Introduction to Data Wrangling with Jupyter Notebooks

IBM **Developer**

Upkar Lidder
> ulidder@us.ibm.com
> @lidderupk

http://bit.ly/spectra-ibm
https://slides.com/upkar/jupyter

Prerequisites

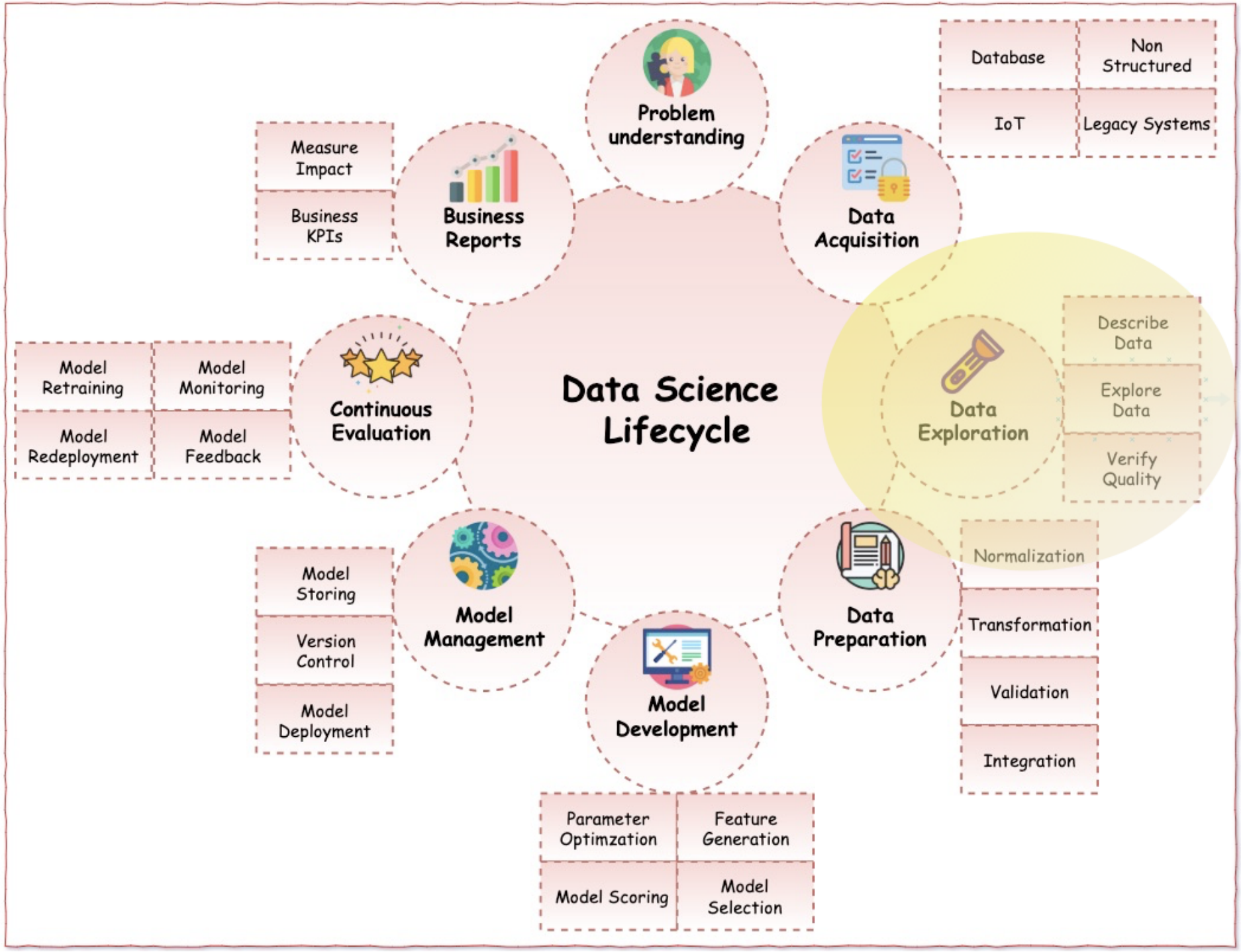
1. Create IBM Cloud Account using THIS URL

http://bit.ly/spectra-ibm

2. Check your email and activate your account. Once activated, log back into your IBM Cloud account using the link above.

