**AZURE DATA FACTORY**

Azure Data Factory is Azure's cloud ETL service for scale-out serverless data integration and data transformation. It offers a code-free UI for intuitive authoring and single-pane-of-glass monitoring and management. You can also lift and shift existing SSIS packages to Azure and run them with full compatibility in ADF. SSIS Integration Runtime offers a fully managed service, so you don't have to worry about infrastructure management

**Create a Azure data factory using azure portal**

1. **Azure subscription**

If you don't have an Azure subscription, create a free account before you begin.

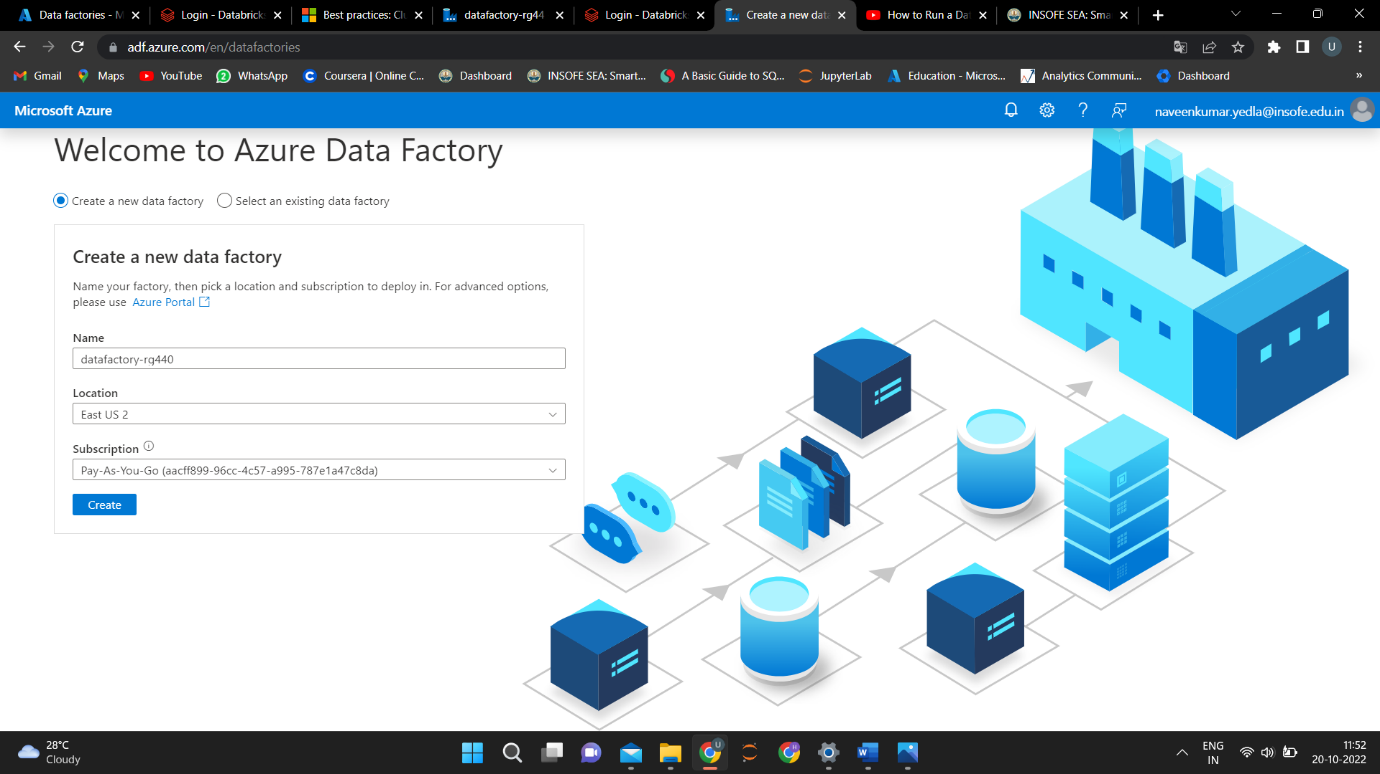
1. **Azure roles**

To learn about the Azure role requirements to create a data factory, refer to Azure Roles requirements

