

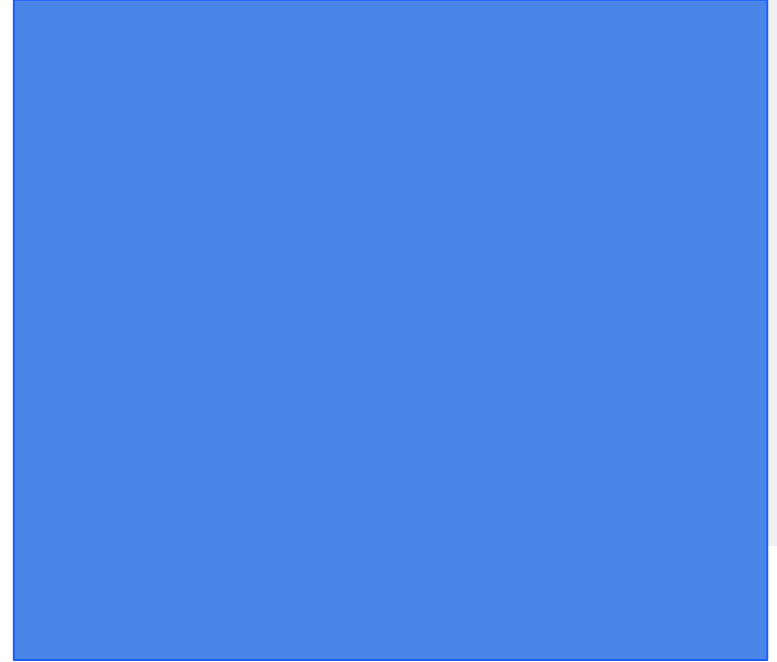


# Big Data Analytics

Dr Sirintra Vaiwsri | Email: [sirintra.v@itm.kmutnb.ac.th](mailto:sirintra.v@itm.kmutnb.ac.th)



# Software Preparation

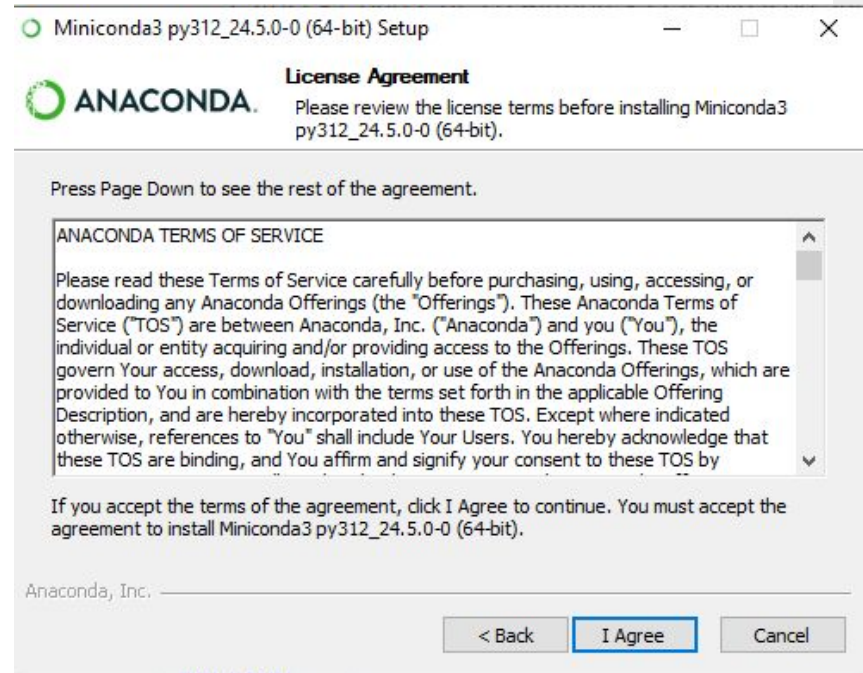
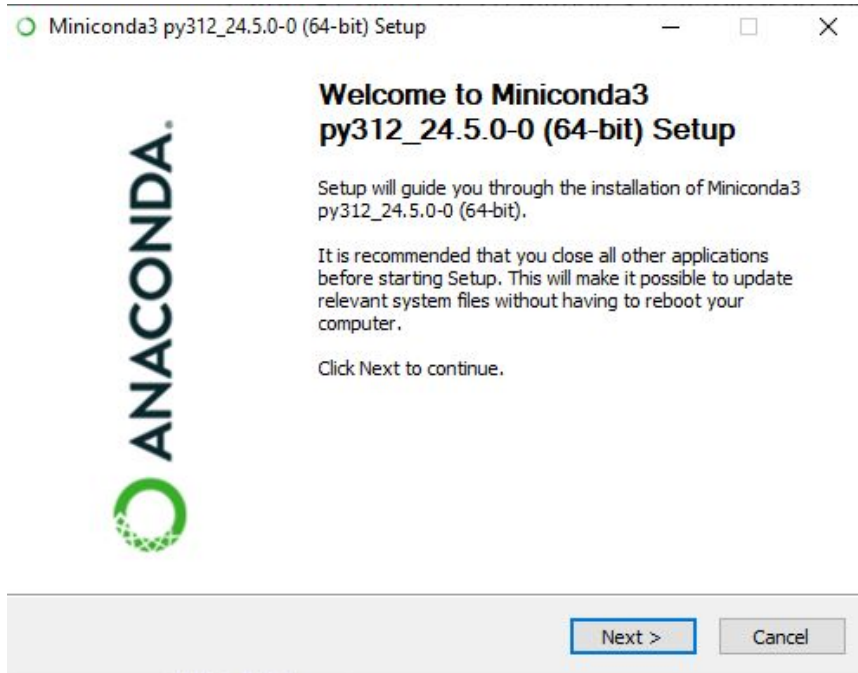


