DSC 102
Systems for Scalable Analytics

Arun Kumar

Sample Midterm from an online-only edition

Exam Instructions Page

Quiz Details

Quiz Instructions:

This is the first **graded** exam in this course. You can take this exam **only once**.

1. This exam is for **80pts** (+ 5pts extra credit). The official time limit is **160min** but the content is tailored for 80min under normal circumstances. You must take the whole exam in one continual sitting, i.e., Canvas's absolute time counter will start the moment you "open" the exam.

2. You can NOT view all questions at once, i.e., you can view only one question at a time. You must answer the question (or skip it) to view the next question. You can NOT go back and change your answer (or non-answer) to a previous question, i.e., your answers (or non-answers) are "locked in."

3. Here is a summary of the questions in order: AI pledge (1pt) + 10 T/F (20pts) + 10 Conceptual MCQ (30pts) + 1 Short Essay (6pts) + 1 Numerical-multi-part-MCQ (8pts) + 1 Numerical-multi-part-MCQ (15pts). There is an optional Extra Credit Question (Short Essay) at the end (5pts).

4. None of the MCQ have partial credits but partial credits are possible for the short essay questions if you provide a clear and detailed step-by-step explanation/derivation in the rich text box. It can be typed text or a legible photo of your handwritten answer.

T/F Qns. 2pts/qn. Pick subset of 8.

1. T	1. A piece of code is a string.
2. F	2. Float16 arithmetic is associative.
3. F	3. A compiler converts assembly code to human-readable code.
4. T	4. Prefetching data from DRAM to cache is an example of exploiting locality of reference.
5. F	5. Reducing memory stalls will never help raise instruction execution throughput.
6. T	6. Being locked in by a cloud vendor is one of the key disadvantages of public clouds.
7. F	7. Full resource disaggregation is the fastest form of cloud computing.
8. F	8. Tensor arithmetic is *not* amenable to SIMD.
9. F	9. Dask scales well to arbitrarily large task graphs.
10. F	10. The number of edges in a task graph is upper bounded by the number of vertices.

T/F Qns. 2pts/qn. Pick subset of 2.

1. T 1. In the scaleup plot (weak scaling), the dataset size is *not* fixed as the factor is varied.
2. F 2. Threads of a multi-threaded process are allotted disjoint address spaces.
3. F 3. CPU caches have a dichotomy for random vs sequential access latency.

MCQs 1.1. 3pts/qn. Pick subset of 7.

Question
Which component(s) of the OS is/are primarily responsible for data persistence?

Question
Which scheduling policy/policies use(s) preemption?

Question
Which of these file formats typically store compressed data?

MCQs 1.2. 3pts/qn. Pick subset of 7.

Question
What is the name of the PL runtime/compiler mechanism to reclaim DRAM space wasted by processes?

Question
Which of the following is an emerging standard for laying out and accessing structured data in DRAM?

Question
Which of these secondary storage hardware typically degrades the fastest when workloads are write-heavy?

MCQs 1.3. 3pts/qn. Pick subset of 7.

Question
Which of the following is considered FaaS?

Question
Amdahl's Law helps give an upper bound on the utility of adding more of what kind of resource?

Question
Which of these terms refer(s) to data not being present at a higher level of the memory hierarchy, forcing a read from a lower level?

Question
Which of these data types is typically used to store the weights of a logistic regression model?

MCQs 2. 3pts/qn. Pick subset of 2.

Question
Which of the following is *not* a typical file format to store structured data?

Question
Which of the following tools is custom-designed for tensor dataflow graphs?

Question
Which of the following is typically considered SaaS in cloud jargon?

MCQs 3. 3pts/qn. Pick 1 qn.

Question
How many numbers can be represented with upto 6 digits of base 7 representation?

Question
How many numbers can be represented with upto 7 digits of base 6 representation?

