

M. Mahesh Kumar

89-280-9605

LSU Department of Computer Science

Fall 2010 Final Exam

CSC7700 Scientific Computing

December 6th 2010, 5.30pm to 7.30pm

General Instructions

- This is a closed book exam.
- No calculators or electronic devices.
- Part I of the exam covers all the five course modules and is designed to take 80 minutes to complete. Part II of the exam is for the Networks and Data module and is designed to take 40 minutes to complete.
- Part I is worth 20% of the final grade. Each module includes 5 questions. All questions have equal weight. Answer all questions.
- Part II is worth 10% of your final grade. Answer only four out of five questions. If you answer all five, only the lowest graded four will be taken into consideration. Questions have two parts, you need to answer both parts of the four questions you select.

Part I

Module A: Basic Skills

1. Provide two reasons why the same text file can look different when viewed on different systems or within different tools.

Same text can look different when viewed on different system or different tools because parameter used are different

2. In the context of numerical simulations, explain what is meant by discretization and why it is used.

④ Discretization is nothing but, it is an approximation or else it is an error.
It is used for minimize an error or remove error.

3. Briefly describe what a pseudo random-number generator is, and name three disadvantages over real random-number generators. Name two reasons why pseudo random-number generators are often used despite these disadvantages?

Pseudo random-number generator is imaginary number generator
The numbers are generated randomly.

disadvantages over real random number generators

- * It is accurate whereas in imaginary we did not get accurate result.
- * In pseudo we should predict for value.

4. Name one advantage and two potential disadvantages of the Newton-Raphson method over the bisection method for root-finding.

* Advantage of New-Raphson method over bisection:

→ The time taken for root find in bisection method is large compared to Newton-Raphson method.

→ No. of iterations taken will be more

Main disadvantage:

If first root is known we can find second root

5. Explain the difference between centralized and distributed version control systems, including one advantage and one disadvantage for each. Name one software implementation example for each kind of system.

* Centralized version of control system VS distributed

It is mainly centralized at one place and take at a time whereas in distributed version

control system distribute in to different layers

and reduces it work and make it fast

where we can easily define error in distributed

compare to centralized version.

Module B: Networks and Data

1. List two TCP parameters used in iperf and briefly describe their influence on the performance of TCP.

① -W → It is mainly ^{used} for window size
-t → Used for time taken by process
-l → Used for packet size. max for packet size & len will be the Performance.

2. Briefly describe what the server-side data processing plug-in included in the standard GridFTP installation does and what it can be used for (hint - you used it in your homework)

3. List two benefits that middleware provides to developers of distributed applications.

middleware: It is interface or software between API
and running machine

It helps the developers in developing application or in
code generation.

4. Briefly outline two methods for accessing remote data in a distributed application.

* IRDS is one of the methods for accessing remote data in distributed application.

5. Briefly outline two methods of doing remote visualization (based on distribution of the visualization pipeline)

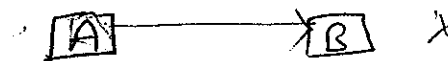
Module C: Simulations and Application Frameworks

1. What determines the accuracy of a simulation? List two ways in which accuracy can be improved.

* Accuracy of simulation: The error we get is decreased to get or minimized to get Accurate result is Accuracy of simulation.
Accuracy can be improved by

2. What is MPI, and what is it used for? Assume there are two processes, and process A needs to access an array element stored on process B. Schematically, how does this work?

* MPI defines message passing interface. The name itself suggests that it is an interface for passing of messages.



3. What is a software framework? Name one software framework, and provide three characteristic elements of a software framework.

Software framework is steps by step divided in to frames i.e. the total work done in step by step process divided in to number of frames

Accuracy: we get Accurate result

4. What are CCL files in Cactus? List which CCL files exist, and what they define.

⑧ Interface.ccl
Schedule.ccl
Param.ccl } These are 3 types of ccl files
Exist in cactus

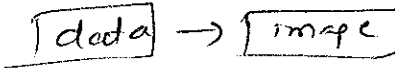
Param.ccl is mainly for run process
Interface.ccl implements our program for compiling
& we define key word for cxx: (Kenetic::KE)

5. Name and briefly describe five tools that support code development in large, distributed, international collaborations.

