ibm.com

e-business

Second data center considerations

Frank Kyne - ITSO Poughkeepsie
(kyne@us.ibm.com)

Redbooks

International Technical Support Organization

Objective

The objective of this session is to help you identify the considerations for running systems spanning multiple data centers, and especially multi-site sysplexes. The aim is to help you ensure that the project will deliver the expected benefits and meet your SLAs.

We will NOT give you all the answers - every situation is different, meaning that the right answer is different for everyone.

But we *will* highlight the things you need to worry about, and tell you where you can get more information about each.

Topics

WHY are you planning a second data center?

Remote copy considerations

Supported distances

Connectivity considerations

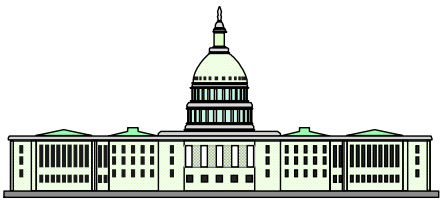
Performance considerations

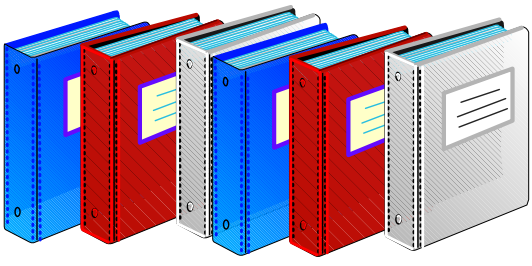
Systems Management considerations

Other sources of information

WHY do you want a second data center?

Disaster recovery





Government regulations for financial institutions

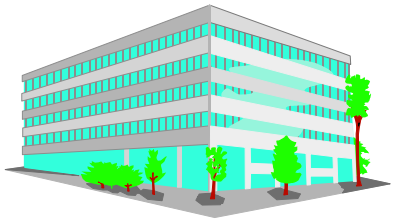
WHY do you want a second data center?

Continuous availability

A recent survey put data center outage costs at:

- $125,000/hr for pay per view TV service
- $2,600,000/hr for credit card validation
- $6,400,000/hr for online retail brokerage

WHY do you want a second data center?



You merged with another company and want to benefit from the data center investments

WHY do you want a second data center?

Software cost savings through aggregation of systems in both sites into a smaller number of sysplexes

Redbooks

ibm.com/redbooks

IBM

Remote copy considerations

Redbooks

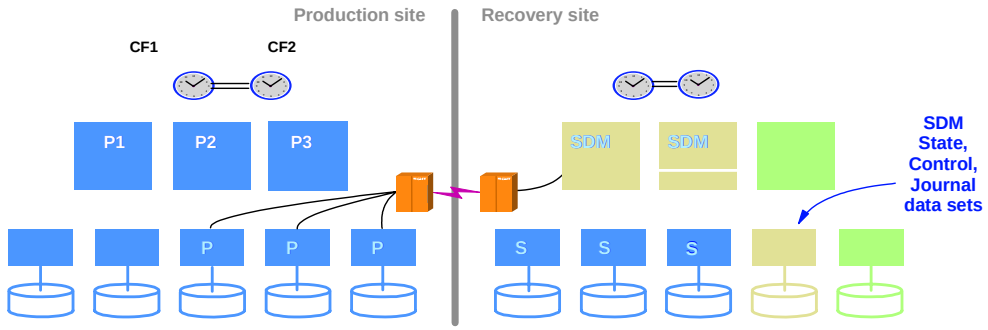
International Technical Support Organization

Remote copy considerations

The type of remote copy you can use (synchronous or asynchronous) depends on:

- Recovery point objective: how much data can you afford to recreate
- Recovery time objective: how long can you afford to be without your systems
- Likelihood of a regional disaster - is there a minumum distance that the second site must be from the first one
- Is your objective solely disaster recovery, or do you need continuous availability as well?

Asynchronous Remote Copy - XRC



Recovery Point Objective:

- Depends on many factors - distance, bandwidth, SDM capacity, CU type, etc
- Generally data in recovery site is less than one minute behind production site

Recovery Time Objective:

- Depends on level of automation, tested procedures, etc
- *Potentially* as low as less than 2 hours

Asynchronous Remote Copy - XRC

Pros:

- Negligible impact on reponse time
- Unlimited distance between primary and secondary sites
- XRC guarantees time consistency of remote DASD
- Supports very large configurations (1000s of volumes)

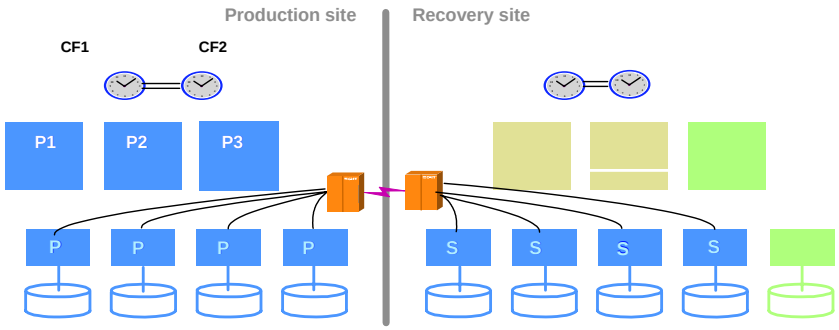
Cons:

- Requires z/OS capacity to drive data movement (System Data Mover (SDM))
- Guaranteed that you will have to recreate some data

More information:

- XRC Planning and Installation Guide, GC35-0481
- XRC Reference Information for Advanced Users, GC35-0482

