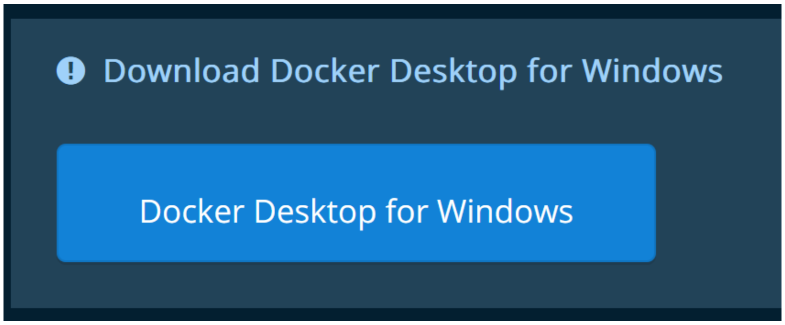
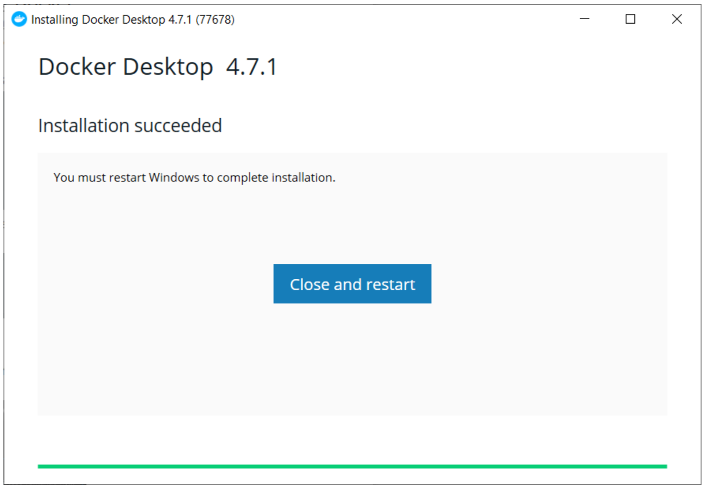
How to install and set up PyLucene using Docker

1. Install Docker Desktop (from Install Docker Desktop on Windows)



Download Docker Desktop for Windows

2. Once installed, you will get a message to restart your computer.



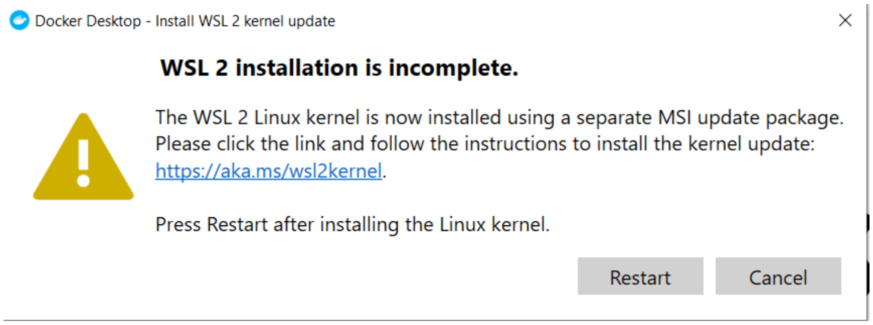
Installing Docker Desktop 4.7.1 (77678)

Docker Desktop 4.7.1

Installation succeeded

You must restart Windows to complete installation.

3. After restart, if not already installed, Docker will ask you to install WSL2. After the installation of WSL2, you can click on “Restart”.



