

pyspark 윈도우 설치 메뉴얼

7zip

- <https://www.7-zip.org/download.html>

윈도 패키지 매니저 choco

- <https://chocolatey.org/install>
- @"%SystemRoot%\System32\WindowsPowerShell\v1.0\powershell.exe" -NoProfile -InputFormat None -ExecutionPolicy Bypass -Command "iex ((New-Object System.Net.WebClient).DownloadString('https://chocolatey.org/install.ps1'))" && SET "PATH=%PATH%;%ALLUSERSPROFILE%\chocolatey\bin"
- 관리자 권한으로 설치

스파크 2.3.1

- <https://www.apache.org/dyn/closer.lua/spark/spark-2.3.1/spark-2.3.1-bin-hadoop2.7.tgz>
- 버전 맞춰야 하니 choco로 설치하지 말것
- 7zip으로 압축해제
- SPARK_HOME → C:\spark\spark-2.3.1-bin-hadoop2.7

하둡 바이너리 (winutil.exe)

- <https://github.com/stevetloughran/winutils/blob/master/hadoop-2.7.1/bin/winutils.exe>
- C:\spark\spark-2.3.1-bin-hadoop2.7\bin 에 넣기
- HADOOP_HOME → C:\spark\spark-2.3.1-bin-hadoop2.7

스칼라(optional)

- choco로 설치하면 jre7 깔아버리고 환경변수 수정해버림
- 애 먼저 설치하고 jre8재설치 하는게 좋을듯, 버전 안맞으면 에러남
- SCALA_HOME → C:\bin\scala-2.11.4

자바 설치

- jdk >= 8.0
- jre >= 8.0
- JAVA_HOME → C:\Program Files\Java\jre1.8.0_181

아나콘다 설치

- <http://joonyon.tistory.com/25>
- 직접 다운받거나 choco로 다운
 - 관리자 권한으로 설치
- 애는 환경변수 패스 직접 추가해줘야 함
 - C:\tools\Anaconda3
 - C:\tools\Anaconda3\Scripts
 - C:\tools\Anaconda3\Library\bin
- ~~python3.7 업데이트 필요~~

system PATH

- C:\Program Files\Java\jdk1.8.0_181\bin
- C:\Program Files\Java\jre1.8.0_181\bin
- C:\spark\spark-2.3.1-bin-hadoop2.7\bin
- C:\bin\scala-2.11.4\bin

pyspark 설치

- conda install pyspark