

Nội dung

- Tìm hiểu về Hadoop Map Reduce, cài đặt hệ thống phân tán Hadoop và triển khai ứng dụng với MapReduce
- Tìm hiểu và cài đặt hệ thống Spark

Hướng dẫn và yêu cầu

Phần 1: Tìm hiểu về Hadoop và cài đặt hệ thống Hadoop với 1 Master Node và 3 Slave Node.

Tìm hiểu về Hadoop:

<https://link.springer.com/book/10.1007/978-1-4842-1910-2>

Cài đặt hệ thống Hadoop

<https://hadoop.apache.org/docs/stable/hadoop-project-dist/hadoop-common/SingleCluster.html>

Phần 2: Triển khai ứng dụng với Hadoop MapReduce

1. Ví dụ WordCount

<https://helpex.vn/article/chuong-trinh-dem-tu-voi-mapreduce-va-java-5c6b03a3ae03f628d053aa72>

2. Cài đặt một số giải thuật song song dựa trên Hadoop: *tính số PI, sắp xếp chẵn lẻ, nhân hai ma trận*

Phần 3: Tìm hiểu và cài đặt hệ thống phân tán Spark

Link: <https://codingpearls.com/huong-dan-tim-hieu-trien-khai-apache-spark-tren-windows-linux>