



**CATE Storage and Backup Engineering
Citi Architecture and Technology Engineering**

Guidelines for Host Spanning

Policy

GDSE-08003-P

1.1

1 Version Control

Version	Date	Author	Reviewed	Comments
0.1	5-Feb-08	Ludwig Derodel	Paul Ratinov	Peer Review
0.2	14-Nov-08	Ludwig Derodel	Chris Clayton	Added section on Thin volumes
0.3	07-Jan-09	Ludwig Derodel	Eliot Wilson	Added Overview Recommendation section
0.4	14-Jan-09	Ludwig Derodel	Chris Clayton	Moved the recommendation section.
0.5	17-Feb-09	Ludwig Derodel		RFC
1.0	5-Mar-09	Ludwig Derodel	Global Stakeholders	Final
1.1	3-Oct-11	Ludwig Derodel	Chris Clayton	Added FAST-VP related statements

2 Legal Notice

Information contained herein is for Internal Use and may be used only for business purposes authorized by Citigroup.

CITI, CITIBANK, CITICORP, CITIGROUP and the Arc Device are trademarks and service marks of Citicorp and are used and registered throughout the world. Citibank, Citicorp, Citigroup and their subsidiaries also claim rights in certain other trademarks and service marks contained in these pages.

3 Contact

Please address any comments or questions regarding the information contained in this document via email to the “*GT Global Distributed Storage Engineering” distribution list.

4 Latest Versions

It is recommended that the latest version of the document be used from the Global Distributed Storage Engineering website <https://catecollaboration.citigroup.net/domains/platstor/gdse/products/default.aspx>.
Newer versions may include expanded scope, or comments on issues discovered after deployment.

5 Table Of Contents

1	VERSION CONTROL.....	2
2	LEGAL NOTICE.....	3
3	CONTACT.....	3
4	LATEST VERSIONS.....	3
5	TABLE OF CONTENTS.....	4
6	OVERVIEW	5
6.1	VERSION HISTORY	5
7	RECOMMENDATIONS.....	5
7.1	OVERVIEW OF RECOMMENDATIONS	7
8	SPANNING RISKS	8
8.1	SINGLE ARRAY WITH DIFFERENT FE PORTS – RISK LEVEL 1	8
8.2	MULTIPLE BUT SAME VENDOR STORAGE ARRAYS – RISK LEVEL 2	8
8.3	MULTIPLE BUT DIFFERENT VENDOR STORAGE ARRAYS – RISK LEVEL 3.....	10
9	SUMMARY	11

6 Overview

This document is put together for storage and system administrators by the platform and storage engineering in order to aid in understanding the risks levels associated with spanning host data across multiple storage arrays. Given today's storage growth rate of many applications, it has become nearly impossible to confine a host to a particular set of Front End (FE) ports or a single storage array without performing periodic host migrations. The viable alternative is that of host spanning. Host spanning is the process of allocating storage to a host using more than one set of FE ports or multiple storage array units. Although host spanning is less involved than host migration, it does require some careful planning and consideration.

The main considerations with host spanning revolves around the following: performance associated with RAID and volume types, array-based replication and host level volume management as it relate to data management and array outages. This document is organized to clarify the risk exposure as it relate to the level of spanning.

