Run spark notebooks

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.

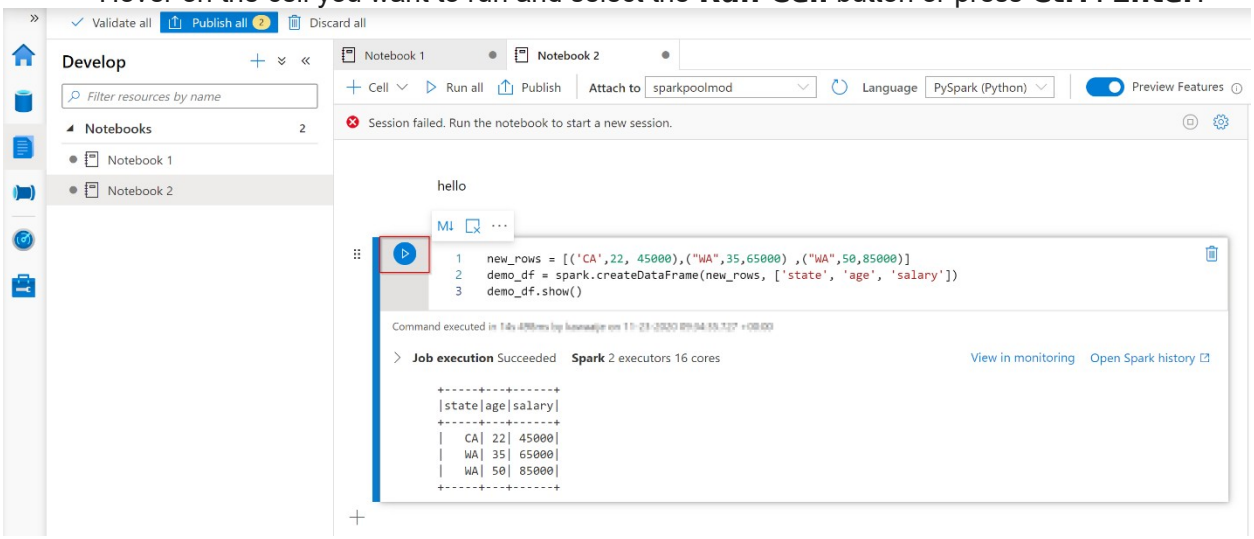
In the following steps, you will explore the run features of a notebook:

- Run a cell
- Run all cells
- Cancel a running cell
- Cell Status indicator
- Spark progress indicator

Run a cell

There are several ways to run the code in a cell.

- Hover on the cell you want to run and select the **Run Cell** button or press **Ctrl+Enter**.

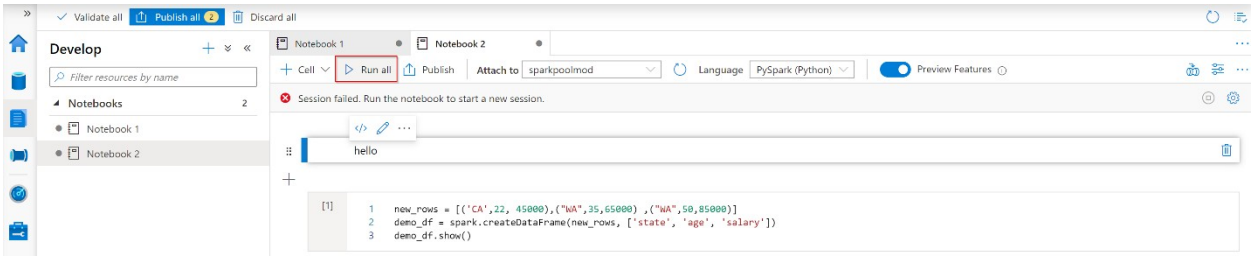


Run cell in notebook

- In addition, you can press **Shift+Enter** to run the current cell and select the cell below.
- Alternatively press **Alt+Enter** to run the current cell and insert a new cell below.

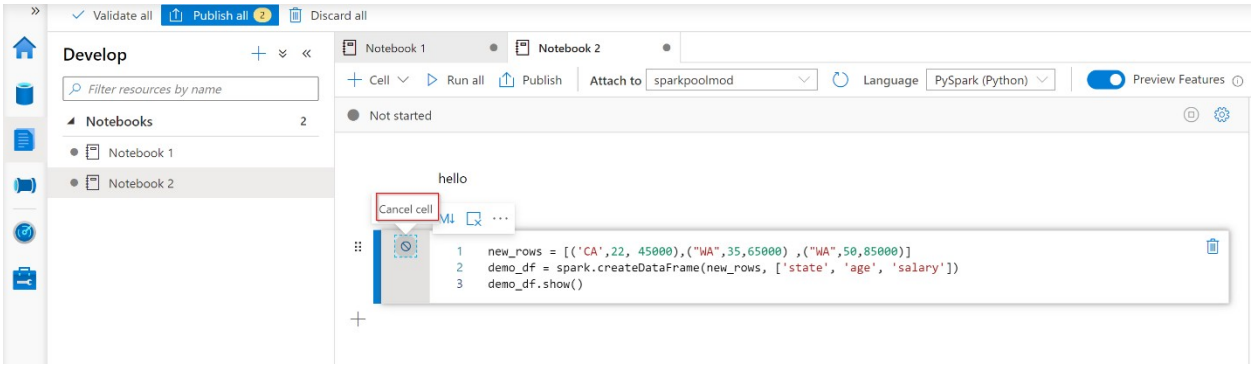
Run all cells

Select the **Run All** button to run all the cells in current notebook in sequence.



