

**Answers sheet. Only answers on this sheet will be corrected.
Only those multiple choice answers circled here will be graded.**

CANDIDATE NUMBER: _____

1. Fundamentals

- 1.1. A **B** C D
- 1.2. A B **C** D
- 1.3. A B C **D**
- 1.4. A B C **D**

2. Memory Hierarchy

- 2.1. **A** B C D
- 2.2. **A** B C D
- 2.3. A B **C** D
- 2.4. **A** B C D

3. Caches

- 3.1. A **B** C D
- 3.2. **A** B C D
- 3.3. A B **C** D
- 3.4. A **B** C D

4. Instruction-Level Parallelism

- 4.1. **A** B C D
- 4.2. A B C **D**
- 4.3. A B **C** D
- 4.4. **A** B C D

5. Data-Level Parallelism

- 5.1. A B **C** D
- 5.2. A **B** C D
- 5.3. A B C **D**
- 5.4. A B C **D**

6. Thread-Level Parallelism

- 6.1. A B **C** D
- 6.2. A B C **D**
- 6.3. A B **C** D
- 6.4. **A** B C D

7. Warehouse-Scale Computers

- 7.1. A B C **D**
- 7.2. A B **C** D
- 7.3. **A** B C D
- 7.4. A **B** C D

8. Networks

- 8.1. A **B** C D
- 8.2. A B C **D**
- 8.3. A B C **D**
- 8.4. **A** B C D

9. Prefetching

- 9.1. A **B** C D
- 9.2. A **B** C D
- 9.3. A **B** C D
- 9.4. A B C **D**

10. Articles

- 10.1. A B **C** D
- 10.2. **A** B C D
- 10.3. A B C **D**
- 10.4. **A** B C D

**Answer each of the five open questions as concisely as possible.
Only answers on this sheet will be graded so use the space wisely.**

Answer 1: _____
Energy proportionality is when the power consumption of a core is proportional to the performance, e.g., 10% power consumption at 10% performance and 90% power consumption at 90% performance.
Many contemporary systems operate at less than 100% performance, e.g., datacenters often operate at less than 50% of peak performance. Without an energy proportional system the operations per watt is not maximized.

Answer 2:

A web search can be performed by splitting the search index across multiple machines and then have each machine produce a partial result from its part of the index, i.e., the map functionality in MapReduce.

The partial results can be collected and, e.g., each word in the search can be allocated to a machine to rank the pages for that particular word. The ranked pages are then globally ranked to produce the final result, i.e., the reduce functionality in MapReduce.

Answer 3:

False sharing is when it appears like to processors share the same data when they in reality does not.

Two processors sharing the same cache line but they write to two different words.

Answer 4:

Modified - only this cache has the data and it's dirty

Exclusive - only this cache has the data and it's clean

Shared - the data resides in multiple caches and it's clean

Invalid - the data in this cache is not valide and have to be fetched

Answer 5:

The address is split into a tag, index, and offset.

The index is used to access the correct set and reads the tag and data.

The tag (page) is translated from virtual to physical address space by the TLB.

The translated tag is compared against the stored tag.

If the tags match (hit in the cache) then the offset is used to access the correct word/half-word/byte in the cache line and the data is returned to the pipeline.

If the tags do not match (miss in the cache) then the data is fetched and installed in the cache and the data is returned to the pipeline.