# Übung 1 – 3 Fortgeschrittene HDFS Beispiele

Im Folgenden werden drei fortgeschrittene HDFS Beispiele kurz erläutert und mit einem Beispiel hinterlegt:

## Checksum

Mit dem Befehl Checksum kann die Prüfsumme seiner Datei im HDFS ermittelt werden:

Beispiele:

*hadoop fs -checksum hdfs://quickstart.cloudera /user/cloudera/pg14591.txt #1*

*hadoop fs -checksum <file:///etc/hosts> #None*

## Snapshot

Mit dem Snapshot Befehl können Snapshots des HDFS erstellt und verwaltet werden. Snapshots sind eine Momentaufnahme eines Datenbestandes oder Zustand. In diesem Fall wird eine Kopie des HDFS gezogen, die als Backup später wieder eingespielt werden kann oder auch auf ein anderes System dupliziert werden kann.

Außerdem können Snapshots beispielsweise auch verglichen werden oder es kann nur einzelne Dateien oder Verzeichnisse als Snapshot gespeichert werden. Voraussetzung ist hierfür, dass für ein Verzeichnis oder eine Datei mit „allowSnapshot“ gekennzeichnet ist.

Beispiele:

*sudo -u hdfs hdfs dfsadmin -allowSnapshot /user/cloudera/*

*hdfs dfs -createSnapshot /user/cloudera/ SNAPSHOT1*

*hdfs snapshotDiff /user/cloudera/ SNAPSHO1 SNAPSHOT2*

## Crypto

Crypto dient dazu, Dateien auf dem HDFS zu verschlüsseln und entschlüssen und somit eine Datensicherheit zu gewährleisten.

Als Abfolge muss zunächst ein Schlüssel (Key) und ein Ordner angelegt werden.

Daraus wird eine sogennante Zone – letztendlich sowas wie ein Verschlüsselungsbereich, in dem alle Dateien verschlüsselt abgelegt werden.

Beispiele:

*sudo –u hdfs hadoop fs -mkdir /zone*

*hadoop key create testkey*

*hdfs crypto –getFileEncryptionInfo –path /user/cloudera/pg14591.txt*

*sudo –u hdfs hdfs crypto -createZone -keyName testKey -path /zone*