Cancel a running cell

In order to cancel a running cell, hover over left side of the cell that is running and select "**cancel cell**".

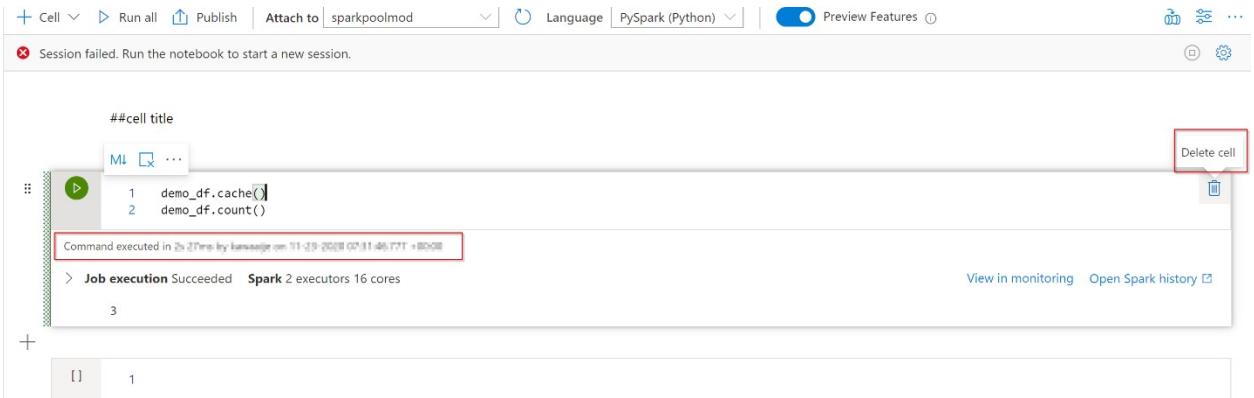


Cancel a running cell

Cell status indicator

A step-by-step cell execution status is displayed beneath the cell to help you see its current progress.

Once the cell run is complete, an execution summary with the total duration and end time are shown and kept there for future reference.



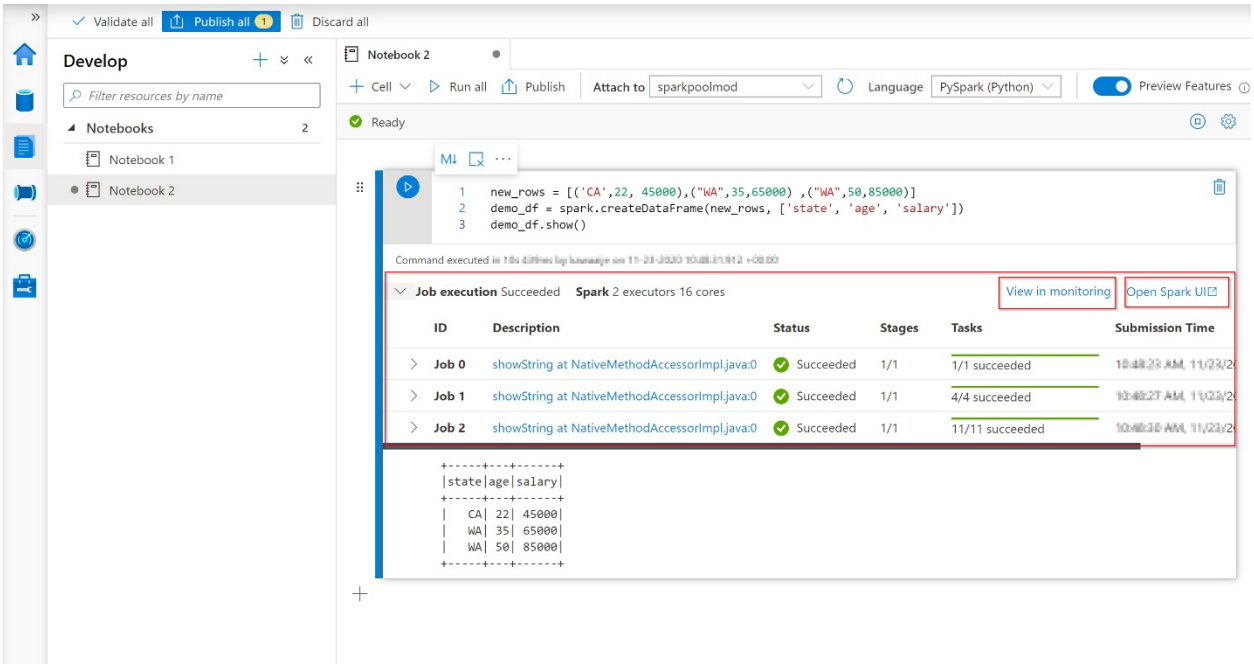
Cell status Indicator

Spark progress indicator

Azure Synapse Studio notebook is purely Spark based. Remotely, the code cells are executed on the serverless Apache Spark pool.

