**Cassandra Datenspeicherung und Komprimierungsverfahren**

Cassandra verwendet zur Datenspeicherung SSTables(Sorted String Tables) welche unveränderlich im Speicher abgelegt werden. Werden Daten verändert, so wird nicht wie bei traditionellen relationalen Datenbanken eine Dekomprimierung mit anschließendem Überschreiben und Komprimieren der Daten durchgeführt, sondern weitere SSTable im Speicher abgelegt, welche dann von Zeit zu Zeit mit bereits bestehenden SSTables verglichen und verbunden werden um Daten aktuell zu halten und Speichernutzung zu minimieren.

Datenbanktabellen können mittels verschiedener Verfahren komprimiert werden. Cassandra unterstützt dabei folgende Komprimierungsverfahren:

* LZ4-Komprimierung
* Snappy-Komprimierung
* Deflate-Komprimierung

Man kann sowohl beim Erstellen von Tabellen als auch beim Verändern festlegen, welche Methode zum Komprimieren verwendet werden soll.

*CREATE TABLE Tabelle1(*

*…*

*)*

*WITH compression = { ‘class’ : ‘LZ4Compressor’|‘DeflateCompressor’|‘SnappyCompressor’ };*

*ALTER TABLE Tabelle1*

*WITH compression = { ‘class’ : ‘LZ4Compressor’|‘DeflateCompressor’|‘SnappyCompressor’ };*

Somit ist es möglich das Komprimierungsverfahren an die in der Tabelle geführten Daten anzupassen oder sogar zu deaktivieren.

Der Algorithmus den die **Snappy**-Komprimierung verwendet ist auf hohe Komprimierungs- und Dekomprimierungsgeschwindigkeit ausgelegt mit einer folglich schlechteren Komprimierungsrate.

Die Deflate-Komprimierung zielt darauf hinaus Daten verlustlos zu komprimieren wie dies zum Beispiel bei ZIP-Dateien der Fall ist.

Die Stärken beider vorher genannten Verfahren werden in der **LZ4**-Komprimierung vereint. Hierbei werden extrem hohe Komprimierungs- und Dekomprimierungsgeschwindigkeiten (50-80% schneller im Vergleich zu an anderen Verfahren) mit einer verlustlosen Kompression verbunden und trotzdem eine sehr gut Kompressionsrate erreicht, weshalb LZ4 auch das am meisten verwendete Verfahren ist.