Quiz 1 - Results

Attempt 1 of 1

Written Sep 22, 2025 3:13 PM - Sep 22, 2025 3:16 PM

Attempt Score **22 / 23 - A**

Overall Grade (Highest Attempt) **22 / 23 - A**

Question 1 1 / 1 point

Scalable System Software reduces operating system interrupts by stripping down OS running on compute nodes.

- ✔ (●) True
- ( ) False

Question 2 1 / 1 point

Which of the following is true?
(⊆ denotes subset. A⊆B implies A is subset of B)

- ( ) Artificial Intelligence ⊆ Machine Learning ⊆ Deep Learning
- ( ) Artificial Intelligence ⊆ Deep Learning ⊆ Machine Learning
- ( ) Machine Learning ⊆ Deep Learning ⊆ Artificial Intelligence

✔◉ Deep Learning ⊆ Machine Learning ⊆ Artificial Intelligence

Question 3 1 / 1 point

What are the important features of high performance computing architecture?

✔☑ Power

✔☐ Sequentiality

✔☑ Reliability

✔☑ Speed

✔☑ Efficiency

Question 4 1 / 1 point

What changes have we seen in the world of machine learning due to the advent of high performance computing?

✔☑ Homogeneous to Heterogeneous computing

✔☑ Ability to handle and use terabytes of data

✔☑ Moved to InfiBand like computer networking communications from standard networks

Question 5 1 / 1 point

With strong scaling, as we add more and more compute resources we can continue to get speedup since the amount of work per compute resource gets smaller and smaller.

◯ True

✔ ◉ False

Question 6 1 / 1 point

You need to train a machine learning model. When the dataset is small you can easily train on one compute node in a reasonable amount of time. When the dataset is larger training on one node is very time consuming.

Your friend suggests using multiple nodes for training with larger dataset by dividing the dataset equally among those nodes. When you did this, you were able to train using a larger dataset in roughly the same time as training using the smaller dataset on a single node. This is an example of _____ scaling.

✔ ◉ weak

○ strong

○ hybrid

Question 7 1.5 / 1.5 points

The list below shows the time (in sec) to complete 10000 floating point operations on an HPC compute node in 10 different runs.

12, 13, 10, 12.5, 11, 11.25, 10.25, 14, 13.75, 14.25

What is approximately the average throughput using the harmonic mean of throughputs obtained for individual runs?

○ 82 FLOPS

✔ ◉ 820 FLOPS

○ 8.2 FLOPS

○ 8200 FLOPS

Question 8 1 / 1 point

Show below is the output of a test program ran with time command on a linux machine.

```
(base) root@ffl-robust-robust2:~# time ./a.out

real    0m2.376s
user    0m2.372s
sys     0m0.004s
```

What is the difference between CPU time and total elapsed time ?

- ○ 2.372 sec
- ○ 0.004 sec
- ○ 2.368 sec
- ✓ ◉ 0.000 sec

Question 9 1.5 / 1.5 points

Consider a program using parallel processing and running on 5 CPUs in parallel. The total CPU time is 800 secs. What is the total elapsed time assuming the work is evenly distributed on each CPU and no wait is involved for I/O or other resources.

- ○ 4000 secs
- ✓ ◉ 160 secs
- ○ 800 secs
- ○ 805 secs

Question 10 2 / 2 points

You need to create an application to process a database of customers for a bank and identify the top 1000 customers who can be targeted for marketing a new investment fund.

