Quiz 2: do it yourself; use course notes - Results

Attempt 1 of 1

Written May 23, 2024 7:56 PM - May 23, 2024 9:47 PM

Attempt Score	8 / 8 - 100 %
Overall Grade (Highest Attempt)	8 / 8 - 100 %

Question 1 — 1 / 1 point

In your own words: a) What is the principle of temporal locality? b) how do we make use of it?

The principle of temporal locality states that for any word being requested by the CPU, it is likely that the same word will be soon re-requested again by the CPU. Thus, using the principle of temporal locality, the cache tries to keep the recently requested words in the cache memory or even in the CPU registers for the near future; by preventing them from being overwritten. This speeds up the cache's ability to serve words faster to the CPU.

The correct answer is not displayed for Written Response type questions.

Question 2 — 1 / 1 point

In your own words: a) What is the principle of spatial locality? b) how do we make use of it?

The principle of spatial locality assumes that computer programs tend to run in linear order. Therefore, for any word that the CPU requests, it is likely the CPU will also begin to request the

surrounding words soon. Thus, using the principle of spatial locality, the cache fetches every word as a block or group of words from the main memory. Containing both the originally requested word and the surrounding words. This speeds up the cache's ability to serve words faster to the CPU.

The correct answer is not displayed for Written Response type questions.

Question 3 — 1 / 1 point

Describe why we would make a pipeline LONGER. Show an actual example of your own how this would be a benefit compared to a shorter one

One benefit of making a pipeline longer is to help us increase the bandwidth of our pipeline.

For example, consider a 4 stage pipeline with 1 minute stages (1min+1min+1min+1min).
Such a pipeline will have an instructional bandwidth of 1 instruction/min.

If we take the same pipeline and make it into a 8 stage pipeline with 30 second stages.
(30 sec + 30 sec + 30 sec + 30 sec + 30 sec + 30 sec + 30 sec + 30 sec).
Now, such a pipeline will have an instructional bandwidth of 2 instructions/min.

Therefore, by lengthening our pipeline we were able to increase bandwidth from 1 instruction/min to 2 instructions/min.

The correct answer is not displayed for Written Response type questions.

Question 4 — 2 / 2 points

A pipeline has four stages which take 100nanoseconds, 0.3microseconds, 0.5milliseconds, and 400,000 picoseconds. What is the latency and bandwidth of this pipeline? Always show all your work.

Latency:

100 ns = 0.0001 ms +
0.3 us = 0.0003 ms +
0.5 ms = 0.5 ms +
400,000 ps = 0.0004 ms
= 0.5008 ms

Thus, latency is equal to 0.5008 ms/unit (or 0.5008 ms/instruction)

Bandwidth:

1 unit / slowest stage

Thus, bandwidth is equal to 1 unit/0.5 ms (or 2 instruction/ms)

The correct answer is not displayed for Written Response type questions.

Question 5 **1 / 1 point**

What a pipeline and what is the purpose of a pipeline in our computer?

A computer pipeline is a set of multiple hardware devices that Fetch, Decode, and Execute instructions in parallel and in series. This is done in order to maximize the instructional bandwidth. Bandwidth being the number of instructions finishing execution per unit of time.

A pipeline is also similar to an assembly line, it can process instructions in a way that is simpler, cheaper and faster.

The correct answer is not displayed for Written Response type questions.

Question 6 — 2 / 2 points

Very clearly explain what happens when the CPU requests a word from main memory, in your own words

When the CPU requests for a word to be fetched from main memory, that request gets intercepted by the cache. The cache will then first check itself to see if it has the word that the CPU requested. If the cache has the word, then that word is forwarded to the CPU. If the cache does not have the word, the cache will then fetch that word and the surrounding words from main memory as a block of words and forward just the requested word to the CPU.

Overall, the CPU is not aware that the cache exists and just requests for a word to be fetched from main memory. The cache however actually forwards the requested word to the CPU or fetches it from main memory.

The correct answer is not displayed for Written Response type questions.

Done