

Introduction into Spark

# TUTORIAL PREPARATION

For this tutorial, this will be the base:

- Preparation:
  - Go to spark website and download latest base release:  
<http://spark.apache.org/downloads.html>
  - Unpack on your local machine (in user space; hint: use a unique location for this tutorial in your workstation)
  - **Testing:** run a small example - Pi-estimation

```
$ ./bin/run-example SparkPi 10
```

- Install Java JRE:  
<http://www.oracle.com/technetwork/java/javase/downloads/index.html>  
(check environment variables in '**Path**' for '**Java\_Home**' on your system)
- Install Scala: <http://www.scala-lang.org/download/>
- Install SBT: <http://www.scala-sbt.org/download.html>
- Install python 2.7.13 from here: <https://www.python.org/downloads/windows/>  
and add location to '**PATH**' environment variable
- Add tool "winutils.exe" to path
  - Download from <http://public-repo-1.hortonworks.com/hdp-win-alpha/winutils.exe>
  - Create a directory on your system, e.g. 'C:\Hadoop' and copy "winutils.exe" into subdirectory 'bin'
  - Set environment variable on your system to point to the actual location (HADOOP\_HOME=c:\Hadoop)

- Download Spark (evtl. need for zipper, e.g. 7zip, to unpack release)
  - Go to spark website and download latest base release:  
<http://spark.apache.org/downloads.html>
  - Create tutorial directory somewhere and unpack release
- Open command line and run example from the release home dir
  - **Testing:** run a small example - Pi-estimation

```
$ .\bin\run-example SparkPi 10
```