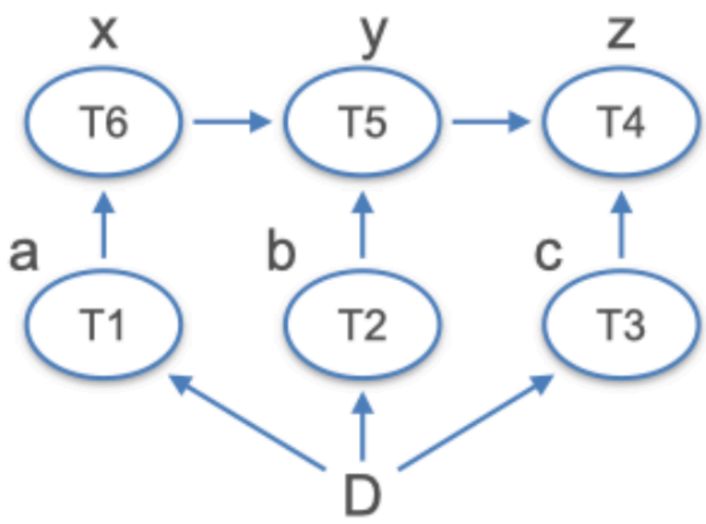
Question
How many numbers can be represented with upto 5 digits of base 8 representation?

Short essay. 6pts/qn. Pick 1 qn.

1. Briefly explain 1 pro and 1 con of “On-Demand” instances vs. “Spot” instances on AWS.
2. Briefly explain 1 pro and 1 con of “Reserved” instances vs “Spot” instances on AWS.
3. Briefly explain 1 pro and 1 con of “Reserved” instances vs “On-Demand” instances on AWS.

Numerical 1.1. 8pts/qn. Pick 1 qn.

Consider this task graph for a workload. You are given (a,b,c) = (10,20,10) and (x,y,z) = (20,10,15)



1) (2pts) What is the degree of parallelism in this workload? Type the exact numeral(s).

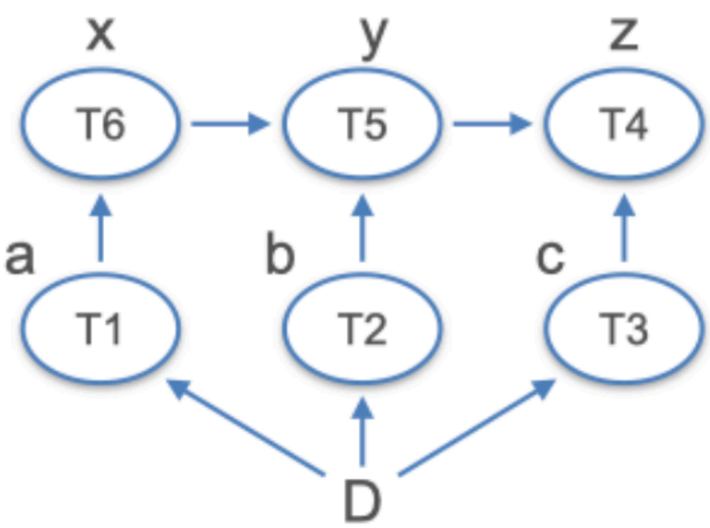
[A1] 3

2) (6pts) What is the lowest possible completion time on a 2-worker cluster? Type only the exact numerals of the time units.

[A2] 55

Numerical 1.2. 8pts/qn. Pick 1 qn.

Consider this task graph for a workload. You are given (a,b,c) = (15,30,15) and (x,y,z) = (30,10,15)



1) (2pts) What is the degree of parallelism in this workload? Type the exact numeral(s).

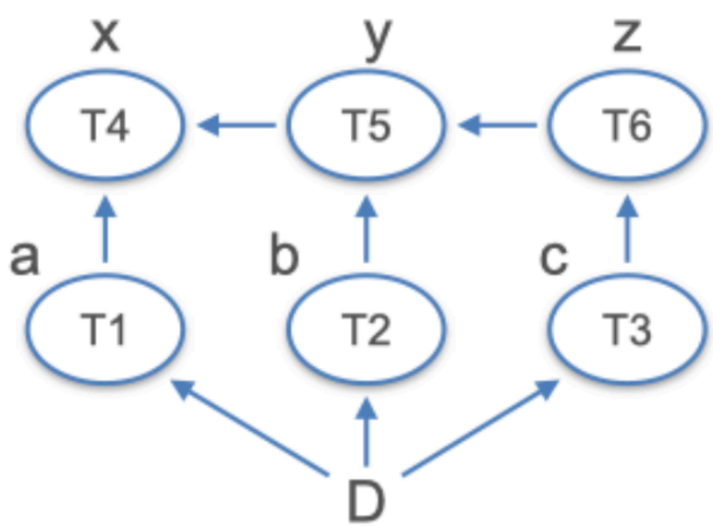
[A1] 3

2) (6pts) What is the lowest possible completion time on a 2-worker cluster? Type only the exact numerals of the time units.

