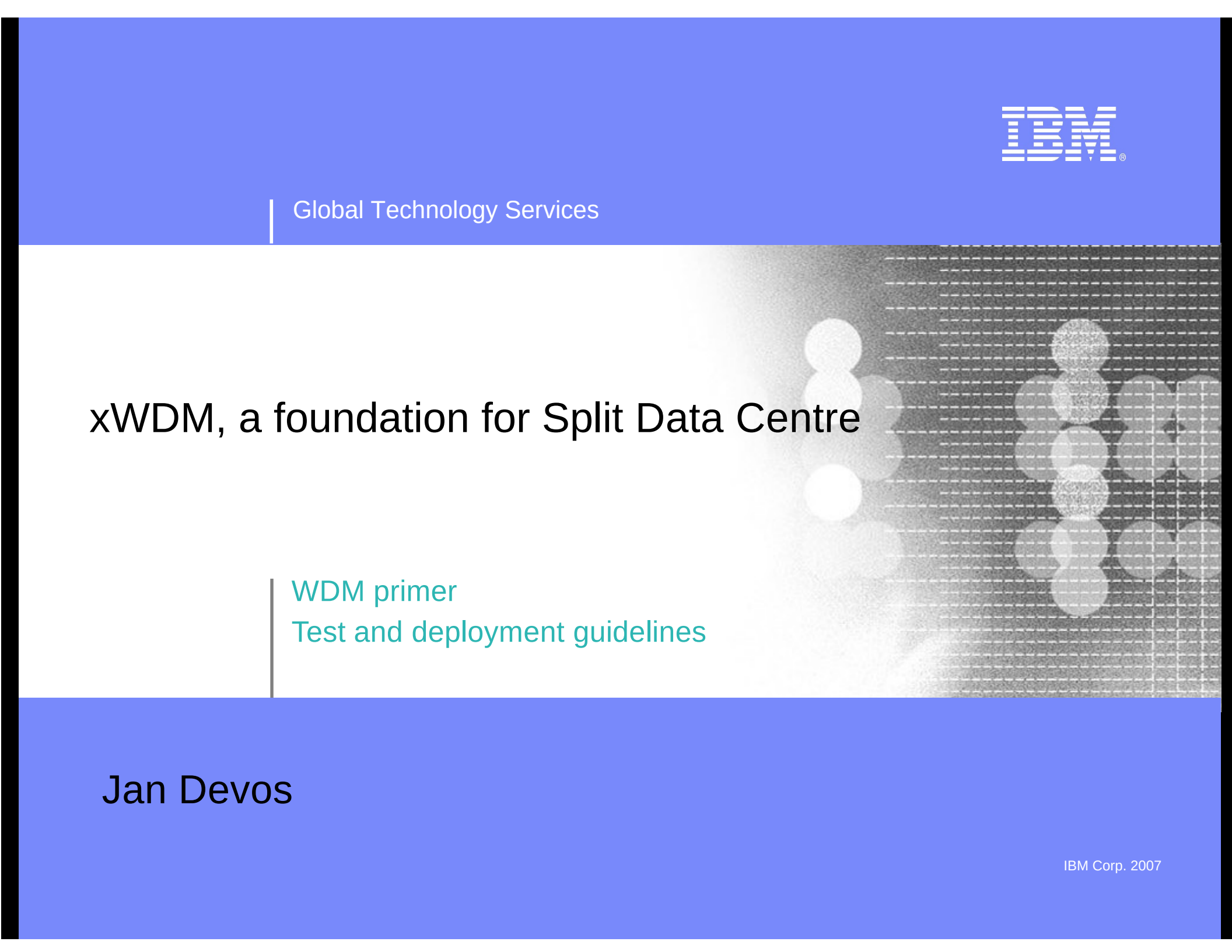

IBM
Global Technology Services
xWDM, a foundation for Split Data Centre
WDM primer
Test and deployment guidelines
Jan Devos
IBM Corp. 2007

