

Intelligent Load Balancing with IBM Multi-site Workload Lifeline

Mike Fitzpatrick – mfitz@us.ibm.com
IBM Raleigh, NC

Thursday, February 7th, 1:30pm
Session: 12860

Trademarks, notices, and disclaimers

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both:

- | | | | | |
|---|---|---|--|--|
| <ul style="list-style-type: none"> • Advanced Peer-to-Peer Networking® • AIIX® • alphaWorks® • AnyNet® • AS/400® • BladeCenter® • Candle® • CICS® • DataPower® • DB2 Connect • DB2® • DRDA® • e-business on demand® • e-business (logo) • e business (logo)® • ESCON® • FICON® | <ul style="list-style-type: none"> • GDDM® • GDPS® • Geographically Dispersed Parallel Sysplex • HiperSockets • HPR Channel Connectivity • HyperSwap • i5/OS (logo) • i5/OS® • IBM eServer • IBM (logo)® • IBM® • IBM zEnterprise™ System • IMS • InfiniBand® • IP PrintWay • IPDS • iSeries • LANDP® | <ul style="list-style-type: none"> • Language Environment® • MQSeries® • MVS • NetView® • OMEGAMON® • Open Power • OpenPower • Operating System/2® • Operating System/400® • OS/2® • OS/390® • OS/400® • Parallel Sysplex® • POWER® • POWER7® • PowerVM • PR/SM • pSeries® • RACF® | <ul style="list-style-type: none"> • Rational Suite® • Rational® • Redbooks • Redbooks (logo) • Sysplex Timer® • System i5 • System p5 • System x® • System z® • System z9® • System z10 • Tivoli (logo)® • Tivoli® • VTAM® • WebSphere® • xSeries® • z9® • z10 BC • z10 EC | <ul style="list-style-type: none"> • zEnterprise • zSeries® • z/Architecture • z/OS® • z/VM® • z/VSE |
|---|---|---|--|--|

* All other products may be trademarks or registered trademarks of their respective companies.

The following terms are trademarks or registered trademarks of International Business Machines Corporation in the United States or other countries or both:

- Adobe, the Adobe logo, PostScript, and the PostScript logo are either registered trademarks or trademarks of Adobe Systems Incorporated in the United States, and/or other countries.
- Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license there from.
- Java and all Java-based trademarks are trademarks of Sun Microsystems, Inc. in the United States, other countries, or both.
- Microsoft, Windows, Windows NT, and the Windows logo are trademarks of Microsoft Corporation in the United States, other countries, or both.
- InfiniBand is a trademark and service mark of the InfiniBand Trade Association.
- Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel SpeedStep, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.
- UNIX is a registered trademark of The Open Group in the United States and other countries.
- Linux is a registered trademark of Linus Torvalds in the United States, other countries, or both.
- ITIL is a registered trademark, and a registered community trademark of the Office of Government Commerce, and is registered in the U.S. Patent and Trademark Office.
- IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency, which is now part of the Office of Government Commerce.

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

Refer to www.ibm.com/legal/us for further legal information.

Complete your sessions evaluation online at SHARE.org/SFEval

Agenda

- ❑ What is Multi-site Workload Lifeline?
- ❑ Providing Continuous Availability
- ❑ Providing Graceful Switching
- ❑ Multi-site Workload Lifeline
- ❑ Appendix: Configuration Statements



Disclaimer: All statements regarding IBM future direction or intent, including current product plans, are subject to change or withdrawal without notice and represent goals and objectives only. All information is provided for informational purposes only, on an “as is” basis, without warranty of any kind.

Multi-site Workload Lifeline

⇒ ***What is Multi-Site Workload Lifeline?***

Providing Continuous Availability

Providing Graceful switching

Multi-site Workload Lifeline

Appendix: Configuration Statements

Multi-site Workload Lifeline Benefits

- Multi-site Workload Lifeline plays a key role in solving 2 major problems in the Enterprise
 - Continuous availability for critical workloads
 - Downtime for planned outages

- Enables intelligent load balancing of TCP/IP workloads across two sites at unlimited distances to provide nearly continuous availability
 - Increased performance: Response time is reduced by ensuring new connections for a workload are distributed to the applications and systems most capable of handling them
 - Increased availability: New connections for a workload can be routed to available applications even in the event of application, system, or site outages
 - Increased scalability: Application instances can be added on demand
 - Analytic capability: Network Management Interface (NMI) provides access to workload, application, and site status information
 - Improved recovery time: Reduction of Recovery Time Objective from hours to minutes

Multi-site Workload Lifeline Benefits...

- Enables movement of workloads from one site to another by providing graceful rerouting
 - Workload migration: Ability to move workloads from one site to the other with minimal disruption
 - Increased availability: Outages for maintenance updates or other planned events can be minimized
 - Verification of disaster recovery procedures: Simpler, nondisruptive testing of disaster recovery procedures by validating workloads remain accessible on the recovery site without requiring a site outage on the production site

Multi-site Workload Lifeline

What is Multi-Site Workload Lifeline?



Providing Continuous Availability

Providing Graceful switching

Multi-site Workload Lifeline

Appendix: Configuration Statements

What are customers doing today at metro distances?



- GDPS/PPRC, based upon a multi-site Parallel Sysplex and synchronous disk
 - Workloads can withstand site and/or storage failures
- GDPS/PPRC supports two configurations:
 - active/standby
 - active/active
- Some customers have deployed GDPS/PPRC active/active configurations
 - All critical data must be PPRC'd and HyperSwap enabled
 - All critical CF structures must be duplexed
 - Applications must be parallel sysplex enabled
 - Signal latency will impact OLTP thrupt and batch duration resulting in the sites being separated by no more than tens of KM (fiber)
- Low recovery time and zero data loss
- Issue: the GDPS/PPRC active/active configuration does not provide enough

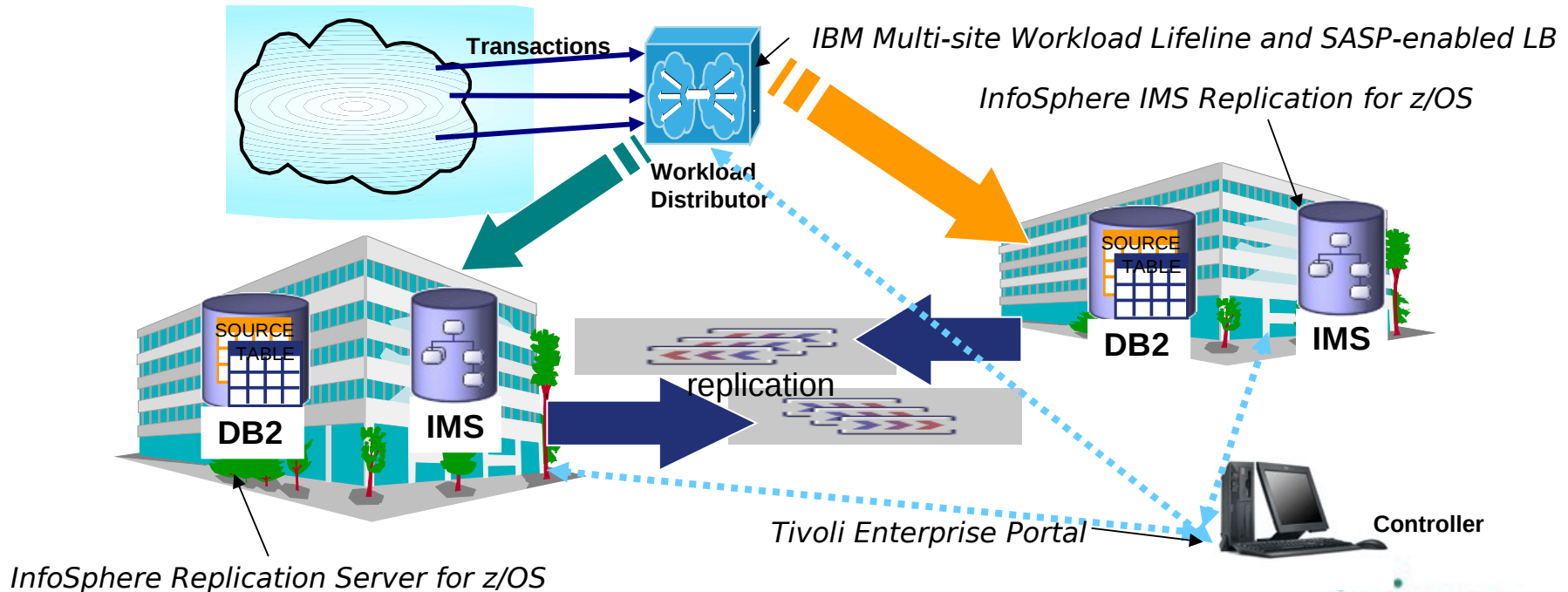
What are customers doing today at global distances?