[A2] 70

Numerical 1.3. 8pts/qn. Pick 1 qn.

Consider this task graph for a workload. You are given (a,b,c) = (20,40,20) and (x,y,z) = (15,20,40)



1) (2pts) What is the degree of parallelism in this workload? Type the exact numeral(s).

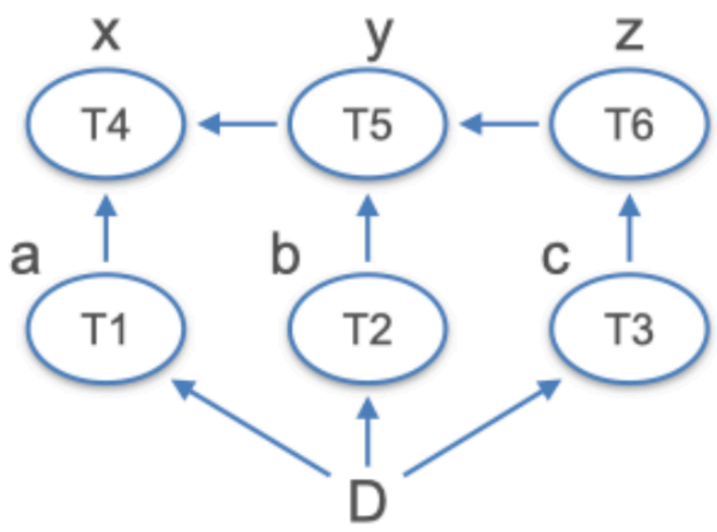
[A1] 3

2) (6pts) What is the lowest possible completion time on a 2-worker cluster? Type only the exact numerals of the time units.

[A2] 95

Numerical 1.4. 8pts/qn. Pick 1 qn.

Consider this task graph for a workload. You are given (a,b,c) = (30,20,30) and (x,y,z) = (15,20,20)



1) (2pts) What is the degree of parallelism in this workload? Type the exact numeral(s).

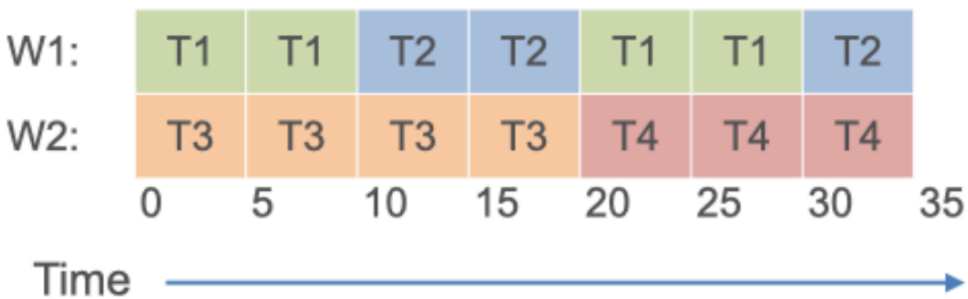
[A1] 3

2) (6pts) What is the lowest possible completion time on a 2-worker cluster? Type only the exact numerals of the time units.

[A2] 85

Numerical 2.1. 8pts/qn. Pick 1 qn.

Tasks T1-T4 arrive at times 0, 10, 0, and 10, respectively. No dependencies exist among these tasks. You are given the following task-parallel schedule to execute this workload on a 2-worker cluster.



1) (5pts) What is the average turnaround time with the given task-parallel schedule? Round to the nearest whole number and type only the exact numerals of the time units. (Example: If the answer is 15.34, type only 15; if it is 15.79, type only 16).

[A1] 25

2) (5pts) One crude way to think about fairness in scheduling is to divide each task's turnaround time by its length; the lower this ratio, the better the deal a task got in the schedule. Viewed this way, which task got the *best* deal in the giver schedule? Type only the numeral of the task number (Example: If the answer is T1, type only 1).

