IBM Systems – Enterprise Networking Solutions

Enabling Continuous Availability with IBM Multi-site Workload Lifeline

SHARE Session 17086

Lin Overby – overbylh@us.ibm.com
IBM Raleigh, NC

Disclaimer

IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion. Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchase decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code, or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

Agenda

- Disaster Recovery and Continuous Availability
- Do I need continuous availability for my workloads?
- Steps to achieving continuous availability
  - Database replication
  - Multi-site Workload Lifeline
  - GDPS Active-Active Sites

Multi-site Workload Lifeline

⇨ ***Disaster Recovery and Continuous Availability***

Do I need continuous availability for my workloads?

Steps for achieving continuous availability

- Database replication
- Multi-site Workload Lifeline
- GDPS Active-Active Sites

Business Continuity Definitions

- Disaster recovery (DR)
  - The reconstruction of the current physical production site in an alternate physical site, following the loss of the production site.
  - The process of bringing up systems and applications, utilizing the disk-replicated data, in order to support the business from the alternate site.
- Continuous Availability (CA)
  - The duplication of the current primary environment in order to restore the business following the loss of the primary site with minimal downtime.
- Recovery Point Objective (RPO)
  - Amount of data loss following a loss of the primary site when the business is restored.
- Recovery Time Objective (RTO)
  - Duration of time following a loss of the primary site until the business is restored.

Existing DR/CA solutions at metro distances

- GDPS/PPRC (Metro Mirror), based upon a multi-site Parallel Sysplex and synchronous disk replication, is a metro area Continuous Availability (CA), Disaster Recovery (DR) solution
  - Workloads can withstand site and/or storage failures
- Low recovery time and zero data loss
- **Issue: This GDPS product does not provide enough site separation for some enterprises**

Existing DR solutions at global distances

- GDPS/XRC and GDPS/GM, based upon separate sysplexes and asynchronous disk replication, are unlimited distance Disaster Recovery (DR) solutions
  - Disaster recovery for metro-region interruptions
  - Longer recovery time with “seconds” of data loss
- These GDPS products require the failed site’s workload to be restarted in the recovery site and this typically will take an hour or longer
- **Issue: These GDPS products will not achieve a recovery time of seconds being requested by some enterprises**

Continuous Availability at any distance

- 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
  - Allows customers to pursue a gradual, incremental approach to continuous availability that focuses on the most critical workloads first
- Can be used for both unplanned and planned workload outages
- Paradigm shift: failover model => near continuous availability model
  - For critical workloads requiring continuous availability
  - Not a replacement for disaster recovery of non-critical workloads

What is a continuous availability workload?

- 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) or SNA application resource names (e.g., NETA.APPL1)

Multi-site Workload Lifeline

- Disaster Recovery and Continuous Availability
- ⇨ ***Do I need continuous availability for my workloads?***
- Steps for achieving continuous availability
  - Database replication
  - Multi-site Workload Lifeline
  - GDPS Active-Active Sites

How much interruption can your business tolerate?

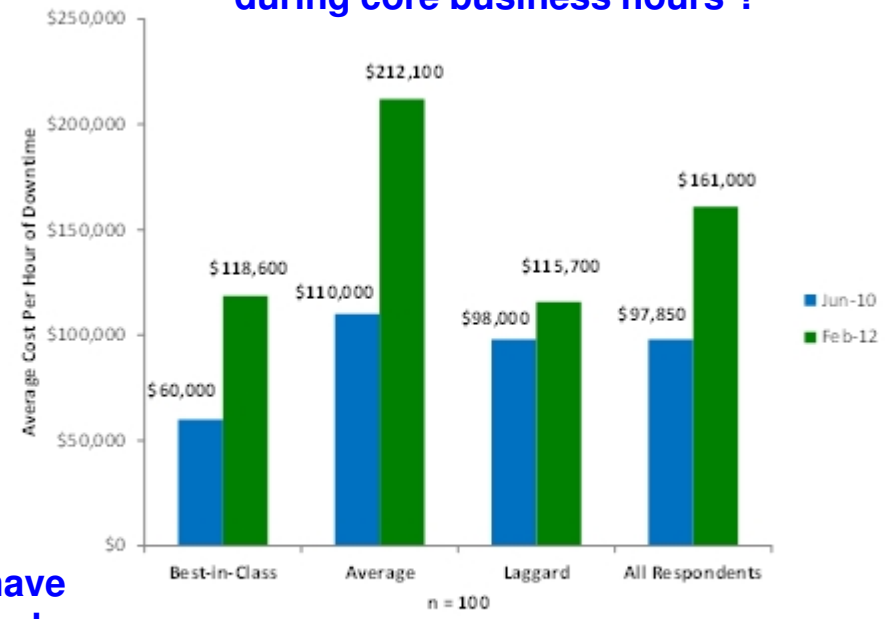
Standby

Active/Active

Ensuring Business Continuity:

- Disaster Recovery
  - Restore business after an unplanned outage
- High Availability
  - Meet Service Availability objectives, e.g., 99.9% availability or 8.8 hours of down-time a year
- Continuous Availability
  - No downtime (planned or not)

What is the cost of 1 hour of downtime during core business hours ?