PPRC Global Mirror - Asynchronous PPRC



Recovery Point Objective:

- Depends on a number of factors - distance, bandwidth
- Data in recovery site might be as little as 3-5 seconds behind production site

Recovery Time Objective:

- Depends on level of automation, tested procedures, etc
- Potentially as low as less than 2 hours

Asynchronous PPRC - PPRC Global Mirror

Pros:

- Negligible impact on reponse time
- Unlimited distance between primary and secondary sites
- Guaranteed time consistency of remote DASD
- Removes requirement (and cost!) of an SDM

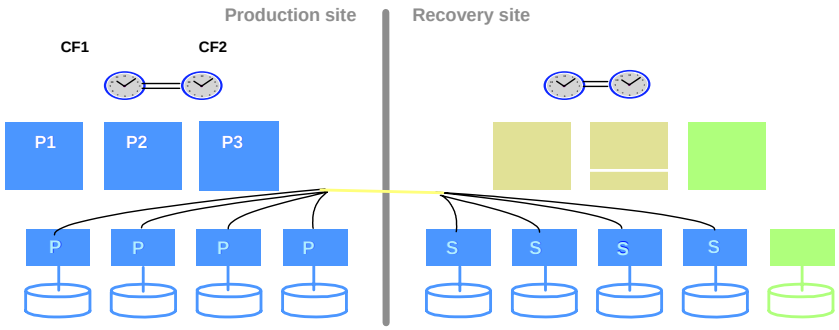
Cons:

- Currently supports a maximum of 8 ESSs in total (Primary and secondary)
- Guaranteed that you will have to recreate some data
- Requires 3 copies of data - Primary, Secondary, FlashCopy

More information:

- Implementing ESS Copy Services on S/390 or zSeries Hosts, SG24-5680-04

PPRC Metro Mirror - Synchronous PPRC



Recovery Point Objective:

- As low as zero data loss

Recovery Time Objective:

- Depends on level of automation, tested procedures, etc
- Potentially as low as less than 1 hour

Synchronous PPRC - PPRC Metro Mirror

Pros:

- Potential for zero data loss
- Potential for continuous availability (HyperSwap)
- Supports very large configurations
- No requirement for an SDM

Cons:

- Will have an impact on response times
- Limited distance compared to asynch - up to 303km with ESS, less if both sites are to be in the same sysplex
- Possibly needs more bandwidth than asynch options

More information:

- Implementing ESS Copy Services on S/390 or zSeries Hosts, SG24-5680

Remote copy considerations

Question - how will you guarantee consistency of secondary DASD across multiple CUs?

- With XRC, the System Data Mover (SDM) ensures that data is applied to secondary DASD in consistent manner
- With Asynch PPRC, the PPRC microcode manages the data to ensure the secondary DASD are updated in a consistent manner
- With Synch PPRC..... you need something on top of PPRC, such as Freeze and automation

Remote copy considerations

Question - how will you decide that you have a disaster?

- If you continue to update the primaries after the remote copy relationship is broken, and this turns out to be a real disaster, you will have to recreate data, regardless of whether you are using PPRC or XRC
- This requires automation AND *prior* management decision on how to react to failure situations

Remote copy considerations

And don't forget that every time you break the relationship between the primary and secondary devices, the secondary devices contain inconsistent data while you resynchronize.....

Answer is to take a copy of the secondaries BEFORE you start to resync, so at least you have a consistent set of secondary DASD, even if they are aged. And if you want to be able to resynch in both directions, you need FlashCopy devices at both ends

AND..... don't forget that every PPRC secondary and every FlashCopy target device takes up a subchannel.....

Redbooks
© 2004 IBM Corporation
ibm.com/redbooks

Supported distances

OK, so what is the supported distance for synchronous remote copy and for asynchronous remote copy?

Answer: It depends (remember, this IS an IBM presentation!)

Supported distances

The *maximum* supported distance depends on what you want to do:

- Do you just want asynchronous remote copy or offsite tape vault?
  - Is the moon far enough? See http://www.transorbital.net/

TRAILBLAZER®

THE FIRST COMMERCIAL VENTURE TO THE MOON: Transorbital® is the only private company to be authorized by the US State Department and NOAA for commercial flights to the Moon. The TrailBlazer® lunar orbiter will be the first delivery service to the Moon. Delivered to the Moon surface in a special capsule will be your certificates, business cards, cremated remains, jewelry, artwork and many other items of choice. The Trailblazer® satellite will deliver commercial and scientific projects and experiments to lunar orbit, as well as conduct lunar exploration and mapping.

Supported distances

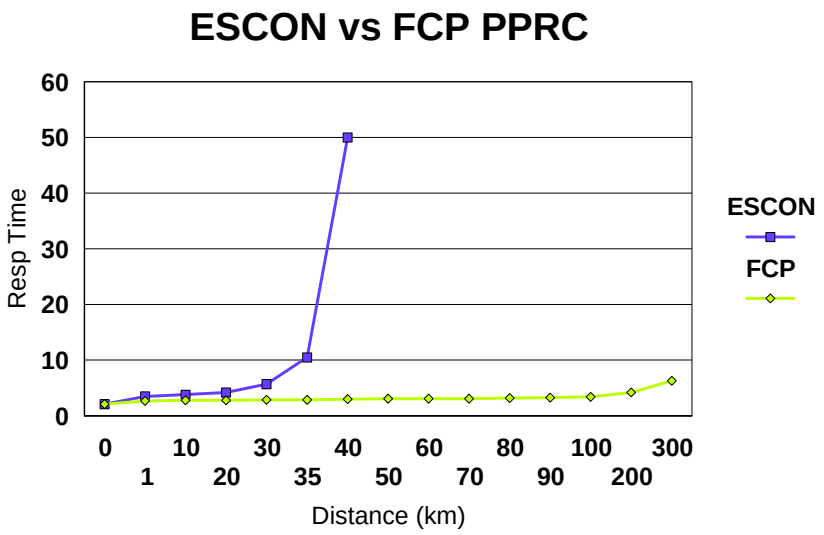
Do you want synchronous remote copy, with all production CPCs in one site, and only backup CPCs in the remote location?