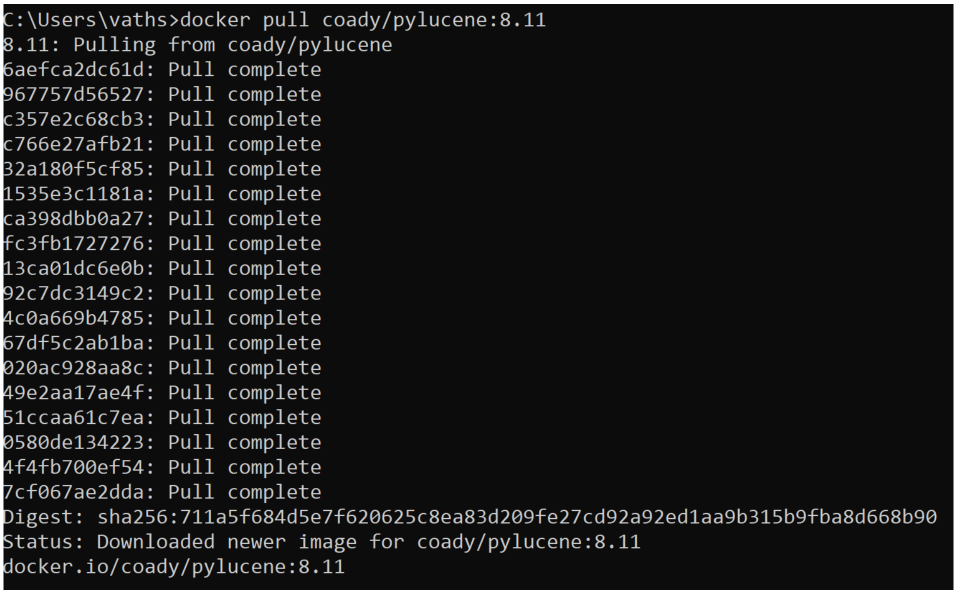
Docker Desktop - Install WSL 2 kernel update

WSL 2 installation is incomplete.

The WSL 2 Linux kernel is now installed using a separate MSI update package. Please click the link and follow the instructions to install the kernel update: https://aka.ms/wsl2kernel.

Press Restart after installing the Linux kernel.

4. After restart, open the command prompt and run: `docker pull coady/pylucene:8.11`
   After installing all required packages, and successfully pulling, you will see “`Status: Downloaded newer image for coady/pylucene:8.11`”



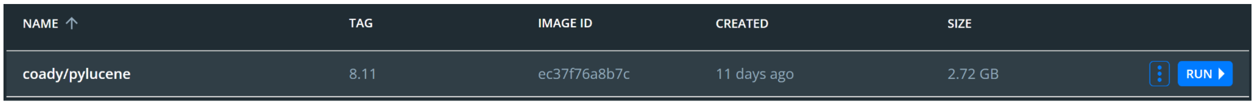
```
C:\Users\vaths>docker pull coady/pylucene:8.11
8.11: Pulling from coady/pylucene
6aefca2dc61d: Pull complete
967757d56527: Pull complete
c357e2c68cb3: Pull complete
c766e27afb21: Pull complete
32a180f5cf85: Pull complete
1535e3c1181a: Pull complete
ca398dbb0a27: Pull complete
fc3fb1727276: Pull complete
13ca01dc6e0b: Pull complete
92c7dc3149c2: Pull complete
4c0a669b4785: Pull complete
67df5c2ab1ba: Pull complete
020ac928aa8c: Pull complete
49e2aa17ae4f: Pull complete
51ccaa61c7ea: Pull complete
0580de134223: Pull complete
4f4fb700ef54: Pull complete
7cf067ae2dda: Pull complete
Digest: sha256:711a5f684d5e7f620625c8ea83d209fe27cd92a92ed1aa9b315b9fba8d668b90
Status: Downloaded newer image for coady/pylucene:8.11
docker.io/coady/pylucene:8.11
```

Dockerfile for the pull command can be found here:

Dockerfile

```
ARG VERSION=latest
FROM python:$VERSION

RUN apt-get update \
    && apt-get install -y default-jdk ant

WORKDIR /usr/lib/jvm/default-java/jre/lib
RUN ln -s ../../lib amd64

WORKDIR /usr/src/pylucene
RUN curl https://downloads.apache.org/lucene/pylucene/pylucene-8.11.0-src.tar.gz
\
    | tar -xz --strip-components=1
RUN cd jcc \
    && NO_SHARED=1 JCC_JDK=/usr/lib/jvm/default-java python setup.py install
RUN make all install JCC='python -m jcc' ANT=ant PYTHON=python NUM_FILES=8

WORKDIR /usr/src
RUN rm -rf pylucene
```