3. If you already have an account, use the above URL to sign into your IBM Cloud account.

Data Science Lifecycle



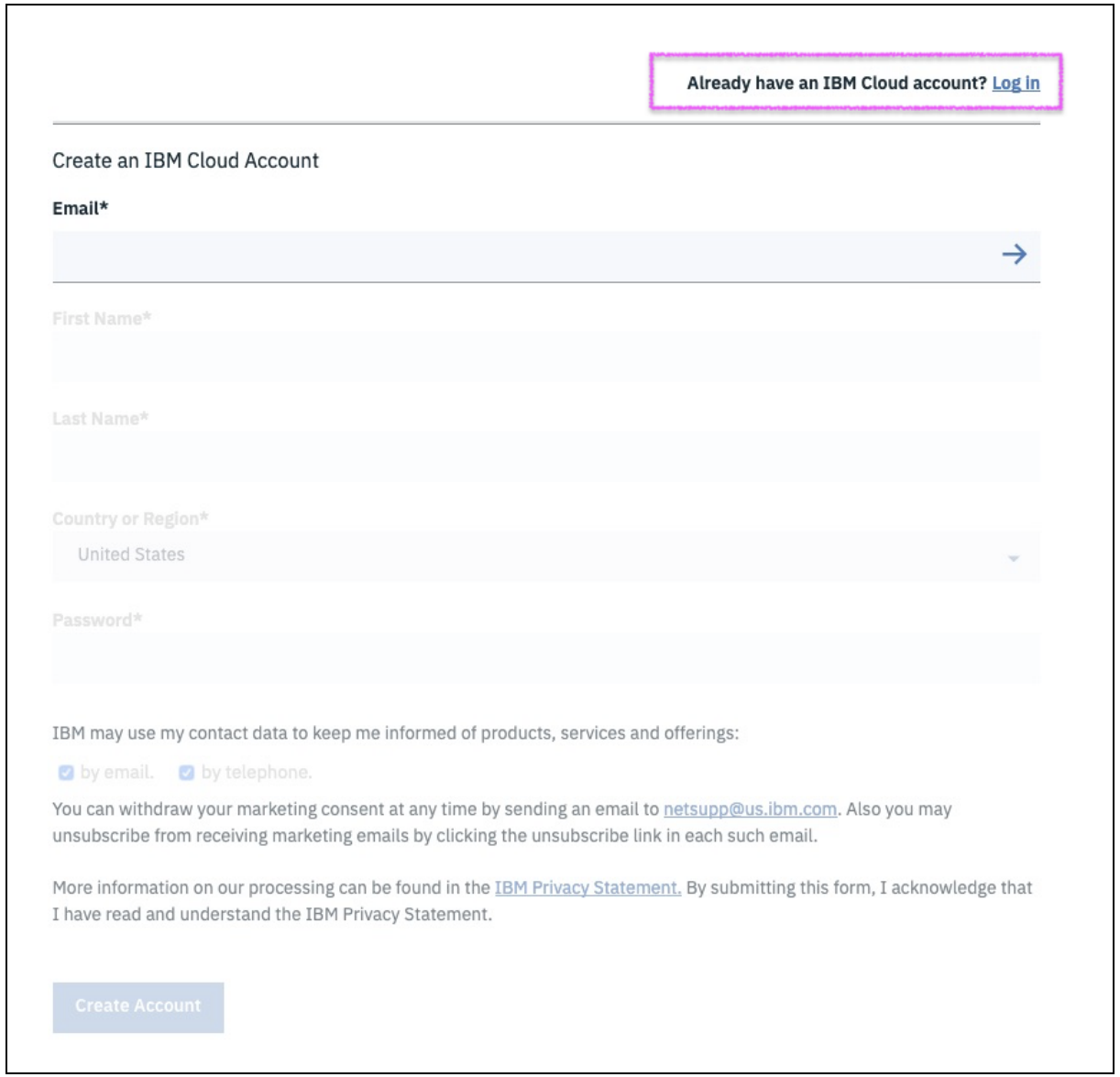
Workshop - Goals

Get acquainted with Pandas and Jupyter Notebooks on the cloud and analyze a movies dataset!

Steps

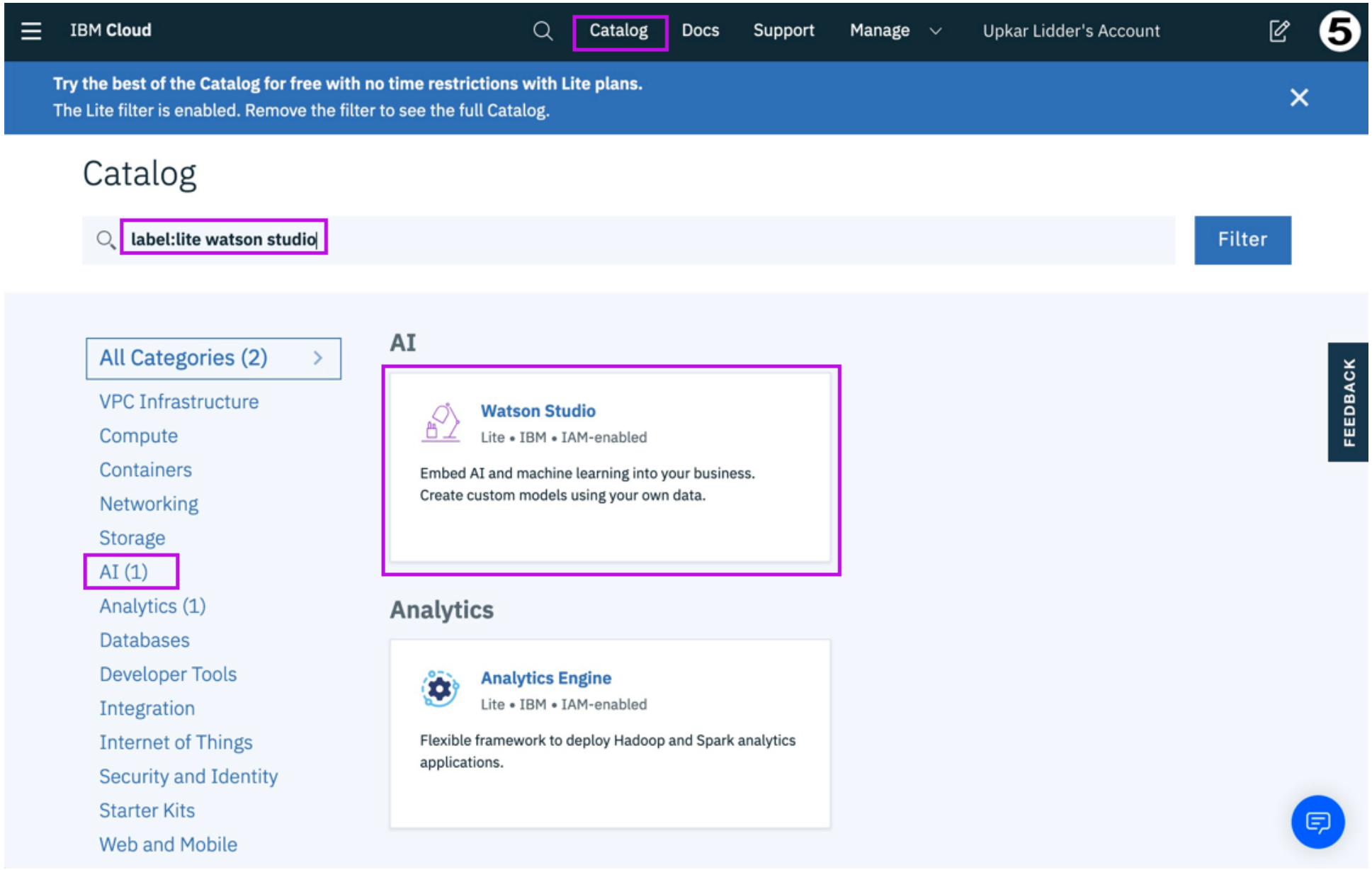
1. Sign up / Log into *IBM Cloud* - http://bit.ly/spectra-ibm
2. Create *a Watson Studio Service*
3. Create a *new project*
4. Import the sample notebook to your project
5. *RUN* the cells and explore data!