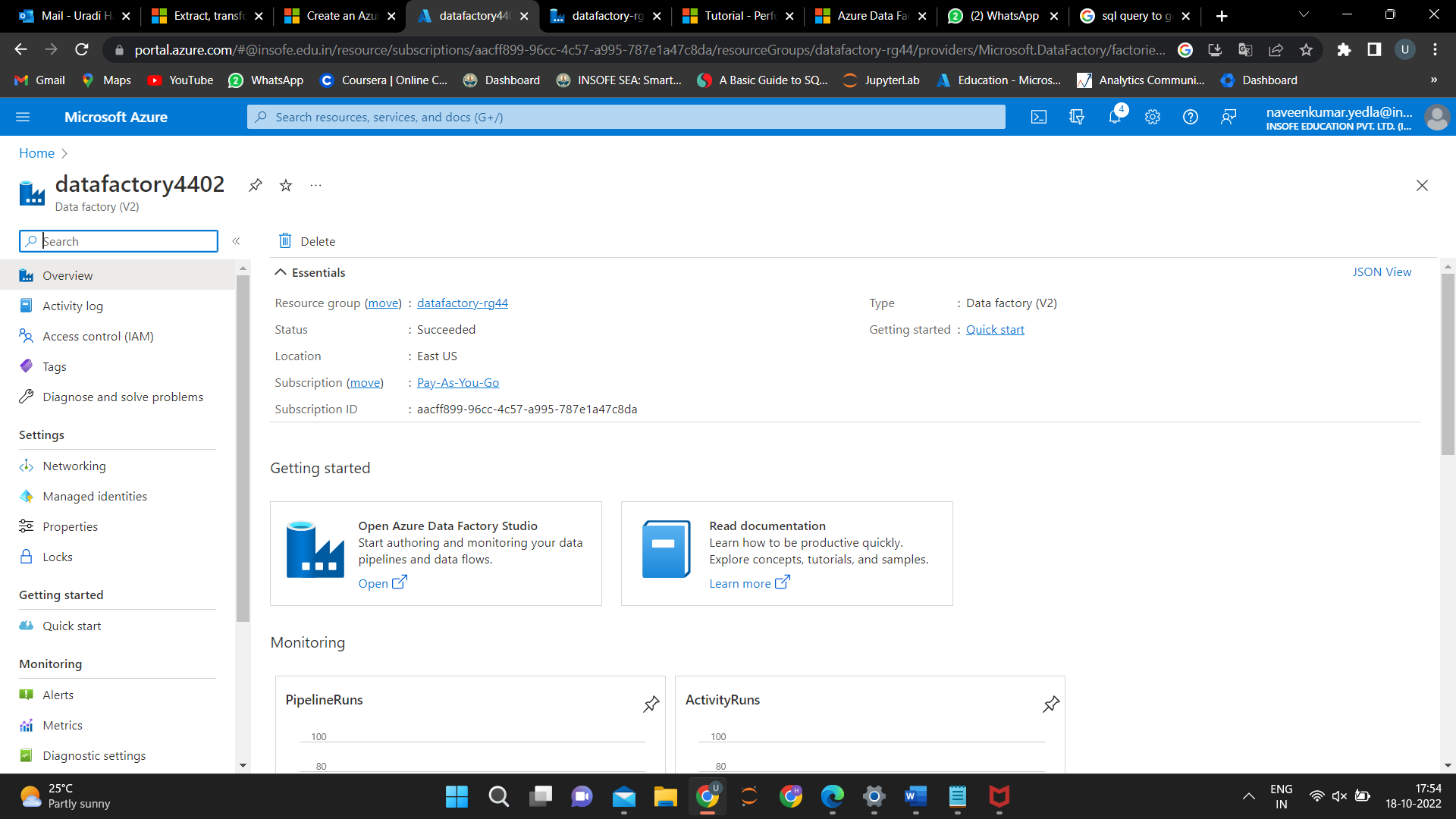
1. **Create a data factory**

A simple creation experience provided in the Azure Data Factory Studio to enable users to create a data factory within seconds. More advanced creation options are available in Azure portal.

1. **Simple creation in the Azure Data Factory Studio**
2. Launch Microsoft Edge or Google Chrome web browser. Currently, Data Factory UI is supported only in Microsoft Edge and Google Chrome web browsers.
3. Go to the Azure Data Factory Studio and choose the Create a new data factory radio button.
4. You can use the default values to create directly, or enter a unique name and choose a preferred location and subscription to use when creating the new data factory

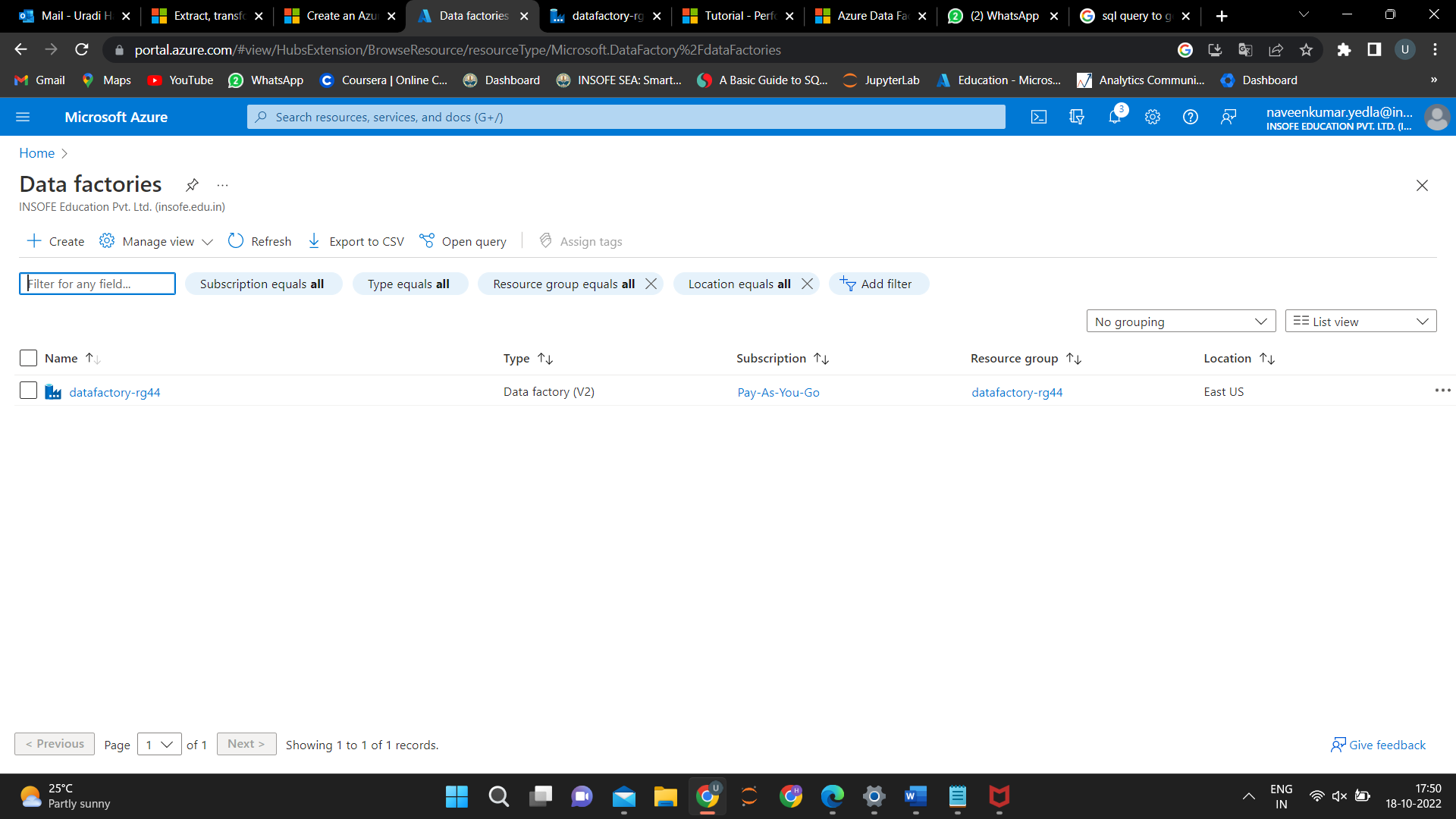


1. After creation, you can directly enter the homepage of the Azure Data Factory Studio.



**Advanced creation in the Azure portal**

1. Launch Microsoft Edge or Google Chrome web browser. Currently, Data Factory UI is supported only in Microsoft Edge and Google Chrome web browsers.
2. Go to the Azure portal data factories page.
3. After landing on the data factories page of the Azure portal, click Create.



1. After landing on the data factories page of the Azure portal, click Create.

For Resource Group, take one of the following steps:

a. Select an existing resource group from the drop-down list.

