## 引 “在宇宙中一切事物都是互相关联的，宇宙本身不过是一条原因和结果的无穷的锁链。〔法〕霍尔巴赫”

## 为什么是centos7？

hbase作为分布式kv数据库，具有高吞吐，低延迟等特性，企业内部越来越受大家的青睐。随着越来越多的用户使用，集群规模也不断扩大，我们在hbase容器化上进行了探索实践，特别是现在炙手可热的k8s技术，由于k8s推荐内核3.10以上版本，之前我们一直使用的centos6.5不升级内核版本无法满足需求，因此新版本集群我们使用了公司内部centos7.2系统。



系统内核优化沿用了centos6.5 sysctl配置

初始化完系统，对网络io，和磁盘写进行了基准测试，测试结果write在300MB/s +

网络1024MB/s +

安装hadoop-2.7.1，hbase-1.1.6，jdk1.7.0\_67

启动集群观察hbase 监控check put TP99 达到30ms，而线上同版本hbase集群 put TP99仅有3ms。为什么性能差这么多？分析对比两个集群差异，不同在于操作系统，以及hadoop版本。

## 针对以上两个不同点，我们设计了两个测试用例：

### 基于同版本操作系统centos7.2,使用不同版本hadoop-2.6.1，hadoop-2.7.1测试观察PUT TP99：

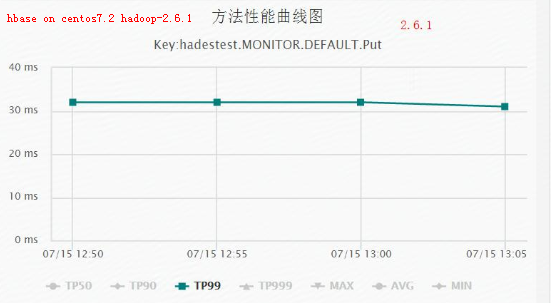


图 1

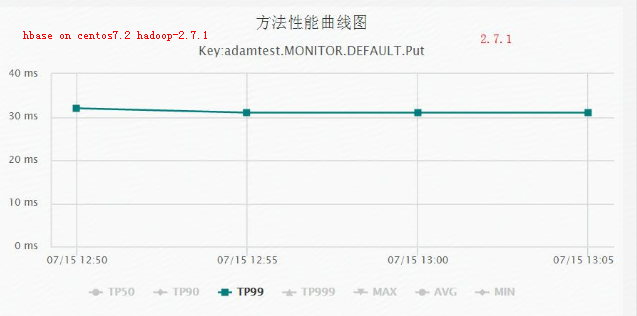


图 2

图 1基于hadoop2.6.1，图2基于hadoop2.7.1 对比PUT TP99指标无明显差异，排除hadoop性能瓶颈

### 同批测试服务器重装系统centos6.5，hbase hadoop版本不变观察PUT TP99：

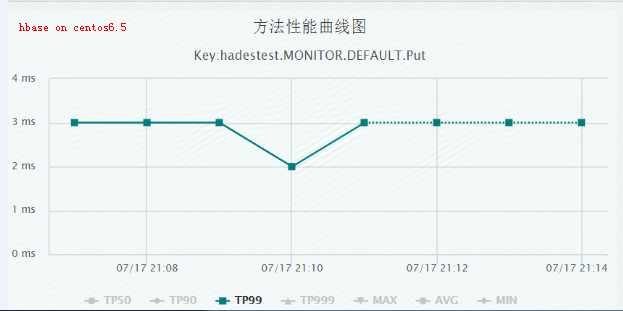


图 3

bingo！这下可以确定是操作系统因素导致TP99 性能差异

接下来我们针对系统优化

### 使用cpupower工具，查看cpu工作模式处于powersave(省电模式)状态：

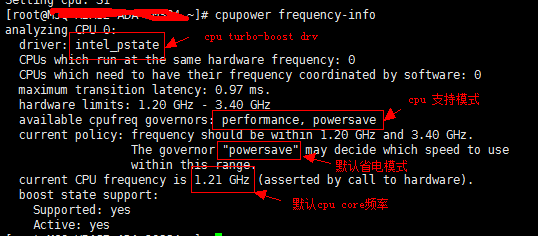


图 4

如图4 cpu 超频驱动支持 intel\_pstate,acpi\_cpufreq,如果driver unknown可以安装相应模块modprobe加载模块

### 设置cpu 为performance(高性能)模式，观察TP99：

cpupower -c all frequency-set -g performance