Step 1 - sign up/ log into IBM Cloud

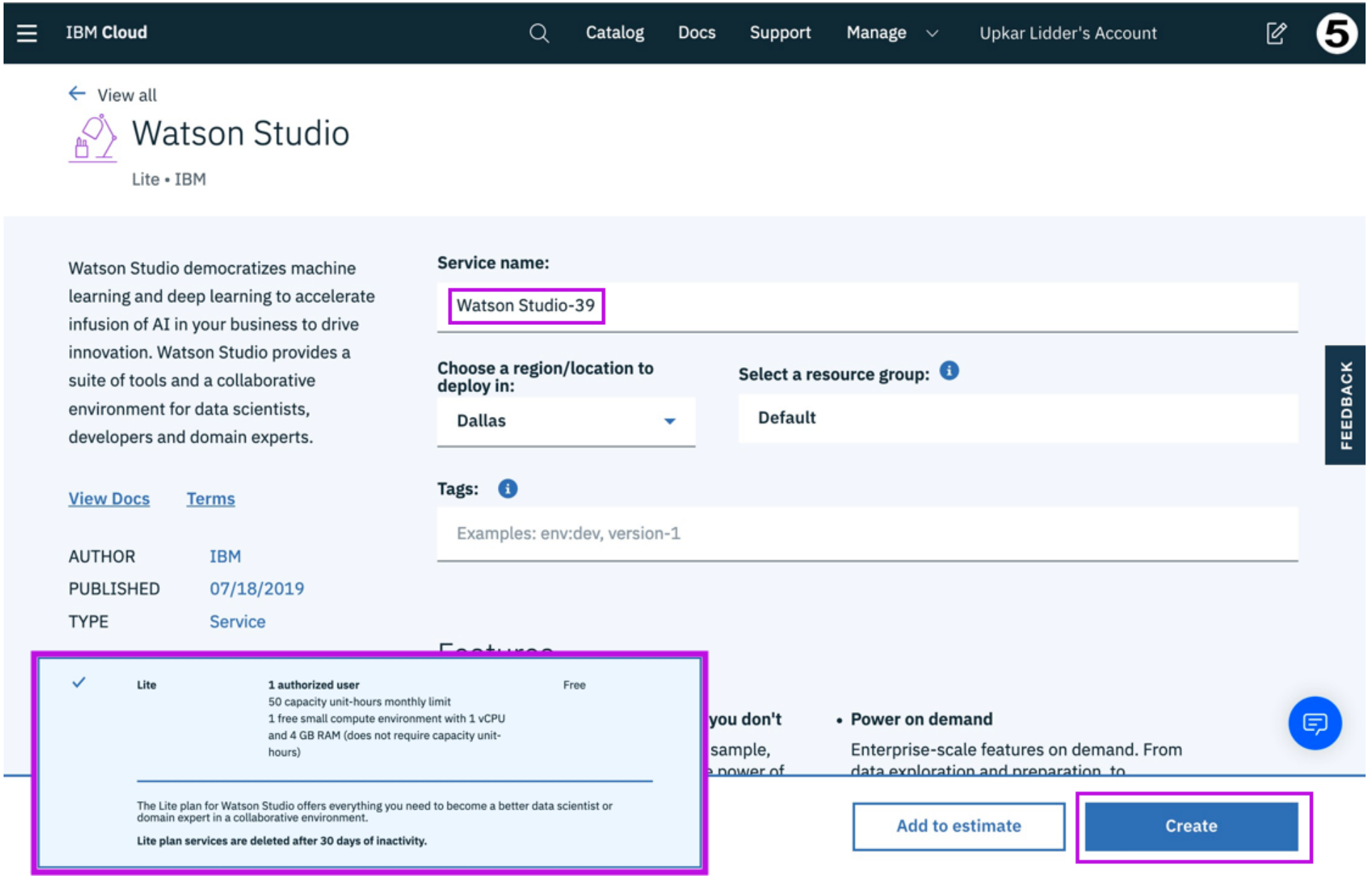


http://bit.ly/spectra-ibm

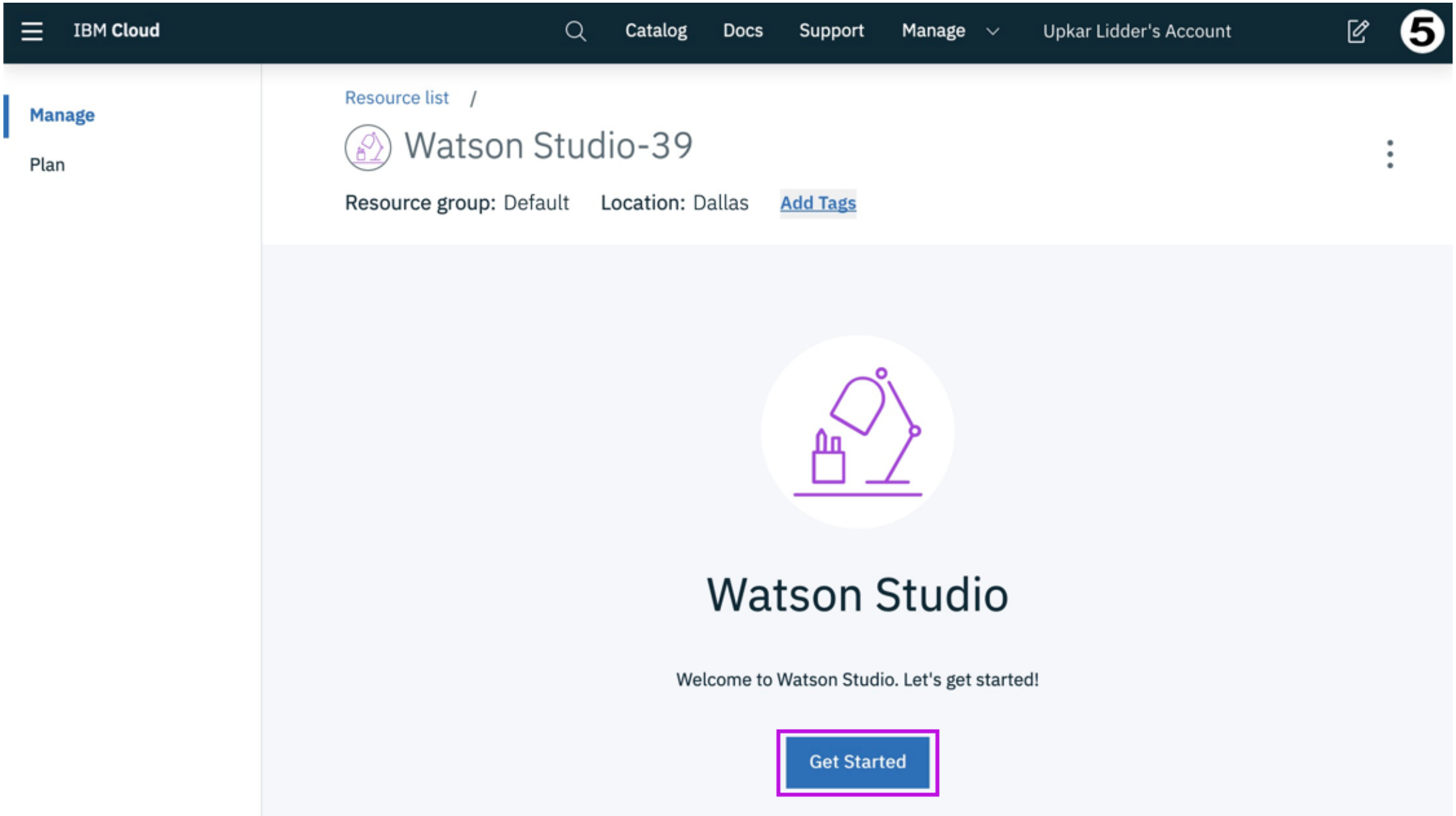
Step 2 - locate Watson Studio in the catalog