- GDPS/XRC and GDPS/GM, based upon separate Sysplexes and
 - Recovery with “seconds” of data loss
 - Disaster recovery for out of region interruptions
- The current GDPS asynchronous replication products require the failed site’s
 - Power fail consistency
 - Transaction consistency
- There are no identified extensions to the existing GDPS async replication
- Issue: GDPS/XRC and GDPS/GM will not achieve an RTO of seconds being

GDPS Active-Active Sites – What is it?

- Two or more sites, separated by *unlimited* distances, running the same applications and having the same data to provide cross-site workload balancing and Continuous Availability / Disaster Recovery
- Access data from any site (unlimited distance between sites)
- Provide workload distribution between sites
- Provide application level granularity
- Paradigm shift: failover model => near continuous availability model
 - For critical workloads requiring continuous availability



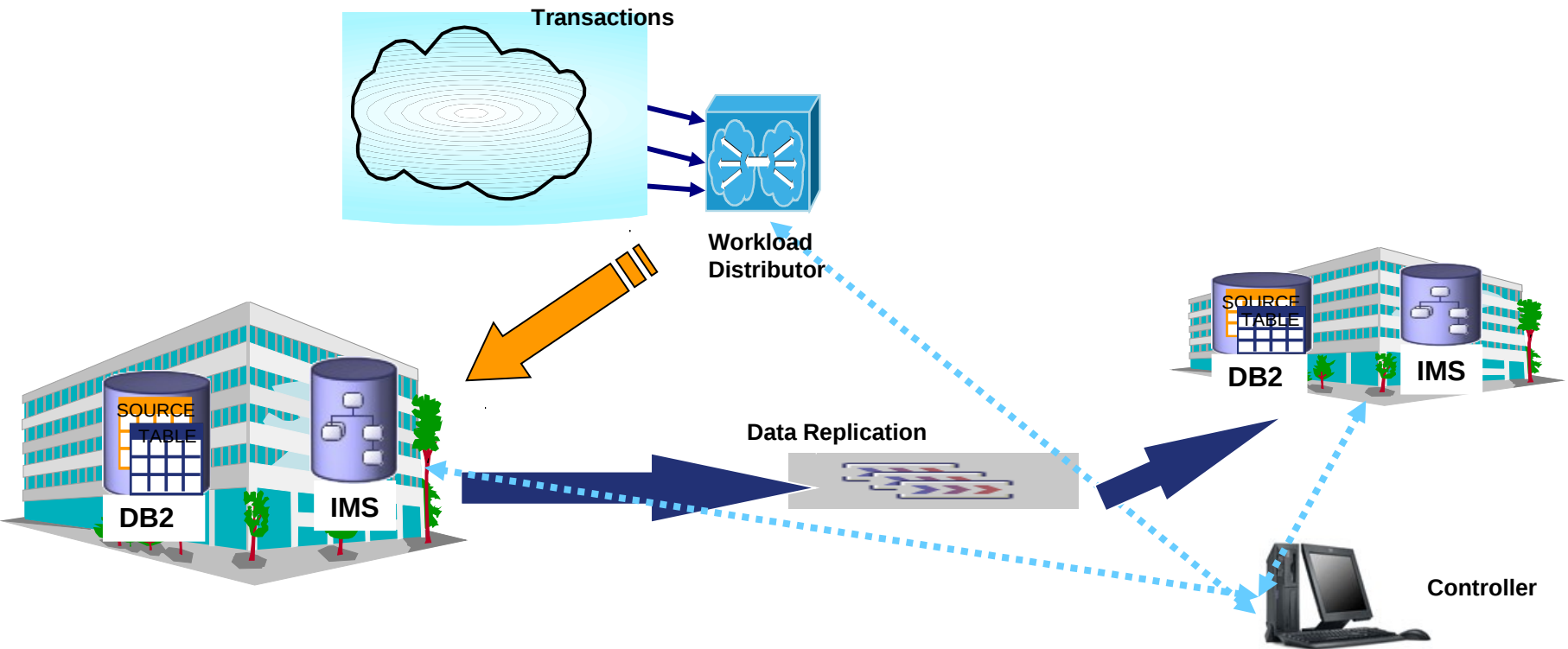
GDPS Active-Active Sites Configurations

- Configurations
 - Active/Standby
 - Active/Query (future)
- A configuration is specified on a workload basis
- A workload is the aggregation of these components
 - **Software:** applications (e.g., COBOL program) and the middleware run time environment (e.g., CICS region & DB2 subsystem)
 - **Data:** related set of objects that must preserve transactional consistency (e.g., DB2 Tables)
 - **Network connectivity:** one or more TCP/IP addresses & ports (e.g., 10.10.10.1:80)

Active/Standby Configuration

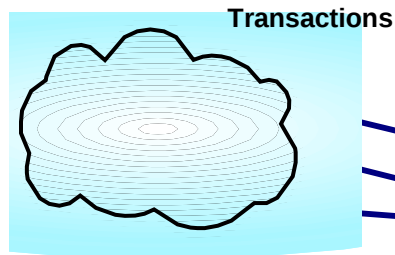
Site 1
Workload A active

Site 2
Workload A standby

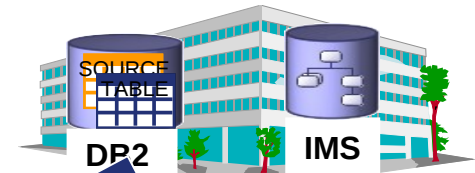


Active/Standby Configuration...

