Started timing

Java, Streaming PY, Streaming R, HadoopR Spark SparkR

SparkR Documentation

<http://amplab-extras.github.io/SparkR-pkg/rdocs/1.2/index.html>

Detailed information for Hadoop Streaming

hadoop jar hadoop-streaming-2.6.0.jar -info

HOME = /user/dotcz12/

390.7 K data/book.txt

4.2 M data/book10.txt

15.3 M data/book100.txt.gz

38.5 M data/book100.txt

382.0 M data/book1000.txt

3.7 G data/book10000.txt

4.0 M data/enwiki-latest-stub-meta-current1.xml

63.1 M data/enwiki-latest-stub-articles8.xml

181.8 M data/enwiki-latest-pages-meta-current1.xml

227.2 M data/enwiki-latest-all-titles-in-ns0

390.2 M data/enwiki-latest-pages-articles3.xml

671.6 M data/enwiki-latest-all-titles

708.7 M data/enwiki-latest-abstract1.xml

4.4 G data/enwiki-latest-pages-meta-current20.xml

48.1 G data/enwiki-latest-pages-articles.xml

262.8 G data/enwiki-latest-stub-meta-history.xml