University of Rome «La Sapienza» Course of Business Intelligence - 2021

Course Lab 1

Mastering data analysis with IBM Watson Studio

Lab Speaker



Eleonora Picca

Data Scientist
IBM Technology Garage
picca.eleonora@gmail.com

Target & Scope

- Familiarize with a «real» software used in large enterprises
- Accomplish small but significant use cases in BI arena
- Understand the impact of tools over team productivity
- Introduce advanced topics like the use of "non structured" information

Lab sessions will be held on Thursday, starting from October 7, 2021 Session schedule: from 4 pm to 6 pm

Labs Schedule

Presentation of the tool and its basic features

Lab 1: Introduction to Watson Studio

Understanding and use of the tool features or conducting BI use cases

Lab 2: Working with Data

Lab 3: Analyze, Cleanse and Process

Lab 4: Report and Visualization

Use of the tool for Advanced Analytics

Lab 5: Machine Learning & AI (for dummies)

Lab 6: Content Analytics & Social Media

Lab 7: Putting all together

Today's topic is highlighted!

Reference Materials

IBM Watson Studio Documentation

https://dataplatform.cloud.ibm.com/docs/content/wsj/landings/wsl.html?context=analytics

Official product documentation, including both overview and tutorials to get started and detailed description of all product features

IBM Watson Studio Learning Center

https://www.youtube.com/playlist?list=PLzpeuWUENMK3u3j_hffhNZX3-Jkht3N6V

A YouTube collection, curated by IBM Developer Channel, that introduces viewers to the use of the product with simple examples (Note: it also contains advanced features for data scientist)

Today's Contents: Introduction to Watson Studio Cloud

- 1. From Data to Business Insights
- 2. What is Watson Studio?
- 3. Register to IBM Academic Initiative
- 4. Access Watson Studio on IBM Cloud
- 5. Watson Studio project interface

1. From DataTo Business Insights

The role of data into enterprise

Organizations need to use data collected and conveyed from several sources to:

- Describe and understand the status of business processes and phenomena
- Perform analysis of causes
- Got answers and actionable insight
- Make informed decisions

The three souls of Data Analytics

Data Analytics flavor	Description	Answers to questions like
Descriptive	Uses data mining to provide insight into the past	WHAT has happened?
Predictive	Uses statistical models and forecasts techniques to explain behaviors and understand the future	WHY did it happen? WHAT could happen tomorrow?
Prescriptive	Uses optimization and simulation algorithms to advice on possible outcomes	HOW should we do?

The rise of AI, ML & Cognitive Computing

- Artificial Intelligence (AI) is an umbrella term for a family of techniques that allow machines to learn from data and to act simulating human thought processes
 - Machine Learning (ML), Natural Language Processing (NLP), vision, and human-computer interaction

 We prefer to talk about augmented intelligence rather than "artificial intelligence", meaning that the idea is to enhance and scale human expertise rather than attempt to replicate human intelligence

Data + Al = Better Decisions

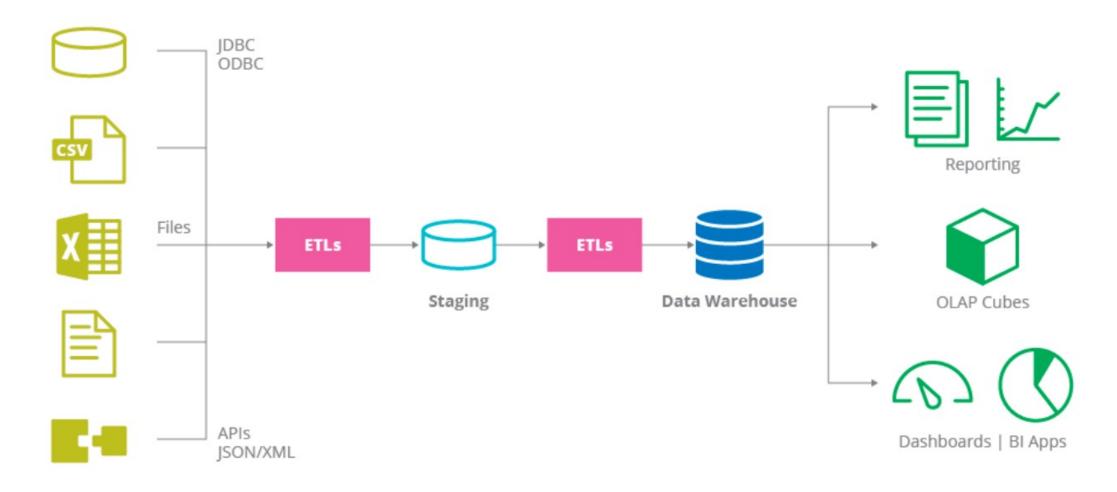
We expect from "intelligent" systems the possibility to have faster:

- Predictions
- Automations
- Optimization

In two word: democratize innovation



From traditional Business Intelligence schema...



