

**Big Data Management and Analytics**  
WS 2016/17

**Tutorial 6: Spark**

**Assignment 6-1**     *Spark & PySpark*

In this assignment we are going to become a bit more familiar with Spark

- (a) First make sure that Java ( $\geq 1.7$ ) is installed. Download Spark from <http://spark.apache.org/downloads.html> and install it on your local machine. It already includes the Spark Python API PySpark.
- (b) Implement the word count example using PySpark.

**Assignment 6-2**     *Matrix-Matrix multiplication with Spark*

Recap: In tutorial 4 you performed a matrix-matrix multiplication using MapReduce.

- (a) Download the code template *matrixMultiplication\_template.py* and become familiar with it.
- (b) Implement the two functions *map1*, *map2* and replace *None* with your code
- (c) Test your implementation by checking the result for multiplying the matrices

$$A := \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \text{ and } B := \begin{pmatrix} 7 & 8 \\ 9 & 10 \\ 11 & 12 \end{pmatrix}$$