PySpark Environment Setup

Java Related Settings:

Windows 10 Java Environment Variables:

* JAVA\_OPTIONS=-Xmx512M -Xms512M
* JAVA\_HOME=D:\Java\jdk1.8.0\_241

Java Installation Verification:

D:\Temp>java -version

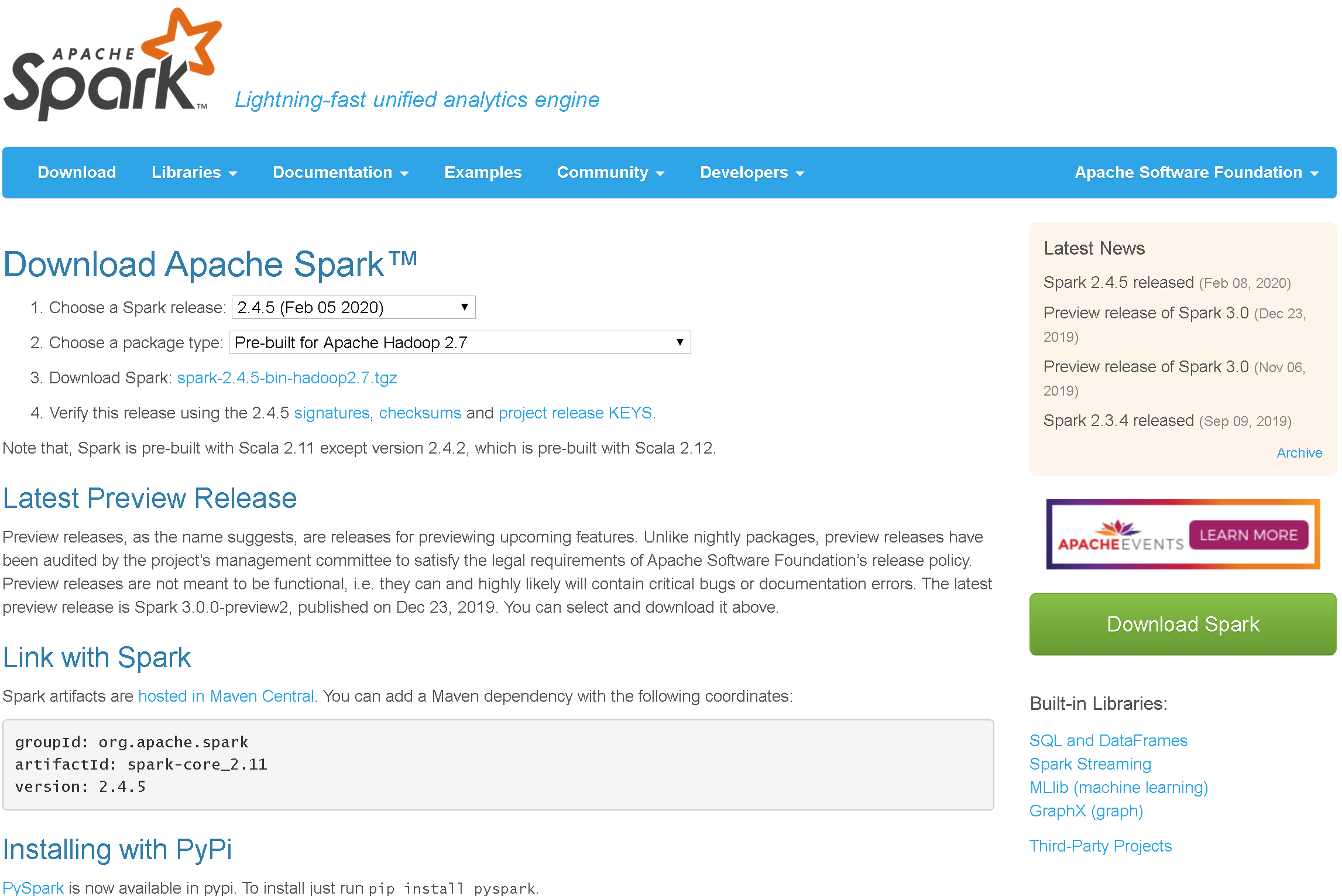
java version "1.8.0\_241"

Java(TM) SE Runtime Environment (build 1.8.0\_241-b07)

Java HotSpot(TM) 64-Bit Server VM (build 25.241-b07, mixed mode)

Spark Related Settings

Sample spark download URL: <https://www.apache.org/dyn/closer.lua/spark/spark-2.4.5/spark-2.4.5-bin-hadoop2.7.tgz>.



## Resources

1. UCSD instructions: <https://mas-dse.github.io/DSE230/installation/windows/>.
2. Created around 3/2018: <https://medium.com/big-data-engineering/how-to-install-apache-spark-2-x-in-your-pc-e2047246ffc3>.
3. Created around 5/2019: <https://www.knowledgehut.com/blog/big-data/how-to-install-apache-spark-on-windows>.
4. A good overview from 1/2019: <https://medium.com/@naomi.fridman/install-pyspark-to-run-on-jupyter-notebook-on-windows-4ec2009de21f>.
5. Some winutils replacement background: <https://dataintoresults.com/post/apache-spark-on-windows-without-winutils-exe/>.