Ingest Transform Analyze

...to modern Al-Powered Knowledge Management



Structured



Unstructured



Visual Recognition



Sentiment



Entities



Concepts



Language Translation



Element Classification



Document Understanding



And More...



Ingest

Enrich

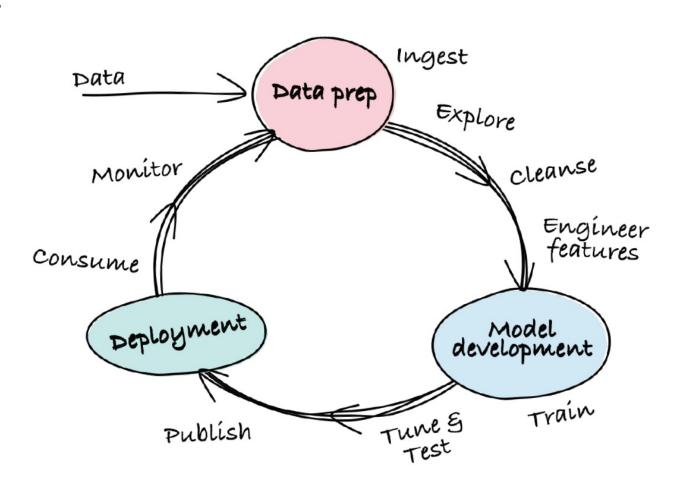
Ask

The need for a Data Architecture

Organizations needs a clear "data-driven" vision in order to adapt to new business requirements and platform technologies.

This vision should include:

- Data strategy,
- Data topology (classification, organization and flow),
- Data governance
- Data design

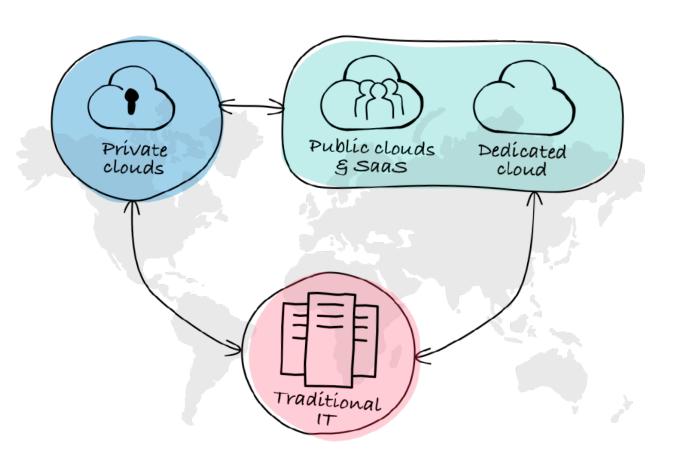


(Multi)Cloud-based models

Cloud computing is a consumption and delivery model inspired by consumer internet services and based on

- On-demand self-service
- Broad network access
- Location independent resources
- Rapid elasticity

Cloud models gives a broad degree of flexibility and choice, giving the best combination of repositories environments, and tools for data science, data engineering, and building applications.



Quick Recap

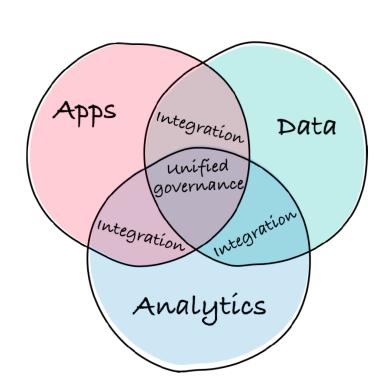
NEEDS

Collect and aggregate data

Analyze, Search and Discovery

Visualize analysis, patterns and trend

Use cognitive models



OUTCOMES

Answers from data

Actionable Insight

Informed Decision

Intelligent Processes

2. What is Watson Studio?

What is Watson Studio?

- Watson Studio is the data analysis service included in *IBM Cloud Pak for Data*, that provides users with the environment and tools to solve their business problems by collaboratively working with data.
- Users can choose the tools they need to analyze and visualize data, to cleanse and shape data, to ingest streaming data, or to create and train machine learning models.
- You can use Watson Studio to organize and prepare your data, create visualizations, quickly discover patterns and meaning and apply models— all on your own!

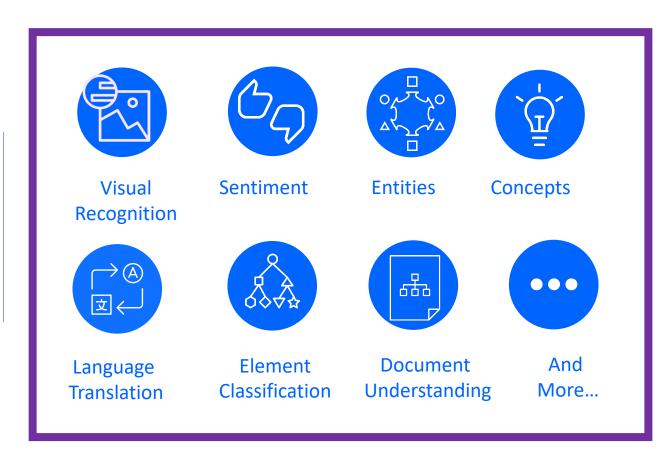
...Watson users can choose the tools they need...