About the speaker …

- IBM-Certified Networking and Connectivity Specialist within IBM Global Technology Services
- Specialty in ZSeries connectivity, Routing and Switching and Optical Networking
- Role and responsibilities : architecture, (re-)design and (re-)build enterprise networks
  - Greenfields
  - Rejuvenation of legacy networks (SOA driven, capacity upgrades, availability improvement etc.)
    - Network and Server consolidation
  - ZSeries network access and integration
  - Connectivity within the Data Centre(s)
  - Any other networking and connectivity related mission

Agenda

- WDM Primer
- Split Data Centre : drivers and issues
- WDM deployment and test approaches
- Conclusion

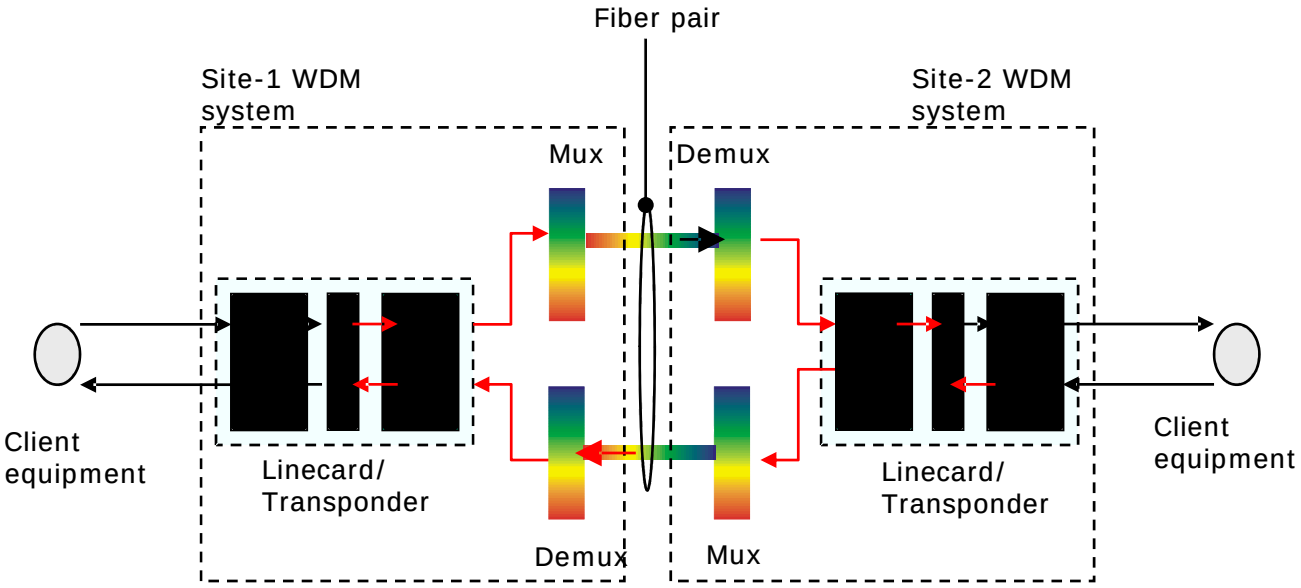
WDM Primer

- xWDM principles
- A generic xWDM device
- Transponders and Muxponders
- Optical protection

xWDM is a technique to transport several connections over one single fiber pair.