- Maximum distance is about 103 kms using ESCON sync PPRC, 303km with FCP sync PPRC
- Must carefully consider impact of long distances on primary DASD response time, especially if using ESCON for PPRC connections
- Does not provide continuous availability capablity

Supported* vs *realistic

- Just because something is *supported* does not necessarily mean you can successfully implement it in your environment - you must check its validity for your configuration and workload

Comparison of ESCON and FCP PPRC Links



Sample configuration using 2105-800 doing 2100 I/O per sec and 2 ESCON PPRC or 2 FCP PPRC links, no PAV

Supported distances

Do you want a multi-site sysplex?

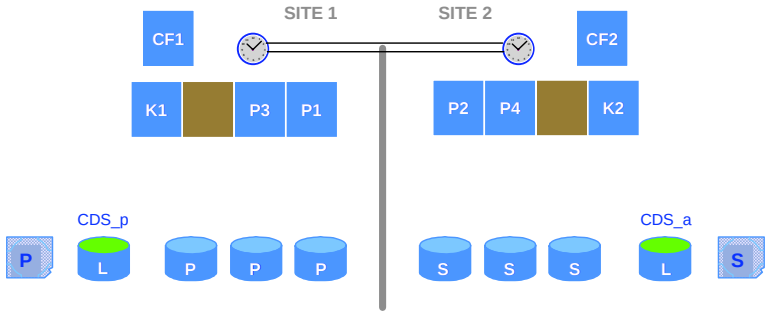
- Max distance is 100 kms (sort of)

At very large distances, need to consider impact of distance on response times, even for "shamplexes". Some things (like GRS and XCF) are sysplex-wide and all systems will be impacted by the response time impact of large distances

Supported distances

Do you want to do muti-site sysplex data sharing?

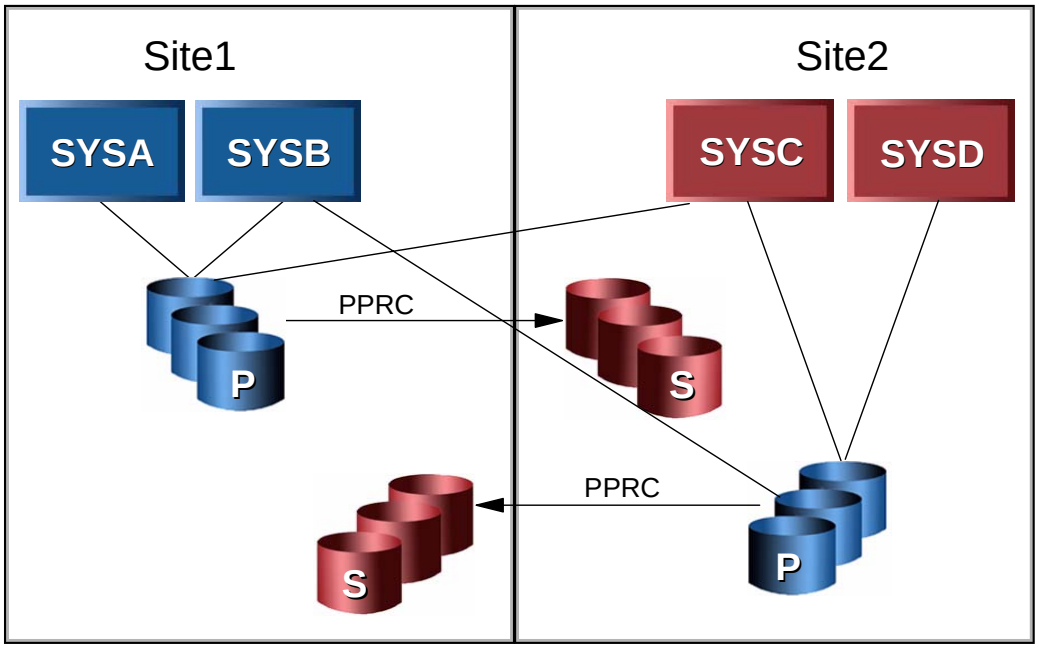
- Realistic limit is *about* 10 km.
- THIS IS ONLY A RULE OF THUMB
  - The actual realistic distance depends on your workloads and how they are impacted by increased response times
  - Remember batch jobs as well as online - large batch update jobs are more likely to be impacted by longer response times



Supported distances

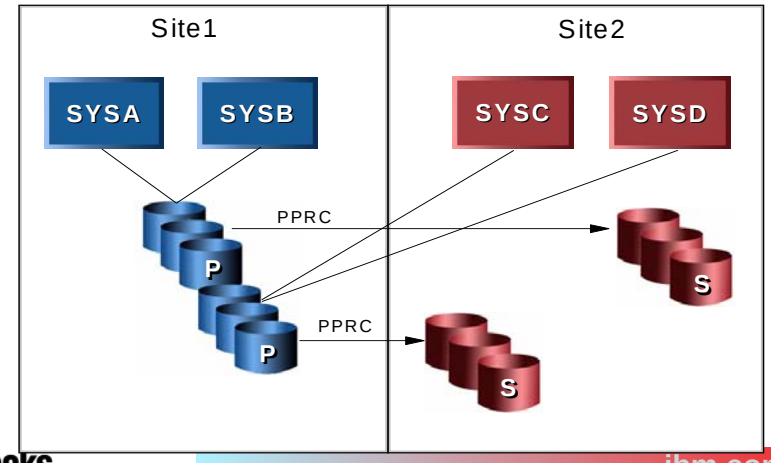
AND.... for any synchronous remote copy implementation, you need to consider where all your primary DASD will be.....