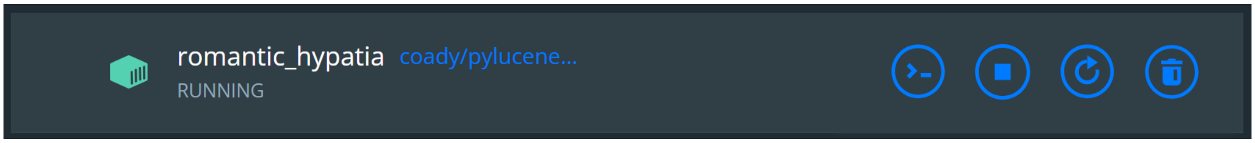
5. Once completed, open Docker Desktop. Go to “Images” and click on the blue “Run” button.



NAME ↑	TAG	IMAGE ID	CREATED	SIZE	
coady/pylucene	8.11	ec37f76a8b7c	11 days ago	2.72 GB	⋮ RUN ▸

When you click on run, the Container for coady/pylucene is created.

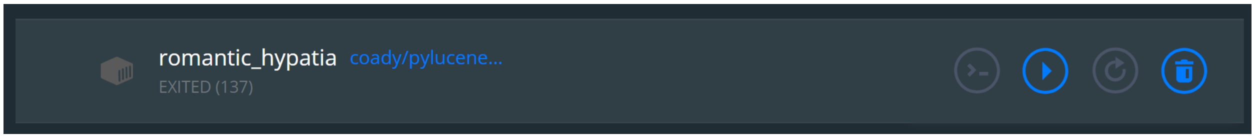
6. Go to the “Container” tab, and you will see that the container is already running.



romantic_hypatia coady/pylucene...

RUNNING

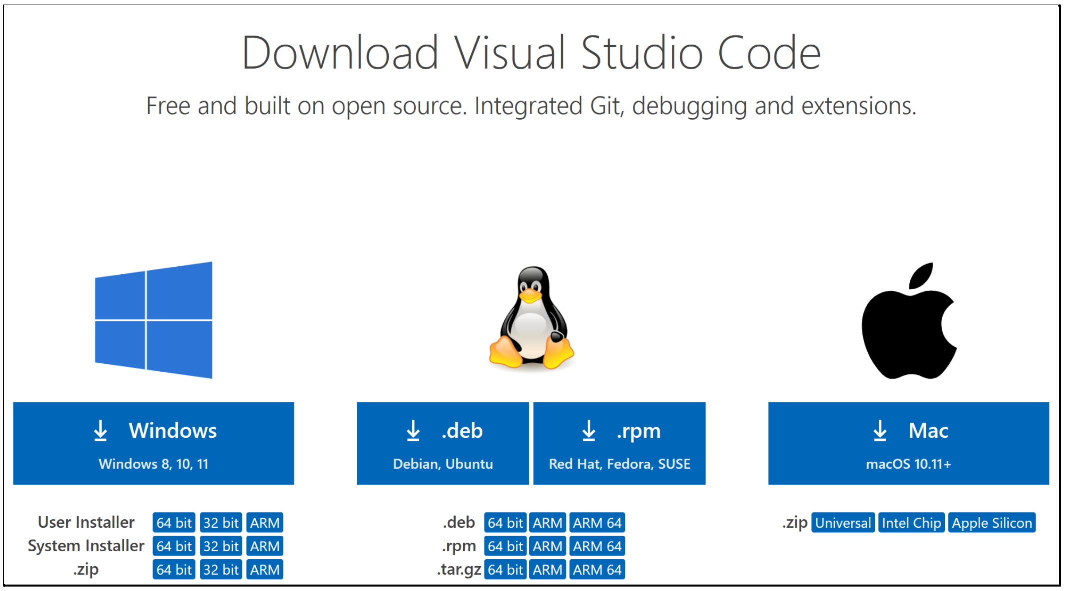
From the second time onwards, you will have to click on “start”



romantic_hypatia coady/pylucene...

