
UNIVERSITI SAINS MALAYSIA

Second Semester Examination
2015/2016 Academic Session

June 2016

CST433 – Advanced Computer Organisation & Architecture
[Organisasi & Seni Bina Komputer Termaju]

Duration : 2 hours
[Masa : 2 jam]

INSTRUCTIONS TO CANDIDATE:

[ARAHAN KEPADA CALON:]

- Please ensure that this examination paper contains **FIVE** questions in **FOUR** printed pages before you begin the examination.

*[Sila pastikan bahawa kertas peperiksaan ini mengandungi **LIMA** soalan di dalam **EMPAT** muka surat yang bercetak sebelum anda memulakan peperiksaan ini.]*

- Answer **ALL** questions.

*[Jawab **SEMUA** soalan.]*

- You may answer the questions either in English or in Bahasa Malaysia.

[Anda dibenarkan menjawab soalan sama ada dalam bahasa Inggeris atau bahasa Malaysia.]

- In the event of any discrepancies, the English version shall be used.

[Sekiranya terdapat sebarang percanggahan pada soalan peperiksaan, versi bahasa Inggeris hendaklah diguna pakai.]

1. Server farms such as Google and Yahoo! provide enough compute capacity for the highest request rate of the day. Imagine that most of the time these servers operate at only 60% capacity. Assume further that the power does not scale linearly with the load; that is, when the servers are operating at 60% capacity, they consume 90% of maximum power. A new system has been proposed that allows for a quick restart but requires 20% of the maximum power while in this “barely alive” state.

Ladang-ladang pelayan seperti Google dan Yahoo! menyediakan kemampuan komput untuk kadar permintaan tertinggi harian. Bayangkan hampir sepanjang masa pelayan tersebut beroperasi 60% daripada kemampuan sebenar. Anggap bahawa kuasa tidak berskala linear dengan bebanan; apabila pelayan tersebut beroperasi 60% daripada kemampuan sebenar, mereka menggunakan 90% daripada kuasa maksimum. Satu sistem baru dicadangkan untuk membolehkan mula semua dengan pantas tetapi memerlukan 20% daripada kuasa maksimum semasa sistem itu di dalam keadaan “separa hidup.”

- (a) How much power savings would be achieved by turning off 60% of the servers?

Berapakah penjimatan kuasa yang dicapai dengan menutup 60% daripada pelayan tersebut?

(8/100)

- (b) How much power savings would be achieved by placing 60% of the servers in the “barely alive” state?

Berapakah penjimatan kuasa yang dicapai dengan meletakkan 60% daripada pelayan tersebut dalam keadaan “separa hidup”?

(8/100)

- (c) How much power savings would be achieved by placing 30% of the servers in the “barely alive” state and 30% off?

Berapakah penjimatan kuasa yang dicapai dengan meletakkan 30% daripada pelayan tersebut dalam keadaan “separa hidup” dan 30% ditutup?

(8/100)

2. Explain the following concepts. Provide **one (1)** example for each concept.

*Terangkan konsep-konsep berikut. Berikan **satu (1)** contoh untuk setiap konsep tersebut.*

- (a) Classes of computers and their common use.

Kelas-kelas komputer dan kegunaan umum mereka.

(6/100)

- (b) Types of parallelism that can be exploited in computer hardware.

Jenis keselarian yang boleh dieksploitasikan oleh perkakasan komputer.

(6/100)

- (c) Evaluation metrics that classify ten advanced optimization of cache performance.

Matrik-matrik penilaian yang mengklasifikasikan sepuluh pengoptimuman termaju bagi prestasi cache.

(6/100)

- (d) Data dependence and name dependence.

Kebergantungan data dan kebergantungan nama.

(6/100)

3. One of the important enablers of Warehouse-Scale Computer (WSC) is ample request-level parallelism, in contrast to instruction or thread-level parallelism. This question explores the implication of different types of parallelism on computer architecture and system design.

Salah satu pengupaya penting komputer berskala gudang ialah keselarian tahap-permintaan yang besar, berbeza dengan keselarian tahap-arahan atau tahap-bebenang. Soalan ini meninjau kesan kepelbagaian jenis keselarian pada seni bina komputer dan reka bentuk sistem.

- (a) Discuss **one (1)** scenario where improving the instruction or thread-level parallelism would provide greater benefits than achievable through request-level parallelism.

*Bincangkan **satu (1)** senario yang melibatkan keselarian tahap arahan dan beban yang mendatangkan lebih faedah daripada pencapaian melalui keselarian tahap-permintaan.*

(8/100)

- (b) What is the software design implication when the volume of request-level parallelism is increased?

Apakah implikasi reka bentuk perisian apabila keselarian tahap-permintaan ditingkatkan?

(8/100)

- (c) What is the potential drawback when the volume of request-level parallelism is increased?

Apakah potensi kelemahan apabila keselarian tahap-permintaan ditingkatkan?

(8/100)

4. List and describe **four (4)** factors that influence the performance of GPU kernels. In other words, which runtime behaviours that are caused by the kernel code due to a reduction in resource utilization during kernel execution?

*Senaraikan dan huraikan **empat (4)** faktor yang mempengaruhi prestasi kernel GPU. Dalam kata lain, perilaku masa jalan yang manakah yang disebabkan oleh kod kernel menyebabkan pengurangan dalam penggunaan sumber semasa pelaksanaan kernel?*

(16/100)

5. Describe one of the ways to change the code of an application to avoid false sharing? What might be done by a compiler and what might require programmer directives?

Terangkan salah satu cara menukar kod aplikasi bagi mengelakkan perkongsian palsu? Apa yang boleh dilakukan oleh pengkompil dan apa yang memerlukan arahan pengatur cara?

(12/100)