What's wrong with this picture?

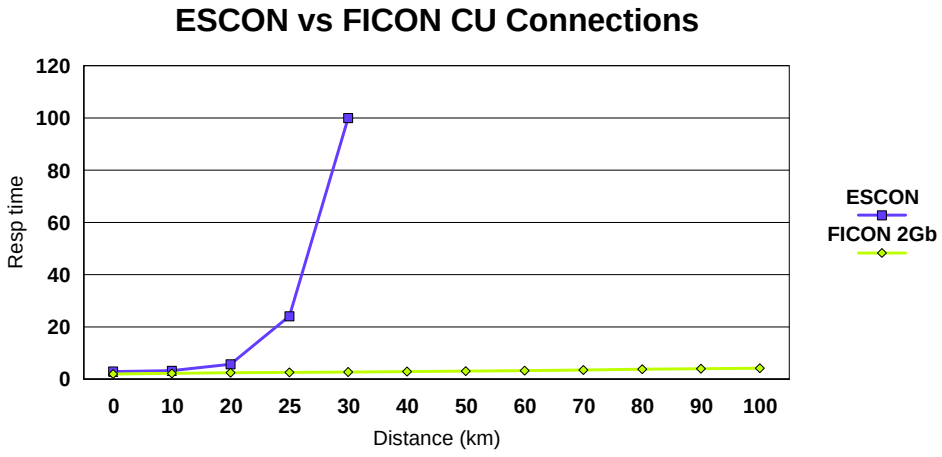


Supported distances

In a multi-site sysplex, and especially when doing multi-site DASD sharing, ALL the primary DASD must be in the same site. This means that some systems will have the cost of long connectivity *in addition* to the PPRC cost.



Comparison of ESCON and FICON CU connectivity



Sample configuration using 2105-800 doing 2100 I/O per sec and 8 ESCON or 4 FICON channels

Bandwidth

The bandwidth REQUIRED depends on:

- The type of channel - FICON provides much more bandwidth than ESCON, for example
- FCP PPRC links provide more bandwidth AND more efficient use of that bandwidth compared to ESCON
  - ESS provides more efficient use of PPRC links (that is, better performance) than pre-ESS CUs.
  - PPRC links between an SSID pair must be ALL ESCON or ALL FCP
  - ESCON PPRC links can only operate in one direction, but FCP PPRC links can operate in both direction concurrently
  - In nearly all cases, two FCP links should provide acceptable performance and availability
- Use RMF Magic and Disk Magic to calculate DASD Write Rate MB/sec and estimate impact of various channel types and numbers

Bandwidth

More bandwidth considerations.....

- Are you going to use P/DAS or HyperSwap or similar function? Do you plan on swapping between DASD in the two sites?
  - If so, you need to plan for sufficient CPU to CU bandwidth, not just PPRC bandwidth
  - If you are going to use HyperSwap, that assumes that the systems will continue to run in Site1, requiring full connectivity to secondary DASD from Site1 CPCs.
- If you have to use the secondary DASD, where will the systems run - in the normal site or that site?
  - If in their normal site, you need production level bandwidth between the two sites
  - If in the second site, you need sufficient spare capacity AND connectivty from those CPCs to the secondary DASD
- Will you run systems in Site 2 off DASD in Site1?
  - If so, you need to plan for sufficient CPU to CU bandwidth from Site 2 back to Site 1

Bandwidth

More bandwidth considerations.....

- I/O Rate (how many per second)
- I/O Size (how long is the channel busy for each request)
- Performance characterists of connected device
  - Tape can provide acceptable performance at higher channel utilizations than DASD
- If link is used for remote copy, the write intensiveness of the primary DASD is critical
- Need to consider single points of failure
  - Diverse routing
  - Two DWDMs
  - If you lose one DWDM/Path, can you continue operating on half capacity?
- Is the length of both paths similar and within related limits?

Bandwidth

The bandwidth AVAILABLE depends on:

- The connectivity you provide
  - Dark fiber:
    - ►Just ESCON and FICON directors?
    - ►DWDM
  - Telecoms lines
    - ►Channel extenders
- The cost!
  - Varies hugely from country to country.
  - Availability also varies - in some countries, dark fibre not available

Connectivity options

Connectivity considerations

Discuss HOW you can connect to remote devices (which technology can be used):

- Connectivity options
- CPU to DASD, Tape, Printers, 3x74, etc.
- CPU to Coupling Facility and Coupling Facility to Coupling Facility
- Sysplex Timer to Sysplex Timer, Sysplex Timer to CPU
- DASD to DASD and Tape to Tape (PtPVTS)
- XCF Signalling
- HMC LAN
- Consoles
- Network

