**AZURE DATA BRICKS**

The Azure Databricks Lakehouse Platform provides a unified set of tools for building, deploying, sharing, and maintaining enterprise-grade data solutions at scale. Azure Databricks integrates with cloud storage and security in your cloud account, and manages and deploys cloud infrastructure on your behalf.

Azure Databricks **provides the latest versions of Apache Spark and allows you to seamlessly integrate with open-source libraries**. Spin up clusters and build quickly in a fully managed Apache Spark environment with the global scale and availability of Azure.

## **What is Azure Databricks used for?**

Our customers use Azure Databricks to process, store, clean, share, analyze, model, and monetize their datasets with solutions from BI to machine learning. You can use the Azure Databricks platform to build many different applications spanning data personas. Customers who fully embrace the lake house take advantage of our unified platform to build and deploy data engineering workflows, machine learning models, and analytics dashboards that power innovations and insights across an organization.

The Azure Databricks workspace provides user interfaces for many core data tasks, including tools for the following:

* Interactive notebooks
* Workflows scheduler and manager
* SQL editor and dashboards
* Data ingestion and governance
* Data discovery, annotation, and exploration
* Compute management
* Machine learning (ML) experiment tracking
* ML model serving
* A feature store
* Source control with Git

**Create the workspace resources you need to get started with Azure Data Bricks.**

The workspace is the top-level resource for your machine learning activities, providing a centralized place to view and manage the artifacts you create when you use Azure Machine Learning. The compute resources provide a pre-configured cloud-based environment you can use to train, deploy, automate, manage, and track machine learning models.

**Prerequisites**

An Azure account with an active subscription.

**Create the workspace**

If you already have a workspace, skip this section and continue to Create a compute instance.

If you don't yet have a workspace, create one now:

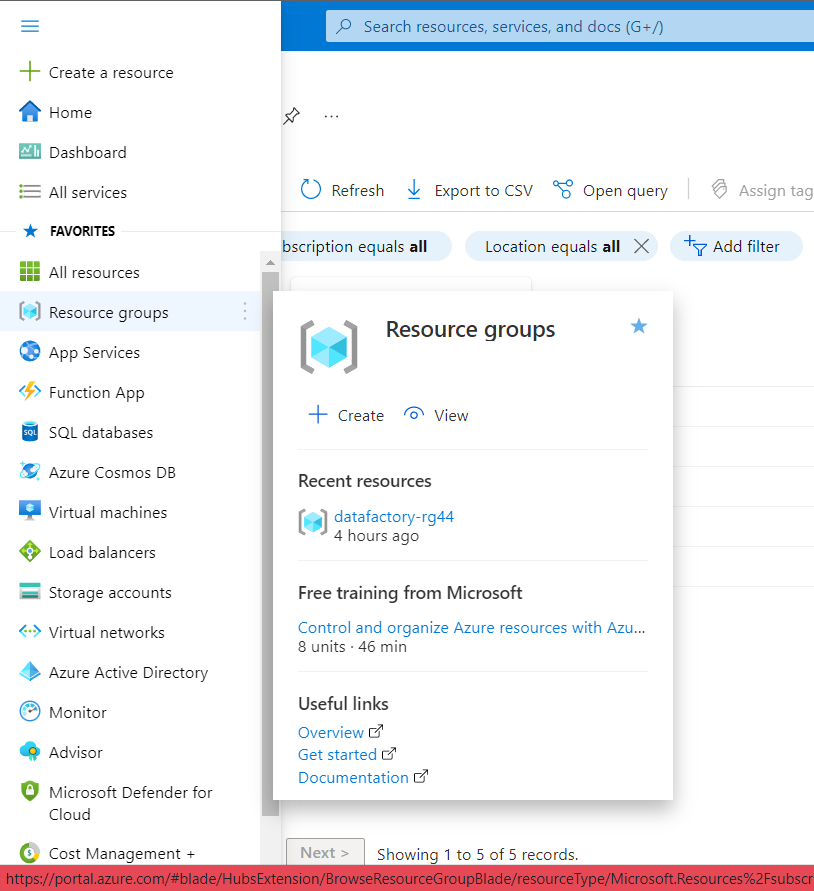
* Sign in to Azure Machine Learning studio
* Select Create workspace
* Provide the following information to configure your new workspace:

|  |  |
| --- | --- |
| Field | Description |
| Workspace name | Enter a unique name that identifies your workspace. Names must be unique across the resource group. Use a name that's easy to recall and to differentiate from workspaces created by others. The workspace name is case-insensitive. |
| Subscription | Select the Azure subscription that you want to use. |
| Resource group | Use an existing resource group in your subscription or enter a name to create a new resource group. A resource group holds related resources for an Azure solution. You need contributor or owner role to use an existing resource group. For more information about access, see Manage access to an Azure Machine Learning workspace. |
| Region | Select the Azure region closest to your users and the data resources to create your workspace. |

* Select Create to create the workspace.

**Create a Resource group**

* Select Resource groups
* Select Create

****



* Subscription: Select your Azure subscription.
* Resource group: Enter a new resource group name.
* Region: Select an Azure location, such as Central US, East US.

Graphical user interface, text, application, email

Description automatically generated



* Select Next
* Select Review + create
* Select Create. It takes a few seconds to create a resource group

