

QUESTION: What name did you give to your Azure VM?

ANSWER:

IP address of the VM ubuntu008.westus2.cloudapp.azure.com

Username: pava

pava@ubuntu008.westus2.cloudapp.azure.com

QUESTION: When you ran 'docker --version' what did you get as output?

ANSWER:

Docker version 18.09.7, build 2d0083d

* A Quick Tour of the Linux Shell

QUESTION: When you run 'apropos delete' why do you not see the man page for 'rm' listed? (Hint: look at the man page for 'apropos' to see what it claims it does, and then look at the contents of the man page for 'rm')

ANSWER: Apropos is used to search the name sections of all manual pages, it is often the wrapper for man -k command.

- 1) rm - removes files or directories.
- 2) Apropos delete it searches for delete sections of manual pages for string "delete". Outputs a list containing variety of delete operations that can be performed.
- 3) Since rm does not delete directories.
- 4) Files removed using rm can be recovered.
- 5) Hence apropos delete does not include rm.

QUESTION: Try to list the contents of the directory '/var/run/docker' using the command 'ls /var/run/docker'. Why does it fail? How can you make it not fail? What does your fixed command display?

ANSWER: when /var/run/docker command is used it displays permission is denied on the command prompt. This is because if the file set on content are identical from the host as well as container. It can be solved by changing ownership permission using following commands.

\$ sudo chown var/run/docker

```
pava@UbuntuServer008:/$ sudo chown pava /var/run/docker
pava@UbuntuServer008:/$ cd /var/run/docker
pava@UbuntuServer008:/var/run/docker$
```

```
pava@UbuntuServer008:/var/run/docker$ ls -lF
total 0
drw----- 2 root root 60 Oct 6 04:07 libnetwork/
srwxr-xr-x 1 root root 0 Oct 6 04:06 metrics.sock=
drwx----- 2 root root 40 Oct 6 04:06 plugins/
drwx----- 2 root root 40 Oct 6 04:07 swarm/
pava@UbuntuServer008:/var/run/docker$
```