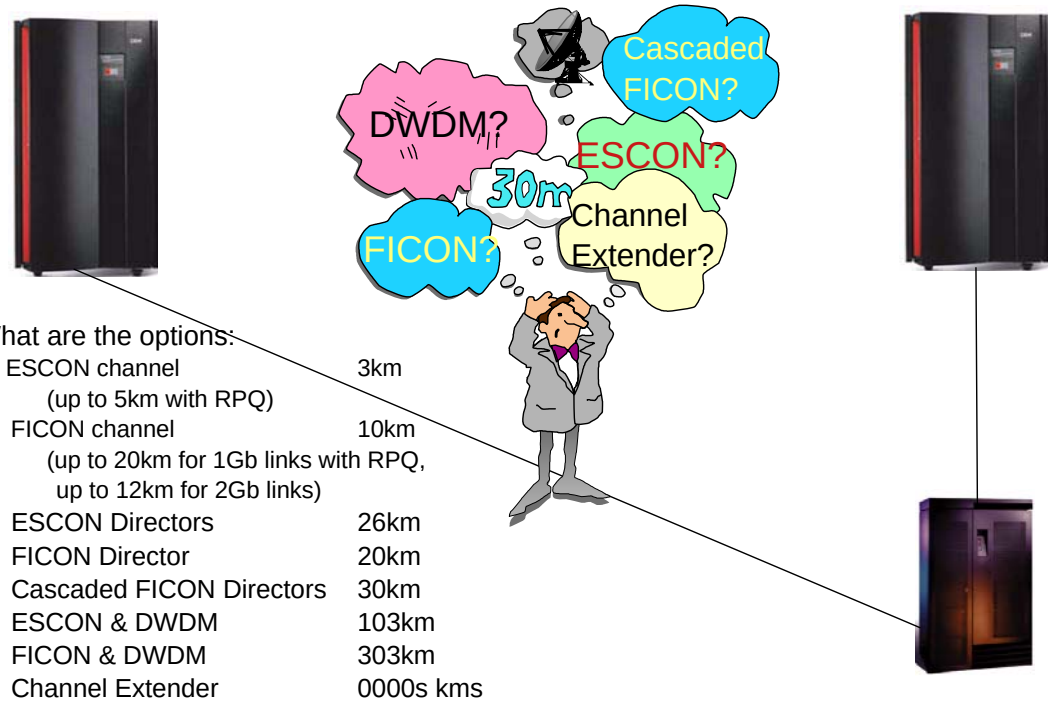
I presume Parallel channels are no longer being used.....
(but if they are, remember that converters can't be used with FICON Bridge (FCV) channels)

Connectivity considerations

Connectivity options

- There are a variety of ways to connect to a "device" over distances
  - Direct channel attach (point-to-point)
  - ESCON and FICON directors and switches
  - Use of repeaters such as IBM 9036
  - DWDMs
  - Channel extenders
- The distance supported for each option varies - by manufacturer, by device at the other end of the link, and especially by link quality
- You must contact the vendors to get the latest information and capabilities - this varies a lot by time and vendor. For example, one DWDM supports between 50 and 175km, depending on which features you select
- Not all devices perform acceptably over long distance
  - For example, ESCON CTC is terrible at 100 km, but ESCON VTS is fine at this distance

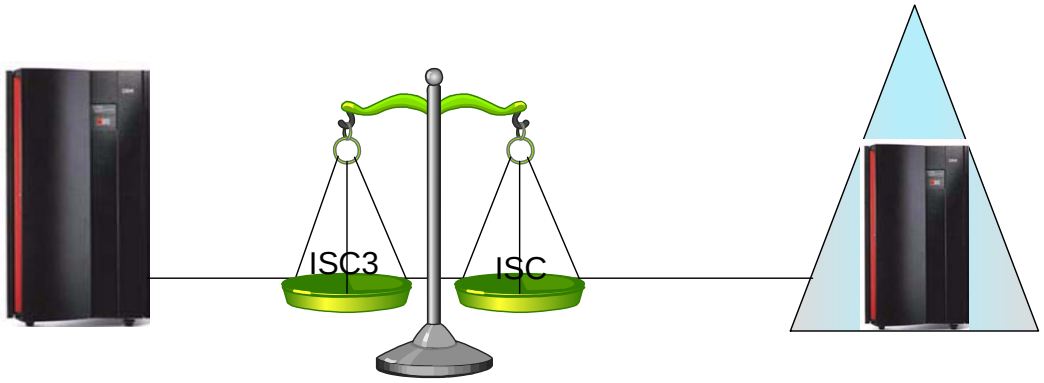
CPU to DASD, Tape, Cons controller, etc. connectivity



What are the options:

ESCON channel (up to 5km with RPQ)	3km
FICON channel (up to 20km for 1Gb links with RPQ, up to 12km for 2Gb links)	10km
ESCON Directors	26km
FICON Director	20km
Cascaded FICON Directors	30km
ESCON & DWDM	103km
FICON & DWDM	303km
Channel Extender	0000s kms