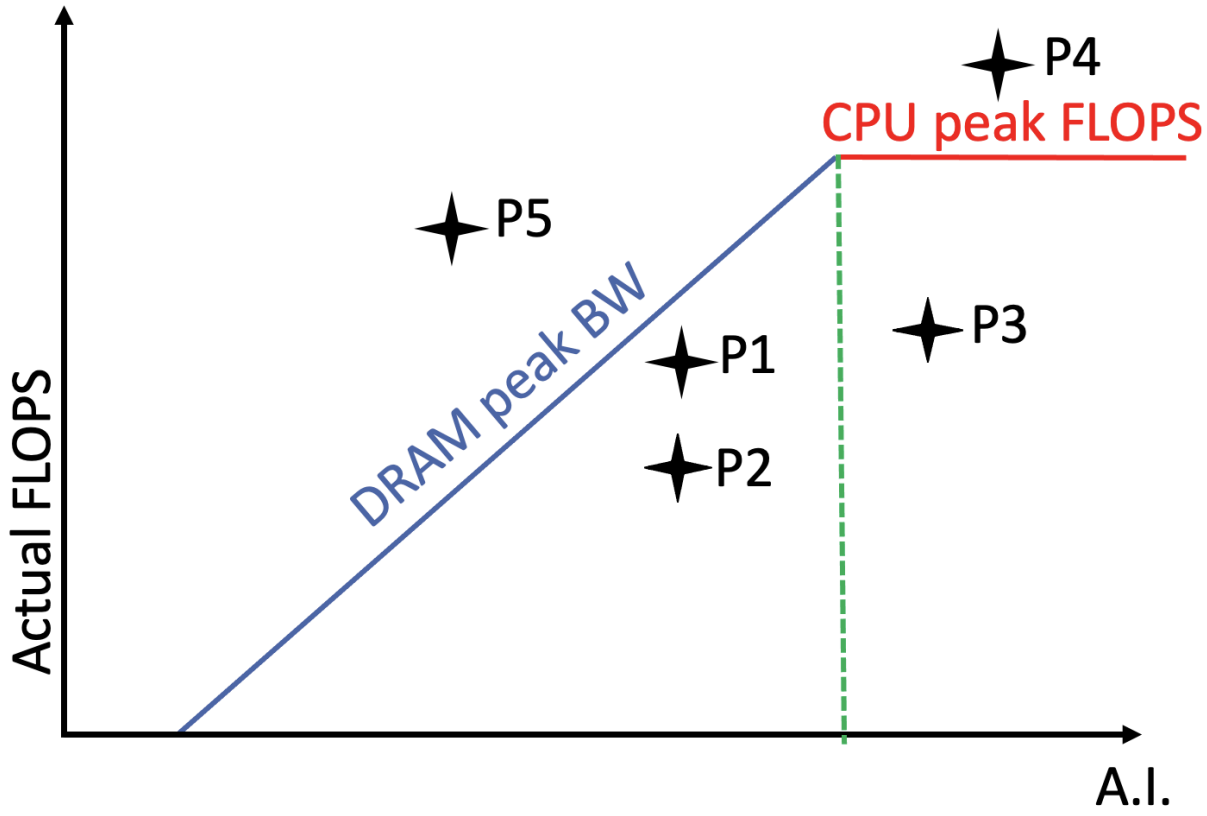
The application should query the database, process the query results, present the list of customers as an excel sheet with charts showing distribution of different statistics.

You quickly wrote the application code and handed it to the performance testing team. The team identified that it takes around 15 minutes to run your code end-to-end. Further profiling revealed that 30% of the time is spent in running the database query, 50% in processing the query results, and the remaining 20% in creating the excel sheet. The product team says that the code will only be deployed if the runtime is brought down to 9 mins. Suppose you target to optimize the code to process the query results. How much speedup is needed in this part of the code so that you can meet the runtime target of 9 mins?

- ○ 0.8x
- ○ It cannot be determined from the information provided
- ○ 1.67x
- ✓ ◉ 5x
- ○ 4x

Question 11 2 / 2 points

Consider 5 different codes with measured performance marked by P1, P2, P3, P4, and P5 in the roofline performance model shown below:



What can be inferred from this chart? Select all that apply.

- [x] P4 is not feasible with current CPU configuration in the system
- [x] P1 and P2 are compute-bound but P3 and P4 are memory-bound
- [x] P1 and P2 are memory- bound and P3 is compute-bound.
- [x] P5 is not feasible with current DRAM in the system

Question 12 1 / 2 points

Cache blocking and SIMD are two techniques to improve software performance. Select all that is true about these techniques.

- [ ] If we enable SIMD we may see an increase in FLOPS but the Arithmetic intensity remains unchanged.
- [x]

If we enable SIMD we may increase the Arithmetic intensity.

If we use cache blocking we may increase the Arithmetic intensity.

If we use cache blocking we may see an increase in FLOPS but the Arithmetic intensity remains unchanged.

Question 13 2 / 2 points

Give the following code:

```
for(k=1;k<N;k++){
   for(j=1;j<N;j++){
      for(i=1;i<N;i++){
         int ijk = i + j * jStride + k * kStride;
         new[ijk] = -6.0 * old[ijk] + old[ijk-1] + old[ijk+1] + old[ijk-jStride] +
old[ijk+jStride] + old[ijk-kStride] + old[ijk+kStride];
         new[ijk] = -8.15 * new[ijk]
      }
   }
}
```

The code is executed on a system with DRAM bandwidth 51.2 GB/s and a 2-core processor with peak 81.3 GFLOPS per core. What is true about its Arithmetic Intensity (A.I.) and bottleneck with double precision floating point?