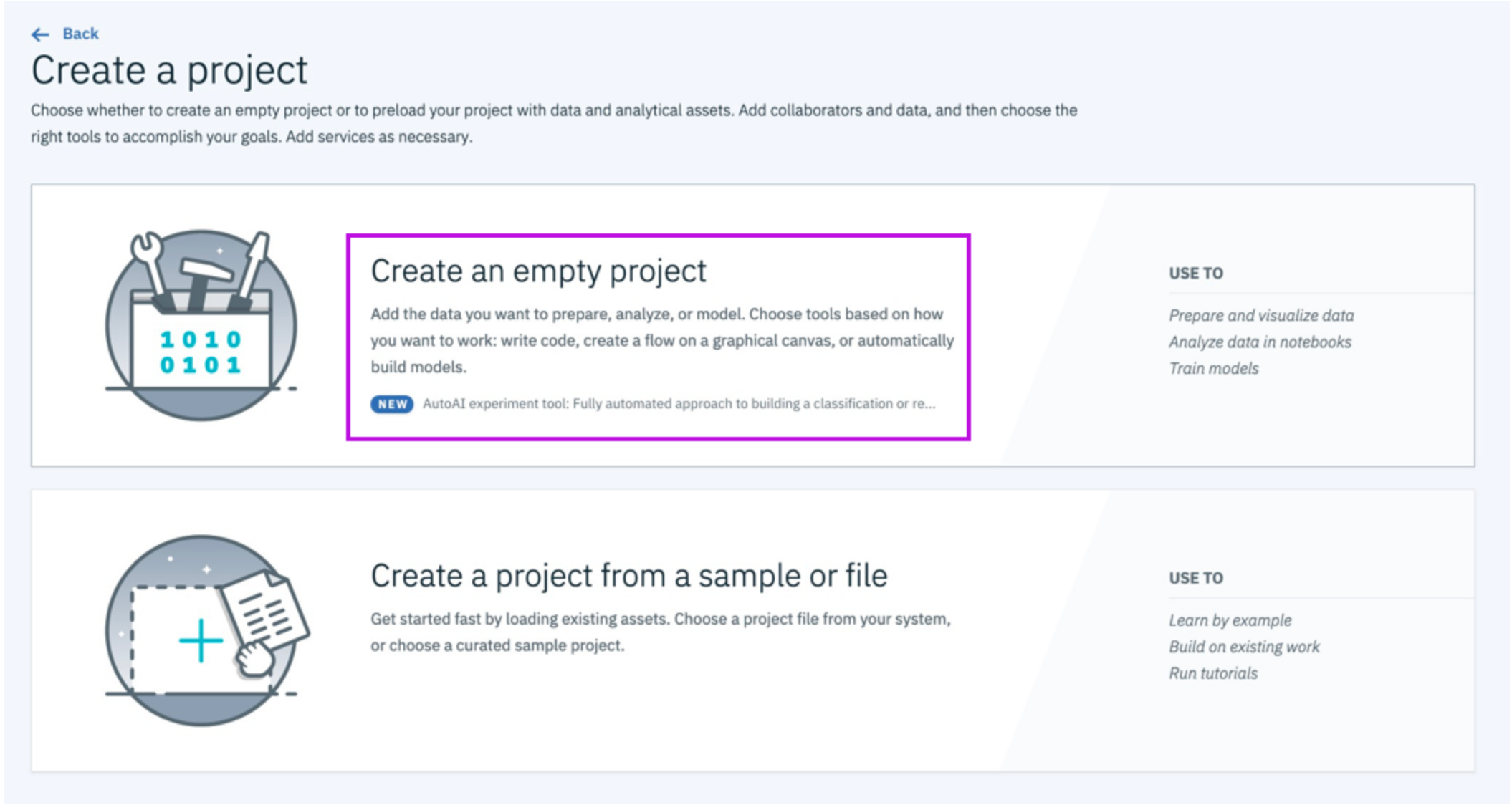
Step 3 - create new watson studio service