Enterprises that operate across time-zones no longer have any 'off-hours' window, continuous availability is required

Yearly Cost Metrics	Best-in-Class	Industry Average	Laggards
Business interruption events	.3	2.3	4.4
Time per business interruption event (hours)	.1	1	9
Total disruption (hours)	.03	2.3	39.6
Average cost per hour of disruption	$101,600	$181,770	$99,150
Total cost of business interruption events	**$3,048**	**$418,071**	**$3,926,340**

Source: Aberdeen Group, February 2012

Source: Aberdeen Group, February 2012

Disruptions affect more than the bottom line …

August 18, 2013
Google total eclipse sees 40 percent drop in Internet traffic

August 22, 2013
Nasdaq: 'Connectivity issue' led to three-hour shutdown

July 20, 2013
DMV Computers Fail Statewide, Police Can't Access Database

April 16, 2013
American Airlines Grounds Flights Nationwide

… with enormous impact on the business

- Downtime costs can equal up to 16 percent of revenue [1]
- 4 hours of downtime severely damaging for 32 percent of organizations [2]
- Data is growing at explosive rates – growing from 161EB in 2007 to 988EB in 2010 [3]
- Some industries fined for downtime and inability to meet regulatory compliance
- Downtime ranges from 300–1,200 hours per year, depending on industry [1]

1 Infonetics Research, The Costs of Enterprise Downtime: North American Vertical Markets 2005, Rob Dearborn and others, January 2005
2 Continuity Central, "Business Continuity Unwrapped," 2006, http://www.continuitycentral.com/feature0358.htm
3 The Expanding Digital Universe: A Forecast of Worldwide Information Growth Through 2010, IDC white paper #206171, March 2007

Evolving customer requirements

- Shift focus from failover model to ***near-continuous availability*** model (RTO near zero)
- Access data from ***any site*** (unlimited distance between sites)
- Provide ***workload distribution between sites*** (route around failed sites, dynamically select sites based on ability of site to handle additional workload)
- Provide ***application level granularity***
  - Some workloads may require immediate access from every site, other workloads may only need to update other sites every 24 hours (less critical data)
  - Current solutions employ an all-or-nothing approach (complete disk mirroring, requiring extra network capacity)
- Ensure successful recovery via ***automated processes*** (similar to GDPS technology today)
  - Can be handled by less-skilled operators

Multi-site Workload Lifeline

Disaster Recovery and Continuous Availability

Do I need continuous availability for my workloads?

⇨ ***Steps for achieving continuous availability***

- ***Database replication***
- Multi-site Workload Lifeline
- GDPS Active-Active Sites

Database Replication

- What is data replication?
  - A solution for replicating transactions between databases, typically residing in different sites
  - Emphasizes the replicating 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 (for example, a subset of tables)
- Why use data replication?
  - Offload read-only transactions to replicated database
    - Read-only database provides near-real time reporting
  - Continuous (High) Availability
    - Failover to replicated database during disaster recovery

Database Products

- InfoSphere Data Replication for DB2 for z/OS
  - A solution for replicating DB2 table or database updates
- InfoSphere Data Replication for IMS for z/OS
  - A solution for replicating IMS database updates
- InfoSphere Data Replication for VSAM for z/OS
  - A solution for replicating VSAM file updates

Multi-site Workload Lifeline

Disaster Recovery and Continuous Availability

Do I need continuous availability for my workloads?

⇨ ***Steps for achieving continuous availability***

- Database replication
- ***Multi-site Workload Lifeline***
- GDPS Active-Active Sites

What is Multi-site Workload Lifeline?

- Lifeline is a product that enables very high availability with load balancing and workload rerouting

- Lifeline plays a key role in solving 2 major problems in the Enterprise
  - Providing continuous availability for critical workloads during unplanned outages
  - Reducing downtime for planned outages

- Lifeline solution requires:
  - Data replication products
  - SASP capable load balancer

Multi-site Workload Lifeline Benefits …

Providing continuous availability for critical workloads

- Provides intelligent load balancing advice for TCP/IP workloads across two sites at unlimited distances to provide nearly continuous availability
  - Increases availability: New connections are routed away from failing applications, systems and sites in the event of outage.
  - Increases performance: Reduces response time by routing new connections to applications and systems with most capacity for additional work
  - Allows scalability: Additional application instances can be added on demand
  - Improves recovery time: Reduces recovery time from hours to minutes
- Improves analytic capability: Network Management Interface (NMI) provides access to workload, application and site status

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, non-disruptive testing of disaster recovery procedures by validating workloads remain accessible on the recovery site without requiring a site outage on the production site

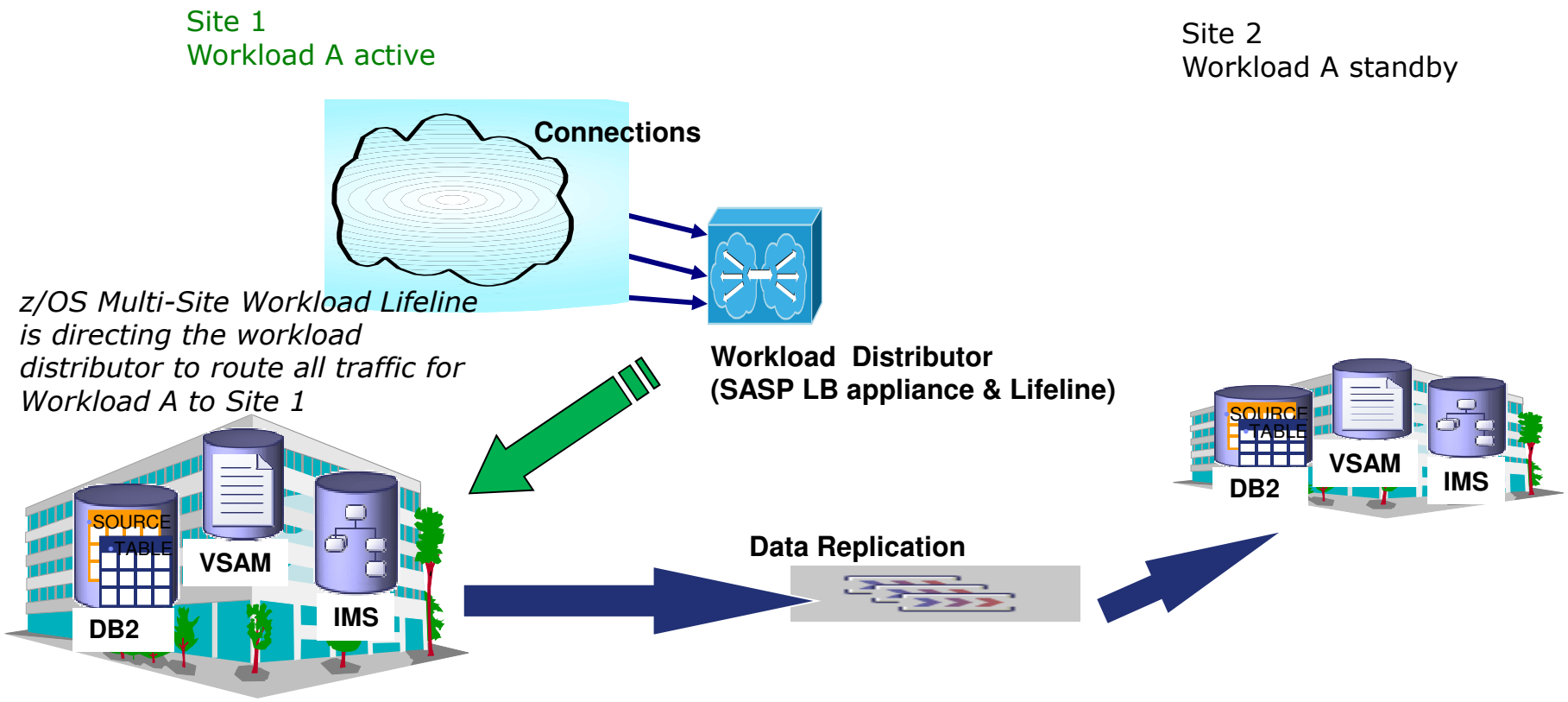
What makes Multi-site Workload Lifeline different?

