

Introduction to Data Wrangling with Jupyter Notebooks

IBM Developer

Upkar Lidder

<http://upkarlidder.com/talks/>

> ulidder@us.ibm.com

> @lidderupk

<http://bit.ly/01-22-2020-ibm>

<http://bit.ly/jupyter-pandas-git>

<https://slides.com/upkar/jupyter>

Prerequisites

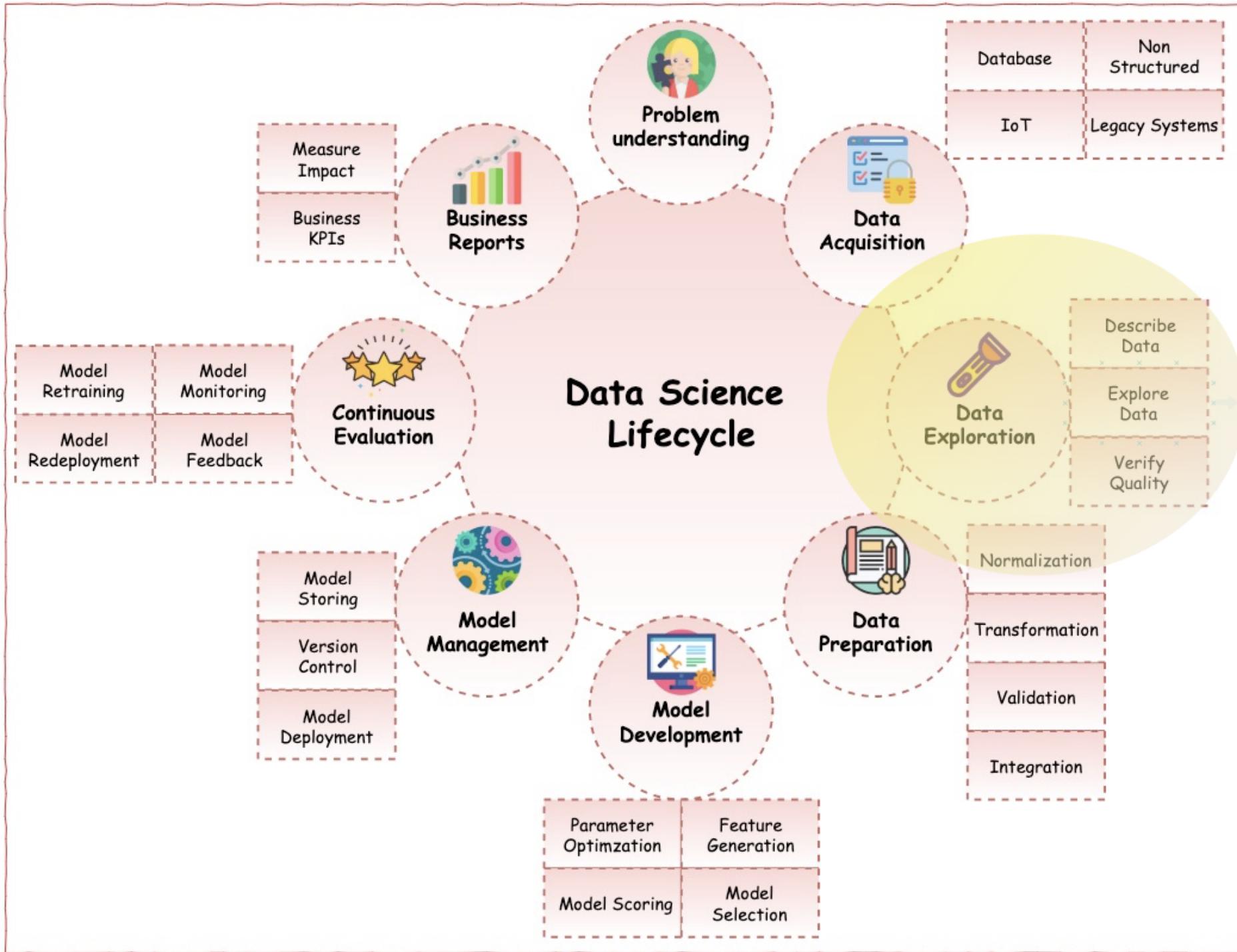
1. Create IBM Cloud Account using THIS URL

<http://bit.ly/01-22-2020-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/01-22-2020-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

Already have an IBM Cloud account? [Log in](#)

Create an IBM Cloud Account

Email* →

First Name*

Last Name*

Country or Region* United States

Password*

IBM may use my contact data to keep me informed of products, services and offerings:

by email. by telephone.

