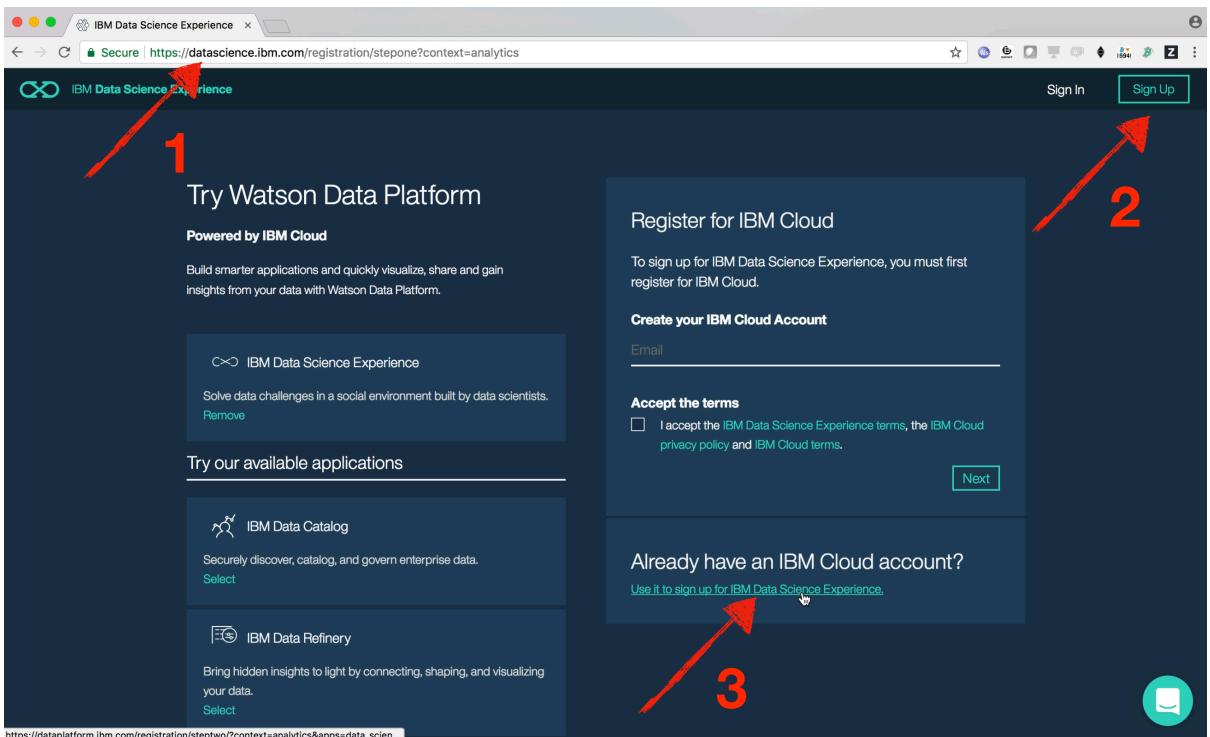


Create and setup a Data Science Experience (DSX) project for jupyter and Apache Spark in the IBM Cloud

Step 1: Signup for DSX

1. Start the process



- a. Open the following URL (1)
<http://datascience.ibm.com>
*HINT: In case you've accidentally created your resource in United Kingdom instead of US South you can just use the following link instead:
<https://eu-gb.datascience.ibm.com/>.*
- b. Click on “Sign Up” (2)
- c. Click on “Use it to sign up for IBM Data Science Experience (3)

2. Make use all dropdowns are filled

Select Organization and Space

Confirm your IBM Cloud organization and space information below.
[Or create new organization and space](#)

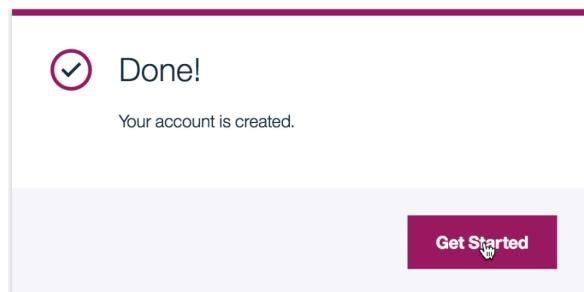
Select IBM Cloud account
Dummyuser23.ormium@spamgourmet.com ▾

