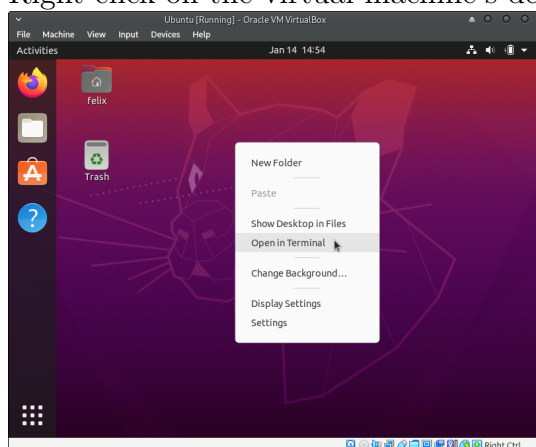


This step-by-step guide shows you how to set up everything required to work on Bonus Project 2 on a Linux system like a Ubuntu virtual machine. If you do not have access to a Linux system, please first refer to the guide on how to set up a Ubuntu virtual machine and, after completion, come back to this one. Note that we cannot support other operating systems like Windows or Mac OS. Also note that setting up Spark on Windows or Mac OS can be quite cumbersome, so we strongly recommend you using a Linux system or virtual machine.

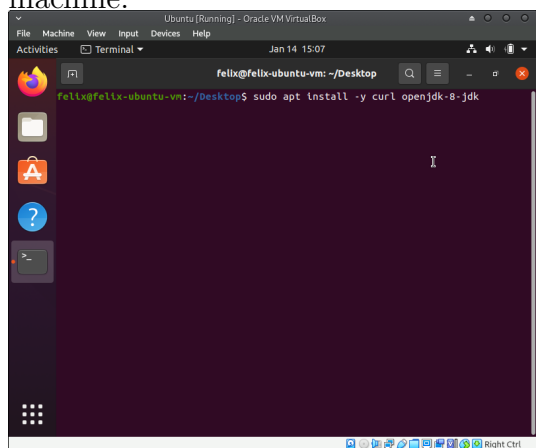
Right-click on the virtual machine's desktop and open a terminal:



To install the Curl and the Java 8 JDK, type in the following command and press **ENTER**:

```
sudo apt install -y curl openjdk-8-jdk
```

You will be prompted to enter the password you have set when installing the virtual machine.

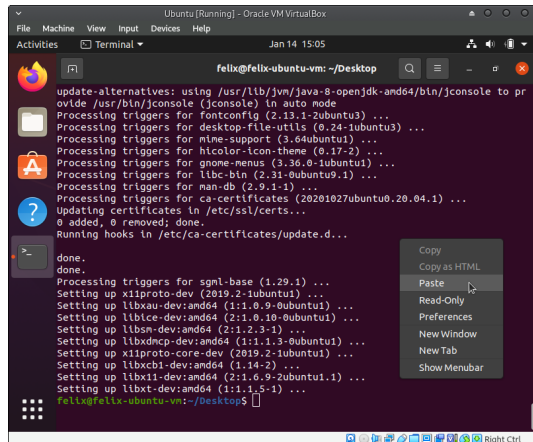


Next, to install Sbt, go to this website, copy the commands listed there, paste them **one-by-one** into the terminal, and press ENTER each time:

[https://www.scala-sbt.org/release/docs/Installing-sbt-on-Linux.html#](https://www.scala-sbt.org/release/docs/Installing-sbt-on-Linux.html#Ubuntu+and+other+Debian-based+distributions)

Ubuntu+and+other+Debian-based+distributions

Note that CTRL+V does not work in the terminal. Instead, right click and select “Paste”:



Next, we install Spark. To do this, first download the latest version starting with 2. (not the one starting with 3.; package type should be Pre-built for Apache Hadoop) from here:

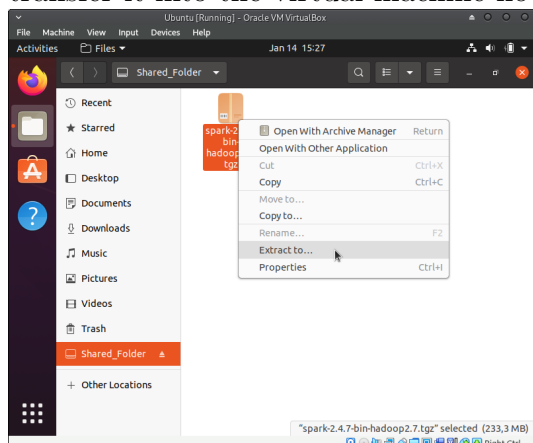
<https://spark.apache.org/downloads.html>:

Download Apache Spark™

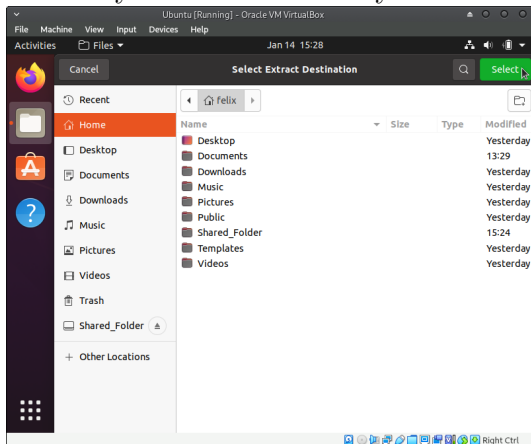
1. Choose a Spark release: 2.4.7 (Sep 12 2020) ▾
2. Choose a package type: Pre-built for Apache Hadoop 2.7 ▾
3. Download Spark: [spark-2.4.7-bin-hadoop2.7.tgz](#)
4. Verify this release using the 2.4.7 [signatures](#), [checksums](#) and [project release KEYS](#).

Note that, Spark 2.x is pre-built with Scala 2.11 except version 2.4.2, which is pre-built with Scala 2.12. Spark 3.0+ is pre-built with Scala 2.12.

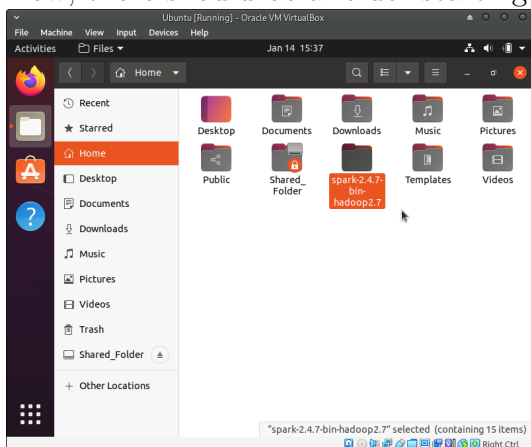
If you have downloaded the file on your local computer and not in the virtual machine, transfer it into the virtual machine now, e.g., via the shared folder. Then unpack it:



Choose your home directory as the extract destination:

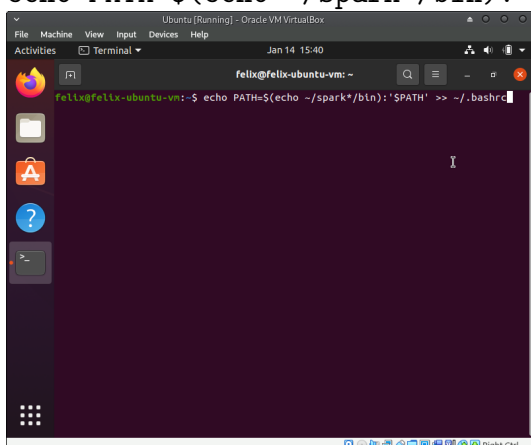


Now, there should be a folder starting with spark in your home directory:



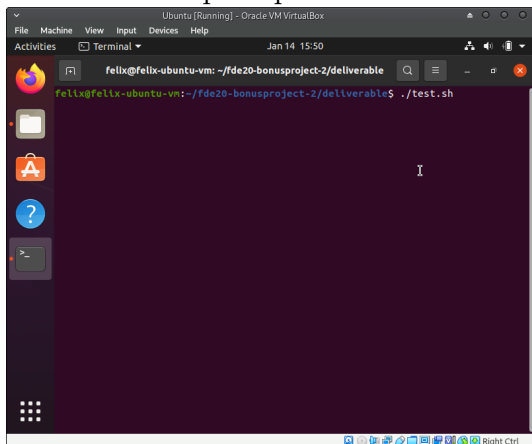
Next, open a terminal again and execute the following command, which makes Spark universally available on the system:

```
echo PATH=$(echo ~/spark*/bin):'$PATH' >> ~/.bashrc
```

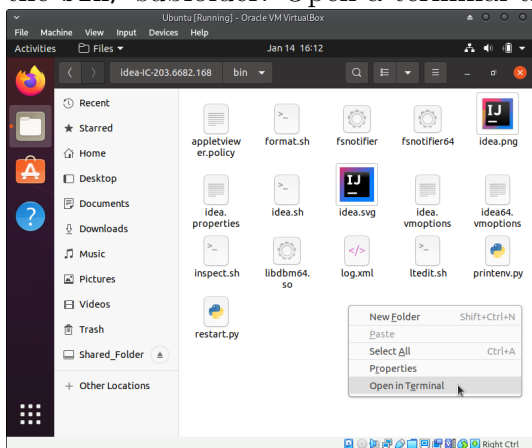


Now, use Git to clone your fork of the bonus project repository into some folder on the virtual machine. Navigate into the bonus project folder, then into the `deliverable/` subfolder and open a terminal there. Execute the command `./test.sh`, which compiles your code and tests whether it works. This will of course initially fail, as you haven't edited your code yet. You can now start working on the project and later use this command to test whether your code works.

