

Übungsblatt 1

1 HDFS

Machen Sie sich mit der Funktionsweise der Hadoop-Filesystem-Tools vertraut:

1. Finden Sie den „Hadoop File System Shell Guide“ im Web.
2. Melden Sie sich mit Ihrem Account bigdataXYZ auf einer der Maschinen `infbd06...11.fh-trier.de` per SSH an, z. B.:

```
ssh bigdata030@infbd09.fh-trier.de
```
3. Legen Sie ein Verzeichnis `caesar` in Ihrem User-Verzeichnis im HDFS an (das ist nicht das Verzeichnis `/home/bigdataXYZ` auf dem Rechner `infbd09...`!).
4. Kopieren Sie die Datei `/data/caesar.txt` im HDFS in das neue Verzeichnis. Setzen Sie dabei den Replikationsfaktor auf 5.
5. Geben Sie `caesar.txt` aus.
6. Geben Sie den jeweiligen Replikationsfaktor des originalen `caesar.txt` und Ihrer Kopie aus.
7. Listen Sie alle Dateien in Ihrem Benutzer-Verzeichnis `/user/bigdataXYZ` und allen seinen Unterverzeichnissen auf.

Für die Punkte 4 (Replikationsfaktor setzen) und 6 (Replikationsfaktor anzeigen) müssen Sie vermutlich eine Internetrecherche bemühen.

2 HDFS-API

- Schreiben Sie ein Java-Programm, das 1.000 Zufallszahlen zwischen 0 und 1.000.000 zeilenweise in eine Datei im HDFS schreibt! Benutzen Sie dafür das HDFS-API in Java (und nicht etwa die Kommandozeilen-Tools).
- Kopieren Sie die Datei mit den Zufallszahlen auf das lokale Dateisystem und zeigen Sie sie an.

Hinweise:

- Sie finden ein Projekt-Gerüst für Eclipse im Stud.IP in `bdt-uebung-1-hdfsexample.zip`.
- Ergänzen Sie den Code in `WriteRandomNumbers.java`.
- Exportieren Sie Ihren Code aus Eclipse als Jar-File: File → Export → Runnable JAR File. Das geht erst, nachdem Sie Ihr Programm einmal lokal gestartet haben.
- Alternativ können Sie Ihr Projekt mit Maven bauen:
 - `mvn package` auf einer der `infbd`-Maschinen

- Run as → Maven build → Goal: "package" in Eclipse
- Kopieren Sie das Jar-File auf einen Rechner infbdt06...11 und starten Sie Ihr Programm dort mit `yarn jar <meinprogramm.jar>`.