Cloudera Flume Installation

Für die Installation von Flume verwenden wir die Cloudera Distribution.

Dazu fügen wie zuerst die Paketquellen für die entsprechende Version hinzu. Zum Zeitpunkt des Wiki Beitrags ist die CDH3 aktuell. Ihr könnt das Beispiel natürlich auch auf eine andere Version anwenden:

DISTRO müsst ihr durch eure verwendete Version ersetzen. Verwendet ihr Ubuntu 10.10, muss es also maverick lauten.

Quellcode

1. sudo bash -c 'cat <<EOF > /etc/apt/sources.list.d/cloudera.list
2. deb http://archive.cloudera.com/debian DISTRO-cdh3 contrib
3. deb-src http://archive.cloudera.com/debian DISTRO-cdh3 contrib
4. EOF'

Anschließend signieren wir den APT Key von Cloudera:

Quellcode

1. wget -q -O - http://archive.cloudera.com/debian/archive.key | sudo apt-key add -

Danach können wir die Paketlisten aktualisieren

Quellcode

1. sudo apt-get update

Im Beispiel wollen wir den Flume Master und den Flume Node auf einem Computer installieren, ihr könnt das natürlich für euer Setup anpassen

Quellcode

1. sudo apt-get install hadoop-0.20

Nun müssen wir uns überlegen welche Rollen der Hadoop Node denn übernehmen soll. Im Beispiel entscheiden wir uns für ein Single Node Setup. Das heißt, dass alle Hadoop Komponenten erstmal auf einem Computer laufen sollen. Dazu installieren wir die folgenden Init Skripte.

Quellcode

1. sudo apt-get install flume-node flume-master

== Flume Weboberfläche ==
Ihr könnt den Flume Master auf Port 35871 erreichen. Für localhost ruft ihr den Flume Master Service wie folgt auf:
localhost:35871/
Ihr könnt den Flume Node auf Port 35862 erreichen. Für localhost ruft ihr den Flume Node Service wie folgt auf:
localhost:35862/

== Flume als Entwicklungssystem ==
Wenn ihr euren Rechner eher selten für Flume nutzt, dann könnt ihr verhindern, dass Flume bei jedem Hochfahren des

Rechners startet indem ihr die Initscripte aus den Runlevels entfernt:

Quellcode

1. sudo update-rc.d -f flume-node remove
2. sudo update-rc.d -f flume-master remove

Um alle Services zu starten oder stoppen könnt ihr dann z.B. folgenden Code nutzen:

Quellcode

1. for service in /etc/init.d/flume-*; do sudo $service restart; done

== Literatur ==

- wiki.cloudera.com/display/DOC/Flume+Installation