基于 Poisson 过程的 Hadoop 集群平均寿命计算方法

李佳宣，郑晓薇，张　永

（辽宁师范大学 计算机与信息技术学院，辽宁 大连 116081）

摘　要： Hadoop 集群在连续长时间执行任务的过程中，集群故障会对集群造成不同程度的冲击与损害，是集群性能下降、系统故障频发的主要原因，更会导致集群平均寿命的大大缩短.为解决此问题，提出了一种基于 Poisson 过程的 Hadoop 集群平均寿命预测方法.通过预测计算，及时对集群进行调整优化，使系统性能保持在正常状态.同时在三种不同特征类型的 Hadoop 集群的实验中观察集群在任务执行过程中平均寿命的变化.实验结果表明，该方法能有效地判断 Hadoop 集群故障发生的频率，预测出集群的平均寿命，对集群性能优劣进行评价并及时对集群性能进行优化

关键词： 集群故障；Poisson 过程；集群平均寿命

A Calculation Method of Hadoop Cluster Average Life Based on Poisson Process

LI Jia-xuan ， ZHENG Xiao-wei ，ZHANG Yong

(College of Computer and Information Technology，Liaoning Normal University，Dalian 116081 China)

Abstract： After a long task execution on Hadoop cluster，as cluster faults would hurt Hadoop cluster on different strength，which became the main reason of cluster performance degradation and more frequent cluster faults，more serious would cause the decrease of cluster average life. To solve the problem，this paper proposes a method of Hadoop cluster average life prediction based on Poisson process. By prediction calculation，adjust and optimize the cluster timely to keep the cluster performance in a normal state. Then make several experiments on three different feature types of Hadoop cluster to observe the frequncy of cluster average life. According to the experiment results，this method is the significant to Hadoop cluster performance evaluation and optimization.

Key words： cluster faults; poisson process; cluster average life

作者简介：

李佳宣　女，（1990-），硕士研究生. 研究方向为并行计算、云计算、分布式集群系统.E-mail:lijiaxuan09@aliyun.com.

郑晓薇　女，（1957-），教授.研究方向为并行计算、云计算、分布式集群系统.

张　永　男，（1975-），副教授.研究方向为云计算、数据挖掘.