# Distcopy

Mit dem Befehl distcp kann ein Verzeichnis rekursiv kopiert werden. Beispiel:

hadoop distcp books books2

hadoop fs –ls books2

* distcp nutzt zum Kopieren MapReduce.
* Wird zum Kopieren von großen Datenmengen verwendet. Innerhalb des Clusters aber auch zwischen verschiedenen Clustern.

# Merge Files

Mit dem Befehl getmerge können mehrere Dateien zu einer Datei zusammengefügt werden. Beispiel:

cd /home/cloudera/Desktop/Übungen/fh-muenster-bde-lesson-4/src/main/resources

hadoop fs –put books/ books

hadoop fs –getmerge books/ ./allbooks.txt

cat allbooks.txt

# Snapshots

Mit Snapshots können ReadOnly-Kopien von Dateien oder Ordnerstrukturen erstellt werden, welche z.B. als Backup genutzt werden können. Dabei wird neuer Speicherplatz nur benötigt, wenn anschließend Änderungen an den originalen Dateien vorgenommen werden. Änderungen werden in rückwärtiger chronologischer Reihenfolge gespeichert. Ursprungsdaten werden rekonstruiert indem die Änderungen von den aktuellen Daten „subtrahiert“ werden. Blocks der Daten werden nicht kopiert, es wird eine Liste der Blocks und Dateigrößen gespeichert.

hadoop fs –allowSnapshot books

hadoop fs –createSnapshot books booksnapshot

hadoop fs –ls /user/cloudera/books/.snapshot

hadoop fs –renameSnapshot books booksnapshot newBookSnapshot

hadoop fs –ls /user/cloudera/books/.snapshot

hadoop fs –deleteSnapshot books newBookSnapshot