Graphical user interface, text, application

Description automatically generated

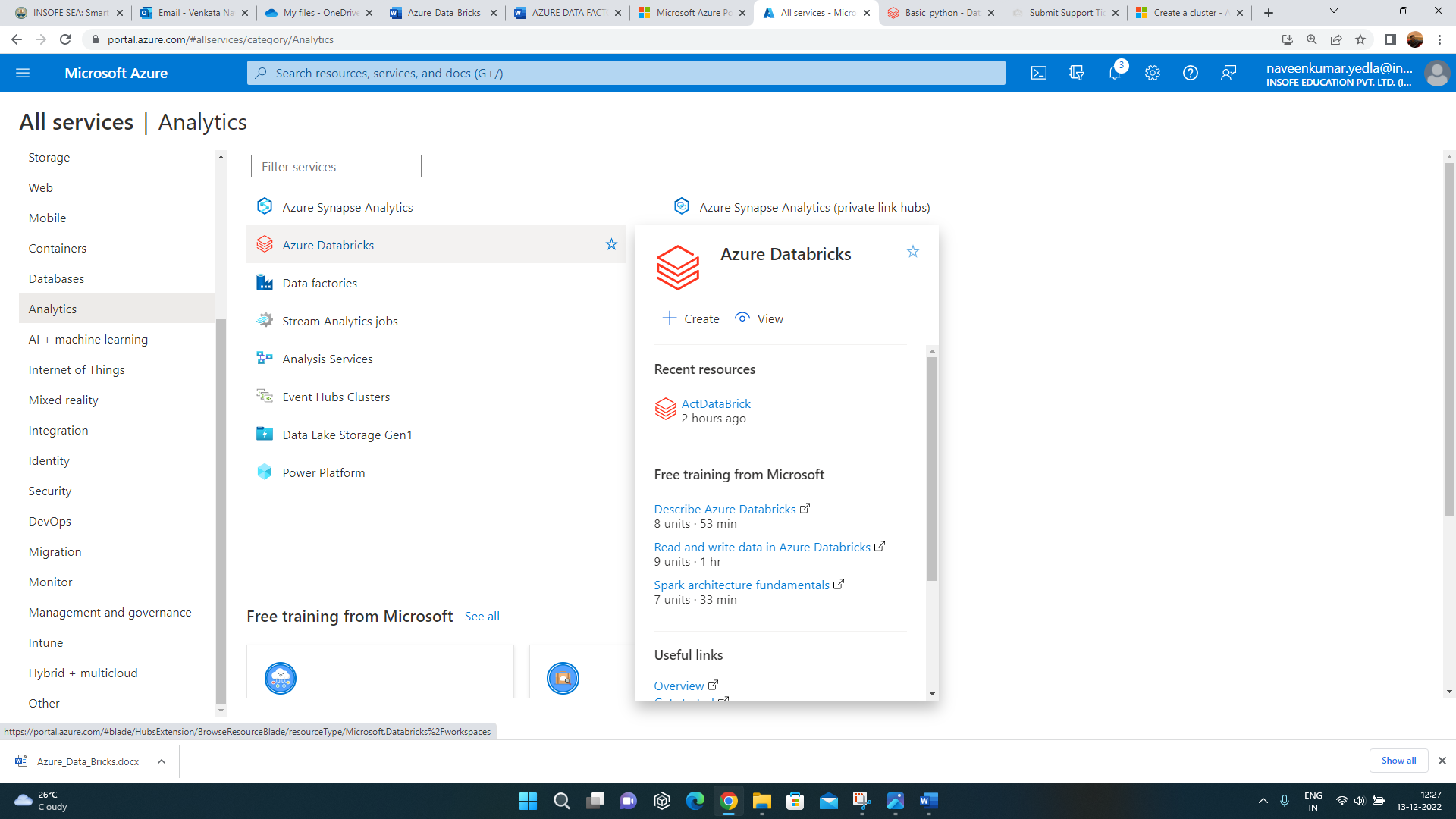
* Select Refresh from the top menu to refresh the resource group list, and then select the newly created resource group to open it. Or select Notification(the bell icon) from the top, and then select Go to resource group to open the newly created resource group