You can withdraw your marketing consent at any time by sending an email to netsupp@us.ibm.com. Also you may unsubscribe from receiving marketing emails by clicking the unsubscribe link in each such email.

More information on our processing can be found in the [IBM Privacy Statement](#). By submitting this form, I acknowledge that I have read and understand the IBM Privacy Statement.

[Create Account](#)

http://bit.ly/01-22-2020-ibm

Step 2 - locate Watson Studio in the catalog

The screenshot shows the IBM Cloud Catalog interface. At the top, there is a navigation bar with links for IBM Cloud, Catalog (which is highlighted), Docs, Support, Manage, and Upkar Liddar's Account. A notification bar at the top states: "Try the best of the Catalog for free with no time restrictions with Lite plans. The Lite filter is enabled. Remove the filter to see the full Catalog." On the left, a sidebar lists categories: All Categories (2), VPC Infrastructure, Compute, Containers, Networking, Storage, AI (1) (which is highlighted with a purple box), Analytics (1), Databases, Developer Tools, Integration, Internet of Things, Security and Identity, Starter Kits, and Web and Mobile. The main content area shows the results for the search term "label:lite watson studio". It includes a section titled "AI" which contains one item: "Watson Studio" (Lite • IBM • IAM-enabled). Below this is a section titled "Analytics" which contains one item: "Analytics Engine" (Lite • IBM • IAM-enabled). A "Filter" button is located in the top right of the search bar area. A "FEEDBACK" button is also visible on the right side of the page.

IBM Cloud

Catalog Docs Support Manage Upkar Liddar's Account

Try the best of the Catalog for free with no time restrictions with Lite plans.
The Lite filter is enabled. Remove the filter to see the full Catalog.

X

Catalog

label:lite watson studio

All Categories (2)

- VPC Infrastructure
- Compute
- Containers
- Networking
- Storage
- AI (1)
- Analytics (1)
- Databases
- Developer Tools
- Integration
- Internet of Things
- Security and Identity
- Starter Kits
- Web and Mobile

AI

Watson Studio
Lite • IBM • IAM-enabled
Embed AI and machine learning into your business.
Create custom models using your own data.

Analytics

Analytics Engine
Lite • IBM • IAM-enabled
Flexible framework to deploy Hadoop and Spark analytics applications.

Filter

FEEDBACK

Step 3 - create new watson studio service

IBM Cloud Catalog Docs Support Manage Upkar Lidder's Account 5

View all ←

Watson Studio

Lite • IBM

Watson Studio democratizes machine learning and deep learning to accelerate infusion of AI in your business to drive innovation. Watson Studio provides a suite of tools and a collaborative environment for data scientists, developers and domain experts.

[View Docs](#) [Terms](#)

AUTHOR IBM
PUBLISHED 07/18/2019
TYPE Service

Service name:

Choose a region/location to deploy in:

Select a resource group:

Tags:

Features

<input checked="" type="checkbox"/>	Lite	1 authorized user 50 capacity unit-hours monthly limit 1 free small compute environment with 1 vCPU and 4 GB RAM (does not require capacity unit-hours)	Free
-------------------------------------	------	---	------

The Lite plan for Watson Studio offers everything you need to become a better data scientist or domain expert in a collaborative environment.
Lite plan services are deleted after 30 days of inactivity.

you don't

- **Power on demand**

Enterprise-scale features on demand. From data exploration and preparation to

[Add to estimate](#) [Create](#)

FEEDBACK

Step 4 - launch Watson Studio

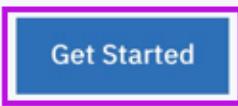
IBM Cloud Catalog Docs Support Manage Upkar Lidder's Account **5**

Resource list /
 Watson Studio-39
Resource group: Default Location: Dallas [Add Tags](#) 



Watson Studio

Welcome to Watson Studio. Let's get started!