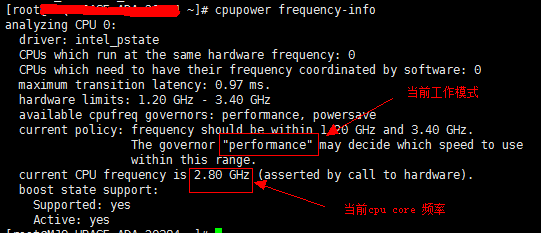


图 5

如图5 cpu频率从1.21GHz提升到2.80GHz，TP99 下降至20ms左右

### 优化perf-bias：

cpupower -c all set -b 0

默认perf-bias 为6 perf-bias官方解释(The range of valid numbers is 0-15, where 0 is maximum performance and 15 is maximum energy efficiency)

这个优化项对tp99并没有明显改观，此参数提供参考

### cpu还有优化空间？使用centos7自带tuned工具

tuned是RHEL7推荐系统动态调优工具，依赖tuned守护进程，通过profile切换预定义的模式，来动态改变优化硬件资源，或通过sysctl参数改变静态设置

开启tuned：

systemctl start tuned;systemctl enable tuned

查看当前模式：

tuned-adm active默认处于balance模式

查看支持的模式

tuned-adm list

切换为低延时高性能模式：

tuned-adm profile latency-performance

查看TP99 指标并没有改善

查看/usr/lib/tuned/latency-performance/tuned.conf

[cpu]

governor=performance

energy\_perf\_bias=performance

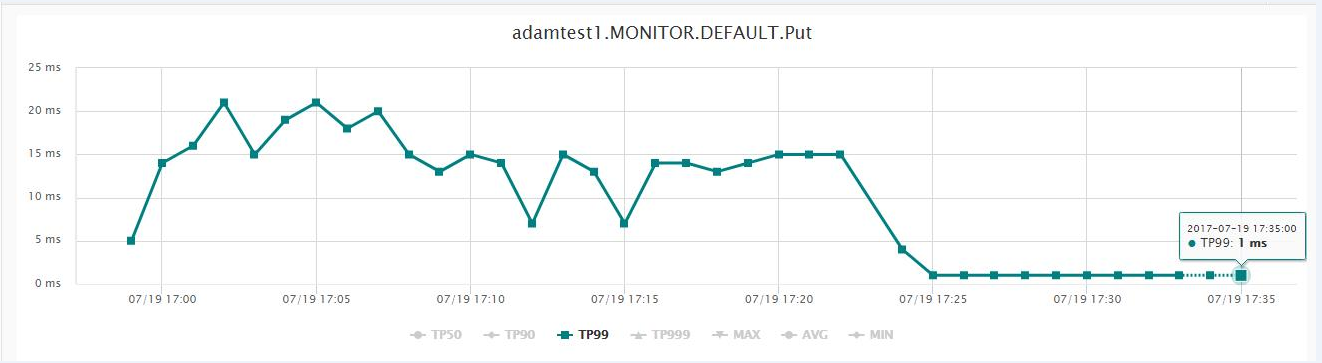
这两个配置其实已经对cpu设置高性能模式

ok到这里可以确定cpu性能已经被我们充分挖掘了

### 抱着一试的态度，我们调整network-latency观察一下

tuned-adm profile network-latency

见证奇迹的时刻！



TP99 降低至1ms！困扰了2天的问题终现端倪

最终确定关闭numa\_balancing使得put TP99延时降低了

kernel.numa\_balancing=0

## 问题：

hbase应用是否适和开启numa？如何优化？接下来会针对以上问题进行深入了解，欢迎各位交流指导。

## 结语：

种下的因，得到的甜果还是苦果，只有体会了才能知其滋味

## 附：

numa 全称Non-Uniform Memory Access ,cpu被划分为node，系统内存被划分到 NUMA *节点，*并且与 socket 相对应，或与特定某一组与本地系统内存子集具有相同访问延迟的 CPU 相对应