EXITED (137)

7. Install Visual Studio Code (VS Code) (from Download Visual Studio Code - Mac, Linux, Windows).



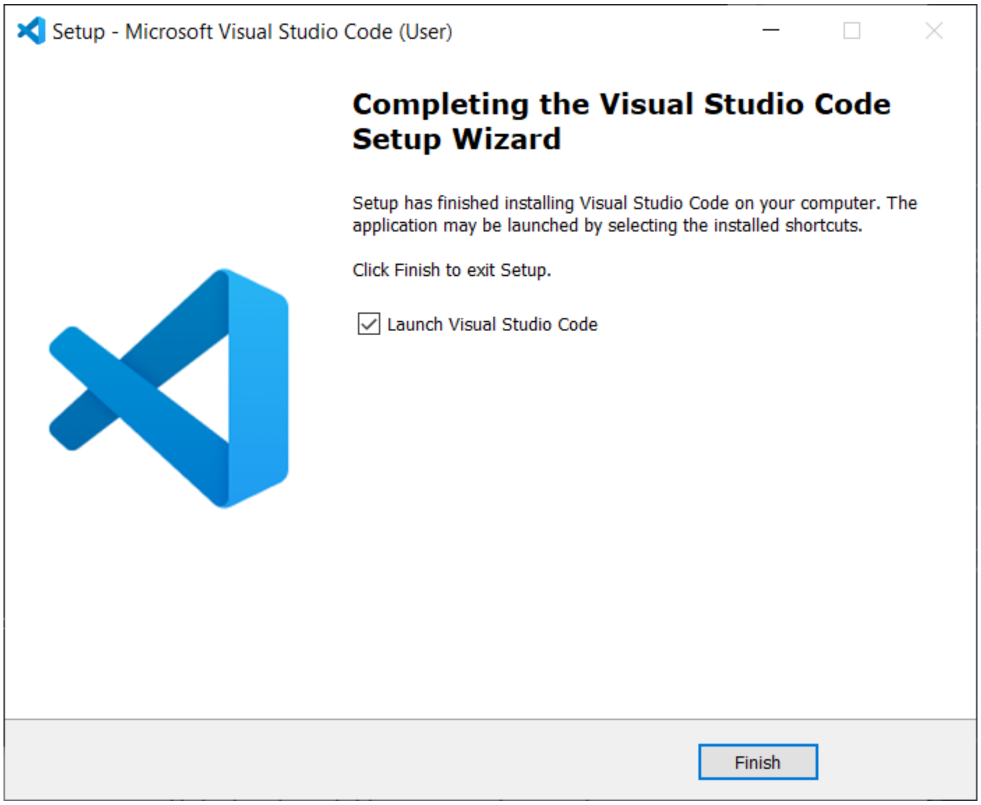
Download Visual Studio Code

Free and built on open source. Integrated Git, debugging and extensions.

User Installer 64 bit 32 bit ARM
System Installer 64 bit 32 bit ARM
.zip 64 bit 32 bit ARM

.deb 64 bit ARM ARM 64
.rpm 64 bit ARM ARM 64
.tar.gz 64 bit ARM ARM 64

.zip Universal Intel Chip Apple Silicon



Setup - Microsoft Visual Studio Code (User)