- Lifeline is not an all-or-nothing solution
  - Lifeline helps allow customers to pursue a gradual, incremental approach to continuous availability that focuses on the most critical workloads first
- Lifeline supports routing workloads to the alternate site, reducing the strain on the primary transaction system and allowing organizations to get more value from their secondary site investment
- Lifeline requires no configuration changes to:
  - Server Applications
  - Clients
  - The network topology

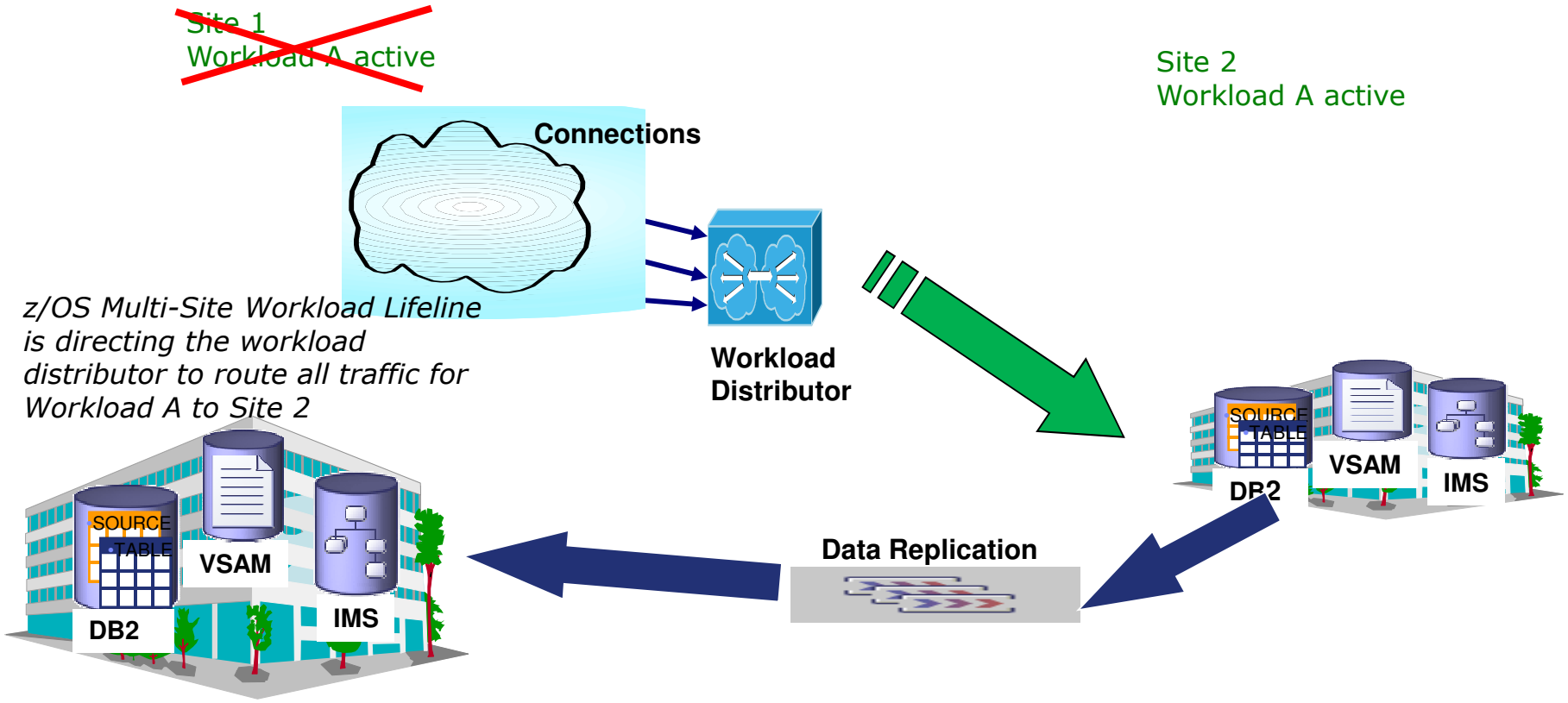
Continuous Availability Configurations

- Configurations
  - Active/Standby
  - Active/Query
- Configuration is specified on a workload basis
- Supported workloads
  - TCP/IP workloads
  - Linux on System z workloads
  - SNA workloads

