

Lab 3 - Spark

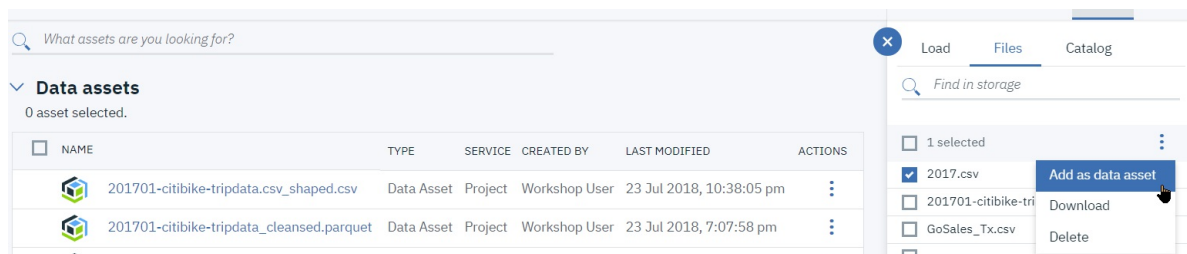
This hands-on lab section will be essentially driven by the instructions in the `Lab3-WatsonStudio_Apache_Spark_Basics_cleared.ipynb` notebook.

We will use a rather large (>1GB) data file, named `2017.csv`, retrieved from its source on the internet.

The first section of the notebook will fetch this file and store it into your Cloud Object Storage. It will then be added as Data Asset to the project.

Setting up the Hands-On lab notebook: file download

1. Go to your `WatStud_Workshop` project, add a new notebook from the Assets tab, using the `From file` option and selecting the `Lab3-WatsonStudio_Apache_Spark_Basics_cleared.ipynb` file.
NOTE: Make sure your notebook has a Spark runtime selected (named `spark-xy`).
2. Execute the first set of instructions in the notebook to fetch data over HTTP then store it into COS.
3. Close the notebook and switch back to your project's assets page, add the newly added file from COS as data asset:



Run the Hands-On through the notebook

1. Open the notebook again, and start executing the cells after the file download section.



2. Follow the instructions