- Wave Division Multiplexing : multiplex several *transport channels over one single fiber pair.*
  - Each channel corresponds to a wavelength
- Channels are bitrate-independent : 100 Mbps, 1,2,4,10 Gbps …, and used to transport client connections or ‘services’.
- Wavelength spectrum is standardized by ITU-T, the G.694-grid.
- C(oarse) WDM : 8 wavelengths with 20nm (2500 GHz) spacing in 1470 - 1610 nm range
- D(ense) WDM :
  - initially 32 wavelengths with 200 GHz spacing
    - 16 in 1530 – 1560 nm range (C-Band)
    - 16 in 1575 – 1605 nm range (L-band)
  - Ever increasing demand for more connections on one fiber
    - Use more dense spacing (100 GHz, 50 GHz) => Results in up to 144 wavelengths/channels per fiber pair.
    - Transport several ‘lower speed’ client connections on one wavelength/channel => e.g 4 x 2 Gbps on one 10G channel
  - An Optical Supervisor Channel (OSC) is reserved for Inter-Node Out-of-Band management traffic
- DWDM allows for more compact optical backbones than CWDM, at an extra cost
  - Dense wavelength spacing requires more accurate and hence more expensive lasers and filters.
  - Trade-off between cost for extra fibers plus O&M (CWDM), against more expensive equipment (DWDM)

A generic WDM system contains Transponders and Filters

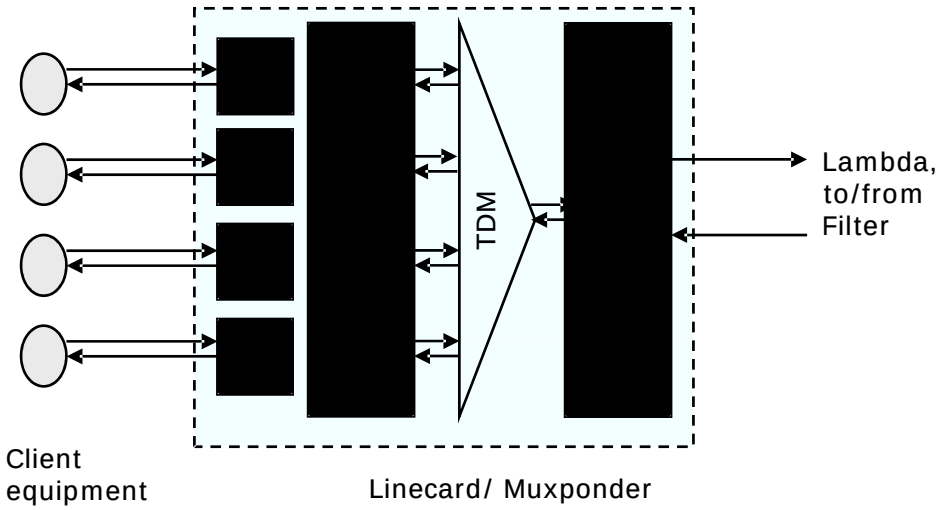


Client equipment = LAN/SAN sw, CF, Syspl. Tmr …

OEO = Optical-Electrical-Optical : Convert client signal on SM (1310nm) or MM (850nm) to an electrical signal, and again to an optical signal in the ITU-T range (1550-1600 nm). Same vice-versa.

- Transponder or Linecard
  - where client equipment connects to WDM network
  - Provides optical-electrical-optical (OEO) conversion : client signal from a 'neutral' wavelength (850 nm MM or 1310 nm SM) put on a wavelength of the ITU-T grid for upfront multiplexing
  - Other term is Linecard
  - First gen. transponders are wavelength specific and bitrate specific (!!)
  - Modern transponders are configurable for client bitrate and/or wavelength
- Filter
  - Mux : combines several wavelengths forthcoming from # of linecards into one egress composite signal
  - DeMux : splits ingress composite signal into several individual wavelenghts for hand-off to the linecards
  - Mux/Demux on same or different module
  - Support a series of adjacent wavelengths of the ITU-T grid
  - Filters are passive, imagine a filter as a PRISMA
- Transponders connect into the filters via provisionable patch cords (SM).