Structured



Unstructured







Ingest

Enrich



... and Ask to use them!

Who uses Watson Studio?

Anyone who is interested in finding business insights and develop predictive models:

- Business analysts
- Data scientists
- Application developers
- Subject Matter Experts (SMEs)
- Students

Isn't Watson Studio too complicated for me?

- Watson Studio also includes capabilities to develop data models using neural networks and/or models developed with data scientists' specific tools (Python, R, Jupyter notebooks, etc.)
- Our Lab will leverage just features that are related with data analysis:
 - Data loading and refinement
 - Data visualization and exploration
 - Report creations
 - Use of built-in tools for non-structured data analysis (social media)
- But if you are interested, you are welcome to explore and use advanced features too!!

3. What kind of data may I use with Watson Studio?

Data asset constraints

Data assets (= data sets) can be loaded if they meet:

✓ size requirements.

✓ structure requirements

Size requirements

- The size of each file cannot be larger than 5 GB.
- You can use up to 50 assets in the Lite plan. Paid plans have larger limits
 - Standard plan: up to 500 assets
 - Enterprise plan: Unlimited assets

File structure requirements

- Watson Studio accepts several file formats, however for our scope we will consider these kinds of file:
 - Comma Separated Values files (.csv)
 - Excel files (.xls, .xslx)
 - JSON files (.json)
- Other data asset are not appropriate (even if they contains just tabular data!)
 - Word documents containing both text and tables
 - A table represented on a PDF document

General considerations on data structure

 For the sake of structured data analysis, you are supposed to load «rectangles» of data, i.e. data in tabular format.

- Is your data asset OK? Look at these fast facts
 - You can easily identify Rows and Columns
 - Each column contains data of uniform types
 - Each column has a (recommended) **title** (or header) that represents the description of meaning of data contained in column itself

Data assets vs Databases

- If you are familiar with relational databases (DBs), you may think of your data asset as a DB table
 - Rows -> DB Table Records
 - Columns -> DB Table Columns
- Column headers represents names of attributes of records
- Terminology mapping
 - header is equivalent to: attribute, field name, key in database terminology
- Your CSV data asset could come from the export of some operation (e.g., SELECT) from some operational databases!



DATA	VALUE			
Complete Green Shaded Fi	ields, Only			
PURCHASE PRICE	\$10,000.00			
PERCENT DOWN PAYMENT	7.00%			
TOTAL DOWN PAYMENT	\$700.00			
LOAN AMOUNT	\$9,300.00			
ANNUAL INTEREST RATE	4.50% 2 Monthly			
LENGTH OF LOAN IN YEARS				
PAYMENT FREQUENCY				
FIRST PAYMENT DATE	5/1/18			
MONTHLY INTEREST RATE	0.375%			
MONTHLY PAYMENT	\$405.92			
NUMBER OF PAYMENTS	24			
INTEREST TOTAL	\$442.21			

LOAN AMORTIZATION SCHEDULE									
PYMT DATE OF NUMBER PAYMENT	DATE OF	BEGINNING BALANCE	SCHEDULED PAYMENT	ADDITIONAL PAYMENT	INTEREST	PRINCIPAL	BALANCE \$9,300.00		
	PAYMENT				MIERESI	PRINCIPAL			
1	5/1/18	\$9,300.00	\$405.92		\$34.88	\$371.04	\$8,928.96		
2	6/1/18	\$8,928.96	\$405.92		\$33.48	\$372.44	\$8,556.52		
3	7/1/18	\$8,556.52	\$405.92		\$32.09	\$373.83	\$8,182.69		
4	8/1/18	\$8,182.69	\$405.92		\$30.69	\$375.23	\$7,807.46		
5	9/1/18	\$7,807.46	\$405.92		\$29.28	\$376.64	\$7,430.82		
6	10/1/18	\$7,430.82	\$405.92		\$27.87	\$378.05	\$7,052.77		



1	А	В	С	D	E	F	K	L	М
1	company_id	wave	id_scheda	ateco	inquadramento	occupation	cod_area_	area_fun	ISTAT4
2	13023555	ago	1	81	3	OPERAIO CAVATORE GENERICO	1		6112
3	13438606	ago	1	81		Supervisore cava	1		3151
4	13504100	ago	1	81	3	operaio generico	1		7441
5	13019612	ago	1	81	3	autista operaio	18	Altro	
6	13618690	ago	1	81		operaio	1		
7	14109819	ago	1	10	3	addetti al servizio di pulizia	18	Altro	8143
8	14109819	ago	2	10	2	impiegata amministyrativa	9		4112
9	13504744	ago	1	101		web marketing	5		3335
10	14216932	ago	1	101	2	IMPIEGATO AMMINISTRATIVO	5		4112
11	14216932	ago	3	101	2	ADDETTO RICERCA E SVILUPPO	2		3154
12	13504479	ago	1	101	2	ADDETTO LOGISTICA	12		4312
13	13504355	ago	1	101	3	operaio qualificato	1		7321
14	12923716	ago	1	101	3	Operaio Alimentare	1		7321
15	13024109	ago	1	101	3	AUTISTA MAGAZZINIERE	12		7423
16	13023978	ago	1	101	3	Operaio generico e autista	1		7329
17	13504582	ago	1	101	3	operaio addetto alla produzione	1		7321
18	13438667	ago	1	101	2	impiegato par-time	17		4112
19	13504479	ago	2	101	2	ADDETTO QUALITA'	3		3154
20	12923693	ago	1	101	3	Addetto confezionamento	1		7281
21	12923828	ago	1	101	3	operaio produzione	10		7281
22	13618977	ago	1	101	2	IMPIEGATO	9		3312



