IBM **Spectrum Storage**

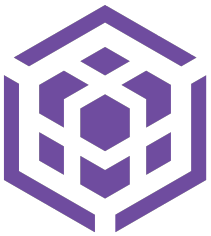
Achieving High Availability with IBM Spectrum Virtualize

Clarence POUTHIER, IBM Systems Data Architect, Europe

clarence_pouthier@fr.ibm.com

+33 6 71 92 26 25


cpouthier21



IBM
Spectrum
Virtualize

10/05/2017

Agenda

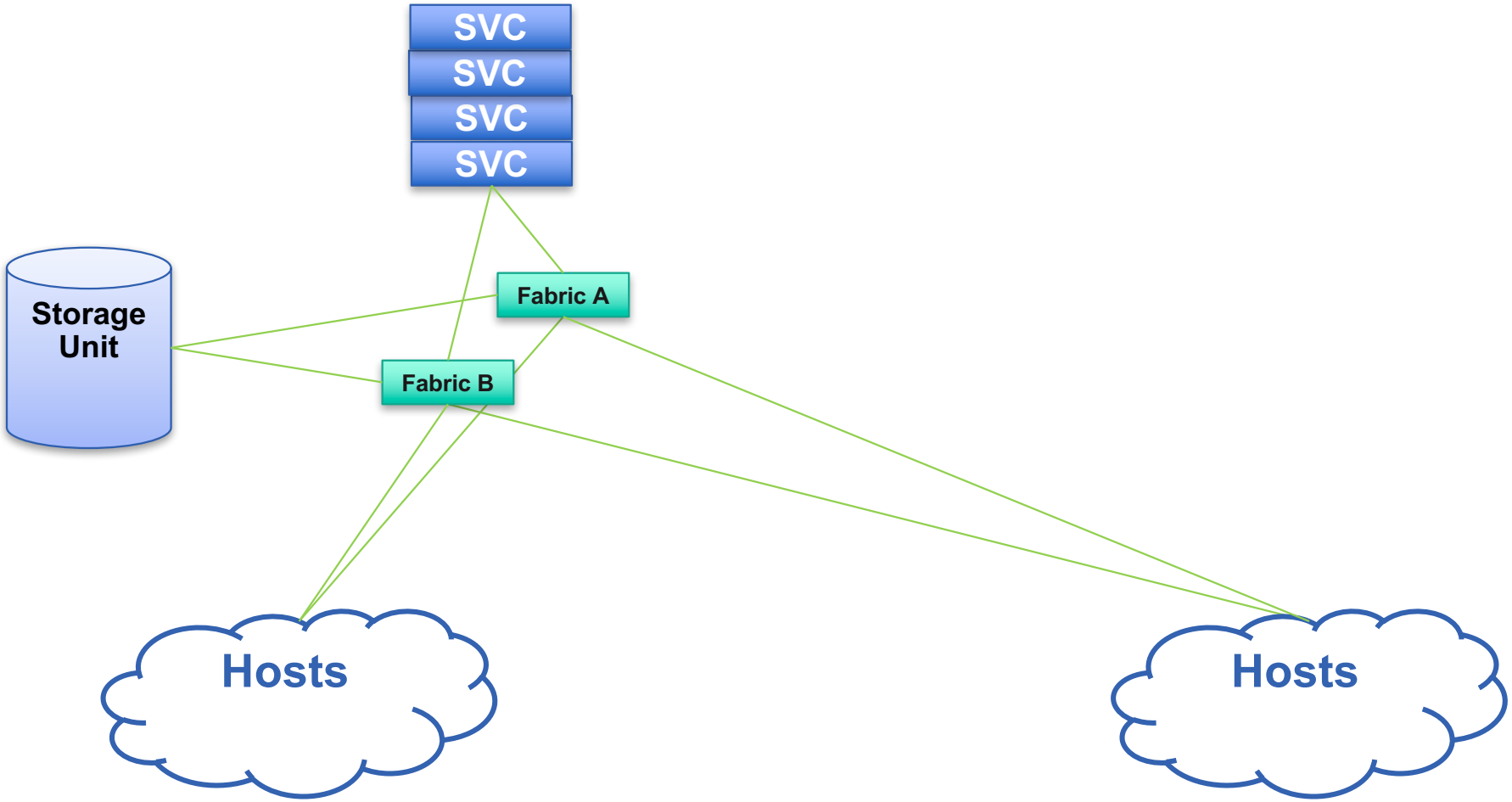
- Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture
- Implementing IBM Spectrum Virtualize with Hyperswap
- Considerations and comparison

Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

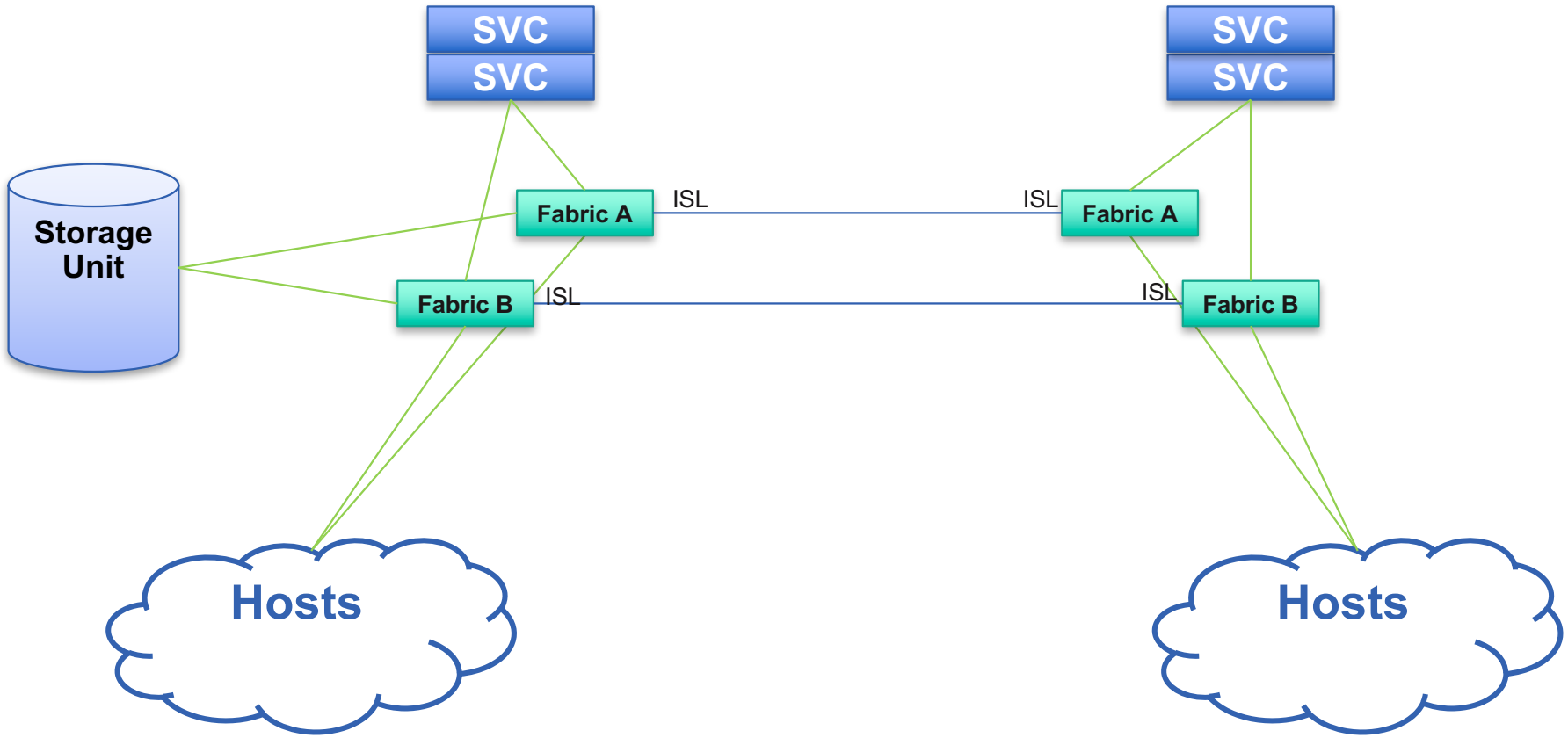
Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

- It allows Spectrum Virtualize hardware to be geographically distributed. Stretched cluster deployments can span as short a distance as two racks in a data center, two buildings in a campus, across a city, or as far as 100 km or potentially more.
- The key benefit of a stretched cluster, compared to Metro Mirror, is that it allows for fast non-disruptive failover in the event of small scale outages.
- Another aspect of this attractiveness is that SAN Volume Controller always has an automatic quorum to act as a tie-break. This means no external management software or human intervention is ever required to perform a failover except particular case with override quorum command to restart half of the nodes on the surviving site which is only available in a ESC configuration.
- A key benefit of a ESC configuration vs. classical stretched cluster, is that it allows local cache destage avoiding sending many times the same write I/O over intersite links

Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

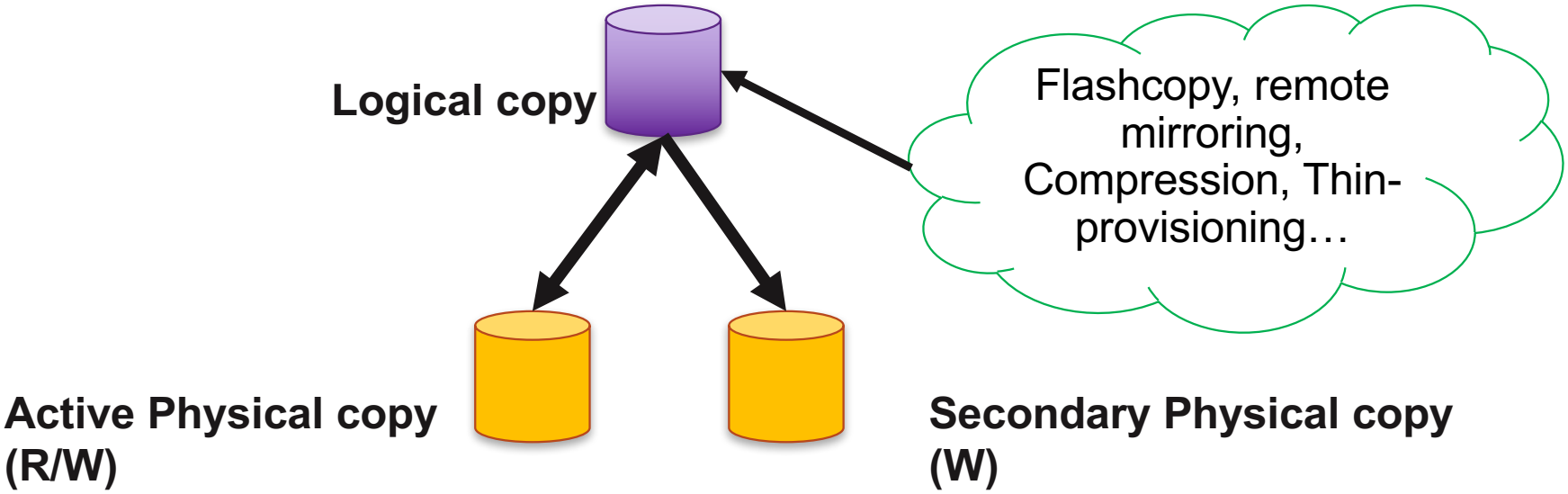


Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

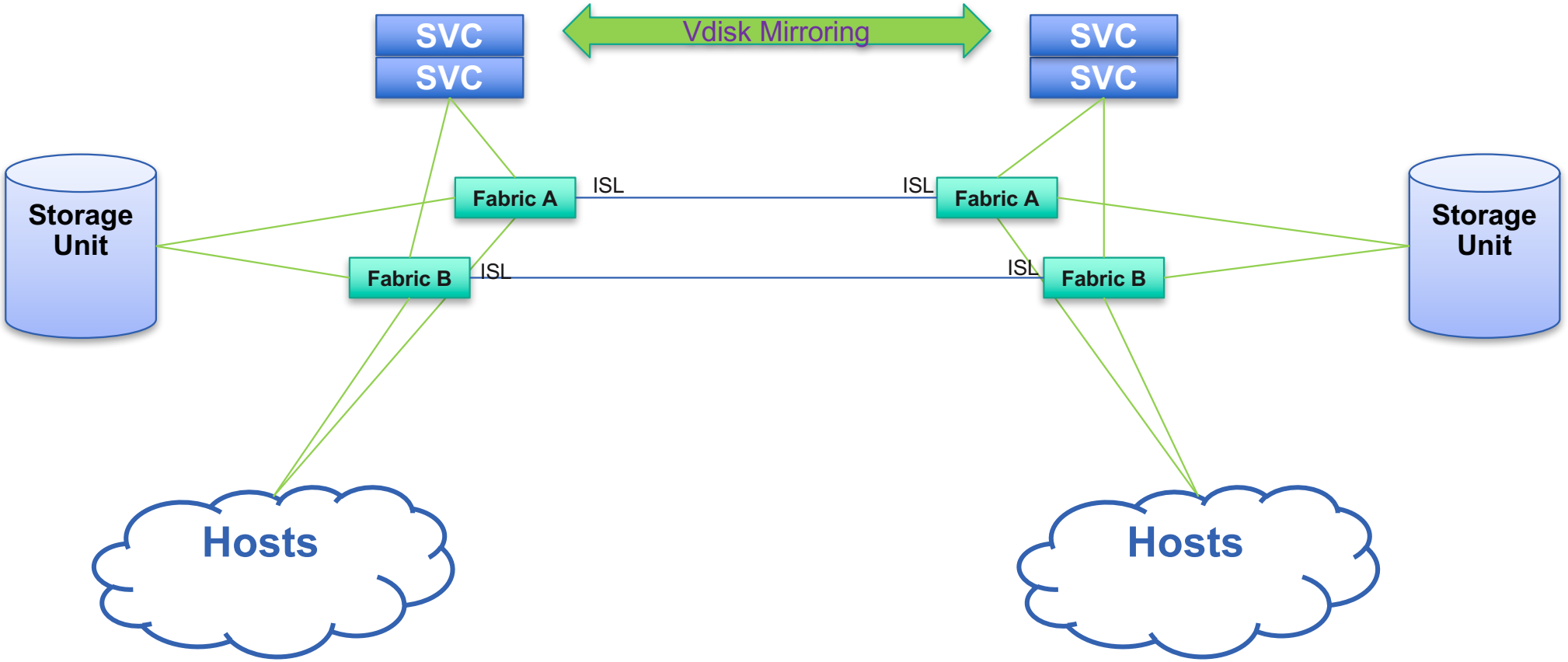


Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

- The Enhanced Stretched Cluster configuration uses the Spectrum Virtualize volume mirroring function, which allows the creation of one volume with two copies.
- The resynchronization between both copies is incremental and is started automatically.
- In the Spectrum Virtrualize software stack, volume mirroring is below the cache and copy services. A mirrored volume has the same functions and behavior as a standard volume.



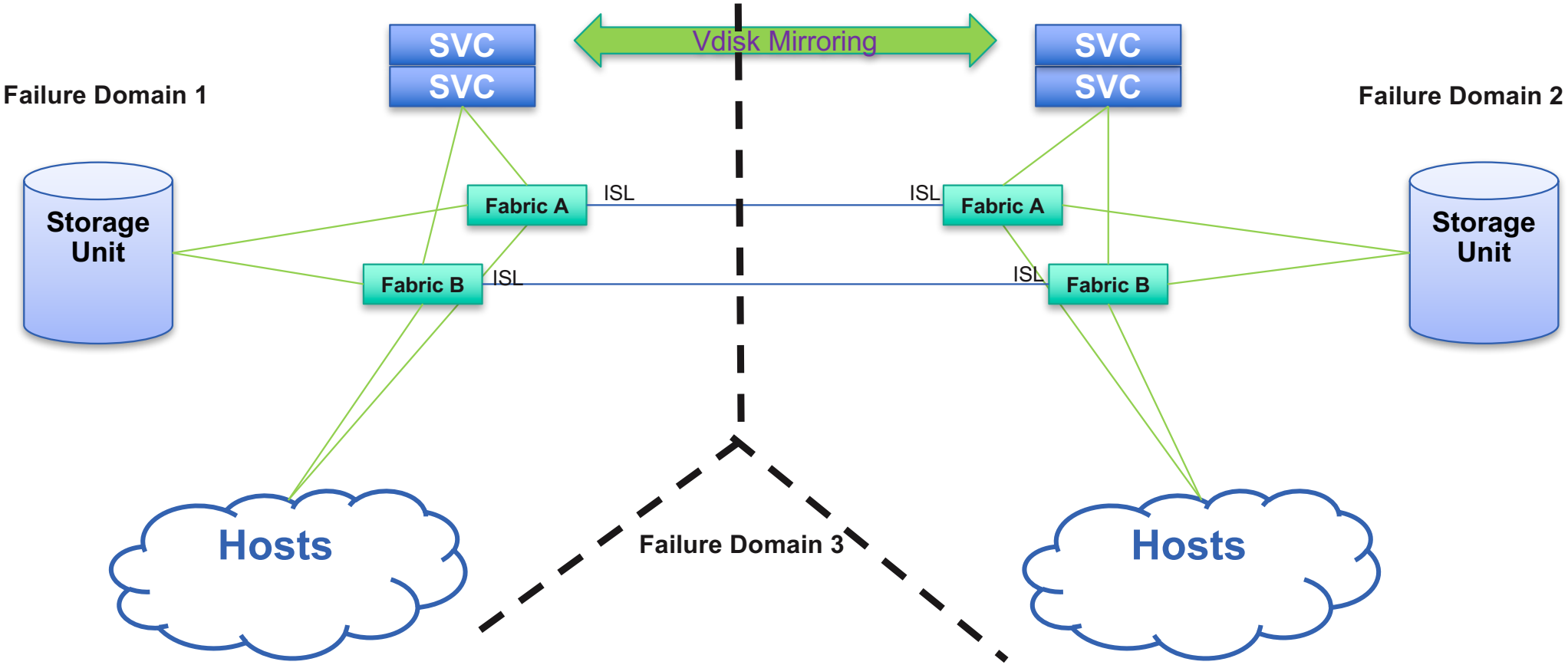
Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture



Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

- In an ESC configuration, the term *failure domain* is used to identify **components of the Spectrum Virtualize cluster that are contained within a boundary** so that any failure that occurs (such as power failure, fire, or flood) **is contained within that boundary**.
- Failure domains are typically **areas or rooms in the data center, buildings on the same campus, or even buildings in different towns**. Different kinds of failure domains protect against different types of failure conditions:
  - If each failure domain is an area with a separate electrical power source within the same data center, the cluster can **maintain availability if any single power source fails**.
  - If each site is a different building, the cluster can **maintain availability if a loss of any single building were to occur (for example, power failure or fire).**
- Ideally, **each of the three failure domains** that are used for the ESC configuration is in a separate building and powered by a separate power source. **This configuration offers the highest level of protection** against all possible failure and disaster conditions.