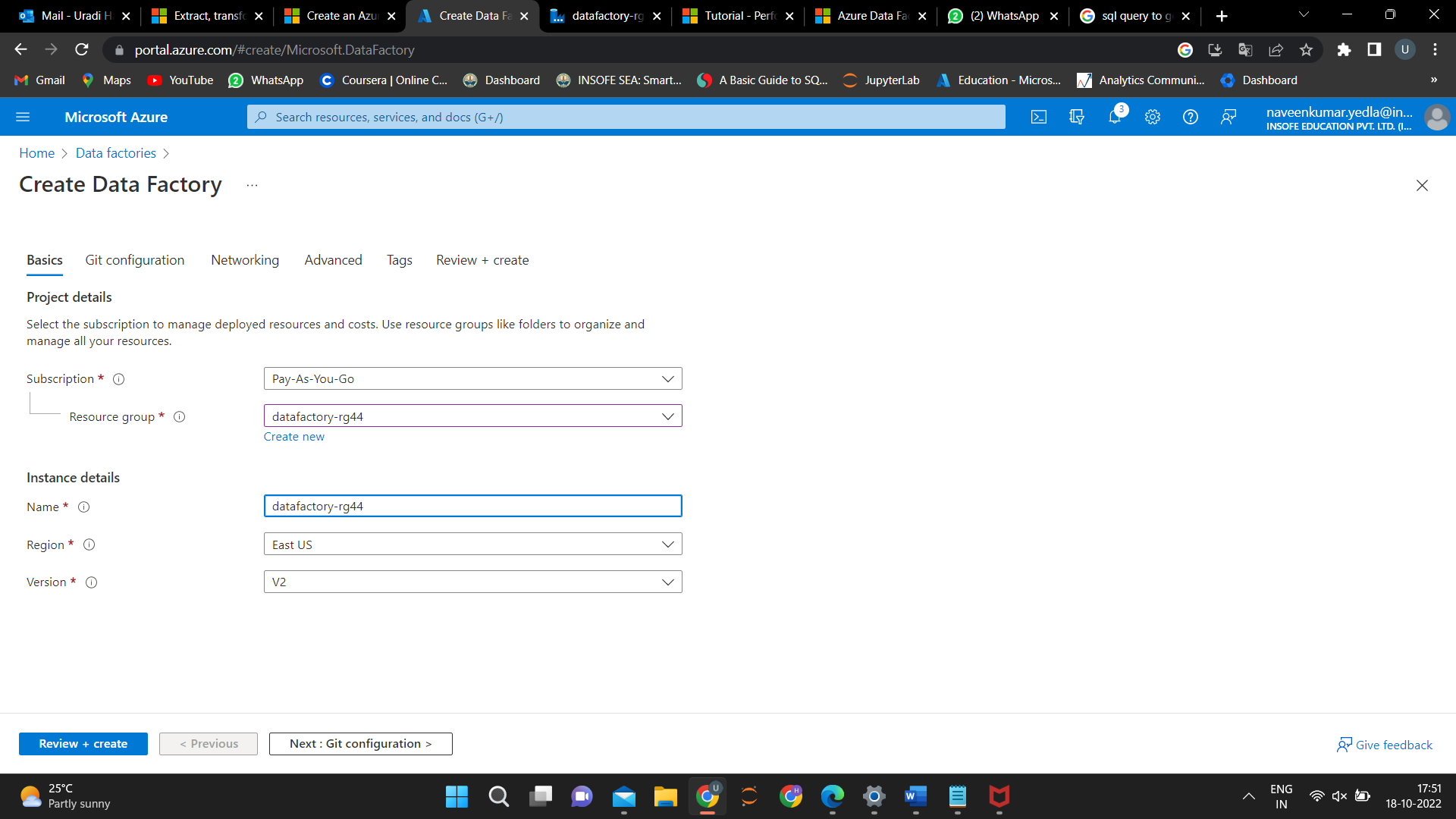
b. Select Create new, and enter the name of a new resource group.

1. For Region, select the location for the data factory.

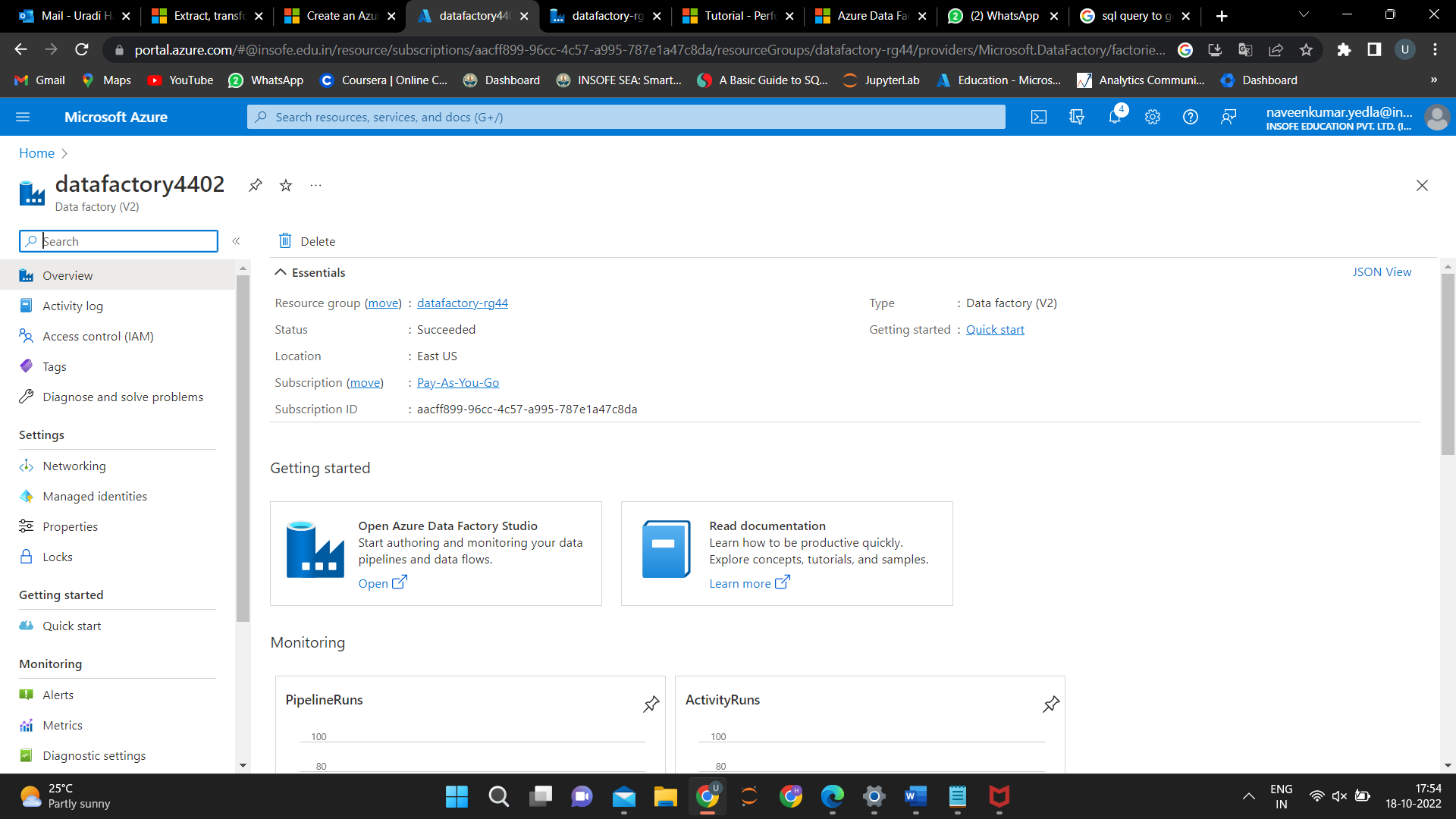
The list shows only locations that Data Factory supports, and where your Azure Data Factory meta data will be stored. The associated data stores (like Azure Storage and Azure SQL Database) and computes (like Azure HDInsight) that Data Factory uses can run in other regions.

1. For Name, enter ADFTutorialDataFactory.

The name of the Azure data factory must be globally unique. If you see the following error, change the name of the data factory (for example, <yourname>ADFTutorialDataFactory) and try creating again. For naming rules for Data Factory artifacts, see the Data Factory - naming rules article.