IBM Cloud Organization
Dummyuser23.ormium@spamgourmet.com ▾

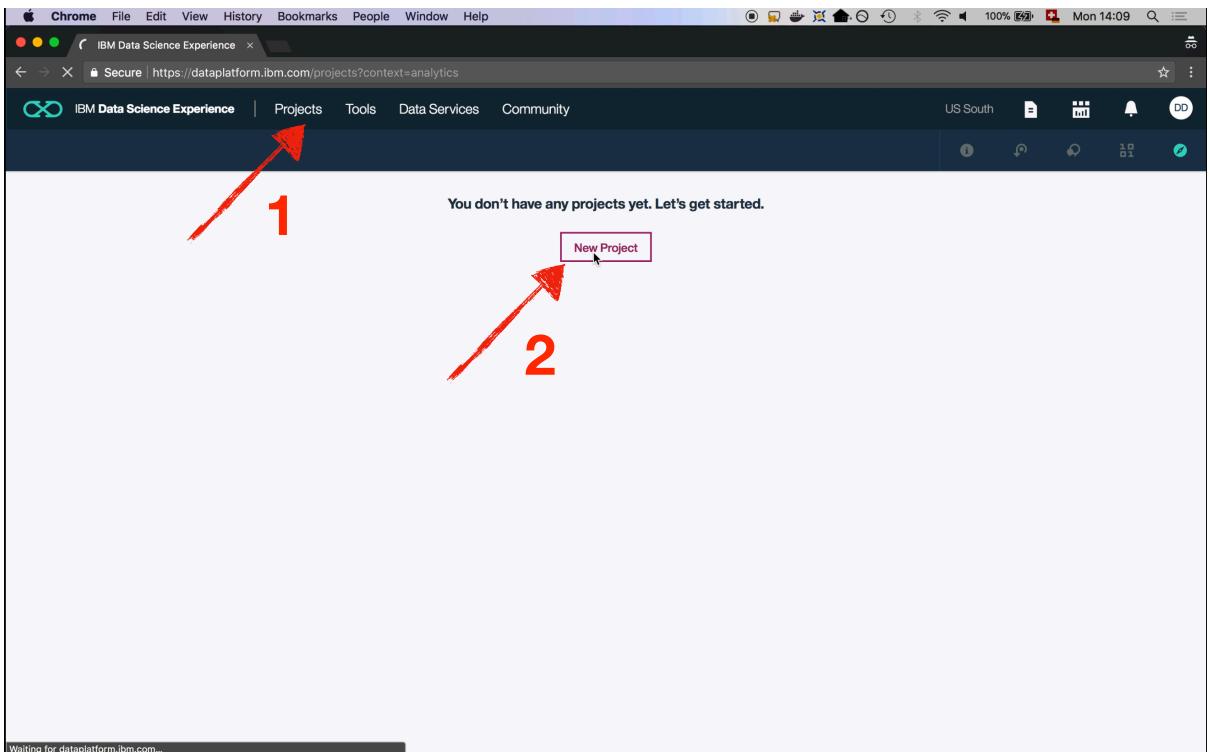
IBM Cloud Space
dev ▾

 Continue

3. Once the process is finished (it may take up some time) click on “Get Started”



4. In the DSX UI, create a new Project



- a. Click on “Projects”
- b. Click on “New Project”

5. Configure the project

New project

Define project details

Name
default

Description
Project description

Choose project options

Restrict who can be a collaborator

Define storage

① Select storage type

Object Storage (Swift API) IBM Cloud Object Storage

② Add

Add an object storage instance and then return to this page and click Refresh.

③ Refresh

Define compute engine

① Select Spark service

Spark service
Add

Add IBM Analytics for Apache Spark, then re Refresh.

