PowerHA Tools
Full System Copy Manager

—

David Painter
Senior Managing Consultant
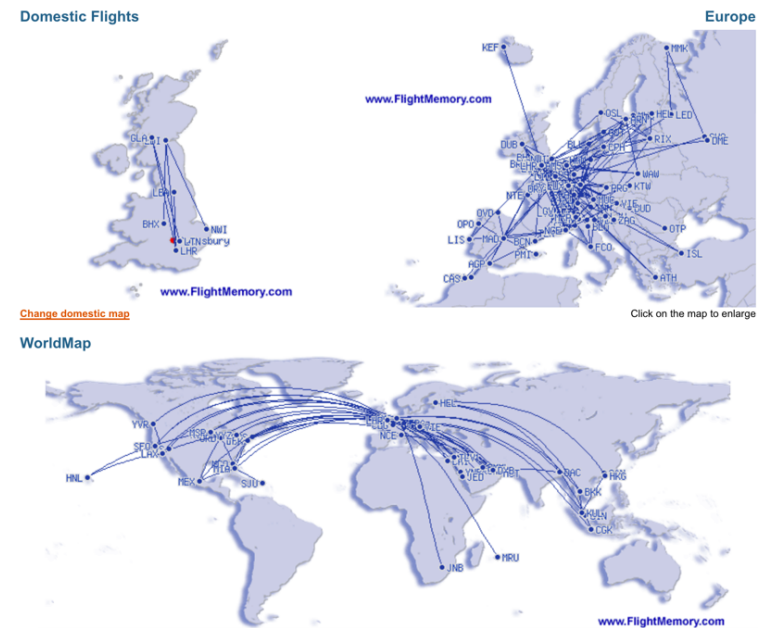
IBM Systems Lab Services
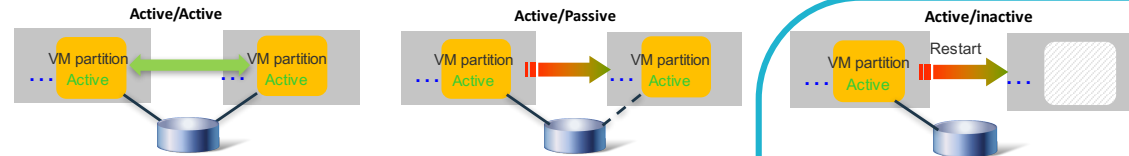## Pan-IOT HA Practice

Agenda

This presentation covers the options available for a Full System IBM i system image

This includes a brief overview of CA/HA/DR options but is primarily focused on the FlashCopy capabilities

Standard Industry High Availability Classifications & Positioning

Technology	Active/Active Clustering	Active/Passive Clustering	Active/Inactive
Definition	All VMs in the cluster have access to the production data	One VM in the cluster has access to the data, two logical OS instances, two physical copies	One VM in the cluster has access to the data, one logical OS, two physical copies
Positioning	Mission Critical Applications	Mission Critical Applications	Non-Mission Critical Applications
Outage Types	All outage types	All outage types	Primarily hardware outages
OS integration	Inside the OS	Inside the OS	OS agnostic
RPO	Sync/Async	Sync/Async	Sync/Async
RTO	0	Fast (minutes)	Fast Enough (VM Reboot)
Licensing*	N+N	N+1 licensing	N+0 licensing
Industry Examples	Oracle RAC, Sysplex, HANA DB2 Mirror for IBM i	PowerHA, Redhat HA, Linux HA	Vmware SRM, GDR, FSR

* N = number of licensed processor cores on each system in the cluster

     illustrations represent two node shared storage configurations for definitional simplicity

     there are many other topologies and data resiliency combinations

What options exist from IBM for Full System IBM i availability

Continuous Availability

This is provided by the DB2 Mirror capabilities at IBM i 7.4, this provides an Active/Active solution with two systems working with the same database, therefore this provides continuous availability

High Availability

No options exist, as an IPL is unavoidable

Disaster Recovery

This is provided by PowerHA Tools – Full System Replication manager, and provides a DR capability with either synchronous (zero loss of data) or asynchronous (some loss of data) replication

Backup window reduction

This is provided by PowerHA Tools – Full System Copy manager, and provides for a zero impact copy of the production system for use independently, thus eliminating an outage for backups etc.

Continuous Availability - IBM Db2 Mirror for i

Application

Db2 Mirror

Requires POWER8 or later and IBM i 7.4

New IBM i LPP: 5770-DBM
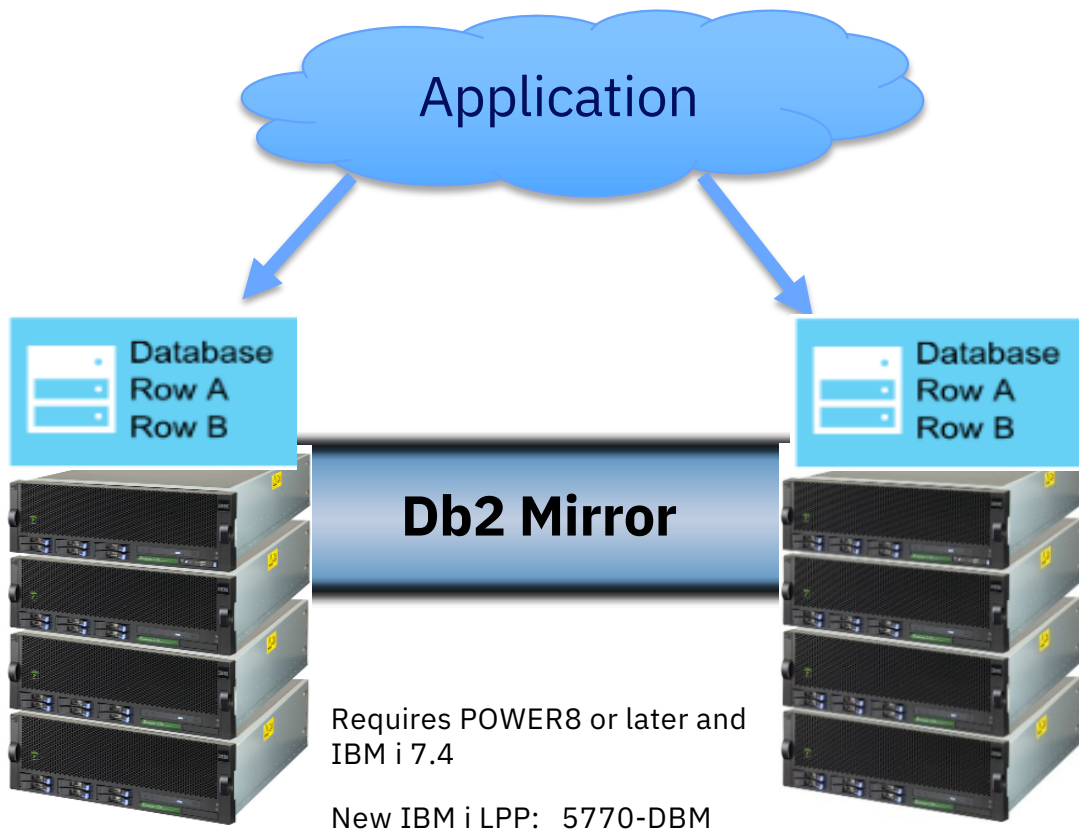
IBM Db2 Mirror for i

