此处不方便透露主机名及IP地址，所涉及所有截图中的IP及主机名将被抹去

一：测试环境

1. 软件

压力测试工具dnsperf

测试命令：dnsperf -d domain.list -s IP -l300

测试域名在domain.list中配置



测试目标sample/dns\_server



参数说明:

-d: 指定网卡设备

-i：指定dns的服务IP

-p：指定dns的服务端口

-g：指定加载的ebpf的字节码成勋

-Q：指定网卡队列，同时worker数量和网卡队列数相同

1. 硬件

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | CPU | 核数 | 内存 | 网卡 | 角色 | 运行程序 |
| 机器A | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation 82599ES 10-Gigabit SFI/SFP+ | 被打压机器 | sample/dns\_server |
| 机器B | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64 | Intel Corporation 82599EB 10-Gigabit SFP+ | 打压机器 | dnsperf |
| 机器C | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation 82599ES 10-Gigabit SFI/SFP+ | 打压机器 | dnsperf |
| 机器D | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation 82599ES 10-Gigabit SFI/SFP+ | 打压机器 | dnsperf |
| 机器E | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation 82599ES 10-Gigabit SFI/SFP+ | 打压机器 | dnsperf |
| 机器F | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation 82599ES 10-Gigabit SFI/SFP+ | 打压机器 | dnsperf |
| 机器G | Intel(R) Xeon(R) CPU E5-2620 v2 @ 2.10GHz | 24 | 64G | Intel Corporation Ethernet 10G 2P X520 Adapter | 打压机器 | dnsperf |
| 机器H | Intel(R) Xeon(R) CPU E5-2630 v2 @ 2.60GHz | 24 | 64G | Intel Corporation 82599EB 10-Gigabit SFI/SFP+ | 打压机器 | dnsperf |

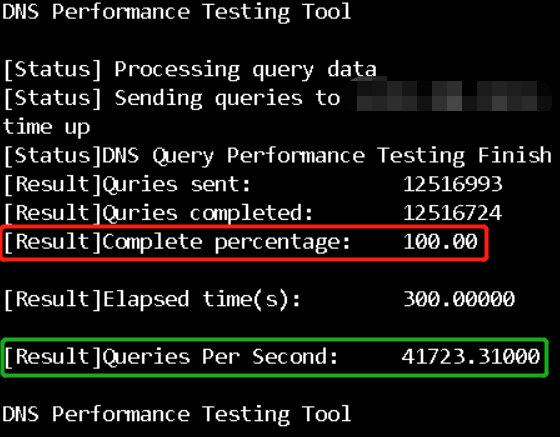
1. sample/dns\_server

此测试程序只是做了简单的dns解包和封包处理。目的是为了接近收发包计算量，为了简单没有单独开发发包工具，用了dnsperf当做打压工具，所以写了个简单的dns\_server。同时因为收报为轮询式的，CPU会直接占满，为了可以观察CPU的利用率所以在收到包为0后进行了nanosleep 1000ns。

xdp\_packet的工作线程命名方式为”xdp\_ wrk\_” + worke\_id

三、测试方法

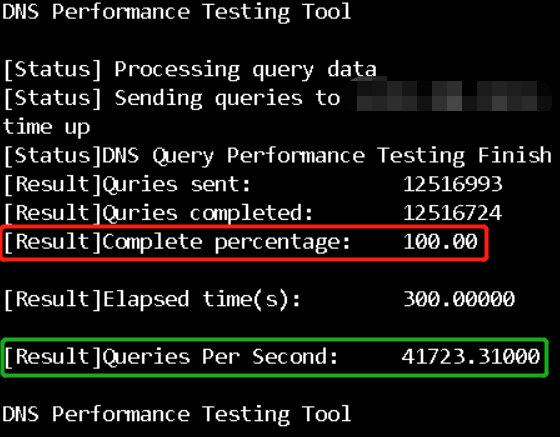
１、在打压机器上逐渐增加dnsperf的进程数，观察所有dnsperf的输出结果如果，如果Complete percentage不是100.00则代表压力超出承受范围，需要通过减少dnsperf进程数的方式来减少压力



2、统计每个dnsperf的输出结果，以单个进程输出结果为例

红框中“Complete percentage”必须为100代表所有请求都得到了正确的响应。

绿框中为qps，将多个同时运行的dnsperf的输出结果中的”Queries Per Second”加在一起就是测试结果的qps



1. 单个打压机器所能产生的压力是有限的，所以需要增加打压机器，并且需要保证所有的机器的输出结果”Complete percetnage”为100.00时达到最大每秒发包数，将同时运行的所有的打压机器上的所有的dnsperf的输出结果中的”Queries Per Second”加在一起即为被打压机器的处理能力

