

Distributed Data Management

Sheet I I (until 15.07.2010)

This exercise will be performed in multiple parts. This is the first part, the second will be released in a day or two. Please hand in the full solution in two weeks.

Part I. (20 pts)

Obtain a Bittorrent client (e.g. μ Torrent or Vuze) and download the following Torrent (might take some time...have patience): http://www.ifis.cs.tu-bs.de/webfm_send/505

This will download the Cloudera appliance (a preconfigured virtual machine with all [Hadoop](#) applications readily installed and configured.)

Unpack the downloaded archive. Obtain a copy of a VMware compatible virtualization software (e.g. the free VMware Player).

Open the downloaded virtual machine image within that software and start it (you may ignore all error messages about outdated VMware tools / or copied machines).

Login user is "training", password is also "training".

Open a terminal window and call: "sudo -u hbase hbase shell" (password is still "training").

Congratulations, you just started an [HBase](#) shell (which is an Open Source clone of Google BigTable). Using the shell, create a new table named "superhero" to store... superheroes! (with two column families "identity" and "power"). Use the internet to learn how to use HBase...).

For each hero, the first name, last name, and the hero alias should be stored in the "identity" column family, and for the most notable power the name and a short description should be stored in the "power" family (e.g. "Christoph", "Lofi", "Captain DB", "Coffee Immunity", "Cannot be harmed by coffee overdose").

Enter yourself/yourself as records in that table. Use the get-command to return the row/s and copy/paste the HBase output as your exercise solution.