- Data Center Solution for **continuous availability**

- High speed synchronous replication of Db2 for I

- Does not utilize Db2 journaling

- Access Db2 objects from either LPAR

Application Availability Enablement

- Two Nodes read and write to the same DB Files

- Enables quickly moving all work to one node, for planned maintenance or node failure

Non Disruptive System Upgrades

- Nodes can be at different OS levels

- Nodes can be on different Power Hardware Generations

- Rolling upgrades for no downtime

- Roll a node back a release with minimal impact if Active/Active applications are deployed
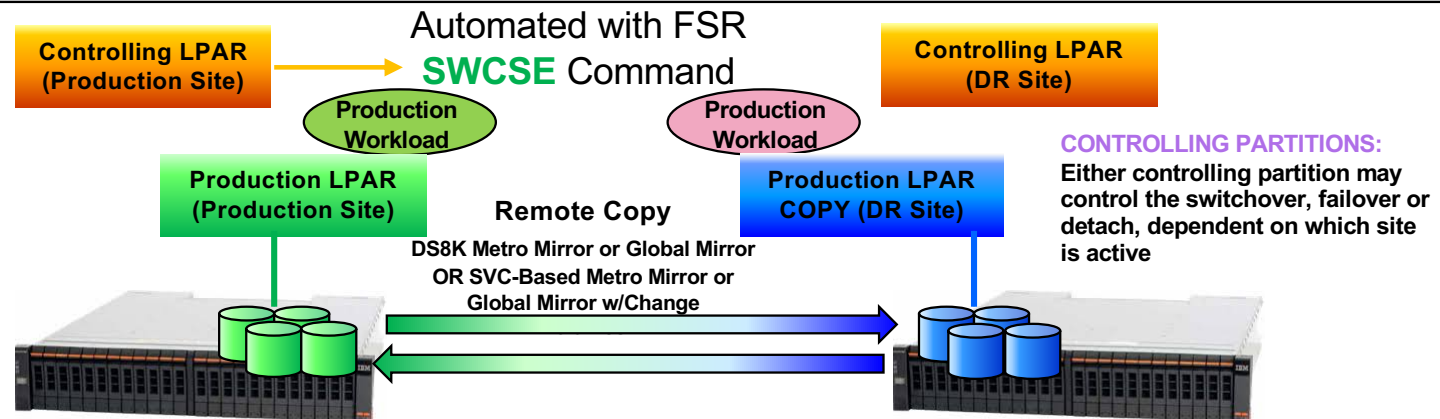
Full System Replication (FSR) for i – Automated Failover Operations

End applications/jobs on production site

Monitor/wait for all jobs to end

Issue the shut-down command

Monitor/wait for shut-down to complete

Log in to storage as Admin

Verify/wait for disk synchronization

Issue the switch operation

**MANUAL STEPS AND LOG-INS NO LONGER REQUIRED!**

Log in to HMC at DR site as superadmin

IPL the DR site partition in manual mode

Modify "autostart" objects (lines, interfaces, devices, applications) to not start

Correct comm resources, storage resources, IP interfaces, TCP routes, etc.

Apply license keys

Complete startup processes

Automated with FSR
SWCSE Command

Controlling LPAR (Production Site)

Production Workload

Production LPAR (Production Site)

Remote Copy

DS8K Metro Mirror or Global Mirror
OR SVC-Based Metro Mirror or
Global Mirror w/Change

Controlling LPAR (DR Site)

Production Workload

Production LPAR COPY (DR Site)

CONTROLLING PARTITIONS:
Either controlling partition may control the switchover, failover or detach, dependent on which site is active

Backup Window Reduction

FlashCopy – What is it?

- A way to make a copy (or what behaves like a copy) of a full-system image in a matter of a few seconds

- A "point in time" copy – not continuous replication of changes

- Copies the DISK units - anything in memory that hasn't been sent to disk isn't copied.

- Requires external storage

- Copy is performed by the storage system

Production System (LPAR)

External Disks

FlashCopy Disks

Flash Copy!

DS8000 / SVC
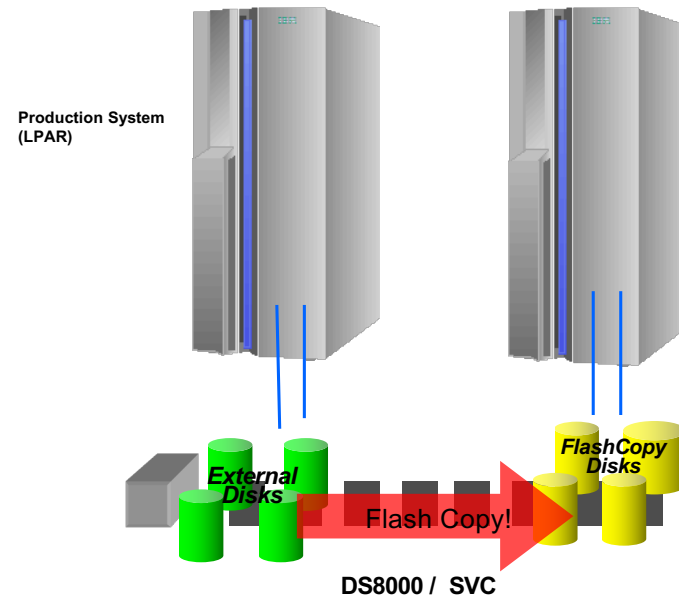
FlashCopy - What can you do with it?