14109913	ago	1	103	3	operaio generico	18	Altro	8131
14217005	ago	1	103	3	Operaio qualificato	1		6513
13504796	ago	2	103	2	amministrativa	16		4112
13979211	ago	1	103	3	Manipolazione prod.aliment.	1		7329
14109920	ago	3	103	3	addetto alle pulizie	18	Altro	8143
14256873	ago	1	103	3	Raccoglitori	18	Altro	8311
13979227	ago	1	103	3	operaie generiche	1		7281
13310166	ago	1	103	3	operai stagionali	1		7329
13504796	ago	1	103	3	operaia generica	1		7281
13438703	ago	1	103	3	operaio generico	1		7211
14256860	ago	1	103	3	Addetto alla manipolazione in r	18	Altro	7281
13979227	ago	2	103	2	amministrativi	16		3312
14257020	ago	1	104	3	frantoiano-mulettista	1		7312
14110212	ago	1	104	3	OPERAIO STAGIONALE	18	Altro	8311
14110433	ago	1	104	3	SEGRETARIE	18	Altro	4111
14256953	ago	3	104	3	Autista trasporto olive e reflui	12		7423
14110282	ago	1	104	3	OPERAIO GENERICO	18	Altro	8431



APPENDIX J. PESTICIDE RESIDUES BY COMMODITY (Pairs With Residue Detections in at Least 10 Percent of Samples)

			% of	Number of	Number of	Range of	Mean of	EPA
		Pest.	Samples with	Samples	Samples with	Detections,	Detections,	Tolerance
Commodity / Pesticide		Type	Detections	Analyzed	Detections	ppm	ppm	ppm
1	Apples							
	Acetamiprid	1	33.1	744	246	0.002 - 0.11	0.021	1.0
	Azinphos methyl	1	17.9	744	133	0.005 - 0.73	0.041	1.5
	Boscalid	F	18.1	744	135	0.005 - 0.16	0.041	3.0
	Carbendazim (MBC)	F	19.9	534	106	0.002 - 0.19	0.033	7.0
	Chlorantraniliprole	1	28.8	534	154	0.003 - 0.095	0.01	0.30
	Diphenylamine (DPA)	F	73.7	744	548	0.005 - 5.3	0.35	10
	Endosulfan II	IM	12.7	739	94	0.007 - 0.19	0.023	1.0
	Fenpyroximate	A	13.1	534	70	0.002 - 0.12	0.015	0.40
	Formetanate hydrochloride	1	14.7	744	109	0.0002 - 0.026	0.003	0.50
	Imidacloprid	1	16.9	744	126	0.002 - 0.051	0.005	0.5
	Methoxyfenozide	1	12.5	744	93	0.002 - 0.11	0.008	1.5
	Phosmet	1	14.0	744	104	0.005 - 1.2	0.062	10
	Pyraclostrobin	F	16.0	744	119	0.002 - 0.11	0.02	1.5
	Pyrimethanil	F	57.4	744	427	0.002 - 6.0	0.48	14
	Tetrahydrophthalimide (THPI)	FM	15.9	744	118	0033-0.2	0929//	25.0
	Thiabendazole	F	75.4	744	561	0.002 - 5.9	-0:38 / -	5.0



3. Register to IBM Academic Initiative

--- IMPORTANT NOTES

 To complete all following steps, you MUST use the email address that has been provided to you by Sapienza University

xxxxxxxx@studenti.uniroma1.it

Otherwise, you will **NOT BE ABLE** to obtain 12-month product license

 When working in the Faculty Lab, you will need some credentials to access Internet. Ask faculty staff if you do not have them