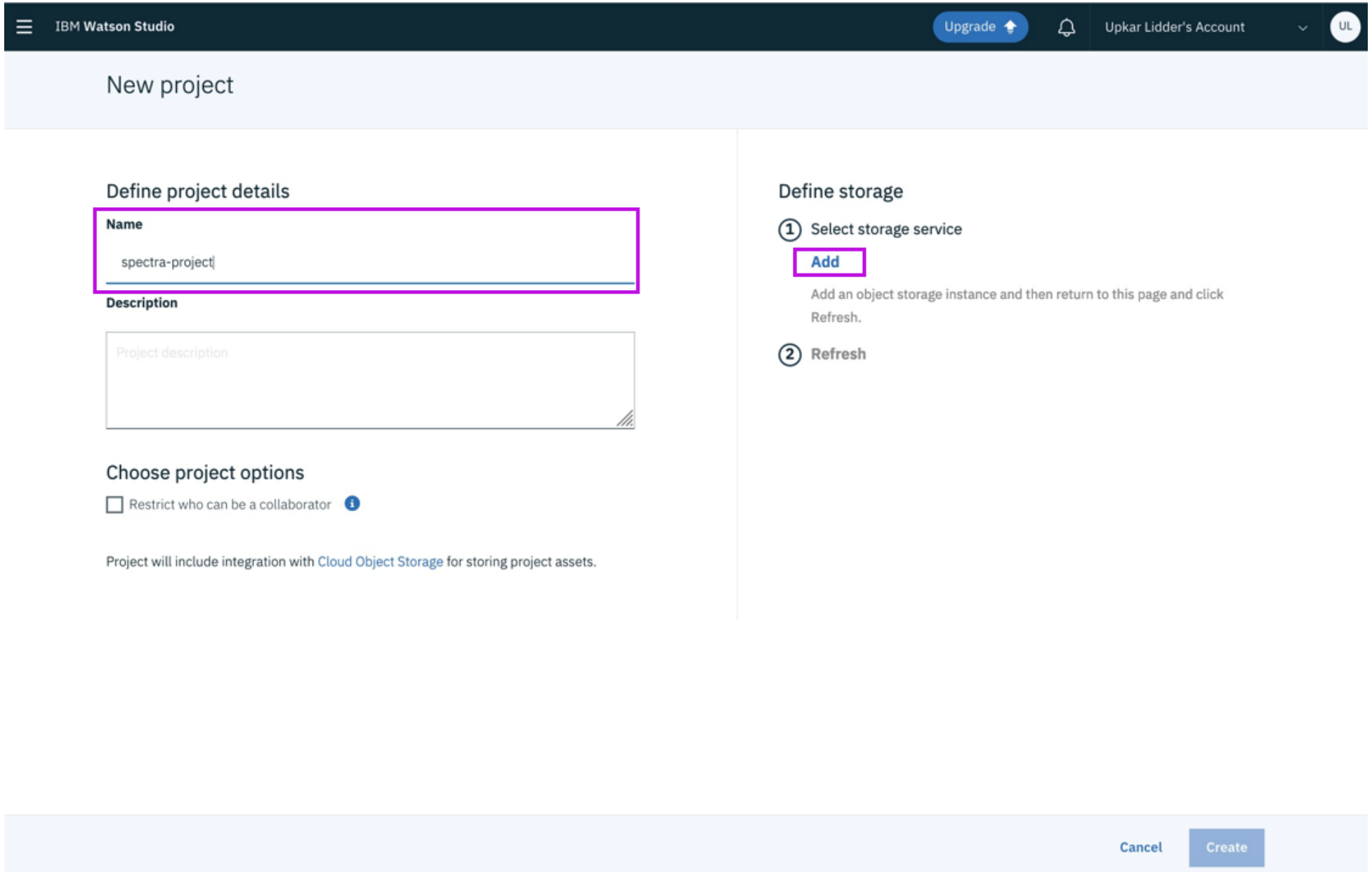
Step 4 - launch Watson Studio



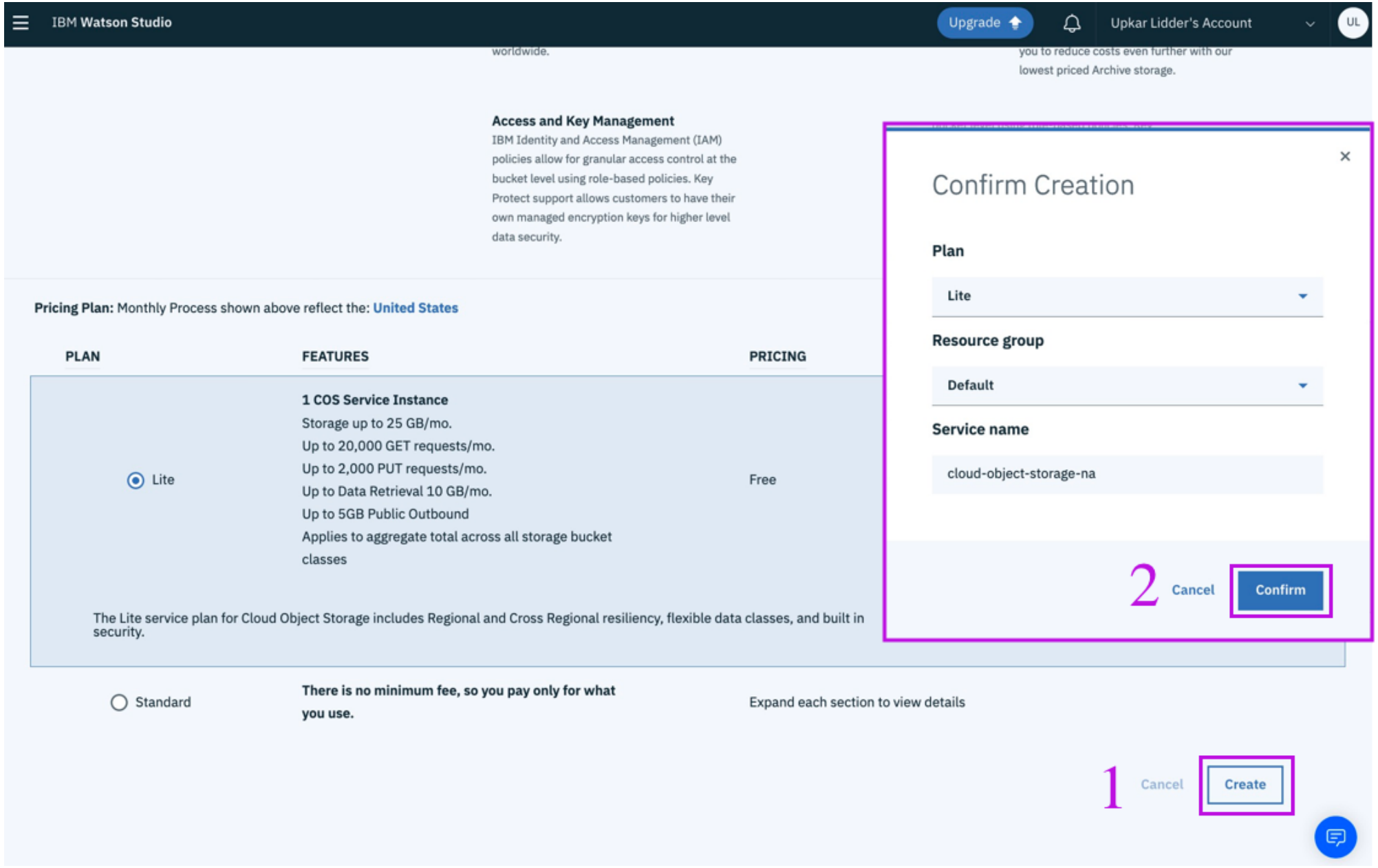
Step 5 - create new project and pick empty template



