1 Part I

1.1 Module A

1. Why can the same simple text file look different on different systems or within different tools? Name at least two reasons.Possible answers: *EOL, TABS, Syntax highlighting, Character encodings*

2. What is discretization and why is it used?Possible answers: *model of continuous function on distrete set; because continuous space would require unlimited memory*

3. What are pseudo randomnumber generators, and name at least three disadvantages over real randomnumber generators. Why are are they often used nevertheless (name at least two reasons)? Possible answers: *generate sequences of numbers with properties close to real random numbers; finite size sequences, correlations, non-uniform distribution; cheap, fast, reproducable*

4. Which different communication channels can improve interaction of distributed teams? Name at least three types with their advantages and disadvantages. Possible answers: *personal meetings, email, phone, online chat, video conferences*

5. Which two different types of version control systems exist, and what are their differences? Name at least one example for each type. Possible answers: *central(cvs,svn)/decentral(git,darcs,mercurial);...*

1.2 Module B

1.

2.

3.

4.

5.

1.3 Module C

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

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?

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

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

5. What tools exist to support code development in large, distributed, international collaborations? Name at least five such tools.

1.4 Module D

1. Define and describe a "Visualization Pipeline".

2. Specify the difference between the "push model" and the "pull model".

3. Describe the three atomic elements in a visualization network.

4. Define and describe the purpose of a Bi-Vector.

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

1.5 Module E

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? Challenges &/or success in distributing them?

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 magnitue 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. What are some of the factors – technological and 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. What is (are) the primary difference(s) between predominantly HTC and HPC Grids? Provide a "real production" example of a HPC and HTC Grid?

2 Part II