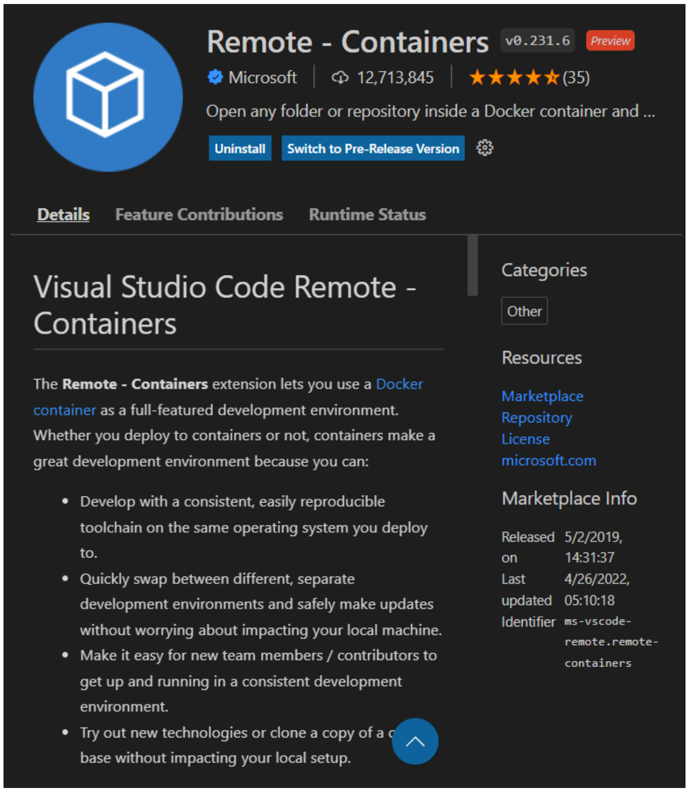
Completing the Visual Studio Code Setup Wizard

Setup has finished installing Visual Studio Code on your computer. The application may be launched by selecting the installed shortcuts.

Click Finish to exit Setup.

☑ Launch Visual Studio Code

Open Visual Studio Code, go to “Extensions”, search for “Remote container Extension” and install it.



Remote - Containers v0.231.6 Preview

Microsoft | 12,713,845 | ★★★★★(35)

Open any folder or repository inside a Docker container and ...

Visual Studio Code Remote - Containers

The **Remote - Containers** extension lets you use a Docker container as a full-featured development environment. Whether you deploy to containers or not, containers make a great development environment because you can:

- Develop with a consistent, easily reproducible toolchain on the same operating system you deploy to.
- Quickly swap between different, separate development environments and safely make updates without worrying about impacting your local machine.
- Make it easy for new team members / contributors to get up and running in a consistent development environment.
- Try out new technologies or clone a copy of a c base without impacting your local setup.

Categories

Other

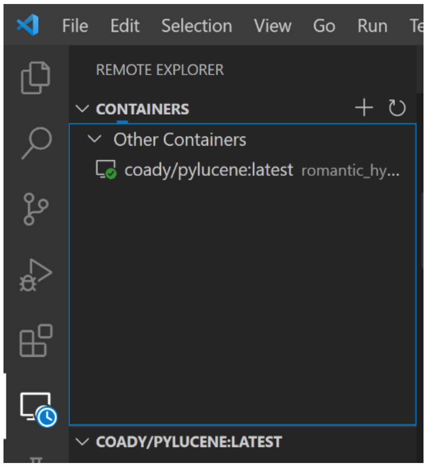
Resources

Marketplace
Repository
License
microsoft.com

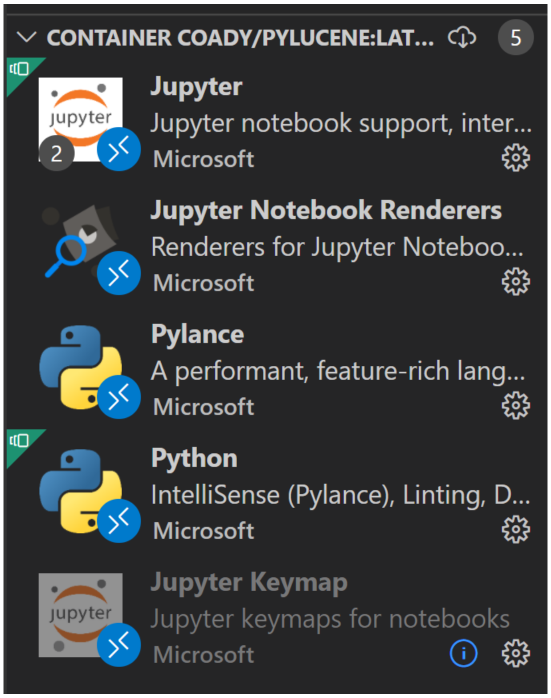
Marketplace Info

Released on	5/2/2019, 14:31:37
Last updated	4/26/2022, 05:10:18
Identifier	ms-vscode-remote.remote-containers

8. Now, you will be able to see a new icon called “Remote Explorer” click on it. You should see “Containers” under which you can find “coady/pylucene”



9. Right-click on “coady/pylucene” and select “Attach to container”. This will open a new VS Code window where coady/pylucene is contained.
10. In this VS Code window, install Python, Jupyter Notebook, and any additional libraries.



