

HBase-Writer 0.20.3 Released

Contributed by Ryan Smith
Monday, 16 February 2009

HBase-Writer version 0.20.3 has been released and is available for download now. This version of HBase-Writer has a new runtime dependency: ZooKeeper. This is because HBase-0.20.X now depends on ZooKeeper to manage configuration and connection information. This version has been tested on a few Heritrix2-2.0.2 crawls on Hadoop 0.20.1, HBase 0.20.1 and ZooKeeper 3.2.1. and works fine as far as my tests go. The main difference you will have to be aware of when upgrading from 0.19.x to 0.20.x are 2 things:

- In the global sheet configuration for your heritrix job, There is no "master" address for HBaseWriterProcessor anymore. Instead you need to provide a comma-separated list of zookeeper hosts that make up the zookeeper quorum (zkquorum). Heritrix will talk to ZooKeeper to determine the master address of HBase. This has been done by HBase in 0.20.x to avoid the Master node being a SPOF (single point of failure) Support for an alternate zk client port has been added as well..

- You need to add the zookeeper.jar to the lib/ folder. The zookeeper jar is included with the HBase distribution, or you can download it from the OSM Archive Repository .

The other changes in this version were under-the-hood. The BatchUpdate API has been deprecated in HBase-0.20.x and HBase-Writer is now using the new Put/Get API from HBase to write and manage records when doing crawls. Feel free to create Issues if you want to see support added for something or if you have a bug to report. Thanks for checking it out and Enjoy! :)