 Get Started

Step 5 - create new project and pick empty template

[← Back](#)

Create a project

Choose whether to create an empty project or to preload your project with data and analytical assets. Add collaborators and data, and then choose the right tools to accomplish your goals. Add services as necessary.



Create an empty project

Add the data you want to prepare, analyze, or model. Choose tools based on how you want to work: write code, create a flow on a graphical canvas, or automatically build models.

NEW AutoAI experiment tool: Fully automated approach to building a classification or re...

USE TO

- Prepare and visualize data
- Analyze data in notebooks
- Train models



Create a project from a sample or file

Get started fast by loading existing assets. Choose a project file from your system, or choose a curated sample project.

USE TO

- Learn by example
- Build on existing work
- Run tutorials

Step 6a - name your project and create storage service

The screenshot shows the 'New project' creation interface in IBM Watson Studio. The top navigation bar includes 'IBM Watson Studio', 'Upgrade', 'Upkar Lidder's Account', and a user icon. The main area is titled 'New project'.

Define project details

Name: spectra-project (highlighted with a pink box)

Description: Project description (empty text area)

Choose project options

Restrict who can be a collaborator i

Project will include integration with [Cloud Object Storage](#) for storing project assets.

Define storage

① Select storage service: Add (highlighted with a pink box)
Add an object storage instance and then return to this page and click Refresh.

② Refresh

At the bottom right are 'Cancel' and 'Create' buttons.

Step 6b - add storage opens a new page

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio' on the left, an 'Upgrade' button, a notification bell, 'Upkar Lidder's Account', and a user icon on the right. Below the navigation bar, there's a message about reducing costs with lowest priced Archive storage.

In the main content area, there's a section titled 'Access and Key Management' which describes IAM policies and Key Protect support. Below this, a table lists the 'Pricing Plan: Monthly Process shown above reflect the: United States'.

PLAN	FEATURES	PRICING
<input checked="" type="radio"/> Lite	1 COS Service Instance Storage up to 25 GB/mo. Up to 20,000 GET requests/mo. Up to 2,000 PUT requests/mo. Up to Data Retrieval 10 GB/mo. Up to 5GB Public Outbound Applies to aggregate total across all storage bucket classes	Free
<input type="radio"/> Standard	There is no minimum fee, so you pay only for what you use.	Expand each section to view details

A note at the bottom of the table states: 'The Lite service plan for Cloud Object Storage includes Regional and Cross Regional resiliency, flexible data classes, and built in security.'

To the right, a modal window titled 'Confirm Creation' is displayed. It contains fields for 'Plan' (set to 'Lite'), 'Resource group' (set to 'Default'), and 'Service name' ('cloud-object-storage-na'). At the bottom of the modal, there are two buttons: 'Cancel' and 'Confirm'. The 'Confirm' button is highlighted with a pink box and a large number '2' is placed to its left.

At the bottom right of the main interface, there's a 'Create' button with a pink box around it and a large number '1' to its left. A blue speech bubble icon is also present.

At the very bottom left, it says 'IBM Developer' and at the bottom right, it says '@lidderupk'.

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

The screenshot shows the 'New project' interface in IBM Watson Studio. It consists of two main sections:

- Define project details**:
 - Name**: spectra-project
 - Description**: Project description
 - Choose project options**:
 - Restrict who can be a collaborator i
- Define storage**:
 - Select storage service**:
 - Add**: Add an object storage instance and then return to this page and click Refresh.
 - ② Refresh**: A purple box and arrow point to this button.
 - Storage**:
 - cloud-object-storage-na

At the bottom right of the interface, there are two buttons: **Cancel** and **Create**. The **Create** button is highlighted with a purple box.