三、单核单网卡队列

1、测试过程

情况一：

单核单网卡队列压力测试，网卡、CPU核、内存属于同一numa节点的情况

第一步：搭建及配置测试环境

* 一个被打压机器机器A：
* 三个打压机器分别是机器B、机器C、机器D,每个机器上同时运行9个dnsperf进程

通过反复测试，如果在增加进程数”Complete percentage”将低于100.00

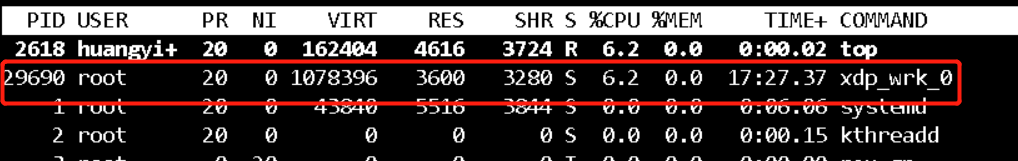
第二步: 在机器A上启动dns\_server

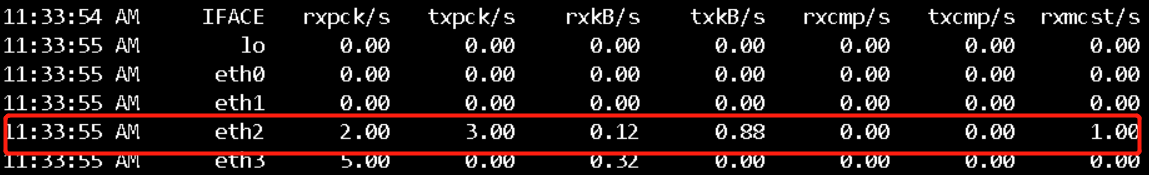


-i 是测试网卡eth2的IP地址

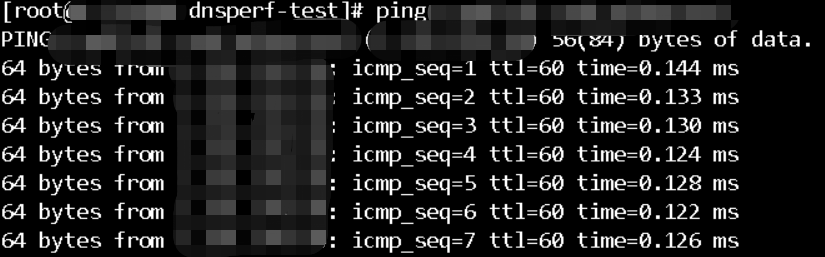
第三步：记录测试之前，各机器状态

* 机器A：CPU及网卡情况

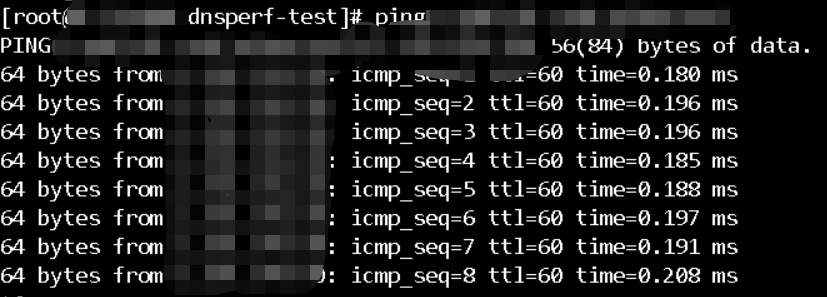




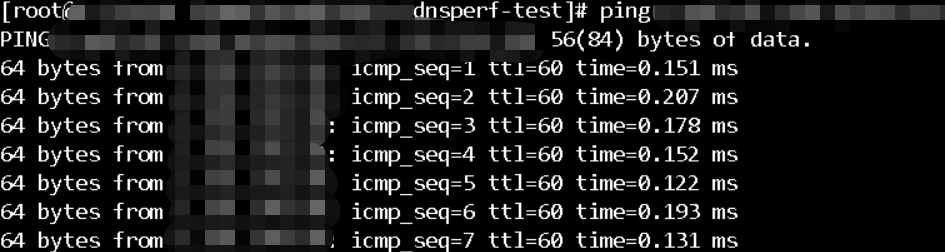
* 机器B到机器A的延时



* 机器C到机器A的延时



* 机器D到机器A的延时



第四步：启动测试

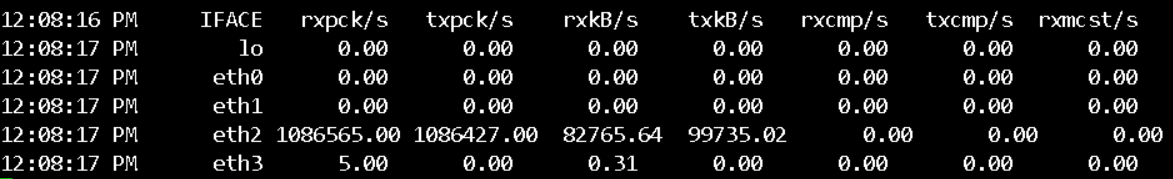
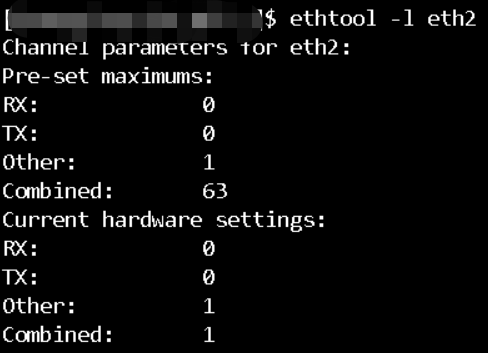
同时在机器B、C、D、上分别启动9个dnsperf并将结果保存在dnsperf.log中，脚本如下

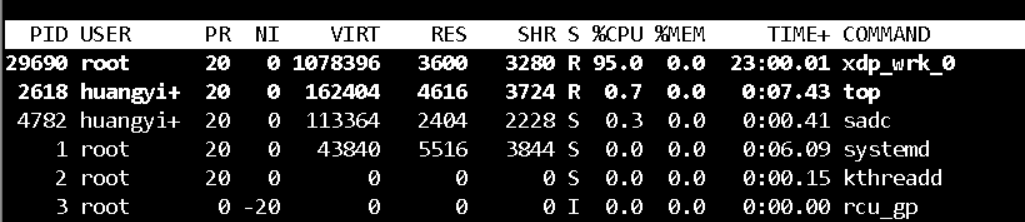


测试域名如下



第五步：观察测试过程中机器A的状态

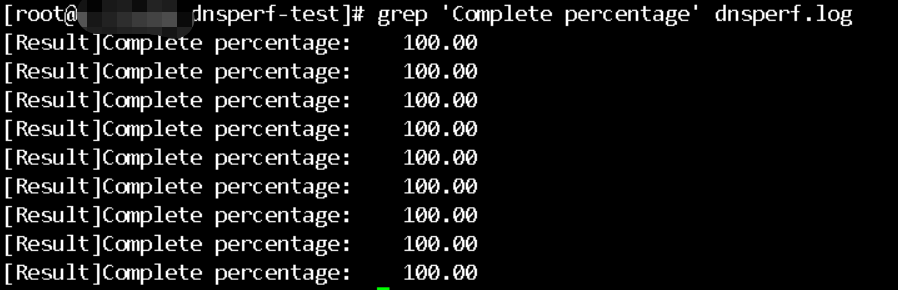




第六步：统计dnsperf的输出结果并计算总QPS

* 首先检查“Complete percentage”

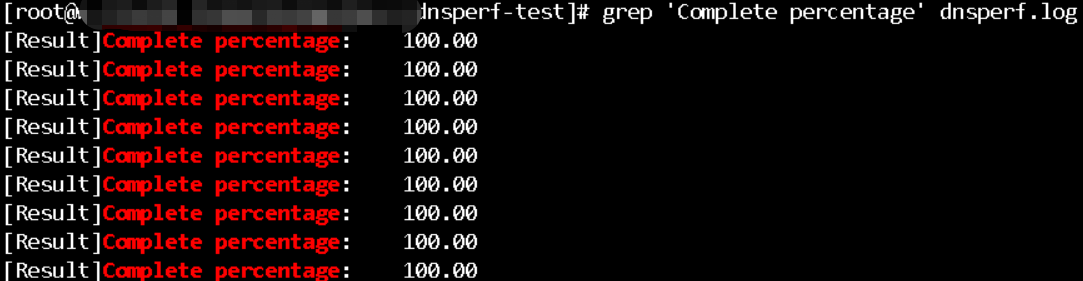
机器B



机器C



机器D



所有机器上的所有9个dnsperf都是100.00

* 然后计算总QPS

机器B



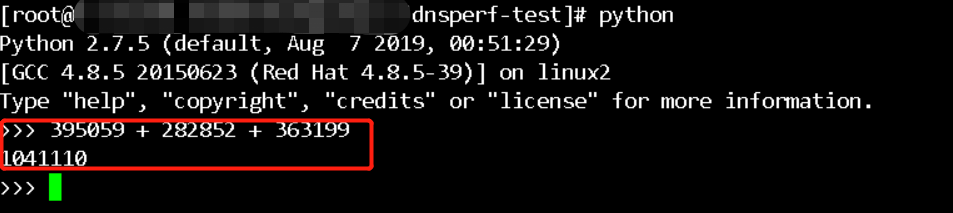
机器C



机器D



把三个机器的QPS加在一起即为总QPS



**小结:**



**此测试网卡与xdp\_wrk\_0属于同一numa节点，且在单核单网卡队列的情况下， sample/dns\_server的CPU使用率95%左右，最大可承载1041110QPS即100万+QPS，但在接下来的过程中发现了bug,这个bug会导致在分配内存时设置numa策略失败，但程序可以正常执行，numa策略设置失败时，使用的仍然是本地优先分配的策略**

