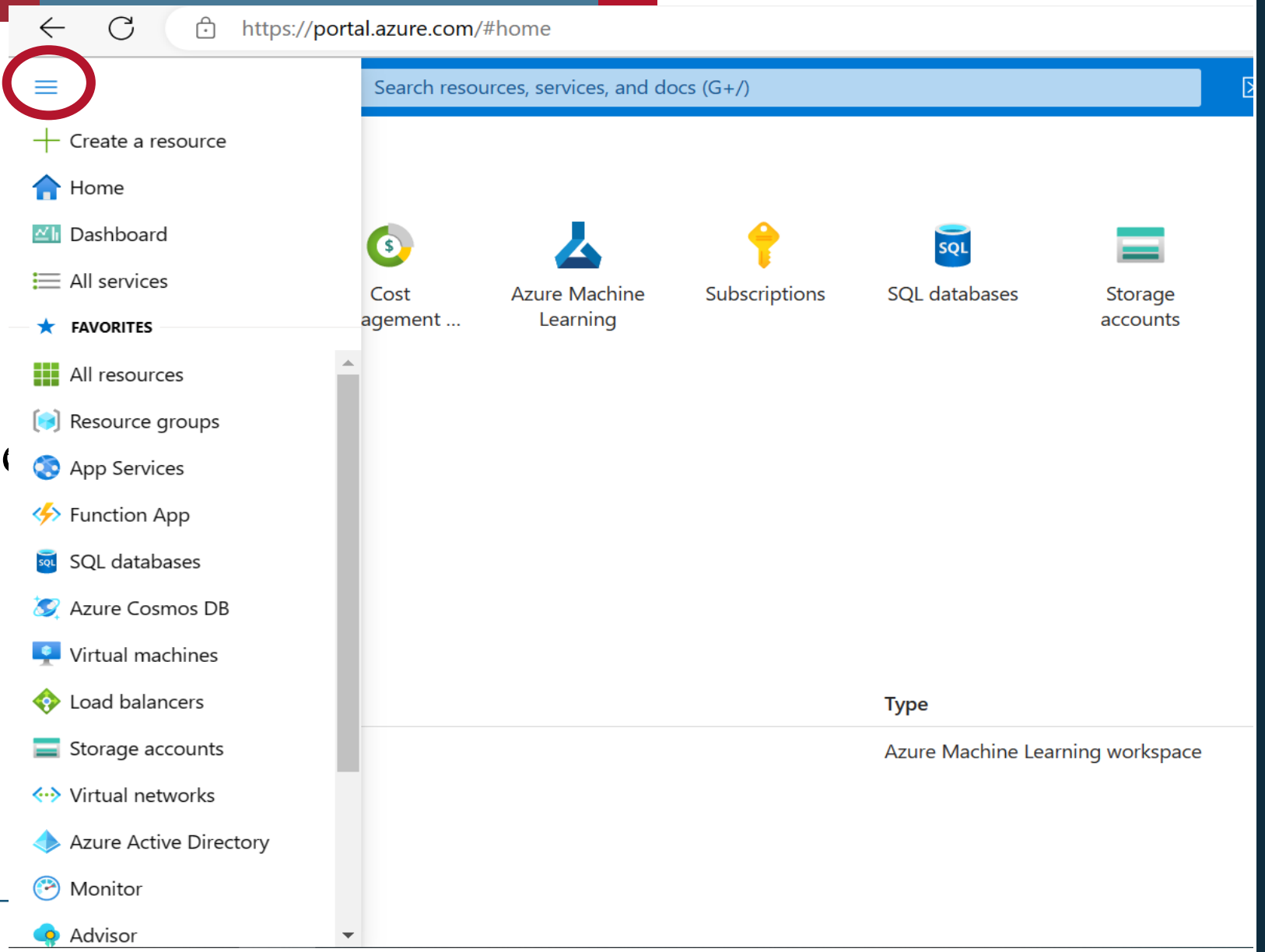
1. Create a workspace
2. Go to Azure Machine Learning studio
3. Create and load dataset
4. Configure job
5. Configure task
6. Explore the models