Step 7 - add notebook feature to your project

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio', 'Upgrade', 'Upkar Lidder's Account', and a user icon. Below the navigation bar, the main header says 'My Projects / spectra-project'. The top menu has tabs for 'Overview', 'Assets' (which is highlighted with a purple box and labeled '1'), 'Environments', 'Jobs', 'Bookmarks', 'Deployments', 'Access Control', and 'Settings'. To the right of the menu, there are icons for 'Load', 'Files', and 'Catalog'. A central search bar asks 'What assets are you looking for?'. Below it, a section titled 'Data assets' is expanded. On the right side of the screen, there's a large modal window titled 'Choose asset type' with the heading 'AVAILABLE ASSET TYPES'. Inside the modal, there are several asset types listed in a grid:

- Data
- Connection
- Connected data
- AutoAI experiment NEW
- Notebook 3 (highlighted with a purple box)
- Dashboard
- Visual Recognition ...
- Natural Language Cl...
- Watson Machine Lea...
- Experiment
- Modeler flow
- Data Refinery flow
- Streams flow
- Synthesized neural n...

At the bottom of the modal, there's a 'Close' button. At the very bottom of the page, there's a call-to-action: 'Start analyzing data' and 'To create an analytic asset like a notebook or a model, click [Add to Project](#)'.

Step 8 - import notebook, get link from github

IBM Watson Studio

My Projects / spectra-project / Add Notebook

New notebook

Blank From file From URL

Name: spectra-notebook (25 characters remaining)

Description (optional): Type your Description here (500 characters remaining)

Select runtime: Default Python 3.6 XS (2 vCPU and 8 GB RAM)

The selected runtime has 2 vCPU and 8 GB RAM and consumes 1 capacity unit per hour.
[Learn more](#) about capacity unit hours and Watson Studio pricing plans.

Notebook URL: <https://github.com/lidderupk/lidderupk-ibmdevelopersf-jupyternotebooks/blob/main/asset/spectra-pandas.ipynb>

Create Notebook

<https://raw.githubusercontent.com/lidderupk/lidderupk-ibmdevelopersf-jupyternotebooks/master/asset/spectra-pandas.ipynb>

Step 9 - Let's look at data now!

The screenshot shows the IBM Watson Studio Jupyter Notebook interface. The top navigation bar includes 'IBM Watson Studio', 'Upgrade', 'Upkar Lidder's Account', and a user icon. The notebook path is 'My Projects / spectra-project / spectra-notebook'. The toolbar includes file operations like 'Upload' and 'Download', and various cell type and format selection buttons. The status bar indicates 'Not Trusted | Python 3.6'.

Some Python if you need a quick refresher

Let's look at some basic data structures in Python

- str (string)
- List (mutable, order maintained)
- Tuple (immutable)
- Set (unordered, unique)
- Dictionary (mutable, key-value pairs)

That's it! Let's focus on code now!

```
In [38]: #this is a string (str)
hackathon = "spectra"
print('length of the string:', len(hackathon))
print('first character of the string:', hackathon[0])
print('last character of the string:', hackathon[-1])
print('capitalize string: ', hackathon.capitalize())
print('upper case string: ', hackathon.upper())
print('find tra in the string: ', hackathon.find('tra')) #start counting from 0
# Many more methods available. Read the docs - https://docs.python.org/3/library/stdtypes.html#string-methods
```

```
length of the string: 7
first character of the string: s
last character of the string: a
capitalize string:  Spectra
upper case string:  SPECTRA
find tra in the string:  4
```

```
In [39]: #this is a list
states = ["Alabama", "Alaska", "Arizona", "Arkansas", "California", "Colorado", "Connecticut", "Delaware", "Florida"]
print('number of states:', len(states))
print('first state: ', states[0])
print('last state: ', states[-1])
```

<https://raw.githubusercontent.com/lidderupk/lidderupk-ibmdevelopersf-jupyter-notebooks/master/asset/spectra-pandas.ipynb>

Some links for the workshop

IBM Cloud account - <http://bit.ly/01-22-2020-ibm>

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

Datasets

Casts - <https://ibm.box.com/shared/static/nwvh08c06wtk48rsunk05qi9cruzkx4l.csv>

Release Dates - <https://ibm.box.com/shared/static/bm5kazwww0deze2a9r05l3mcqz5fgyvt.csv>

Titles - <https://ibm.box.com/shared/static/bvzl6da3t5vdav9xev4ype6sh78jhq7b.csv>

Film Locations - <https://ibm.box.com/shared/static/1kqel8vcya6r9wgxj4bvcsblp5xmyqsj.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



Build Smart
↳



Part of **IBM Developer** – 44 groups [?](#)

IBM Developer SF Bay Area

San Francisco, CA

8,451 members · Public group [?](#)

Organized by Angie K and 6 others

IBM Developer



Share: [f](#) [t](#) [in](#)

THU, JAN 30, 9:30 AM

Online Meetup: Deep Learning Master Class I - Introduction

Needs a location



Please do not forget to register on Crowdcast and join us using Chrome browser via Crowdcast on the event date! RSVP on Crowdcast NOW: <https://www.crowdcast.io/e/deep-learning-master-class-i-introduction>
This series takes engineers from zero to expert in machine learning in 5 sessions. The world...