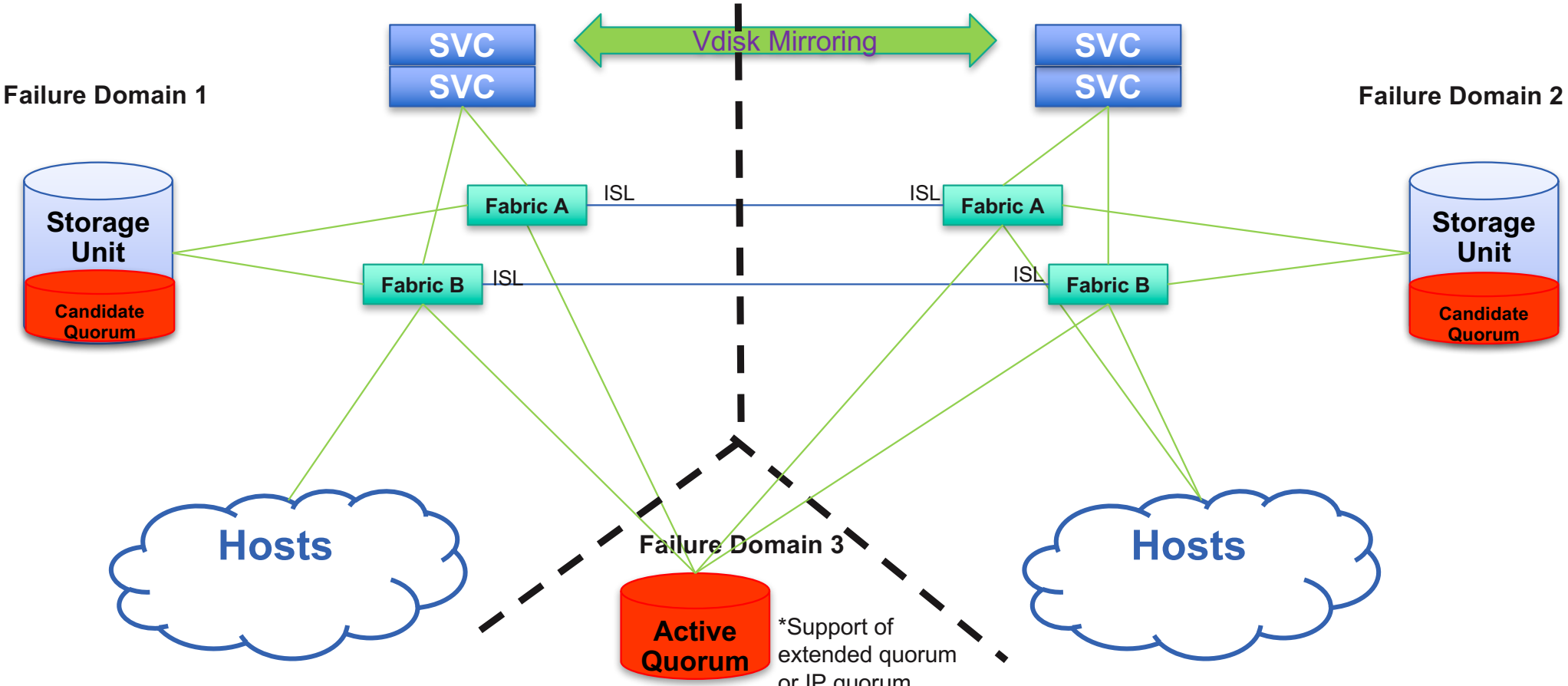
Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture



Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

- The quorum disk fulfills two functions for cluster reliability:
  - **Acts as a tiebreaker** in split brain scenarios
  - **Saves critical configuration metadata**
- The SAN Volume Controller quorum algorithm distinguishes between the **active** quorum disk and quorum disk **candidates**.
- A quorum must be placed **in each of the three failure domains**.
- The quorum can be manually selected (ESC) in case of a rolling disaster (to restart the half surviving cluster).