1. Create a workspace

1.1. Sign in to azure portal (portal.azure.com)

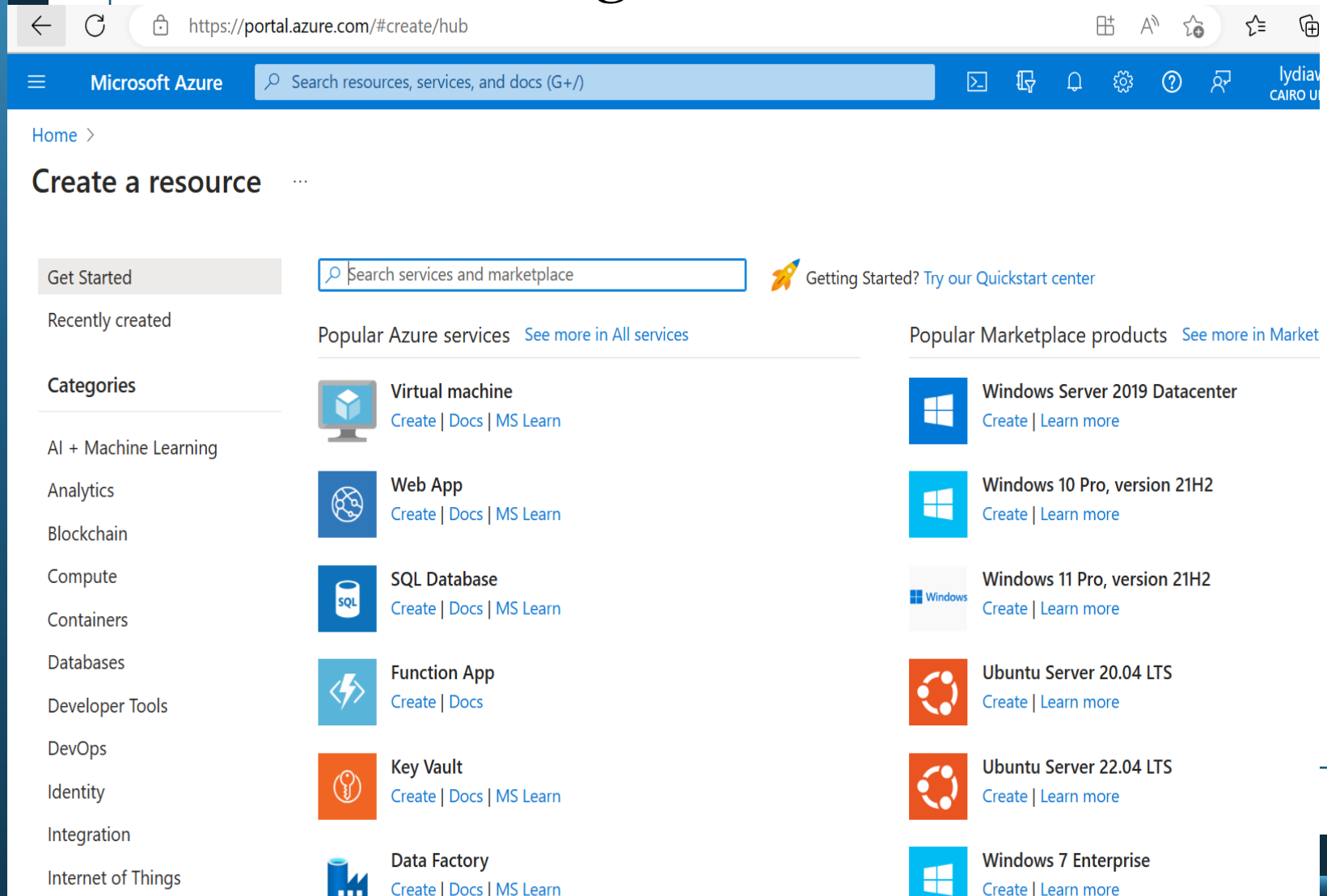
1.2. Create a resource

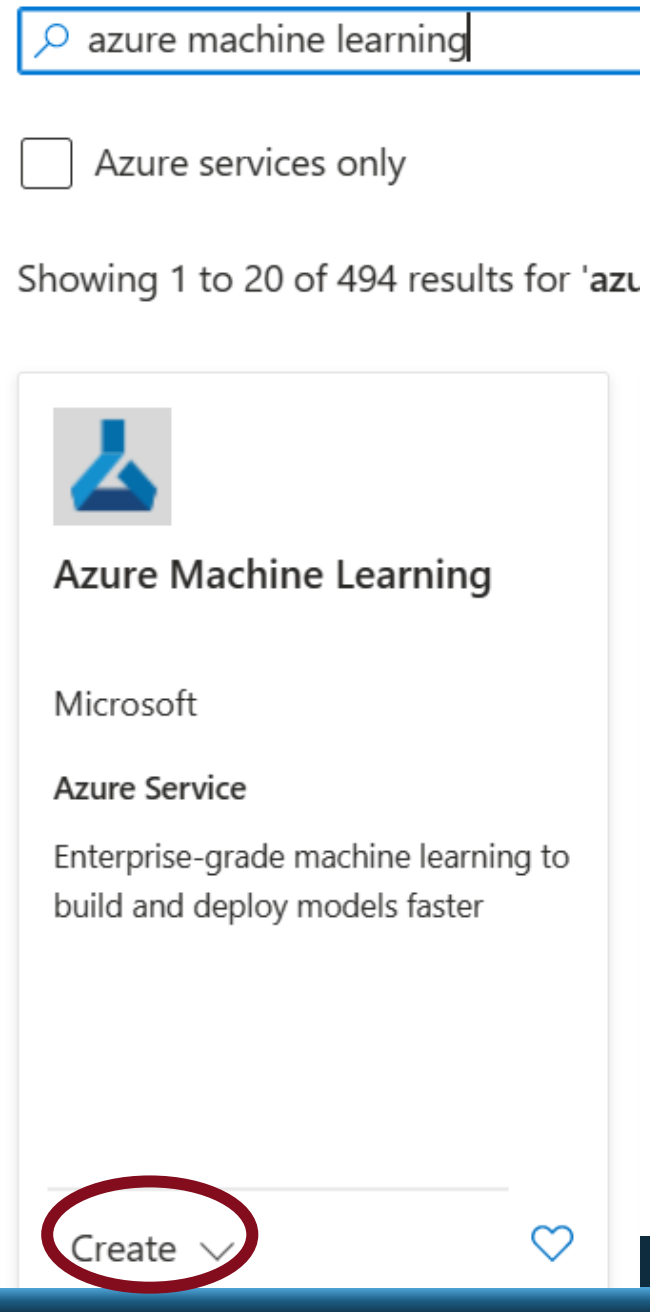




1. Create a workspace

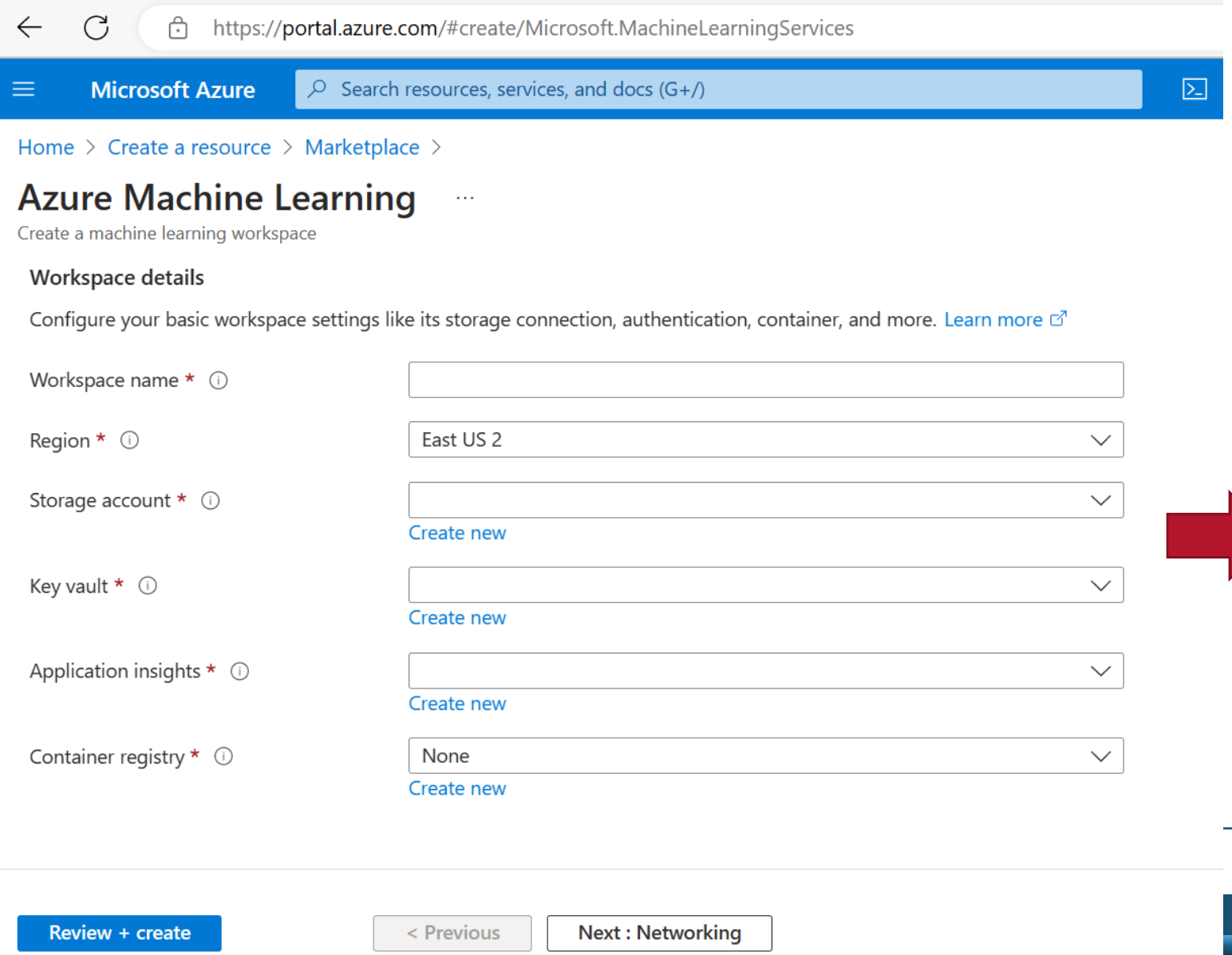
- **1.3.** Use the search bar to find and create **Azure Machine Learning**.

