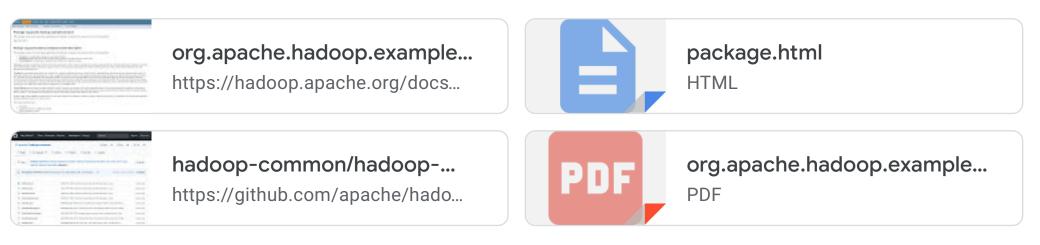






Prof. Jigar Pandya • Jul 29 (Edited Jul 31)

Benchmarking is a way to have insight about prototype work success. There can be various rubrics/matrices factors based on which certain benchmarking competition can be hosted. Terasort is one of the ways to benchmark cluster computing.



4 class comments



Prof. Jigar Pandya Jul 29

I was involved in High Performance Computing Linpack Benchmarking with ROCKS cluster more than a decade ago.

More information about ROCKS at http://www.rocksclusters.org



Prof. Jigar Pandya Jul 29

To start MapReduce with Apache Hadoop refer https://hadoop.apache.org/docs/stable/hadoop- mapreduce-client/hadoop-mapreduce-client-core/MapReduceTutorial.html



Prof. Jigar Pandya Jul 29

Source code to download and understand partitioning, etc.

https://github.com/apache/hadoop-common/tree/trunk/hadoop-mapreduce-project/hadoopmapreduce-examples/src/main/java/org/apache/hadoop/examples/terasort



Prof. Jigar Pandya Jul 31

In-memory sorting techniques at application level, external merge sort used by databases, and terms non-decreasing and ascending while non-increasing and descending, etc shall be explored as other related topics.



Add class comment..