Graphical user interface, text, application, email

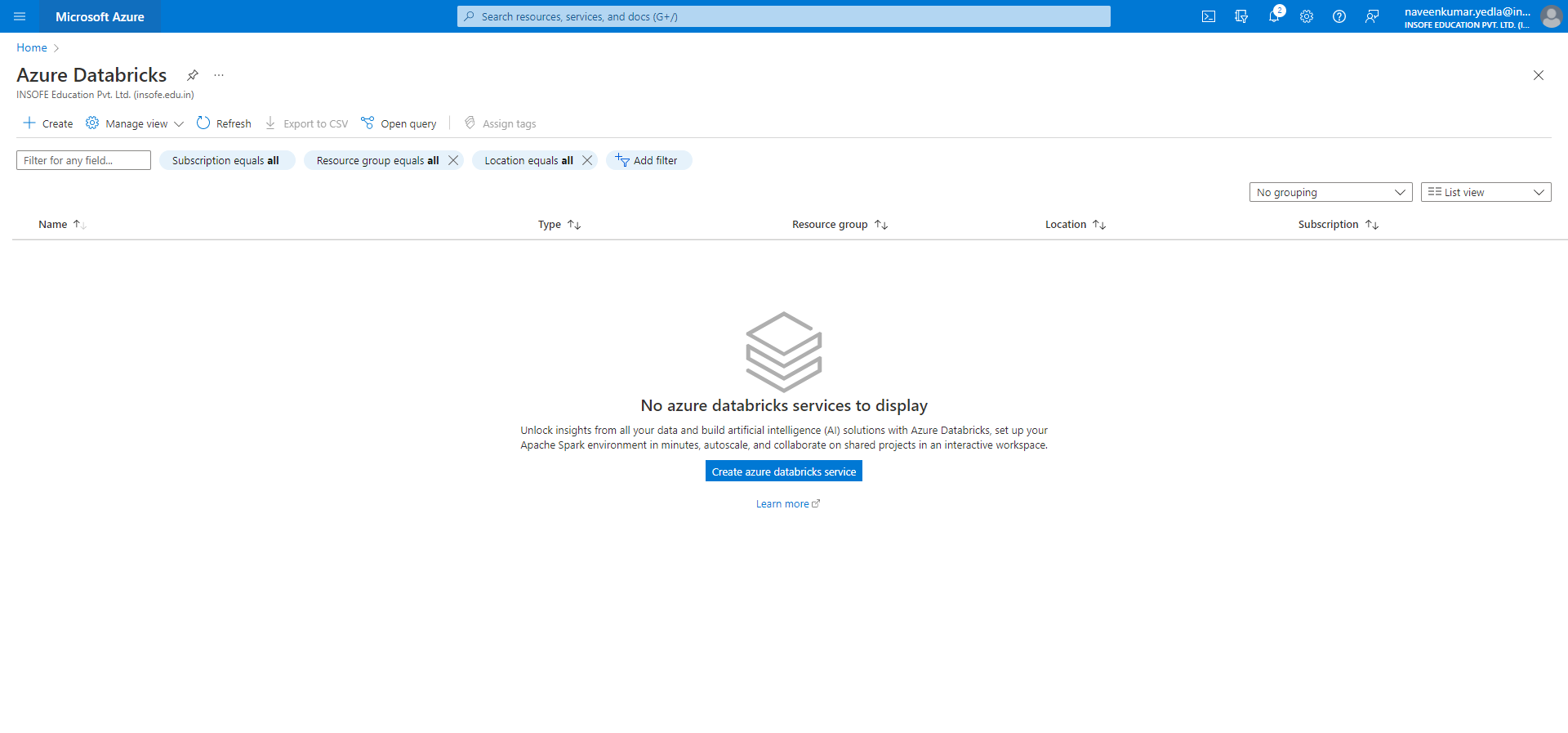
Description automatically generated

**Create Azure Databricks.**

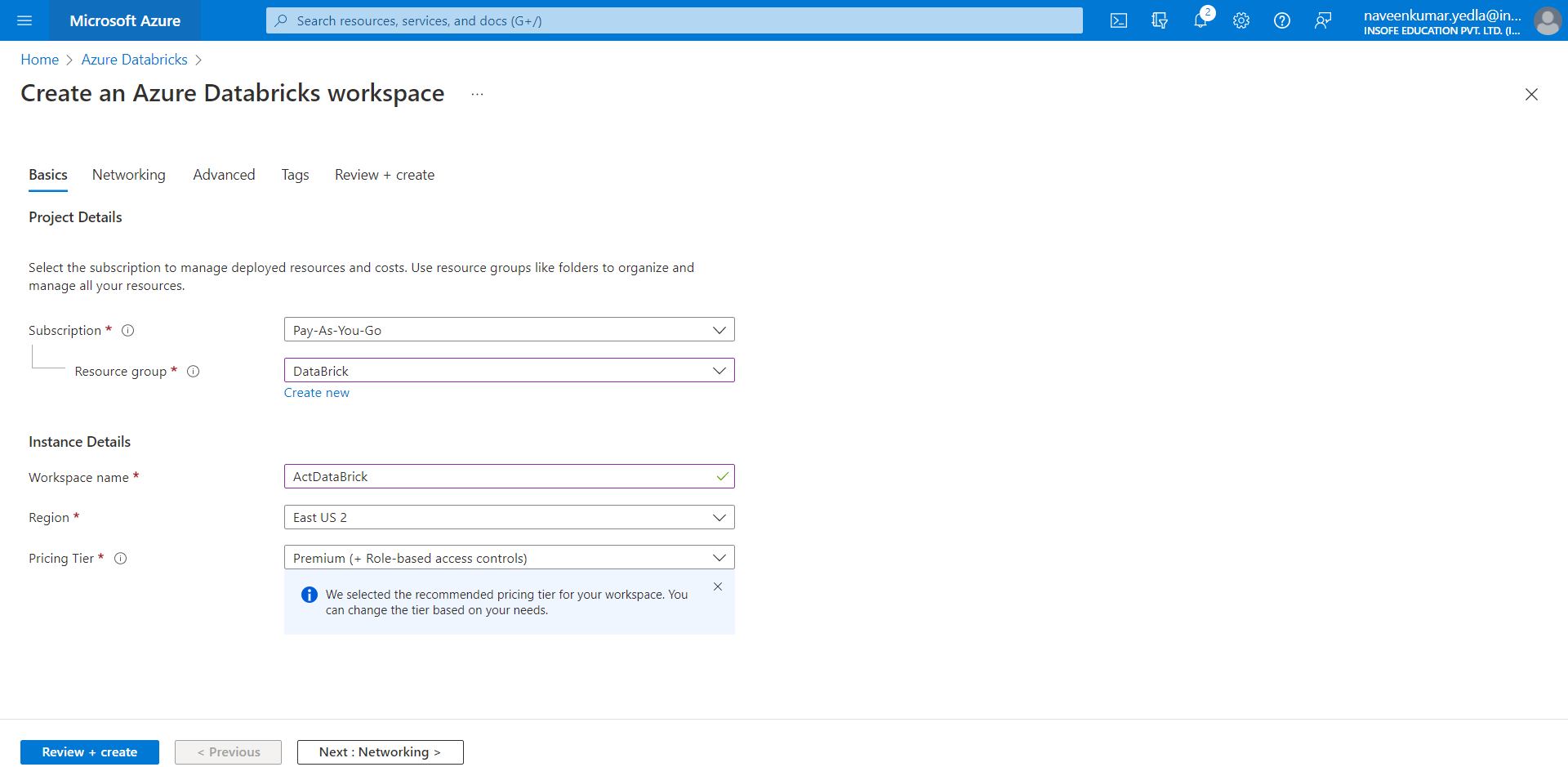
* Select All services
* Select Analytics
* Select Azure Databricks