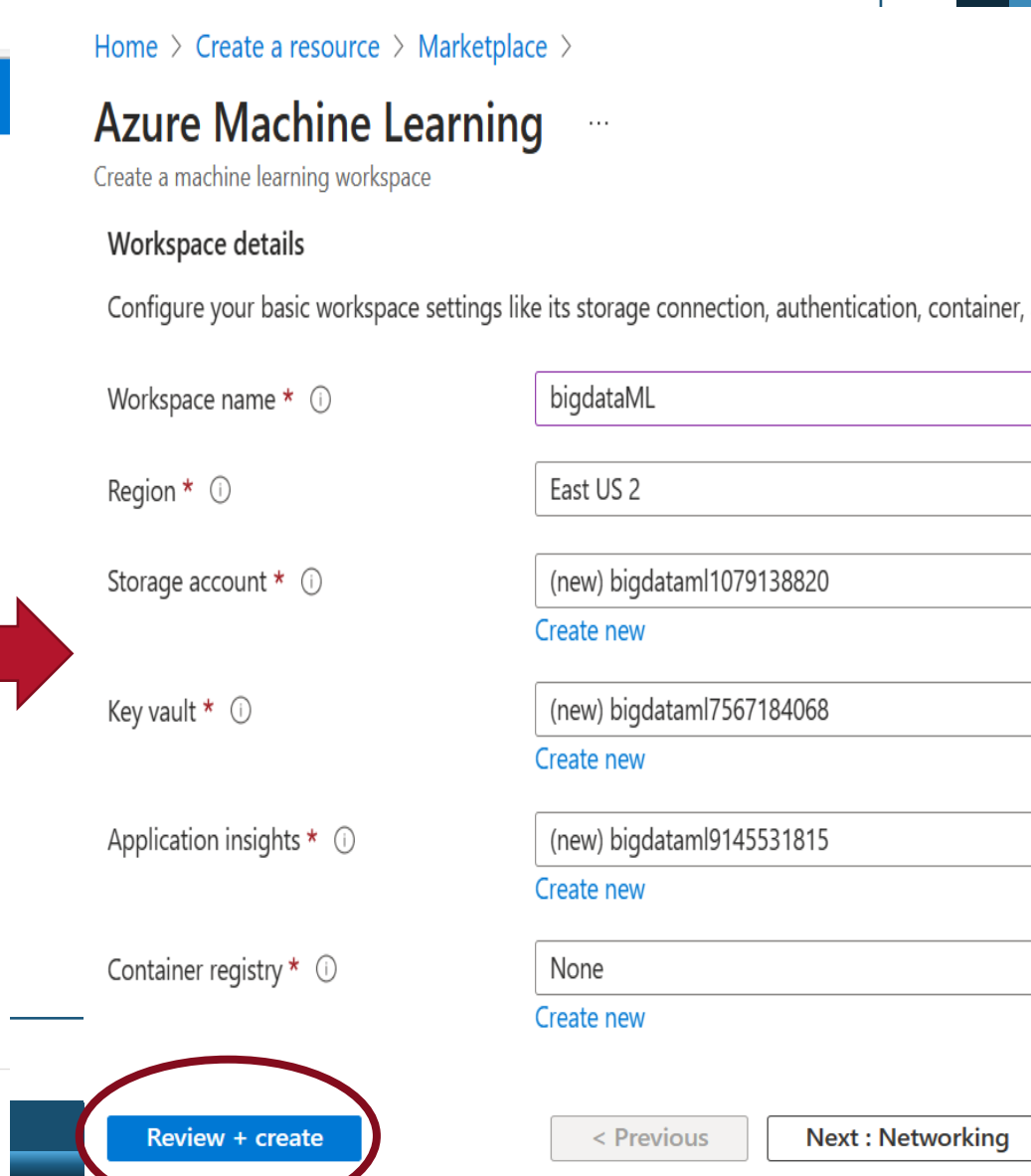




1. Create a workspace

➢1.4. Enter the details of the workspace to create it





Microsoft Azure | Search resources, services, and docs (G+/)

Home > Create a resource > Marketplace >

Azure Machine Learning

Create a machine learning workspace

Validation passed

Workspace name	bigdataML
Storage account	(new) bigdataml1079138820
Key vault	(new) bigdataml7567184068
Application insights	(new) bigdataml9145531815
Container registry	None

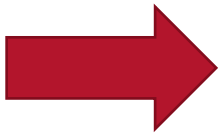
Networking

Connectivity method	Enable public access from all networks

Advanced

Identity type	System assigned
Encryption type	Microsoft-managed keys
Enable HBI Flag	Disabled

Create | < Previous | Next > | Download a templ



Your deployment is complete

Deployment name: Microsoft.MachineLe... Start t

Subscription:

Resource group:

Corre

Deployment details

Next steps

Go to resource

2. Go to Azure Machine Learning studio

Home > Microsoft.MachineLearningServices | Overview >

bigdataML

Azure Machine Learning workspace

Search

- Overview
- Activity log
- Access control (IAM)
- Tags
- Diagnose and solve problems
- Events

Settings

- Networking
- Properties
- Locks

Monitoring

- Alerts

Download config.json Delete

Subscription ID
16780f4e-459d-49fc-a347-c33ecb23cfb4

Storage
bigdataml1079138820

Application Insights
bigdataml9145531815

MLflow tracking URI
azureml://eastus2.api.azureml.ms/mlflow/v1.0/subscriptions/16780f4...

Work with your models in Azure Machine Learning Studio

The Azure Machine Learning Studio is a web app where you can build, train, test, and deploy ML models. Launch it now to start exploring, or learn more about the Azure Machine Learning Studio