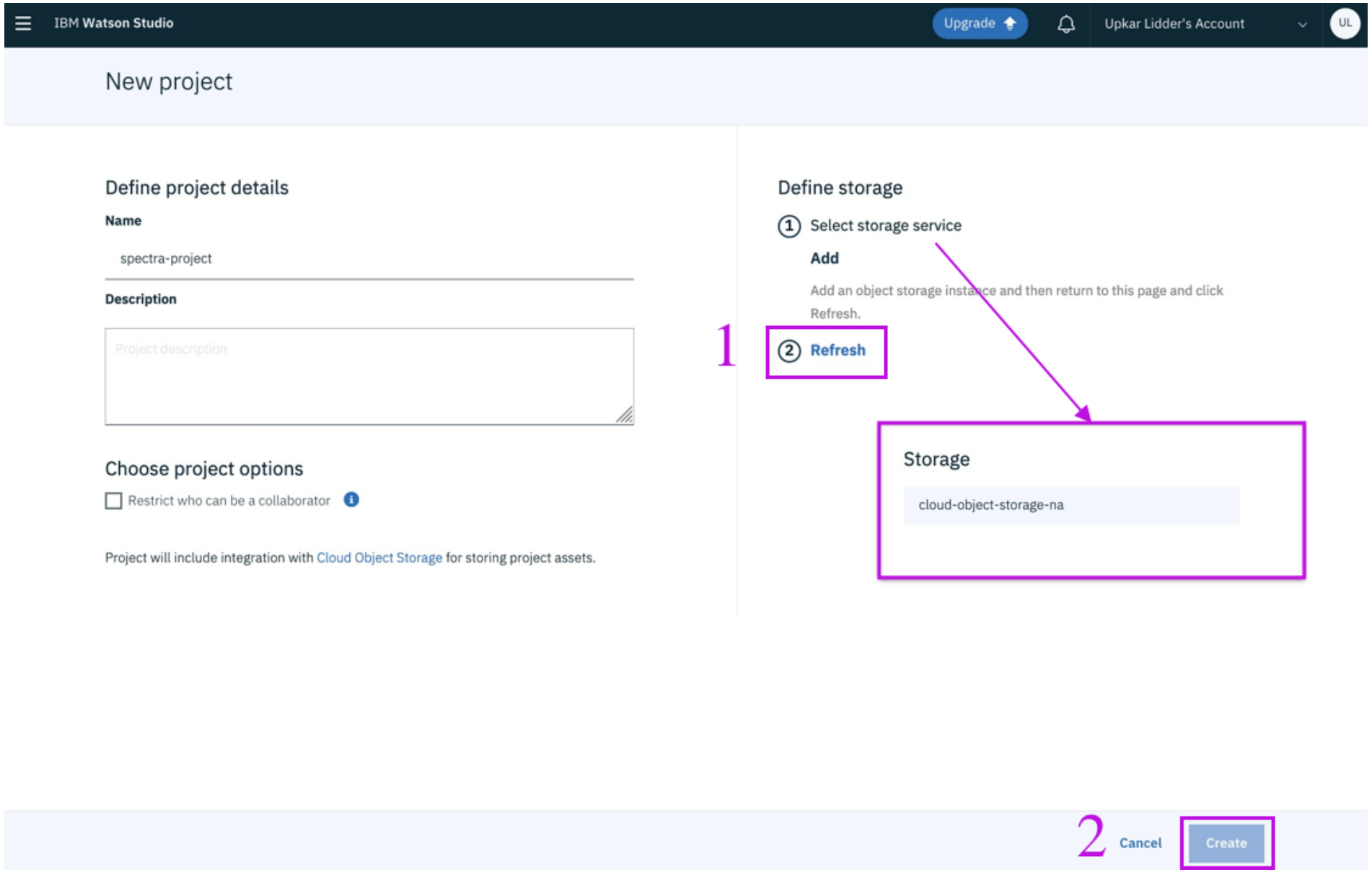
Step 6a - name your project and create storage service