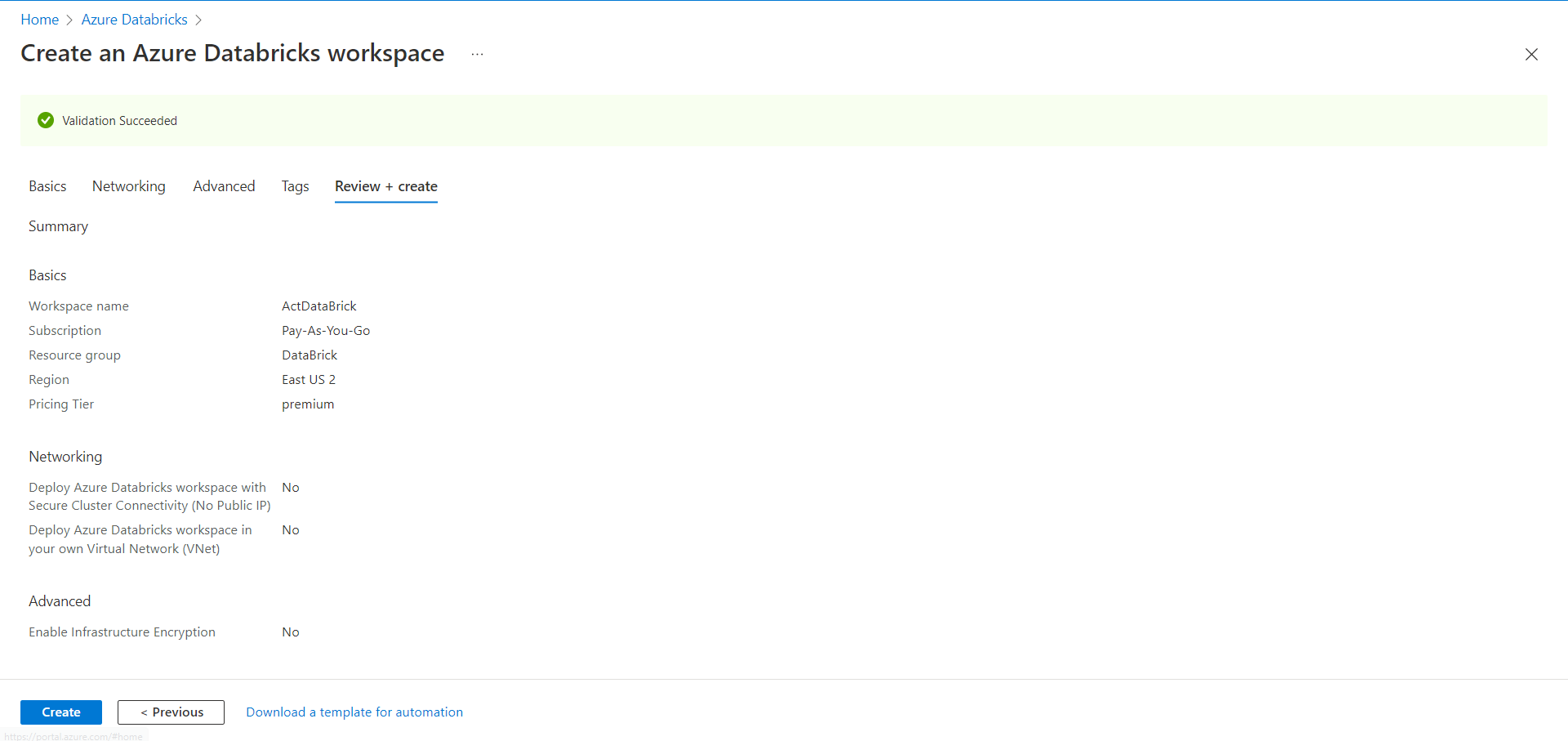
* Select Create

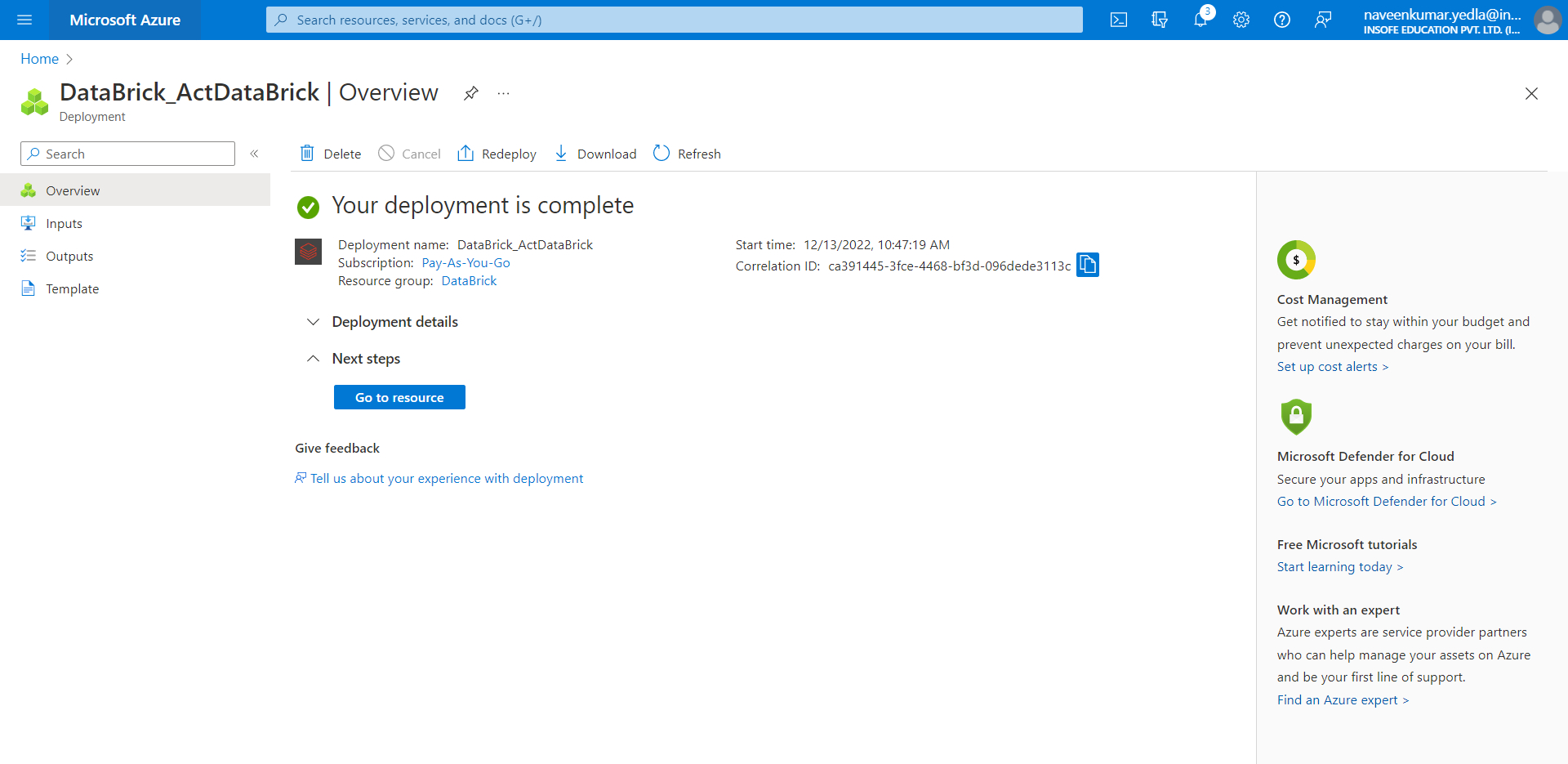


* Subscription: Select your Azure subscription.
* Resource group: Enter a new resource group name.
* Workspace name: Enter a new workspace name.
* Select on Review+create

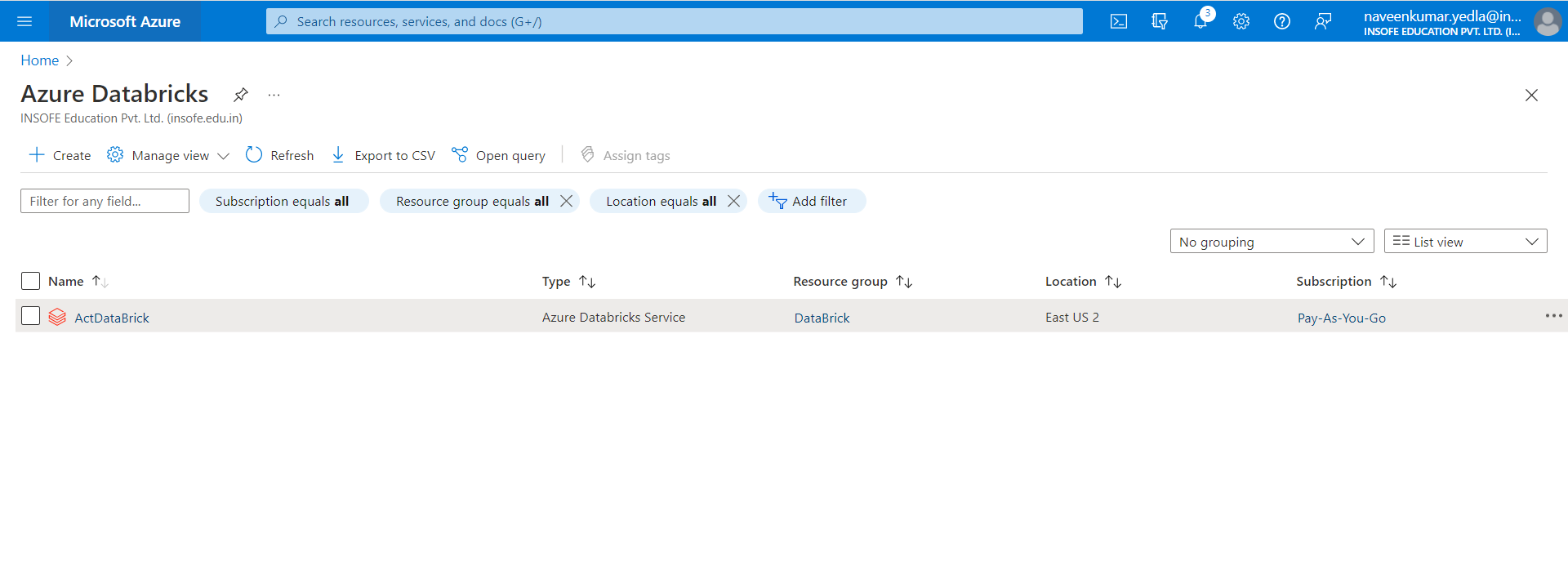


* Select create. It takes a few seconds to create Databricks workspace.