1. For Version, select V2.
2. Select Next: Git configuration, and then select Configure Git later check box.
3. Select Review + create, and select Create after the validation is passed. After the creation is complete, select Go to resource to navigate to the Data Factory page.
4. Select Open on the Open Azure Data Factory Studio tile to start the Azure Data Factory user interface (UI) application on a separate browser tab.



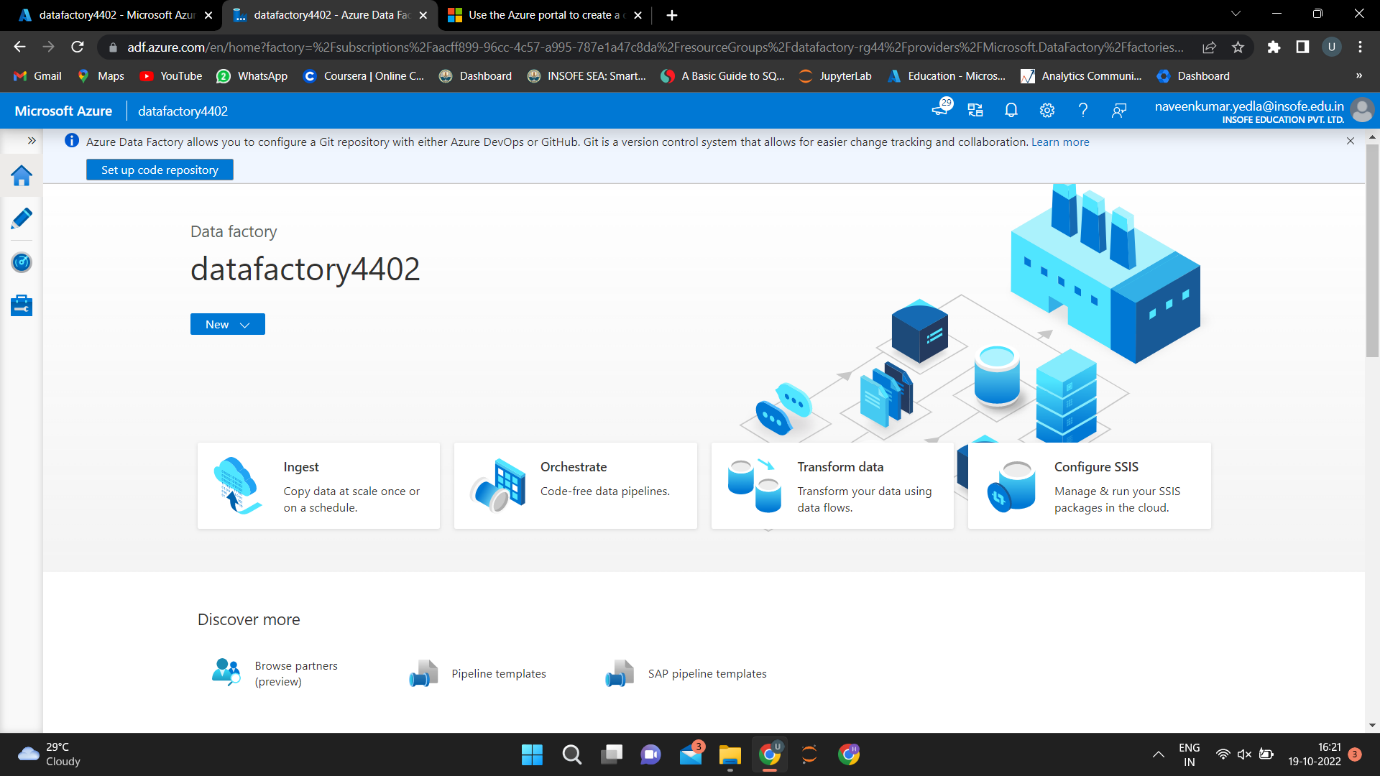
**DATABRICKS**

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

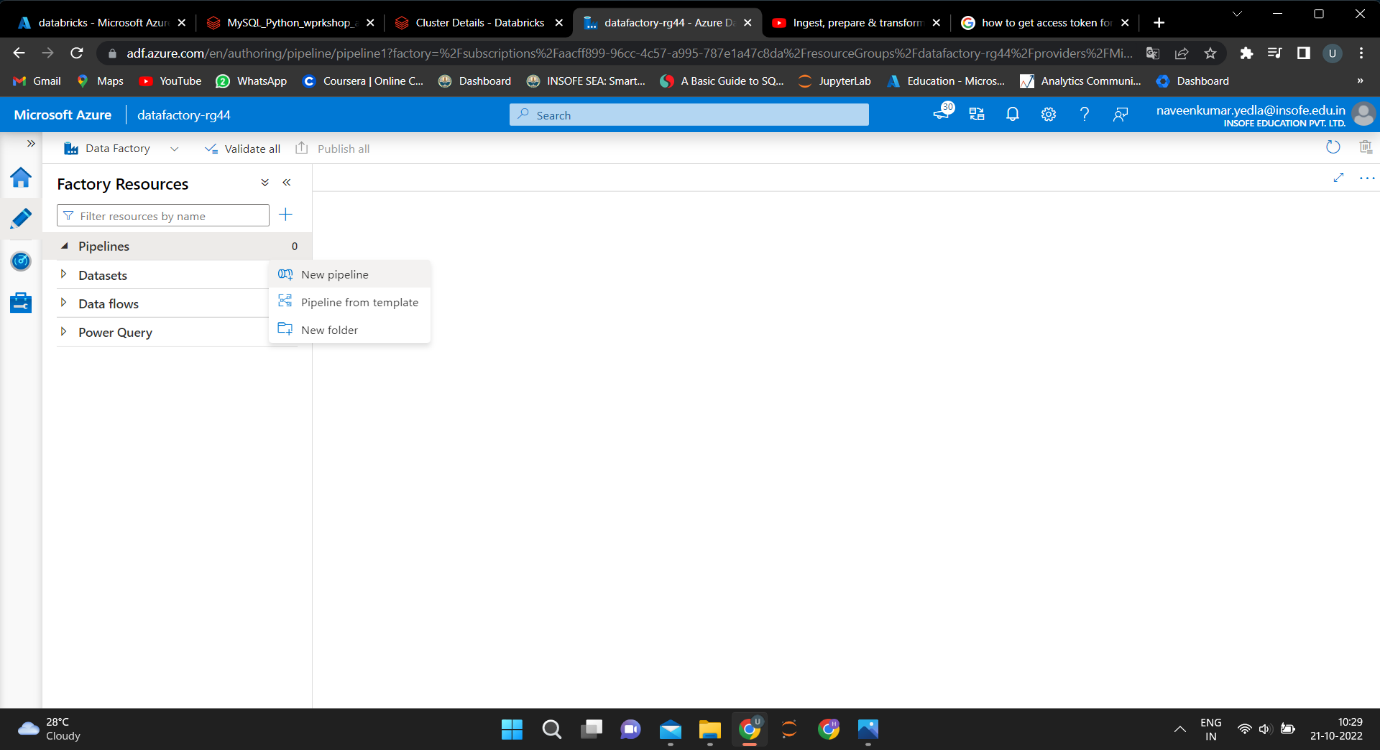
**Databricks Notebook Activity in Azure Data Factory**

In this activity, you use the Azure portal to create an Azure Data Factory pipeline that executes a Databricks notebook against the Databricks jobs cluster.

