

Gestão de Centros de Dados

Date: 14/Janeiro/2011

Duration: 1h45m

Note: Not everything is fully specified. This is intentional, and you may assume reasonable options and then formulate your answers. Times and grade values are approximate.

Q-1 [30m, 5 val.] Considering the time you have available, write a 1 page (or slightly less) executive summary to “sell”, to a company’s manager/top executive (CEO, CFO, or other non-IT expert), the idea of using hypervisor-based virtualisation on the (non HPC) DC. Do not forget to convey the advantages (and, if you find relevant, the disadvantages – at least referring situations where virtualisation should not be used) paying attention to what an executive regards as fundamental objectives, such as cost control/reductions and user (both external and employees) satisfaction, among other considerations...

Q-2 [30m, 5 val.] A bank’s IT manager has spotted the following: 1) the bank’s storage devices are disk array racks that are completely filled “up to the top” with fast (FC de 10K rpm) disks which, however, are currently unable to satisfy the on-line transaction rate (as needed by the RDBMS and its users); and, 2) in spite of the efforts to selectively copy only “fresh” information, the backup window is too narrow because all the DB must be copied (as it resides in a “single” file). Currently, each host submits a request to the backup server to access the tapes; then data is moved by the host, through the backup server, into the tape.

An analysis has been carried out and it was found that only 10% of the DB are accessed for on-line transaction processing; the remaining data are accessed just for reading, with 30% corresponding to data which is up to 2 years old, and is seldom accessed, while another 30% correspond to data which is 2 to 5 years old and is rarely accessed; finally, the remaining 30% correspond to data which is 5 to 10 years old and is kept for legal reasons. The DB is stored in the disks that occupy a full rack. A benchmark, using a copy of “young” data and the bank’s online application was carried out with SSD disks (which are compatible with the existing disk arrays) and showed that their performance is so good that it solves the bank’s problem.

- a) Find a solution for the performance problem that also takes the overall cost into account.
- b) Find a solution for the backup problem.

Q-3 [20m 5 val.] In the previous question the issue of benchmarking with a small portion of the bank's data while using the real application was mentioned.

- a) Do you think that their choice was adequate, or would it be better if they have used one (or more) standard benchmarks instead? Why?
- b) Now, suppose that we could not use the bank's data and application; what would be a good standard benchmark to simulate online transactions against an RDBMS to see if the SSD disks would solve the problem? Why? [Note: if you do not remember the benchmark's name just describe its characteristics – i.e., just answer the why above]

Q-4 [25m 5 val.] As you know (and was presented in a seminar), unifying the infrastructures that are used to interconnect hosts and devices in an IT (not HPC) DC is the way to go today... [Note: you can organize your answer in a table, where a part corresponds to (a) while another corresponds to (b)]

- a) Briefly mention (as this subject was already covered in the previous GCD test) the two infrastructures commonly used in DC today. Describe the protocols that are used, including the physical layer (cabling, signalling – electrical vs. optical – usual transmission speeds/rates, “cable” lengths).
- b) Do describe now the characteristics of the new unified (or converged) infrastructure in a way one can easily spot the differences among (b) and (a).
- c) What are the reasons behind this movement to change from the technologies described in (a) to those of (b)? What are the advantages? Are there any disadvantages? (cost? others?)