~~Site 1
Workload A active~~



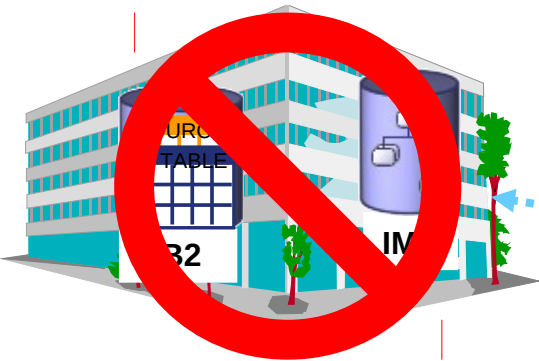
Site 2
Workload A active



Data Replication



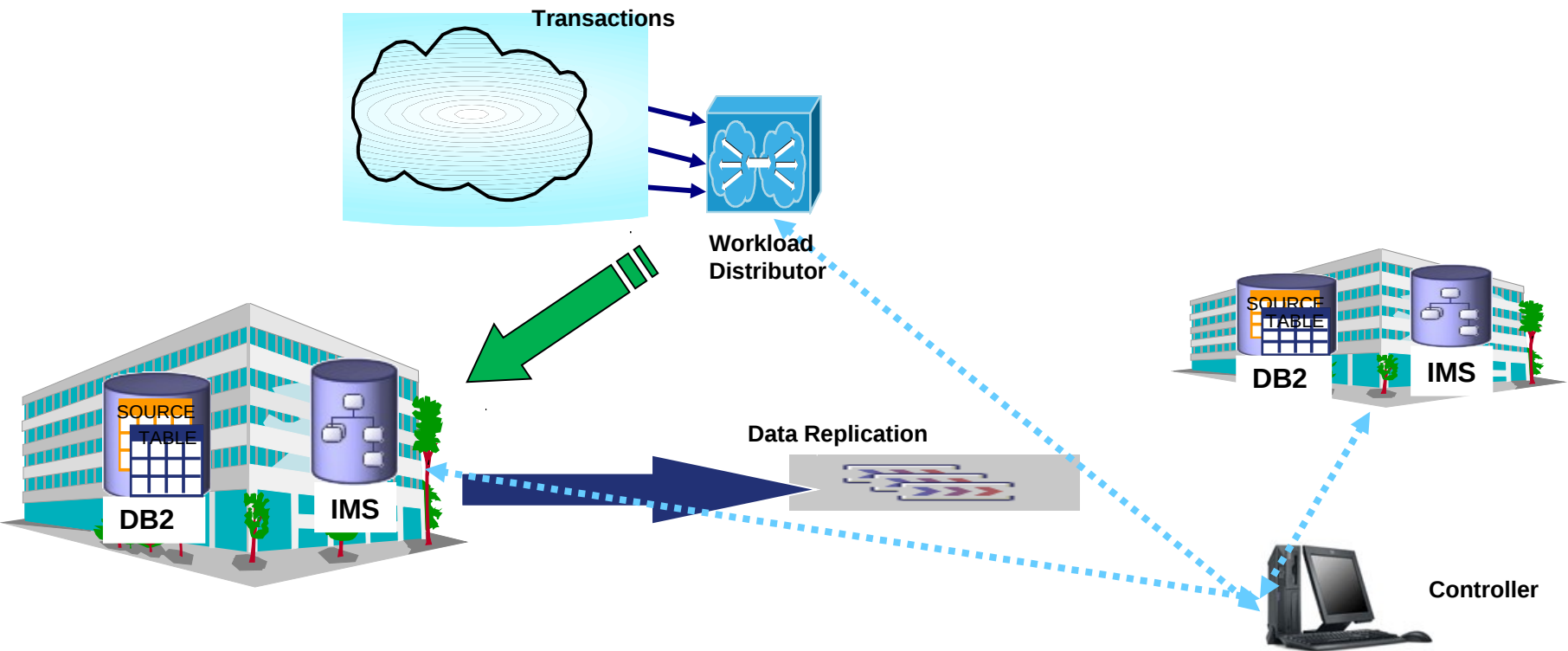
data queued



Active/Standby Configuration (multiple workloads)

Site 1
Workload A active

Site 2
Workload A standby



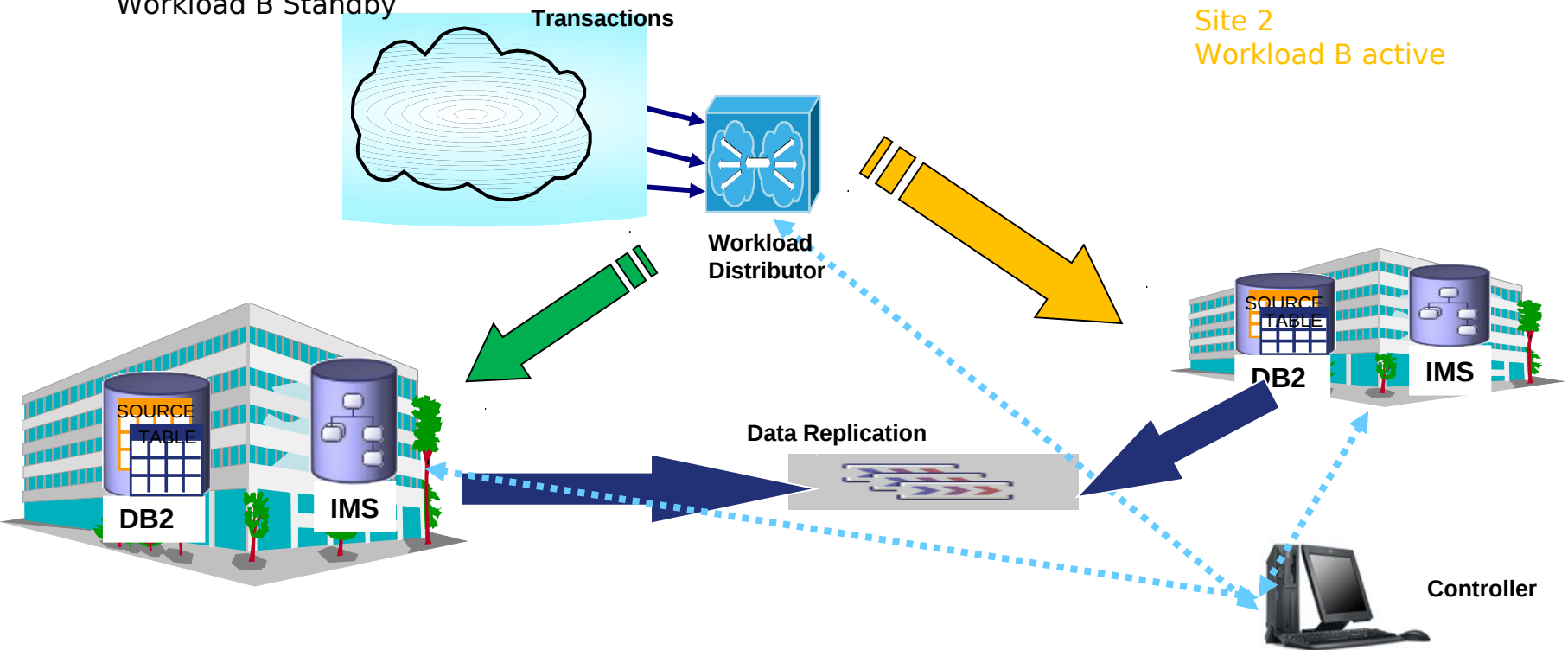
Active/Standby Configuration (multiple workloads)...

