Data Storages

Monday, November 12, 2007
2:46 PM

- Price of individual disks gets cheaper.
- Price of connecting and managing disks are getting higher.

- Managing Storage
    - Think of it as a community effort.
    - Storage server should be allocated on group-by-group basis.
        - Particularly effective when the cost of storage comes from own group's budget.
    - Start by asking customers what they want to do. What does and doesn't work in the current arrangement. Ask about projected disk usage as well.
    - As much as possible, arrange customers by departments rather than usage.
        - Partition shared storage if groups are to share harddisks.
    - Separate general storage spares (for desktop computers), and infrastructure spares (for servers). Stock CPU and RAM for server upgrades as well.
    - Have a storage standard to avoid one-off equipments.
    - Erase old hard drives fully. Company information may be in there!

- Storage Service
    - Storage SLA
        - Specifiy availability: how much space is usable, how much is reachable, to whom?
        - Specify response time: mean time to repair in case a storage fails.
        - Use standard benchmark tools to test equipments so that you can write SLA with confidence.
    - Reliability
        - Keep spare disks for use when one disk in your RAID fails.
        - NAS should have multiple network interfaces. Each interface should connect to different network switch.
        - Reliability is expensive. Educate the management of the return of investment in reliability.
    - Backups
        - RAID is not backup strategy.
        - Use RAID mirrors to back up: disconnect one mirror and use that one mirror to transfer data to tapes or other devices. Then, reattach.
        - If you have more than 3 mirrors, backups can be faster, and the system doesn't slow down much during backup time.
        - NAS makes backups easier because users will store only shared data on NAS servers.
        - SAN makes backups easier in two ways.
            - Tapes can be part of SAN.
            - Backups do not interfere with normal traffic.
    - Monitoring: You should look for
        - Disk failures
        - Other outages, such as network outages
        - Space used and space free
        - Rate of change: for future planning
        - I/O local usage
        - Network local interface
        - Network bandwidth usage
        - File service operations
        - Lack of usage
        - Individual resource usage.
    - SAN caveats
        - SAN is new and is always changing.
        - Stick with one or two vendors.
        - Test extensively.

- Performance
  - Mantra
    - Measure, optimize, and measure again.
  - Tools
    - Add more RAM.
    - Add more HDDs.
  - RAID
    - RAID 3 is fast for streaming applications.
    - RAID 4 can have faster reads than RAID 5 in a tuned file system.
  - NAS optimizes performance by separating file service workload from other servers.
  - SAN move file service traffic from the main network. It also gives security advantage since files are separated from other services.

- Common problems
  - Power and cooling
    - Make sure that your data center can cool the storage well, and can give enough power to storage.
      - □ Storage uses more power when starting up than when running.
  - Timeouts
    - Changing in network configuration can cause TCP timeouts that can drop NAS requests.
  - Saturation behavior
    - Look out for what happen when the load is high.