

# Prerequisites

If you want to run all exercises "live" and complete this workshop effectively, there are some pieces of software that you will need to install on your computer. Remember that it is not madatory at all, but you will be limited to a normal reading if you decide not to install them.

## 1.- bash

You will need to execute commands using your terminal and they use the conventions of the [bash shell](https://www.gnu.org/software/bash/). All Linux distributions and MacOS include bash or a compatible command-line interface. On Windows, you can try to setup the [Windows-Subsystem for Linux (WSL)](https://learn.microsoft.com/en-us/windows/wsl/) on your computer. If it is not possible to install bash, you will have to figure out an alternative way of issuing the commands you will see during all the modules. If you are not familiar with bash, it is recommendable to visit online resources to get a basic knowledege. [Here](https://www.coursera.org/projects/coding-for-beginners-an-easy-introduction) you can see one of the many examples you could find.

## 2.- Jupyter Notebooks and Python

This workshop has been developed with [Jupyter Notebooks](https://jupyter-notebook.readthedocs.io/en/latest/) technology in order to maximize the user experience. While it is not strictly necessary, you should [install](https://docs.jupyter.org/en/latest/install/notebook-classic.html) it on your local computer. Otherwise, you will have to copy and paste the bash commands and execute them in a terminal.

There are parts of the workshop that have been developed with [python](https://www.python.org/) which is a requirement of several technologies we will cover. There is no needed to be a python programmer but a basic knowledge is highly recommendable. Please review the introduction of any tutorial you may find in internet. The course mentioned [above](https://www.coursera.org/projects/coding-for-beginners-an-easy-introduction) may also be a good start.

## 3.- bash kernel for Jupyter

Jupyter is well-know for python development but it is also possible to extend its default kernel to include bash capabilities. Once you have installed Jupyter, you will need to run a very simple command in your bash terminal:

pip install bash\_kernel  
python -m bash\_kernel.install

Notice that instaling Jupiter Notebooks implies an underlying python environment, as indicated during the [installation instructions](https://docs.jupyter.org/en/latest/install/notebook-classic.html) mentioned above.

By the way: there are several variations of bash kernels for jupyter. Take a look at the [web site](https://github.com/jupyter/jupyter/wiki/Jupyter-kernels) where all possible kernels are listed, search for bash and pick one that may fit better for you.

After the installation of the bash kernel, you need to activate it just one time if you want to click on the cells to run the commands. You can do it this way if you are using Microsoft Visual Studio Code:



Alternatively, if you use the native jupyter notebook tool:



## 4.- Docker

We will run some operations with containers like tagging or transferring them to a different repository. In order to do that we will need [docker](https://www.docker.com/) or [podman](https://podman.io/) as an alternative. A very convenient way of getting docker is downloading [docker desktop](https://www.docker.com/products/docker-desktop/). Alternatively, you may want to install docker with the command-line. In that case, you may want to run the following commands on RHEL8:

First of all, get a clean python environment:

sudo yum install curl wget git  
sudo yum install gcc openssl-devel bzip2-devel libffi-devel zlib-devel  
sudo yum install python3.8  
sudo yum install python38-devel

Then download, install and setup docker:

sudo yum install -y yum-utils  
sudo yum-config-manager --add-repo https://download.docker.com/linux/rhel/docker-ce.repo  
sudo sed -i 's~/rhel/~/centos/~g' /etc/yum.repos.d/docker-ce.repo  
sudo yum --noplugins install docker-ce docker-ce-cli containerd.io docker-compose-plugin  
  
sudo systemctl enable --now docker  
sudo systemctl start docker  
  
sudo usermod -aG docker $USER  
newgrp docker  
  
sudo pip3 install docker-compose

Don't forget to start docker from the user interface if you installed docker desktop:



Verify that docker is running by displaying the version

docker version

In case you are not familiar with Docker, [here](https://www.coursera.org/learn/ibm-containers-docker-kubernetes-openshift#syllabus) is a good introductory course that will help you to begin your journey into the world of containers and OpenShift.

## 5.- oc and helm

The [OpenShift command-line interface (oc)](https://docs.openshift.com/container-platform/4.10/cli_reference/openshift_cli/getting-started-cli.html) and the [helm](https://helm.sh/) utility are also neccesary to run some commands. An alternative to oc is [kubectl](https://kubernetes.io/docs/tasks/tools/) but this worshop has not been tested with it.

It may be more convenient to defer the installation of these utilities after having provisioned the hardware because the cluster contains direct links to the downloads you need, but you may want to do it now. In that case, you will need a RedHat account and follow these [instructions](https://docs.openshift.com/container-platform/4.10/cli_reference/openshift_cli/getting-started-cli.html).

The download that you need to choose will be similar to this:



Then, just follow the installation procedure descibed [here](https://helm.sh/docs/intro/install/) to get helm

## 6.- Databand images

There are several media packages of databand and we need the helm chart version of the software for this workshop. This section describes the steps that an IBMer would follow to download the right package. Additionally, this is the point where I need you to encourage to adhere to the terms and conditions of a licensed software like databand. In simple words: do not distribute these images illegally or use them for other purposes other than your own education during this workshop.

* IBMers need to be connected to the IBM intranet or use the Cisco Secure Client in order to access the [IBM Internal DSW Downloads](https://w3south-limited-use.cpc.ibm.com/software/xl/download/ticket.wss).
* Read the authorized and not autorized use of the software you will download. If you agree, you can go on
* Search for databand like in the following picture:



* Look for the helm charts version and download it:



After pressing the blue button, the download will start and you will get a file similar to databand-1.0.19-helm-chart.tar.gz (960MB). Needless to say, names and sizes will change with upcoming versions.

## 7.- DataStage

If you want to excercise with the integration of DataStage, you will need to create an instance in the IBM Cloud. The lite (free) plan is perfectly fine for this workshop.



## 8.- Airflow

It is highly recommendable, but not strictly necessary, to master the basic concepts of [Apache Airflow](https://airflow.apache.org/). There are plenty of tutorials in internet and visiting any introduction would be enough to follow the workshop. [Here](https://www.coursera.org/lecture/etl-and-data-pipelines-shell-airflow-kafka/apache-airflow-overview-Uyfwg) is a good example.

## 9. OpenShift

You need to provision an OpenShift cluster as a part of the hands-on workshop. If you are not familiar with OpenShift, [this](https://learn.ibm.com/course/view.php?id=9394) may be your starting point but, please, review the next chapter [Hardware Provisioning](./1_provisioning.ipynb) for a full description of this pre-requisite.

Next Section: [Hardware Provisioning](./1_provisioning.ipynb)

[Return to main](../README.md)