Site 1
Workload A active

Site 1
Workload B Standby

Site 2
Workload A standby

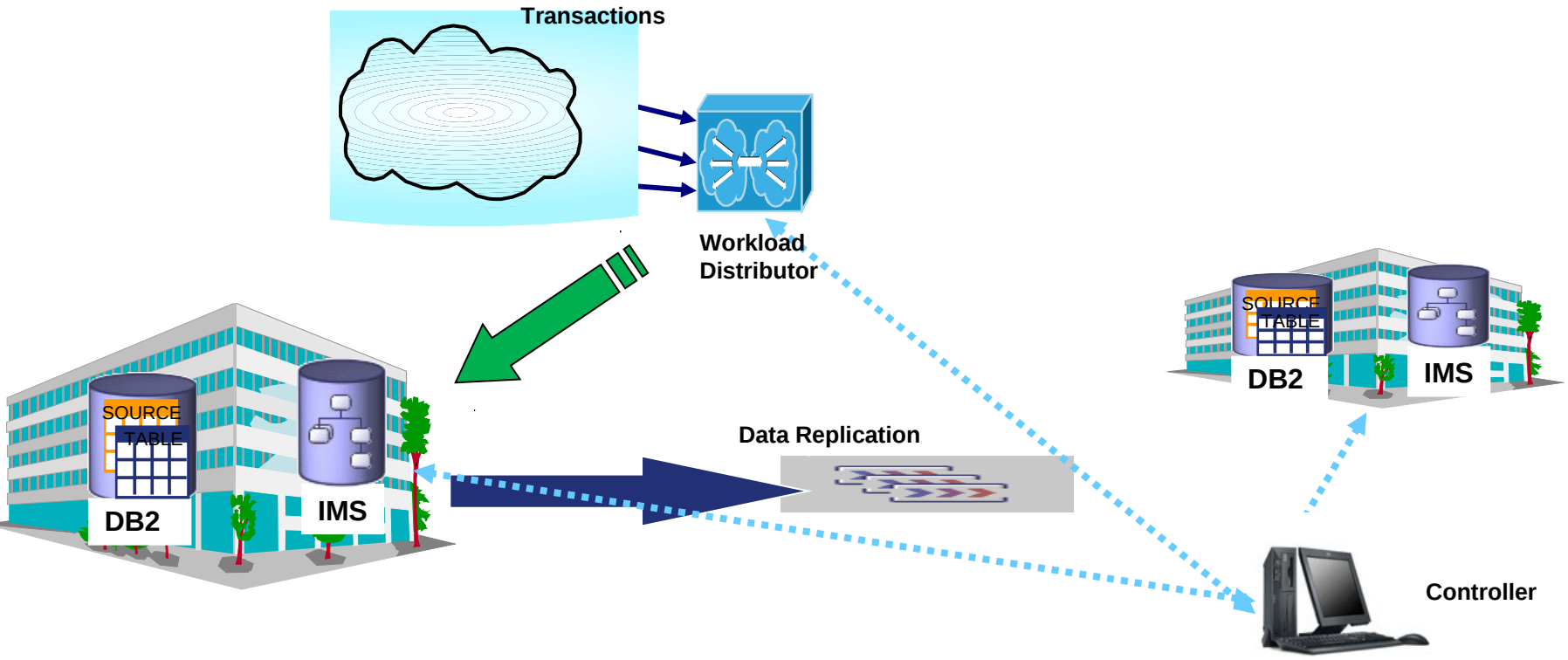
Site 2
Workload B active



Active/Query Configuration

Site 1
Workload A active

Site 2
Workload A standby



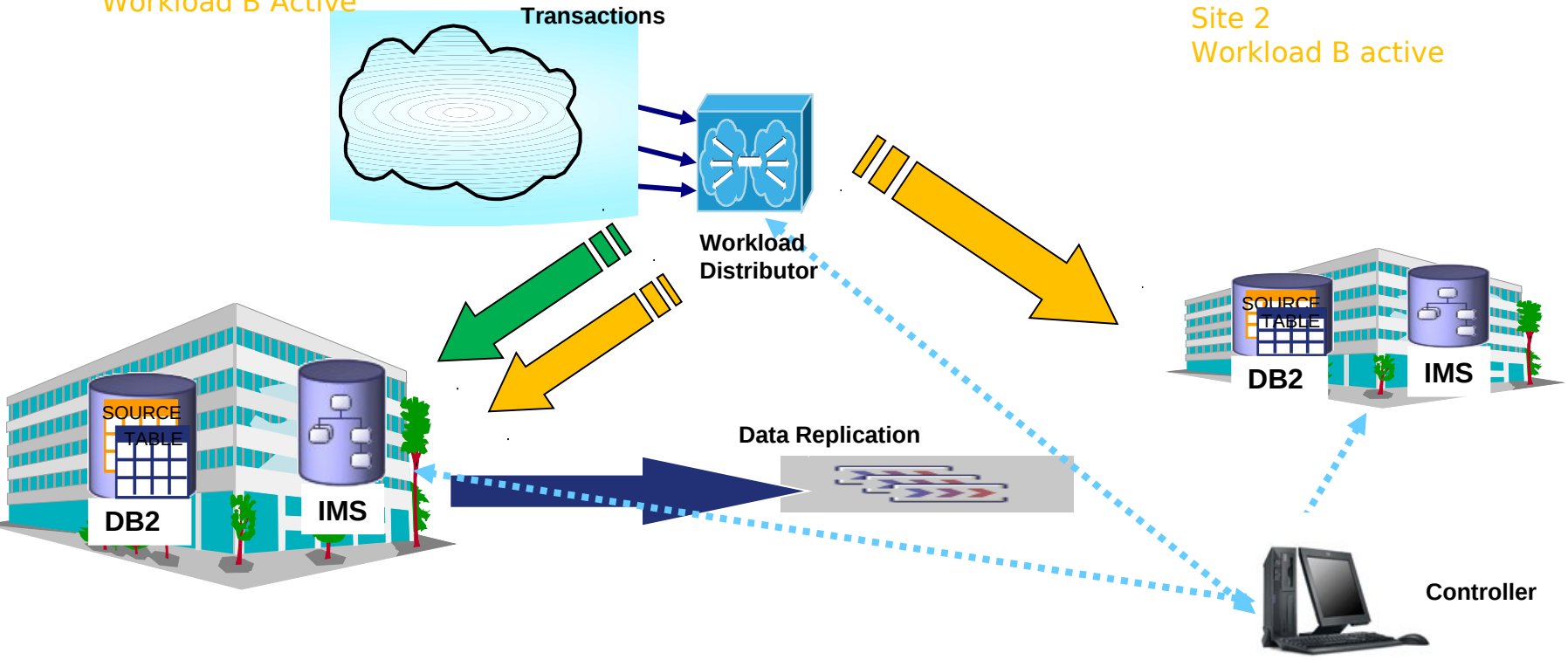
Active/Query Configuration...