Access IBM Academic Initiative site

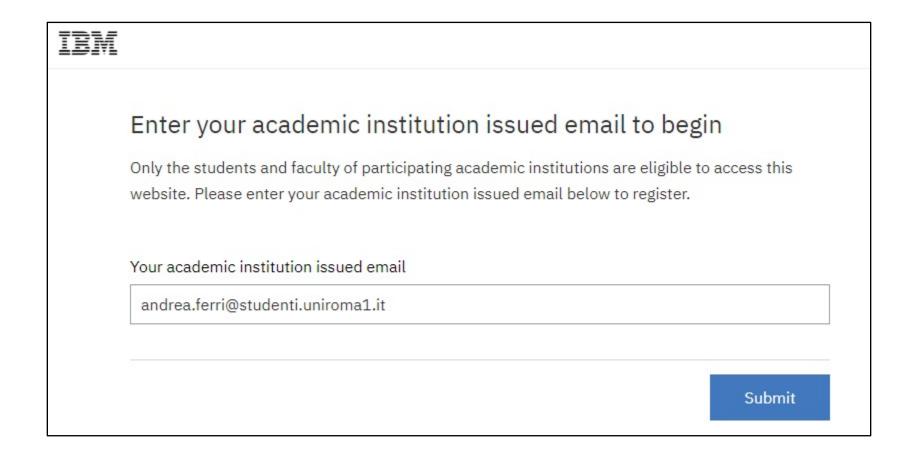
Use a Web browser to access IBM Academic Initiative site:

http://ibm.biz/academic

- If you have already registered, you can click **Log In** link in the right section of the page.
- Otherwise click Register Now button in the center of the page to create a new account
 - If a pop-up window about cookies is displayed, select one of the option to continue («default» or «required cookies» only). This should be requested only once.

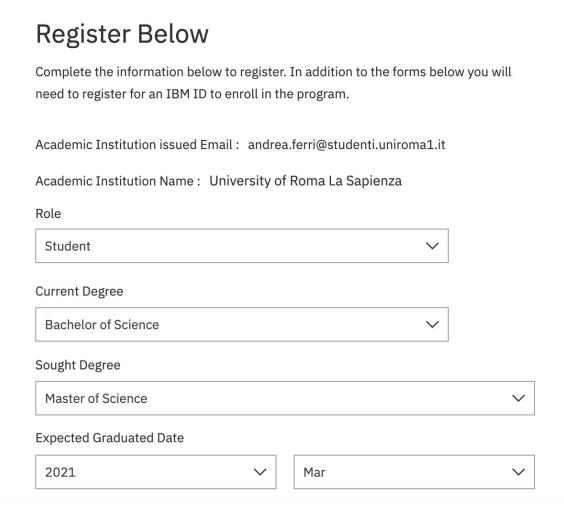
Registration procedure

In the next page, type your faculty-issued email address
 (xxxx@studenti.uniroma1.it) to certificate your eligibility to
 the program and click Submit



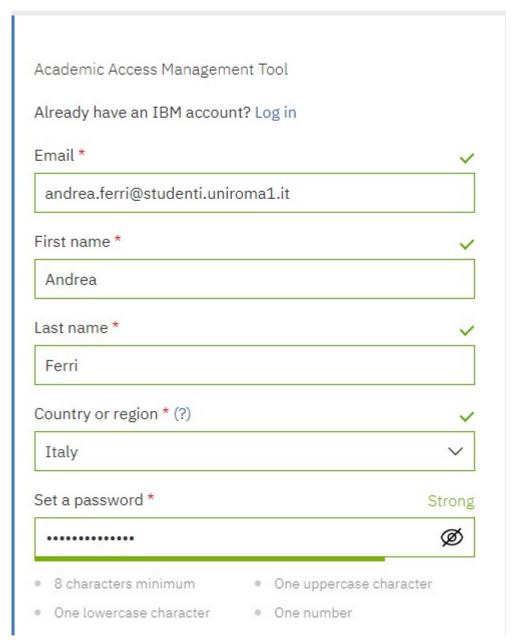
Registration procedure (cont'd)

- Fill out the form, read and accept Terms of Service/ Privacy statements.
- Click Register to complete the process and give confirmation on the subsequent screen



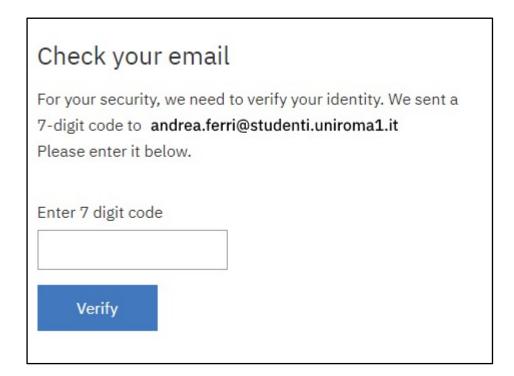
Create an IBM ID

- You will be redirected to the page for the creation of IBM ID. This represent your master «digital identity» through IBM Services
- If you already have an IBM ID, skip this page by clicking Log In link
- Insert all requested data and then click on **Continue** button



Finalize your IBM ID

- Read and understand IBMid
 Account Privacy information. Click
 on Proceed when you are done
- An email with a verification code will be sent by IBM to your faculty inbox. When you receive it, type the code in the subsequent page
- Now you can access your new IBM account with your email as userid and the password you have chosen



Advantages of Academic Initiative

- Watson Studio, like other IBM Cloud offerings, can be used by everyone with a free Lite license, that is enough to accomplish tasks of small/medium complexity. All that you need is your IBM ID to access IBM Cloud site.
- However, being affiliate to Academic Initiative gives you the opportunity to access plenty of IBM Resources, like free courses, seminars and badge programs.
- Moreover, you are entitled to download and use for non-commercial purposes IBM analytics software like SPSS, Cognos Analytics and so on.