情况二：

相同测试环境及相同机器配置下的修正numa设置策略失败问题再次测试。

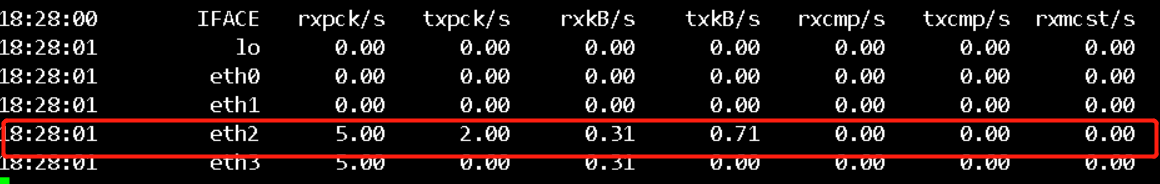
第一步：搭建及配置测试环境，与情况一完全相同

第二步: 在机器A上启动dns\_server



-i 是测试网卡eth2的IP地址

第三步：记录测试之前，机器A的状态



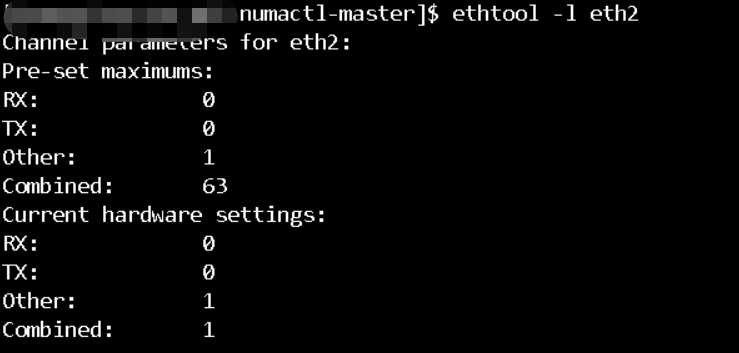


第四步：启动测试

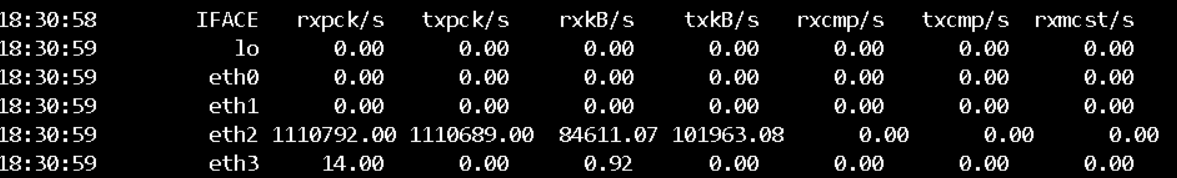
同时在机器B、C、D、上分别启动10个dnsperf并将结果保存在dnsperf.log中



第五步：观察测试过程中机器A的状态



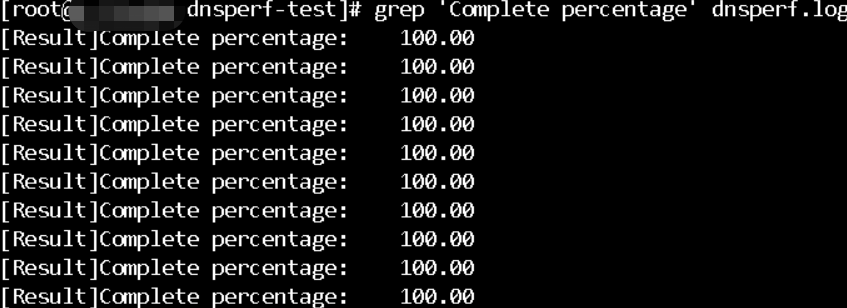




第六步：测试结束统计dnsperf的输出结果并统计QPS

* 首先检查”Complete percentage”

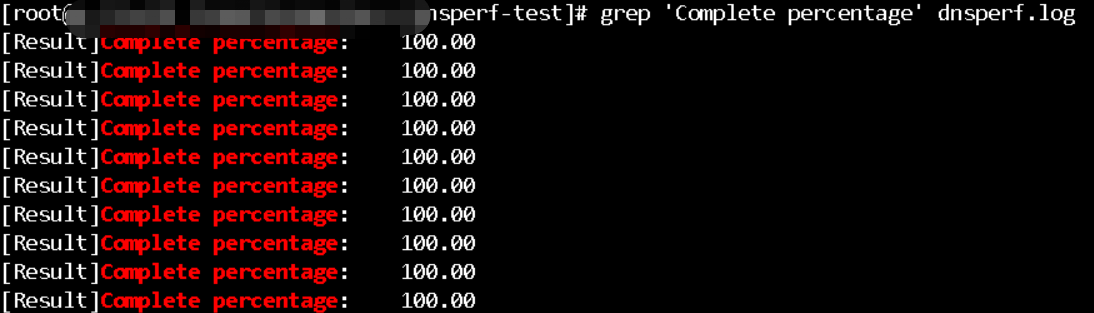
机器B



机器C



机器D



有机器上的所有10个dnsperf都是100.00

* 然后计算总QPS

机器B



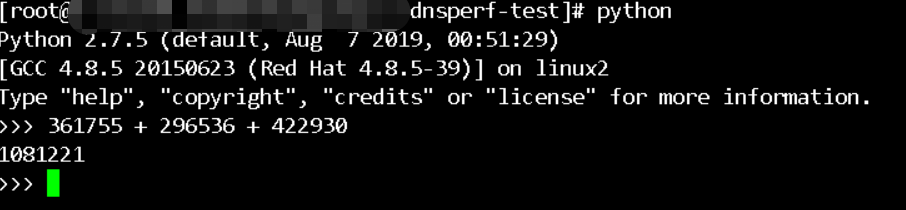
机器C



机器D



把三个机器的QPS加在一起即为总QPS



**小结：**

**在修复了numa策略设置失败的问题之后且同样是单核单网卡队列的配置下QPS是1081221,比修复前的1041110提升了4万，提升并不明显，应该是因为修复前修复后用的numa节点及CPU核与网卡的numa节点都一致造成的。**

情况三:

单核单网卡队列，内存和CPU核在numa节点1上，网卡eth2在numa节点0上

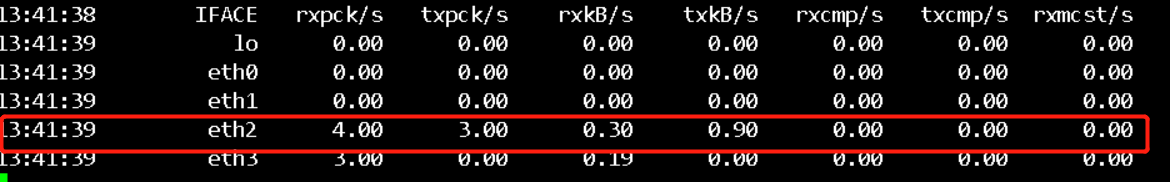
第一步：搭建及配置测试环境，与情况一完全相同

第二步: 在机器A上启动dns\_server，通过-n参数设定numa节点为1

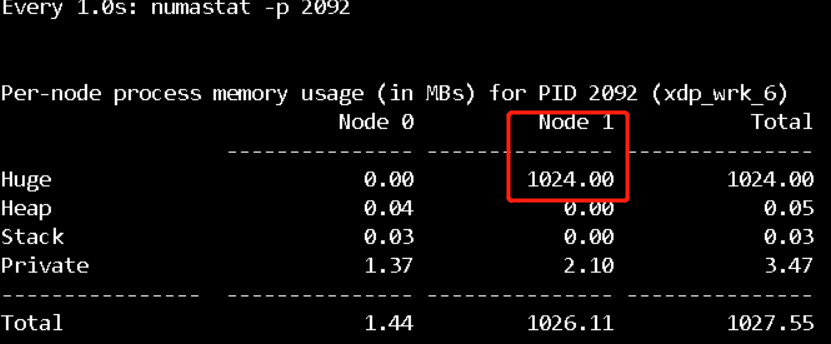


-i 是测试网卡eth2的IP地址

第三步：记录测试之前，机器A的状态





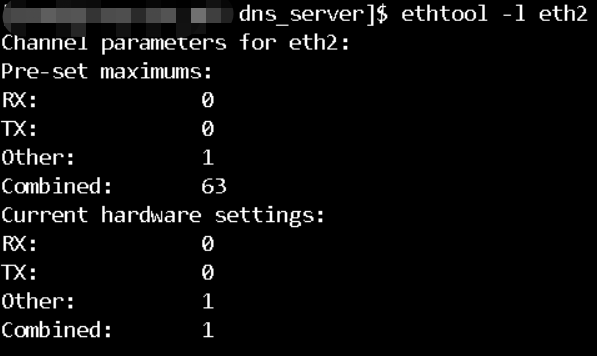


第四步：启动测试

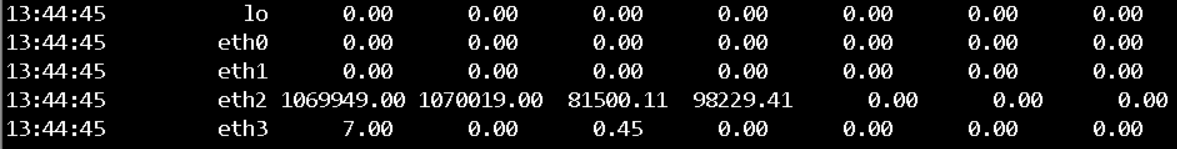
同时在机器B、C、D、上分别启动10个dnsperf并将结果保存在dnsperf.log中



