

Create a spark notebook in Azure Synapse Analytics

Note:

You are not required to complete the processes, tasks, activities, or steps presented in this example. The various samples provided are for illustrative purposes only and it's likely that if you try this out you will encounter issues in your system.

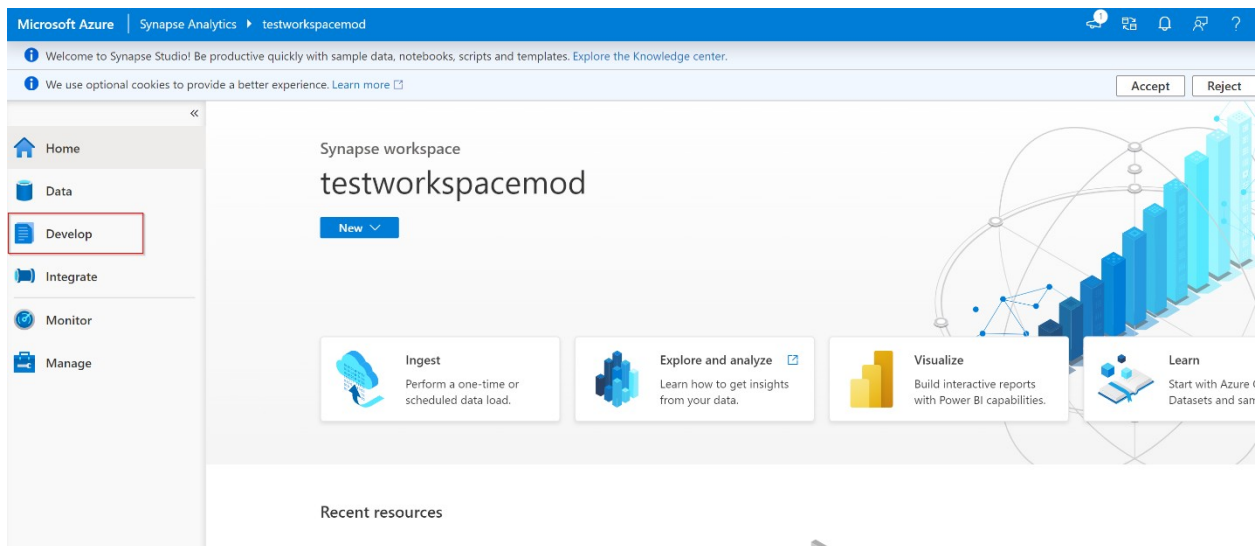
1. In the Azure portal, navigate to the Azure Synapse workspace you want to use, and select **Open Synapse Studio**.

The screenshot shows the Azure Synapse Studio interface for a workspace named 'testworkspacemod'. The left sidebar contains navigation links for Overview, Activity log, Access control (IAM), Tags, Settings, SQL Active Directory admin, Properties, Locks, Analytics pools, Security, Monitoring, Automation, Support + troubleshooting, and New support request. The main content area displays workspace details such as Resource group (testmod), Status (Succeeded), Location (West US 2), Subscription (MSFT CSA Internal Subscription), Subscription ID, Managed virtual network (No), Managed Identity object, Workspace web URL, and Tags. Below this, the 'Getting started' section highlights 'Open Synapse Studio' with a description and an 'Open' button. To the right, there is a 'Read documentation' link. The 'Analytics pools' section includes a search bar and a table listing pools. The 'SQL pools' section shows a 'Built-in' pool of type 'Serverless'. The 'Apache Spark pools' section shows a 'sparkpoolmod' pool of type 'Apache Spark pool'.