- Quickly make a "point-in-time" snapshot copy of the logical disk units for a full-system image.

- Reduce production down time for performing saves.

  - FlashCopy allows you to make a copy, then backup the copy with minimal impact to the production system availability.

  - Off-load save operation to another partition (LPAR) or system

- The FlashCopy is immediately usable to present to another partition and Activate (full-system image) for full read/write use.

- FlashCopy can be taken at time X and brought on-line to the target LPAR hours or days later – actual spinning of data to tape can occur during prime shift, even if FlashCopy is taken during the night or on the weekend.

- May even eliminate the need for save-to-tape in some circumstances.

- Facilitates cloning systems and / or creating data reset points

- Requires external storage

FlashCopy – A simple example

A FlashCopy of a full system image for performing saves of production data from a different partition

The FlashCopy system is just an LPAR definition and hardware resources – no separate IBM i instance for the LPAR exists until the FlashCopy is made.

Production System
(LPAR)

FlashCopy
(LPAR)

Tape

Flash Copy!

*Production
Disks*

DS8000 / SVC
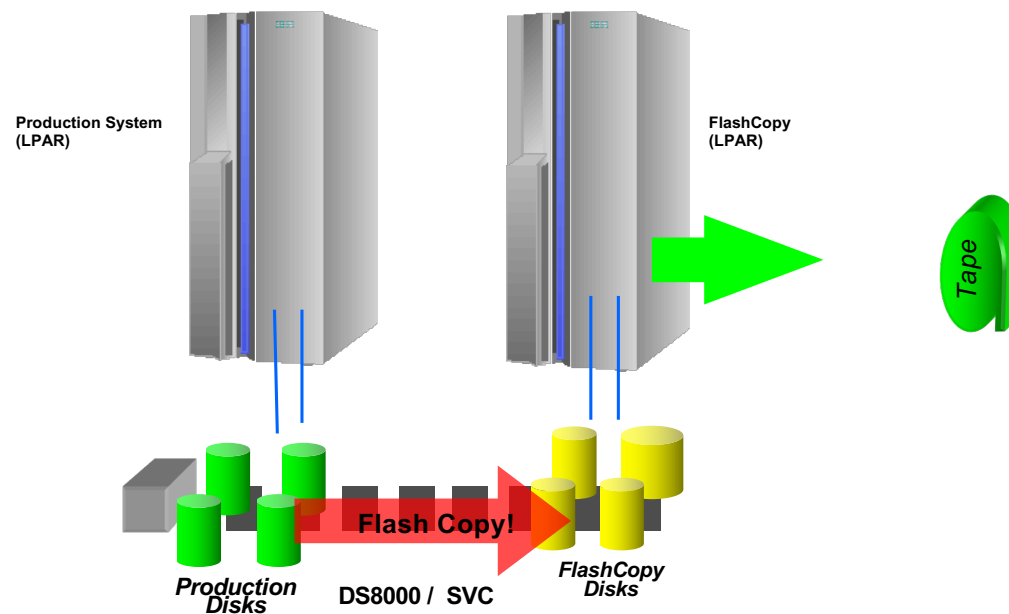
*FlashCopy
Disks*

FlashCopy - How does it work?

FlashCopy creates a copy of a logical volume at a specific snapshot in time and is referred to as a point-in-time (PIT) Copy, an instantaneous copy, or T0 copy (time-zero copy).

The FlashCopy function is only supported within the same storage unit. You cannot establish a FlashCopy relationship between volumes located in different storage units.

When you request a FlashCopy, a relationship is established between a source and target volume. Both are considered to form a FlashCopy pair.

The FlashCopy relationship typically takes only a few seconds to complete regardless of the number of volumes that are flashed, and regardless of the FlashCopy attributes selected
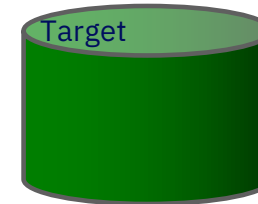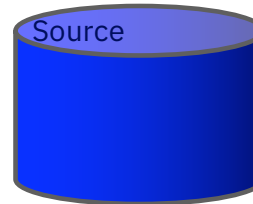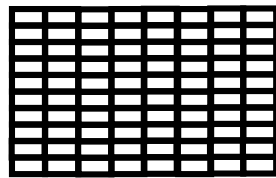
After a FlashCopy relationship is established, the source and target volumes are immediately available for processing with full read/write access to both the source and the target.

Requires an LPAR or system to own the FlashCopy instance in order to bring up the flash and work with it (save it).

A volume can either be the source or the target of a FlashCopy relationship – not both

FlashCopy – under the hood

Initialize target bitmap

Source

Target

When the FlashCopy is started, a relationship between each source and target volume is established within seconds by creating a pointer table, including a bitmap for the target disk tracks.

All bits for the bitmap are set to their initial values which means no data has been copied so far.

At this point, all read I/O against the FlashCopy target tracks will actually be redirected to the source tracks.

FlashCopy – under the hood

An attempt is made to write a source track.

Source write

Target write
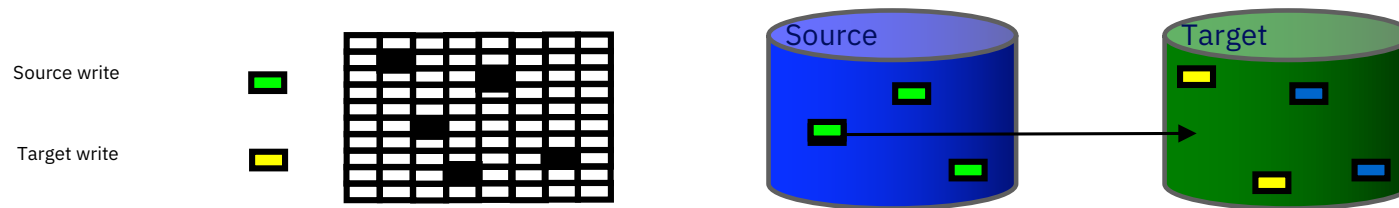
Source

Target

- The request is intercepted and the source track is copied to the target volume before the update occurs. This maintains the Time 0 copy on the target.

- After the copy to the target volume, the bitmap is updated and the new source track value is written .

- An attempt is made to write a track on the target copy.

- Writes to the target volume would proceed and the bitmap is updated to prevent any write to the source track from overlaying the target volume.