第五步：观察测试过程中机器A的状态



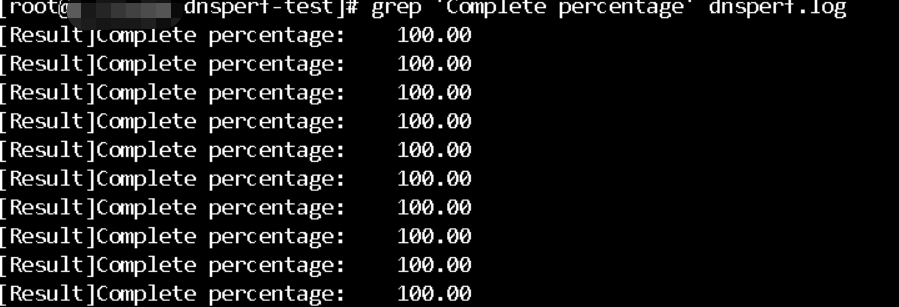




第六步：测试结束统计dnsperf的输出结果并统计QPS

* 首先检查”Complete percentage”

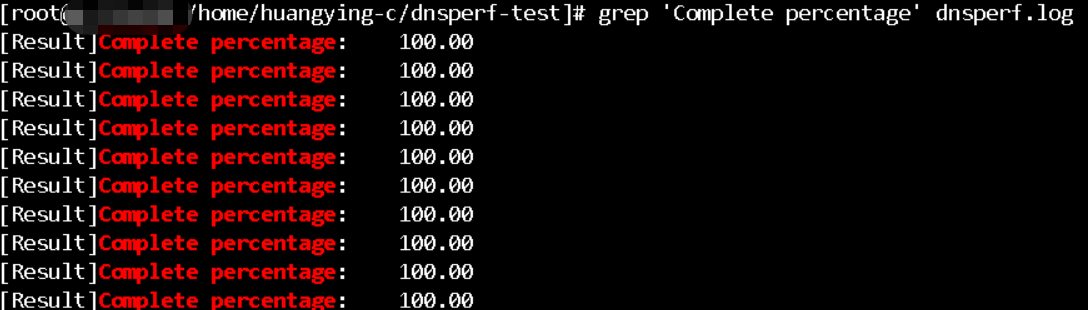
机器B



机器C



机器D



有机器上的所有10个dnsperf都是100.00

* 然后计算总QPS

机器B



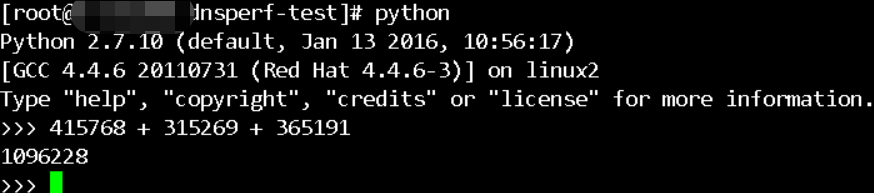
机器C



机器D



把三个机器的QPS加在一起即为总QPS



**小结：**

**在选取了CPU核、内存与网卡不同的numa节点后QPS是1096228,反而比在同一numa节点的测试结果还略高，而且CPU的占用率只有64％，后来又反复通过增加压力进行了测试，发现CPU占用率不升反而略微下降。这点与认知不符。因numa对性能的影响不是此测试的重点。有以下推测需要验证：**

* **对numa的理解不够透彻，需要再更深入的了解numa**
* **因为测试时间段不同，机房交换机在不同的负载状态下对测试的影响可能不同**
* **因为内存和cpu核在同一numa节点，网卡在另外一个numa节点上，再加上预取指令，所以才会有这样的结果**

2、单核单网卡队列性能测试总结：

**不论numa是否一致，单核单网卡队列都可以达到100万QPS，而对于此测试来说一个请求就是一个包，所以可以认为QPS（query per second）约等于PPS(packet per second)，所以简单认为使用了xdp\_packet的sample/dns\_server的PPS是100万**

四、2核2网卡队列

情况一:

2核2网卡队列，内存和CPU核在numa节点0上，网卡eth2在numa节点0上

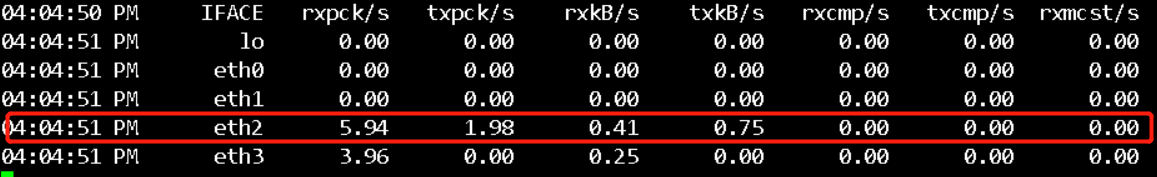
第一步：搭建及配置测试环境，与单核但队列情况一完全相同,另外需要增加三台打压机器分别为E、F、G。