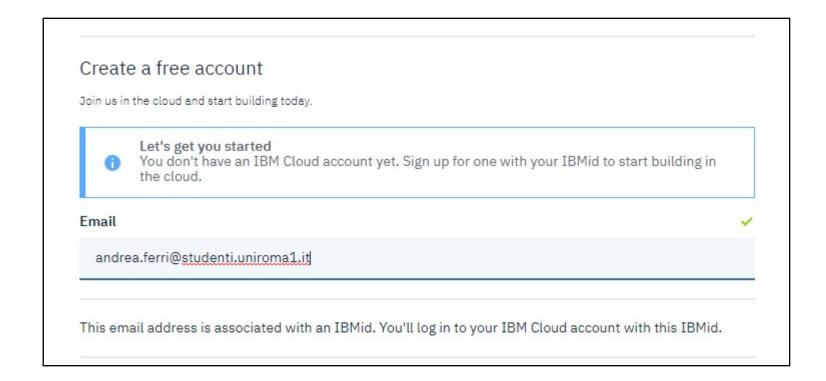
4. Access Watson Studio on IBM Cloud

Access IBM Cloud

- Login to the IBM Cloud https://cloud.ibm.com using your IBM ID credentials (faculty email and password you set)
- Only for the first time, you will be asked to create an IBM Cloud Account. This is the login you will use to access Watson Studio and IBM Cloud services, and will be linked with your IBM ID
- The system should recognize that you already have an IBM ID and will use its data for the creation of the cloud account

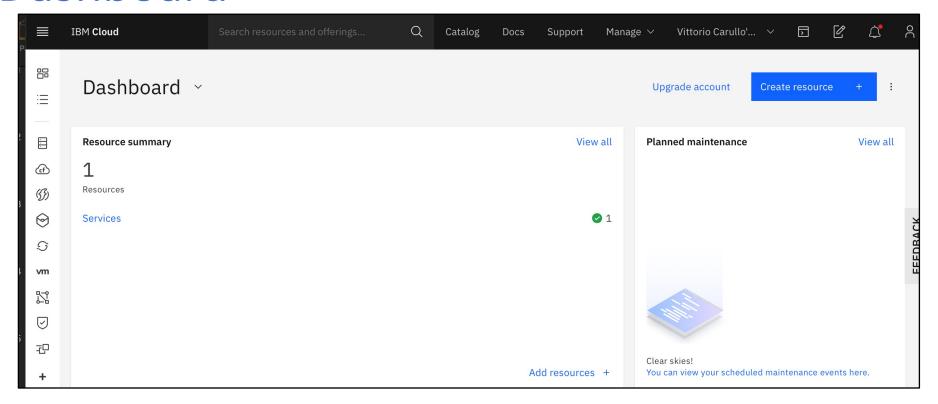
Access IBM Cloud

•



 After clicking on creation button, you will receive an email Click on Confirm Account to validate your identity and enable access to IBM Cloud. At this time you are ready to log in into Cloud home page

Dashboard

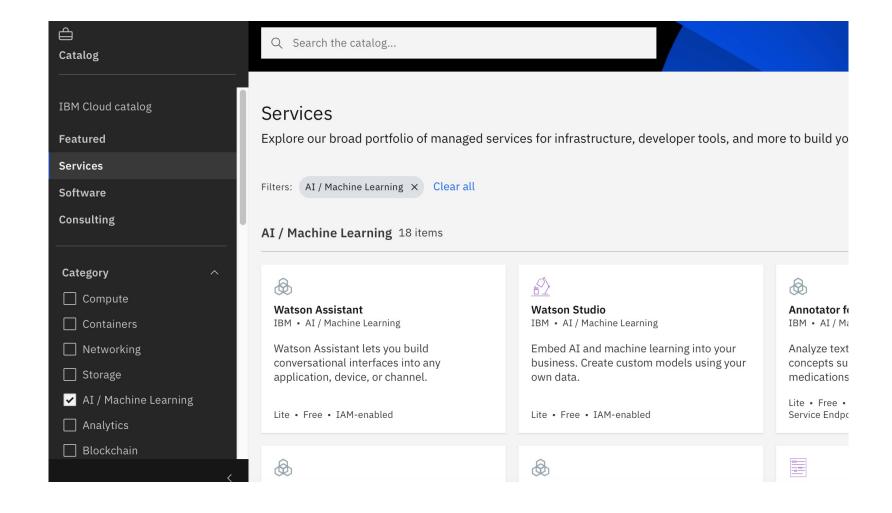


• The **Dashboard** is your home page on IBM Cloud. Here you will find all services and apps you create. From here you will also be able to manage users' access, view status, find documentation and so on

