

## Apache Hadoop

**Apache Hadoop**<sup>1</sup> ist ein Framework (frei und geschrieben in Java) für skalierbare, verteilte Software. Hadoop basiert auf dem *MapReduce*-Algorithmus der von Google erdacht und 2010 patentiert wurde.<sup>2</sup> Weiters wurden Vorschläge für verteilte Dateisysteme vom *Google File System (GFS)*<sup>3</sup> implementiert. Das Hadoop-Framework ermöglicht es somit intensive Rechenprozesse mit großen Datenmengen (Big Data) auf Clustern durchzuführen.

Das Dateisystem von Hadoop nennt sich *Hadoop Distributed File System (HDFS)*, es ermöglicht es mehrere 100 Mio. Dateien auf Dateisystemen mehrerer Rechner (Knoten) abzuspeichern. Die Dateien werden dabei in Datenblöcke fixer Länge zerteilt und redundant auf die teilnehmenden Knoten verteilt. Das System arbeitet nach dem Master-Slave-Prinzip: ein Master-Knoten nimmt eingehende Daten an und verteilt diese auf die Slave-Knoten. Datenlänge und Redundanzlevel sind dabei konfigurierbar.

Für Apache Hadoop existieren diverse Erweiterungen die den Umgang mit den verteilten Rechnern vereinfachen sollen. Die namhaftesten Erweiterungen sind hierbei HBase<sup>4</sup>, Hive<sup>5</sup>, Pig<sup>6</sup>, Chukwa<sup>7</sup> und ZooKeeper<sup>8</sup>.

## DevStack

**DevStack**<sup>9</sup> ist ein dokumentiertes Shell-Skript um komplette OpenStack-Entwicklungsumgebungen<sup>10</sup> zu erstellen. Entwickelt und verteilt wird DevStack über GitHub<sup>11</sup> und dabei werden ausschließlich Ubuntu (11.10 Oneiric oder neuer) und Fedora (F16 oder neuer) als

Betriebssysteme unterstützt. Nach dem klonen des GitHub-Repos von DevStack wird *stack.sh* ausgeführt. Das Shell-Skript konfiguriert dabei verschiedene Kombinationen von *Ceilometer*, *Cinder*, *Glance*, *Heat*, *Horizon*, *Keystone*, *Nova*, *Quantum*, und *Swift*. Für jedes Skript sind auf der offiziellen DevStack-Website Dokumentation zu finden (z. B. <http://devstack.org/stack.sh.html> für die Dokumentation von *stack.sh*).

---

<sup>1</sup> <http://hadoop.apache.org/>

<sup>2</sup> <http://research.google.com/archive/mapreduce.html>

<sup>3</sup> <http://research.google.com/archive/gfs.html>

<sup>4</sup> <http://hbase.apache.org/>

<sup>5</sup> <http://hive.apache.org/>

<sup>6</sup> <http://pig.apache.org/>

<sup>7</sup> <http://incubator.apache.org/chukwa/>

<sup>8</sup> <http://zookeeper.apache.org/>

<sup>9</sup> <http://devstack.org/>

<sup>10</sup> <http://www.openstack.org/>

<sup>11</sup> <https://github.com/openstack-dev/devstack>