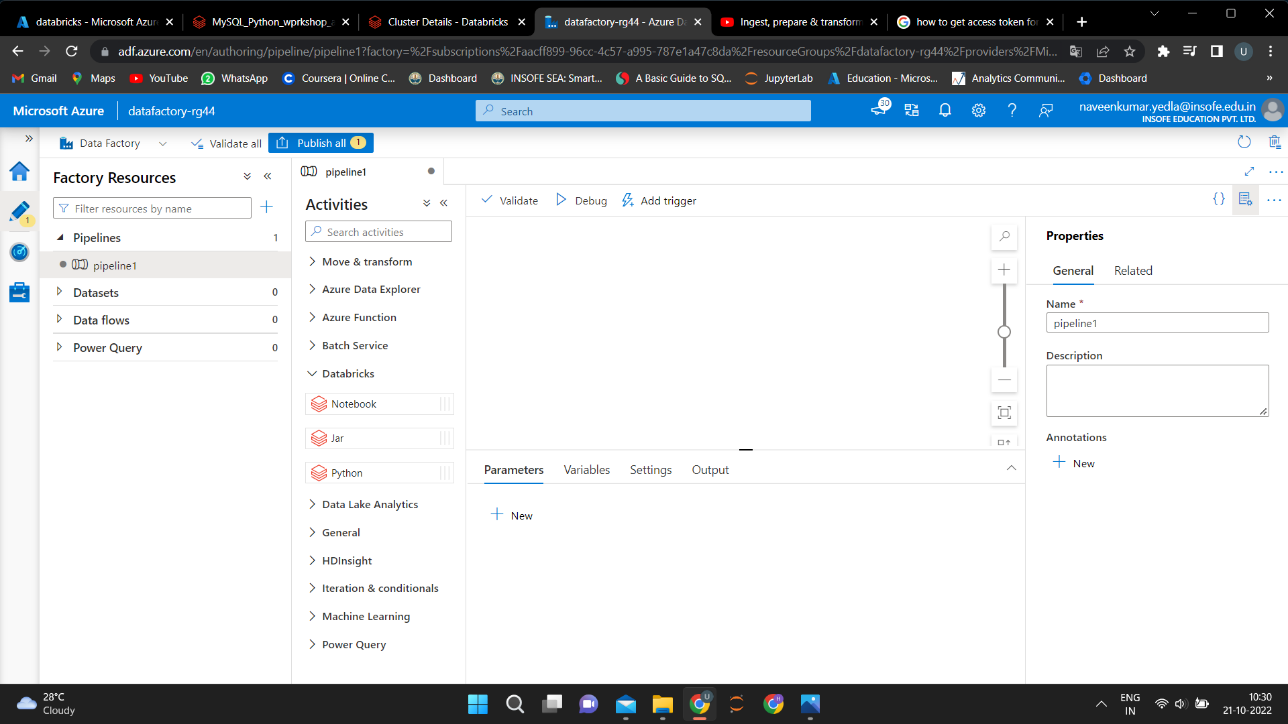
1. Follow above steps to create the azure data factory and after creation of it. Open the ADF that u have created
2. Select orchestrate to create the pipeline, click on it