6.1 Version History

Revision 0.1 – Draft version of policy

Revision 0.2 – Added section for thin provisioning

Revision 0.3 – Enhanced risk level description

Revision 0.4 – Added recommendation overview section

Revision 0.5 – Moved the recommendation section up

Revision 1.0 – Production release

Revision 1.1 – Update to reflect automated storage tiering best practice

7 Recommendations

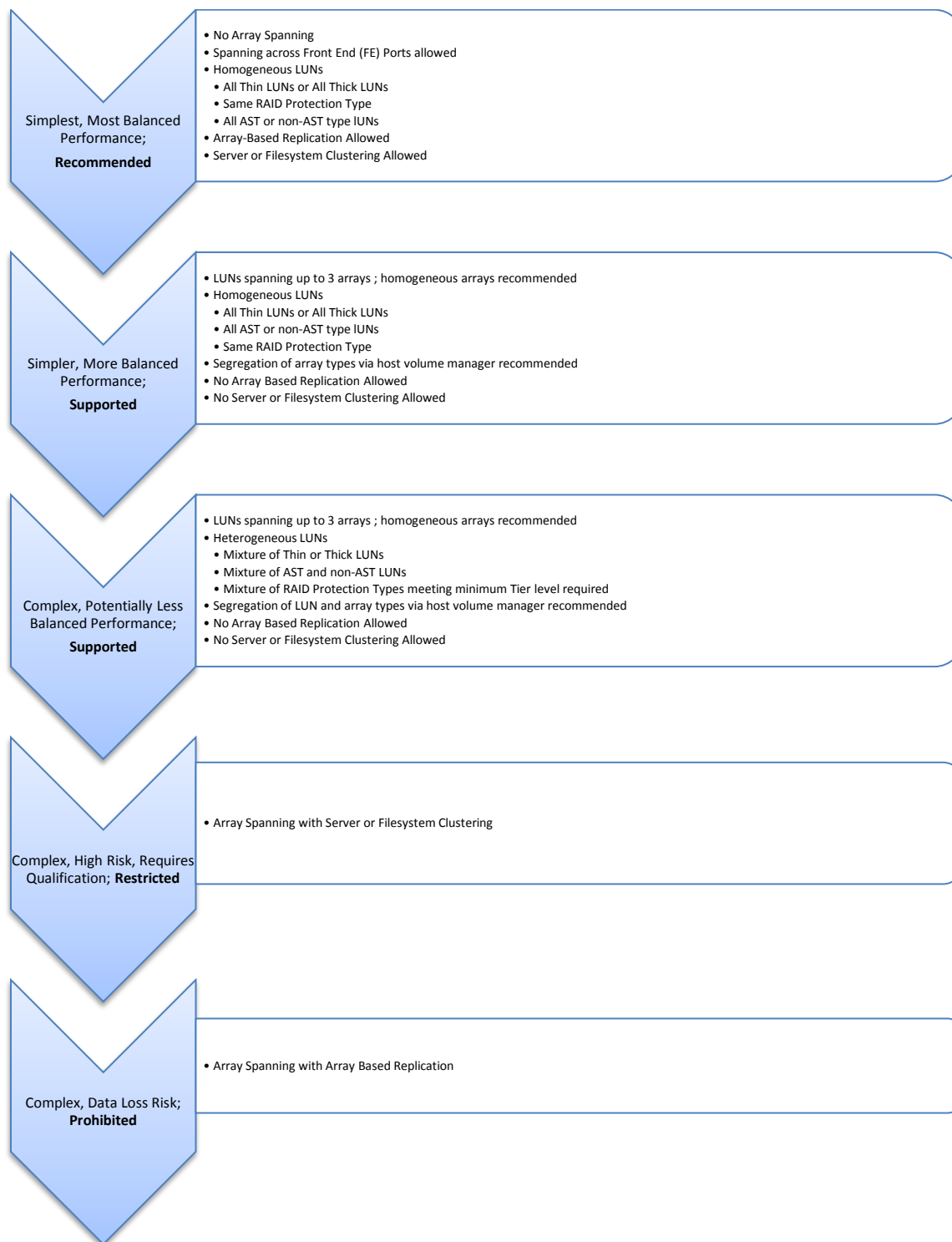
The points listed below (section 8) were put together to highlight the different issues around spanning a host storage across multiple storage arrays or components. It should be noted that there is no substantial technical reasoning or functional limitations to prevent any of configurations mentioned above. Moreover, third party volume manager vendors like Symantec do support such configurations. However, it is our creed to document all known risks associated with storage related configurations to help educate our customers in order to help them chose the proper storage configuration. Complementing the latter statement, the following CATE engineering recommendations were put together to help manage and mitigate risks in either a standalone or host cluster configurations at Citi.

