# Getting Stated – Big Data Infrastructure Setup

## System Requirements

* 8 GB RAM
* X86 Processor
* 50 GB Free Disk Space

To check your system resources:

### On Windows:

1. Click on the Start menu
2. Type About your PC and select the option
3. In the Device specifications section, check Installed RAM and Processor

### On Mac (Intel):

1. Click on the Apple menu on the top left
2. Select About This Mac
3. Here, you can see the Memory and Processor details under Overview.

To check your disk space:

### On Windows:

1. Click on the "File Explorer" icon in the taskbar or press the Windows key + E.
2. Look for your local disk drive (usually C:). The available disk space will be displayed next to the drive icon.
3. Ensure that there is at least 50 GB of free space available on your local drive.

### On Mac (Intel):

1. Click on the Apple icon in the top-left corner of the screen.
2. Click on "About This Mac" from the dropdown menu.
3. In the "Overview" tab, click on the "Storage" tab to view the available disk space on your Mac.
4. Ensure that there is at least 50 GB of free space available on your Mac's local drive.

**If you do not have enough resources (at least 8 GB RAM and 50 GB Free Disk) or if you are running Macs with M1 or M2 processors (i.e., ARM processors), please follow the supplemental instructions for setting up the environment on Google Cloud.**

## Download Pre-Configure OVA

Download the pre-configured virtual image from the course website.

Installation of VirtualBox

Please install VirtualBox on your respective systems. This is a free and open-source hosted hypervisor for x86 virtualization.

### On Windows:

1. Visit the official [VirtualBox website](https://www.virtualbox.org/wiki/Downloads)
2. Download the Windows host by clicking Windows hosts
3. Open the downloaded file and follow the instructions to install

### On Mac (Intel):

1. Visit the official [VirtualBox website](https://www.virtualbox.org/wiki/Downloads)
2. Download the OS X host by clicking OS X hosts
3. Open the downloaded file and follow the instructions to install

## Importing the Virtual Machine

1. Open VirtualBox
2. Click on File -> Import Appliance
3. Navigate to the provided .ova file and click Open
4. Follow the instructions to import

## Starting the Virtual Machine

1. Click on the imported virtual machine in VirtualBox
2. Click Start
3. Login with the username: bigdata and password: bellevue

## Downloading and Running the Setup Script

1. Open the terminal from the side bar

Go to the Documents directory

cd Documents

1. Download the git repository for the class.

git clone https://github.com/bellevue-university/dsc650-infra.git

1. Change into the dsc650-infra directory.

cd dsc650-infra

1. Change the script’s permissions to make it executable:

* chmod +x setup.sh

1. Run the script:

* sudo ./setup.sh
* This will install Docker and Docker Compose, and clone the Bellevue Big Data repository.

## Running the Big Data Software

1. Type cd bellevue-bigdata and hit Enter
2. You should now see several directories: hadoop-hive-spark-hbase, kafka, nifi, and solr. Each contains a docker-compose.yml file except for nifi which contains the software binaries.

***Do not attempt to run docker-compose up -d for all directories at once.***

Follow these steps for the **hadoop-hive-spark-hbase** and **solr** directores:

1. Change into the directory with cd <directory-name>, replacing <directory-name> with the name of the directory.
2. Type *docker-compose up -d* and hit Enter. This will start up the software in that directory.
3. Verify that everything is healthy using *docker ps*.
4. Open Firefox from the left toolbar and navigate to the bookmarked user interface for each software component.
5. For NiFi, please look at the section below titled: Logging into the NiFi Interface.

***You can watch the google cloud video starting with part 2 for assistance with this part. You will not need to do port forwarding as everything is running locally on your virtual machine. The Web Browsers can be accessed through FireFox.***

Follow these steps for the **nifi** directory**:**

1. Change into the nifi directory with cd nifi
2. Start NiFi using the command:

/bin/bash nifi-\*/bin/nifi.sh start

1. Navigate to the user interface for each software component with the instructions provided in the next section, Logging into the NiFi User Interface
2. Stop NiFi using the command:

/bin/bash nifi-\*/bin/nifi.sh stop

### Logging into the NiFi User Interface

1. Your browser may show a warning about the website’s security certificate. This is expected because we are using a self-signed certificate for the NiFi instance. To proceed, click on “Advanced” and then “Accept the Risk and Continue” (the wording may vary depending on your browser).
2. To log in, you will need a username and password. These are generated when the NiFi instance is started and can be found in the instance’s logs.
3. On your VM terminal, go into the nifi directory and run:

* grep Generated nifi-\*/logs/\*

1. Look for the username and password in the output. They will be inside square brackets. For example:

**Generated** Username […]

**Generated** Password […]

1. Use these credentials to log in to the NiFi user interface.

## Shutting Down

1. Ensure that all Docker containers are turned off with docker-compose down for each directory.
2. Ensure NiFi is stopped.
3. You can now shut down your Ubuntu instance:

* Click on the power icon in the top-right corner of the screen.
* A drop-down menu will appear with several options.
* Click on "Power Off" or "Shut Down" (the exact wording may vary depending on the version of Ubuntu you're using).
* Ubuntu will initiate the shutdown process, close all applications, and power off the machine.