- A.I is 0.109 FLOP/byte and its memory-bound
- A.I. is 0.125 FLOP/byte and its compute-bound
- None of the options are correct
- A.I. is 0.125 FLOP/byte and its memory-bound
- A.I. is 0.109 FLOP/byte and its compute-bound

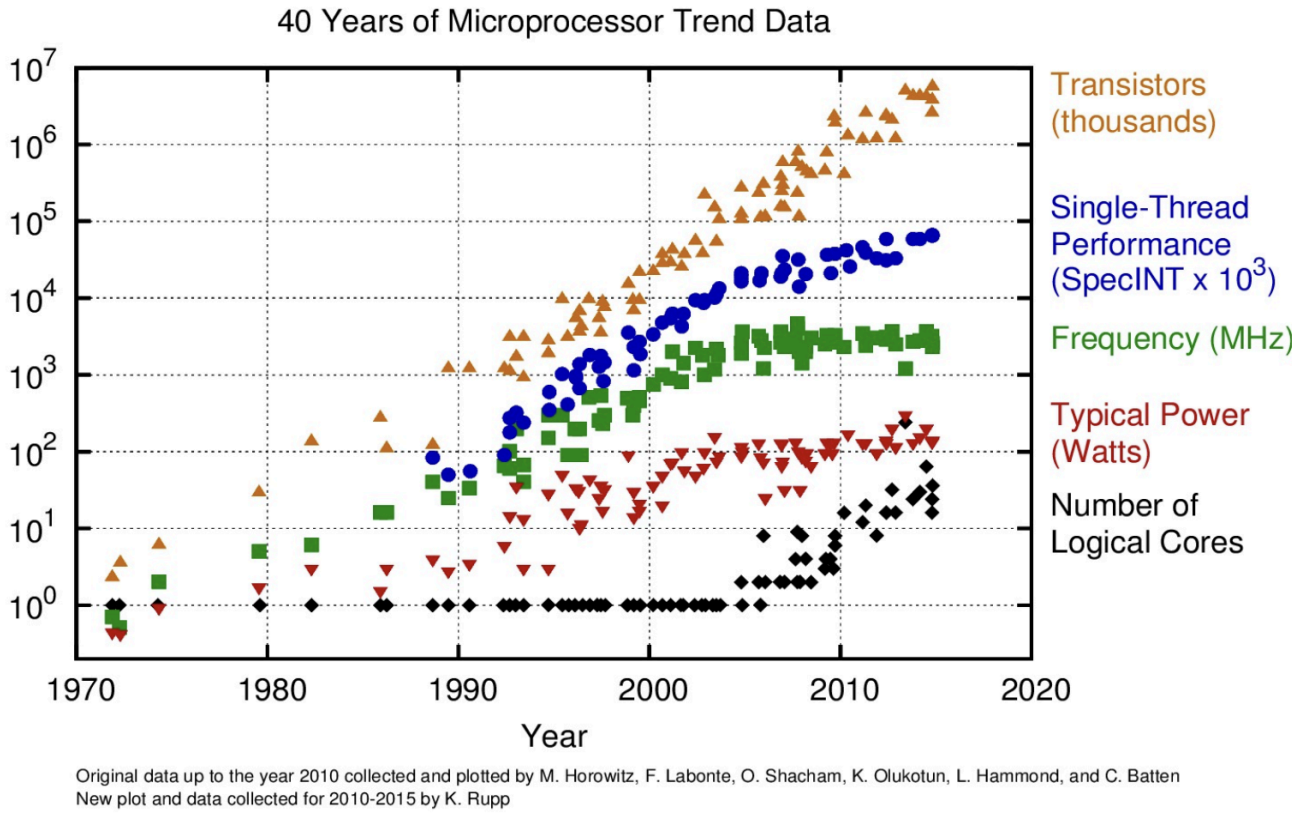
Question 14 2 / 2 points

Consider an Intel Xeon server with 8 cores and 3.5 GHz clock frequency and 32 DP FLOPs/cycle. Here DP stands for double precision (64 bit double). What is true about the peak FLOPS for this server.

- [x] Peak FLOPS can be jumped to about 1.8 TFLOPS (T for tera) by changing to single precision.
- [x] Its peak FLOPS is 896 GFLOPS for double precision arithmetic.
- [x] When running at reduced clock frequency of 2.5 GHz, the peak FLOPS drop to 640 GLOPS
- [ ] Its peak FLOPS is 428 GLOPS for single precision arithmetic.

Question 15 1 / 1 point

Which of the following graph represents Moore's law?



- ○ Thread Performance
- ○ Power

- [ ] Logical Cores
- [x] Transistors

Question 16 2 / 2 points

Table below shows the speedups (as a ratio) obtained when running 10 different jobs on a V100 GPU over an Intel 2.53 GHz CPU.

5, 3, 7.5, 2, 15, 1, 3, 5, 6, 2.5

What is the best approximation for average speedup?

- [ ] 15
- [ ] 5
- [x] sqrt(15)
- [ ] sqrt(5)

Done