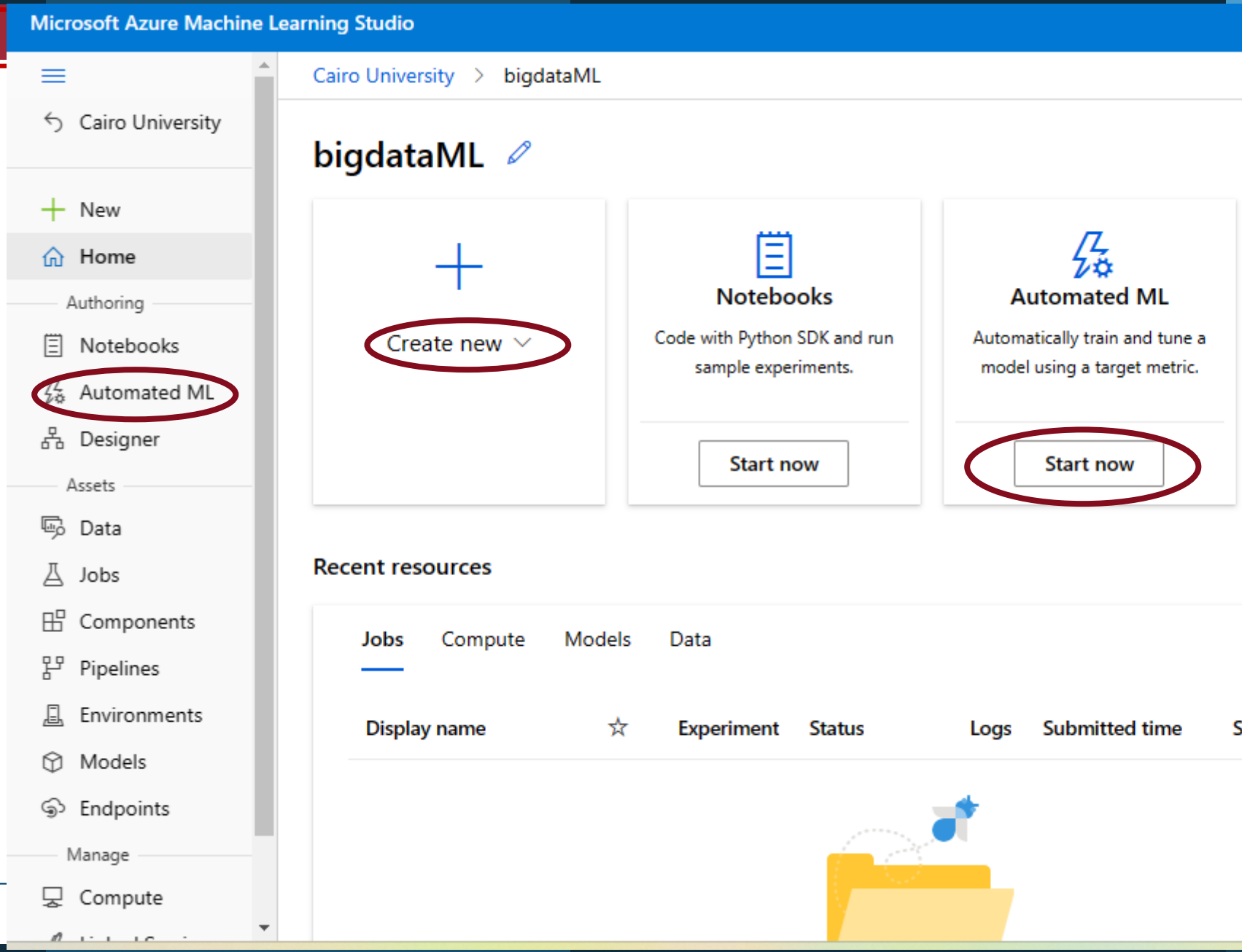
Launch studio

Or go to: ml.azure.com

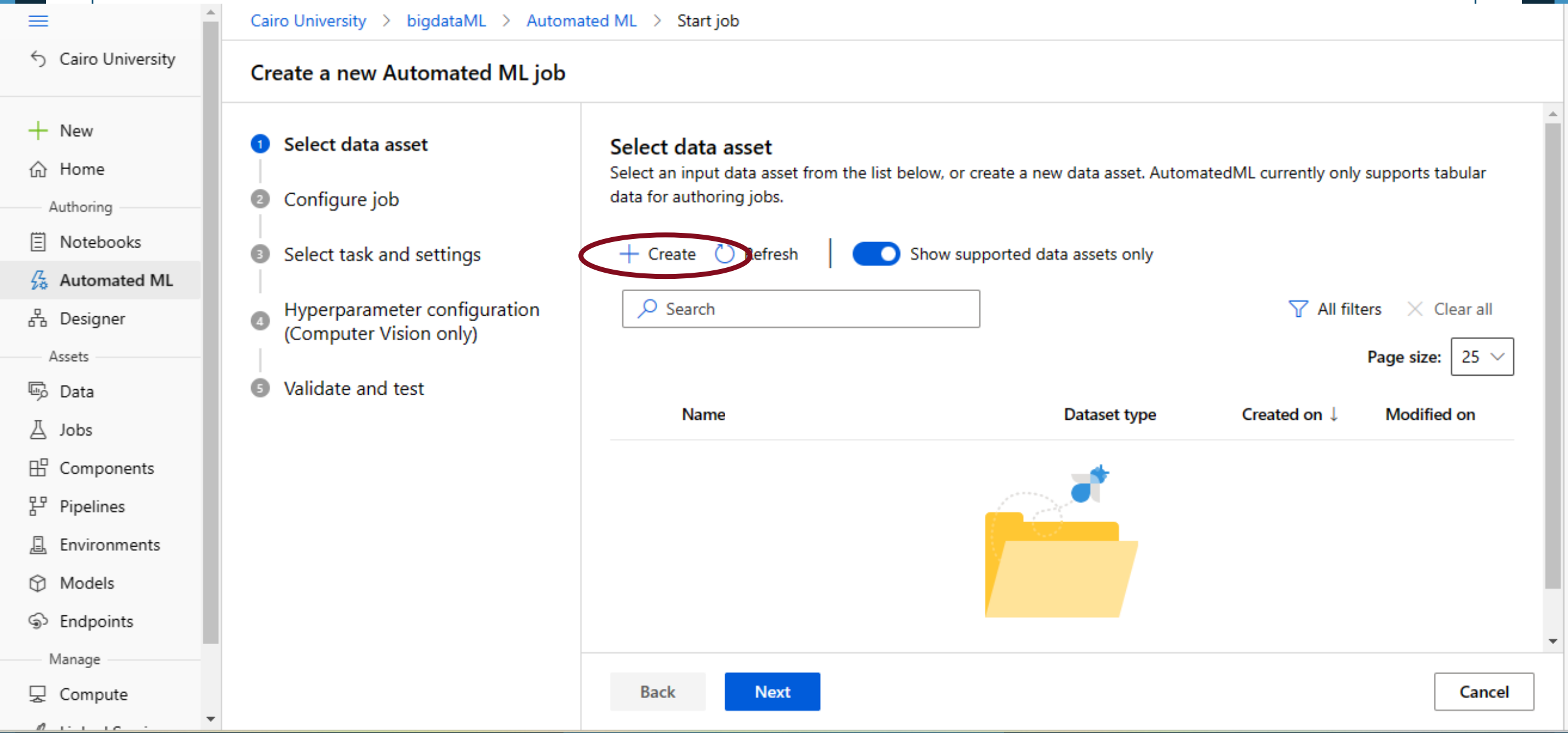
2. Go to Azure Machine Learning studio

2.1 Select your subscription and the workspace you created.

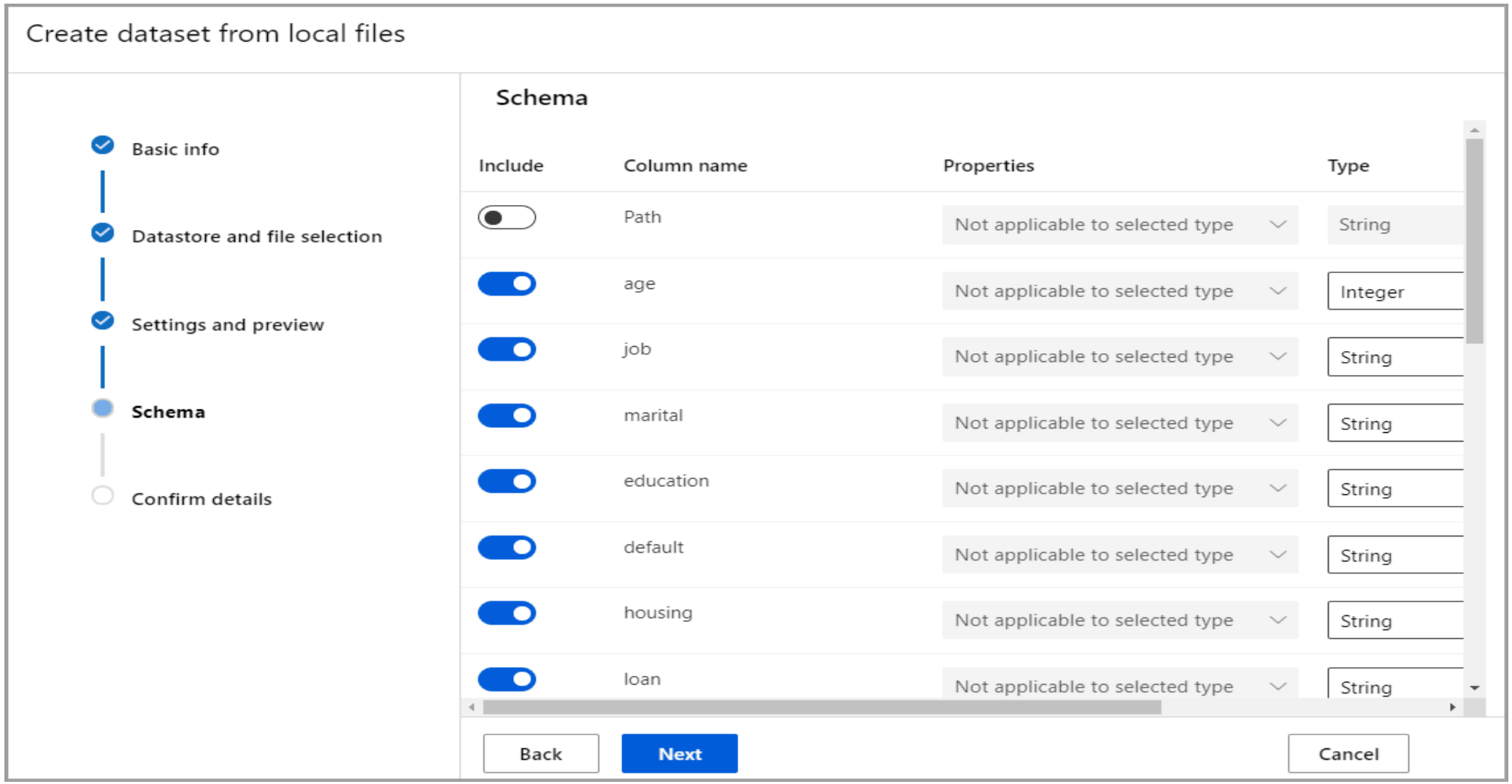
2.2. Create a new AutoML job: In the left pane, select **Automated ML** OR from **Create new** OR Click **Start now** in the Automated ML icon.



