Best Practices for Designing Highly Available Storage – Overall Storage System

This is the last part of a 6 part article where I am talking about Best Practices for Designing a Highly Available Storage System from different perspectives. Today in this last part I am going to talk about what it takes to design a highly available storage system.

Availability refers to the storage system's ability to provide user access to data in the case of a hardware or software fault. Midrange systems are classified as highly available because they provide access to data without any single point of failure. However, the following configuration settings can improve performance under degraded mode scenarios.

Single DAE Provisioning is the practice of restricting the placement of a RAID group within a single enclosure. This is sometimes called horizontal provisioning. Single DAE provisioning is the default method of provisioning RAID groups, and, because of its convenience and High Availability attributes, is the most commonly used method.

In Multiple DAE Provisioning, two or more enclosures are used. An example of multiple DAE provisioning requirement is where drives are selected from one or more additional DAEs because there are not enough drives remaining in one enclosure to fully configure a desired RAID Group. Another example is SAS backend port balancing. The resulting configuration may or may not span backend ports depending on the storage system model and the drive to enclosure placement.

An LCC connects the drives in a DAE to one SP's SAS backend port; the peer LCC connects the DAE's drives to the peer SP. In a single DAE LCC failure, the peer storage processor still has access to all the drives in the DAE, and RAID group rebuilds are avoided. The storage system automatically uses its lower director capability to reroute around the failed LCC and through the peer SP. The peer SP experiences an increase in its bus loading while this redirection is in use. The storage system is in a degraded state until the failed LCC is replaced. When direct connectivity is restored between the owning SP and its LUNs, data integrity is maintained by a background verify (BV) operation.

Request forwarding's advantages of data protection and availability result in a recommendation to horizontally provision. In addition, note that horizontal provisioning requires less planning and labor.

If vertical provisioning was used for compelling performance reasons, provision drives within RAID groups to take advantage of request forwarding. This is done as follows:

RAID 5: At least two (2) drives per SAS backend port in the same DAE. RAID 6: At least three (3) drives per backend port in the same DAE. RAID 1/0: Both drives of a mirrored pair on separate backend ports.

FAST Cache

It is required that flash drives be provisioned as hot spares for FAST Cache drives. Hot sparing for FAST Cache works in a similar fashion to hot sparing for traditional LUNs made up of flash drives. However, the FAST Cache feature's RAID 1 provisioning affects the result.

If a FAST Cache Flash drive indicates potential failure, proactive hot sparing attempts to initiate a repair with a copy to an available flash drive hot spare before the actual failure. An outright failure results in a repair with a RAID group rebuild.

If a flash drive hot spare is not available, then FAST Cache goes into degraded mode with the failed drive. In degraded mode, the cache page cleaning algorithm increases the rate of cleaning and the FAST Cache is read only.

A double failure within a FAST Cache RAID group may cause data loss. Note that double failures are extremely rare. Data loss will only occur if there are any dirty cache pages in the FAST cache at the moment both drives of the mirrored pair in the RAID group fail. It is possible that flash drives data can be recovered through a service diagnostics procedure.

The first four drives, 0 through 3 in a DPE or in the DAE-OS of SPE-based storage system are the system drives. The system drives may be referred to as the *Vault drives*. On SPE-based storage systems the DAE housing the system drives may be referenced as either DAE0 or DAE-OS. These drives contain files and files space needed for the:

1. Saved write cache in the event of a failure
2. Storage system's operating system files
3. Persistent Storage Manager (PSM)
4. Operating Environment (OE) - Configuration database

The remaining capacity of system drives not used for system files can be used for user data. This is done by creating RAID groups on this unused capacity.