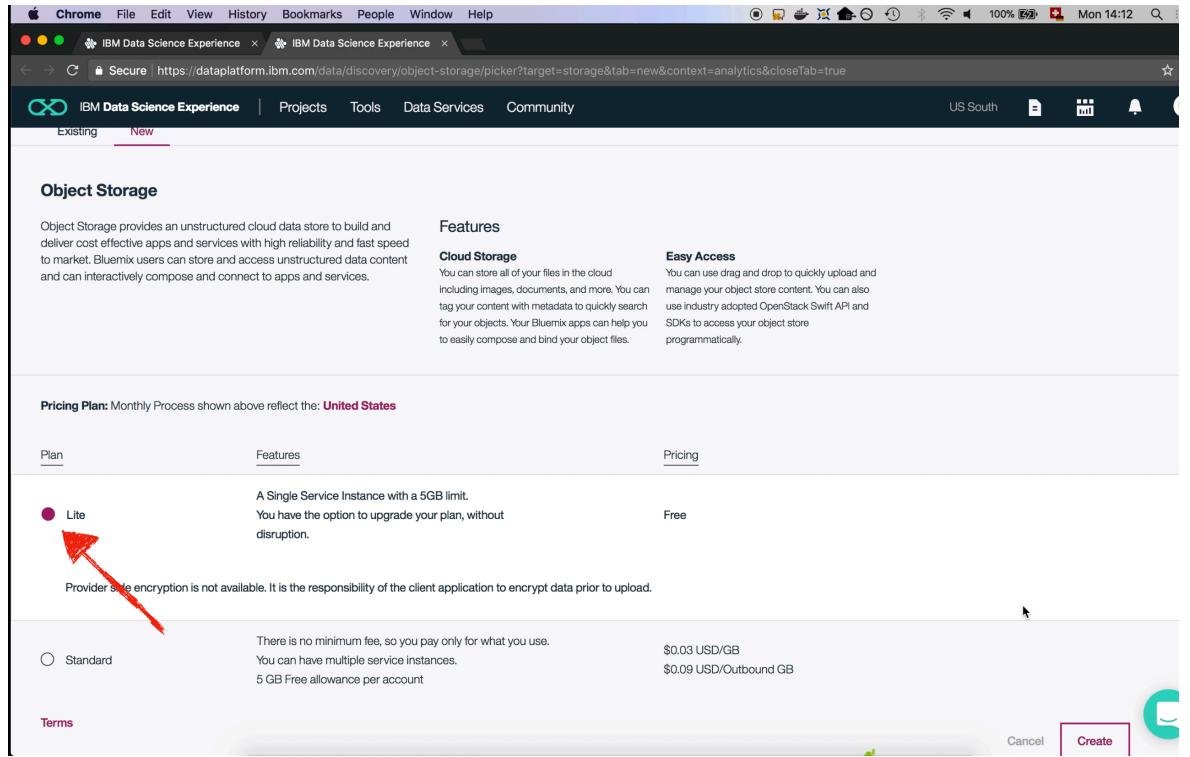
② Refresh

Armand from IBM
Hi
Dummyuser23.ormium@spamgourme...

Cancel Create

- a. Enter a name (1)
- b. Select “Object Storage (Swift API) (2)

c. Click on “Add” (3)



The screenshot shows the 'Object Storage' creation page in the IBM Data Science Experience. At the top, there are tabs for 'Existing' and 'New'. Below the tabs, the title 'Object Storage' is displayed. A brief description of Object Storage follows. To the right, there are two sections: 'Features' and 'Cloud Storage' (under 'Easy Access'). In the 'Features' section, there is a note about provider-side encryption. Below this, a table compares the 'Lite' and 'Standard' plans. The 'Lite' plan is selected, indicated by a red arrow pointing to its radio button. The 'Standard' plan is also listed with its details. At the bottom, there are 'Terms' and 'Create' buttons.

Plan	Features	Pricing
<input checked="" type="radio"/> Lite	A Single Service Instance with a 5GB limit. You have the option to upgrade your plan, without disruption. Provider side encryption is not available. It is the responsibility of the client application to encrypt data prior to upload.	Free
<input type="radio"/> Standard	There is no minimum fee, so you pay only for what you use. You can have multiple service instances. 5 GB Free allowance per account	\$0.03 USD/GB \$0.09 USD/Outbound GB

- i. Select “Lite”
- ii. Click on “Create”

iii. Make sure all fields are filled up and click on “Confirm”

Confirm Creation

Organization: Dummyuser231.ormium@spamgourmet.com

Plan
Lite

Space
dev

Service name
object-storage-kp

[Cancel](#) [Confirm !\[\]\(9233f7b99cc9fda31246240158e818f0_img.jpg\)](#)

d. Click on “Add” again

The screenshot shows the 'New project' page in the IBM Data Science Experience. The left panel contains 'Define project details' (Name: default, Description: Project description) and 'Choose project options' (checkbox for Restrict who can be a collaborator). The right panel contains 'Define storage' (Select storage type: Object Storage (Swift API) selected, IBM Cloud Object Storage available), '② Add' (button highlighted with a red arrow), '③ Refresh', and 'Define compute engine' (Select Spark service: Spark service selected, Add button highlighted with a red arrow). A tooltip for the 'Add' button in the compute engine section indicates: 'Add IBM Analytics for Apache Spark, then re Refresh.' A message from Armand from IBM is visible at the bottom right.

New project

Define project details

Name
default

Description
Project description

Choose project options

① Select storage type

Object Storage (Swift API) IBM Cloud Object Storage

② Add

Add an object storage instance and then return to this page and click Refresh.

③ Refresh

Define compute engine

① Select Spark service

Spark service

Add

Add IBM Analytics for Apache Spark, then re Refresh.