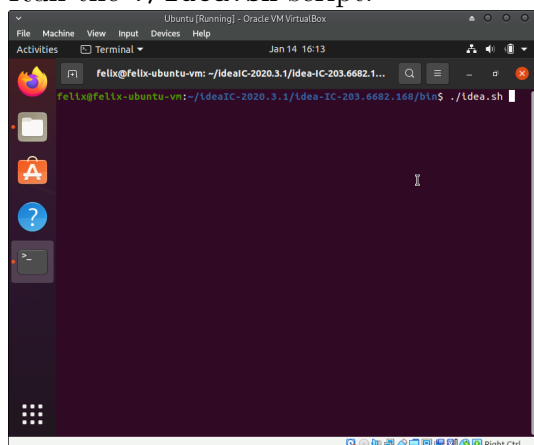
Note that there are a couple of gotchas you have to keep in mind. The command only works if (a) the path to the `test.sh` file does not contain any spaces and (b) you have opened the terminal in the `deliverable/` folder, which is indicated by the blue part of the command prompt.



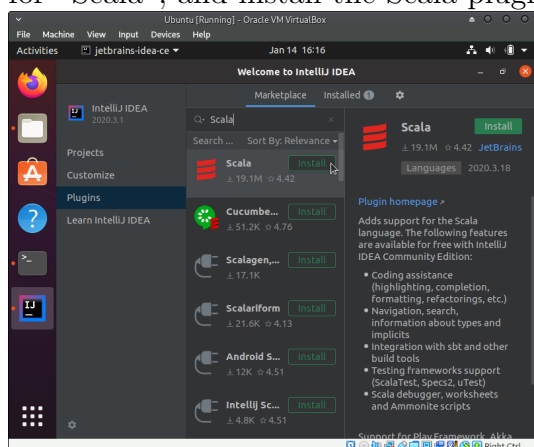
As a bonus, you might want to install IntelliJ IDEA Community Edition to ease coding. To this end, first download the current version for Linux from here: <https://www.jetbrains.com/idea/download/#section=linux>. Then, just like with Spark, extract the downloaded file into your home directory. Navigate to the extracted folder and into the `bin/` subfolder. Open a terminal there:



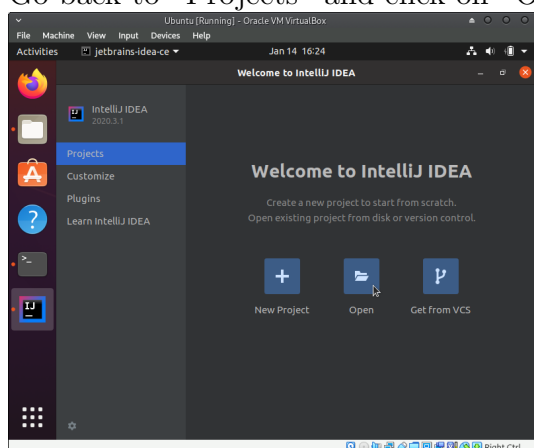
Run the `./idea.sh` script:



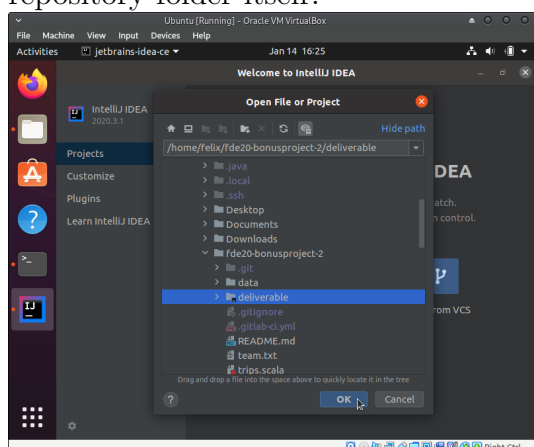
There may be initial setup prompts. If asked, do not import settings from anywhere. Once you see the welcome screen, click on “Plugins”, then click on “Marketplace”, search for “Scala”, and install the Scala plugin:



Go back to “Projects” and click on “Open”. Do not click on “New Project”!



Select the **deliverable/** subfolder in your cloned repository folder. Do not select the repository folder itself!



Now, the project will be imported, which may take some time. Afterwards, you can use IntelliJ to edit the code. To test whether it works, you still need to use the `test.sh` script as shown above.

To make launching IntelliJ easier in the future, click on “Tools” → “Create Desktop Entry”. Once you have done this, you can launch IntelliJ by clicking on “Activities” in the upper left corner and searching for “IntelliJ”.