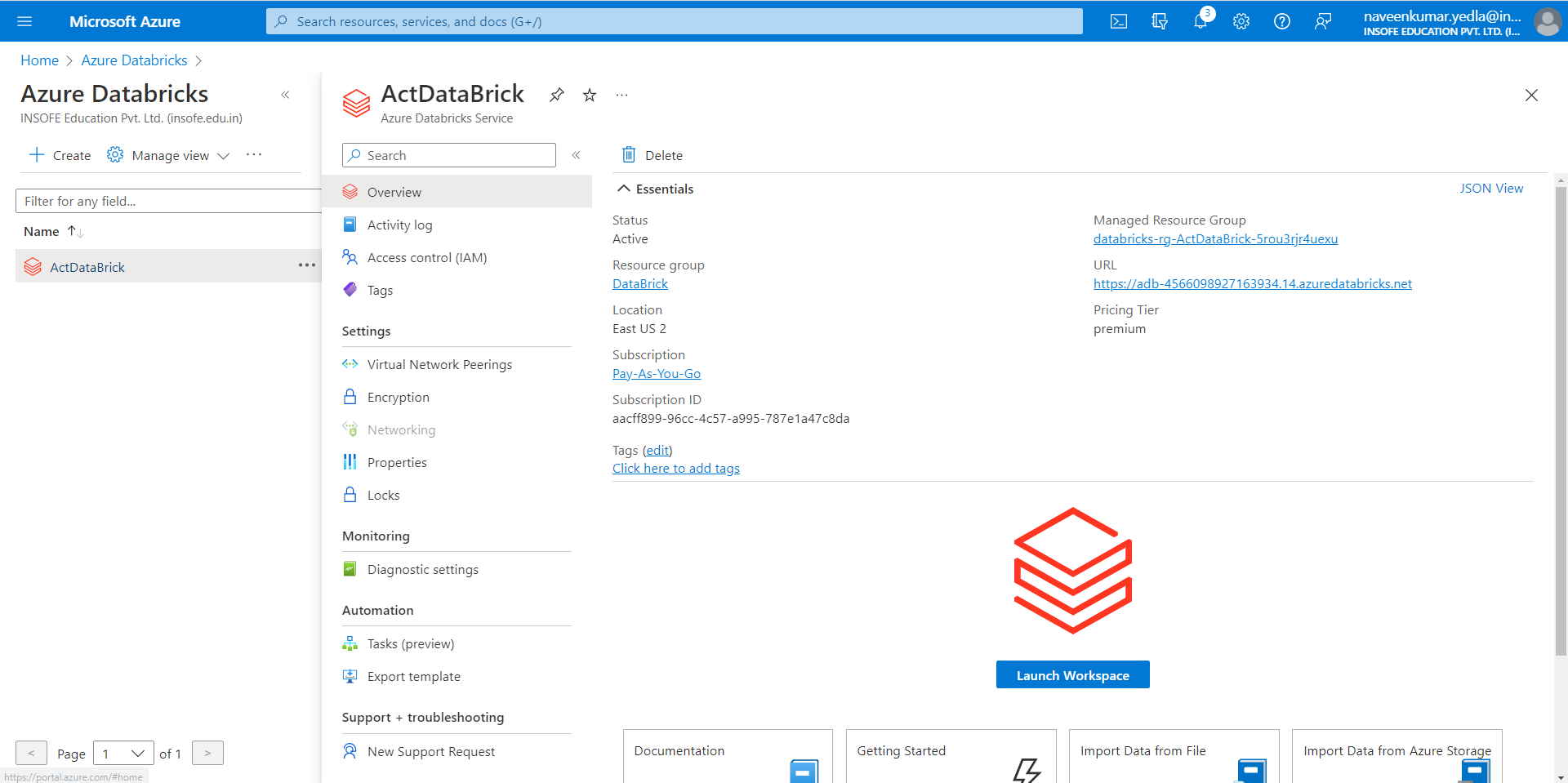




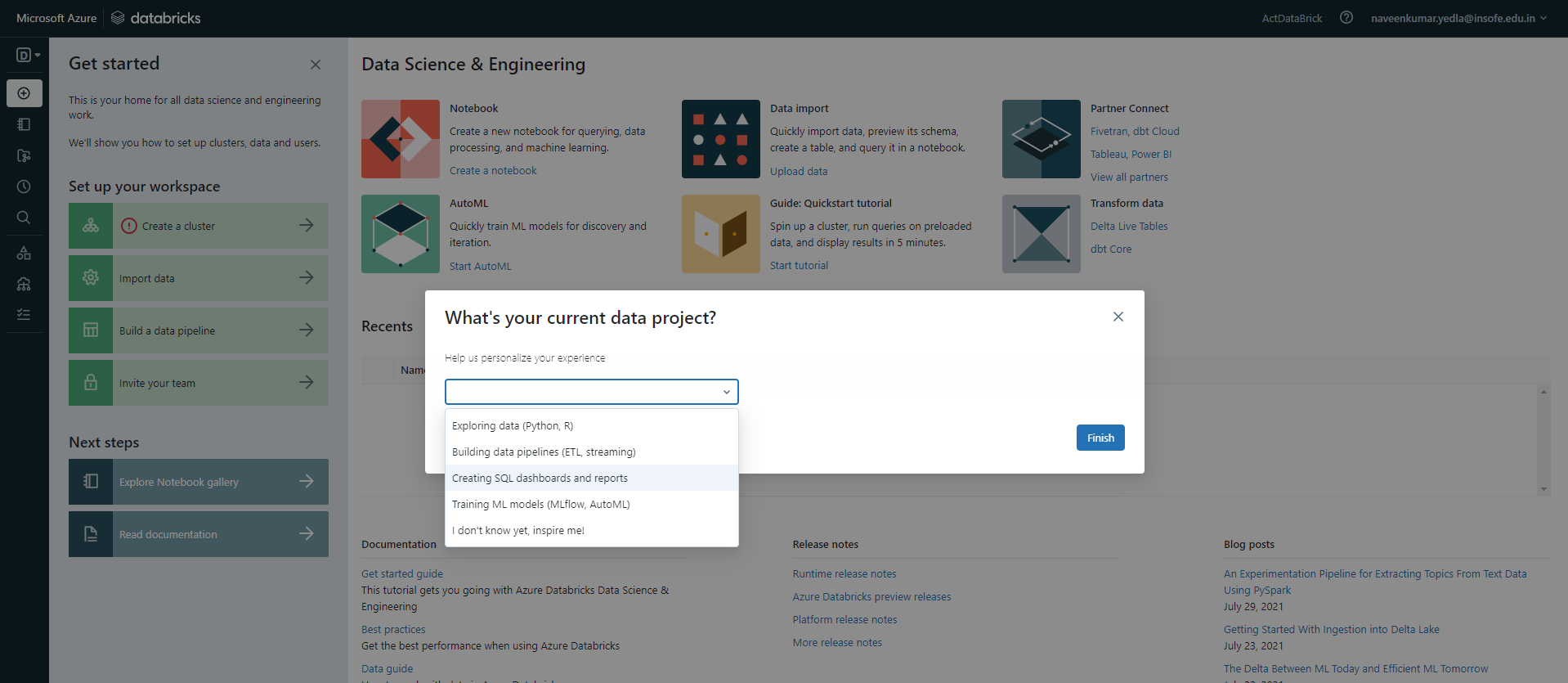
* Select the database ActDataBrick