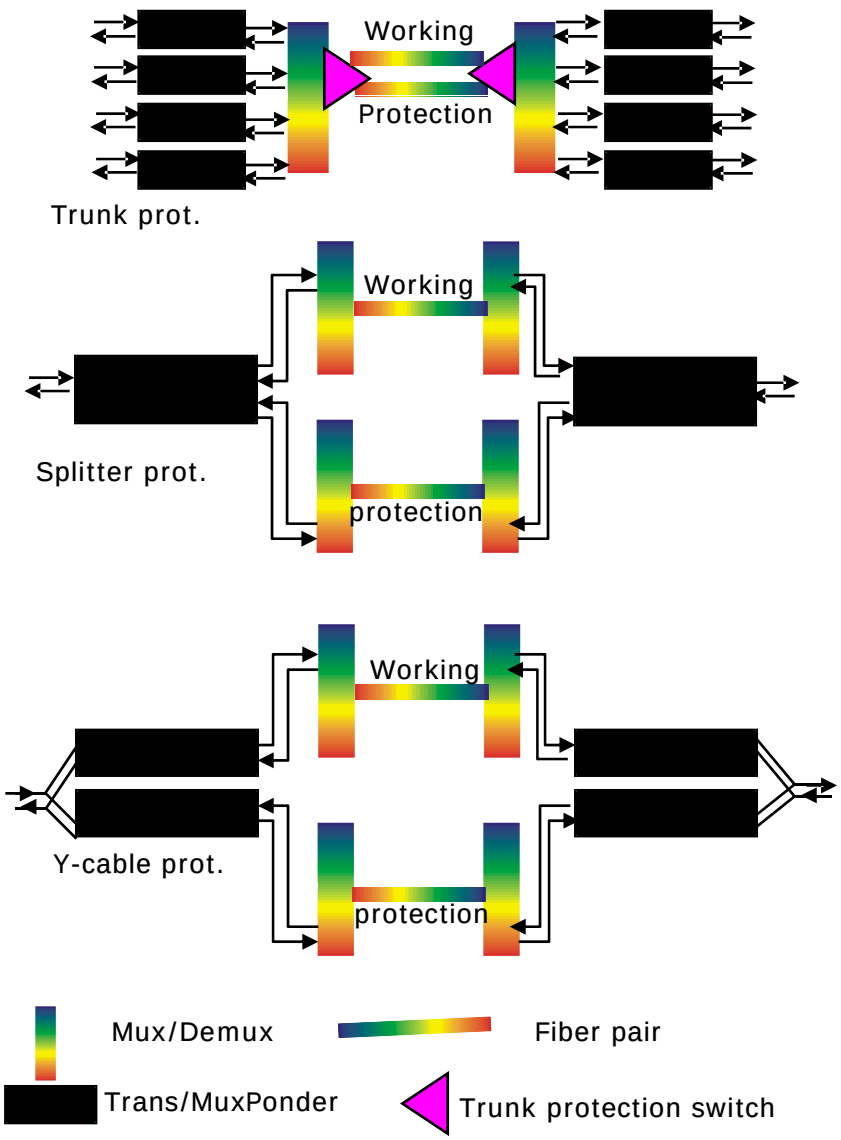
The ever increasing demand for service density is also addressed by multiplexing several client connections over one transport channel. The Transponder becomes a Muxponder.



TDM = Time division multiplex

- Linecard contains logic based on legacy TDM technique
  - to combine several lower-speed client connections onto one single channel/wavelength.
  - Wrap client signal in SDH envelope
  - Multiplex it into next-higher SDH hierarchy
  - Put this signal on a wavelength
  - On the far-end, just the other way around.
- Example : 4 FibreChannel signals of 2 Gbps each are transported over one 10G channel
- Result is a much higher service density
  - Transport channels more efficiently used

The optical layer provides protection upon path or equipment failure, by restoring transport channels within 50 millisec, when an optical path or equipment fails.