85 attendees

Manage [▼](#)

Going

THU, FEB 6, 9:30 AM

Online Meetup: Deep Learning Master Class II - Computer Vision

Needs a location



Please do not forget to register on Crowdcast and join us using Chrome browser via Crowdcast on the event date! RSVP on Crowdcast NOW: <https://www.crowdcast.io/e/deep-learning-master-class-ii-computer-vision>
This series takes engineers from zero to expert in machine learning in 5 sessions. The world...

50 attendees

Manage [▼](#)

Attend

THU, FEB 13, 9:30 AM

DL Master Class III - Model Performance, Quantization, & Hyperparameter Search

Needs a location



Please do not forget to register on Crowdcast and join us using Chrome browser via Crowdcast on the event date! RSVP on Crowdcast NOW: <https://www.crowdcast.io/e/dl-master-class-iii-model-performance-quantization-hyperparameter-search>
This series takes engineers from zero to expert in machine learning in 5 sessions. The world needs...

36 attendees

Manage [▼](#)

Attend

IBM Partners

Enabling Independent Software Vendors (ISVs)
and tech companies for growth

Target audience

- ISVs and tech companies building and selling cloud solutions
- New to IBM Cloud
- Startups who aspire to build and sell their own solutions

Offers to help you get started



Build with up to \$12,000 of free IBM Cloud™ credits (\$1,000 per month for 12 months)

Integrate your solutions with leading-edge IBM Cloud technologies to deliver more innovation and value to your clients. Access more than **130 unparalleled services** including Watson™, Analytics and Security.



Build with 10TB of IBM Cloud Object Storage at no charge

Build data capability into your offering. IBM Cloud Object Storage is designed for high durability, resiliency and security.



Build with IBM Watson Assistant with a 1-year free trial

Receive access to 100K API calls per month plus 10 workspaces. Build and deploy chatbots quickly and efficiently with IBM Watson Assistant's advanced capabilities and seamless interface.

Get started

Experience IBM's countless partner benefits. Start building and selling with IBM today.

Learn more and access offers at ibm.com/partners/start



Build with IBM Cloud Kubernetes Service with a 1-year free trial

Containerize your solution with 1TB of block storage. Ship all your applications in one agile, well-defined structure with IBM Cloud Kubernetes Service.



Build with IBM Blockchain with a 6-month free trial

Build a network with up to 3 organizations to prototype. Build a secure business transaction network for your clients using blockchain and smart contracts.



Finished building and testing? Go-to-market with IBM

Access Provider Workbench, attend an orientation session and join the premier network of over 400 partners who are already listing their solutions on the IBM Marketplace.



Is your business a Startup? Build with up to \$120,000 in IBM Cloud credits

If your business revenue in the last 12 months is less than \$1M and you've been in business for fewer than five years, then you may qualify for Startup with IBM.

What is the IBM Insights Forum?

Use this link to join the community:

<https://qrgo.page.link/oNRFJ>



An online community for research with developers

- Explore product concepts, value propositions, messaging, and more
- Activities include forums, surveys, video diaries, etc.
- Invitations to 1-2 activities per month

What do you get?

- Research activities **relevant to your job as a developer**
- Learn from your peers**
- Stay on the **cutting edge of technology**
- Influence IBM products** and solutions
- Earn **monetary incentives** for participating

**All responses are confidential and for research only – your name and contact information are never shared with IBM sales or marketing*

Thank you

Let's chat !

Upkar Lidder, IBM

@lidderupk

<http://upkarlidder.com/talks/>

ulidder@us.ibm.com

#BehindtheCode



**It's about time
developers got
some attention.**

IBM
Developer

@lidderupk

IBM Developer