FlashCopy dictionary

Temperature of the flash:
Indicates the amount of disruption to the source system (production), and the quality of the FlashCopy target image, when any given FlashCopy is made.

- Cold

- Quiesced

- Warm

Type of FlashCopy relationship:
FlashCopy relationship attributes that are Internal to the storage system and affect how the storage system manages the copy. These attributes are irrelevant to how IBM i sees and accesses the FlashCopy.

- Full copy

- No copy

- Persistent / Incremental

Type of FlashCopy volumes:
How the FlashCopy target logical disk units are created and allocated on the storage system. These attributes are irrelevant to how IBM i sees and accesses the FlashCopy.

- Fully provisioned volumes

- Space Efficient volumes / Thin Provisioned volumes

FlashCopy temperature - cold

Temperature of the flash:
Indicates the amount of disruption to the source system (production), and the quality of the FlashCopy target image, when any given FlashCopy is made.

A FlashCopy ONLY copies what has been sent to disk. Any updates sitting in memory on IBM i are not known to the storage system, and are not included in the FlashCopy image.

- **Cold**: All memory is flushed to disk; guaranteed

  - If full system flash, power down the source system before performing the flash

  - If IASP flash, vary off the source IASP before performing the flash

  - The ONLY way to guarantee all writes held in memory are sent out to disk and included in the FlashCopy image

  - The ONLY FlashCopy temperature that will result in a "normal" bring-up of the copy on the target LPAR.

  - **IBM Recommended technique, if outage for shutdown/startup or vary off/vary on is acceptable – will result in best FlashCopy quality, but most source system disruption**

FlashCopy temperature - warm

Temperature of the flash:
Indicates the amount of disruption to the source system (production), and the quality of the FlashCopy target image, when any given FlashCopy is made.

A FlashCopy ONLY copies what has been sent to disk. Any updates sitting in memory on IBM i are not known to the storage system, and are not included in the FlashCopy image.

  – **Warm:** No memory is flushed to disk; guaranteed

  • Source system has zero disruption or impact when the FlashCopy is taken

  • NO writes held in memory are pushed to the storage system, and as such, are excluded from the FlashCopy target image

  • Crash quality copy is made, as objects may have been partially paged out to disk when the FlashCopy is taken

  • Will always result in an "abnormal" bring-up of the copy on the target LPAR.

  • **May result in lower quality FlashCopy target image. Proper journaling for data recoverability will improve the quality.**

Warm FlashCopy expectations

Database / IFS Objects

Data in the DB will not have integrity without journaling and anything in memory is lost.

If DB/IFS objects are journaled and damage occurs, a call to IBM Support is valid.

System (IBM) Objects

System should recover on vary-on or IPL. If not, a call to IBM support is valid

If using FlashCopy for off-line backups, object recoverability on FlashCopy target image is critical. Journal!

FlashCopy temperature – quiesced

Temperature of the flash:
Indicates the amount of disruption to the source system (production), and the quality of the FlashCopy target image, when any given FlashCopy is made.

A FlashCopy ONLY copies what has been sent to disk. Any updates sitting in memory on IBM i are not known to the storage system, and are not included in the FlashCopy image.

- **Quiesced:** Some memory is flushed to disk

  - Source system has minimal disruption or impact when the FlashCopy is taken. I/O is suspended for a configurable number of seconds, but no outage occurs.

  - Attempt to flush writes held in memory out to the storage system and optionally attempt to suspend DB I/O and reach commitment control boundaries.

  - No guarantee that memory will be flushed completely, or that commitment control boundaries are reached.

  - Crash quality copy is made, could result in damaged objects, but will be a lower probability. "Cleaner" copy than a warm flash.

  - Will always result in an "abnormal" bring-up of the copy on the target LPAR.

  - **May result in FlashCopy target image of better quality than warm flash. Proper journaling for data recoverability will improve the quality.**

ASP quiesce function

- IBM i suspends transactions & operations to ensure that as much in-flight data as possible is written to disk
  - Optionally places transactions at database boundaries if possible
    - Intended for use with applications running commitment control
  - Copy will still be seen as 'abnormal' but much friendlier
- Applications without commitment control *might* be OK, but assessment and testing is recommended
- If quiesce is unable to flush memory, or reach commitment control boundaries, no information is available on what was not flushed. FlashCopy continues.
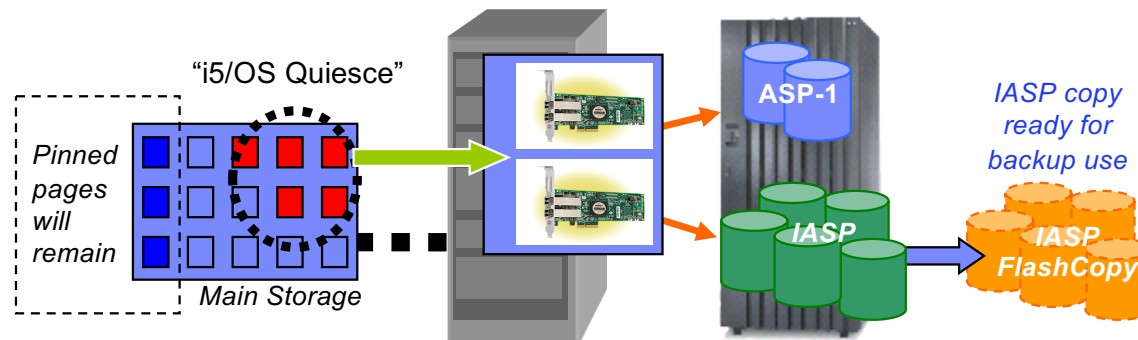
Quiesced Flash – CHGASPACT parameters

CHGASPACT command introduced in IBM i 6.1 – performs the quiesce of an ASP

OPTION (Option) –

- *SUSPEND – suspends DB transactions and DB and IFS file change operations. This does not suspend non-database activity.

- *RESUME – resumes suspended activity.

- *FRCWRT – forces changed pages in memory to disk. Does not suspend transactions. Should improve vary-on time, but not as effectively as *SUSPEND.

SSPTIMO (Suspend Timeout) – number of seconds to attempt to get to a good commit boundary for DB transactions

SSPTIMOACT (Suspend Timeout Action) – if all transactions could not be suspended within the timeout value, this parameter says to either –-

- *CONT : Continue the quiesce. Leave everything suspended which could be suspended. This should improve the vary-on time of the FlashCopy, although there will be some recovery expected

- *END : End the quiesce. All transactions will be resumed. This effectively get the same result as if a CHGASPACT was not attempted.