QUESTION: Run 'htop' and look at the output. Consult the (non-living) resource of your choice to find out what the "Load average" indicator means. What do the numbers reported next to "Load average" mean? What source did you consult to learn this?

```
pava — pava@UbuntuServer008: ~ — ssh pava@ubuntu008.westus2.cloudapp.azure.com — 181x54

1  [ ] 0.7% Tasks: 33, 40 thr: 1 running
2  [ ] 0.0% Load average: 0.00 0.00 0.00
3  [ ] 0.0% Uptime: 00:18:53
4  [ ] 0.7%
Mem[||||] 377M/15.6G
Swp[ ] 0K/0K

      PID USER      PR  NI  VIRT  RES  SHR  S CPU% MEM%   TIME+  Command
2444 pava        20   0 37972 6040 3688  S  0.0  0.0  0:03.07 htop
1614 root        20   0 231M 24328 19216  S  0.0  0.1  0:05.31 python3 -u bin/WALinuxAgent-2.2.42-py2.7.egg --run-exthandlers
1 root        20   0 77964 9284 6748  S  0.0  0.1  0:03.07 /sbin/init
510 root       19  -1 94848 11688 10980  S  0.0  0.1  0:00.40 /lib/systemd/systemd-journald
515 root       20   0 12016 2648 1868  S  0.0  0.0  0:00.51 /usr/lib/linux-tools/5.0.0-1020-azure/hv_kv_daemon -n
525 root       20   0 97704 1812 1640  S  0.0  0.0  0:00.00 /sbin/lvmtdad -f
534 root       20   0 44324 6272 3092  S  0.0  0.0  0:00.78 /lib/systemd/systemd-udevd
868 systemd-t  20   0 1308M 3240 2708  S  0.0  0.0  0:00.00 /lib/systemd/systemd-timesyncd
258 systemd-t  20   0 1308M 3240 2708  S  0.0  0.0  0:00.00 /lib/systemd/systemd-timesyncd
964 systemd-n  20   0 80164 8884 5148  S  0.0  0.0  0:00.03 /lib/systemd/systemd-networkd
992 systemd-r  20   0 70624 6276 4716  S  0.0  0.0  0:00.22 /lib/systemd/systemd-resolved
1375 syslogd    20   0 261M 4520 3728  S  0.0  0.0  0:00.01 /usr/sbin/rsyslogd -n
1376 syslogd    20   0 261M 4520 3728  S  0.0  0.0  0:00.00 /usr/sbin/rsyslogd -n
1377 syslogd    20   0 261M 4520 3728  S  0.0  0.0  0:00.03 /usr/sbin/rsyslogd -n
1225 syslogd    20   0 261M 4520 3728  S  0.0  0.0  0:00.05 /usr/sbin/rsyslogd -n
1312 root        20   0 95536 1556 1420  S  0.0  0.0  0:00.00 /usr/bin/lxcfs /var/lib/lxcfs/
1316 root        20   0 95536 1556 1420  S  0.0  0.0  0:00.00 /usr/bin/lxcfs /var/lib/lxcfs/
1238 root        20   0 95536 1556 1420  S  0.0  0.0  0:00.01 /usr/bin/lxcfs /var/lib/lxcfs/
1278 root        20   0 107M 2060 1844  S  0.0  0.0  0:00.00 /usr/sbin/irqbalance --foreground
1242 root        20   0 107M 2060 1844  S  0.0  0.0  0:00.04 /usr/sbin/irqbalance --foreground
1563 root        20   0 166M 17168 9372  S  0.0  0.1  0:00.00 /usr/bin/python3 /usr/bin/networkd-dispatcher --run-startup-triggers
1245 root        20   0 166M 17168 9372  S  0.0  0.1  0:00.09 /usr/bin/python3 /usr/bin/networkd-dispatcher --run-startup-triggers
1250 root        20   0 4812 772 708  S  0.0  0.0  0:00.00 /usr/lib/linux-tools/5.0.0-1020-azure/hv_vss_daemon -n
1261 daemon      20   0 28328 2352 2148  S  0.0  0.0  0:00.00 /usr/sbin/atd -f
1267 root        20   0 72800 20740 8748  S  0.0  0.1  0:00.24 /usr/bin/python3 -u /usr/sbin/waagent --daemon
1335 root        20   0 281M 6940 6064  S  0.0  0.0  0:00.02 /usr/lib/accounts-service/accounts-daemon
1464 root        20   0 281M 6940 6064  S  0.0  0.0  0:00.00 /usr/lib/accounts-service/accounts-daemon
1277 root        20   0 281M 6940 6064  S  0.0  0.0  0:00.04 /usr/lib/accounts-service/accounts-daemon
1279 root        20   0 70576 6164 5468  S  0.0  0.0  0:00.04 /lib/systemd/systemd-logind
1287 messagebus  20   0 50040 4452 3772  S  0.0  0.0  0:00.04 /usr/bin/dbus-daemon --system --address=systemd: --nofork --nopidfile --systemd-activation --syslog-only
1598 root        20   0 183M 20176 12268  S  0.0  0.1  0:00.00 /usr/bin/python3 /usr/share/unattended-upgrades/unattended-upgrade-shutdown --wait-for-signal
1455 root        20   0 183M 20176 12268  S  0.0  0.1  0:00.08 /usr/bin/python3 /usr/share/unattended-upgrades/unattended-upgrade-shutdown --wait-for-signal
1462 root        20   0 31744 3212 2920  S  0.0  0.0  0:00.00 /usr/sbin/cron -f
1703 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.07 /usr/bin/containerd
1708 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.00 /usr/bin/containerd
1709 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.00 /usr/bin/containerd
1710 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.00 /usr/bin/containerd
1819 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.03 /usr/bin/containerd
1843 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.07 /usr/bin/containerd
1844 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.06 /usr/bin/containerd
1845 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.00 /usr/bin/containerd
1898 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.04 /usr/bin/containerd
1899 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.00 /usr/bin/containerd
1902 root        20   0 1004M 43056 22636  S  0.0  0.3  0:00.02 /usr/bin/containerd
```

ANSWER: htop is a tool to see how long the system is running. It gives information about tasks, load average, uptime, Information on process etc.

Load average is the average system load over a period of time. First number is 1min load average, second number is 5 min load average and third number is 15 min load average.