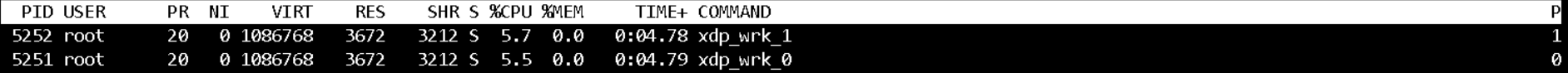
第二步: 在机器A上启动dns\_server，通过-Q参数设定为2

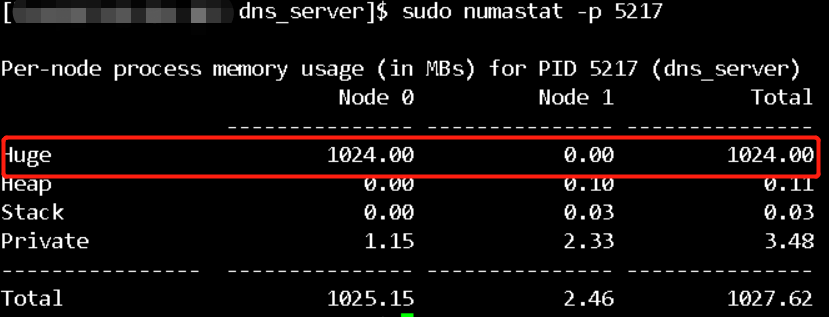


-i 是测试网卡eth2的IP地址

第三步：记录测试之前，机器A的状态

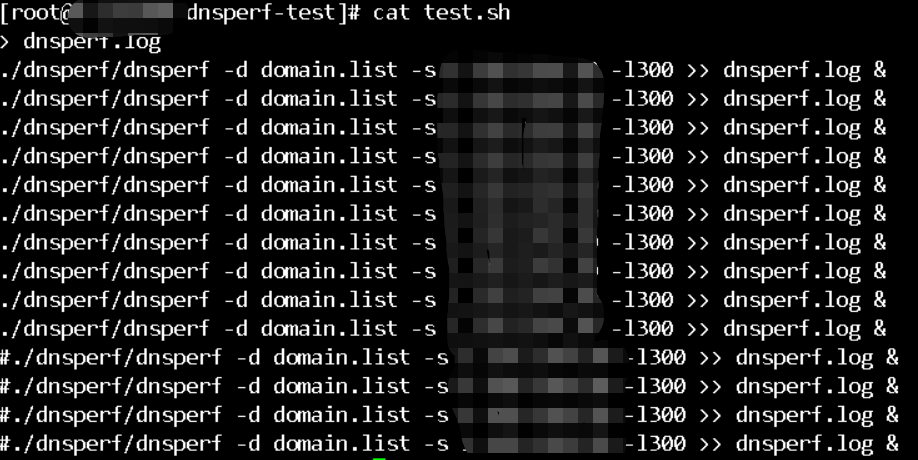






第四步：启动测试

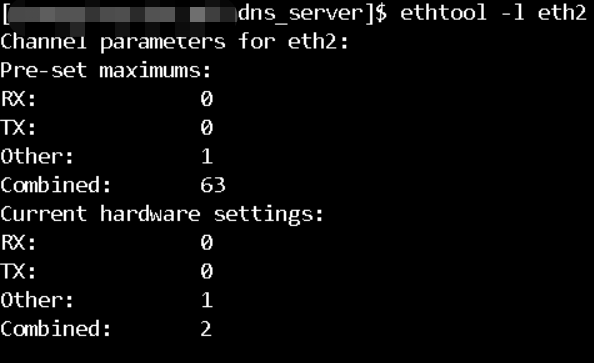
同时在机器B、C、D、E、F、上分别启动10个dnsperf并将结果保存在dnsperf.log中



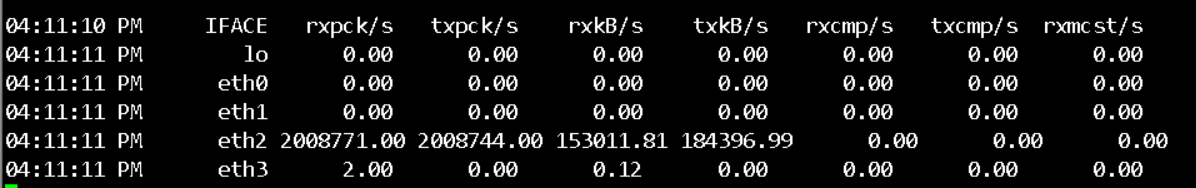
在机器G上启动5个dnsperf



第五步：观察测试过程中机器A的状态



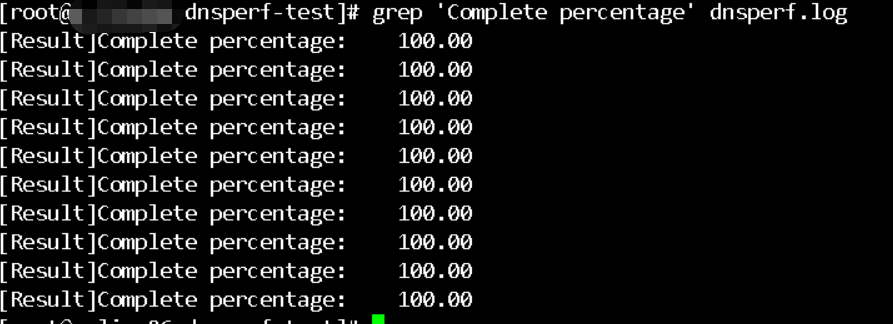




第六步：测试结束统计dnsperf的输出结果并统计QPS

* 首先检查”Complete percentage”

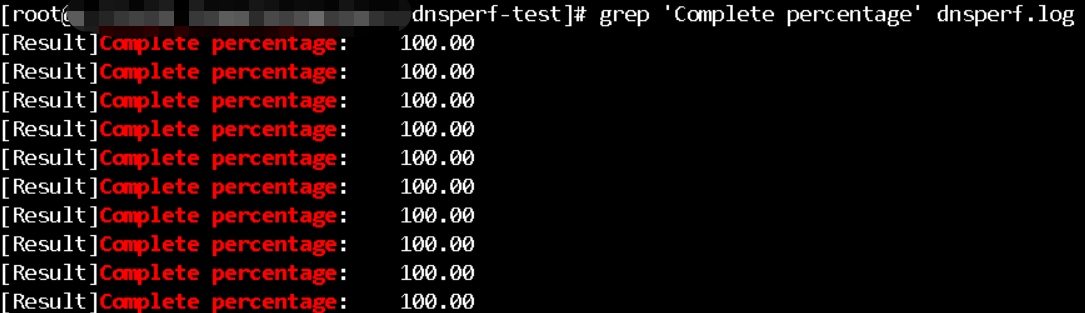
机器B



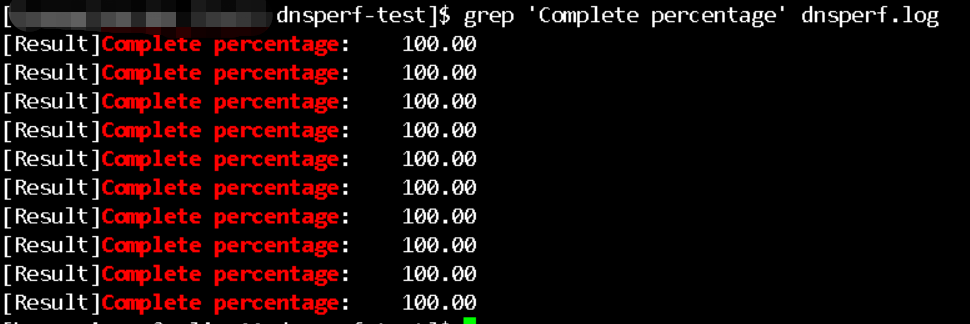
机器C



机器D



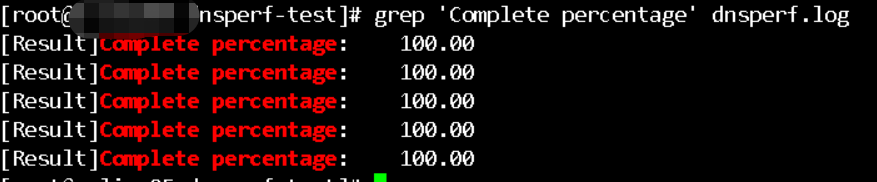
机器E



机器F



机器G



所有机器上的”Complete percentage”都是100.00，因为再增加dnsperf进程”Complete percentage”就会出现99.99的情况，所以在机器G上只有5个dnsperf进程同时运行

* 然后计算总QPS

机器B



机器C



机器D



机器E



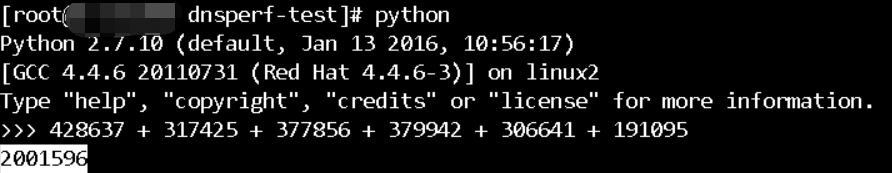
机器F



机器G



把6个机器的QPS加在一起即为总QPS



**小结：**

**在设置成2个CPU核，两个网卡队列后的总QPS为2001596,相对1个CPU核1个网卡队列的情况几乎是线性增长，即性能随CPU核及网卡队列的增加而成倍增长。因为线程模型使用的是run to complete模型，同时需要注意的是数据包的四元组在经过网卡RSS计算出的hash值要足够均匀，可以让数据包均匀的分布在CPU核上。**