# Download Miniconda

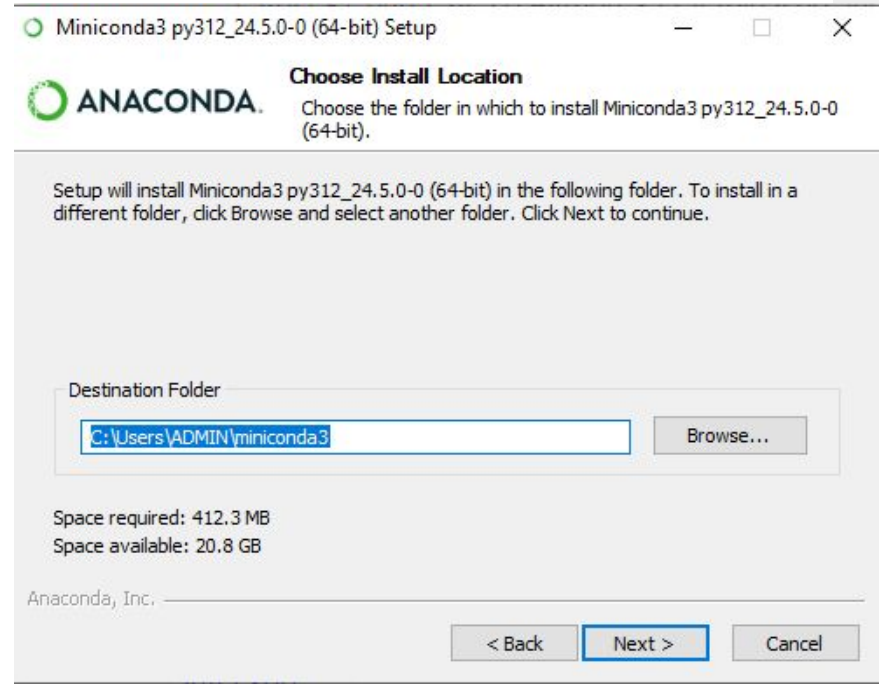
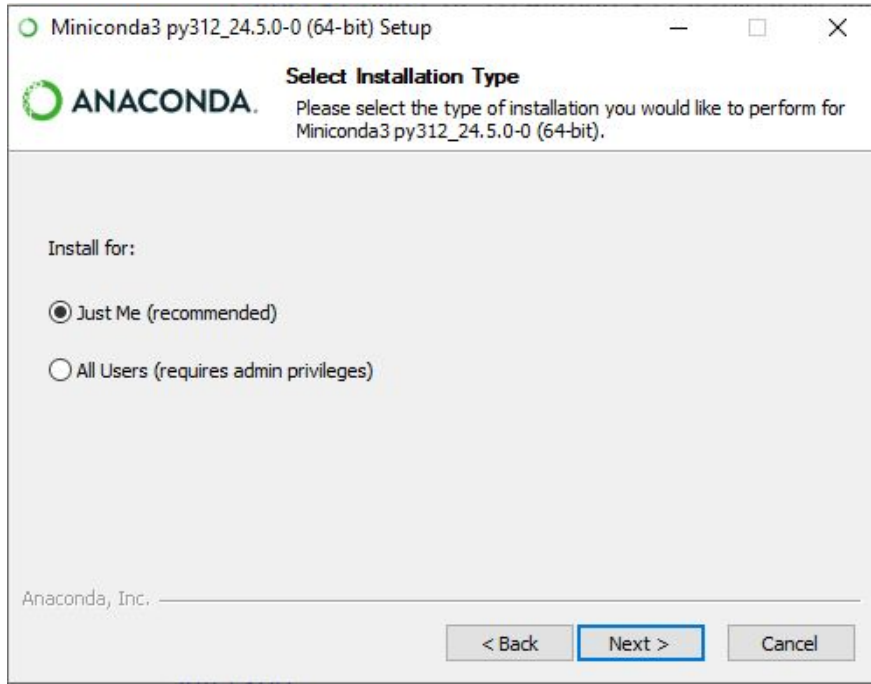
- Download Miniconda [here](#)

Platform	Name	SHA256 hash
Windows	<a href="#">Miniconda3 Windows 64-bit</a>	<code>b1ce11a339c8246010e898065f6fa6feb1940a55fef550b57a</code>
macOS	<a href="#">Miniconda3 macOS Intel x86 64-bit bash</a>	<code>b1c87c8334ea878d30a9976c1860b1004e6d55bdec5228089fe</code>