Name	Type
Built-in	Serverless
sparkpoolmod	Apache Spark pool

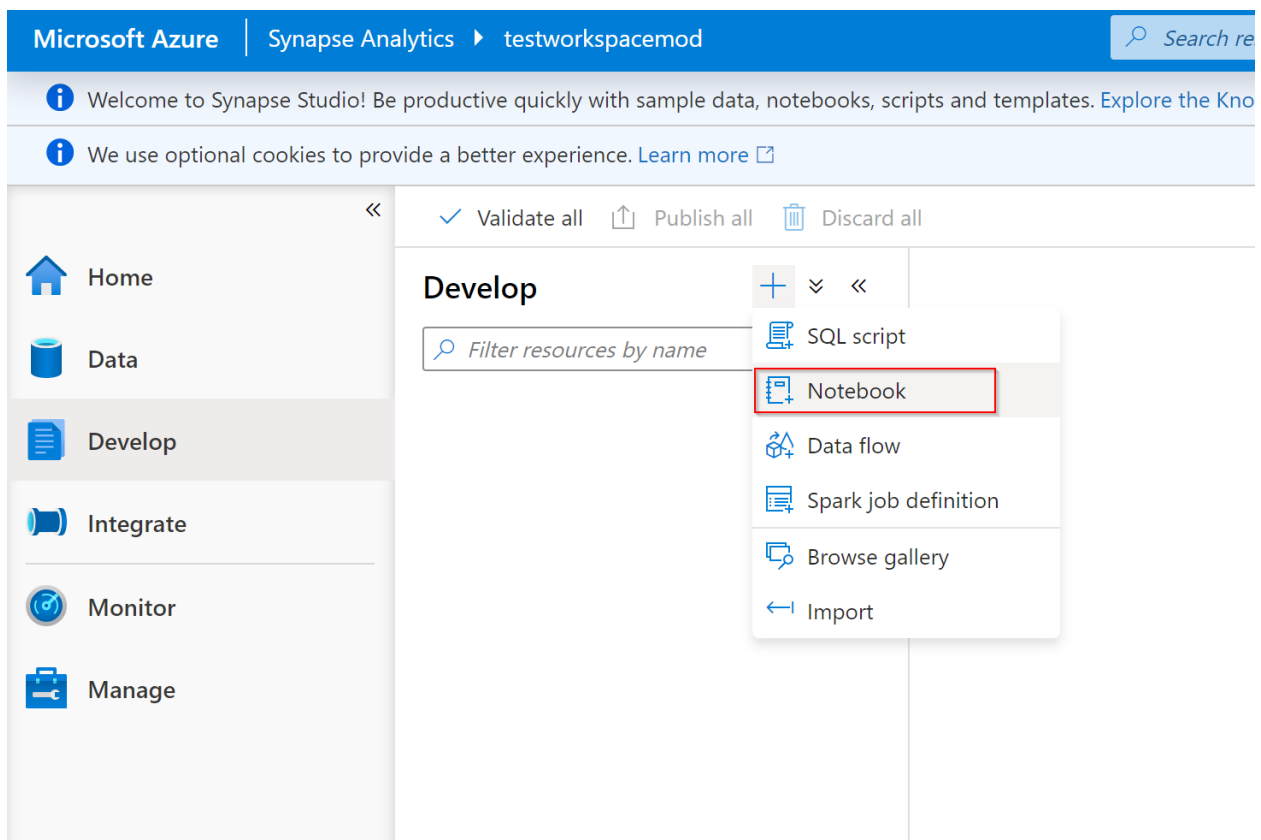
Open Synapse Studio to Create a Notebook

