



VERITAS™

BUSINESS WITHOUT INTERRUPTION™

SPC-2 PR Comment Persistent Reservations



Roger Cummings

THE DATA AVAILABILITY COMPANY™

Background

- ≡ **VERITAS is an cross-platform APPLICATION vendor**
 - With a 32 node shared-nothing SAN-based cluster server (VCS) product available on 3 platforms
- ≡ **Until now VCS has used reserve/release in some cluster management algorithms**
 - Works OK, but code to handle Hard Resets is tricky
 - Would much prefer to use Persistent Reservations for the future
- ≡ **VCS architecture designed to handle 100s of nodes**
 - And operate for years without total reboot!

Data
Access

Data
Protection

Scope

☐ SPC-2 r19 5.5.3 1st sentence says:

“The persistent reservations management method is the mechanism specified by this standard for use by multiple initiators that require operations to be protected across initiator failures, which usually involve hard resets.”

☐ This is absolutely and precisely correct...

- But would you read this as also protecting operations across scheduled initiator maintenance, or a manual initiator offline, where some cleanup is performed????

Data
Access

Data
Protection

PR Comment Response

☐ Stated that:

- The intent of T10 is that releasing a reservation should be part of a controlled software shutdown process during which cooperating members of a cluster should use Persistent Reservation commands to gracefully transfer reservations ownership to remaining cluster members. The behavior described in SPC-2 is consistent with that intent.

☐ That's certainly what we would like, but its not what the current definition does..

Data
Access

Data
Protection

PR Comment Response

Also contained:

- Furthermore, changing the definition in this way at this late date will materially harm those who worked with T10 during the process of defining Persistent Reservations and as such would be a disservice to the industry.
- If the new behavior proposed is desired, it should be proposed for SPC-3 as a new Persistent Reservation Type (e.g., Write Exclusive - All Registrants).

Reasons why comment approach chosen:

- Did not want to introduce additional function
- Thought change minor, self-contained
- Don't believe present definition - release when deregister - is "must have"

Data
Access

Data
Protection

Detailed Operation

- ▼ **During cluster startup all or subset of nodes registers with each storage device**
 - Designated owner node creates Write Exclusive Registrants Only reservation
- ▼ **When a node is to be taken offline, it deregisters**
 - If a persistent reservation is “associated with the unregistered initiator” (per 5.5.3.6.1), the reservation is released, all registrants get Unit Attention
 - While UA is being processed by the former registrants another node could access the device, causing data corruption
 - All remaining nodes have to arbitrate to identify the source of a new reservation (or they all have to PreEmptAbt)
 - Consumes resources at a time when Cluster very busy
 - And all that is really wanted is the original reservation back!

Data
Access

Data
Protection

Why not leave registration?

Node being taken offline could not deregister, but....

- When it comes back online, how does it identify its own registration (without assuming unique keys)??

Leaving the registration leaves an opportunity for problems

- If the node is reinitialized, but does not rejoin the cluster, it still has access to cluster storage!
- And it just feels wrong!

How many registrations can a device support?

- 100s?
- Because that is what we're going to need!
- And how is this number discovered?

Data
Access

Data
Protection

Needs

Two choices going forward:

- Define a new Persistent Reservation type which maintains a Registrants Only reservation until the last Initiator deregisters
 - Should this be in addition to the existing type or as a replacement ?
 - Is the existing behavior a “must have” for anyone?
- Define a new service action for Persistent Reserve In to enable both Initiator identifier & key to be read for existing reservation
 - Allows Initiator going offline to not deregister, and not need to register again on returning online

All of the above defined in detail in 01-204r0

VERITAS preference is new reservation type as a replacement

Comments??

Data
Access

Data
Protection

Summary

- ≡ **VERITAS believes that the proposed modification to Registrants Only Persistent Reservations makes them useful in cluster applications**
- ≡ **IF this is the only substantive change, would not expect it to be included in SPC-2**
 - We have a workaround for the current definition, albeit ugly
 - Will submit one of the previous proposals for SPC-3

Data
Access

Data
Protection