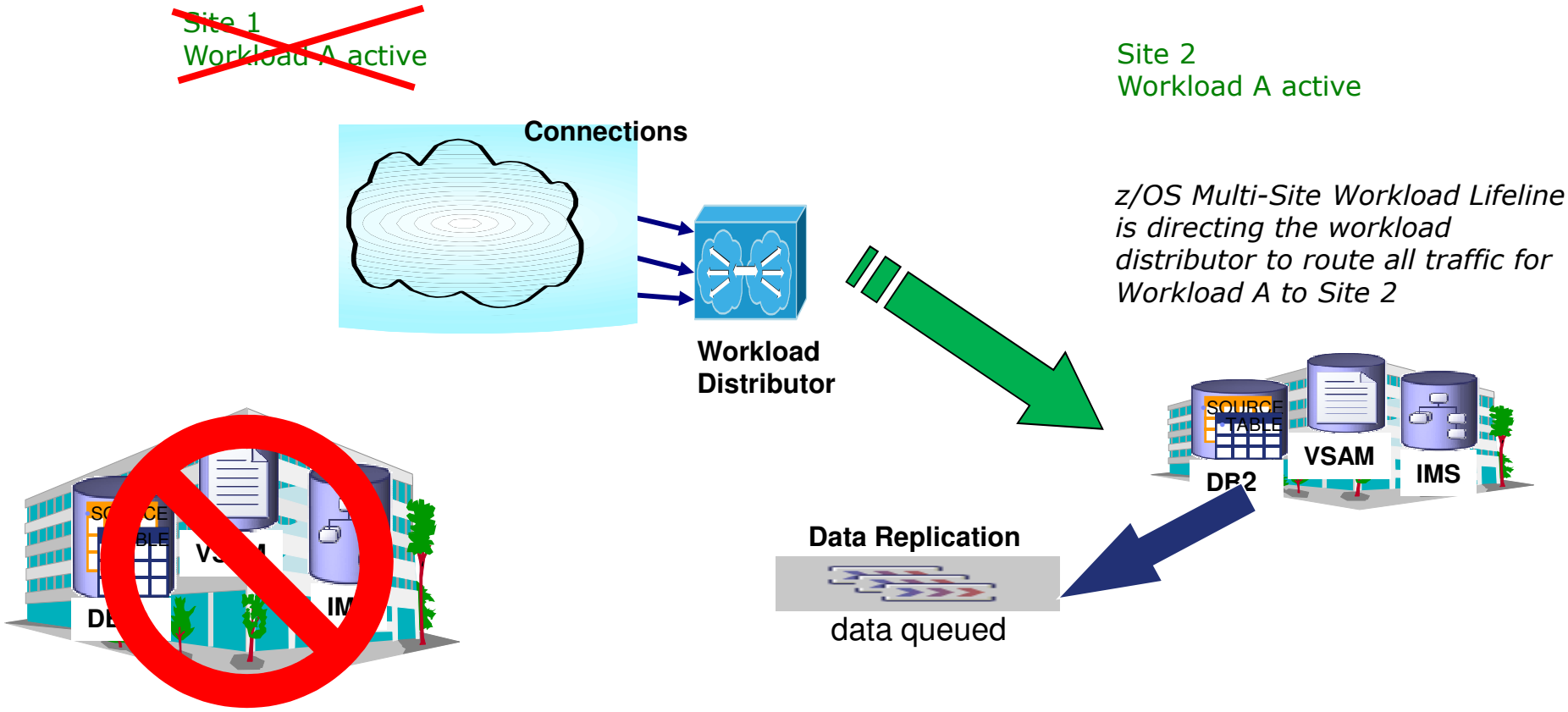
Scenario 1: Active/Standby Configuration – Prior to workload outage



Scenario 1: Active/Standby Configuration
Workload outage / Site available



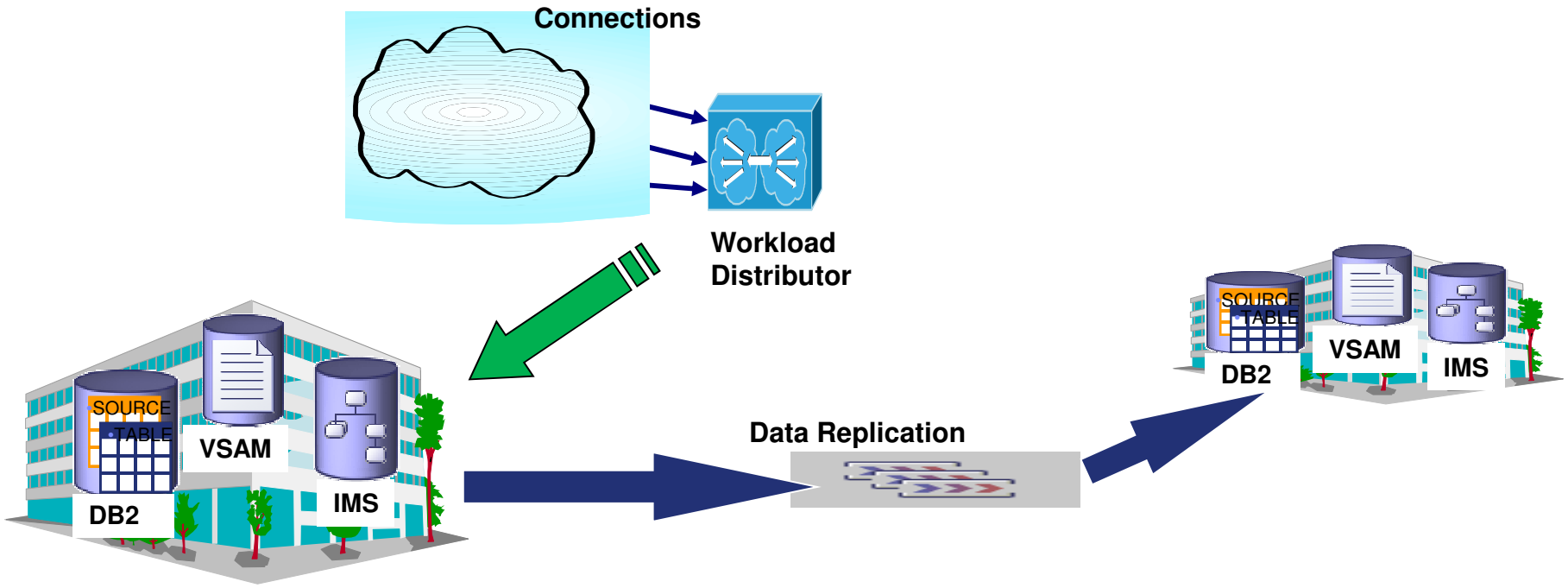
Scenario 1: Active/Standby Configuration
Workload outage / Site outage