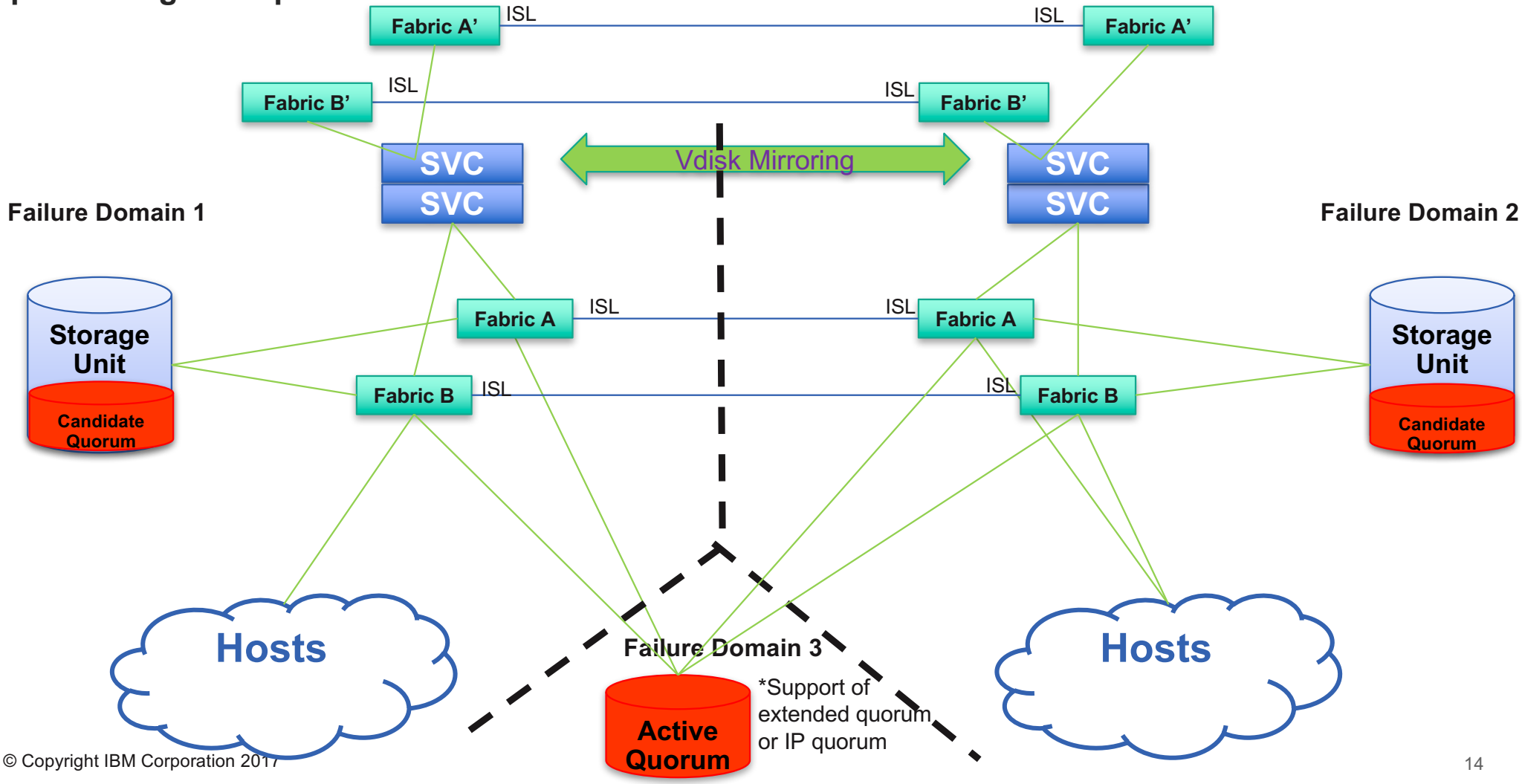
Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture



Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

- Enhanced stretched cluster supports three different approaches for node-to-node intracluster communication between failure domains:
  - **No ISL configuration**
  - **ISL configuration**
  - **FCIP configuration**
- Without ISL: may need to use **LW SFPs** to connect nodes to opposite fabrics
- With ISL: need to have **1 dedicated ISLs per independant (private) fabric** to ensure node-to-node communication (use of xISL allowed under specific support request)

Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture



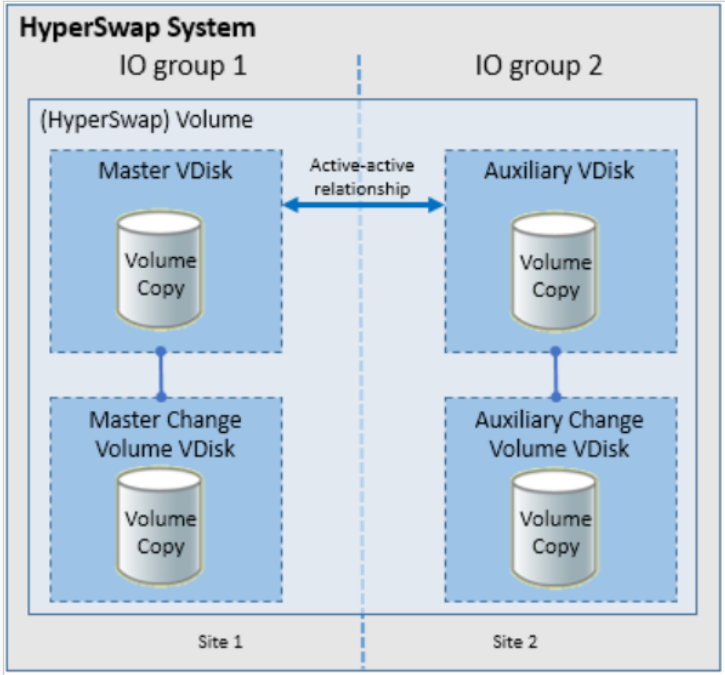
Implementing IBM Spectrum Virtualize in an Enhanced Stretched Cluster architecture

Failure domain 1 (Node 1)	Failure domain 2 (Node 2)	Failure domain 3 (Quorum Disk)	Cluster status
Operational	Operational	Operational	Operational, optimal
FAILED	Operational	Operational	Operational, write cache disabled
Operational	FAILED	Operational	Operational, write cache disabled
Operational	Operational	FAILED	Operational, Active quorum disk moved
Split Brain	Split Brain	Operational	Quorum race
Operational	FAILED	FAILED	Stopped, override on surviving site possible
FAILED	Operational	FAILED	Stopped, override on surviving site possible

Implementing IBM Spectrum Virtualize with Hyperswap

Implementing IBM Spectrum Virtualize with Hyperswap

- Hyperswap has been introduced in 7.5 code version for IBM Spectrum Virtualize, IBM Storwize V7000 and V5000.
- Hyperswap was originally designed to provide high availability on Storwize systems avoiding small customers to stack IBM Spectrum Virtualize upfront.
- As code is shared between Storwize and Spectrum Virtualize, Hyperswap is available for both systems.
- Same architecture pre-requisites as Enhanced Stretched Cluster (i.e. Private and Public SAN)
- As for Enhanced Stretched Cluster, hosts accessing storage through iSCSI or SAS (V5000 only) won't take advantage of high availability function.



HyperSwap volume

Implementing IBM Spectrum Virtualize with Hyperswap

- Unlike Enhanced Stretched Cluster architecture, Hyperswap is not about dividing 2 nodes form the same IO group on two sites, but separating 2 IO groups of the same cluster.
- HyperSwap caches in 2 IO separate groups from the same cluster
  - If one site is unavailable, the remaining site has a full IO group with caching available
  - No drop in performance in a disaster
- HyperSwap adds application awareness
  - Group together multiple volumes to ensure application-wide failover
  - Ensures at least one site has latest copy of all volumes in group
- HyperSwap retains old consistent data during resynchronization
  - Allows sites to continually provide disaster recover protection through lifecycle
- The system will use ALUA to advertise a path on the node or nodes on the same site as the host as the optimized (preferred) path.
- In 7.5, all configuration tasks are done through CLI. With 7.6 configuration steps are embedded through GUI.

Considerations and comparison

Considerations and comparison

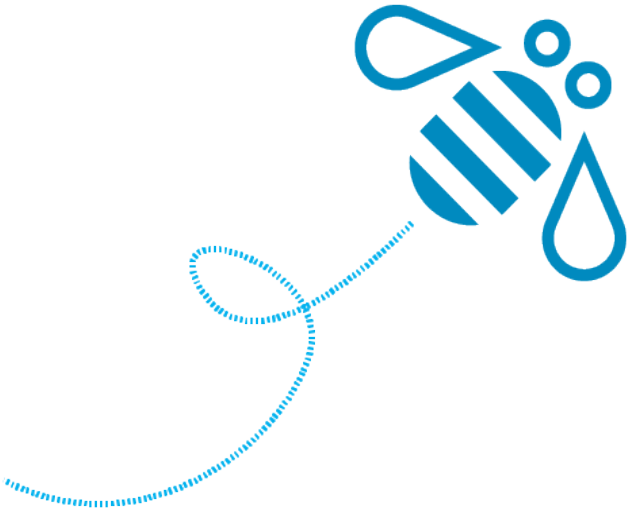
	Enhanced stretched cluster	Hyperswap
Available for	SVC	SVC/Storwize
Two independent sites	Yes	Yes
Two independent servers, one in each site	Yes	Yes
Latency optimization	Local read	Local read
Performances in case of failover	Write cache disabled	No performance impact
Licensing	N/A as Vdisk mirroring in embedded in the base code	Remote mirroring required
3rd site replication	Possible	Impossible

Considerations and comparison

- Enhanced Stretched Cluster and Hyperswap are providing recovery options and IO path optimization.
- Properly configured, Hyperswap can survive most disasters without user interaction, with most remaining disasters giving some usable consistent data.
- High-available architecture should be considered also for servers and applications.
- Keep in mind that enhanced stretched cluster or Hyperswap should be completed with a disaster recovery plan.
- Replication to a 3rd site with Enhanced Stretched Cluster still remains in case one site fails.
- Moving from a dual-site replication storage to Hyperswap is not doable (easily).