CPU to Coupling Facility

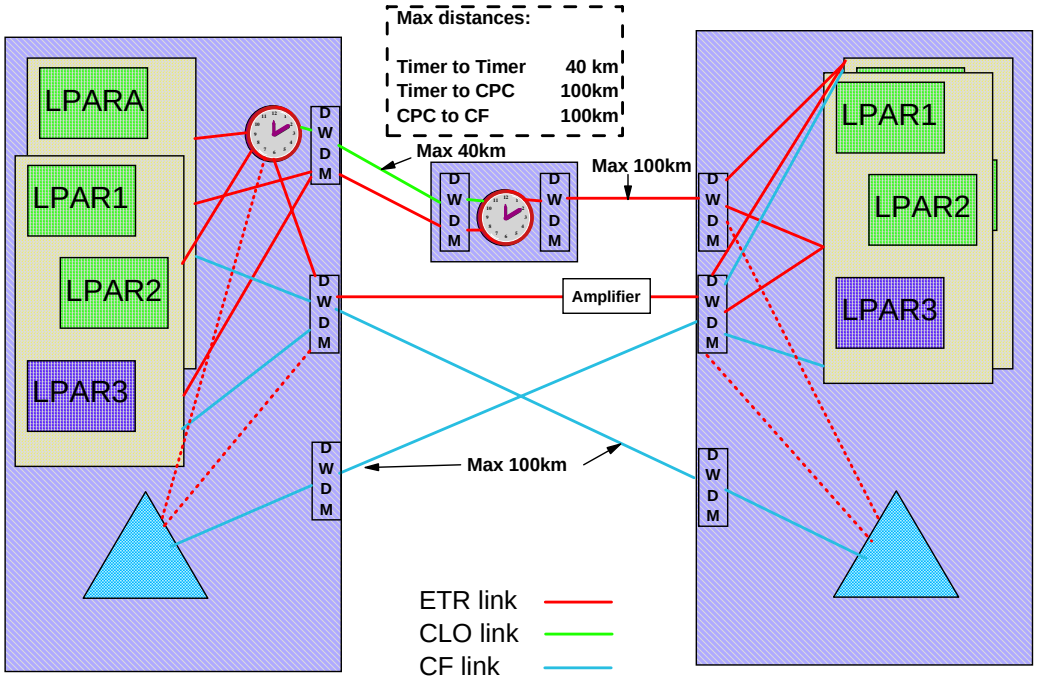


What are your options:

ISC3 peer mode (200MB)	10km (no repeater)
ISC3 peer mode RPQ 8P2197 (100MB)	20km (no repeater)
ISC compat mode (100MB) + DWDM + RPQ	40km
ISC3 peer mode (200MB) + DWDM + RPQ	100km

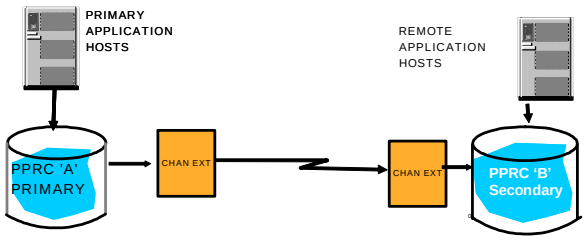
*See item RTA000174949 in viewblue for more information

Sysplex Timer connectivity



*9036 Model 3 supports extending ETR signals, however this device is no longer marketed

Remote copy links



Local Site

With PPRC, you can use:

- Switches, directors
- DWDMs
- Channel extenders (recent models make this more viable)

Remote Site

With XRC, you can use:

- Channel extenders (Most common)
- Switches, directors
- DWDMs

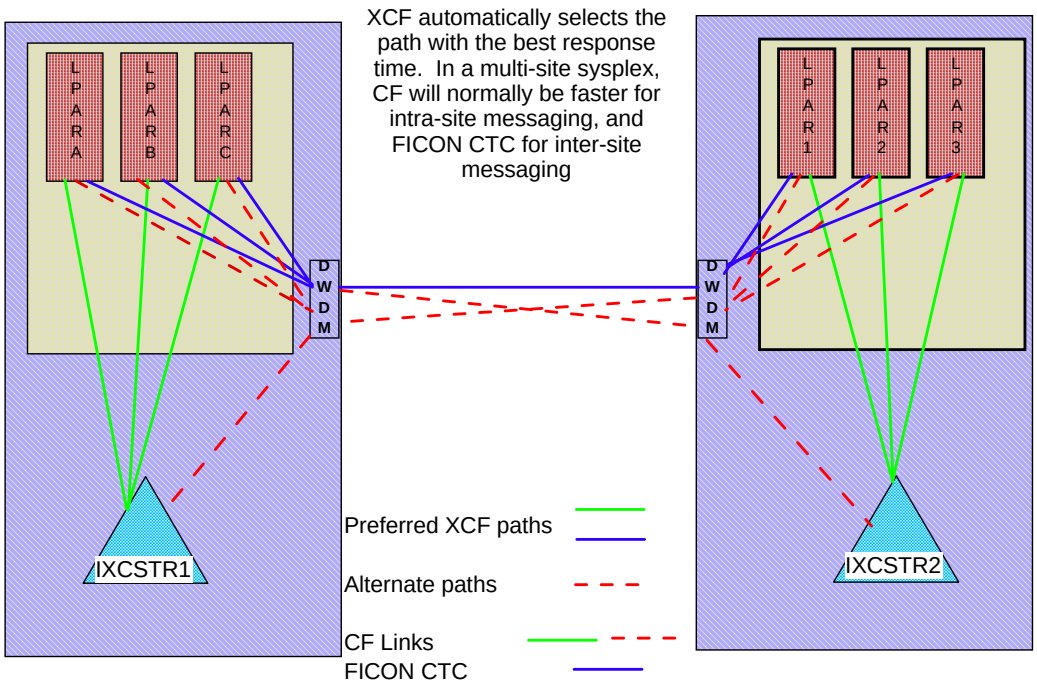
With VTS, you can use:

- Channel extenders
- DWDMs
- Switches, directors

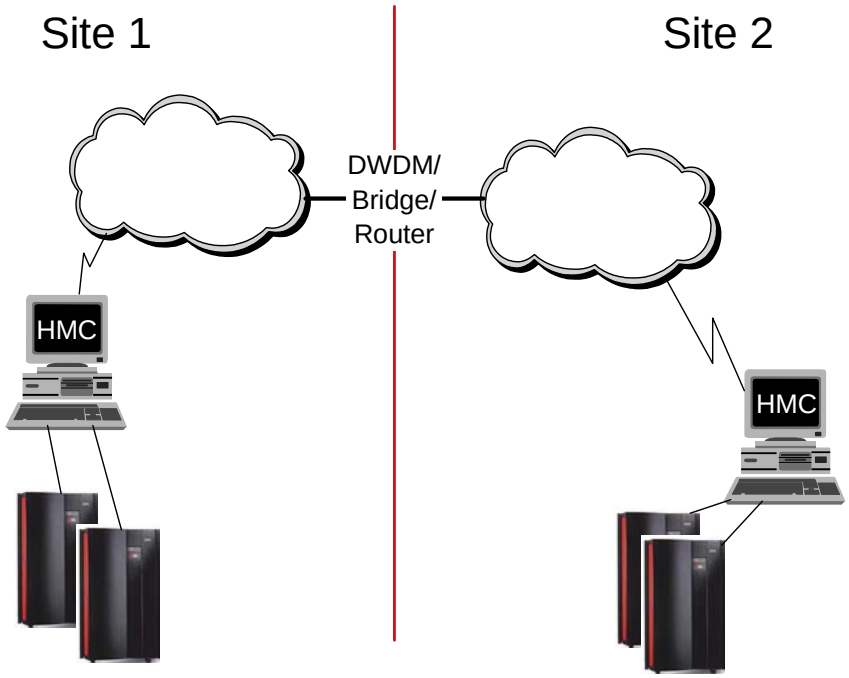
DWDM requires dedicated fiber (dark fiber)

Channel extenders use leased lines and SONET

XCF Signalling



HMC LAN interconnect

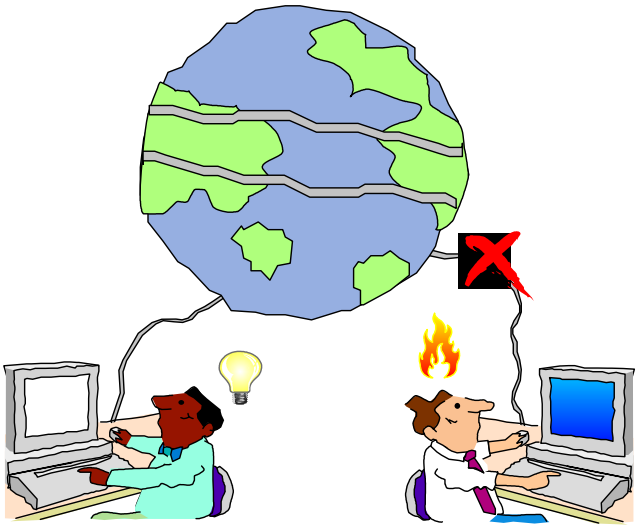


Console connectivity

Console connectivity:

- 2074
  - 2074s should be used to replace 3x74 and similar
  - 2074s are ESCON attached, so can attach via point to point ESCON, ESCON Director, or FICON Bridge
- OSA-ICC
  - Other alternative is OSA-Integrated Console Controller (OSA-ICC), available on z990, z890
  - Simply connect OSA-ICC port to LAN, and as long as you can access that LAN, you can bring up a console on any PC
- SNA Consoles
  - Real MVS console can be brought up on any PC that can access VTAM services
  - Cannot be used for NIP processing - only accessible after VTAM starts

Network connectivity



There is little point in having all the systems recovered and running in Site 2, if all the users are still connected to Site 1.....

In addition to connectivity, you also need automation and/or intelligent routing to switch your users over....

IBM

Performance considerations

Redbooks

International Technical Support Organization

Performance considerations

Performance is dependant on:

- Distance.
  - Degradation is not linear
- Technology.
  - ESS uses ESCON for PPRC much more efficiently than RVA
  - FICON is hugely better over distance than ESCON
- I/O Rate.
  - High channel utilization is much more painful at long distance, AND, long distance (via long response times) drives up channel and UCB utilization
- Bandwidth.
  - You may not be able to afford to provide as much bandwidth to a remote site as you would within the computer room
- Remote copy.
  - Synchronous remote copy may require more bandwidth than asynchronous

Performance considerations

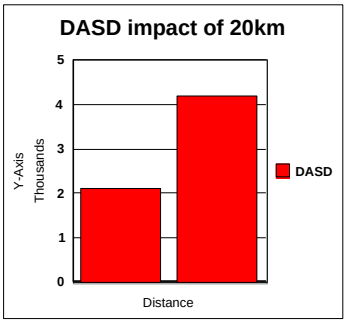
Just because something is technically possible, does NOT mean that it can be made to work in your environment at an affordable cost:

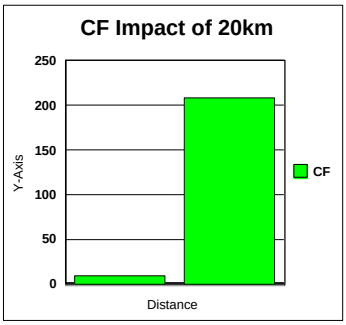
- Long distance between CPU and connected control units impacts response times
  - In the example earlier, adding 20 km doubled ESCON response time
  - You MUST use a tool like Disk Magic to project actual anticipated response times
- Long distance between primary and secondary CU impacts response times (for synch PPRC) or amount of data that needs to be recreated (for async PPRC or XRC)
  - In the example earlier, adding 20 km doubled response time when using ESCON PPRC links
  - Once again, use a tool like Disk Magic to project response times
- Newer technology (FICON and FCP), new features (PAV), and more bandwidth/adaptors can reduce impact of distance to some extent