Scenario 2: Active/Standby Configuration – Part 1 (multiple workloads – mutual continuous availability)

Site 1
Workload A active

Site 2
Workload A standby



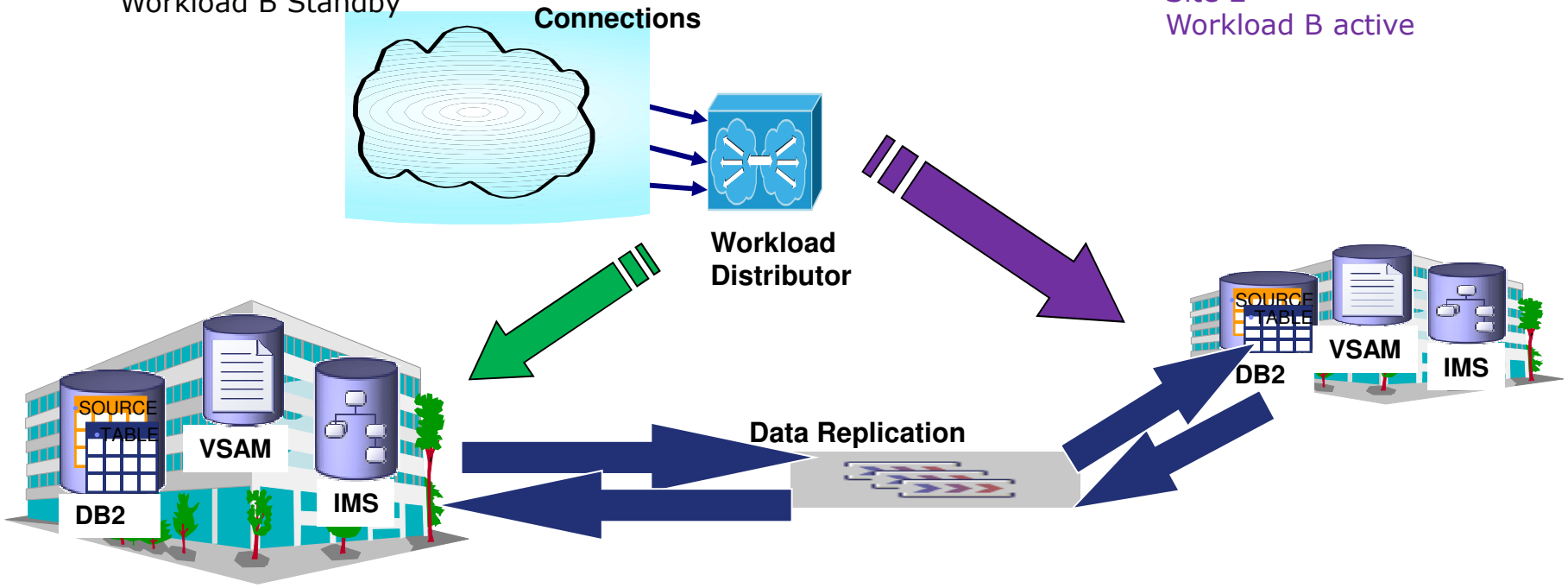
Scenario 2: Active/Standby Configuration – Part 2 (multiple workloads – mutual continuous availability)

Site 1
Workload A active

Site 1
Workload B Standby

Site 2
Workload A standby

Site 2
Workload B active



Active/Query Workloads

- An Active/Standby workload is active on only one site
  - Workload transactions update data on the active site
  - Database changes are replicated to the standby site
- An associated Active/Query workload can be active on both sites
  - Workload transactions access same data being updated by Active/Standby workload
  - Workload transactions only query data
- Active/Query workload connections are distributed to a site based on routing type and average replication latency
  - Dynamic workload routing – distribution between sites based on availability and health of server applications within each site
  - Static workload routing – distribution between sites based on a configured percentage
  - Replication latency – Average delay between when update transactions for a workload to the active site are replicated and applied to the standby site

