

基于 FPGA 的高效数据过滤技术研究

陈宝远, 张秀芝, 梁 状

(哈尔滨理工大学 测控技术与仪器黑龙江省高校重点实验室, 黑龙江 哈尔滨 150080)

摘 要: 为了提高 Hadoop 的数据处理效率, 在 Hadoop 集群中加入 FPGA 对 Hadoop 分布式文件系统中存储的数据进行过滤, 这样 MapReduce 程序只需要对过滤后的数据进行自定义的操作, 从而避免了扫描全部数据. 实验表明处理海量的数据时, FPGA 能够显著地过滤掉无用数据, 从而提高数据处理效率.

关键词: 数据过滤; FPGA; 海量数据; Hadoop

Research on Efficient Data Filtering Based on FPGA

CHEN Bao-yuan, ZHANG Xiu-zhi, LIANG Zhuang

(The Higher Education Key Laboratory for Measuring & Control Technology and Instrumentations of

Heilongjiang Province, Harbin University of Science & Technology, Harbin 150080, China)

Abstract: In order to increase the efficiency of data processing, FPGA was injected into Hadoop cluster. FPGA will filter the data stored in HDFS(Hadoop Distributed File System). In this way MapReduce just need to process the filtered data. A series experiment shows that FPGA could filter useless data and increase the efficiency of data processing.

Key words: data filtering; FPGA; massive data; hadoop

作者简介:

陈宝远 男, (1970-), 副教授, 硕士生导师. 研究方向为通信工程、测试计量技术及仪器.

张秀芝(通讯作者) 女, (1990-), 硕士研究生. 研究方向为通信工程、FPGA.

E-mail: xiuzhi1225@163.com.

梁 状 男, (1990-), 硕士研究生. 研究方向为集成电路设计.