3. Click on pipeline and select new pipeline

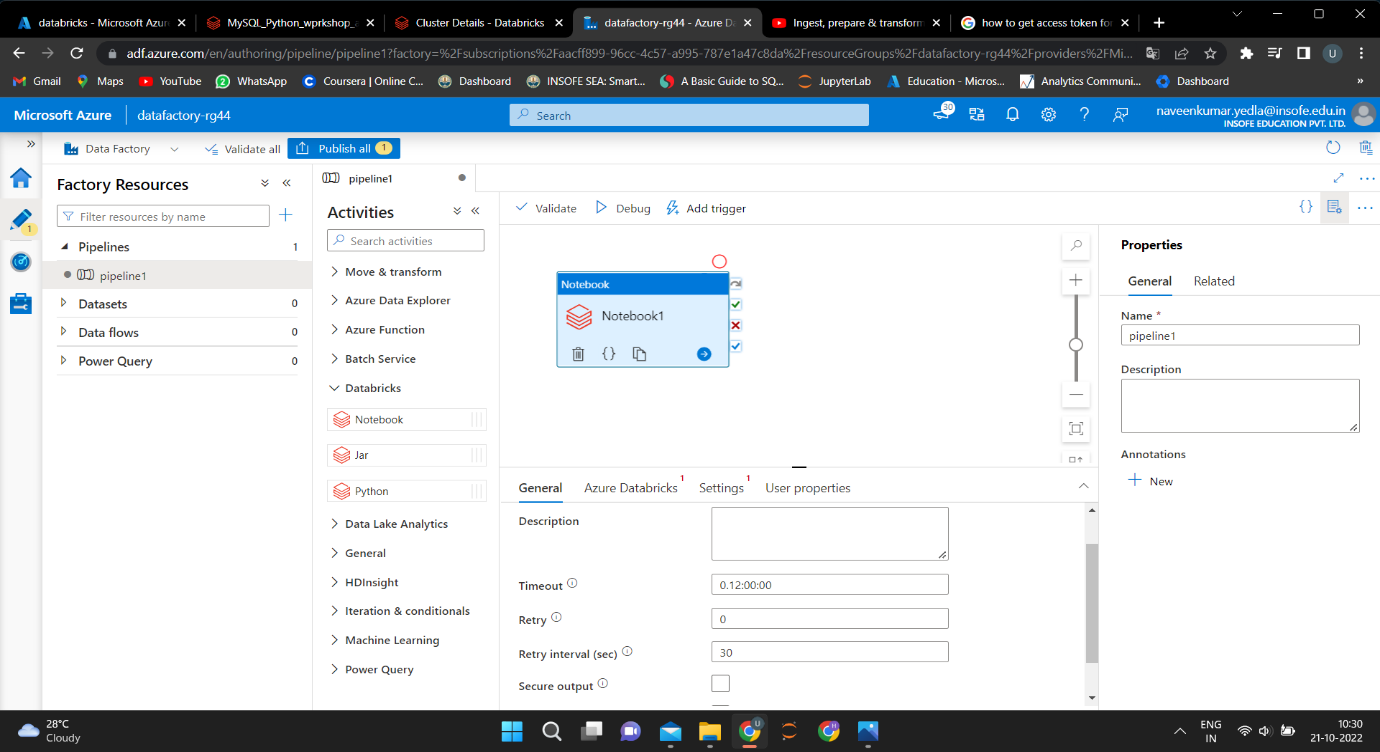


1. After selecting pipeline. Click on Databricks and select notebook.

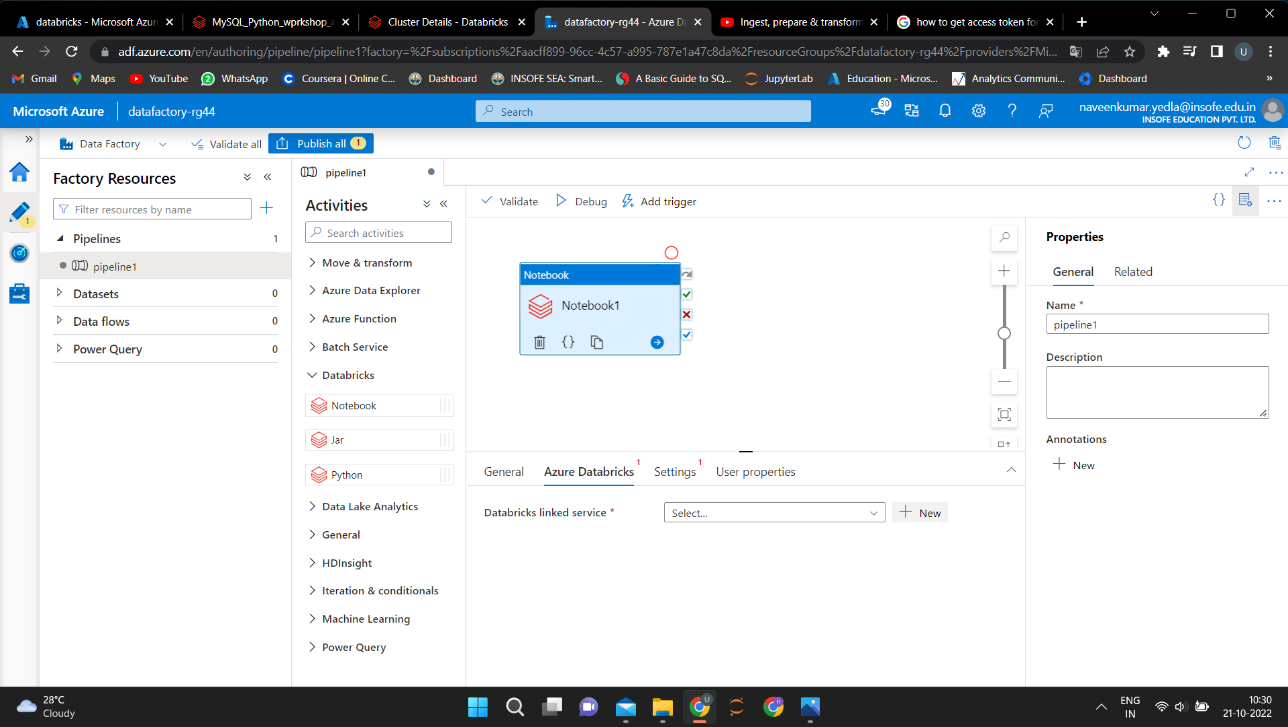


drag and drop the Notebook refer below fig

1. In General blade, add description, give notebook name

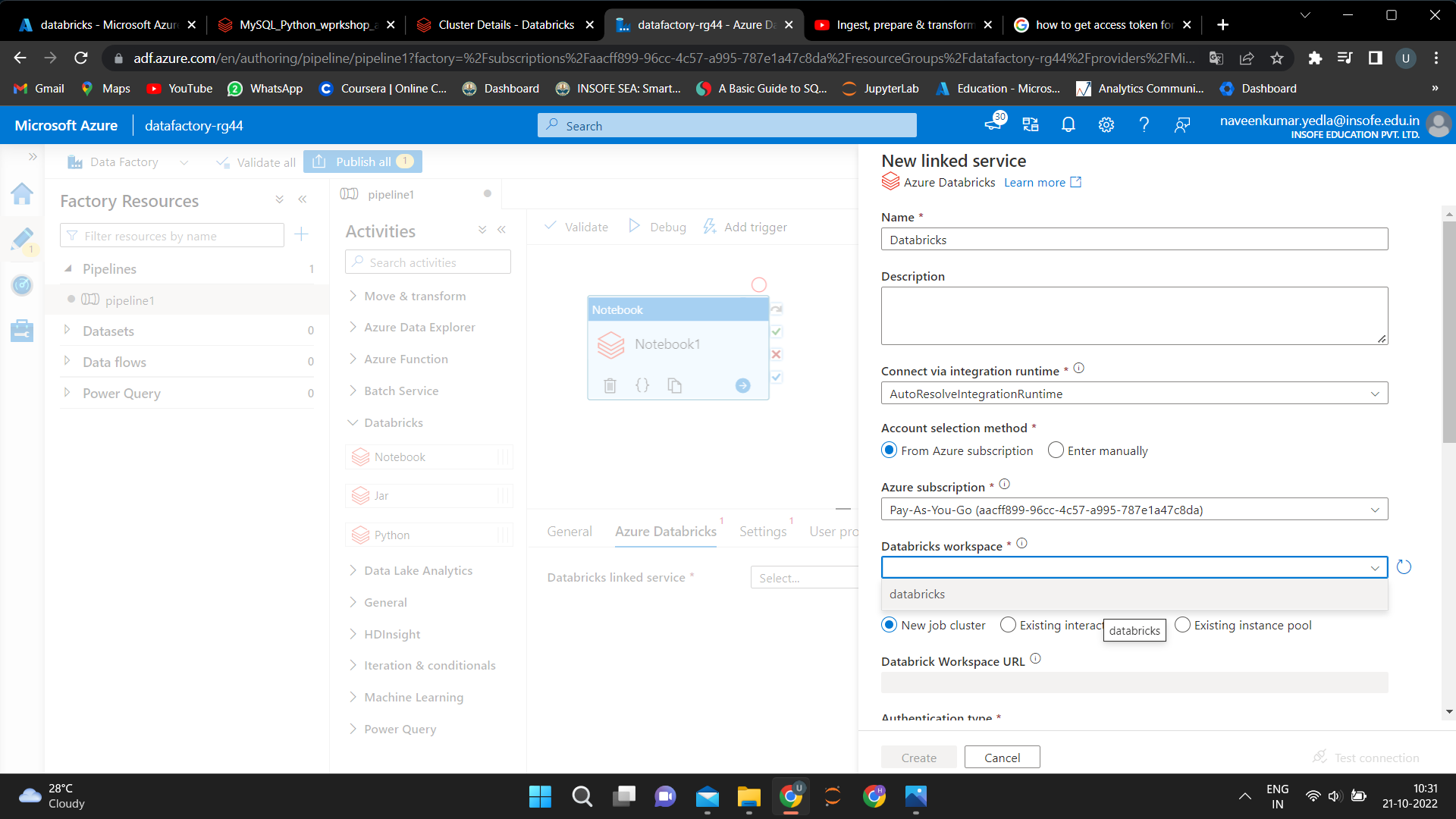


1. In Azure Databricks blade, select linked service if there is an existing one
2. If there is no linked service click on ‘new’