- The protection path ensures survival of a transport channel, should the working path fails
  - Channel is forwarded over both paths, when both are *operational*
  - Receiving end decides which signal to take-in from the network side and hand-off to the client side
- Trunk protection :
  - Protects against fiber cut between two sites
  - Composite signal egress from MUX is split over both paths, receiving end decides which signal to hand-off to the DeMUX and linecard beyond.
  - All wavelengths are switched at once in the event of a fiber cut
  - Cost-effective : a trunk-protection switch is put in front of the filters (at the 'network' side)
- Splitter protection
  - Protects against fiber-cut or failure of Mux/Demux
  - Trans/MuxPonder splits its own channel over two paths on egress, and decides which one to take-in on ingress
  - Requires the 'protected' flavor of Trans/Muxponder (extra cost)
  - Channels are selectively (un)protected
- Y-cable protection
  - Protects against fiber cut or any failure within the WDM system
  - Client signal split over two Trans/Muxponders, forwarding to distinct paths
  - Receiving Trans/Muxponders decide which of them accepts the signal
  - Trans/MuxPonders have to be doubled => expensive solution
- A typical WDM network for a split data centre might not require optical protection
  - For each link, a redundant link is provided, just BAU
  - But Optical Protection provides an automatic restore of the failing links,
- If only one fiber of a pair fails, protection switching is UNIdirectional
  - This is an issue for applications that do not tolerate asymmetric latency (Sysplex Timer)

Split Data Centre : drivers and issues

Increasingly more organizations split their I/T assets over two geographically separate data centers. Split Data Centres mostly address a combination of requirements and face the issue of interconnection.

- Reasons to consider split data centre
  - Increase availability by protecting against local disaster
  - Space constraints : single DC became too small
  - Regulations : SOX, Bazel-2
  - Mergers and acquisitions
- How to interconnect?
  - Maximal transparency
  - Minimal latency
  - Segregation of different sub-networks (LAN, SAN, Mainframe/Sysplex, Server Clustering …)
  - Or in short : which network can provide ‘at an affordable cost’ the same quality, availability and security to a collection of devices that ran ‘close’ to each other, but are from now distanced over a Metropolitan Area?

Currently, the three options for interconnecting Split Data Centres are a bunch of dark fiber, xWDM services or IP/VPNs

The choice between dark fiber, xWDM services or an IP network is greatly imposed by several factors.

- Will my I/T survive when it runs on two DC's distant to each other, rather than on one?
  - What are the implications of adding one type of connectivity between devices (servers, storage, …) that previously ran close to each other?

Type of connectivity	Cost effectiveness	Provisioning speed	Capacity	QoS (incl. availability)	latency
# IP/VPNs	5	4	2	3	2
# xWDM svcs	3	2	5	4	5
# dark fibres	1	1	5	5	5

- Although any case should be studied on its own, some overall guidelines are derived from the above evaluation
  - Dark fiber connectivity is only affordable if you have your own fibers, or can obtain them far below the 'street price'
  - IP/VPNs, though very suitable in the WAN, are likely to hit issues of Capacity, Quality and Latency when used for interconnecting data centres. It will compromise synchronous mirroring and is excluded from use in a Mainframe-DC environment
  - xWDM maintains the strenghts of Dark Fibre, at a more affordable cost and with increased flexibility (provisioning speed)

Focusing on latency, each flow has its own tolerance versus extended distance. Before moving ahead to a Split Data Centre, the flows have to be tested over distance.

- Even when avoiding L2/L3 forwarding latency imposed by switches and routers in an IP/VPN network,
- the light-speed, though very high, is *finite*
  - 10 microsec/km round-trip time
- These round-trips may impact delay-sensitive flows
  - Synchronous Mirroring : “IO-complete” after round-trip from Primary to Mirrored storage unit
  - Server Clustering
  - Typical Mainframe flows
    - Coupling Links
    - Sysplex Timer
- Only way to predict behavior of flows over distance is to run … distance tests!

The ideal distance test is a blind test, running the full production workload. If instead, a controlled test has to be organized, the submitted workloads have to be selected carefully. In both cases, metrics have to be monitored to unambiguously assess the impact of distance.