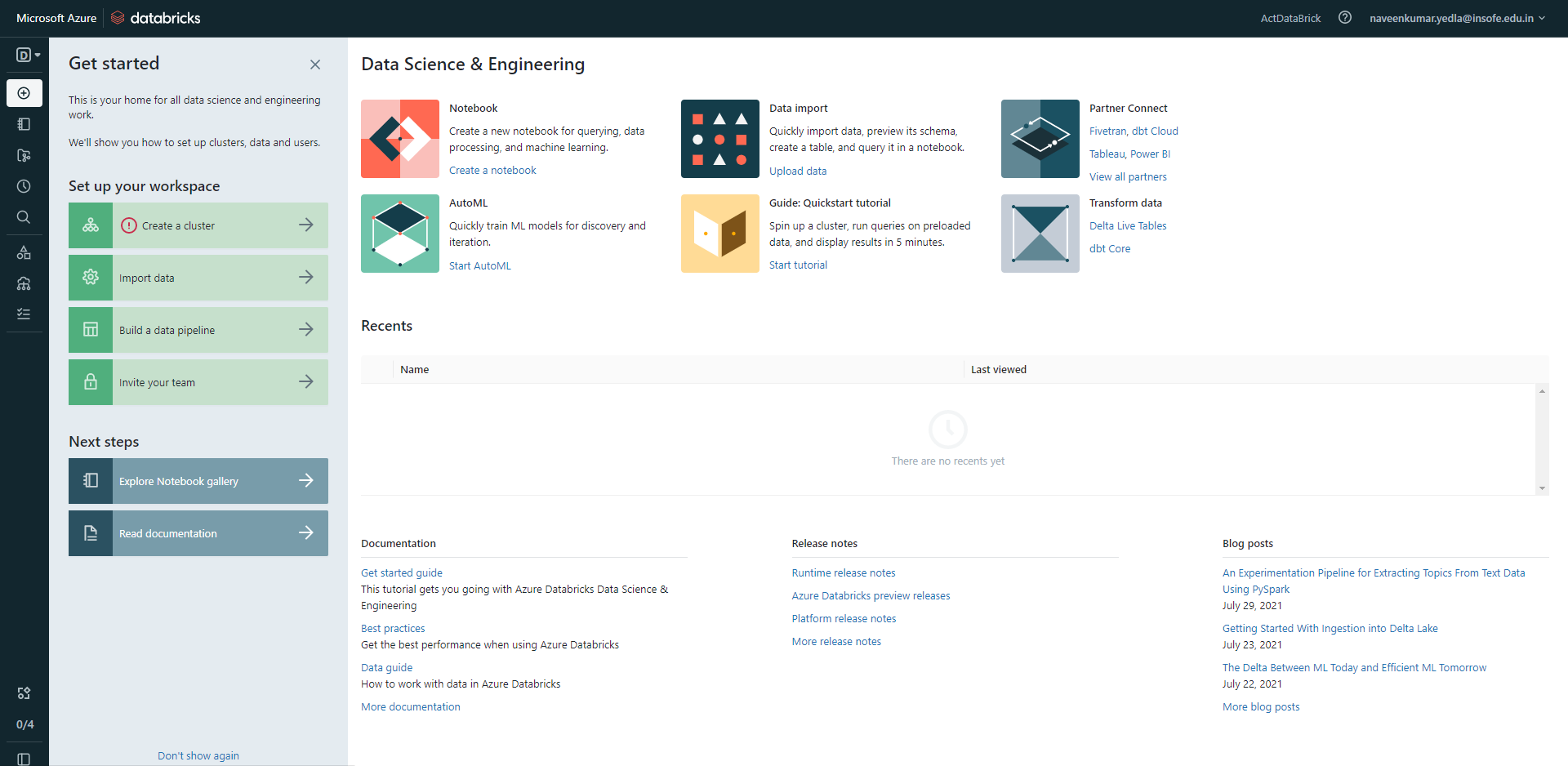
* Select the Launch Workspace a new tab of Azure data bricks workspace will open.