情况二：

2核2网卡队列，内存和CPU核在numa节点1上，网卡eth2在numa节点0上

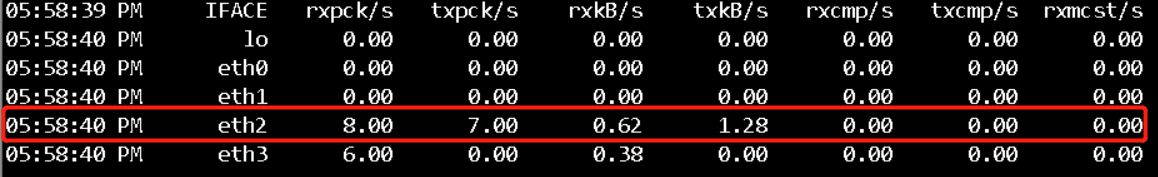
第一步：搭建及配置测试环境，与情况一完全相同

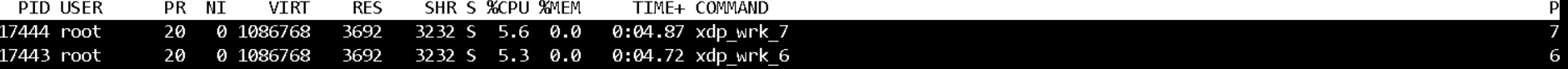
第二步: 在机器A上启动dns\_server，通过-Q参数设定为2

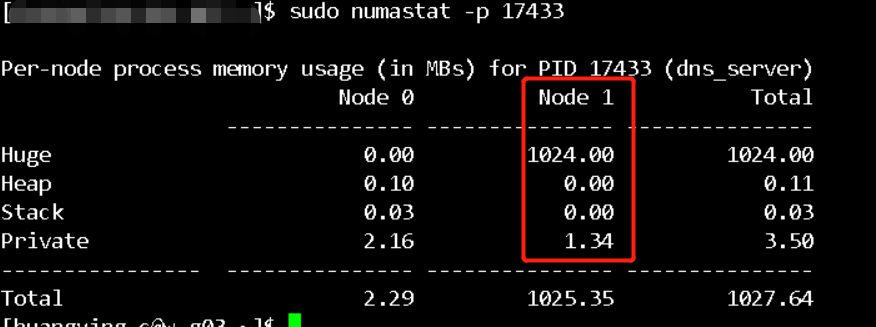


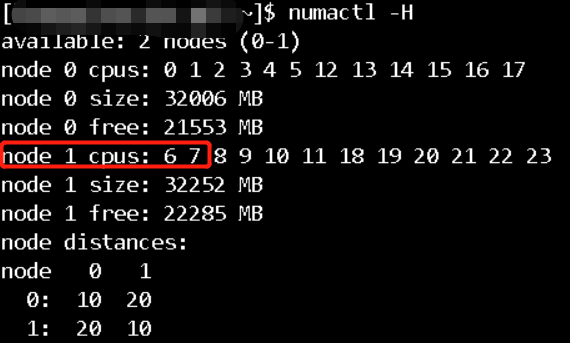
-i 是测试网卡eth2的IP地址，使用-n参数指定numa节点

第三步：记录测试之前，机器A的状态







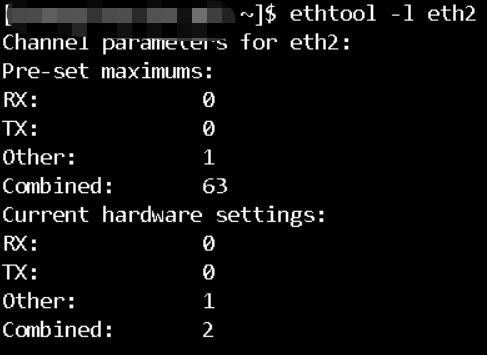




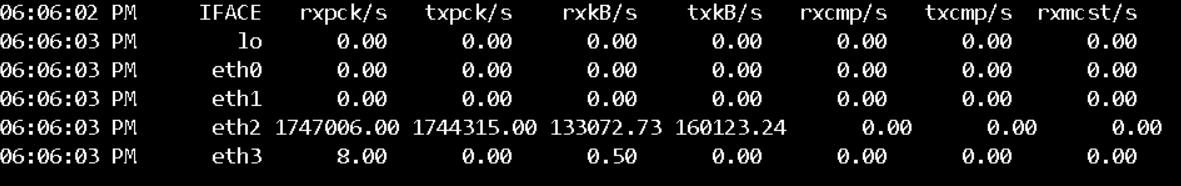
第四步：启动测试

和情况一一样，同时在机器B、C、D、E、F、上分别启动10个dnsperf并将结果保存在dnsperf.log中在机器G上启动5个dnsperf

第五步：观察测试过程中机器A的状态



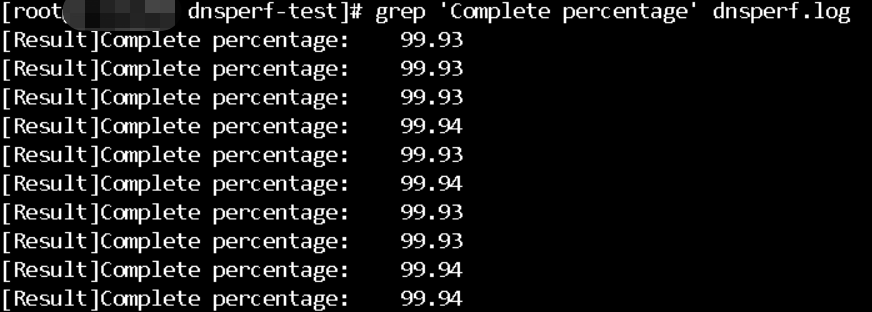




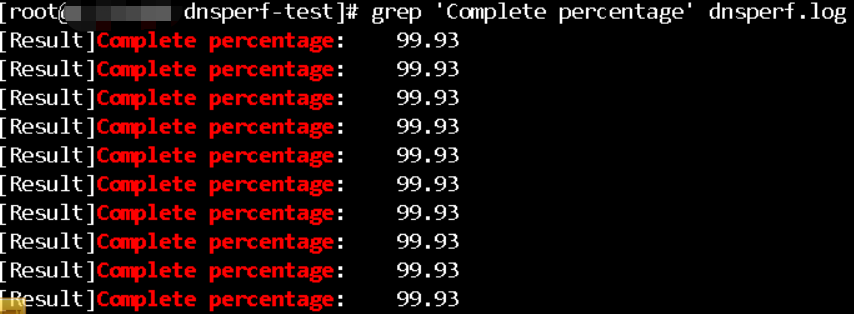
第六步：测试结束统计dnsperf的输出结果并统计QPS

* 首先检查”Complete percentage”