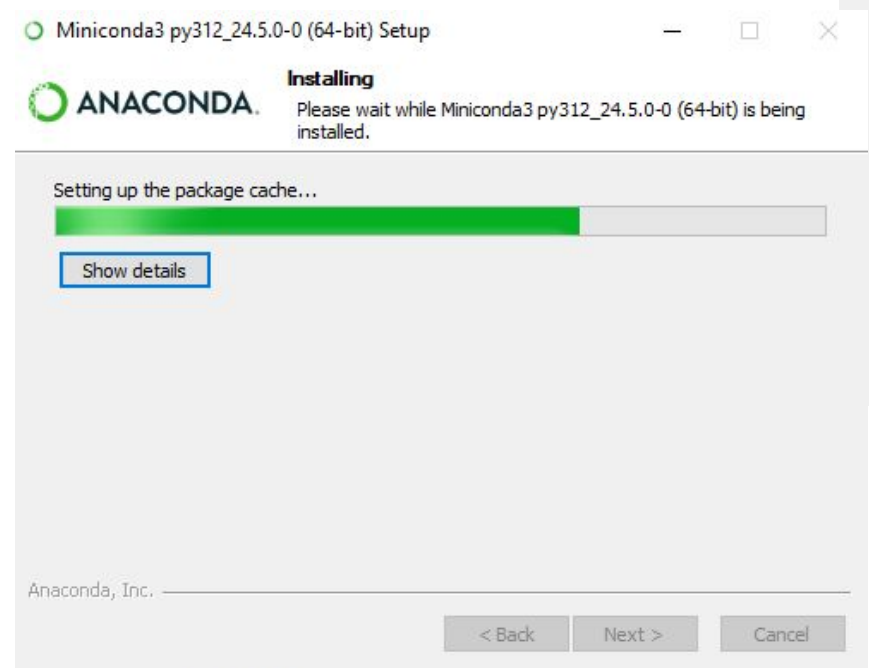
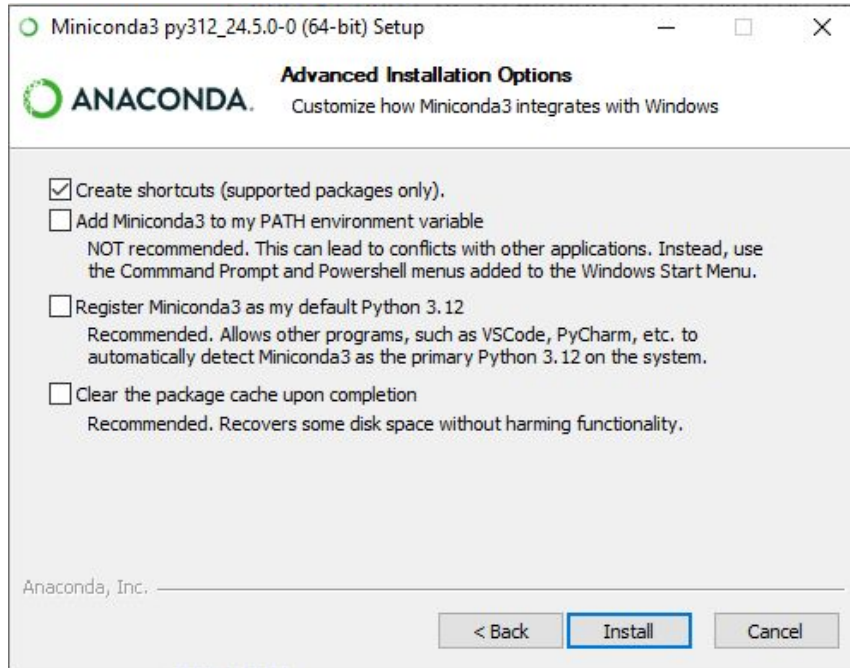
# Install Miniconda