- Provide a testbed
  - Minimal, but sufficient subset of the optical network …
    - xWDM nodes => agree with vendor/supplier on T&C's (mostly Try & Buy)
    - Set of coils to realize distance between the nodes => today available at 'affordable' price
  - … providing connectivity to selected client equipment, especially for flows that are known as latency sensitive
    - Coupling links
    - Storage links (FICON, FibreChannel)
      - Mainframe-only, or open systems as well?
    - Sysplex Timer
    - LAN (GE, 10GE) (?)
    - Other
- Agree on how to measure
  - 'subjective' criteria : feedback from HD, users …
  - Quantified performance figures, response times, CPU load, Memory Utilization (e.g. RMF) etc.
- Select your test approach
  - Either run 'full' production over optical testbed during an adequate time-span
    - Disconnect devices from their direct fiber connections …
    - … and reconnect them over xWDM testbed
    - Have available a solid and fast fail-back procedure, from xWDM connectivity to direct connectivity
  - Or submit controlled workloads off-shift
    - Prepare relevant workloads, typically batch jobs
    - Free your production environment
    - Submit your test-jobs as a baseline run at no distance
    - Add distance and resubmit your jobs
  - Pure LAB-setup, using two test-LPARs and a TestPlex, put at distance
    - Questionable reliability due to high performance fluctuations on Test LPARs

A full-production distance test running under an adequate risk contingency plan, is the best assessment for viability of a (xWDM) network to connect a split-data centre

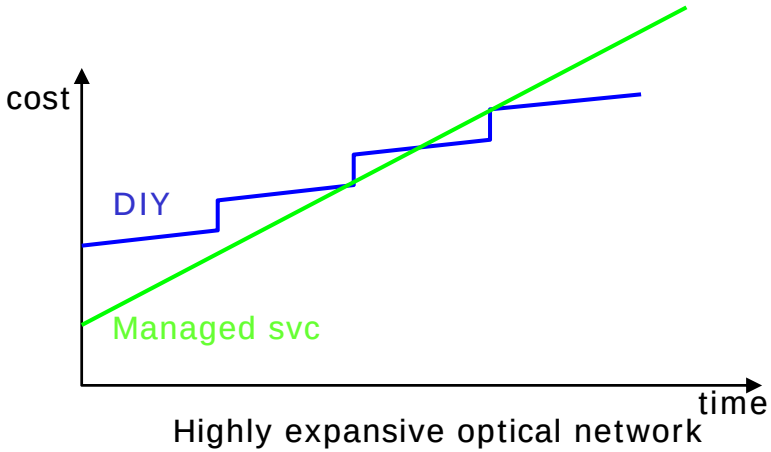
	Full production	Controlled workloads off-shift	Lab setup
Value	Very high	High	Medium-low
Effort	Medium	Very high	Low
Risk	High (*)	Low	Low

(*) Risk assessment, analysis and mitigation plans are critical

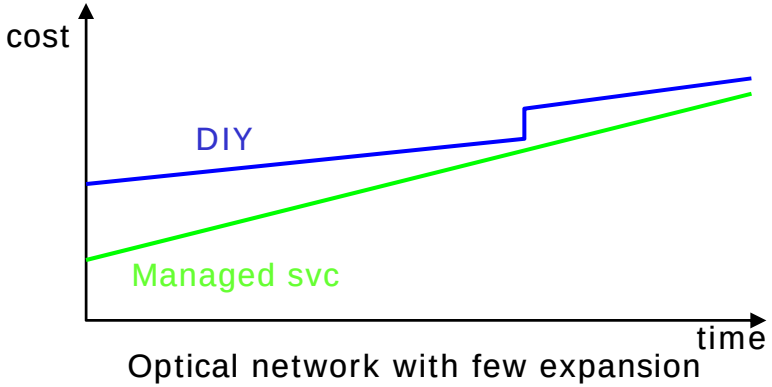
?? Questions ??

Backup slides

To deploy an xWDM network, there are two major approaches : Do It Yourself or Lease a Managed Service



Highly expansive optical network



Optical network with few expansion