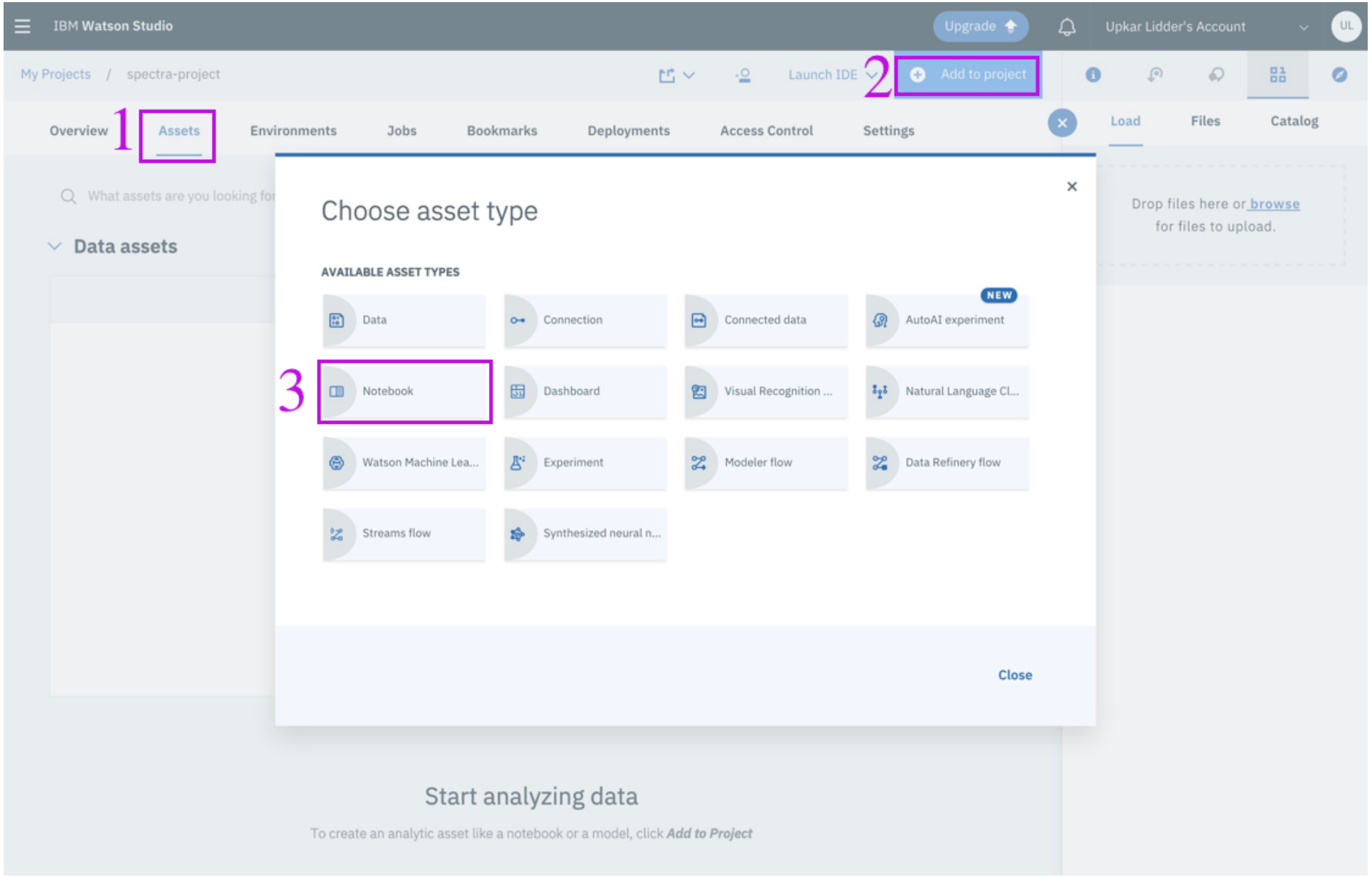
Step 6b - add storage opens a new page



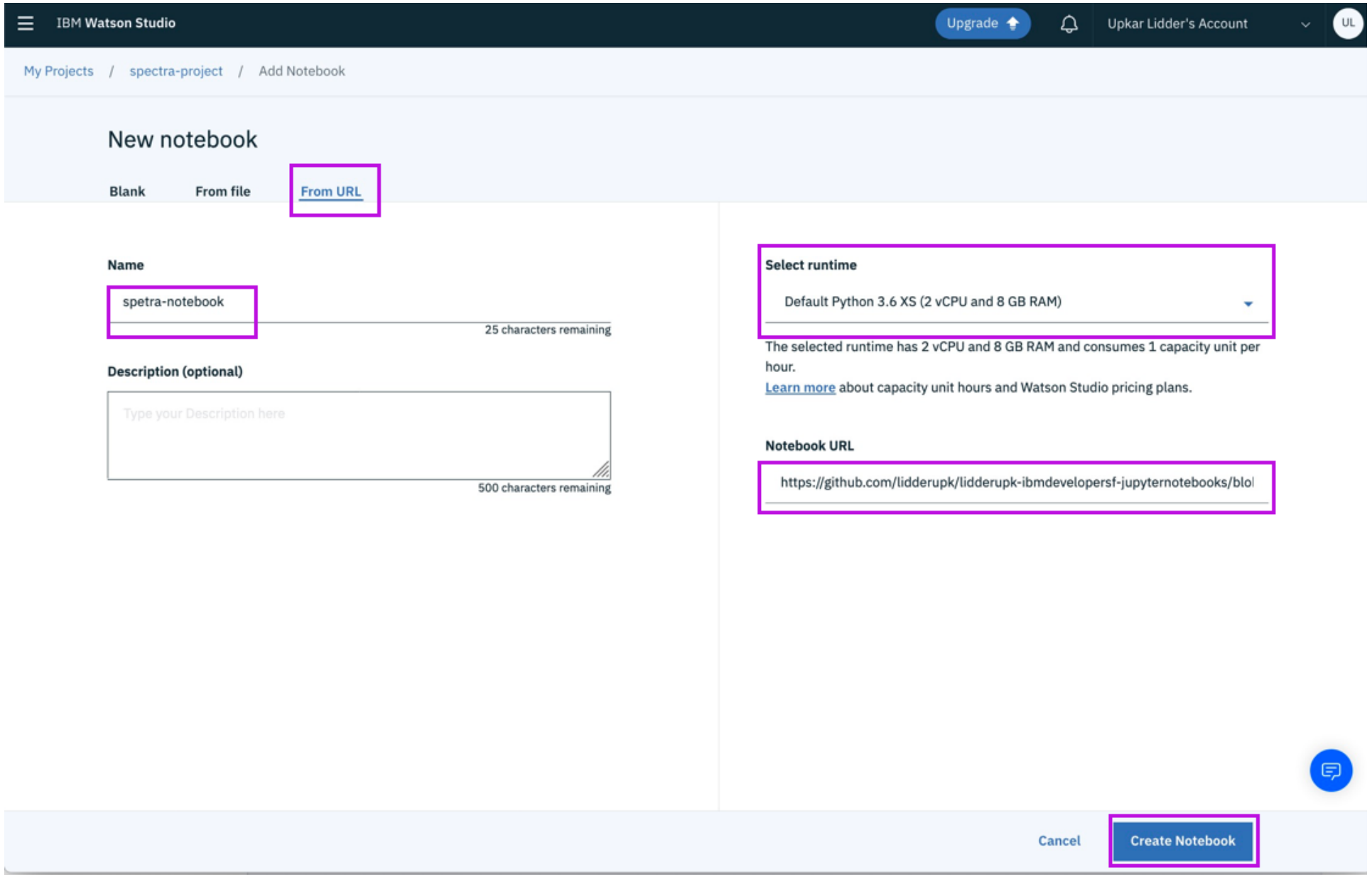
Step 6c - you will be taken back to the first page after storage