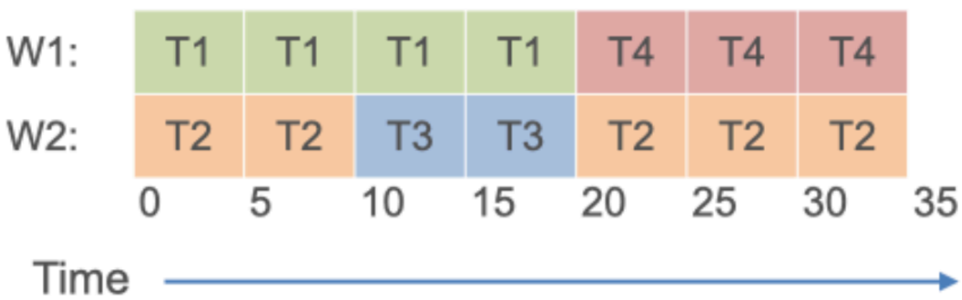
[A2] 3

3) (5pts) Suppose you have only 1 worker and use SCTF scheduling policy. What is the average turnaround time now? Round to the nearest whole number and type only the exact numerals of the time units.

[A3] 39

Numerical 2.2. 8pts/qn. Pick 1 qn.

Tasks T1-T4 arrive at times 0, 0, 10, and 10, respectively. No dependencies exist among these tasks. You are given the following task-parallel schedule to execute this workload on a 2-worker cluster.



1) (5pts) What is the average turnaround time with the given task-parallel schedule? Round to the nearest whole number and type only the exact numerals of the time units. (Example: If the answer is 15.34, type only 15; if it is 15.79, type only 16).

[A1] 23

2) (5pts) One crude way to think about fairness in scheduling is to divide each task's turnaround time by its length; the lower this ratio, the better the deal a task got in the schedule. Viewed this way, which task got the *worst* deal in the given schedule? Type only the numeral of the task number (Example: If the answer is T1, type only 1).

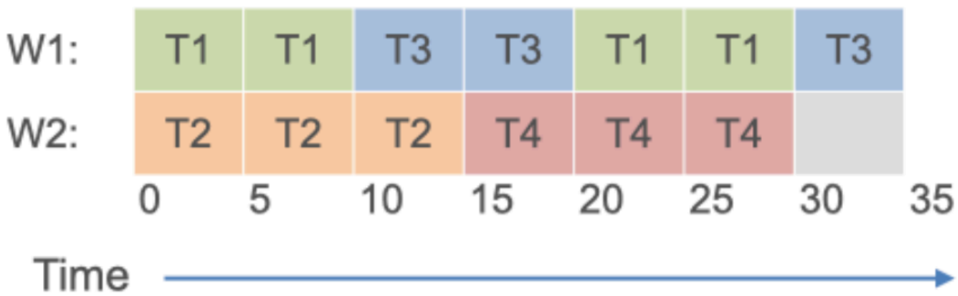
[A2] 4

3) (5pts) Suppose you have only 1 worker and use SCTF scheduling policy. What is the average turnaround time now? Round to the nearest whole number and type only the exact numerals of the time units.

[A3] 36

Numerical 2.3. 8pts/qn. Pick 1 qn.

Tasks T1-T4 arrive at times 0, 0, 10, and 10, respectively. No dependencies exist among these tasks. You are given the following task-parallel schedule to execute this workload on a 2-worker cluster.



1) (5pts) What is the average turnaround time with the given task-parallel schedule? Round to the nearest whole number and type only the exact numerals of the time units. (Example: If the answer is 15.34, type only 15; if it is 15.79, type only 16).

[A1] 23

2) (5pts) One crude way to think about fairness in scheduling is to divide each task's turnaround time by its length; the lower this ratio, the better the deal a task got in the schedule. Viewed this way, which task got the *best* deal in the given schedule? Type only the numeral of the task number (Example: If the answer is T1, type only 1).

[A2] 2

3) (5pts) Suppose you have only 1 worker and use SCTF scheduling policy. What is the average turnaround time now? Round to the nearest whole number and type only the exact numerals of the time units.

[A3] 34

Extra Credit Qn. 5pts/qn. Pick 1 qn.

Question

(Optional) Extra Credit Question:

You are given a matrix **A** represented as a relation with one tuple per cell in the following schema: *A (row, column, value)*. Write an SQL query (SELECT ...) to compute the Gramian matrix $\mathbf{A}^T \mathbf{A}$. Write each clause in a new line.

Question

(Optional) Extra Credit Question:

You are given a matrix **A** represented as a relation with one tuple per cell in the following schema: *A (row, column, value)*. Write an SQL query (SELECT ...) to compute the matrix outer product $\mathbf{A}\,\mathbf{A}^T$. Write each clause in a new line.