

Linux性能分析工具

丁静 dingjingdjdj@gmail.com



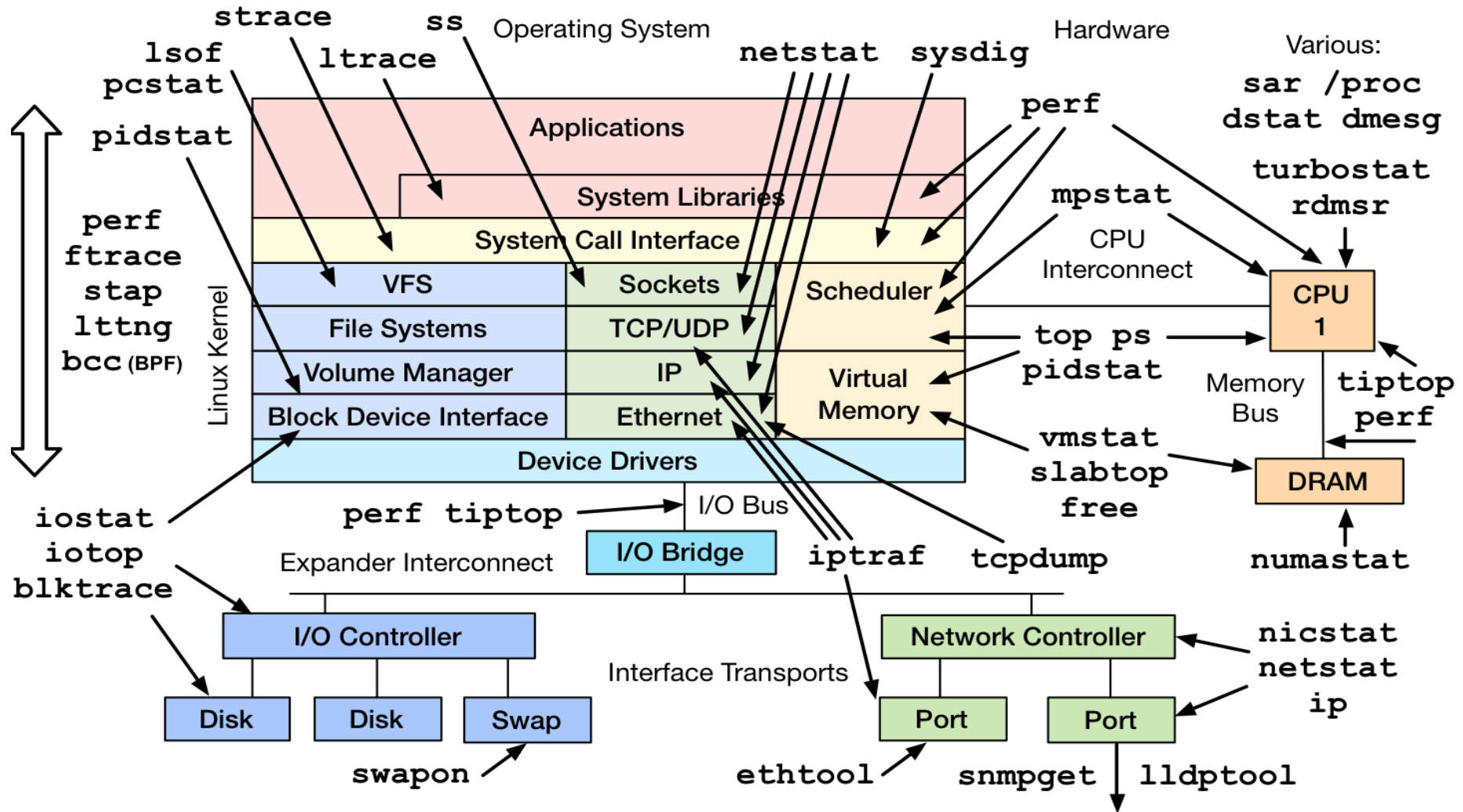
* CPU

* 内存

* I/O

* 网络

Linux Performance Observability Tools



Top- 系统概览

* 第一行重点关注 load average: 0.09, 0.07, 0.05: 三个数字分别表示1分钟, 5分钟, 15分钟的系统负载情况

us — 用户空间占用CPU的百分比。

sy — 内核空间占用CPU的百分比。

ni — 改变过优先级的进程占用CPU的百分比

id — 空闲CPU百分比

wa — IO等待占用CPU的百分比

hi — 硬中断 (Hardware IRQ) 占用CPU的百分比

si — 软中断 (Software Interrupts) 占用CPU的百分比

st — (Steal time)。虚拟 CPU 等待实际 CPU 的时间的百分比。如果你的服务是运行在主机上的虚拟机中, 需要关注此值。过高的此值, 说明主机上运行的虚拟机过多, 引起了cpu时间竞争

top- 系统概览

PR — 进程优先级

NI — nice值。负值表示高优先级，正值表示低优先级

VIRT — 进程使用的虚拟内存总量，单位kb。 $VIRT = SWAP + RES$

RES — 进程使用的、未被换出的物理内存大小，单位kb。

$RES = CODE + DATA$

SHR — 共享内存大小，单位kb

S — 进程状态。D=不可中断的睡眠状态 R=运行 S=睡眠 T=跟踪/停止 Z=僵尸进程

%CPU — 上次更新到现在的CPU时间占用百分比

%MEM — 进程使用的物理内存百分比

进程使用的内存用量可以通过 $RES - SHR$ 简单的估算出来。也可以通过ps命令监测进程。

CPU性能分析

- * `mpstat -P ALL 1` ----- cpu 负载均衡
- * `strace` ----- 系统调用监测
- * `sudo strace -cfp [pid]`
- * `pidstat 1` ----- 进程使用cpu状况