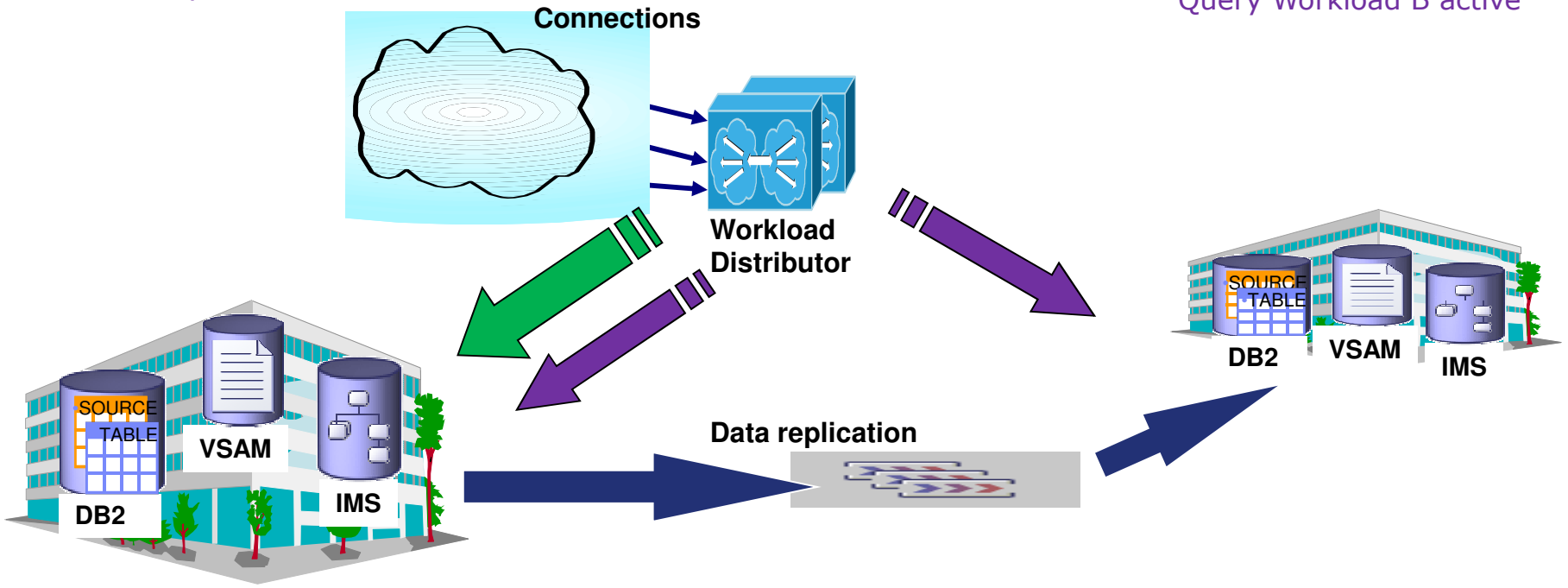
Scenario 3: Active/Query Configuration – Distribution of query workload across two sites

Site 1
Update Workload A active
Replication Latency: 'Normal'

Site 1
Query Workload B active

Site 2
Update Workload A standby

Site 2
Query Workload B active



Replication latency, site capacity, and application health influence the distribution of query workload between sites.

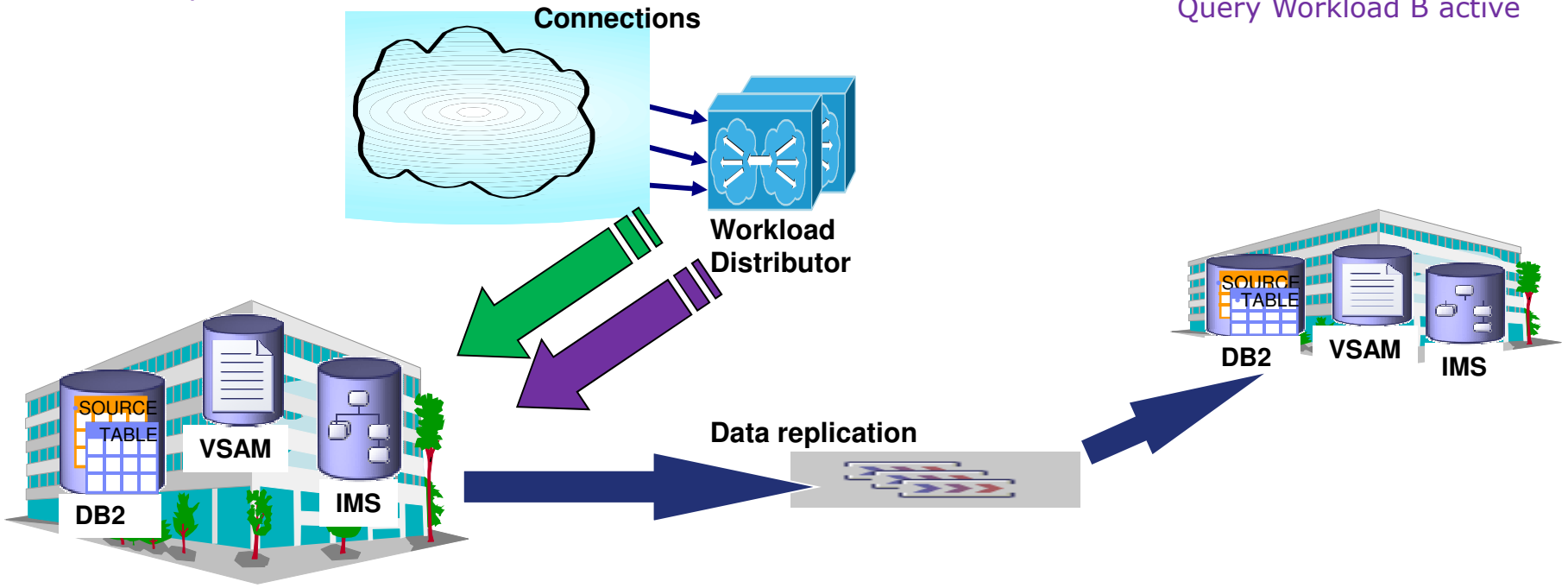
Scenario 3: Active/Query Configuration –
Replication latency rises above maximum configured causing all query workload to be routed to site 1

Site 1
Update Workload A active
Replication Latency: 'Acute'

Site 1
Query Workload B active

Site 2
Update Workload A standby

Site 2
Query Workload B active



Replication latency, site capacity, and application health influence the distribution of query workload between sites.