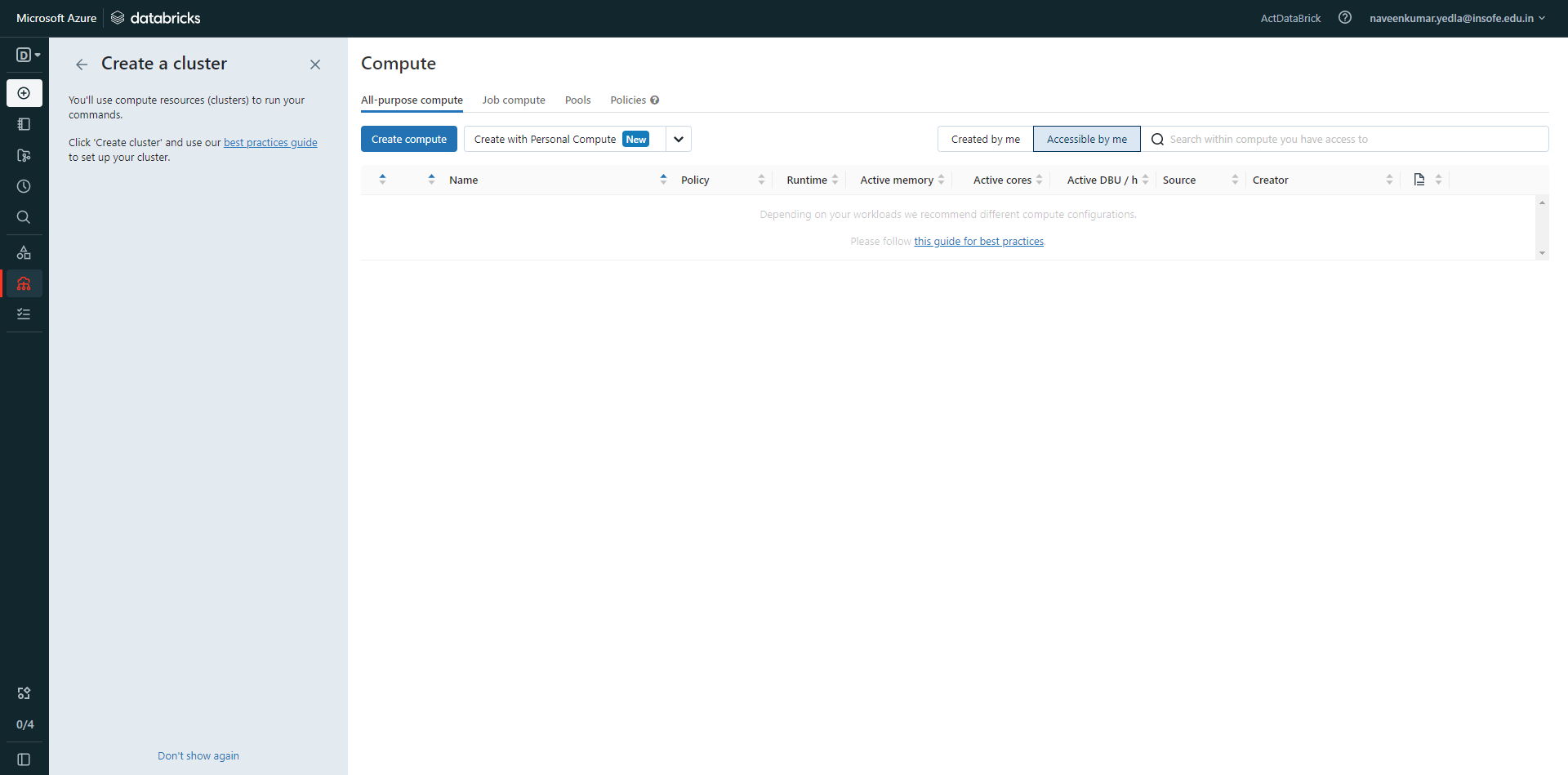
* Select the current data project which is required
* In this case just select the Exploring data (Python, R)
* Select finish



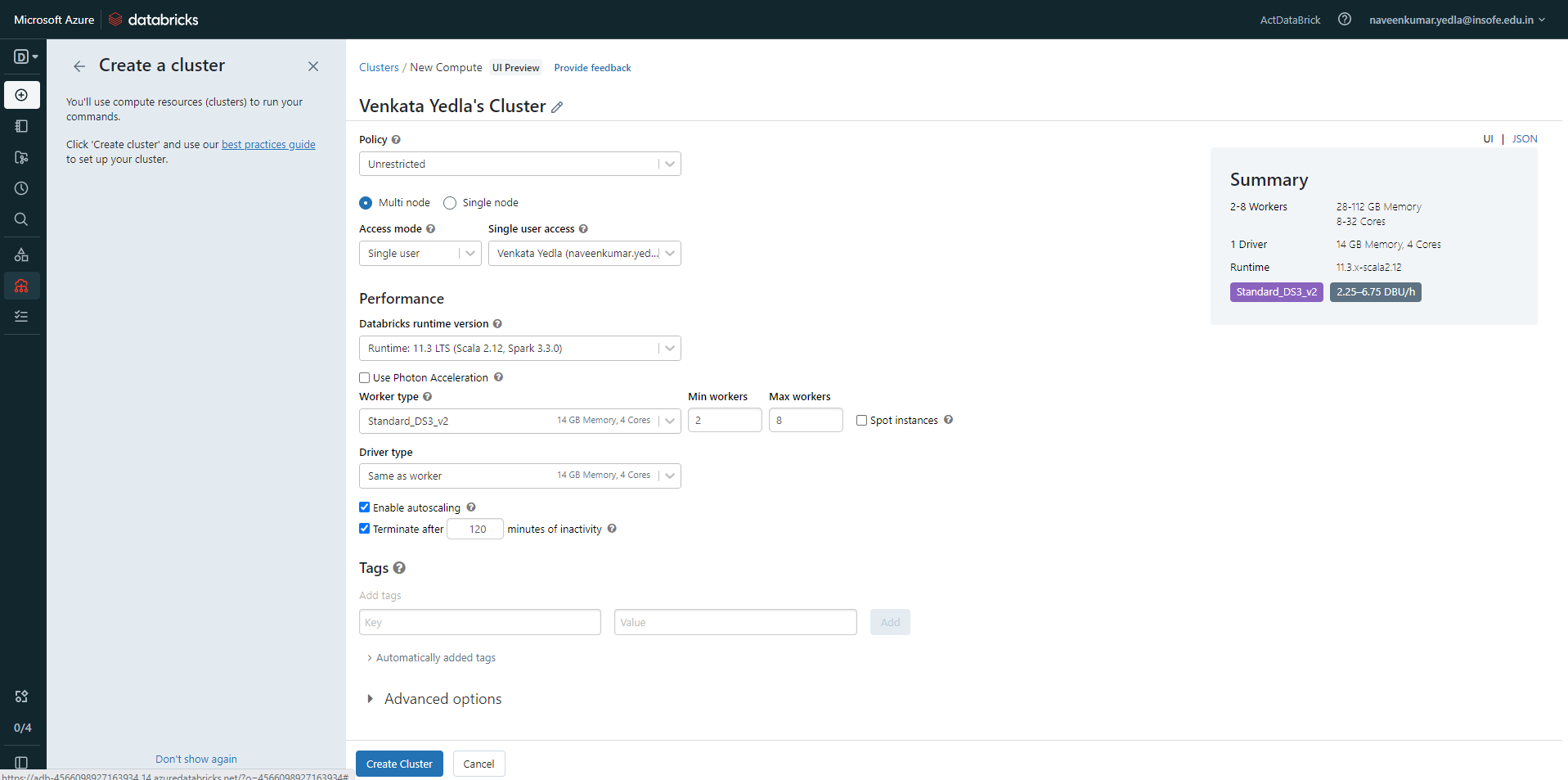
* A new workspace of Azure Databricks will open
* Select Create a Cluster