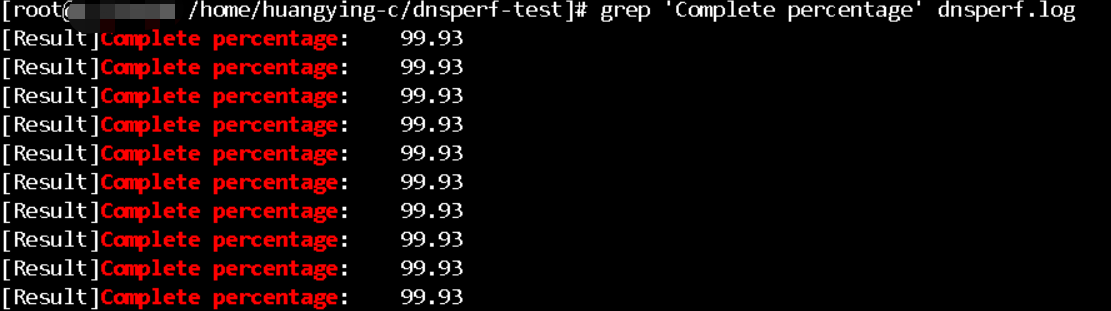
机器B



机器C



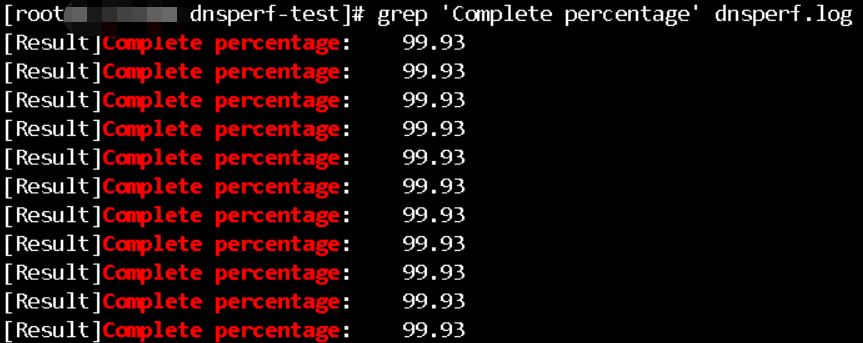
机器D



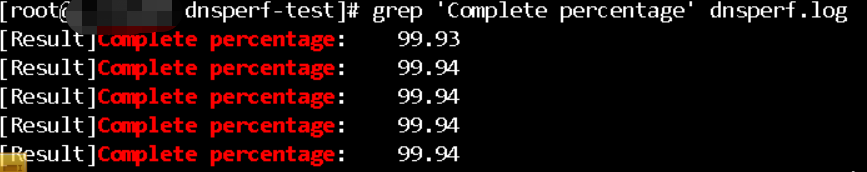
机器E



机器F



机器G



所有机器上的”Complete percentage”都是100.00，因为再增加dnsperf进程”Complete percentage”就会出现99.99的情况，所以在机器G上只有5个dnsperf进程同时运行

* 然后计算总QPS

机器B



机器C



机器D



机器E



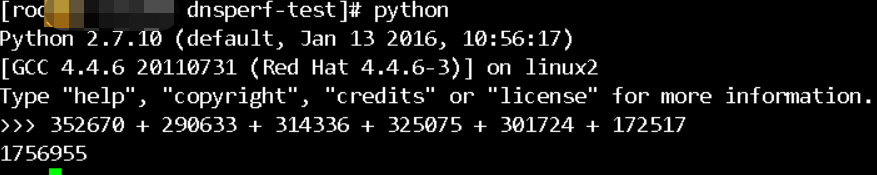
机器F



机器G



把6个机器的QPS加在一起即为总QPS



**小结：**

**在设置成2个CPU核，两个网卡队列且跨numa访问的情况下，同样的压力性能有所下降，QPS为1756955。而且线程负载出现倾斜的情况，相同环境，相同压力，相同的测试机器，相同的网卡RSS算法，为什么会出现CPU负载不均匀的情况？性能下降不仅受numa影响，而且和CPU负载不均匀有很大关系。**

**numa的本地虽然能提高性能但无法充分的利用全部CPU核，之后将考虑测试交错内存策略的numa的方式，力求充分利用CPU。**

五、3核3网卡队列

情况一:

3核3网卡队列，内存和CPU核在numa节点0上，网卡eth2在numa节点0上

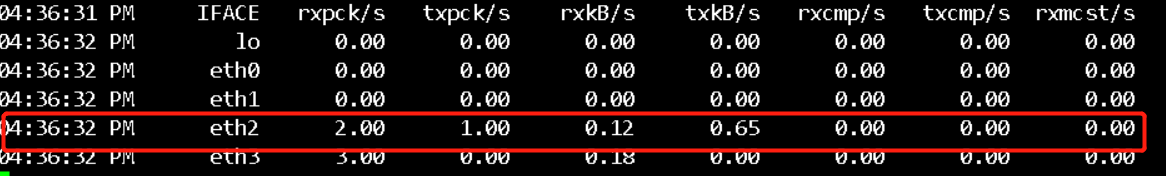
第一步：搭建及配置测试环境，与2核2队列情况一完全相同,另外需要增加1台打压机器为H。

第二步: 在机器A上启动dns\_server，通过-Q参数设定为3

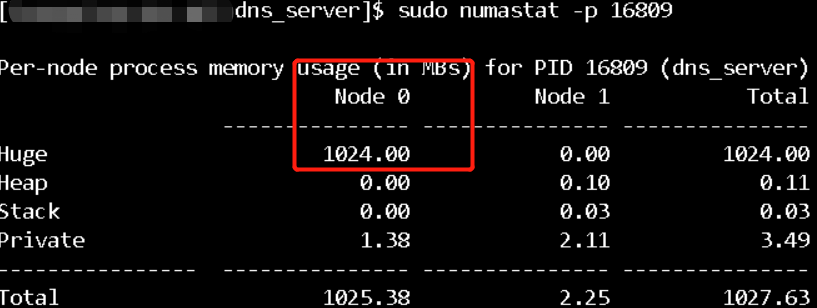


-i 是测试网卡eth2的IP地址

第三步：记录测试之前，机器A的状态

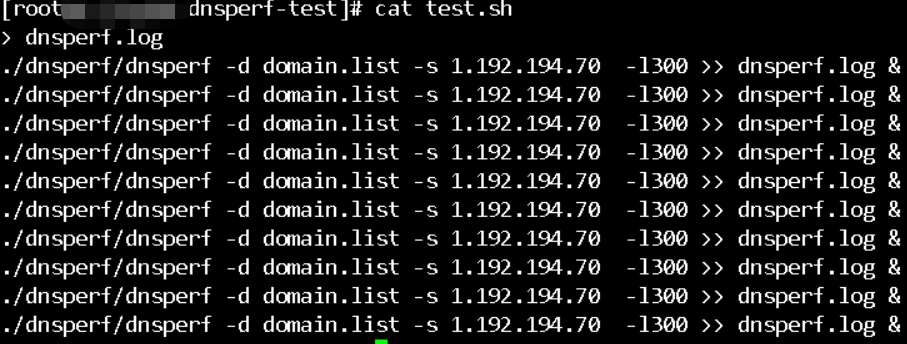






第四步：启动测试

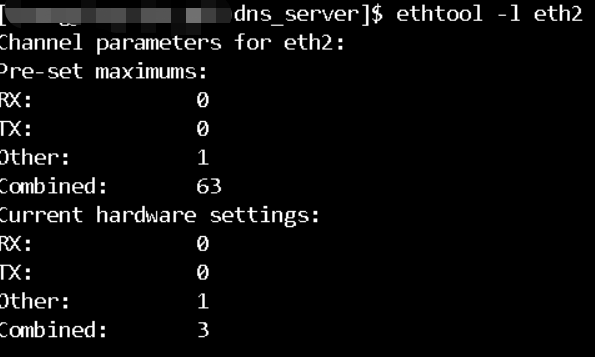
同时在机器B、C、D、E、F、H上分别启动了10个dnsperf，

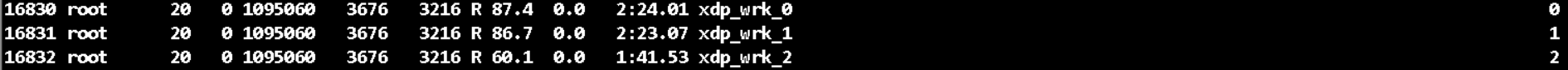


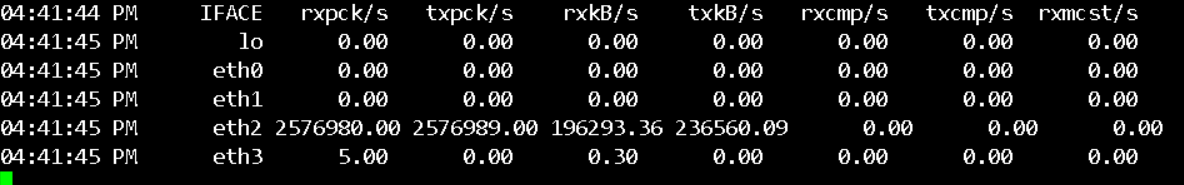
机器G上分别启动8个dnsperf



第五步：观察测试过程中机器A的状态



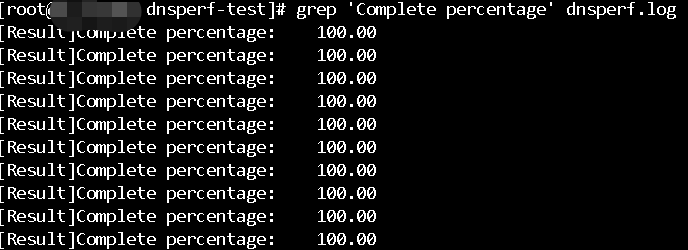




第六步：测试结束统计dnsperf的输出结果并统计QPS

* 首先检查”Complete percentage”