Site 1
Workload A active

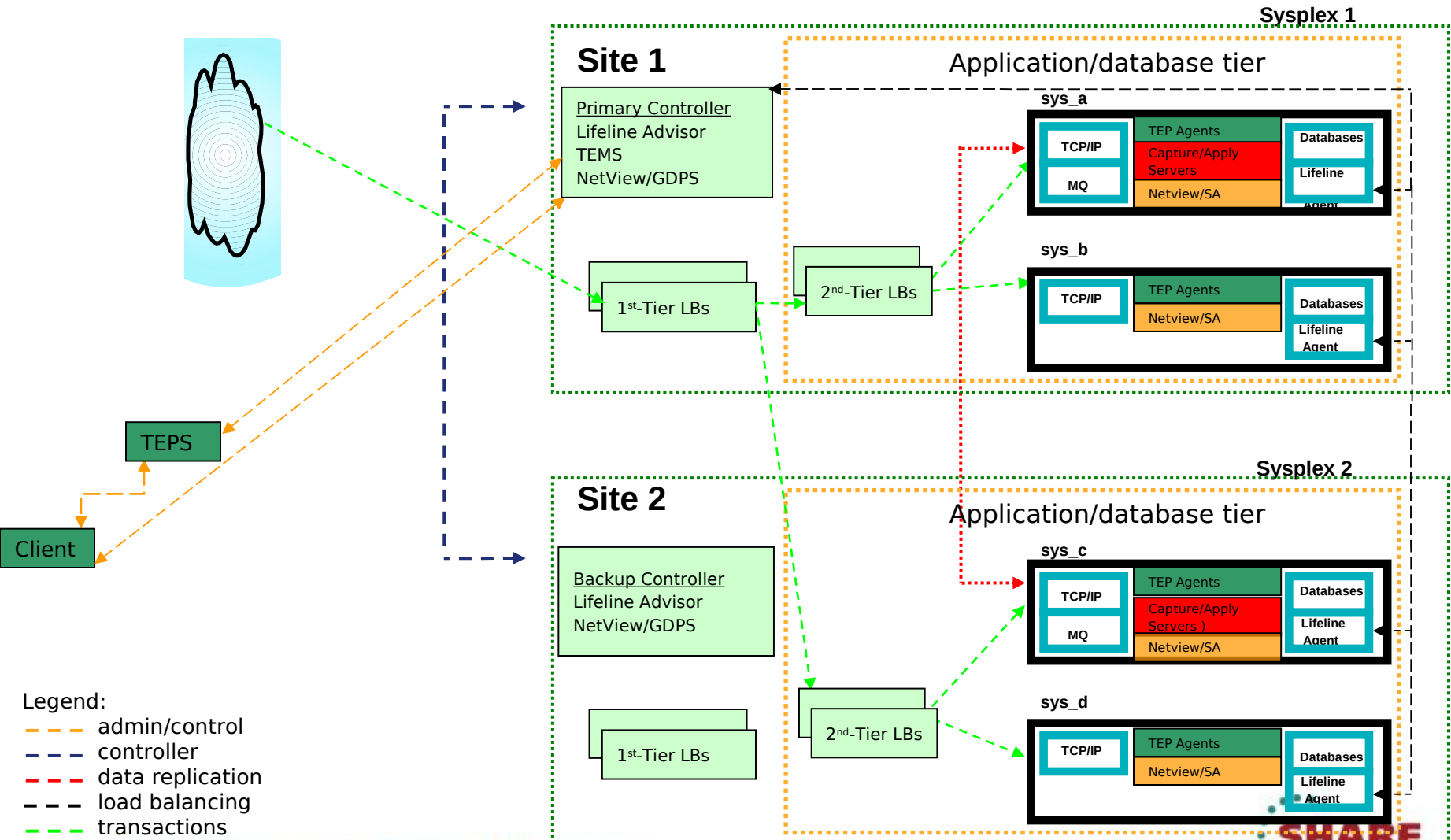
Site 1
Workload B Active

Site 2
Workload A standby

Site 2
Workload B active



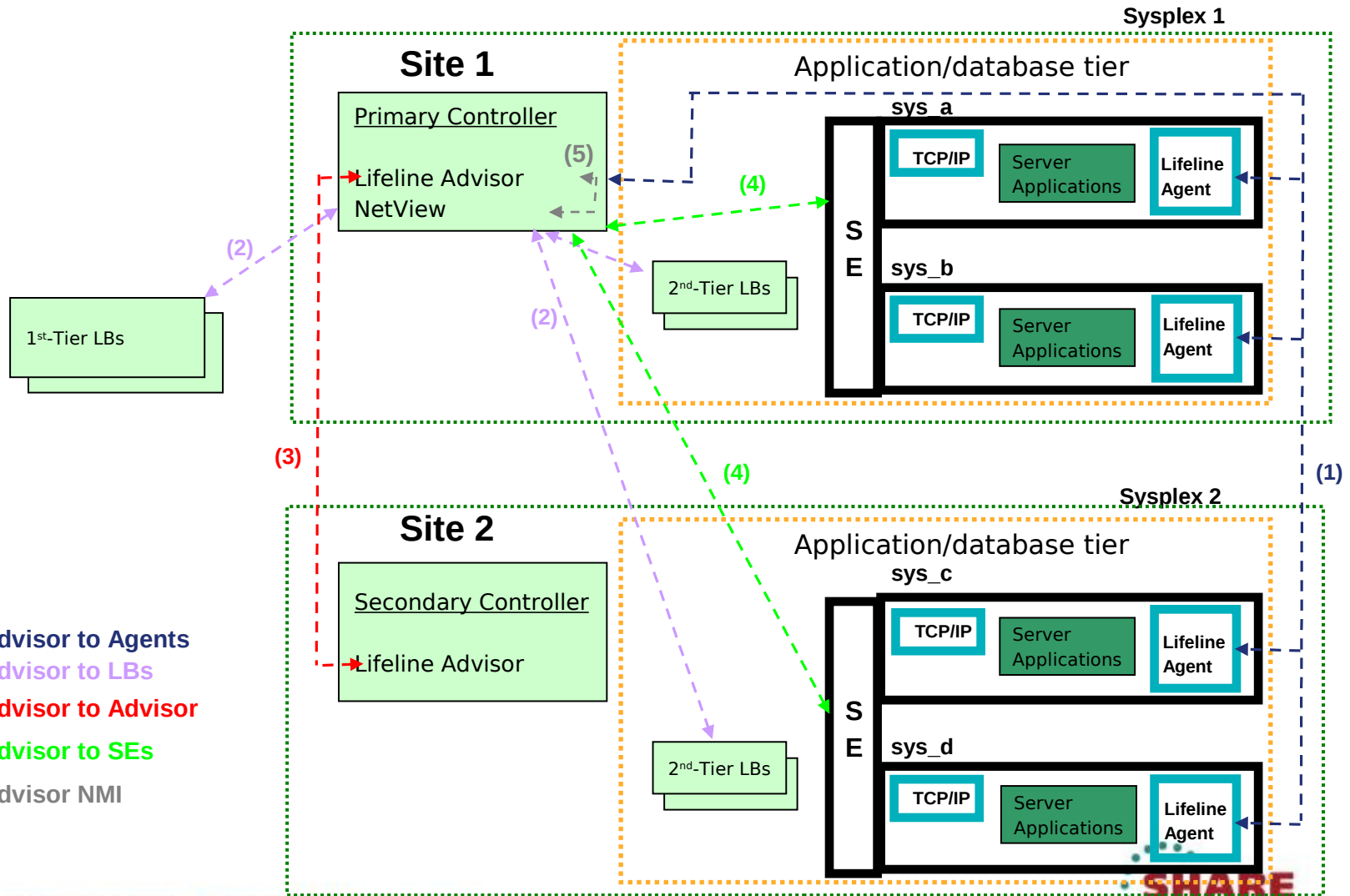
Active/Active Sites Structure



GDPS Active-Active Sites load balancing requirements

- Ability to distribute workloads between sites (and route around failed sites)
 - Based on capacity/health of sites and server application instances within a site
- Ability to detect workload or site failures
- Ability to switch workloads from one site to another site
 - Perform “graceful” failback following a workload or site disaster
- Ability to maintain workload configuration states in event of a workload manager failure
 - Keep a peer workload manager in sync with workload states
- Ability to dynamically add/modify workloads
- Ability to surface routing recommendations to network management agents

Workload Lifeline Structure



Multi-site Workload Lifeline

What is Multi-Site Workload Lifeline?

Providing Continuous Availability

⇒ ***Providing Graceful switching***

Multi-site Workload Lifeline

Appendix: Configuration Statements

Data Replication

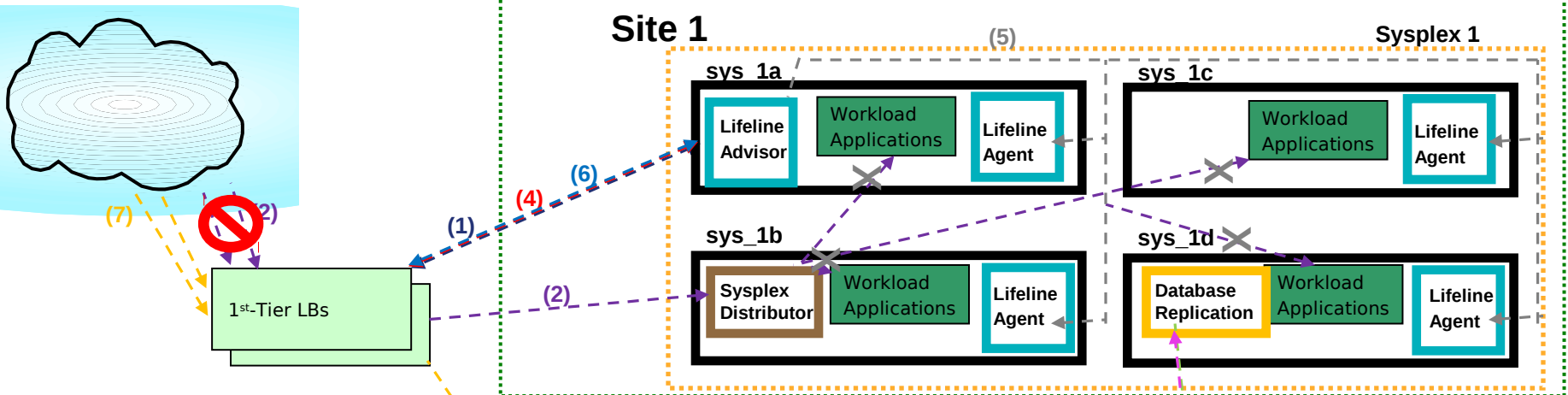
- What is data replication?
 - A solution for copying data between databases, typically residing in different sites
 - Emphasizes the copying of only changed data
 - An application makes updates to a database and these changes are captured locally and applied to a remote database
 - Replication scope
 - An entire database
 - A subset of the database (subset of tables or subset of columns or rows within a table)

- Why use data replication?
 - Offload query workloads to replicated database
 - Read-only database provides near-real time reporting
 - Continuous (High) Availability
 - Failover to replicated database during disaster recovery

Graceful Switch load balancing requirements

- Ability to distribute workloads between sites
 - Based on customer-driven commands
- Ability to switch workloads from one site to another site
 - Perform “graceful” takeover for site maintenance
- Ability to maintain workload configuration states in event of a planned outage
 - Keep a peer workload manager in sync with workload states

IBM Multi-site Workload Lifeline providing graceful workload movement

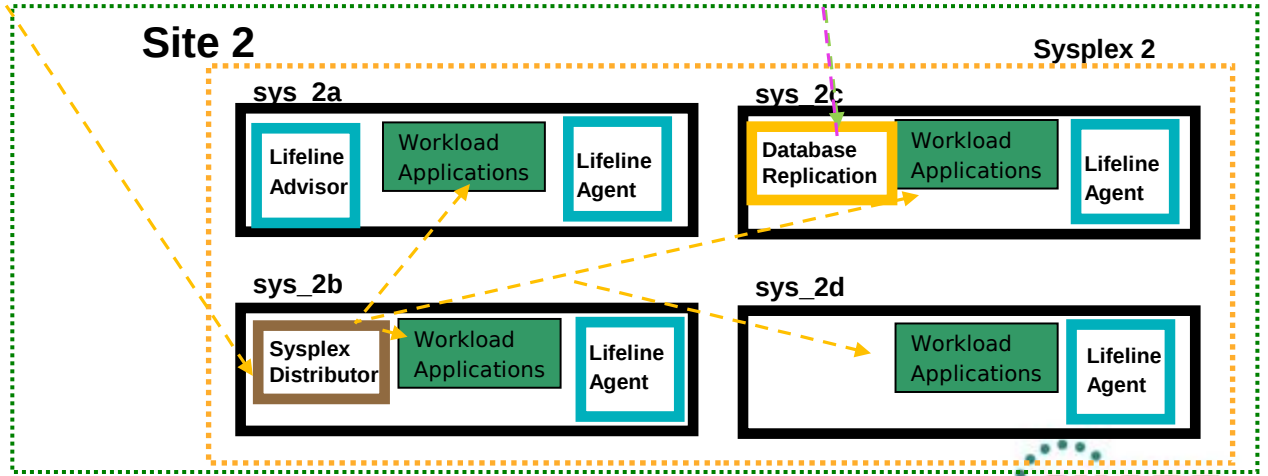


Prior to planned outage

- (1) Advisor notifies LBs about active site (Site 1)
- (2) Client connections distributed to Site 1
- (3) Database updates replicated to Site 2

Initiate graceful movement

- (4) Advisor notifies LBs to stop new connections
- (5) Advisor notifies Agents to drop active connections
- (6) Advisor notifies LBs about new active site (Site 2)
- (7) Client connections distributed to Site 2
- (8) Database updates replicated to Site 1



Multi-site Workload Lifeline

What is Multi-Site Workload Lifeline?