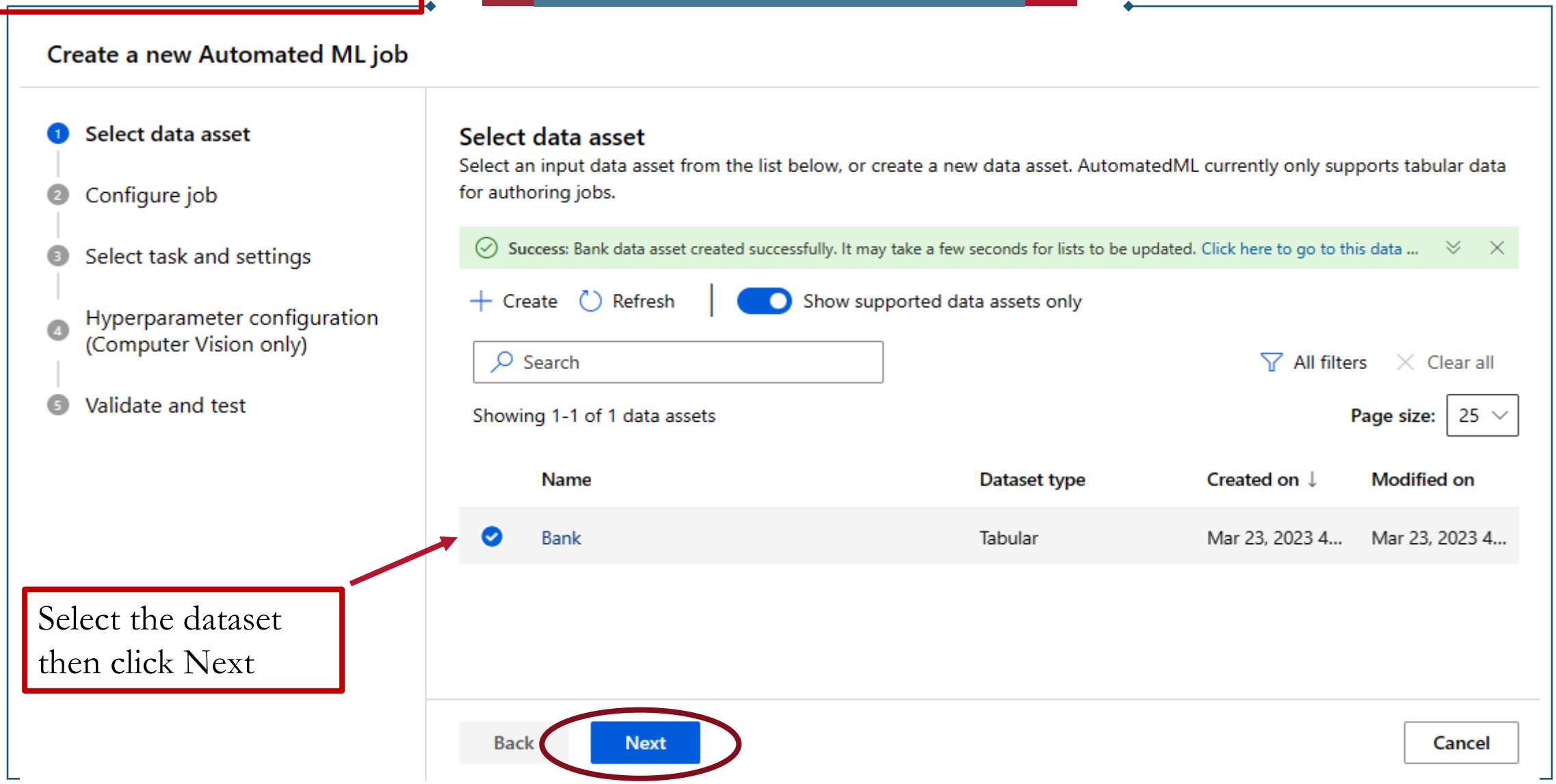
3. Create and load dataset



3. Create and load dataset



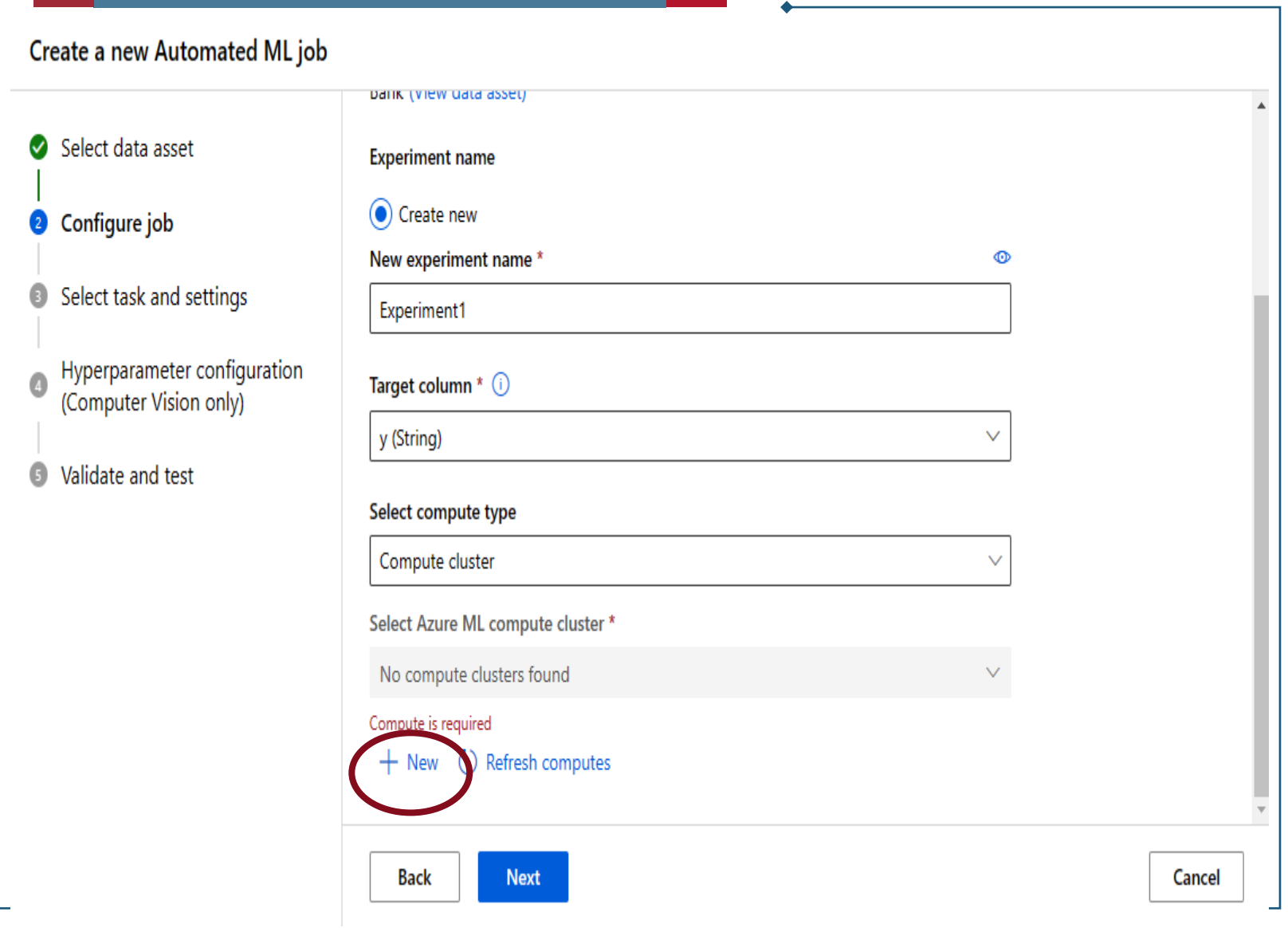
3. Create and load dataset



Select the dataset then click Next

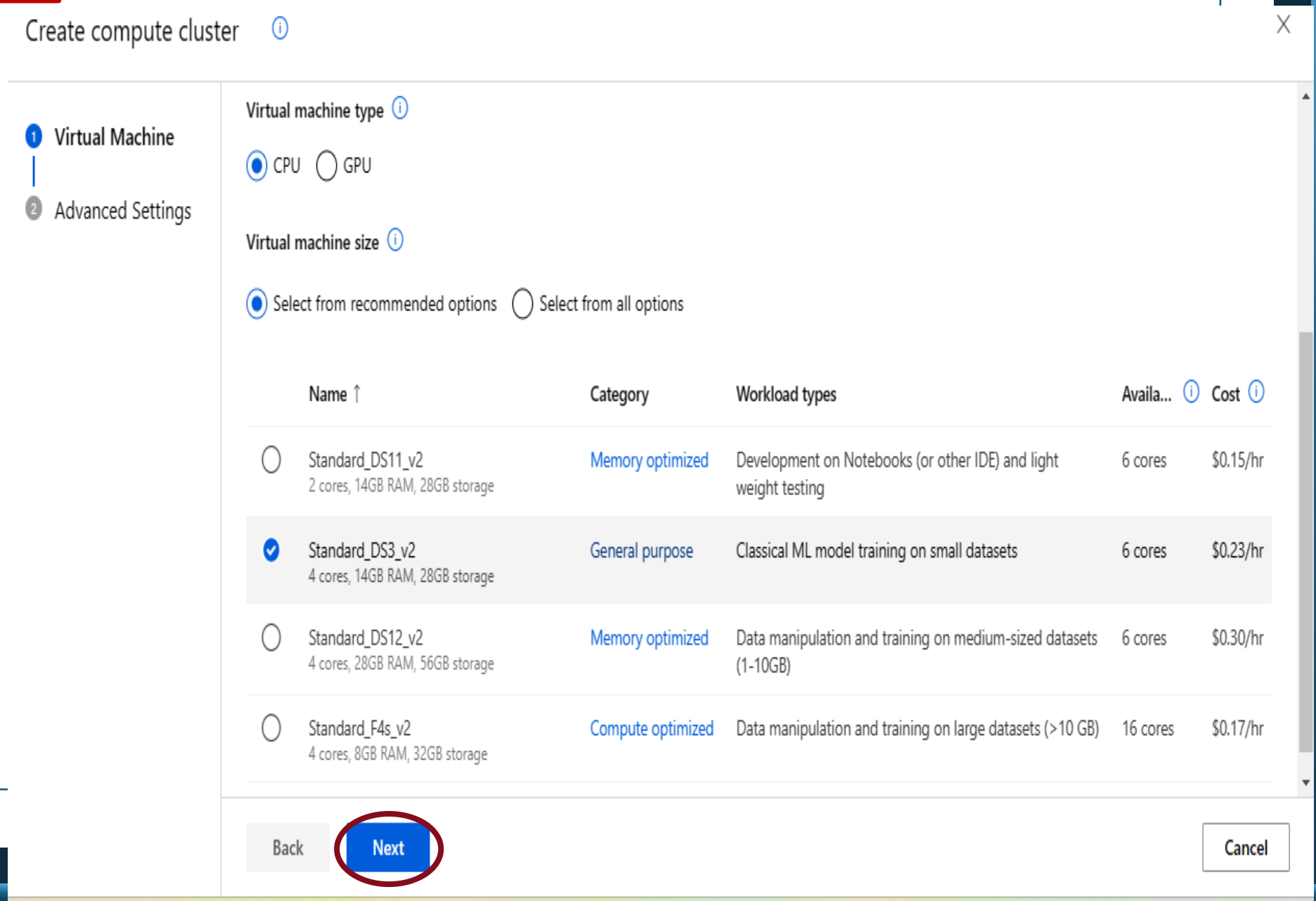
4. Configure job

4.1. Enter the details



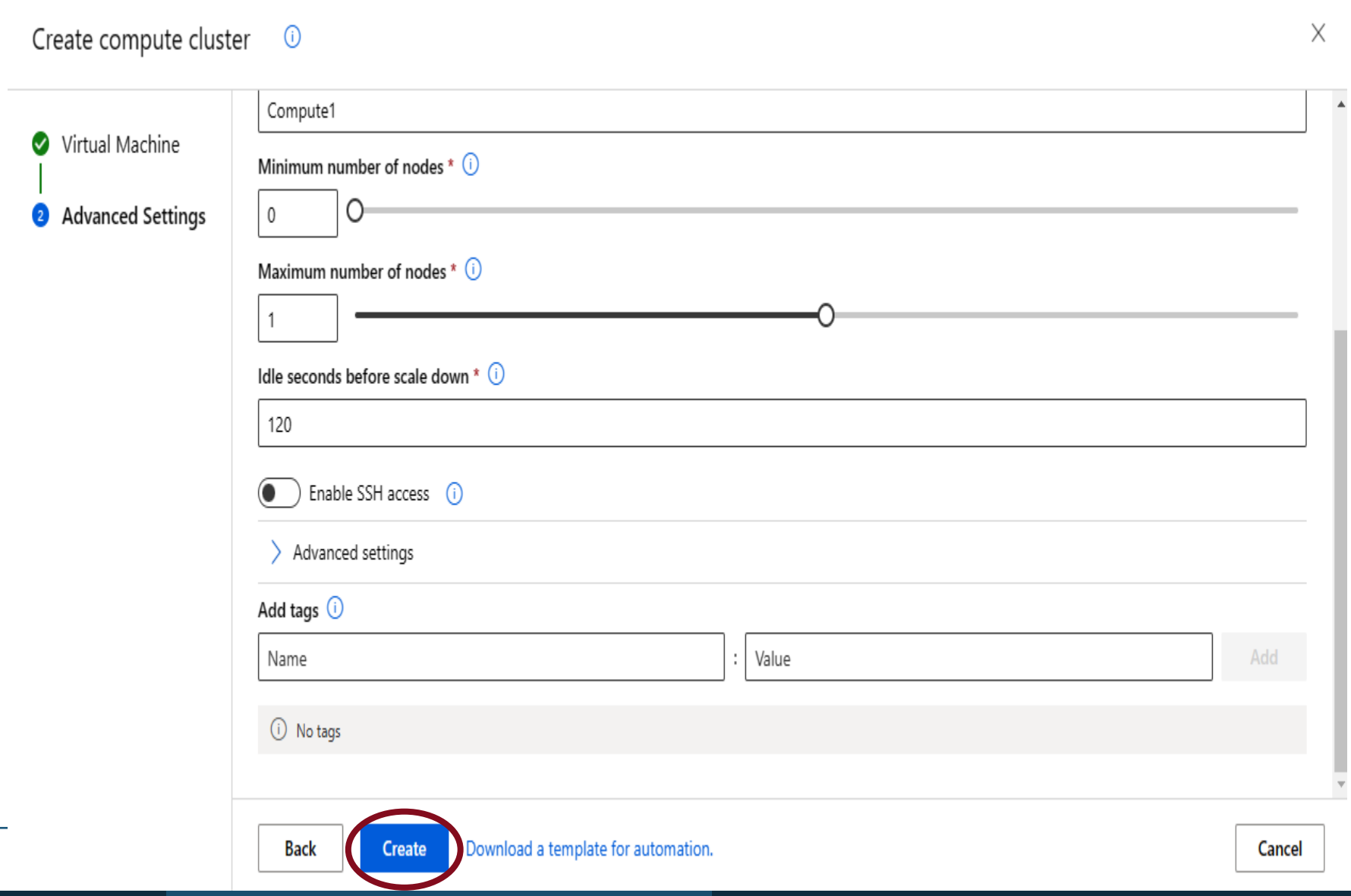