CONTAINER COADY/PYLUCENE:LAT... 5

Jupyter
Jupyter notebook support, inter...
Microsoft

Jupyter Notebook Renderers
Renderers for Jupyter Noteboo...
Microsoft

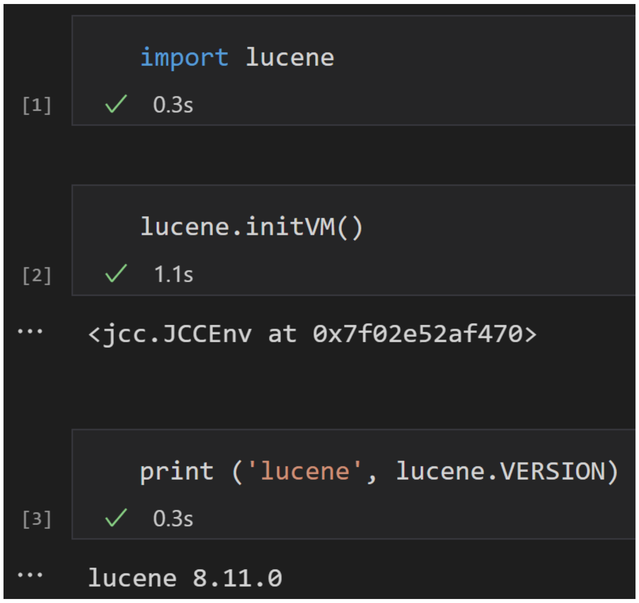
Pylance
A performant, feature-rich lang...
Microsoft

Python
IntelliSense (Pylance), Linting, D...
Microsoft

Jupyter Keymap
Jupyter keymaps for notebooks
Microsoft

11. Once installed, create a new Jupyter Notebook.
12. To test whether PyLucene has been successfully installed, try:

```
import lucene
lucene.initVM()
print ('lucene', lucene.VERSION)
```



```
[1]  import lucene
     ✓ 0.3s

[2]  lucene.initVM()
     ✓ 1.1s
...  <jcc.JCCEnv at 0x7f02e52af470>

[3]  print ('lucene', lucene.VERSION)
     ✓ 0.3s
...  lucene 8.11.0
```

If you get an output similar to the one above, it means everything was successfully installed.

13. The following is a simple example code for using PyLucene tokenizer:

```
#Import
import lucene
from java.io import StringReader
from org.apache.lucene.analysis.ja import JapaneseAnalyzer
from org.apache.lucene.analysis.standard import StandardAnalyzer,
StandardTokenizer
from org.apache.lucene.analysis.tokenattributes import CharTermAttribute

lucene.initVM(vmargs=['-Djava.awt.headless=true'])
```

```
<jcc.JCCEnv at 0x7ff9ea152e30>
```

```
# Basic tokenizer example.
test = "This is how we do it."
tokenizer = StandardTokenizer()
tokenizer.setReader(StringReader(test))
charTermAttrib = tokenizer.getAttribute(CharTermAttribute.class_)
tokenizer.reset()

tokens = []
while tokenizer.incrementToken():
  tokens.append(charTermAttrib.toString())
print(tokens)
```

```
['This', 'is', 'how', 'we', 'do', 'it']
```

```
# StandardAnalyzer example.
analyzer = StandardAnalyzer()
stream = analyzer.tokenStream("", StringReader(test))
stream.reset()
tokens = []
while stream.incrementToken():
  tokens.append(stream.getAttribute(CharTermAttribute.class_).toString())
print(tokens)
```

```
['this', 'is', 'how', 'we', 'do', 'it']
```