Providing Continuous Availability

Providing Graceful switching

⇒ ***Multi-site Workload Lifeline***

Appendix: Configuration Statements

Workload Lifeline role in Continuous Availability



Advisor provides distribution recommendations to multiple tiers of load balancers

- Server-specific WLM metrics and Communications Server weights provided by Agents running in all LPARs across both sites are used to build recommendations

Site recommendations to 1st-tier load balancers

- Direct 1st-tier load balancers to route new connections for a workload to a 2nd-tier load balancer within a site (using SASP – see RFC 4678)
 - F5 Big IP Switch, Cisco CSM, and Citrix NetScaler currently support SASP
- Site selection determined by where the workload is currently active

Server application recommendations to 2nd-tier load balancers

- Direct 2nd-tier load balancers to route new connections for a workload to specific server applications within the site (using SASP)
- Server application selection determined by recommendations provided by the Agents within the site
- Sysplex Distributor may assume role of 2nd-tier load balancer
 - No server application recommendations provided by Advisor in this case

Workload Lifeline distribution recommendations

- Agents provide relative weights per server application instance
 - WLM weight
 - Server-specific WLM recommendations: reflects how much displaceable capacity is available on the target system at the importance level of the server application
 - Communications Server weight
 - This weight is calculated based on the availability of the actual server instances (are they up and ready to accept workload) and how well TCP/IP and the individual server instances process the workload that is sent to them.
 - Prevent stalled server instance from being sent more work (accepting no new connections and new connections are being dropped due to backlog queue full condition)
 - Proactively react to server instance that is getting overloaded (accepting new connections, but size of backlog queue increases over time approaching the max backlog queue size)
- Advisor uses relative weights of all the server application instances for a workload to determine whether the workload is available/healthy within a site or whether a workload failure has occurred

Workload Lifeline role in Continuous Availability...

- Advisor provides ability to group different server applications into a workload
 - Distinguish different workloads and perform different distribution decisions based on the workload (direct each workload to its Active site)
- Advisor responsible for detecting workload failures
 - Monitor the capacity of LPARs within a workload's Active site and availability/health of the server applications that make up the workload
 - Ability to dynamically switch a workload to the alternate site after detecting a failure
- Advisor responsible for detecting site failures
 - Monitor the availability/reachability of the LPARs that make up the site
 - Communication with Agents active on the LPARs verifies IP network connectivity to the site
 - Communication with Support Elements (SE) over HMC network verifies LPAR status
 - Ability to dynamically switch all workloads to the alternate site after detecting a failure

Workload Lifeline role in Continuous Availability...

- Advisor communicates with a peer Advisor
 - Shares workload state information
 - A workload can be inactive
 - A workload can be active to a specific site
 - Peer Advisor takes over responsibilities in the event the primary Advisor fails

- Advisor provides graceful movement of a workload to an alternate site (a 'planned' failure)
 - Prevents new connections for the workload from being distributed to the Active site
 - Terminates any existing connections being distributed to the Active site
 - Reroutes new transactions to the alternate site

- Advisor has ability to dynamically add or modify existing workloads to an active configuration
 - Allows changes without recycling the Advisor

- Advisor provides Network Management Interface (NMI)
 - Surface workload states, distribution recommendations, and component information to network management agents

Workload Lifeline role in Continuous Availability...

- Agents communicate with a Communications Server TCPIP stack
 - Extracts information about available server applications and server application health via documented interfaces

Workload Lifeline role in Graceful Switch

- To ensure graceful movement, a key requirement is that updates to a database (being replicated to another site) can only be occurring on one site at a time
 - Lifeline orchestrates the workload movement to ensure workload connections are directed to only one site (i.e. the active site)
- Prior to planned outage, all workload connections are directed to a single site
 - Accomplished via operator-initiated MODIFY command against Lifeline
- To facilitate graceful movement, the following steps are taken:
 - Any new workload connections must be stopped from being routed to the active site (accomplished via operator-initiated MODIFY command against Lifeline)
 - Wait a period of time to allow any outstanding transactions on existing workload connections time to complete
 - Reset any workload connections that are not completing in a timely manner (accomplished via operator-initiated MODIFY command against Lifeline)
 - Redirect all new workload connections to the alternate site (accomplished via operator-initiated MODIFY command against Lifeline)

Key Advisor Display commands

- **MODIFY advproc,DISPLAY,ADVISOR,DETAIL**
 - When issued on the primary Advisor, displays the role of the Advisor, the connected load balancers (and whether it is a 1st-tier or 2nd-tier), the connected Agents (including system and site name where the Agents are active), and the connected peer Advisor (including the system name where the peer is active)
 - When issued on the peer Advisor, displays the role of the Advisor and the connected primary Advisor (including the system name where the primary is active)

- **MODIFY advproc,DISPLAY,CONFIG**
 - Displays the current active configuration for the Advisor

Key Advisor Display commands...

- **MODIFY advproc,DISPLAY,LB,DETAIL**
 - Displays the connected load balancers, including the list of groups registered by the load balancer, the members within each group, and the distribution recommendations provided for each member
- **MODIFY advproc,DISPLAY,WORKLOAD,DETAIL**
 - Displays the status of all defined workloads, including the status of all the server applications that make up the workload

Display Advisor information

F AQSADV,DISPLAY,ADVISOR,DETAIL

AQS0142I ADVISOR DETAILS

ADVISOR ROLE : PRIMARY

IPADDR : 192.10.1.1

LOAD BALANCERS:

IPADDR : 192.10.1.32

TIER : 1

IPADDR : 192.10.1.64

TIER : 2

AGENTS :

IPADDR : 192.10.110.1

SYSTEM NAME : SYS1 SITE : PLEX1

IPADDR : 192.10.110.2

SYSTEM NAME : SYS2 SITE : PLEX1

IPADDR : 192.20.110.1

SYSTEM NAME : SYS3 SITE : PLEX2

IPADDR : 192.20.110.2

SYSTEM NAME : SYS4 SITE : PLEX2

PEER ADVISOR :

IPADDR : 192.20.1.1

SYSTEM NAME : CNTL2

Display workloads

F AQSADV,DISPLAY,WORKLOAD,DETAIL

AQS0146I WORKLOAD DETAILS

WORKLOAD NAME : WORKLOAD1

STATE : ACTIVE SITE : PLEX2

SERVERS:

IPADDR..PORT : 192.10.110.1..5001

SYSTEM NAME : SYS1 SYSPLEX : PLEX1 STATUS : AVAIL

IPADDR..PORT : 10.20.1.1..5001

SYSTEM NAME : SYS3 SYSPLEX : PLEX2 STATUS : AVAIL

IPADDR..PORT : 192.10.110.2..5001

SYSTEM NAME : SYS2 SYSPLEX : PLEX1 STATUS : AVAIL

IPADDR..PORT : 10.20.1.1..5001

SYSTEM NAME : SYS4 SYSPLEX : PLEX2 STATUS : UNAVAIL

WORKLOAD NAME : WORKLOAD2

STATE : ACTIVE SITE : PLEX1

SERVERS:

IPADDR..PORT : 192.10.111.1..8020

SYSTEM NAME : SYS1 SYSPLEX : PLEX1 STATUS : AVAIL

IPADDR..PORT : 10.21.1.1..8020

SYSTEM NAME : SYS3 SYSPLEX : PLEX2 STATUS : AVAIL

:

:

Key Advisor State Change commands

- **MODIFY advproc,ACTIVATE,WORKLOAD=...,SITE=...**
 - Signals the Advisor to direct 1st-tier load balancers to distribute new connections for the specified workload to the requested site
- **MODIFY advproc,DEACTIVATE,WORKLOAD=...**
 - Signals the Advisor to direct Agents on the site where the specified workload was last active to reset any existing connections for this workload
- **MODIFY advproc,QUIESCE,WORKLOAD=...**
 - Signals the Advisor to direct 1st-tier load balancers to stop distributing new connections for the specified workload to any site
- **MODIFY advproc,REFRESH**
 - Signals the Advisor to reread its configuration file and apply any updates to its active configuration
- **MODIFY advproc,TAKEOVER**
 - Signal the peer Advisor to take over primary Advisor responsibilities from the current primary Advisor

Key Agent Display commands

- **MODIFY ageproc,DISPLAY,CONFIG**
 - Displays the current active configuration for the Agent
- **MODIFY ageproc,DISPLAY,MEMBERS,DETAIL**
 - Displays information about each of the server applications this Agent was asked to monitor, including whether the server application exists, the jobname of the server application, and current state of the server application

Display Members information

F AQSAGE,DISPLAY,MEMBERS,DETAIL

AQS0115I MEMBER DETAILS

LB INDEX : 00 UUID : A67B6699

GROUP NAME : WKLD2_GROUP1

IPADDR..PORT: 10.10.1.1..8020

MATCHES : 001 PROTOCOL : TCP

FLAGS : ANY DISTDVIPA

TCPNAME : TCPIP

JOBNAME : JOB1 ASID : 0034 RESOURCE : 0000096B

GROUP NAME : WKLD2_GROUP2

IPADDR..PORT: 10.10.1.1..8021

MATCHES : 000 PROTOCOL : TCP

FLAGS : DISTDVIPA

TCPNAME : TCPIP

JOBNAME : N/A ASID : N/A RESOURCE : N/A

LB INDEX : 01 UUID : 9A78BE9E

GROUP NAME : TIER2_GROUP1

IPADDR..PORT: 192.10.110.1..5001

MATCHES : 001 PROTOCOL : TCP

FLAGS :

TCPNAME : TCPIP

JOBNAME : JOB3 ASID : 0036 RESOURCE : 0000096D

GROUP NAME : TIER2_GROUP2

IPADDR..PORT: 192.10.110.1..6001

MATCHES : 001 PROTOCOL : TCP

FLAGS :

TCPNAME : TCPIP

JOBNAME : JOB4 ASID : 0037 RESOURCE : 0000096E

Key Agent State Change commands

- **MODIFY ageproc,ENABLE,...**
 - Signals the Agent to enable server applications (make them available to be load balanced to)
 - Server applications bound to a distributable dynamic VIPA must be enabled using the VARY TCPIP,,SYSPLEX,RESUME command

- **MODIFY ageproc,QUEISCE,...**
 - Signals the Agent to quiesce server applications (make them unavailable to be load balanced to)
 - Server applications bound to a distributable dynamic VIPA must be quiesced using the VARY TCPIP,,SYSPLEX,QUIESCE command

Debugging

- All debugging information recorded in syslogd
 - Requires the syslogd daemon be configured and started
- Enable debugging during startup
 - `debug_level` statement in both Advisor and Agent configuration
- Dynamically enable, disable, change debugging while active
 - **MODIFY advproc,DEBUG,LEVEL=...** for Advisor
 - **MODIFY ageproc,DEBUG,LEVEL=...** for Agent
- Display current debug level
 - **MODIFY advproc,DISPLAY,DEBUG** for Advisor
 - **MODIFY ageproc,DISPLAY,DEBUG** for Agent
- Default level traces errors, warnings, and commands

Advisor to Load Balancer communication

- Server Application State Protocol (SASP)
 - Open protocol documented in RFC4678
 - Provides a mechanism for workload managers to give distribution recommendations to load balancers
 - Does not handle the transport or actual distribution of work, only provides recommendations

- 1st-tier load balancers register groups it is interested in load balancing
 - Each group designates a list of 2nd-tier load balancers (members) it will distribute connections to (either another SASP-enabled load balancer or a Sysplex Distributor node)
 - Identified by protocol (TCP/UDP), IP addresses (IPv4/IPv6) of the 2nd-tier load balancers, and the port number used by the server applications that the 2nd-tier load balancer will be load balancing
 - Advisor uses its lb_id_list to verify whether a load balancer is allowed to connect
 - Advisor uses its cross_sysplex_list configuration statement to map groups to a workload

Advisor to Load Balancer communication...

- 2nd-tier load balancers register Groups it is interested in load balancing
 - Each group designates a list of server applications (members) to be load balanced
 - Identified by protocol (TCP/UDP), IP addresses (IPv4/IPv6) of the target systems the server applications reside on, and the port number used by the server applications

- Load balancers can request to receive distribution recommendations using two possible methods
 - Load balancer will periodically “pull” member distribution recommendations from the Advisor
 - Advisor will periodically “push” member distribution recommendations to the load balancer
 - Can be configured to only “push” changed information about members

Advisor to Agent communication

Internal protocol for communication

Agents connect to Advisor

- Each Agent registers its system name, site name (i.e. sysplex name), and LPAR name
- Advisor uses its agent_id_list configuration statement to verify Agent is allowed to connect
- Advisor uses its cross_sysplex_list configuration statement to verify Agent resides in valid site

Advisor sends information about all members it wants the Agent to monitor

- The IP address, port number, and protocol for all server applications that were registered as group members by 2nd-tier load balancers
- The IP address and port number for target server applications being distributed by the Sysplex Distributors node that were registered as group members by 1st-tier load balancers

Advisor to Agent communication...

- Agent sends periodic updates about system to Advisor
 - List of active members (server applications) active on its system
 - Server WLM recommendation for each member
 - Communications Server health on its system
 -
- Advisor sends requests to reset connections to Agents
 - In response to a DEACTIVATE command, Advisor sends a list of server applications (that make up a workload) to Agents to direct them to reset any active connections for these server applications

Advisor to Advisor communication

- Internal protocol
- Peer Advisor (secondary) connects to Advisor (primary)
 - Peer Advisor registers its system name and LPAR name
 - Primary Advisor uses its `advisor_id_list` configuration statement to verify Advisor is allowed to connect
- Primary Advisor sends information about its active configuration to peer Advisor
 - Verifies configurations are identical between the two Advisors in case the peer Advisor needs to become primary Advisor
- Primary Advisor sends workload state changes to peer Advisor
 - Primary Advisor builds a list of commands that will QUIESCE or ACTIVATE workloads based on their current states
 - Peer Advisor replays this list of commands in event it becomes primary Advisor so that all workloads remain in the same state

