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

ANSWER:
 IP address of the VM ubuntuserver008.westus2.cloudapp.azure.com
Username: pava
pava@ubuntuserver008.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?

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.

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 x 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 teh "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)