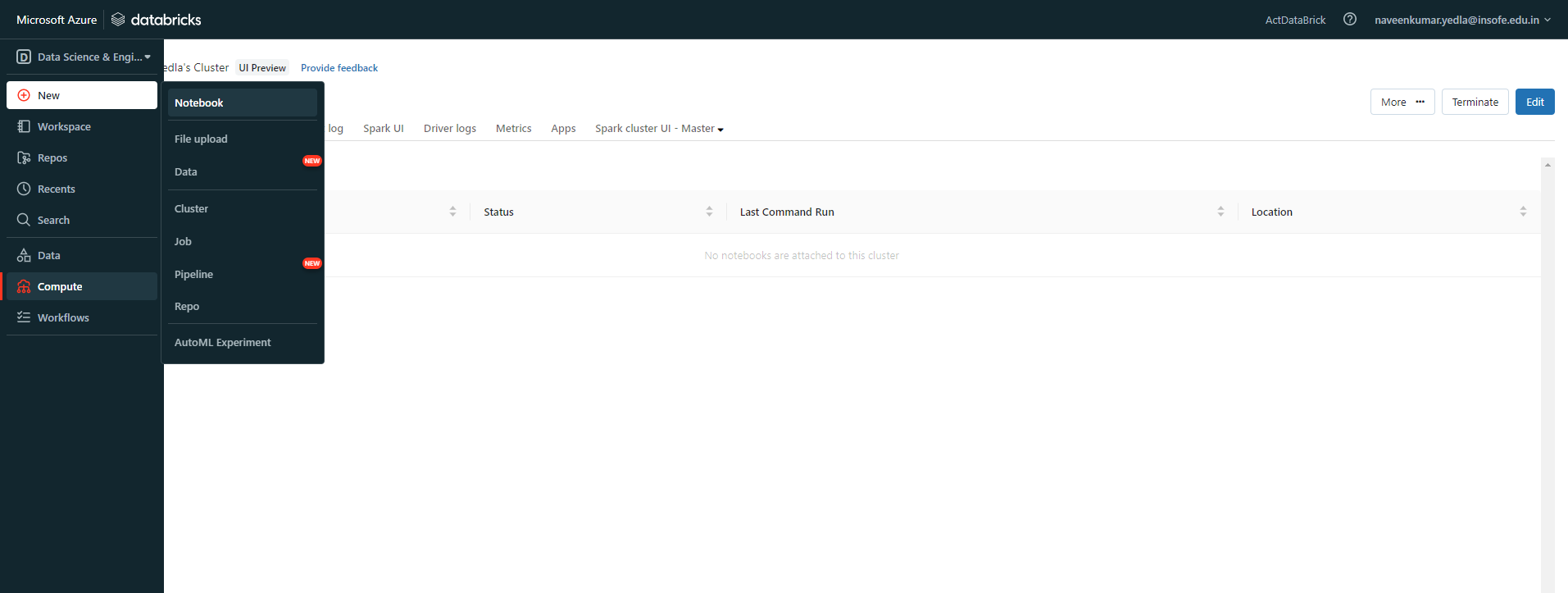
* Select Create Compute



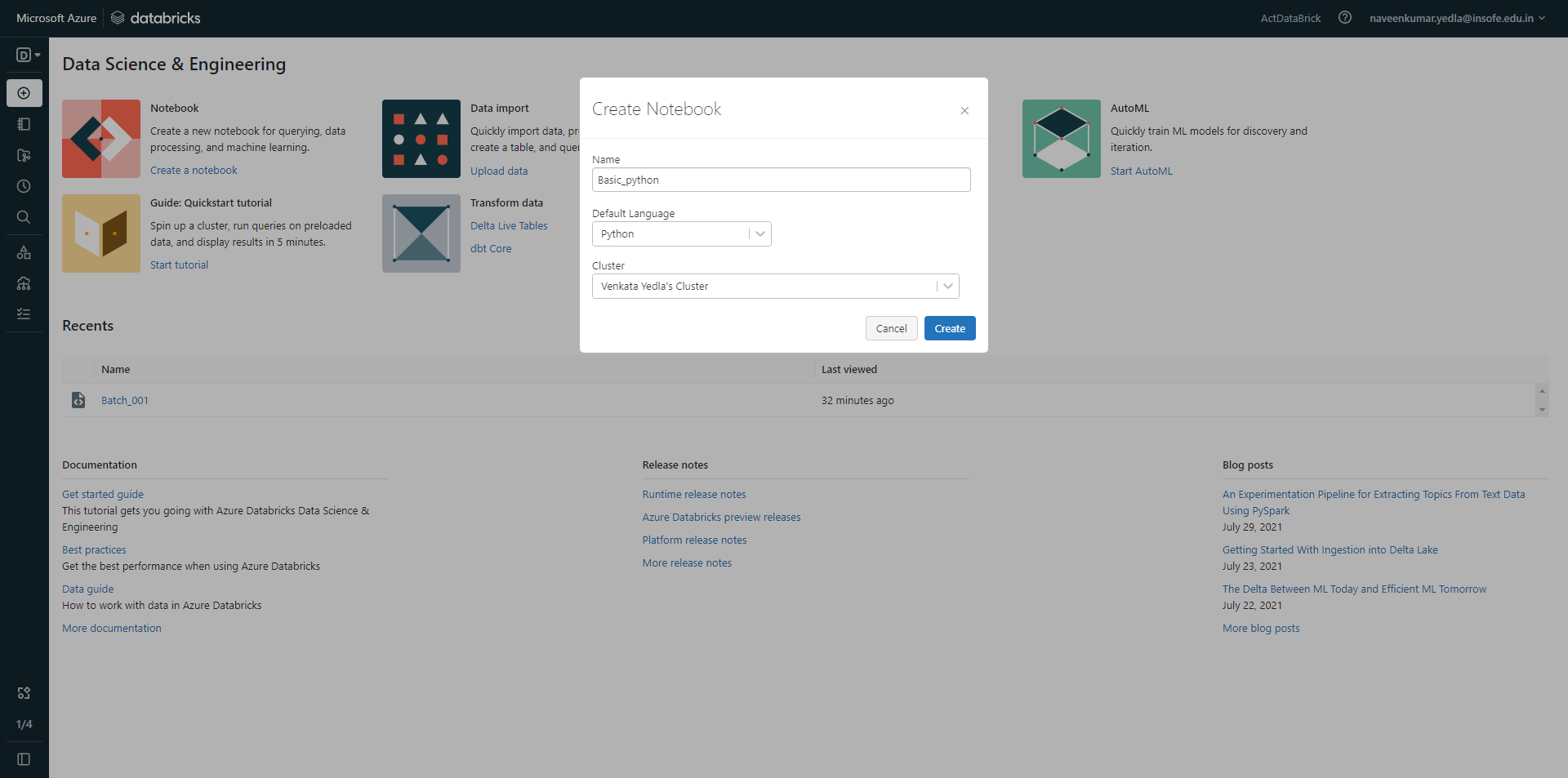
* Select Create Cluster



* Select New
* Select Notebook



* Enter the new name for the notebook
* Default Language: Python
* Cluster shows default your account name
* Select Create



* A new python Notebook will open.
* We can execute the python code.