Advisor to Support Element (SE) communication

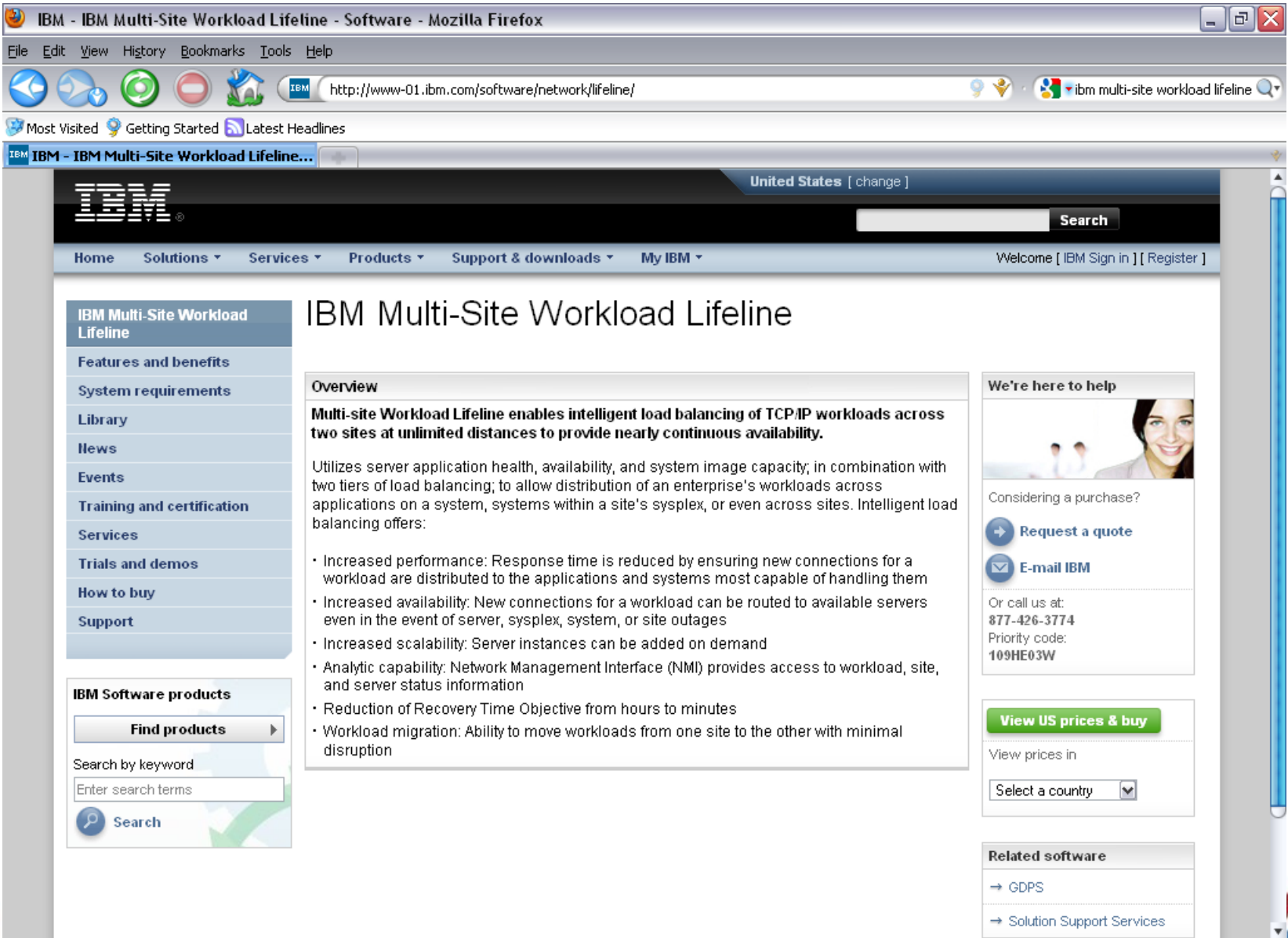
- Base Control Program Internal Interface (BCPii)
 - Documented IBM protocol
 - Allows communication between LPAR where Advisor is active and all interconnected Central Processor Complexes (CPCs)
 - Each CPC can be queried to extract list of LPARs and their status
 - Communication occurs over a Hardware Management Console (HMC) network
 - Typically resides on a different physical network than network used for IP communication

- Advisor uses BCPii address space as a bridge to the SEs
 - New address space shipped in V1R11
 - Advisor uses LPAR names received from peer Advisor and Agents to build list of LPARs to query status information

Advisor to Network Management App communication

- Network Management Interface (NMI)
 - Documented interface
- Advisor creates AF_UNIX socket and accepts connections from network management applications
 - Supplies workload state information, site information, load balancer group registrations, connected load balancers and Agents and peer Advisor, and distribution recommendations for server applications

For more information...



The screenshot shows a Mozilla Firefox browser window with the address bar displaying `http://www-01.ibm.com/software/network/lifeline/`. The page content includes the IBM logo, a navigation menu with items like Home, Solutions, Services, Products, Support & downloads, and My IBM. The main heading is "IBM Multi-Site Workload Lifeline".

IBM Multi-Site Workload Lifeline

Overview

Multi-site Workload Lifeline enables intelligent load balancing of TCP/IP workloads across two sites at unlimited distances to provide nearly continuous availability.

Utilizes server application health, availability, and system image capacity; in combination with two tiers of load balancing; to allow distribution of an enterprise's workloads across applications on a system, systems within a site's sysplex, or even across sites. Intelligent load balancing offers:

- Increased performance: Response time is reduced by ensuring new connections for a workload are distributed to the applications and systems most capable of handling them
- Increased availability: New connections for a workload can be routed to available servers even in the event of server, sysplex, system, or site outages
- Increased scalability: Server instances can be added on demand
- Analytic capability: Network Management Interface (NMI) provides access to workload, site, and server status information
- Reduction of Recovery Time Objective from hours to minutes
- Workload migration: Ability to move workloads from one site to the other with minimal disruption

We're here to help

Considering a purchase?

- ➔ Request a quote
- ✉ E-mail IBM

Or call us at:
877-426-3774
Priority code:
109HE03W

View US prices & buy

View prices in

Select a country ▼

Related software

- ➔ GDPS
- ➔ Solution Support Services

Please fill out your session evaluation

- Intelligent Load Balancing with IBM Multi-site Workload Lifeline
- Session # 12860
- QR Code:



Find us on Facebook at
<http://www.facebook.com/IBMCommserver>



Follow us on Twitter at
http://www.twitter.com/IBM_Commserver



Read the z/OS Communications Server blog at
<http://tinyurl.com/zoscsblog>



Visit the z/OS CS YouTube channel at
<http://www.youtube.com/user/zOSCommServer>

Multi-site Workload Lifeline

Current Disaster Recovery Solutions

GDPS Active-Active Sites

Multi-site Workload Lifeline

⇒ ***Appendix: Configuration Statements***

Key Advisor configuration statements

- `advisor_id_list`
 - List of IP addresses used by primary Advisor to determine which peer Advisors are permitted to connect to it
 - Used by peer Advisor to select a source IP address when connecting to primary Advisor

- `agent_id_list`
 - List of IP addresses used by the Advisor to determine which Agents are permitted to connect to it

- `cross_sysplex_list`
 - Specifies the IP address of the 2nd-tier load balancer, the site name for that load balancer, the port number of the server application used for the workload, and the workload name
 - Used by the Advisor to map 1st-tier load balancer group registrations with workload names
 - Used by the Advisor to validate the sites where connected Agents reside

Key Advisor configuration statements...

- `failure_detection_interval`
 - Time interval used by Advisor to determine how long to wait before declaring a workload or site failure

- `intermediary_nodes_list`
 - Specifies the IP address of the intermediary node, the site name for that intermediary node, the port number of the server application used for the workload, and the workload name
 - Used by the Advisor to map 1st-tier load balancer group registrations with workload names
 - The Advisor does not provide routing recommendations to this intermediary node, as the intermediary node is responsible for routing within the site

- `lb_connection_v4`
 - Specifies the IPv4 address bound by the Advisor to accept connections from load balancers, Agents, and peer Advisor
 - Recommended to be defined as a VIPARANGE dynamic VIPA so that a peer Advisor can take over the dynamic VIPA (when taking over as primary Advisor) without requiring any load balancer or Agent configuration changes

Key Advisor configuration statements...

- `lb_connection_v6`
 - Specifies the IPv6 address bound by the Advisor to accept connections from load balancers, Agents, and peer Advisor
 - Recommended to be defined as a VIPARANGE dynamic VIPA so that a peer Advisor can take over the dynamic VIPA (when taking over as primary Advisor) without requiring any load balancer or Agent configuration changes

- `lb_id_list`
 - List of IP addresses used by the Advisor to determine which load balancers are permitted to connect to it

- `peer_advisor_id`
 - Specifies the IPv4 or IPv6 address used by the peer Advisor to connect to the primary Advisor
 - Used when each Advisor requires a different `lb_connection_v4` or `lb_connection_v6` address (subnets are not allowed to move between sites)

- `update_interval`
 - Time interval communicated to the Agents to specify how frequently the Advisor should be updated with server application metrics

Key Agent configuration statements

- **advisor_id**
 - The IP address used by the Agent as the destination IP address when connecting to the Advisor
 - Must match the IP address specified in the `lb_connection_v4` or `lb_connection_v6` statement

- **advisor_id_list**
 - The list of IP addresses used by the Agent as the destination IP address when connecting to the Advisor (Agent loops through IP addresses in the list until it successfully connects to an Advisor)
 - Used when each Advisor listens for Agent connections on different IP addresses
 - Must match the IP addresses specified in the `lb_connection_v4` or `lb_connection_v6` statement on each Advisor

- **host_connection**
 - The IP address used by the Agent as the source address when connecting to the Advisor
 - Must match one of the IP addresses specified in the `agent_id_list` statement