Create a Watson Studio instance/1

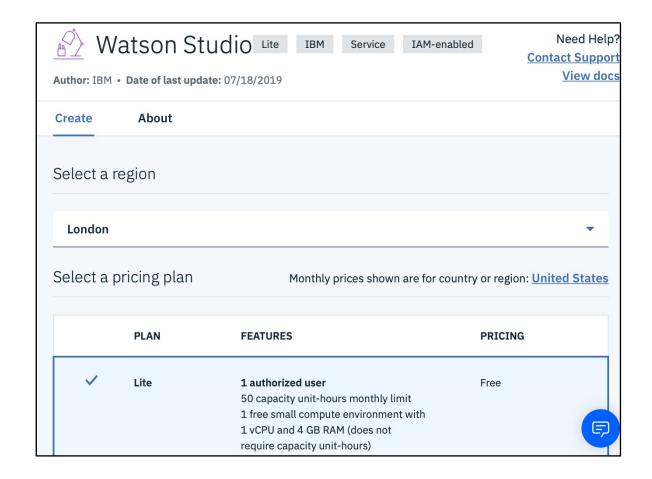
Click on the
 Create Resource
 button to enter
 into Catalog page

 Click on Services and choose AI category in the left section. AIrelated services will be shown



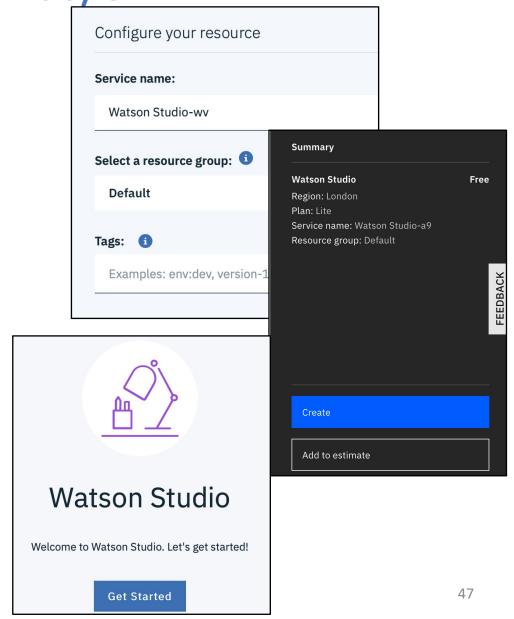
Create a Watson Studio instance/2

- Click on Watson Studio box to open the service creation page
- Look at the proposed options. You can change (if you want) region/location, but ensure that in the Select a Pricing Plan section the Lite plan is selected (should be the first one)

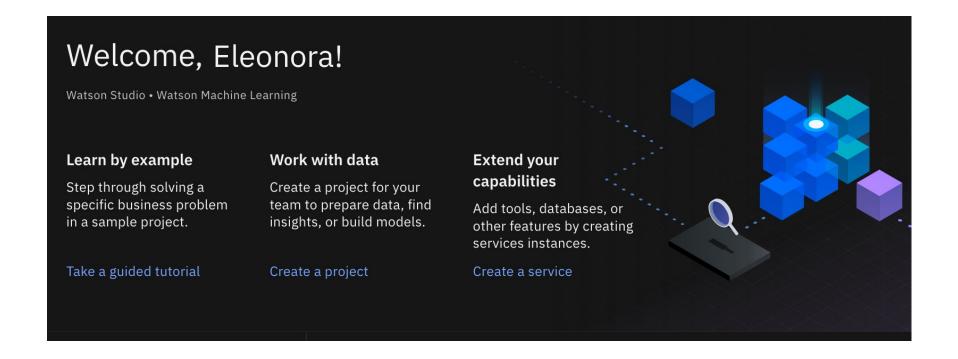


Create a Watson Studio instance/3

- Scroll down the window and change (optionally) the service name. Leave unchanged other values
- Click on Create to start the deploy of the service instance on your account (it should take some seconds)
- On the subsequent Welcome page, click on **Get Started** to access your Studio environment.

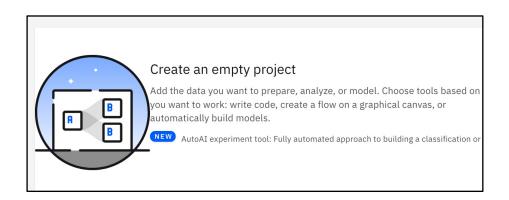


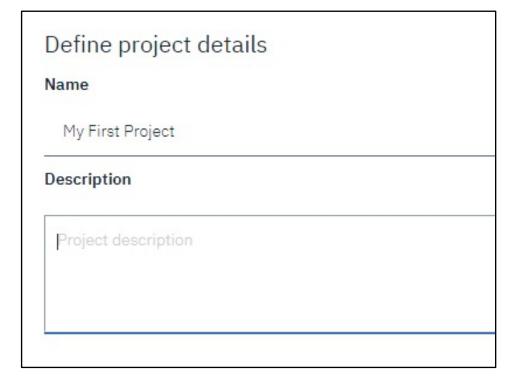
Welcome!