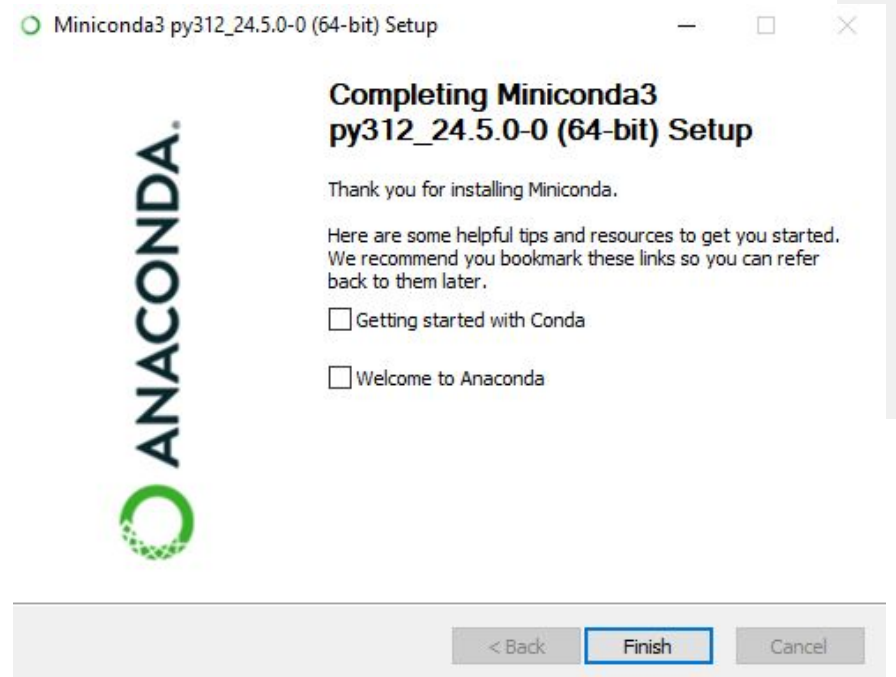
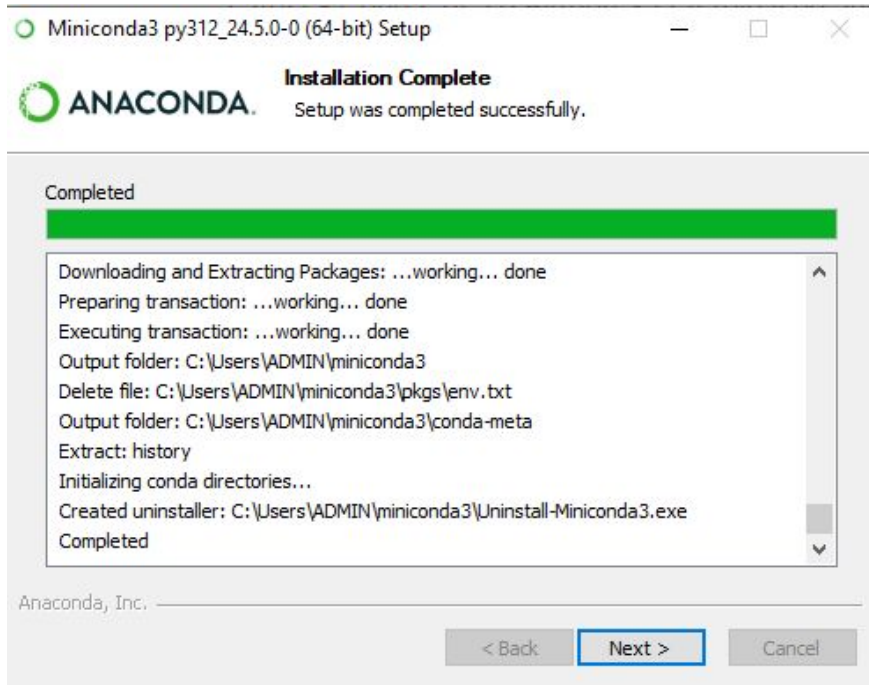
# Install Miniconda