2. Once the Azure Synapse Studio has launched, select **Develop**.



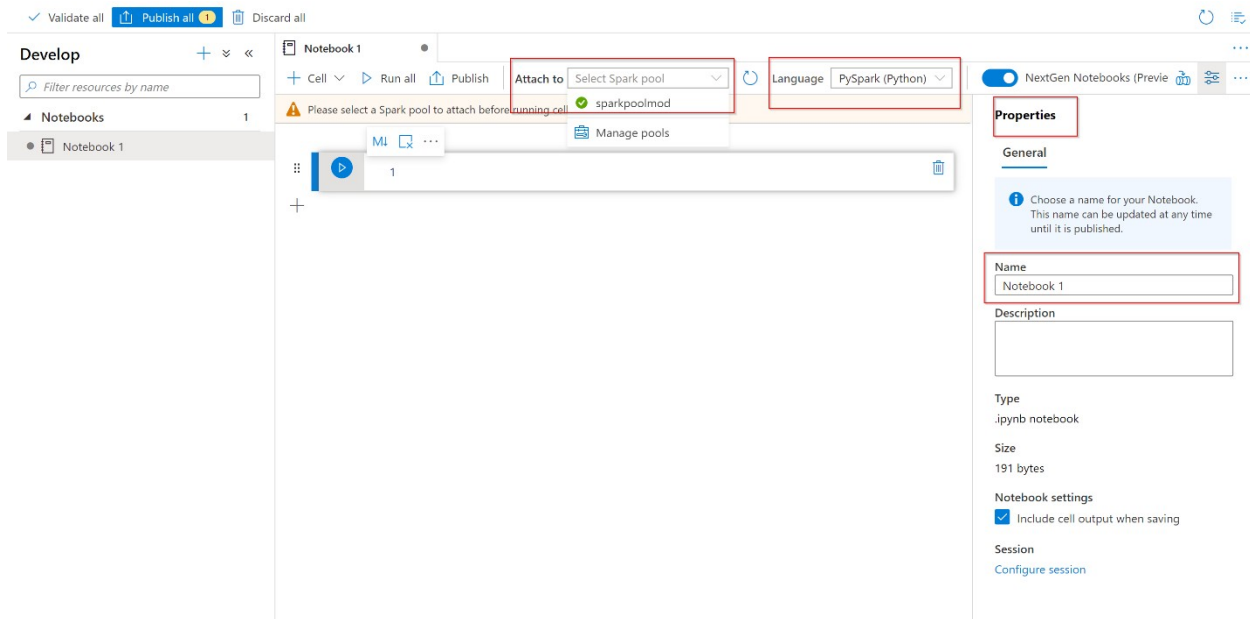
Develop Hub Synapse Studio

3. From there, select the "+" icon, and then select **Notebook**. A new notebook is created and opened with an automatically generated name.



Select Create New Notebook

4. In the **Properties** window, provide a name for the notebook.



Properties for Notebook

5. On the toolbar, select **Publish**.

If there is only one Apache Spark pool in your workspace, then it's selected by default. Use the drop-down to select the correct Apache Spark pool if none is selected. Add code by using the "+" icon or the "+ Cell" icon. Select **Code cell** to input code or **Markdown cell** to input markdown content.

When adding a code cell, the default language is PySpark. In the code cell below, you are going to use Pyspark. However, other supported languages are Scala, SQL, and .NET for Spark.

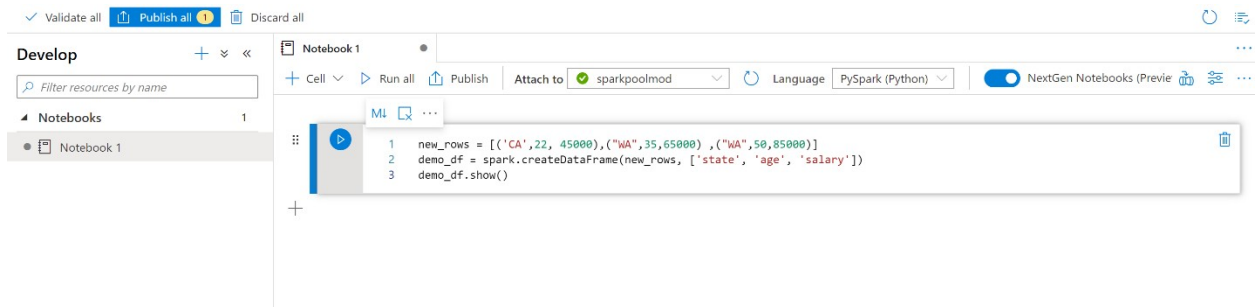
You can select the code below, and paste it into a code cell in your notebook. Run the code to create a Spark DataFrame:

```
new_rows = [('CA', 22, 45000), ('WA', 35, 65000), ('WA', 50, 85000)]
demo_df = spark.createDataFrame(new_rows, ['state', 'age', 'salary'])
demo_df.show()
```

When you want to run a cell, you can use the following methods:

- Press **SHIFT + ENTER**.
- Select the blue play icon to the left of the cell.
- Select the **Run all** button on the toolbar.

1
2
3



Run code within a Notebook

You can see the Apache Spark pool instance status below the cell you are running and also on the status panel at the bottom of the notebook. In the output cell, you see the output.