4. Configure job

4.2. Create compute cluster

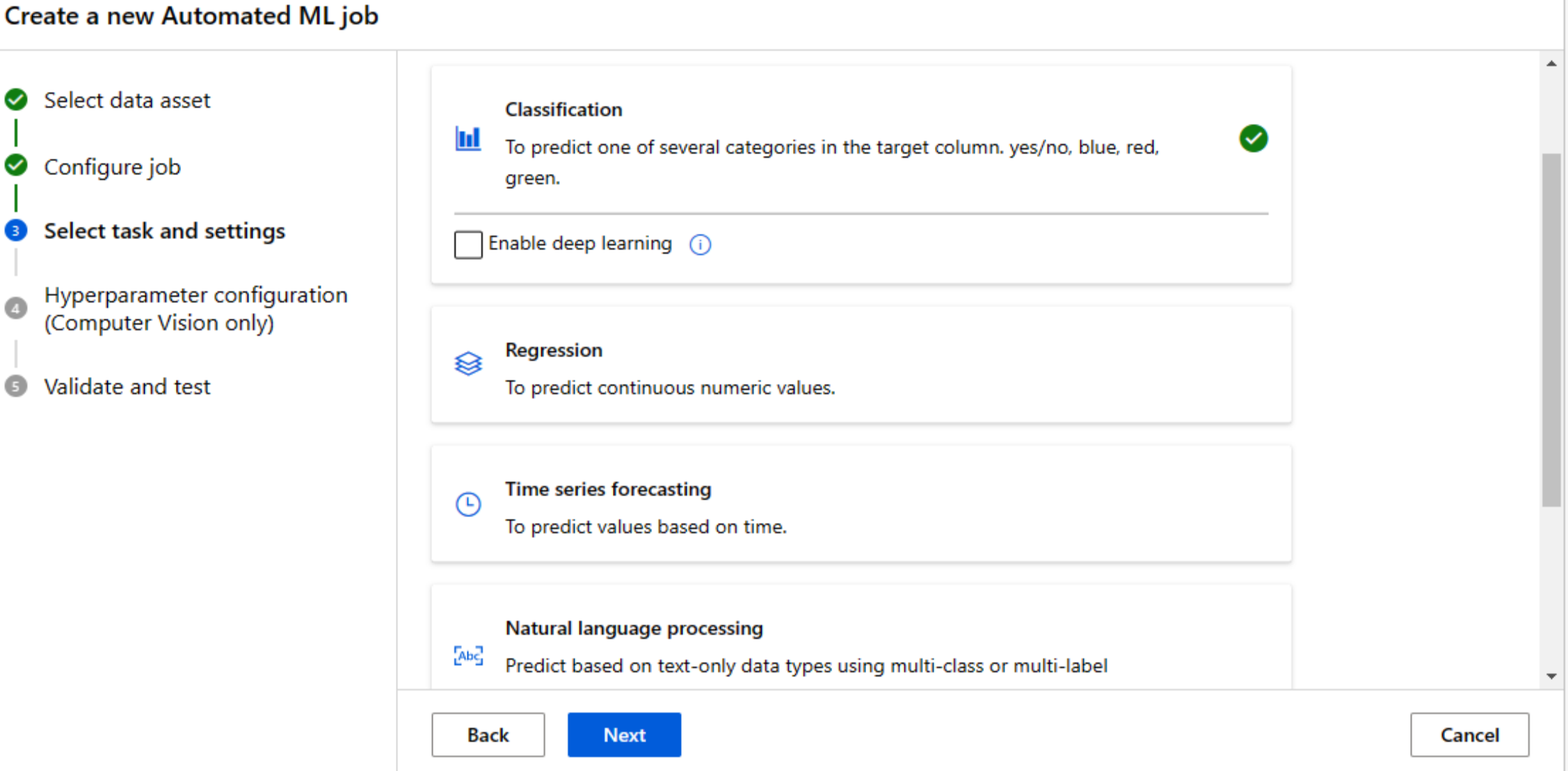


4. Configure job

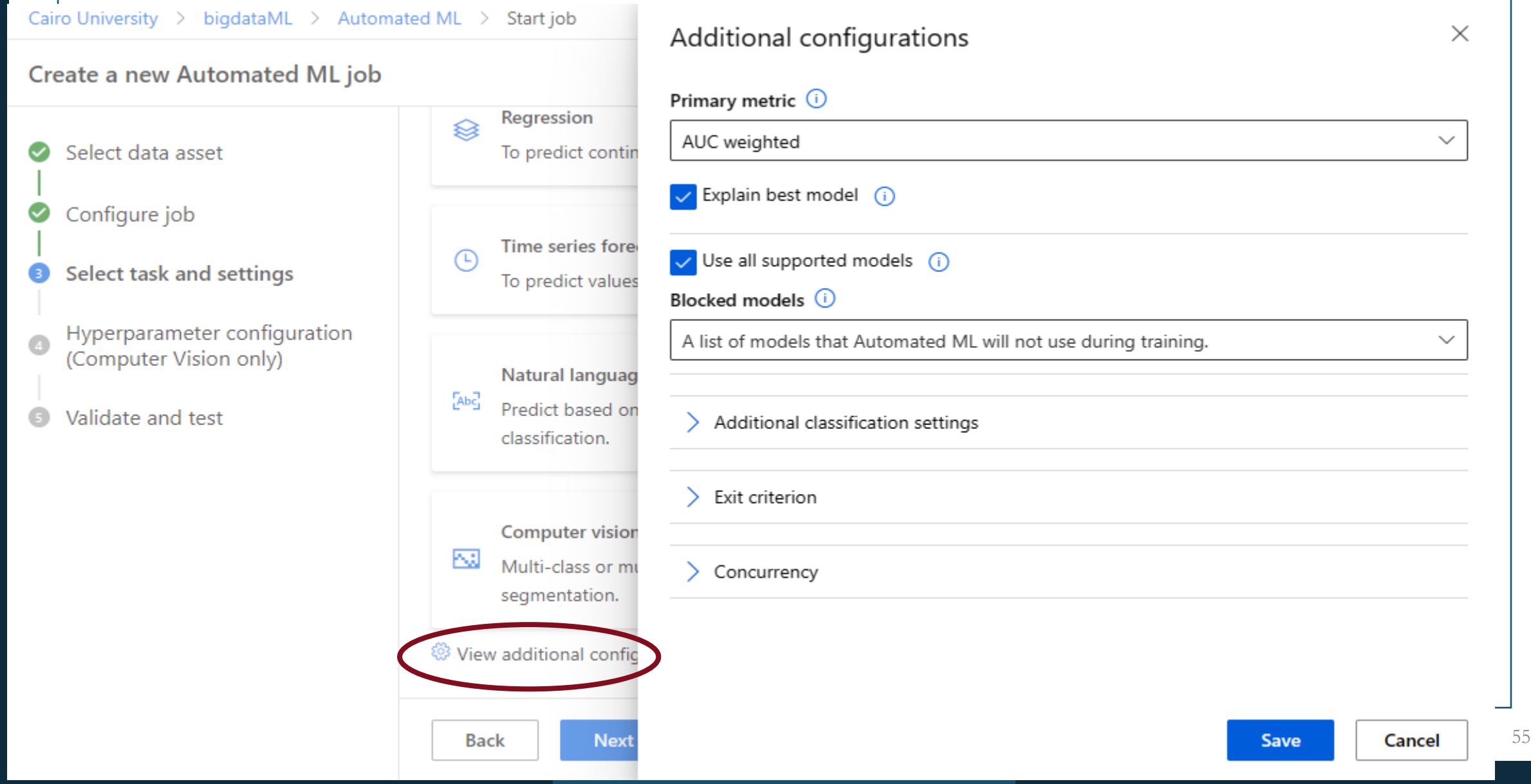
4.3. Configure your cluster and then click create



5. Configure task



5. Configure task



➢The job is currently running

