# 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](https://bellevueuniversity-my.sharepoint.com/:u:/g/personal/nismaily_bellevue_edu/EZf7ziGo5R1MjzNelh75mYgBMvxv-Lge5eAO-XAIdfA2vA?e=UVwEbh) the pre-configured virtual image.

Password: Bellevue1!

## 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

## Running the Big Data Software

1. Open the terminal from the side bar
2. Type cd Documents/bellevue-bigdata and hit Enter
3. You should now see several directories: hadoop-hive-spark-hbase, kafka, nifi, and solr. Each contains a docker-compose.yml file

Follow these steps for each directory:

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. Once you’ve verified that the user interfaces are working correctly, you can shut down the Docker containers for that directory with docker-compose down
6. Return to the parent directory with cd .. and move on to the next directory

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

## Additional Steps for hadoop-hive-spark-hbase

In the hadoop-hive-spark-hbase directory, after running docker-compose up -d, execute the following:

1. docker-compose exec master jps
2. docker-compose exec worker1 jps
3. docker-compose exec worker2 jps

These commands help verify the working status of your application. For other directories, verifying through the user interface alone is sufficient.

## Shutting Down

1. Ensure that all Docker containers are turned off with docker-compose down for each directory
2. You can now shut down your virtual machine in VirtualBox by selecting Machine -> ACPI Shutdown

These instructions assume a basic familiarity with the command line interface. If you’re a first-time user, don’t hesitate to ask for help. It’s part of the learning process!