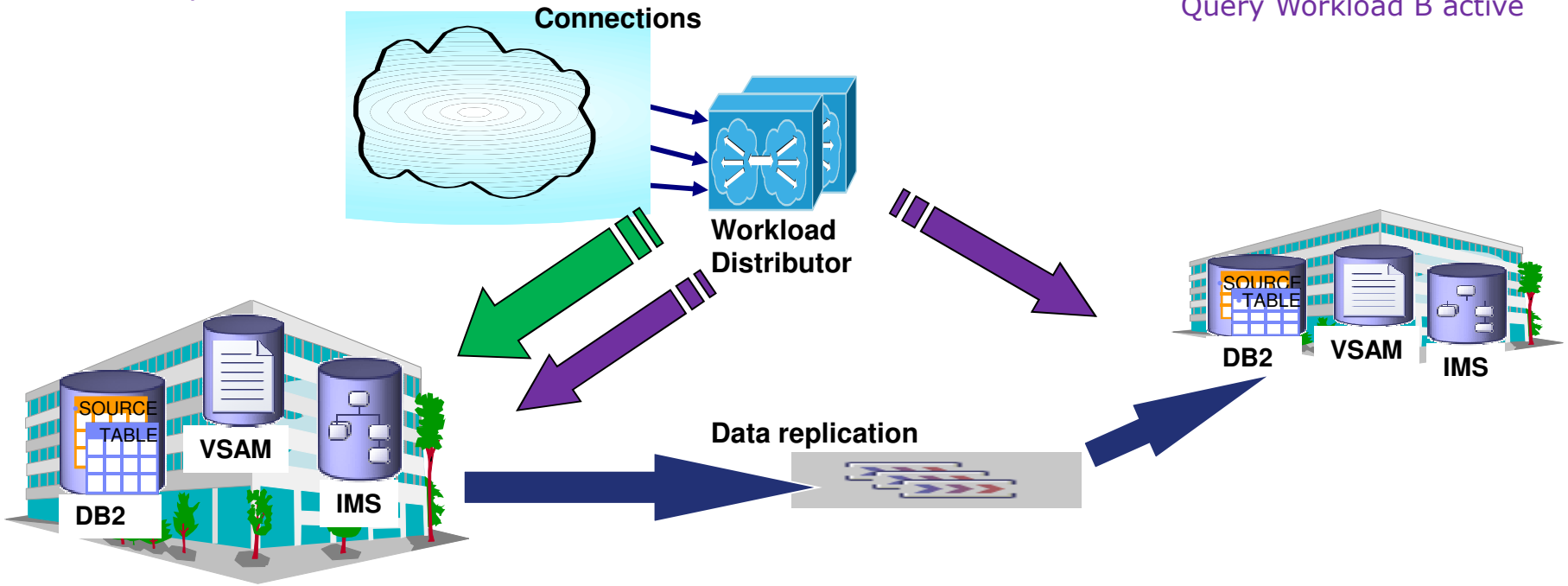
Scenario 3: Active/Query Configuration –
Replication latency falls below minimum configured causing query workload to be distributed across two sites

Site 1
Update Workload A active
Replication Latency: 'Normal'