Cairo University > bigdataML > Automated ML > Experiment1 > polite_sun_4wjl4k40

polite_sun_4wjl4k40 ☆ Running

Overview | Data guardrails | Models | Outputs + logs | Child jobs

Refresh | Edit and submit (preview) | Register model | Cancel | Delete | Compare (preview)

Properties

Status
Running

Setting up the run

Validating run configuration

Created on
Mar 23, 2023 5:49 PM

Start time
Mar 23, 2023 5:49 PM

Compute target
Compute1

Name

Inputs

Input name: training_data
Dataset: Bank:1

Best model summary

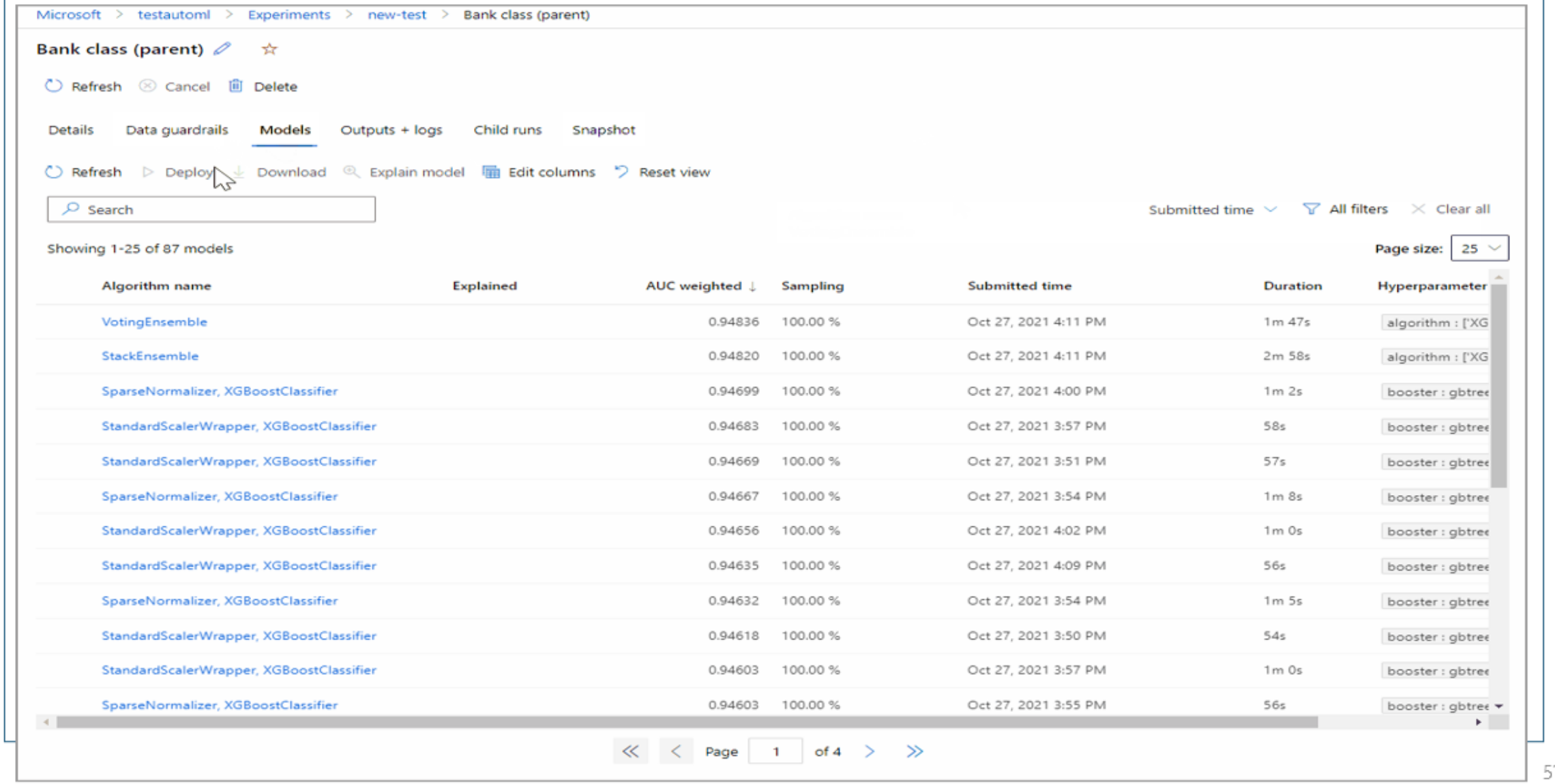
No data

Run summary

Task type
Classification View configuration settings

Featurization
Auto

6. Explore the models



Thank You