



Installatiehandleiding

Introductietraining Spark

Downloads (Windows):

- Java 8 JDK
<http://download.oracle.com/otn-pub/java/jdk/8u181-b13/96a7b8442fe848ef90c96a2fad6ed6d1/jdk-8u181-windows-x64.exe>
- Python 2.7.15
<https://www.python.org/ftp/python/2.7.15/python-2.7.15.amd64.msi>
- Spark 2.3.1
<https://www.apache.org/dyn/closer.lua/spark/spark-2.3.1/spark-2.3.1-bin-hadoop2.7.tgz>
- Visual Studio Code
<https://go.microsoft.com/fwlink/?Linkid=852157>
- Training
<https://github.com/martinsuijs/Introductie-training-spark>

1. Download alle benodigde bestanden
2. Installeer Java 8 JDK in folder c:\Program Files\Java
3. Maak environment variabele JAVA_HOME = c:\Program Files\Java\jdk1.8.0_181
4. Voeg %JAVA_HOME%\bin toe aan de environment variabele Path
5. Installeer Python in folder c:\python27
6. Maak environment variabele PYTHON_HOME=c:\python27
7. Voeg %PYTHON_HOME% toe aan de environment variabele Path
8. Installeer Spark in de folder c:\apache-spark
9. Maak environment variabele SPARK_HOME=c:\apache-spark
10. Voeg % SPARK _HOME%\bin toe aan de environment variabele Path
11. Maak een kopie van het bestand c:\apache-spark\conf\log4j.properties.template en hernoem deze naar log4j.properties
12. Verander de regel log4j.rootCategory=INFO, console naar log4j.rootCategory=ERROR, console
13. Sla het bestand op
14. Maak folder c:\training
15. Download het training-materiaal en sla (unzipped) op in de folder c:\training
16. Maak de nieuwe folder(s) c:\tmp\hive
17. Maak de nieuwe folder(s) c:\hadoop\bin
18. Sla het bestand winUtils.exe (c:\training\installation) op in de folder c:\hadoop\bin
19. Voer het volgende commando uit:
`\hadoop\bin\winUtils chmod 777 \tmp\hive`
20. Installeer Visual Studio Code in folder c:\Program Files
21. In Visual Studio Code open de workspace Spark-introductietraining