Used resources

<https://www.deonsworld.co.za/2012/12/20/understanding-and-using-htop-monitor-system-resources/>

* Numbers Everyone Should Know in Human Terms

QUESTION: For each of the numbers below, perform the alien/computer time scaling exercise discussed in the assignment instructions and report the number in some reasonable units.

ANSWERS:

L1 cache reference: ___0.5ns___
Branch mispredict: ___5ns___
L2 cache reference: ___7ns___
Mutex lock/unlock: ___100ns___
Main memory reference: ___100ns___
Send 2K bytes over 1 Gbps network: ___20,000ns___
Read 1MB sequentially from memory: ___250,000ns___
Round trip within same datacenter: ___500,000ns___
Disk seek: 10,000,000ns___
Read 1MB sequentially from network: 10,000,000 ns
Read 1MB sequentially from disk: 30,000,000 ns
Send package CA -> Netherlands -> CA: 150,000,000 ns

Reference: <https://stackoverflow.com/questions/4087280/approximate-cost-to-access-various-caches-and-main-memory>

QUESTION: Consult a reference for the metric prefixes and give the power of ten associated with each prefix.

ANSWER:

milli: 10exp(-3)
micro: 10exp(-6)
nano: 10exp(-9)
femto: 10exp(-15)

kilo: __10exp(3)____
mega: __10exp(6)____
tera: __10exp(12)____
peta: __10exp(15)____
yotta: __10exp(24)____
zetta: __10exp(21)____

QUESTION: Your colleague, Davey Datascience, has developed a new MapReduce-based algorithm that involves many iterated map and reduce phases performed against network-attached disk storage. Your company, in the finest tradition of modern surveillance capitalism has accumulated vast troves of consumer data that your CEO, Barry Boondoggler, is sure must contain extractable insights that will buy him a much less shabby weekend cottage in the San Juans. Davey estimates the job will take 1.44×10^{17} seconds to run on your Hadoop cluster. Is this a reasonable runtime? (Hint: How many years will the job take? What "system" that we inhabit is currently about that old?)

ANSWER: Its performance is slow. It will take 4×10^{21} years to complete the job.

* Questions on Readings for the Week

Note that for some of these questions there is one, true "correct" answer. They're intended to provoke critical thought and sensitize your awareness.

QUESTION: Answer honestly. Can you credibly imagine Keshav's paper reading tips helping you deal with technical information in your day job? Why or why not? What elements of his strategy resemble or diverge from how you consume written technical material in your regular work life?

ANSWER:

The paper is good on advising how to read research papers on first, second, third pass. It's similar to the format of SAT or GRE RC passages. It's easy to get a brief understanding by reading the abstract conclusion and figures. This tip is useful in daily reading.

QUESTION: In MapReduce: A major step backwards, Stonebraker and DeWitt list five numbered grievances the database community has with the MapReduce paradigm.

ANSWER:

- 1) MapReduce is a step backwards in database access
 - 2) Is poor in implementation
 - 3) Not novel
 - 4) Missing features
 - 5) Is incompatible with the DBMS tools
-

QUESTION: What is the key difference between the "relational view" and the "Codasyl view" of how we should write programs to process data?

ANSWER:

Relational View: By stating what you want- rather than presenting an algorithm on how to get it. High-level languages and relational systems are used and is easy to understand.

Codasyl View: Presented an algorithm for data access. Criticized for using assembly language for DBMS access.

QUESTION: Of the features Stonebraker asserts that MapReduce is missing, which absences do you think would have caused you the most pain in the sort of data processing work you've done up to this point in your life? If you haven't yet done much, which sound like they're going to be the most troublesome to you?

ANSWER:

Low performance to handle large data and incompatibility with DBMS tools

QUESTION: In What Goes Around Comes Around, be sure to read chapter "XI Full Circle closely" after skimming the rest of the paper for flavor and rough history. Why do we worry about avoiding repeating history?

ANSWER:

Additional data models were proposed none of which gained substantial market traction, because they failed to offer leverage in exchange for the increased complexity

Ex: In the past XML was compared with relational or DBMS. Complex model was compared to simple models. It's wise to compare models with same independence and complexity to avoid repeating history.

QUESTION: According to What Goes Around Comes Around, what are the noticeably new concepts to have come along in the last 20 years?

ANSWER:

Code in the data base (from the OR camp)

Schema last (from the semi-structured data camp)