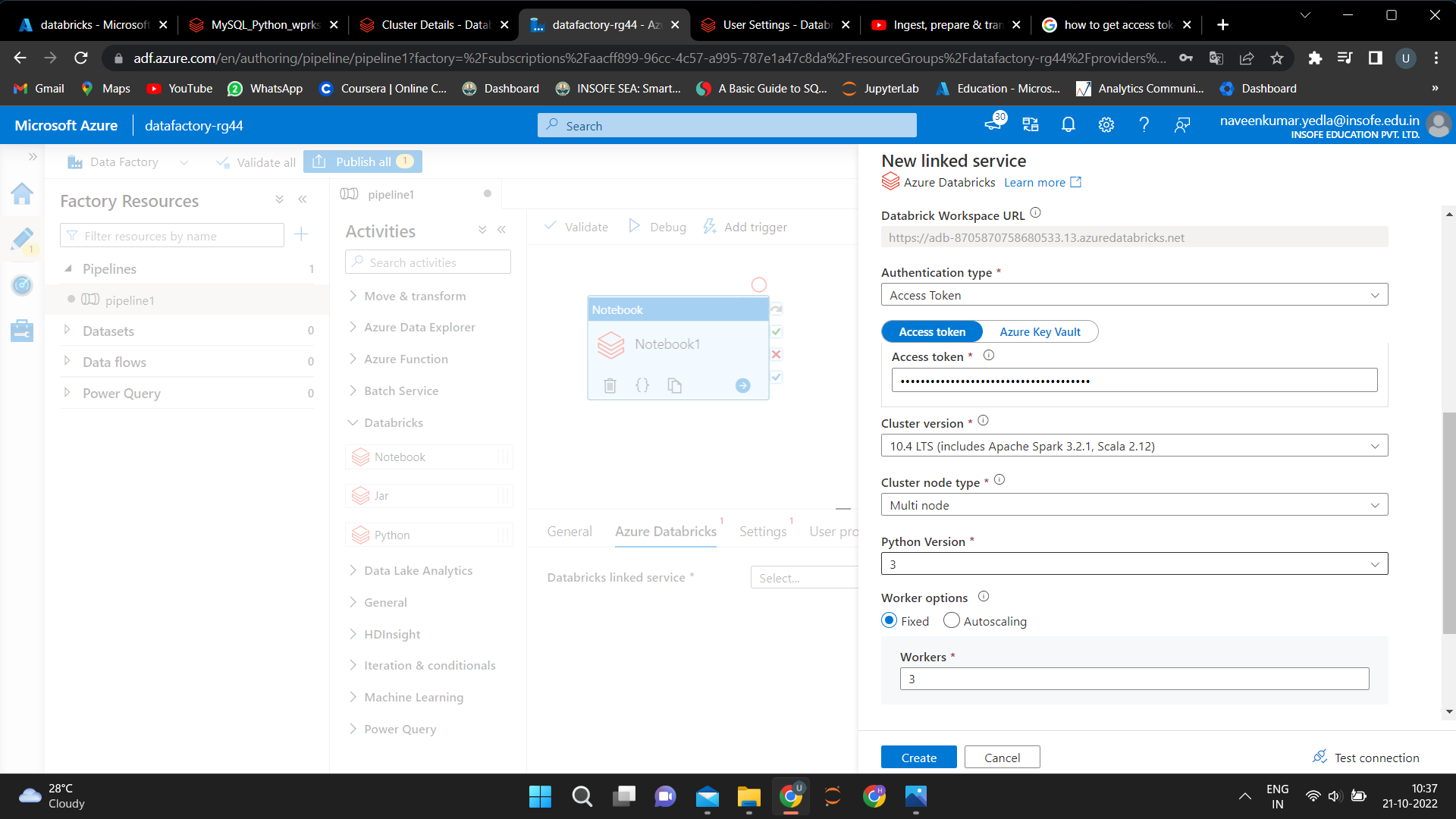


1. Create a new linked service

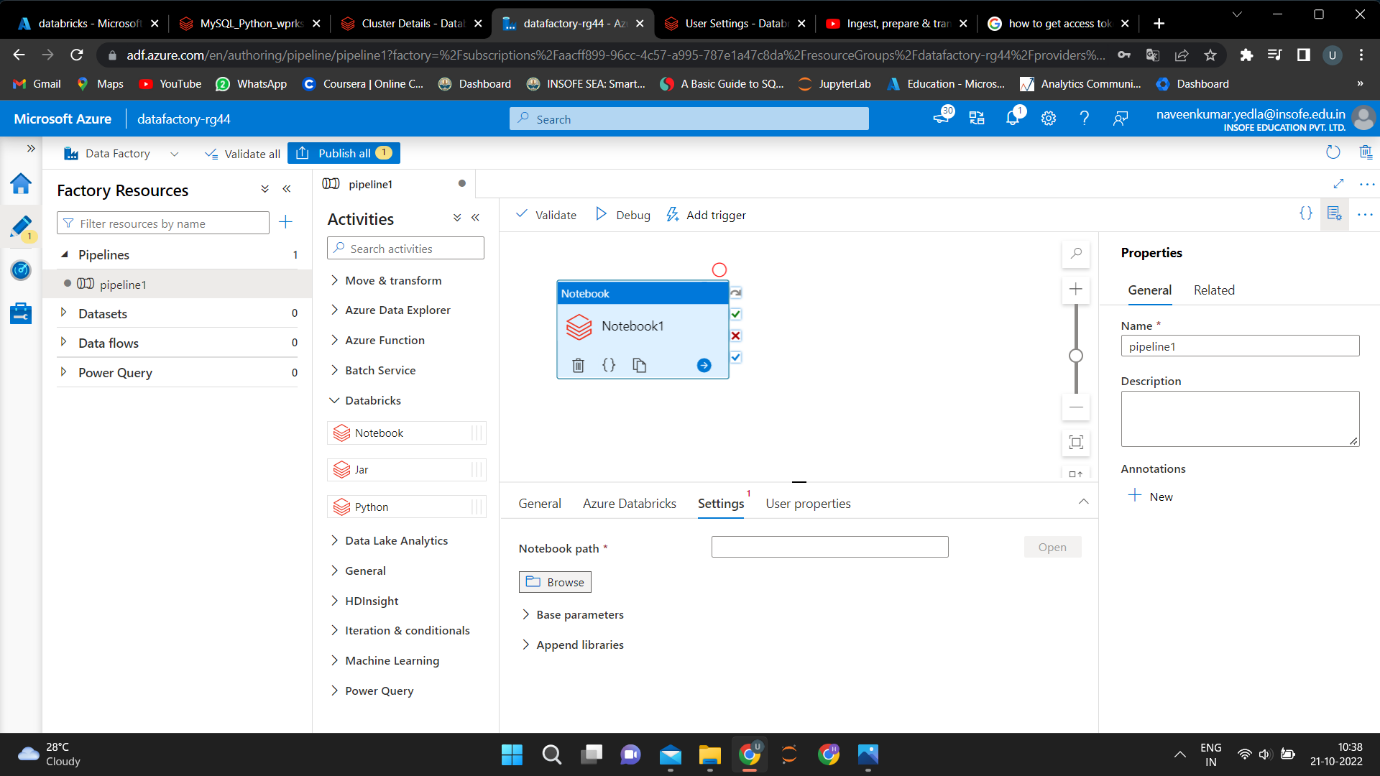
* Give name of your Databricks
* Add your description
* Select AutoResolvingIntegrationRuntime, for connect via integration runtime
* Select from azure subscription method for account selection method
* Select your type of subscription for azure subscription