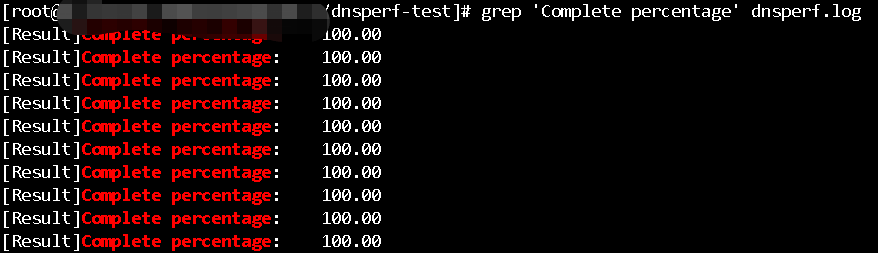
机器B



机器C



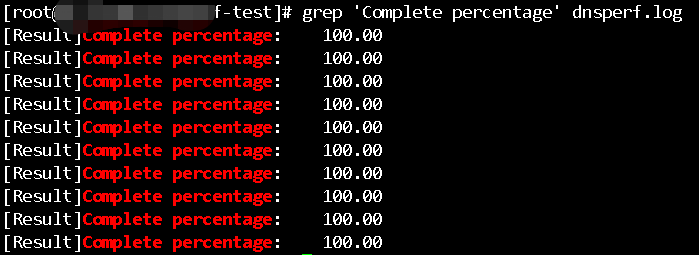
机器D



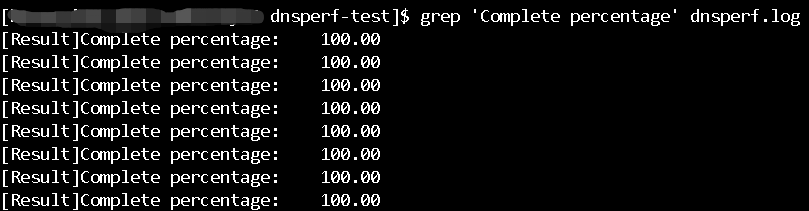
机器E



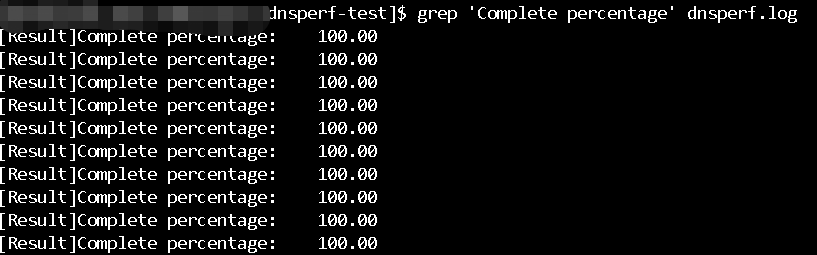
机器F



机器G



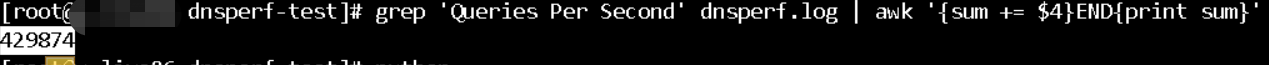
机器H



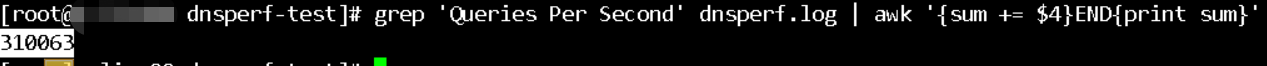
所有机器上的”Complete percentage”都是100.00，因为再增加dnsperf进程”Complete percentage”就会出现99.99的情况，所以在机器G上只有8个dnsperf进程同时运行

* 然后计算总QPS

机器B



机器C



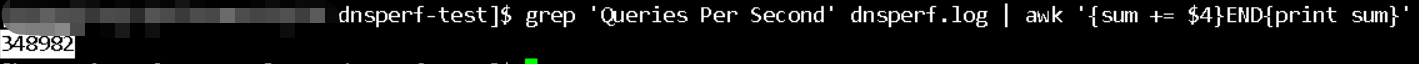
机器D



机器E



机器F



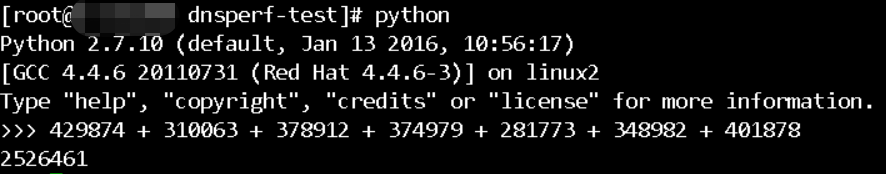
机器G



机器H



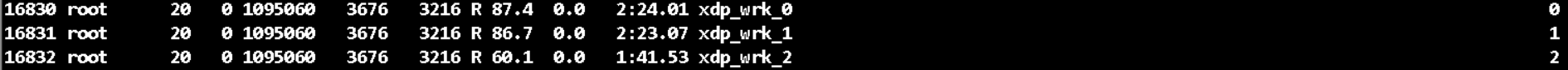
把7个机器的QPS加在一起即为总QPS



**小结：**

**在设置成3个CPU核，3个网卡队列后的总QPS为2526461，但是有如下情况需要说明：**

* **cpu负载不均衡的情况下的测试结果，由打压过程中CPU的截图可以看到,最高的CPU占了87%说明这个CPU核没有跑满。**



* **最低的cpu负载只有60%，说明这个CPU核也没有跑满**
* **CPU占用最低的和最高的相差27%**

**以上可以证明3个CPU核3网卡队列时，压力并没有打满即并不是sample/dns\_server的最大负载能力，造成负载倾斜的原因推测是dnsperf所发的数据包的源端口不够散列造成的。**

**如果假设负载是均匀的，且都按87%计算的话，那么应该在QPS的基础上再加上27%，也就是2526461 + 2526461 \* 27% = 3208605.47，这个结果也是符合现象增长的**

五、总结

基于xdp\_packet的sample/dns\_server的性能在使用本地numa策略的条件下，当增加CPU核的时候，是可以满足线性增长的。在此测试环境中执行sample/dns\_server的机器上属于同一numa节点的cpu核为12个。根据单CPU核单网卡队列的测试结果QPS是1081221。

按照线性增长的话同一nuam节点的12个核的QPS应该是1081221 \* 12 = 12974652，也就是可以达到1000万qps。7台机器已经是此测试可以使用的极限资源了，无法再增加打压机器了，所以测试暂时告一段落。