



MapReduce原理及其主要实现平台分析

亢丽芸, 王效岳, 白如江

山东理工大学科技信息研究所 淄博 255049

Kang Liyun, Wang Xiaoyue, Bai Rujiang

Institute of Scientific & Technical Information, Shandong University of Technology, Zibo 255049, China

- 摘要
- 参考文献
- 相关文章

Download: PDF (944KB) [HTML \(1KB\)](#) Export: BibTeX or EndNote (RIS) Supporting Info

摘要 针对海量数据处理在处理速度、存储空间、容错性、访问时间等方面存在的问题,对Google MapReduce编程模型的原理、执行流程等进行分析研究,介绍4种主要的MapReduce实现平台Hadoop、Phoenix、Disco、Mars,从编程语言、构建平台、功能特点和应用领域4个方面对4种平台进行比较分析,以期对MapReduce编程模型原理及其应用平台有一个较全面的认识。

关键词: [MapReduce](#) [实现平台](#) [Hadoop](#) [Phoenix](#) [Disco](#) [Mars](#)

Abstract: Due to the problems in processing speed, storage space, fault tolerance, access time and others of massive data processing, this paper analyzes principle, implementation process of Google MapReduce programming model, introduces four main MapReduce implementation platforms, including Hadoop, Phoenix, Disco and Mars. Then, it separately compares them in four aspects as the programming language, building platform, functions and features, applications to have a comprehensive understanding of MapReduce programming model and its application platforms.

Keywords: [MapReduce](#), [Implementation platform](#), [Hadoop](#), [Phoenix](#), [Disco](#), [Mars](#)

收稿日期: 2011-09-21;

基金资助:

本文系国家自然科学基金一般项目“海量网络学术文献自动分类研究”(项目编号:10BTQ047)和山东省自然科学基金项目“大规模学术文献并行处理与语义分类研究”(项目编号:ZR2011GL025)的研究成果之一。

引用本文:

亢丽芸, 王效岳, 白如江. MapReduce原理及其主要实现平台分析[J]. 现代图书情报技术, 2012, V28(2): 60-67

Kang Liyun, Wang Xiaoyue, Bai Rujiang. Analysis of MapReduce Principle and Its Main Implementation Platforms[J], 2012, V28(2): 60-67

链接本文:

<http://www.infotech.ac.cn/CN/> 或 <http://www.infotech.ac.cn/CN/Y2012/V28/I2/60>

- [1] 微博-百度百科[EB/OL].(2011-12-21).[2012-01-20]. <http://baike.baidu.com/view/1567099.htm>. (Microblog- Baidu Encyclopedia[EB/OL].(2011-12-21).[2012-01-20]. <http://baike.baidu.com/view/1567099.htm>.)
- [2] Hadoop Archive-淘宝共享数据平台TBDATA.org[EB/OL]. [2012-02-08]. <http://www.tbdata.org/archives/category/cloud-computing/hadoop>. (Hadoop Archive: Taobao Shared Data Platform TBDATA.org[EB/OL]. [2012-01-08]. <http://www.tbdata.org/archives/category/cloud-computing/hadoop>.)
- [3] Hypertable: An Open Source, High Performance, Scalable Database[EB/OL]. [2011-08-19]. <http://hypertable.org/>.
- [4] Disco: Massive Data-minimal Code[EB/OL]. [2011-12-22]. <http://discoproject.org/about>.
- [5] 中国网络媒体论坛[EB/OL]. (2011-11-22). [2011-12-22]. <http://www.chinaz.com>. (China's Internet Media Forum[EB/OL]. (2011-11-22). [2011-12-22]. <http://www.chinaz.com>.)

[1] 赵华茗. 分布式环境下的文档相似度研究与实现[J]. 现代图书情报技术, 2011, 27(7/8): 14-20

[2] 张兴旺, 李晨晖, 秦晓珠. 云计算环境下大规模数据处理的研究与初步实现[J]. 现代图书情报技术, 2011, 27(4): 17-23

[3] 赵华茗. 搭建基于云计算的开源海量数据挖掘平台[J]. 现代图书情报技术, 2010, 26(10): 76-81

Service

- ▶ 把本文推荐给朋友
- ▶ 加入我的书架
- ▶ 加入引用管理器
- ▶ Email Alert
- ▶ RSS

作者相关文章

- ▶ 亢丽芸
- ▶ 王效岳
- ▶ 白如江

- [4] 杨代庆,张智雄.基于Hadoop的海量共现矩阵生成方法*[J].现代图书情报技术,2009,25(4):23-26
- [5] 吴宝贵,丁振国.基于Map/Reduce的分布式搜索引擎研究[J].现代图书情报技术,2007,2(8):52-55