- Feel free to take confidence with Welcome page by following on-screen tips that are displayed (just for the first time) or by taking the guided tour suggested at the end of the tips
- When you are ready, click on Create a Project to start your work

Create your first project/1





- From Create Project wizard, click on Create an empty project
- Type a name and an (optional) description for your project
- Scroll down the page to visualize information about Cloud Object Storage

Create your first project/2

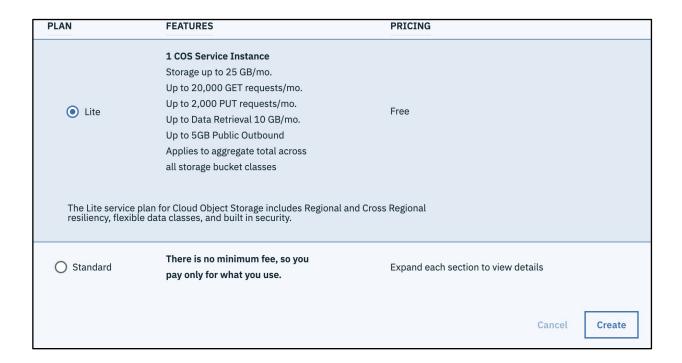
Define storage

1 Select storage service

Add

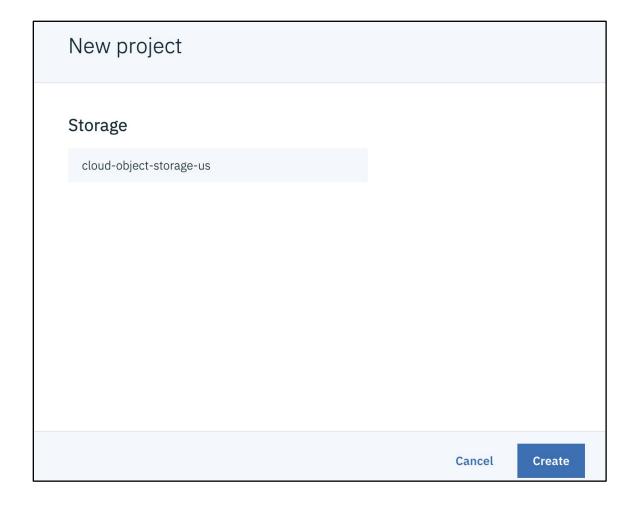
Add an object storage instance and then return to this page and click Refresh.

(2) Refresh



- Under Define Storage, click Add to add a storage (Cloud Object Storage, COS)
- In the following screen, review proposed storage options: scrolling down the page, ensure that the option selected for Plan is Lite
- Click Create button at the end of the page

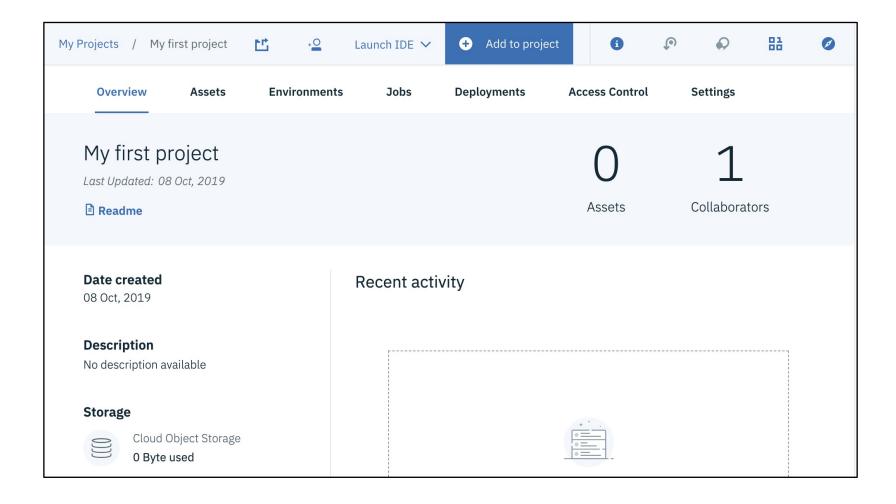
Create your first project/3



- In the confirmation dialog leave all values unchanged and click
 Confirm
- Back to the project creation page, click
 Refresh to see the newly created object storage
- Click Create to finalize the creation of the project
 - Note: The Object Storage instance creation is required only for the first time

5. Watson StudioProject interface

Project Dashboard



What we will use

- Watson Studio include tools for
 - **Prepare data**: cleanse, shape, visualize, organize, and validate data.
 - Analyze data: identify patterns and relationships in data, and display insights.
 - **Build models**: build, train, test, and deploy models to classify data, make predictions, or optimize decisions.
- During these lessons we will focus on first two aspects, data preparation and analysis, so we will use mainly Overview and Asset sections
- Access Control section can be used to let other users access your projects (useful for co-operate on the same work)
- Other sections (Environment, Jobs and Deployment) will not be relevant part of our scope

Create project: Hands On

 Create a new Watson Studio project called «My first project» or something similar

 Take some time to explore the project interface; follow guided tours if you like