
Name: _____

Instructions

There are four (4) questions on the exam. You may find questions that could have several answers and require an explanation or a justification. As we've said, many answers in storage systems are "It depends!". In these cases, we are more interested in your justification, so make sure you're clear. Good luck!

Problem 1 : Short answer. [36 points]

(a) One approach to disaster recovery is to maintain two physically distant replicas, with each update going to both. Another approach is to periodically send modified blocks from the local primary to the remote replica in the background. For each approach, identify a reason why it might be preferred over the other.

(b) File systems (e.g., WAFL and LFS) that remap newly-written blocks into sequential ranges improve performance for RAID 4 and RAID 5 arrays. How?

(c) For distributed file systems that use callbacks, describe how consistency could be compromised if the server reboots.

(d) NFS is a “stateless” protocol, meaning that the server keeps no state about the clients using it. As a result, the consistency guarantees provided to client applications are very weak. How could NFS clients be changed to achieve strong consistency guarantees? What would be the consequence?

(e) Consider a virtualizing switch for NFS servers that replicates all files across two servers in order to improve availability. It does so by simply sending each write-type request (e.g., create, delete, write) to both servers.

(i) How should the switch decide where to send each read if seeking to ensure that, when no writes are in progress, two clients performing reads will see the same answer? (Remember that the switch could crash and reboot.)

(ii) What new capability could be added to the switch to use more aggressive read routing algorithms without violating the above design goal?

Problem 2 : Short answer. [28 points]

Consider a large database system attached to a RAID 4 system with 5 disks, designed for 100 busy order entry workers. Each worker enters one order after another with zero breaks. Each such order requires 10 ms of CPU time, eight cache misses (reads from the RAID array), and a synchronous write to one data block. Data are striped across the data disks such that each request has an equal probability of going to any of the disks. Each individual disk request takes 10 ms.

(a) What is the throughput of the database system?

(b) Which portion (CPU, data disks, parity disk) is the bottleneck? *(Show your work for full credit.)*

(c) What is the ideal number of data disks to make sure that the load on each disk in the array is equal?

(d) The other way to achieve balance among the disks would be to use RAID 5 instead of RAID 4. What would be the throughput of the system (assuming 5 disks in the array) with this change?

Problem 3 : Short answer. [36 points]

(a) Explain the potential value of RDMA in allowing databases running atop distributed file systems to achieve performance competitive with those running atop SAN block-based storage.

(b) Consider a new version of AFS in which clients spread each data file across five servers via RAID 5. What new problem arises in this version and what could be done about it?

(c) Even though the keys could be lost, some users prefer to encrypt files on their client machines even if network transfers to/from the server are unsniffable. Why might this be so?

(d) The SCSI protocol includes no notion of access control, trusting the host operating system fully. Some storage networks support “zoning” to allow only some hosts to communicate with certain devices. As iSCSI (SCSI over TCP/IP) emerges, how could this form of access control be achieved?

(e) Snapshots provide frozen views of storage state at a point in time. Usually, they are implemented via copy-on-write in the storage system. To efficiently support incremental backups, rather than full backups, what information must the storage system expose beyond the snapshot contents? How would it be used?

(f) High reliability disk array controllers replicate cached writes in separate battery-backed cache banks. Why battery-backed and why replicated?

Problem 4 : Bonus questions. [a few bonus points]

(a) How do you do data protection for your personal data? (We'd really like to know.) Have you changed how you do this since taking this course?

(b) If Timmy were invited to give a lecture, which of the topics from this course do you think he is most qualified to present, and why? *(Note: only the most "creative" answers will be considered.)*

This page intentionally left blank in case you need scratch space.

This page intentionally left blank in case you need scratch space.