Performance considerations

Also have to consider Coupling Facility impact:

- Consider that "good" CF response times are about 100 times faster than good DASD response times
- Distance adds a fixed amount - 10 microseconds per km - to response times. Therefore, impact of distance on CF is relatively much higher than on DASD





Performance considerations

Consider:

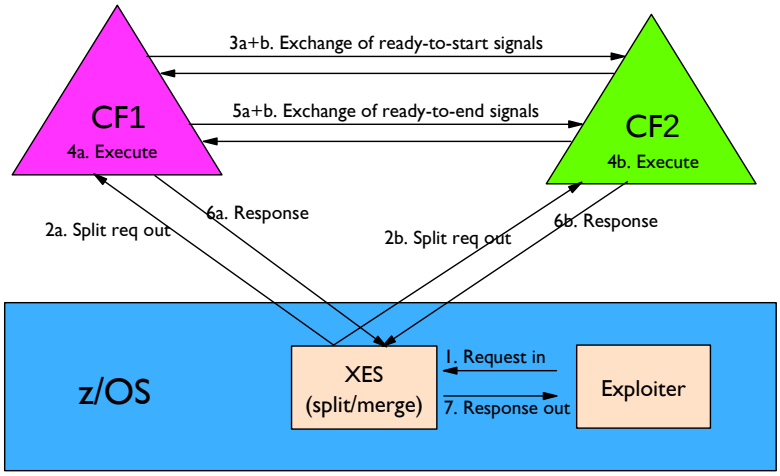
- Who is using the CF
- Synch/asynch algorithm in z/OS 1.2 will cause long sync requests to become asynch, significantly increasing response time for those requests. This change limits the CPU cost of high response times, but this does not help applications using the affected structures
- Peer mode links should be considered a must for longer distances
- It *is* possible to have many requests running against a given structure at the same time - access is not serial
  - One customer is doing 60,000 lock requests per second at 5km distance and about 150 mics. response time
- What is the change in lock response time compared to lock hold times?

Performance considerations

CF Links:

- Subchannels are busy for the whole time of the CF request. For asynch requests, it is still for the whole of the response time
- As distance increases, response time goes up, driving up subchannel utilization, and making reponse times even worse
- To offset this, add CF links and carefully monitor subchannel utilization
  - Use Parallel Sysplex Quicksizer (SPSSZR) to project impact of distance on CF Link utilizations

Performance considerations



System Managed CF Duplexing especially does not like long distances

- Every km travelled adds 10 microseconds to reponse time
  - Distance from CPC to furthest CF * 10 microseconds PLUS
  - Distance between CFs * 4 * 10 microseconds
  - If one CF is 20 km away, impact would be approx 1000 microseconds....

IBM

Automation, processes, and testing

Systems Management considerations

Planning for the hardware and software is only the start

- MUST wrap this in automation, to get the speedy response and consistency that is required
- Ideally automation will handle everything from identifying the disaster trigger through to getting all your systems up and running in the recovery site
  - Automation should also manage and provide operations interface to Remote Copy mechanism and status
- You must test until you get all this working. Then keep on testing to ensure it continues working.
- Give consideration to placement of operations - some installations place ops in a 3rd site.
- If building a new site, don't forget to allow for fitting-out time after the building is complete, but before you can use it.

Systems Management considerations

What else.....

- Expect the unexpected - problems often arise from "trivial" things, like cabling, power supplies, physical access, and so on.
- Calculate your worst case elapsed time to set everything up - then double it! Many of the things you encounter will be outside your control and may take time to rectify.
- The devil is in the details.... for example, while direct attachment from DWDM to CU is possible, some CUs require attachment to a switch because the CU does not provide sufficient credits for the distance to the CPU

More information

IBM Redbooks:

- IBM eServer zSeries Connectivity Handbook, SG24-5444 (new version due this October)
- Implementing ESS Copy Services on S/390, SG24-5680-04
- Planning for IBM Remote Copy, SG24-2595
- IBM Enterprise Storage Server, SG24-5465
- A Disaster Recovery Solution Selection Methodology, REDP3847

IBM Product Manuals:

- Advanced Copy Services
- XRC Planning and Installation Guide
- XRC Reference Information for Advanced Users

More information

IBM Service offerings

- Bandwidth studies - contact local storage specialist
- IBM Disk Magic tool projects impact of distance, remote copy, and connectivity type
- eRCMF - Does not support zSeries
- RCMF - Remote Copy Management Facility
- GDPS - Search IBM Web site for GDPS for White Papers, plus more detailed GDPS Introduction Redbook coming out later this year

More information

GDPS-qualified DWDMs and Channel Extenders

- Adva
- Cisco
- Nortel
- InRange
- CNT
- Ciena
- See individual vendor's Web site for information on these devices.
  - Later this year, there will be a series of RedPapers on qualified DWDMs, with links to all the vendor's Web sites.
- See http://www.storage.ibm.com/disk/ess/supserver.htm for more information on devices that are tested with ESS

Summary

This is a very complex subject, requiring skills in many areas

Technology is changing daily

Before anything else, need to clearly identify the objective:

- Is it Continuous Availability?
- Is it Disaster Recovery?
- Is it SW cost savings?

Involve the experts from the very beginning

Results can be stunning, IF the project is properly planned and managed