

证书号第1524969号



# 发明专利证书

发明名称：一种 MapReduce 系统中的数据采样和划分方法

发明人：姚金字；陈琪；肖臻

专利号：ZL 2012 1 0205841.7

专利申请日：2012年06月18日

专利权人：北京大学

授权公告日：2014年11月26日

本发明经过本局依照中华人民共和国专利法进行审查，决定授予专利权，颁发本证书并在专利登记簿上予以登记。专利权自授权公告之日起生效。

本专利的专利权期限为二十年，自申请日起算。专利权人应当依照专利法及其实施细则规定缴纳年费。本专利的年费应当在每年06月18日前缴纳。未按照规定缴纳年费的，专利权自应当缴纳年费期满之日起终止。

专利证书记载专利权登记时的法律状况。专利权的转移、质押、无效、终止、恢复和专利权人的姓名或名称、国籍、地址变更等事项记载在专利登记簿上。



局长  
申长雨

申长雨





(12) 发明专利

(10) 授权公告号 CN 102799486 B

(45) 授权公告日 2014. 11. 26

(21) 申请号 201210205841. 7

(22) 申请日 2012. 06. 18

(73) 专利权人 北京大学

地址 100871 北京市海淀区颐和园路 5 号北京大学

(72) 发明人 姚金宇 陈琪 肖臻

(74) 专利代理机构 北京君尚知识产权代理事务所(普通合伙) 11200

代理人 余长江

(51) Int. Cl.

G06F 9/50(2006. 01)

H04L 29/08(2006. 01)

(56) 对比文件

CN 101183368 A, 2008. 05. 21,

CN 101226557 A, 2008. 07. 23,

李震等. 云计算环境下的改进型 Map-Reduce 模型. 《计算机工程》. 2012, 第 38 卷(第 11 期),

审查员 王思文

权利要求书 2 页 说明书 6 页 附图 4 页

(54) 发明名称

一种 MapReduce 系统中的数据采样和划分方法

(57) 摘要

本发明涉及一种 MapReduce 系统中的数据采样和划分方法,其步骤包括:1) 客户端向 MapReduce 系统中提交任务请求,所述 MapReduce 系统中的主控节点将 Map 任务划分成采样和普通任务,所述主控节点 Master 将采样任务优先下发到各个分节点 Worker 进行执行;2) 根据各个分节点 Worker 上的 Map 采样任务筛选出样本集合,并将样本上传至主控节点 Master 进行合并;3) 所述主控节点 Master 根据 Map 采样任务结果得到 Reduce 任务工作量,对 Reduce 任务划分键值区间,实现负载均衡,完成采样和划分。在存在数据倾斜的应用场景下,本发明可以较大幅度地提高整个 MapReduce 任务的执行效率,本发明的数据采样和划分算法对于原有 MapReduce 系统带来的额外开销较小,在数据分布比较均匀的应用场景下基本上不影响原系统的执行效率。