内存性能分析

- * `vmstat 1 -Sm` ----- 虚拟内存的统计信息
- * `r`, 可运行队列的线程数, 这些线程都是可运行状态, 只不过 CPU 暂时不可用
- * `swpd`, 已使用的 SWAP 空间大小;
- * `free`, 可用的物理内存大小;
- * `buff`, 物理内存用来缓存读写操作的 buffer 大小;
- * `cache`, 物理内存用来缓存进程地址空间的 cache 大小;
- * `si`, 数据从 SWAP 读取到 RAM (swap in) 的大小;
- * `so`, 数据从 RAM 写到 SWAP (swap out) 的大小;
- * `bi`, 磁盘块从文件系统或 SWAP 读取到 RAM (blocks in) 的大小, block 为单位;
- * `bo`, 磁盘块从 RAM 写到文件系统或 SWAP (blocks out) 的大小, block 为单位;
- * `in`, 被处理过的中断数
- * `cs`, 系统上正在做上下文切换的数目

I/O性能分析

- * `iostat -xkdz 1` ----- I/O统计信息
- * `rrqm/s`:每秒合并放入驱动请求队列的读请求数
- * `wrqm/s`:每秒合并放入驱动请求队列的写请求数
- * `r/s`:每秒发给磁盘设备的读请求数
- * `w/s`:每秒发给磁盘设备的写请求数
- * `rkB/s`:每秒从磁盘设备读取的KB数
- * `wkB/s`:每秒向磁盘设备写入的KB数
- * `argrq-sz`:平均请求大小，单位为扇区（512B）
- * `qvgqu-sz`:在驱动请求队列和在设置中活跃的平均请求数
- * `await`:平均I/O响应时间，包括在驱动请求队列里等待和设置的I/O响应时间（ms）
- * `svctm`:（推断）磁盘设备的I/O平均响应时间（ms）
- * `%util`:设备忙处理I/O请求的百分比（使用率）

网络性能分析

- * tcpdump ----- 网络抓包
- * HTTP, TCP/IP 协议
- * 网络状态(LISTEN, ESTABLISHED, TIME_WAIT, CLOSE_WAIT, FIN_WAIT1, FIN_WAIT2)
- * wireshark, tcptrace ----- 网络分析工具
- * netstat -s ----- 网络统计信息
- *

sar-系统活动监视器

- * `sar -P ALL` 每个CPU利用率
- * `sar -q` CPU运行队列长度
- * `sar -r` 内存利用率
- * `sar -n DEV` 网络接口统计信息
- * `sar -n TCP,ETCP 1` TCP状态统计

perf

- * 基于采样统计
- * 不用更改应用行为
- * 负载低，线上直接运行，对现有应用影响小

性能计数器：

```
sudo perf stat -a sleep 3
```

perf

采样统计:

```
sudo perf record -F 99 -a --call-graph dwarf sleep 3
```

```
sudo perf record -F 99 -p [pid] -a --call-graph dwarf sleep 3
```

结果分析:

```
sudo perf report
```

```
sudo perf report -n -stdio
```

火焰图:

```
sudo perf script | ./stackcollapse-perf.pl > out.perf-folded  
./flamegraph.pl out.perf-folded > ae.svg
```

USE-性能分析方法

* 使用率

在一定时间间隔内，忙于处理请求的线程平均数目，通过top查看

* 饱和度

在一定时间间隔内，请求队列的平均长度，查看loadavg, 或者 running 队列数

* 错误

处于某种原因，请求拒绝或者失败

dmesg | tail

参考

<http://blog.jobbole.com/105610/>

<http://blog.jobbole.com/111795/>

[http: // www.bo56.com /
%e4%b8%80%e6%ac%a1%e7%94%b1%e4%ba%8emtu%e8%ae%be%e7%bd%ae%e4%
b8%8d%e5%bd%93%e5%af%bc%e8%87%b4%e7%9a%84%e7%bd%91%e7%bb%9c%e
8%ae%bf%e9%97%ae%e8%b6%85%e6%97%b6 /](http://www.bo56.com/%e4%b8%80%e6%ac%a1%e7%94%b1%e4%ba%8emtu%e8%ae%be%e7%bd%ae%e4%b8%8d%e5%bd%93%e5%af%bc%e8%87%b4%e7%9a%84%e7%bd%91%e7%bb%9c%e8%ae%bf%e9%97%ae%e8%b6%85%e6%97%b6/)

<http://www.brendangregg.com/index.html>