RECOMMENDATION: CHGASPACT ASPDEV(IASP_NAME) OPTION(*SUSPEND) SSPTIMO(300) SSPTIMOACT(*CONT)

Type of FlashCopy relationship

FlashCopy relationship attributes that are Internal to the storage system and affect how the storage system manages the copy. **These attributes are irrelevant to how IBM i sees and accesses the FlashCopy**.

- **Full copy**

  - In the background, the storage system copies all tracks from the source volumes to the FlashCopy target volumes.

  - FlashCopy relationship ends when full background copy completes, unless persistent

  - Useful for cloning systems or providing some additional protection from failure of the source volumes.

  - Can be used when heavy read I/O is expected against the FlashCopy target volumes (FlashCopy target used for Data Warehouse, i.e.)

  - Should only be used on fully provisioned FlashCopy target volumes – not space efficient / thin provisioned.

Type of FlashCopy relationship

FlashCopy relationship attributes that are internal to the storage system and affect how the storage system manages the copy. **These attributes are irrelevant to how IBM i sees and accesses the FlashCopy**.

- **No copy**

  - The storage system only copies tracks from the source volumes to the FlashCopy target volumes if a track is being written on the source volumes.

  - If a track has not been written by either the source or target system since the T0 FlashCopy instance was established, read I/O for the FlashCopy target volumes will be redirected by the storage system to the source volumes.

  - FlashCopy relationship exists indefinitely until it is ended, or until all tracks on the source volume have been modified

  - Useful for short-lived FlashCopy instances, such as nightly backup images

  - Can be used for either fully provisioned FlashCopy target volumes or space efficient / thin provisioned target volumes

Type of FlashCopy relationship

FlashCopy relationship attributes that are internal to the storage system and affect how the storage system manages the copy. **These attributes are irrelevant to how IBM i sees and accesses the FlashCopy**.

- **Persistent / Incremental**

  - Persistent FlashCopy relationships continue to exist indefinitely – even if all tracks have been copied from the source volume to the target volume

  - Record attribute is also used in conjunction with Persist to record which tracks on the source and target volumes have been changed since the FlashCopy was taken.

  - Together, persist and record attributes allow a FlashCopy relationship to be incrementally refreshed. An incremental flash will only process tracks that have changed since the last time the FlashCopy was taken or resynchronized.

  - Persist and record attributes are also necessary in order to reverse an active FlashCopy relationship and use the FlashCopy target volumes to overlay the FlashCopy source volumes – performing a "reset to T0" for the source volumes.

FlashCopy volume types

FlashCopy target volumes can be created and allocated on the storage system in different ways. The target volumes must appear to be the same size as the source volumes. **These attributes are irrelevant to how IBM i sees and accesses the FlashCopy**.

- **Fully provisioned volumes**

  - Traditional volumes – match the source volumes in actual size allocated. i.e., if source volume is 70GB, target volume is configured as, and actually allocates, 70GB of storage also.

- **Space Efficient / Thin Provisioned volumes**

  - Target volumes appear to the IBM i to be the same size as the source volumes, but no storage is actually allocated for the target volumes at the time they are created.

  - As data is written to the target volumes, storage is allocated from a shared pool of available storage (backing store) and assigned to the host.

  - The actual amount of storage allocated to a 70GB space efficient FlashCopy target LUN may only be 10 or 20GB, depending on write activity

  - Has additional performance overhead to allocate the storage

FlashCopy volume types

FlashCopy target volumes can be created and allocated on the storage system in different ways. The target volumes must appear to be the same size as the source volumes. **These attributes are irrelevant to how IBM i sees and accesses the FlashCopy**.

- **Compressed volumes (SVC/V7000)**

  - Target volumes appear to the IBM i to be the same size as the source volumes, but the actual storage allocated for the target volumes depends on how compressible the data is.

  - As data is written to the target volumes, the data is compressed and written to disk, reducing the amount of actual storage required.

  - The amount of compression achieved is dependent on the data being stored and is therefore variable

  - Have additional performance overhead to allocate the storage

FlashCopy – how do you manage it?

PowerHA Tools - Full System Copy Manager Lab Services Offering (FSCSM)

Supports using a single LPAR to connect to multiple FlashCopy target images – one at a time – either from the same or different source systems
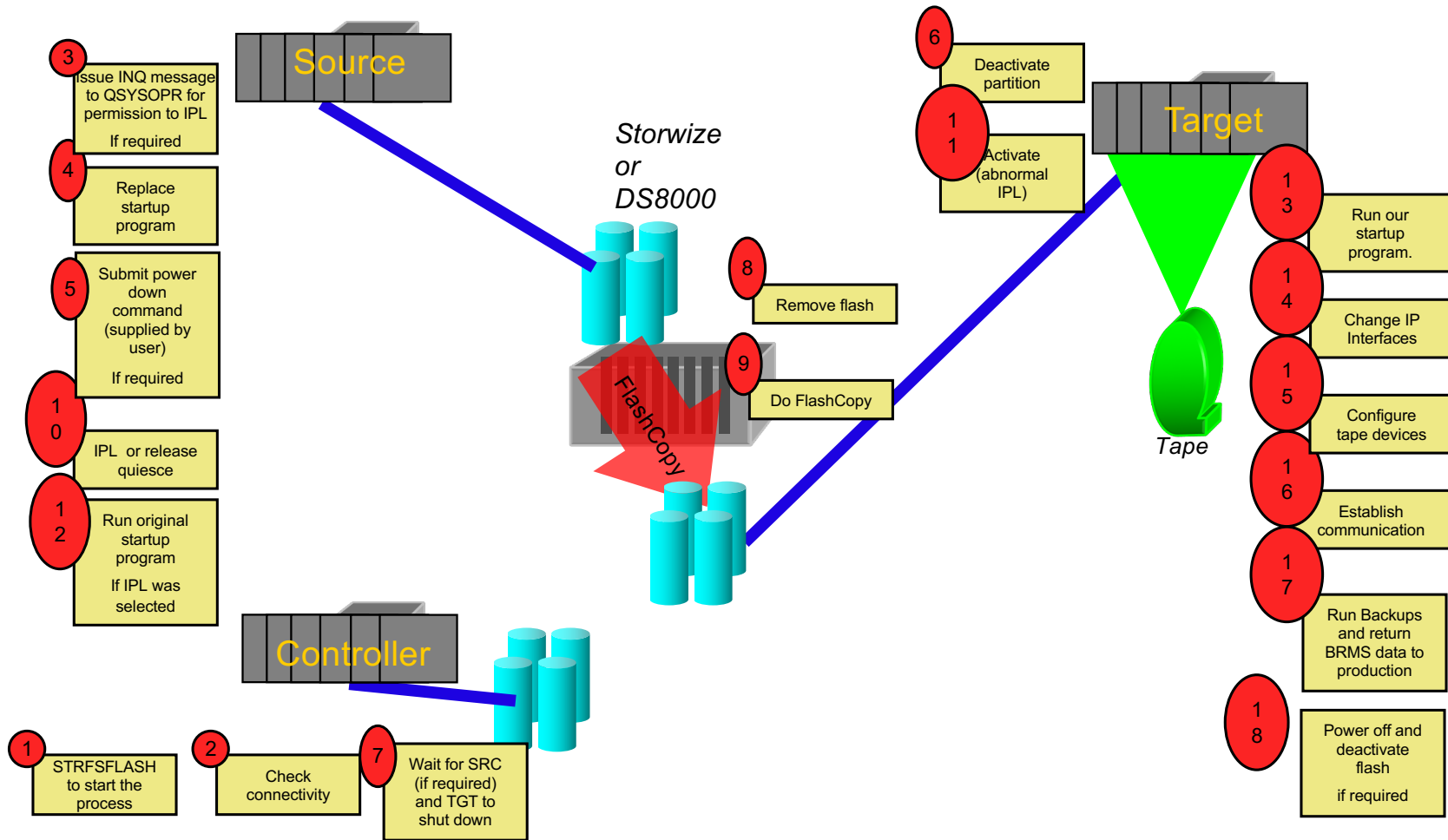
Provides simple interfaces for initiating FlashCopy instances and bringing the target copy on-line.