- Whenever possible a single host or host clusters should be confined to a single storage array.
- If the above is not possible and spanning is to be done, the following should be considered:

- o Storage array should be at the same level of microcode whenever possible.
- o Storage array should be on the same SAN fabric. (If not possible, host migration is recommended)
- o Clustered hosts should be restricted to only certified storage solutions (EMC Symmetrix, HDS and now recently certified EMC VNX).
- o Clustered host should not be shared across HDS and EMC solely because of the complexity of managing complexities associated with Host Multi-pathing support or lack thereof on a single host.
- o Logical volume groups should be storage array dependent whenever possible to avoid complete server outage during maintenance on one of the servicing storage arrays. (i.e. not mixing different array LUNs in the same volume group)
- o Spanning should be limited to at most 2 storage arrays whenever possible. If more than 3 arrays are required, host migration should strongly be considered.
- o Host should always be allocated the same storage RAID types (RAID-10/RAID-10 or RAID-5/RAID-5) whenever possible. If not possible RAID types should be segregated by the volume manager if possible; this will minimize possible future application performance degradation.
- o Thin and Thick volumes can co-exist on the same host as long as care is taken on the host to separate the volume groups when possible.
- o Mixing tiers of storage on a server or cluster lowers the overall reliability and performance expectations to that of the lowest tier.
- o With the introduction of Automated Storage Tiering (AST) where data is dynamically moved across the different storage tiers within the array, it is not recommended to mix AST luns with non-AST luns within the same host volume groups. However, there is no restriction in allocating the different type luns to a server.

7.1 Overview of recommendations

The representation below shows the different configurations as they relate to risk starting from the lowest first to the highest.



8 Spanning Risks

Below are some considerations when spanning a single or cluster of hosts across multiple storage arrays, multiple FE ports and/or different RAID types. The risk level ranks from lowest being 1 and 3 being the highest. The risk criteria as stated briefly above will depend on and not limited to the following:

- Host level management complexity of the solution
- Array based management complexity at solution configuration time
- Performance as it relates to different volume characteristics
- Data consistency as it relates to the array based replication

8.1 *Single array with different FE ports – Risk Level 1*

The risk level of this configuration is rated 1 as it is the simplest and most manageable configuration available. The following points depicts some of the pros/cons

PROS

- Host is bound to a pair of FE ports on single storage array (i.e. any failure is solely dependent on a single storage component which is highly available)
- Array LUN configuration is evenly distributed; where each LUN is presented on at least 2 different physical FE ports for redundancy.
- A single host volume group if spread across FE ports of the same array has a lower probability of being impacted by a failure when compared to that of multiple arrays. (i.e. probability of failure of a single host volume increases with number of independent arrays servicing the volume)
- No issues with array based replication
- Homogeneous LUN types

CONS

- Given an existing storage allocation across 2 FE ports and increasing the lun count across 2 additional FE ports; any volume group created using luns from both FE port pairs will have a higher probability of failure because the number of component has doubled from 2 to 4.

In summary, spanning across different FE ports on the same storage array should always be the first option for any host looking to expand in storage capacity. This is the least intrusive and lowest risk option available to date.

8.2 *Multiple but same vendor storage arrays – Risk Level 2*

The risk level of this configuration is rated 2 as the risk introduced by a completely independent array is higher than that of the FE scenario above. The following points depicts some of the pros/cons

PROS

- Arrays are of the same vendor and have undergone the same level of resiliency tests.
- Storage Arrays are preferred to be running the same level of microcode and family sharing similar standard LUN configuration.
- Segregation of array types or volume types via the host volume manager is necessary to mitigate the outages related to the CONs depicted below.