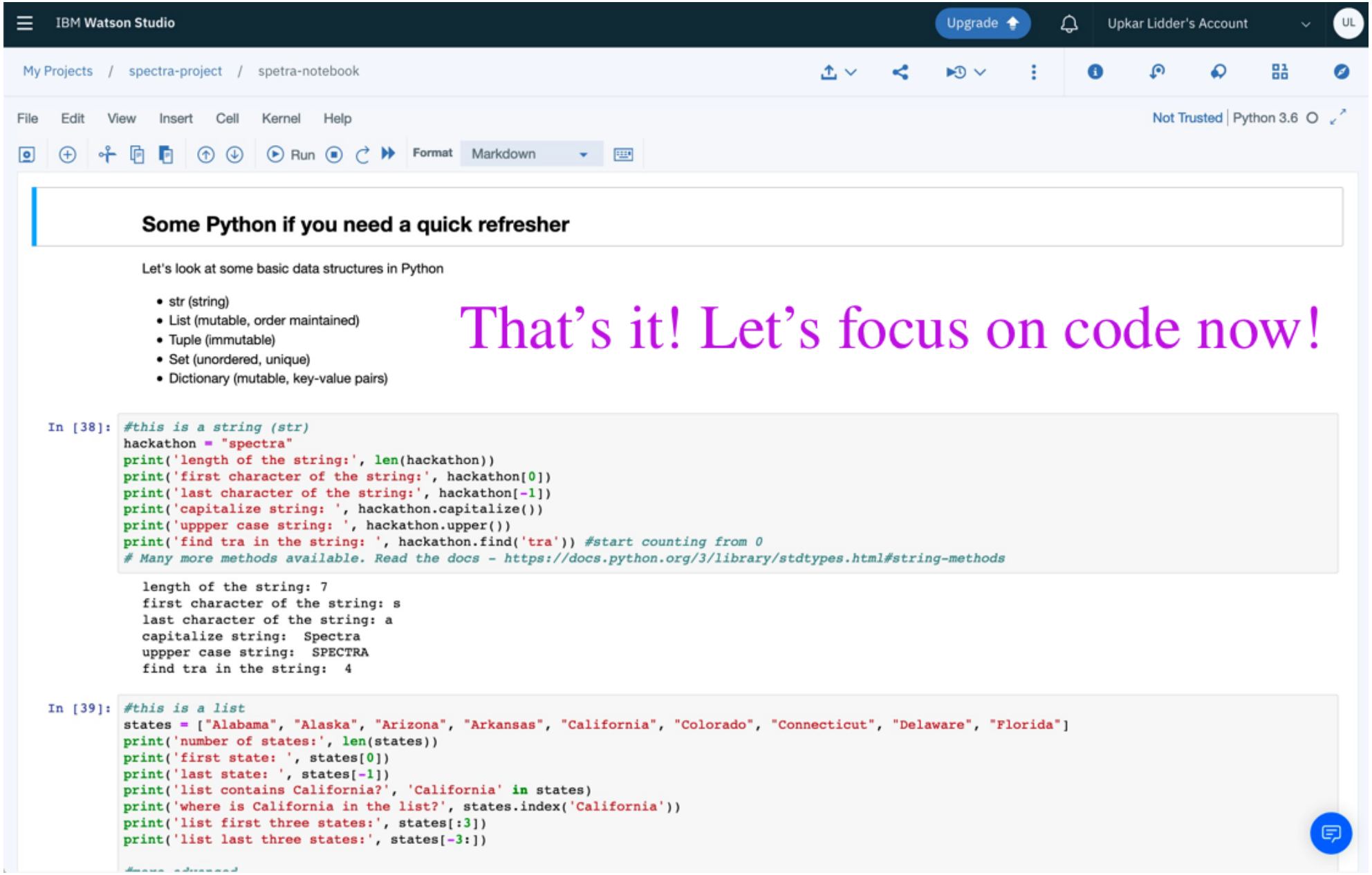
Step 7 - add notebook feature to your project



Step 8 - import notebook, get link from github



Step 9 - Let's look at data now!



Some links for the workshop

IBM Cloud account - http://bit.ly/spectra-ibm

Jupyter Notebook - https://github.com/lidderupk/lidderupk-ibmdevelopersf-jupyternotebooks/blob/master/asset/spectra-pandas.ipynb

Datasets

Casts - https://ibm.box.com/shared/static/569iue5znz5angfxaaojbd7olgegk0bz.csv

Release Dates - https://ibm.box.com/shared/static/fxu6rhfktvjs0uvgtbhjsp5g5k9qgjh1.csv

Titles - https://ibm.box.com/shared/static/cw3wqtzuljiyqz4kbuk26ojrrm9rzfow.csv

Workshop Github - https://github.com/lidderupk/lidderupk-ibmdevelopersf-jupyternotebooks

Some links to get data

US Government Open Data - https://www.data.gov/

San Francisco Open Data - https://datasf.org/opendata/

IBM Data Asset eXchange - https://developer.ibm.com/exchanges/data/

Kaggle Datasets - https://www.kaggle.com/datasets

Google Datasets - https://cloud.google.com/public-datasets/

Curated on Github - https://github.com/awesomedata/awesome-public-datasets

Ryan Anderson Blog - https://dreamtolearn.com/ryan/1001_datasets

07.26.19

SERVERLESS DEVELOPER SUMMIT

GALVANIZE, SAN FRANCISCO

Thank you

Let's chat !

Upkar Lidder, IBM

@lidderupk

https://github.com/lidderupk/

ulidder@us.ibm.com