Questions?

Merci !



Legal Notices

Product data has been reviewed for accuracy as of the date of initial publication. Product data is subject to change without notice. This document could include technical inaccuracies or typographical errors. IBM may make improvements and/or changes in the product(s) and/or program(s) described herein at any time without notice. Any statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only. References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business. Any reference to an IBM Program Product in this document is not intended to state or imply that only that program product may be used. Any functionally equivalent program, that does not infringe IBM's intellectually property rights, may be used instead.

THE INFORMATION PROVIDED IN THIS DOCUMENT IS DISTRIBUTED "AS IS" WITHOUT ANY WARRANTY, EITHER OR IMPLIED. IBM LY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NONINFRINGEMENT. IBM shall have no responsibility to update this information. IBM products are warranted, if at all, according to the terms and conditions of the agreements (e.g., IBM Customer Agreement, Statement of Limited Warranty, International Program License Agreement, etc.) under which they are provided. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products in connection with this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. IBM makes no representations or warranties, ed or implied, regarding non-IBM products and services.

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents or copyrights. Inquiries regarding patent or copyright licenses should be made, in writing, to:

IBM Director of Licensing
IBM Corporation
North Castle Drive
Armonk, NY 1 0504- 785
U.S.A.

Information and Trademarks

Notes:
Performance is in Internal Throughput Rate (ITR) ratio based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput that any user will experience will vary depending upon considerations such as the amount of multiprogramming in the user's job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve throughput improvements equivalent to the performance ratios stated here.

All customer examples cited or described in this presentation are presented as illustrations of the manner in which some customers have used IBM products and the results they may have achieved. Actual environmental costs and performance characteristics will vary depending on individual customer configurations and conditions.
This publication was produced in the United States. IBM may not offer the products, services or features discussed in this document in other countries, and the information may be subject to change without notice. Consult your local IBM business contact for information on the product or services available in your area.
All statements regarding IBM's future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

Prices subject to change without notice. Contact your IBM representative or Business Partner for the most current pricing in your geography.

This presentation and the claims outlined in it were reviewed for compliance with US law. Adaptations of these claims for use in other geographies must be reviewed by the local country counsel for compliance with local laws.

Special notices

This document was developed for IBM offerings in the United States as of the date of publication. IBM may not make these offerings available in other countries, and the information is subject to change without notice. Consult your local IBM business contact for information on the IBM offerings available in your area.

Information in this document concerning non-IBM products was obtained from the suppliers of these products or other public sources. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products.

IBM may have patents or pending patent applications covering subject matter in this document. The furnishing of this document does not give you any license to these patents. Send license inquires, in writing, to IBM Director of Licensing, IBM Corporation, New Castle Drive, Armonk, NY 10504-1785 USA.

All statements regarding IBM future direction and intent are subject to change or withdrawal without notice, and represent goals and objectives only.

The information contained in this document has not been submitted to any formal IBM test and is provided "AS IS" with no warranties or guarantees either expressed or implied.

All examples cited or described in this document are presented as illustrations of the manner in which some IBM products can be used and the results that may be achieved. Actual environmental costs and performance characteristics will vary depending on individual client configurations and conditions.

IBM Global Financing offerings are provided through IBM Credit Corporation in the United States and other IBM subsidiaries and divisions worldwide to qualified commercial and government clients. Rates are based on a client's credit rating, financing terms, offering type, equipment type and options, and may vary by country. Other restrictions may apply. Rates and offerings are subject to change, extension or withdrawal without notice.

IBM is not responsible for printing errors in this document that result in pricing or information inaccuracies.

All prices shown are IBM's United States suggested list prices and are subject to change without notice; reseller prices may vary.

IBM hardware products are manufactured from new parts, or new and serviceable used parts. Regardless, our warranty terms apply.

Any performance data contained in this document was determined in a controlled environment. Actual results may vary significantly and are dependent on many factors including system hardware configuration and software design and configuration. Some measurements quoted in this document may have been made on development-level systems. There is no guarantee these measurements will be the same on generally-available systems. Some measurements quoted in this document may have been estimated through extrapolation. Users of this document should verify the applicable data for their specific environment.