One command manages the whole process

Supports managing the DS8000 and Storwize family (V9000/V7000/V5030/V3700/SVC etc.)

BRMS includes integration for FlashCopy which has been fully integrated into FSCSM

Full FlashCopyFlashCopy Process

Source

Storwize or DS8000

3 — Issue INQ message to QSYSOPR for permission to IPL / If required

4 — Replace startup program

5 — Submit power down command (supplied by user) / If required

10 — IPL or release quiesce

12 — Run original startup program / If IPL was selected

8 — Remove flash

9 — Do FlashCopy

FlashCopy

6 — Deactivate partition

11 — Activate (abnormal IPL)

Target

13 — Run our startup program.

14 — Change IP Interfaces

15 — Configure tape devices

16 — Establish communication

17 — Run Backups and return BRMS data to production

18 — Power off and deactivate flash / if required

Tape

Controller

1 — STRFSFLASH to start the process

2 — Check connectivity

7 — Wait for SRC (if required) and TGT to shut down

BRMS requirements to backup full system FlashCopy

Required: Networking feature (57xxBR1 feature 5101)

- Required to share media inventory and backup history

- Recommended for all IBM i partitions sharing a media library

- Recommended for all HA pairs

  - Allows production partition to be aware of HA partition's backups

Recommended: Advanced Feature (57xxBR1 feature 5102)

- Includes the ability to send save history to a specific partition

  - Reduces amount of data sent around the network

Journaling for recoverability - considerations

Best Practice: Any files, data areas and data queues needed by users or applications should be journaled!!!

Use STRJRNLIB to automatically start journaling new objects that get created/moved/restored in production libraries. Use DSPLIBD to check the status of library journaling

Unless contradicted by the application, journal can be created with at least *MAXOPT2 for journal receiver size to allow it to be spread across many LUNs for better performance. Use *MAXOPT3 if possible.

Minimize entry specific data options can be used without impacting recoverability

Use *SYSTEM journal management setting on all journals – even if you manage them - to ensure journals are changed when needed.

Use of journal caching on a journal will not impact recoverability of journaled objects, but may impact recovery point of data.

Unless contradicted by the application, STRJRNxxx can omit open/close actions, and only need to journal after images.

Access path recovery for logical files

The access path recovery parameter on logical files dictates whether an access path will be recovered during a vary on, after a vary on or at first touch. The default of this parm is either *NO for non-unique keyed files, or *AFTIPL for unique keyed files.

*NO - the access path will not be rebuilt when necessary until it is first touched. This means that the user may experience a hang in the application if they are the first user to touch a logical file.

*YES – this means that access path will be rebuilt if necessary during the vary on – holding up the data access to all data on the IASP.

*AFTIPL - Best Practice: This is the best option for IASP environment. The access paths will be rebuilt aschronously in the background once the IASP is available.

The max number of system jobs dispatched to perform access path recovery for each ASP is equal to number of virtual processors on the LPAR (i.e. 10 virtual processors, SYSBAS + 1 IASP = 20 jobs performing access path rebuild after IPL

SMAPP

System Managed Access Path Protection will assist in avoiding access path recovery by managing the implicit journaling of logical files. The more logical files that are covered by SMAPP, the less recovery after a failover - but at a price of higher I/O on the production system.

SMAPP will keep track of the estimated recovery times of the files, and journaling the files in descending order from the largest down to the smallest - leaving only the smallest files that combined add up to the recovery time set in the SMAPP setting.

The lower that SMAPP is set – the faster the access path recovery will be for the IPL.

SMAPP uses the number of virtual processors as a factor in determining the ability to recover access paths within the desired time window. If the FlashCopy system does not have the same number of virtual processors, and same memory allocations, the SMAPP recovery time chosen on Production may not produce the same access path recovery time on the FlashCopy system.

Best Practice: Review SMAPP settings on Production to ensure rebuild times are acceptable. (DSPRCYAP).

So what else can you do with it?

Pretty much anything within reason!

The FlashCopy is a complete copy of the production node so you could…..

Run reports on it

Test upgrades on it

Run read operations on it during EOD processing

The list is only limited by your imagination

So what else can you do with it? – for example:

A customer is taking a pre-EOD FlashCopy on their production site, and a second FlashCopy on their DR site

The production site FlashCopy is used for a pre-EOD back up to tape and then waits for notification from production that the EOD is complete

The DR site FlashCopy is used for remote data warehouse systems to extract pre-EOD metrics that are required

After the EOD, the production site FlashCopy terminates and another is started at the same time a second FlashCopy is taken at the DR site

This production site FlashCopy is used for a post-EOD backup to tape and then terminates

The second DR site FlashCopy is used for post-EOD extracts for the data warehouse systems once they have finished with the first copy

Full System FlashCopy summary

DS8000, SVC, V7000, V3700 etc.