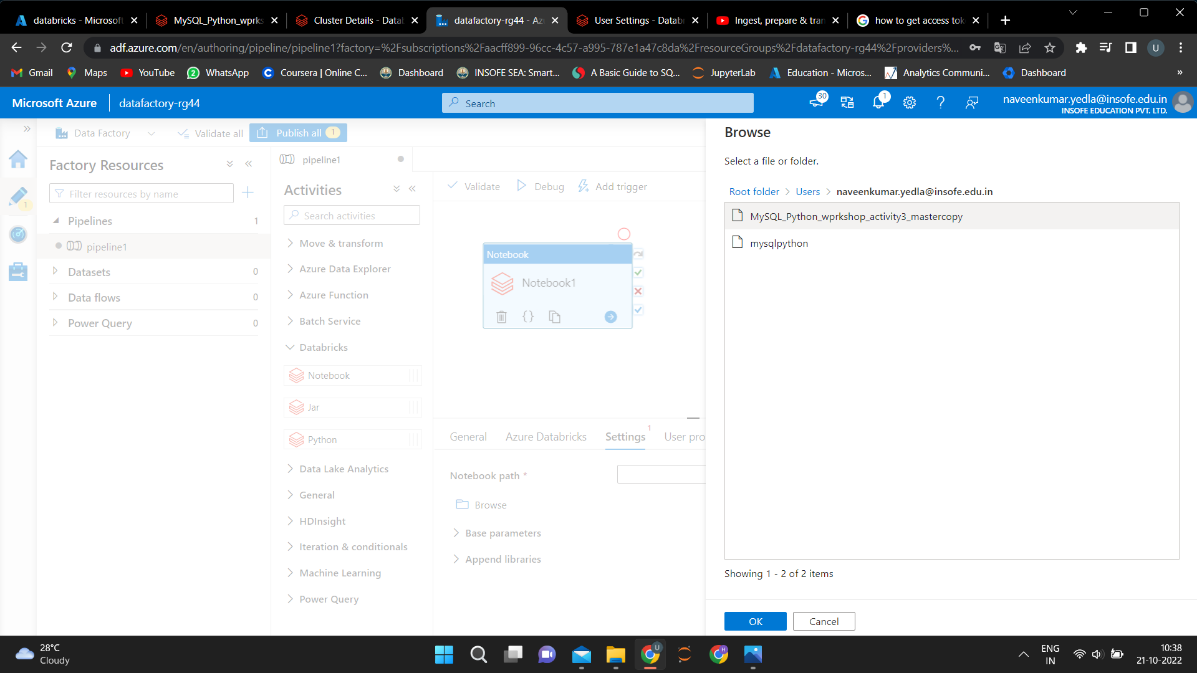
* Give the URL of Databricks that you have created for Databricks workspace URL
* Select access token for authentication type
* Give access token that can be created from the Databricks user account for access token
* Give version of the cluster that you have taken at the time of creation of Databricks for cluster version
* Cluster node type that you have taken at the time of creation of Databricks for cluster node
* Select required Python version
* Select the number of workers
* Then click on create, linked service Databricks is created.

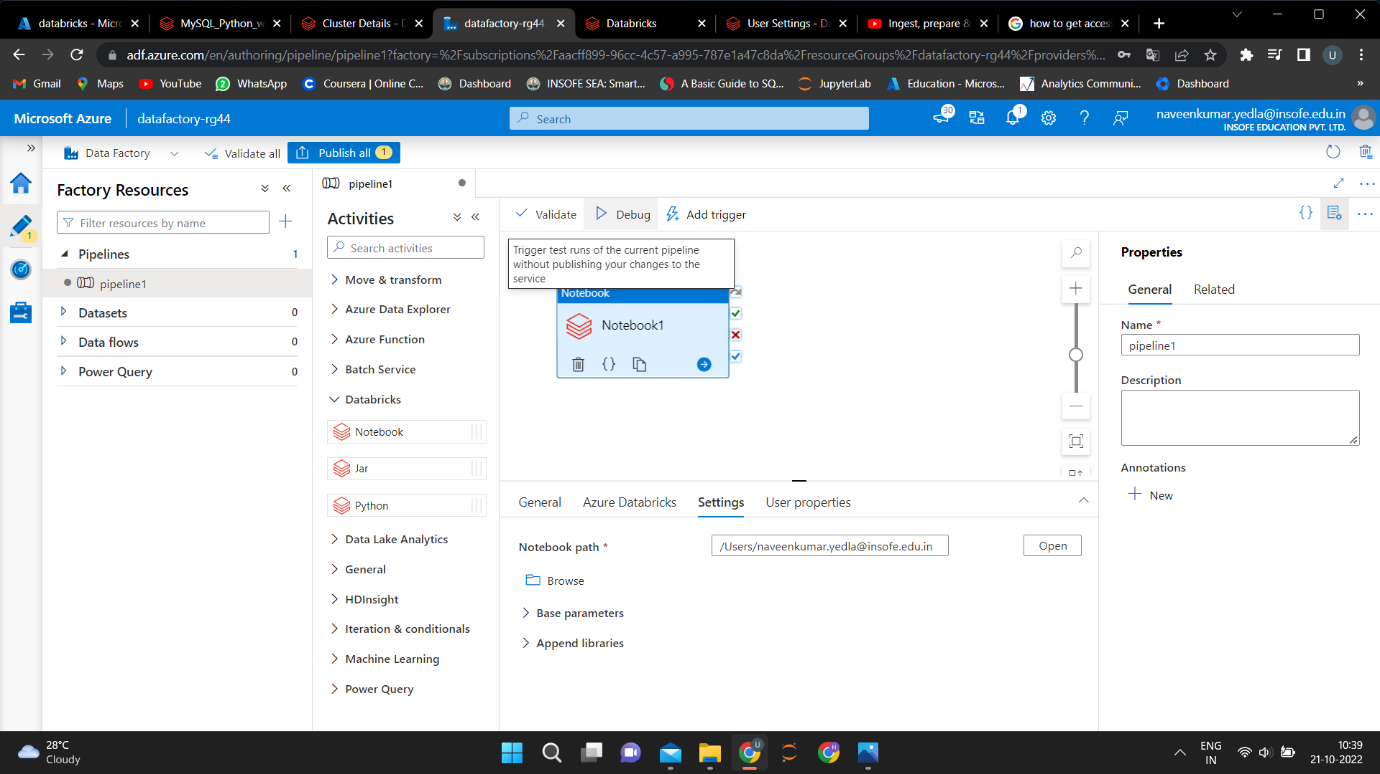


1. In setting blade, browse the notebook from the workspace of Databricks.



1. Select the notebook from the Databricks to execute





1. Now debug the pipeline, pipeline is executed and pipeline ID is generated