Module D. Scientific Visualization

1. Define and describe a "Visualization Pipeline".

Visualization Pipeline : How the data can be visualized by different steps
 pipeline mainly for transferring
 data to image for visualization



2. What is the difference between the "push model" and the "pull model"?

Push model

- 1) Data are made available as early as possible
- 2) transfer data viz when loaded / create

pull model

- 1) Data are made available as late as possible
- 2) It transfer data viz network when rendering time

3. Describe the three atomic elements ("building blocks") in a visualization network.

* Grid, Fiber, Output

Data, Display, output, Grid, Fiber

data is processed for get output

display: image is displayed (output)

Grid: It is set of

4. Define and describe the purpose of a bi-vector.

* Application used for rotating 90° is known as bi vector

5. Which are the three property objects ("communication types") in the "F5" fiber bundle data model that are visible to the end user?

Module E: Distributed Scientific Computing

1. We discussed five applications – Montage, Nektar, Climateprediction.net, SCOOP and Ensemble-based/Replica-Exchange simulations. For any THREE of these (you choose which three), answer any ONE of the following: Why they were distributed? How they were distributed? The Challenges &/or success in distributing them?

Montage: Here DAG is created and enacted for distributed (How) Images are loosely coupled as it is distributed

Nektar: Here 8TB of data is required in factors of 10 (why) so it is distributed

Climateprediction: The prediction comes from different places, so it is easily distributed

2. Estimate to within an order of magnitude the number of jobs that are executed in the Worldwide LHC Computing Grid (WLCG) per day. Estimate to within an order of magnitude the number of bytes of data generated (overall) by the WLCG. Estimate the cost of the LHC Experiment. Therefore what is the cost of generating a byte of data from the LHC experiment?

3. Using your estimate (whatever it was) of number of jobs (on the WLCG) from the previous answer, given that there are approximately 250,000 cores as part of the WLCG, and that it has a typical utilization factor of 50%, estimate the average time each job takes. (assume: each job is a single-core job).

4. List two factors – technological or non-technological, driving Cloud Computing. Provide a “real production” example of a Cloud offering. Is the Cloud offering an example of IaaS, PaaS or SaaS?

5. Provide one difference between predominantly HTC and HPC Grids. Provide a “real production” example of a HPC and HTC Grid.

HPC → High performance Computing Grids

HTC →

MAHESH
KUMAR

Part II

Networks and Data

Question 1

- A) How are layers used in network implementations?

The process of transferring data from client to server or vice versa takes place with the help of different layers. There are 7 layers they are Application layers, presentation layers, session layers, transport layers, Network layer, physical layer, data layer. Application layers for sending files, mails; session layers is for starting or activating session; transport layers is for data transferring; physical layer is mainly for web services; data layer sends data from node to node with station address.

- B) What are the major differences between TCP and UDP?

TCP: It is connection oriented protocol & creates connection between server & client through which data can be passed.

Reliable: It is reliable as we know what is sent & is

Ordered: Here when A & B are sent the data which sends first will receive first. If 'A' is sent first then 'A' will receive first.

* It is heavy weight

* It is byte oriented

UDP: It is connectionless & it does not follow any order. It is just a fixed protocol.

* It is unreliable

* UN ordered

* It is less weight

* It is packet oriented

Question 2

- A) What data transmission protocol would you use for bulk data transmission and why? What protocol would you use for video or audio conference and why?

→ GridFTP is used for bulk data transmission because it has several features supporting: * Parallel transfer: It sends data in different streams so it becomes fast. We can use remote access for all these i.e. Teragrid machines ~~for~~ ^{which} GridFTP is installed.

Video conference and audio conference simultaneously works. There is video conference if there is no audio. The protocol used for this is

- B) Describe circuit network services and their advantage.

Question 3

- A) Describe what a naming service is (in middleware implementations) and what it is used for.

* middleware "is an s/w ^{interface} ~~code~~ between API & running machines.
DNS → Domain Naming Service is used for middleware implementations. It defines on what domain service it is using. It is very important for implementing ~~the~~ middleware.

- B) In your own words, describe the "end-to-end" argument.

There are two types of protocols end-to-end protocol &
Peer-to-peer protocol

How end-to-end protocol :

Question 4

- A) List the usual sequence of operations for accessing data in a distributed file system.
- B) Briefly describe the two possible (and sometimes conflicting) optimization goals of a scheduling system.

Question 5

- A) Describe use case scenarios where remote visualization is useful or needed.

*) → Remote visualization is very useful as we can see visualized the data from different places by remote logging. ~~not~~ here
* ~~can~~ depict correctly whether the data is correct by accessing from different places.

- B) Describe some of the possible benefits of distributed visualization.

*) Benefits of distributed visualization.

→ Here the data to visualized is distributed so that we can clearly identify ~~the~~ errors (if any) & easily depict each & every important reasons for that

→ errors can be minimised.

→ work can be distributed, & minimised

→