Prod LPAR

LPAR-2

(backup)

FlashCopy

Tape Backup

Full System FlashCopy
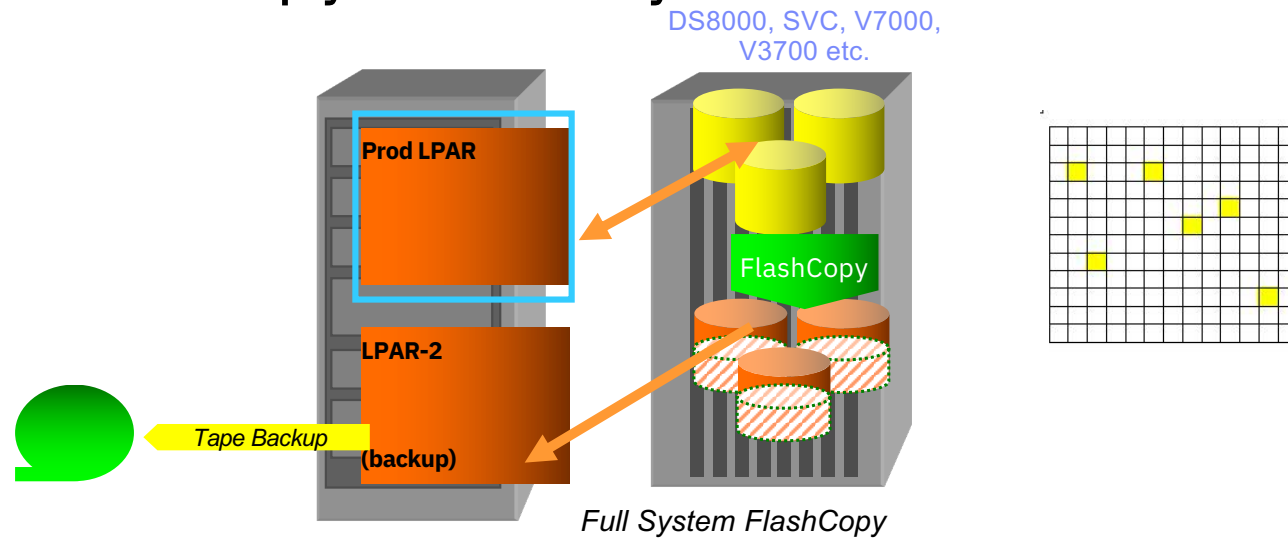
– FlashCopy is a point-in-time-copy of external storage logical volumes that can be established very quickly and with minimal or no disruption or resource on the production LPAR

– FlashCopy options – full copy, no copy, incremental

– Space Efficient / Thin Provisioned FlashCopy volumes can reduce FlashCopy storage requirements by 70-80%

– Use with the Quiesce to eliminate LPAR shut down, journal for object and data integrity of FlashCopy

– Managed with PowerHA Tools - Full System Copy Manager provide automation

– Fully integrated with BRMS

– Supported by IBM Support

– Ideal for off-line backup solutions

PowerHA Tools – Full System Copy Manager

PowerHA Tools - Full System Copy Manager (PowerHA FSCM) is provided by IBM Lab Services

Requires PowerHA and 7.2 or higher

NPIV or Direct-attached storage

– vSCSI is not supported

HMC is required to manage the LPARs

– FSM and ISV is not supported without an HMC
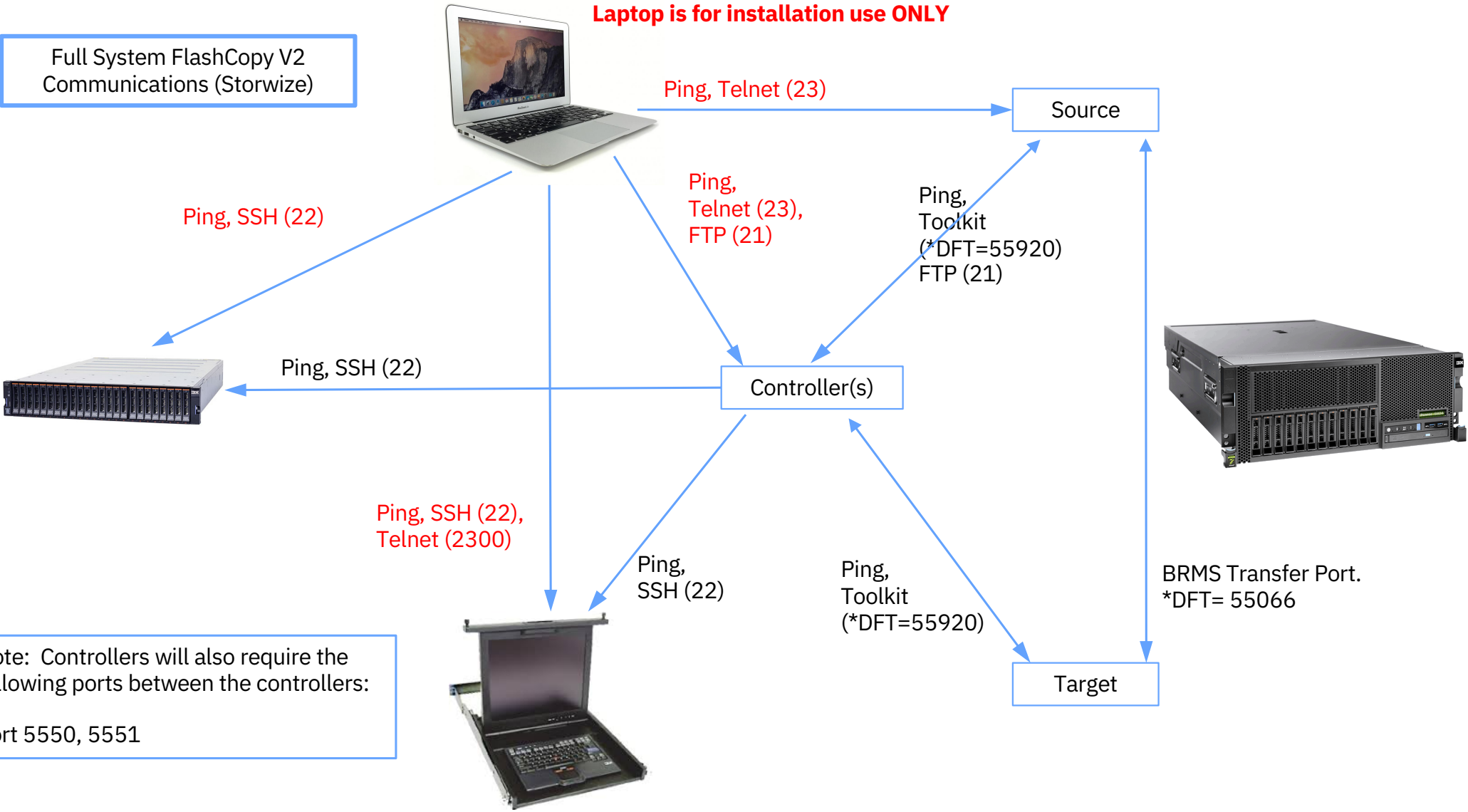
At least one controlling LPAR is required:

– IBM i 7.2+

– PowerHA Standard Edition

– 150 GB, 0.1 CPU and 6 GB RAM
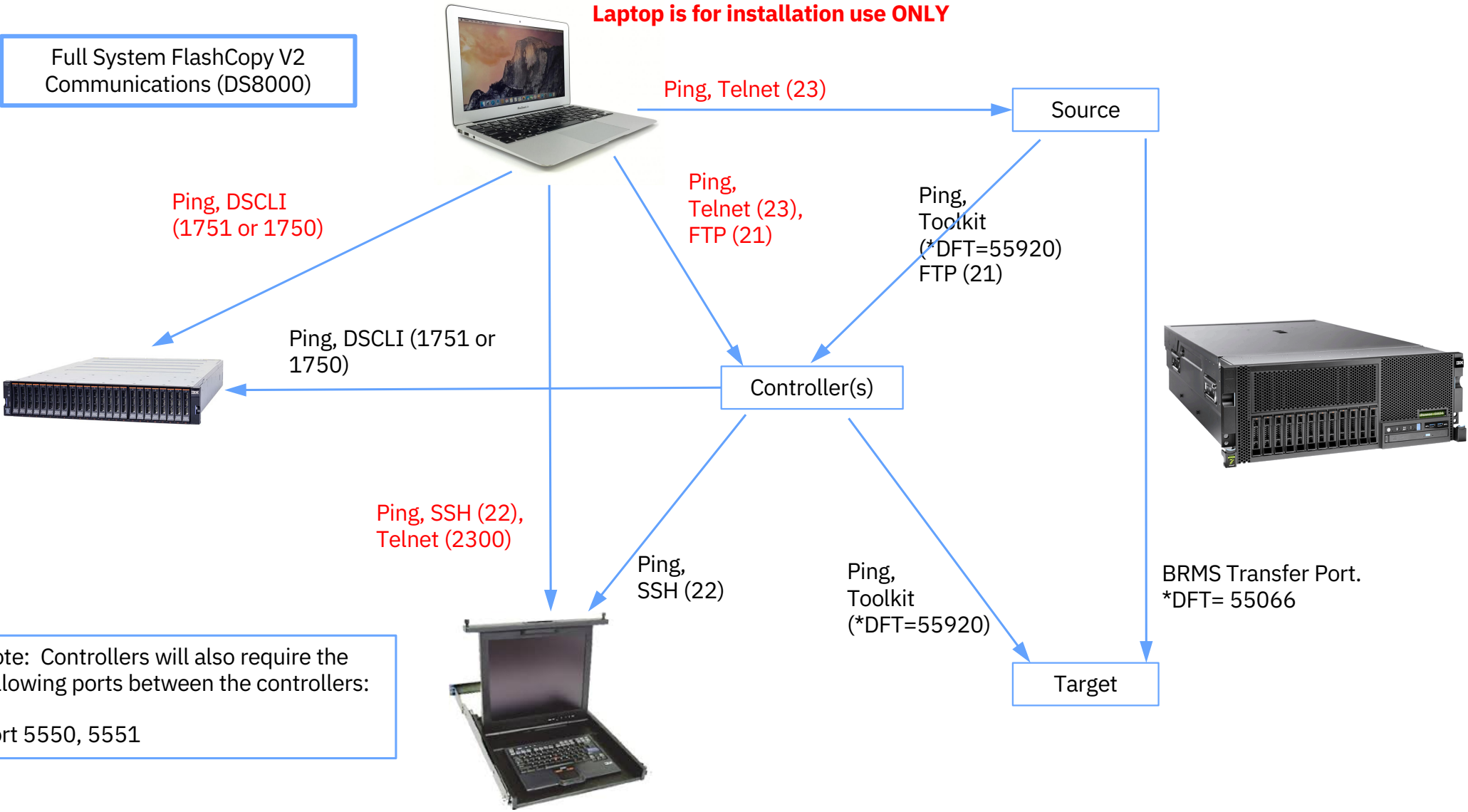
Full System FlashCopy V2
Communications (Storwize)

Laptop is for installation use ONLY

Ping, Telnet (23)

Source

Ping, SSH (22)

Ping,
Telnet (23),
FTP (21)

Ping,
Toolkit
(*DFT=55920)
FTP (21)

Ping, SSH (22)

Controller(s)

Ping, SSH (22),
Telnet (2300)

Ping,
SSH (22)

Ping,
Toolkit
(*DFT=55920)

BRMS Transfer Port.
*DFT= 55066

Note: Controllers will also require the
following ports between the controllers:

Port 5550, 5551

Target

Full System FlashCopy V2
Communications (DS8000)

Laptop is for installation use ONLY

Ping, Telnet (23)

Source

Ping, DSCLI
(1751 or 1750)

Ping,
Telnet (23),
FTP (21)

Ping,
Toolkit
(*DFT=55920)
FTP (21)

Ping, DSCLI (1751 or
1750)

Controller(s)

Ping, SSH (22),
Telnet (2300)

Ping,
SSH (22)

Ping,
Toolkit
(*DFT=55920)

BRMS Transfer Port.
*DFT= 55066

Note: Controllers will also require the
following ports between the controllers:

Port 5550, 5551

Target

Thank you

David Painter
Senior Managing Consultant

david.painter@uk.ibm.com