# Install Miniconda

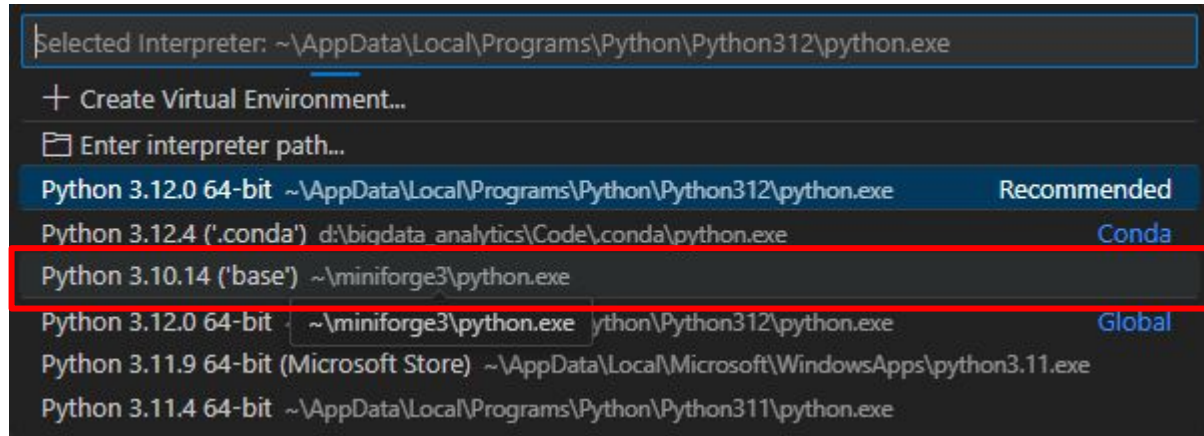


# Install Miniconda



# Select Interpreter

- Type “>” in search or use Ctrl+Shift+P
- Type “Python: Select Interpreter”
- Select Python <version> ('base') ~\miniforge3\python.exe





# Install Pyspark

- Use command:  
conda install pyspark
- Proceed ([y]/n)? - Type y

The following packages will be UPDATED:

certifi 2024.6.2-py312haa95532\_0 --> 2024.7.4-py312haa95532\_0

Proceed ([y]/n)? y

Downloading and Extracting Packages:

Preparing transaction: done

Verifying transaction: done

Executing transaction: done

```
(base) PS D:\bigdata_analytics\bigdata_analytics20> conda install pyspark
Retrieving notices: ...working... done
```

Channels:

- defaults

Platform: win-64

Collecting package metadata (repodata.json): done

Solving environment: done

## Package Plan ##

environment location: C:\Users\ADMIN\miniconda3

added / updated specs:

- pyspark

The following packages will be downloaded:

package	build	
abseil-cpp-20211102.0	hd77b12b_0	1.7 MB
blas-1.0	mkl	6 KB
c-ares-1.19.1	h2bbff1b_0	117 KB
certifi-2024.7.4	py312haa95532_0	160 KB
grpc-cpp-1.48.2	hfe90ff0_1	23.5 MB
libevent-2.1.12	h56d1f94_1	417 KB
libprotobuf-3.20.3	h23ce68f_0	2.2 MB
libthrift-0.15.0	h4364b78_2	777 KB
orc-1.7.4	h623e30f_1	823 KB
python-tzdata-2023.3	pyhd3eb1b0_0	140 KB
re2-2022.04.01	hd77b12b_0	374 KB
six-1.16.0	pyhd3eb1b0_1	18 KB
tbb-2021.8.0	h59b6b97_0	149 KB
Total:		30.3 MB

following NEW packages will be INSTALLED:

abseil-cpp pkgs/main/win-64::abseil-cpp-20211102.0-hd77b12b\_0

# Test Pyspark

- Use command:  
pyspark

```
(base) PS D:\bigdata_analytics\bigdata_analytics2024> pyspark
Python 3.12.4 | packaged by Anaconda, Inc. | (main, Jun 18 2024, 15:03:56) [MSC v.1929 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
24/07/10 14:52:46 WARN Shell: Did not find winutils.exe: java.io.FileNotFoundException: java.io.FileNotFoundException: HADOOP_HOME and hadoop.home.dir are unset. -see https://wiki.apache.org/hadoop/WindowsProblems
Setting default log level to "WARN".
To adjust logging level use sc.setLogLevel(newLevel). For SparkR, use setLogLevel(newLevel).
24/07/10 14:52:47 WARN NativeCodeLoader: Unable to load native-hadoop library for your platform... using builtin-java classes where applicable
24/07/10 14:52:49 WARN Utils: Service 'SparkUI' could not bind on port 4040. Attempting port 4041.
Welcome to

  ____      _
 / ___|  __| | | |
 \___ \  | | | | | |
  ___) | | | | | | |
 |_____|_|_|_|_|_|_| version 3.4.1

Using Python version 3.12.4 (main, Jun 18 2024 15:03:56)
Spark context Web UI available at http://PC02:4041
Spark context available as 'sc' (master = local[*], app id = local-1720597969361).
SparkSession available as 'spark'.
>>> █
```