CONS

- Host or host-cluster is spanning across more than a single storage array increasing the probability of failure by adding another independent array with multiple other variables prone to failure.
- Arrays may also be configured at different microcode level or have the different hardware components.
- Clustered host configurations are not supported on some arrays (i.e. Clustering is not supported on CLARiiON arrays)
- Raid type (RAID-10, RAID-5, RAID-6) and array failover mechanism maybe different (Active/Active vs. Active/Passive for CLARiiON arrays) which may cause some performance degradation.
- Volume characteristics can be very different. For example:
 - Thin storage devices tend to different performance characteristic when compare to thick devices of the same size.
 - In some instances a new OS patchset maybe required to fully support Thin Volumes on some OS platforms (AIX). Furthermore some OS platforms like Windows completely negate Thin Provisioning functionalities due to lack of support from NTFS.
 - Thin devices are allocated from a free pool of storage on the array and if additional physical devices are not added to the pool prior to the pool filing up, any host application residing in the pool will experience and un-planned outage.
 - Spanning across arrays with array-based replication is not supported today.
 - Lack of migration tools from Thick to Thin will negate the benefits of Thin Provisioning

In summary, the risk of spanning a host or host-cluster across multiple arrays is greater than that of a single array and as a result further consideration is required. Although each individual storage array is very resilient and highly redundant in terms of hardware, but outages due happen whether caused by microcode or user interaction. In such scenario, depending on how the volume groups were configured on the host, the outage may have some serious impact on the host and even cause data loss. In some rare instances, it is possible in this situation to have data corruption on the host depending on the resiliency of the volume manager software.

8.3 Multiple but different vendor storage arrays – Risk Level 3

The risk level of this configuration is rated 3 as the risk introduced by a completely independent vendor array is the highest of all the scenarios. The following points depicts some of the pros/cons

PROS

- Ability to allow a host to expand storage requirements without migrating
- Viable solution for applications requiring different tiers of storage (i.e. Tier 2 and Tier 4) where the lower tier will be configured and used for archiving or dump space

CONS

- Host or host-cluster is spanning across different storage array increasing the probability of failure by adding another independent; likely competing technologies.
- Limited deployments and testing by vendors.
- Storage array type and microcode will not be the same and will have different error handling sub-routines.
- Clustered host configurations may not be supported on some arrays (i.e. CLARiiON).
- Cluster specific flags or settings may behave differently based on storage vendor implementation.
- Certain volume manager products (i.e. Veritas) do not recommend using Powerpath as the multipathing product on non-EMC arrays for cluster configurations
- Configuration and support of multiple path management software will increase configuration complexity
- Array-based replication is not supported since 2 different replication methods would need to be deployed in this scenario and therefore impossible to guarantee data consistency
- Volume characteristics can be very different. For example:
 - o Thin storage devices tend to have a slightly lower performance characteristic over all when compare to thick devices of the same size.
 - o In some instances a new OS patchset maybe required to fully support Thin Volumes on some OS platforms (AIX). Furthermore some OS platforms like Windows completely negate Thin Provisioning functionalities due to lack of support from NTFS.
 - o Thin devices are allocated from a free pool of storage on the array and if additional physical devices are not added to the pool prior to the pool filing up, any host application residing in the pool will experience and un-planned outage.
 - o Spanning across arrays with array-based replication is not supported today.

- o Lack of migration tools from Thick to Thin will negate the benefits of Thin Provisioning

In summary, the risks associated with spanning a single host or host-cluster across competing vendor arrays is the highest of all and should only be considered only in extreme cases when all other possible options have been exhausted including data migration. Considering certain cluster configuration, IO fencing may not be feasible due to requirements on voting (quorum) disks. The quorum disks will be created on one array and any outage on that array will cause the entire cluster to fail regardless of the other storage array's availability.

9 Summary

As stated throughout this document, the main theme is to provide the storage management and system administrator teams with some guidelines to apply during spanning considerations. The recommendations provided by engineering were based predominantly around managing risks and not related to any technical reasons. The final decision for risk acceptance should be a collaborative discussion between storage management and the business application team.