If you want to view the progress of a spark job, you can see in real time the job execution status below a cell. The number of tasks per each job or stage help you to identify the parallel level of your spark job.

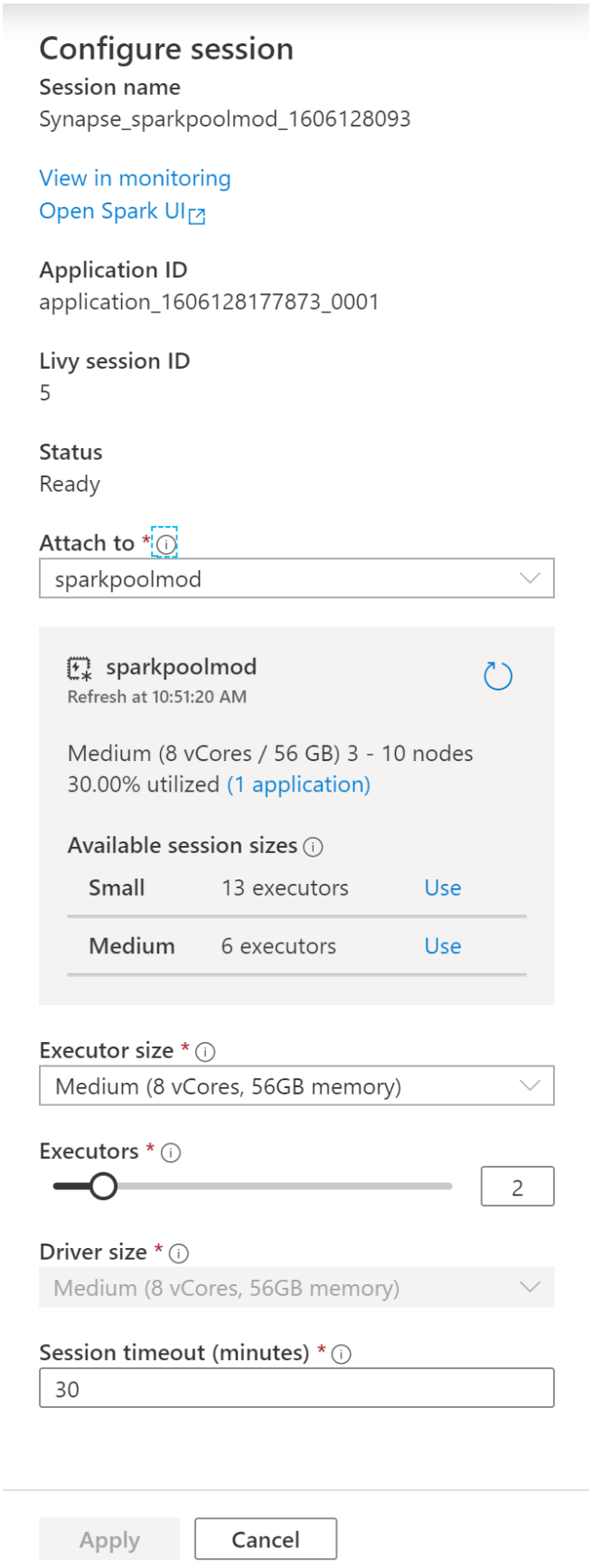
You can also drill deeper to the Spark UI of a specific job (or stage) by selecting the link on the job (or stage) name.



Spark session config

You can specify the timeout duration, the number, and the size of executors to give to the current Spark session in **Configure session**.

Restart the Spark session for configuration changes to take effect. All cached notebook variables are cleared.


Configure session
Session name
Synapse_sparkpoolmod_1606128093
View in monitoring
Open Spark UI
Application ID
application_1606128177873_0001
Livy session ID
5
Status
Ready
Attach to *
sparkpoolmod
sparkpoolmod
Refresh at 10:51:20 AM
Medium (8 vCores / 56 GB) 3 - 10 nodes
30.00% utilized (1 application)
Available session sizes
Small 13 executors Use
Medium 6 executors Use
Executor size *
Medium (8 vCores, 56GB memory)
Executors *
2
Driver size *
Medium (8 vCores, 56GB memory)
Session timeout (minutes) *
30
Apply
Cancel