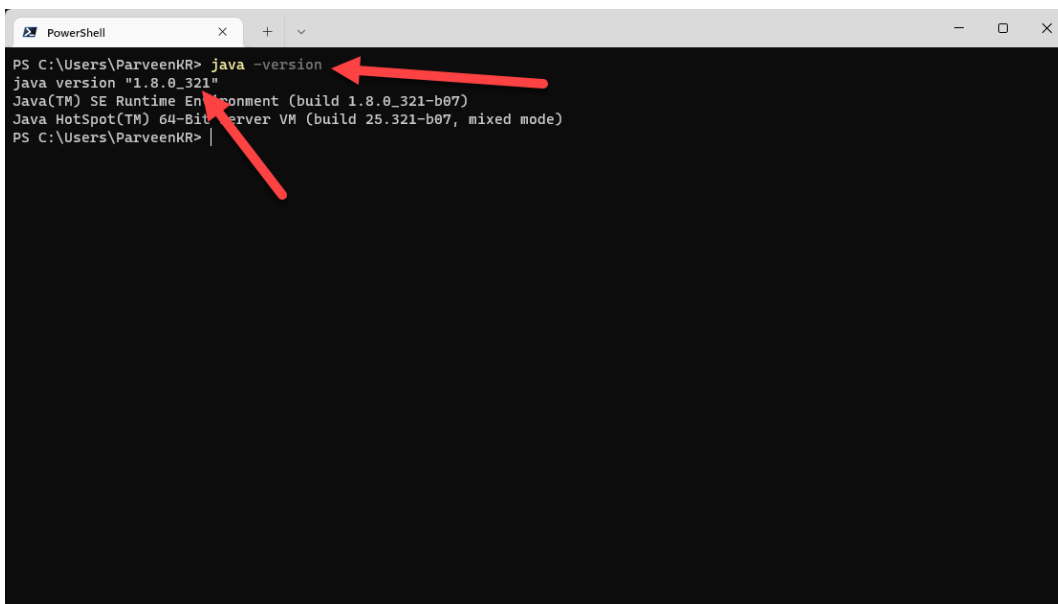


Installing Spark on Windows 10/11

1. Install Java 8: Download Java 8 from the link:
Installers can be downloaded from: <https://bit.ly/391KKzM>
 - a) Set environmental variables:
 - b) User variable:
 - i. Variable: JAVA_HOME
 - ii. Value: C:\Program Files\Java\jdk1.8.0_321
 - c) System variable:
 - i. Variable: PATH
 - ii. Value: C:\Program Files\Java\jdk1.8.0_321\bin
 - d) Test using java -version command on terminal



```
PS C:\Users\ParveenKR> java -version
java version "1.8.0_321"
Java(TM) SE Runtime Environment (build 1.8.0_321-b07)
Java HotSpot(TM) 64-Bit Server VM (build 25.321-b07, mixed mode)
PS C:\Users\ParveenKR>
```

The screenshot shows a PowerShell terminal window with a black background. The command 'java -version' has been entered and executed. The output displays the Java version as '1.8.0_321' and provides details about the runtime environment and the 64-bit server VM. Two red arrows are present: one points to the command 'java -version' and the other points to the first line of the output, 'java version "1.8.0_321"'.

2. Install Scala: Download Scala from the
link: <https://downloads.lightbend.com/scala/2.11.12/scala-2.11.12.msi>
 - a. Set environmental variables:
 - b. User variable:
 - a) Variable: SCALA_HOME;
 - b) Value: C:\Program Files (x86)\scala
 - c. System variable:
 - a) Variable: PATH
 - b) Value: C:\Program Files (x86)\scala\bin

d. Test using scala command on terminal

```
PowerShell 7.2.2
Copyright (c) Microsoft Corporation.

https://aka.ms/powershell
Type 'help' to get help.

PS C:\Users\ParveenKR> scala
Welcome to Scala 2.11.12 (Java HotSpot(TM) 64-Bit Server VM, Java 1.8.0_311).
Type in expressions for evaluation. Or try :help.

scala> |
```

3. Install Spark 1.6.1. Download it from the following link: [Downloads | Apache Spark](https://spark.apache.org/downloads.html)

Download Apache Spark™

- Choose a Spark release: **3.0.3 (Jun 23 2021)**
- Choose a package type: **Pre-built for Apache Hadoop 2.7**
- Download Spark: [spark-3.0.3-bin-hadoop2.7.tgz](#)
- Verify this release using the [3.0.3 signatures](#), [checksums](#) and [project release KEYS](#).

Note that Spark 3 is pre-built with Scala 2.12 in general and Spark 3.2+ provides additional pre-built distribution with Scala 2.13.

Link with Spark

Spark artifacts are hosted in [Maven Central](#). You can add a Maven dependency with the following coordinates:

```
groupId: org.apache.spark
artifactId: spark-core_2.12
version: 3.2.1
```

Installing with PyPi

PySpark is now available in pypi. To install just run `pip install pyspark`.

Convenience Docker Container Images

Spark Docker Container images are available from [DockerHub](#), these images contain non-ASF software and may be subject to different license terms.

Release notes for stable releases

- Spark 3.2.1 (Jan 26 2022)
- Spark 3.1.3 (Feb 18 2022)
- Spark 3.0.3 (Jun 23 2021)

Latest News

- Spark 3.1.3 released (Feb 18, 2022)
- Spark 3.2.1 released (Jan 26, 2022)
- Spark 3.2.0 released (Oct 13, 2021)
- Spark 3.0.3 released (Jun 23, 2021)

APACHECON 2021
September 21-23
[www.apachecon.com](#)

DOWNLOAD SPARK

Built-in Libraries:

- [SQL and DataFrames](#)
- [Spark Streaming](#)
- [MLlib \(machine learning\)](#)
- [GraphX \(graph\)](#)
- [Third-Party Projects](#)

- Extract it to C:\spark
- Set environmental variables:
 - System Variable:
 - Variable: SPARK_HOME
 - Value: C:\spark
 - Variable: HADOOP_HOME
 - Value: C:\spark
- System variable:
 - Variable: PATH