② Refresh

Armand from IBM

Hi

Dummyuser231.ormium@spamgourme...

Cancel Create

- e. Select “Lite” and click on “Create”

The screenshot shows the IBM Data Science Experience interface. At the top, there's a navigation bar with links for Chrome, File, Edit, View, History, Bookmarks, People, Window, Help, and a search bar. Below that is a secondary navigation bar with links for Projects, Tools, Data Services, and Community. On the right side, it shows "US South" and some icons. The main content area has tabs for "Existing" and "New", with "New" being selected. Under "New", there's a section for "Apache Spark". It includes a brief description of Apache Spark, three feature boxes ("Incredibly Fast", "Easy to Use and Powerful", "Convenient Data Storage"), and a "Pricing Plan" table. The table has columns for "Plan", "Features", and "Pricing". It shows one row for the "Lite" plan, which includes "2 Spark Executors" and is listed as "Free". A red arrow points to the "Lite" plan. At the bottom right of the table are "Cancel" and "Create" buttons, with the "Create" button highlighted by a red border.

Plan	Features	Pricing
Lite	2 Spark Executors	Free

Pricing Plan: Monthly Process shown above reflect the: [United States](#)

Features

Incredibly Fast
Apache Spark delivers 100x the performance of Apache Hadoop for certain workloads because of its advanced in-memory computing engine.

Easy to Use and Powerful
Apache Spark's Streaming and SQL programming models backed by MLlib and GraphX make it incredibly easy for developers and data scientists to build apps that exploit machine learning and graph analytics. Because the service is 100% compatible with Apache Spark, developers can build their apps and run them against the IBM managed service to benefit from operational, maintenance, and hardware excellence.

Convenient Data Storage
Object Storage enables a convenient way to upload your data from a file for immediate use by your Spark instance. You can set up Object Storage directly from the Spark service interface.

Pricing Plan: Monthly Process shown above reflect the: [United States](#)

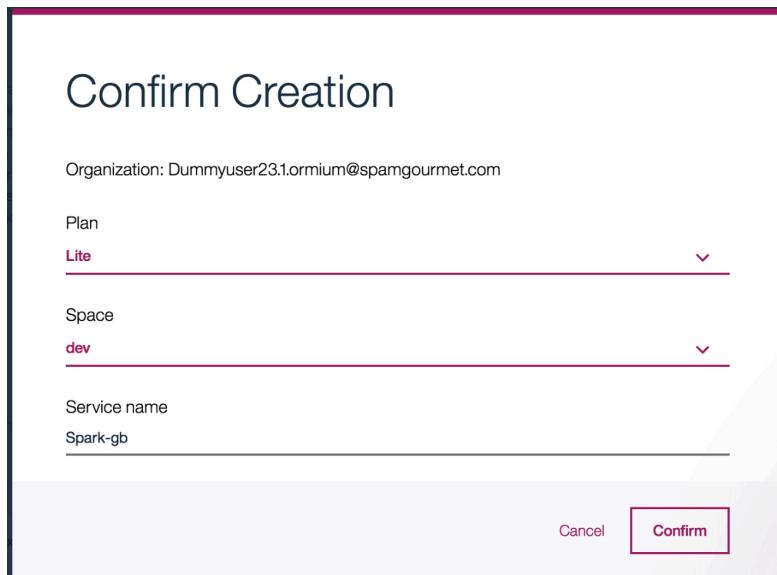
Plan	Features	Pricing
Lite	2 Spark Executors	Free

An entry level plan to run programs using up to 2 Spark executors

[Terms](#)

[Cancel](#) [Create](#) 

- f. Again, make sure all fields are filled and click on “Confirm”



6. Please make sure that an object store and Apache Spark service have been deployed.
You might have to click on “refresh”, then click on “Confirm”

The screenshot shows the 'Create Project' wizard in Watson Studio. It consists of three main sections:

- Define project details:** Includes fields for 'Name' (set to 'default') and 'Description' (with placeholder 'Project description').
- Define storage:** Shows 'Select storage type' (Object Storage (Swift API) is selected) and 'Target object storage instance' (set to 'object-storage-kp').
- Define compute engine:** Shows 'Select Spark service' (Spark service is set to 'Spark-gb') and a warning message: '⚠ If you associate the same Spark service with multiple projects, the Spark history server will display job history information for all the projects.'

At the bottom right, there are 'Cancel' and 'Create' buttons, with 'Create' being highlighted by a red box. Red arrows point from the 'Define storage' and 'Define compute engine' sections towards the 'Create' button.

7. Congratulations, your DSX setup is complete!