Site 1
Query Workload B active

Site 2
Update Workload A standby

Site 2
Query Workload B active

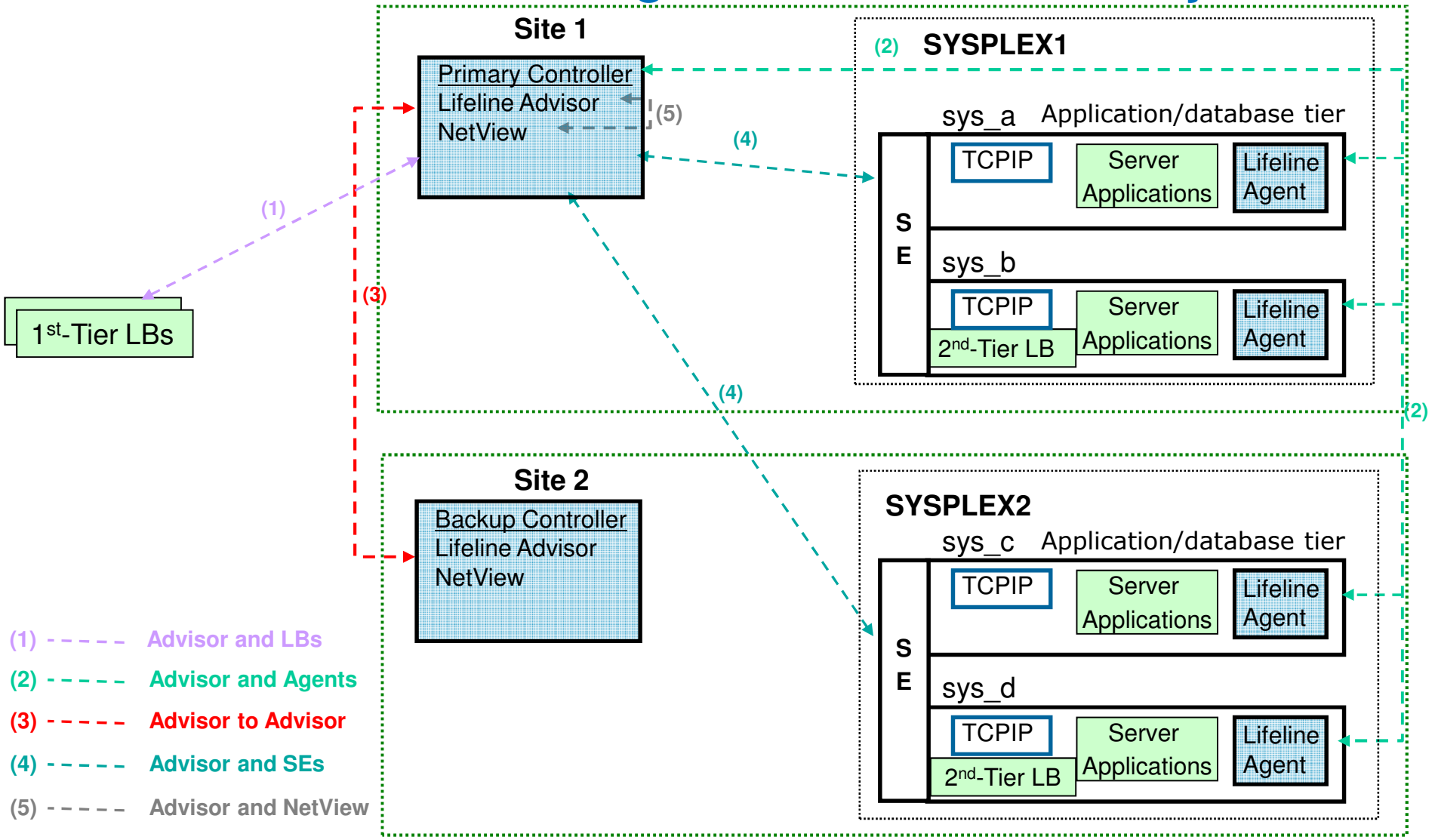


Replication latency, site capacity, and application health influence the distribution of query workload between sites.

Lifeline support for planned and unplanned outages

- 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 perform automatic takeover or prompt for action
- Ability to switch workloads from one site to another site
  - Perform “graceful” takeover for planned outages
  - Perform “graceful” failback following a workload or site disaster
- Ability to maintain workload configuration states in event of a workload manager failure or planned outage
  - 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

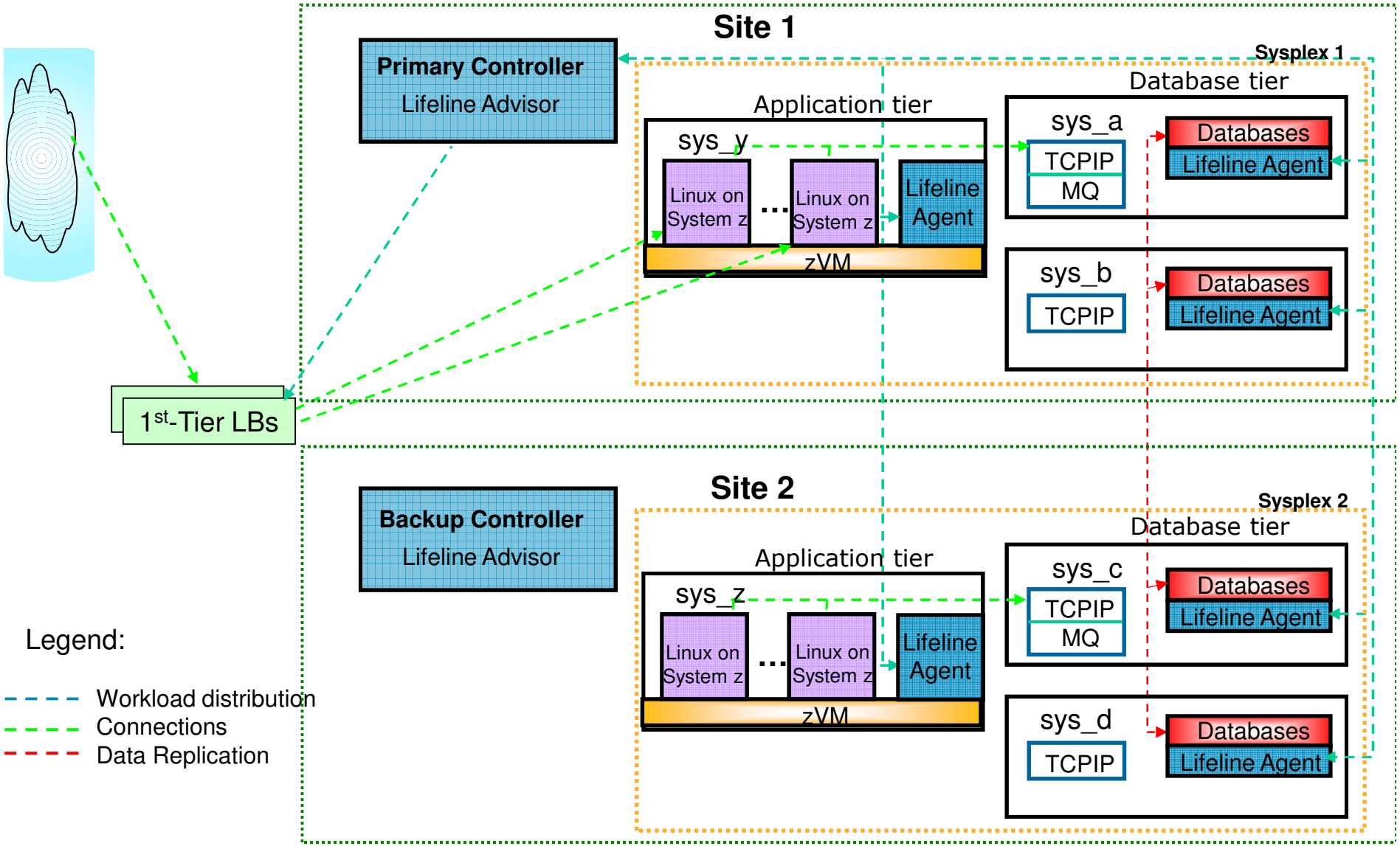
Workload Lifeline Providing Continuous Availability



Linux on System z workloads

- Provide support for workloads that have the application-tier hosted by Linux on System z and data-tier hosted on z/OS
  - End-to-End workload support
- Failure of application-tier results in failure of workload
  - Independent of status/availability of backend z/OS images in site

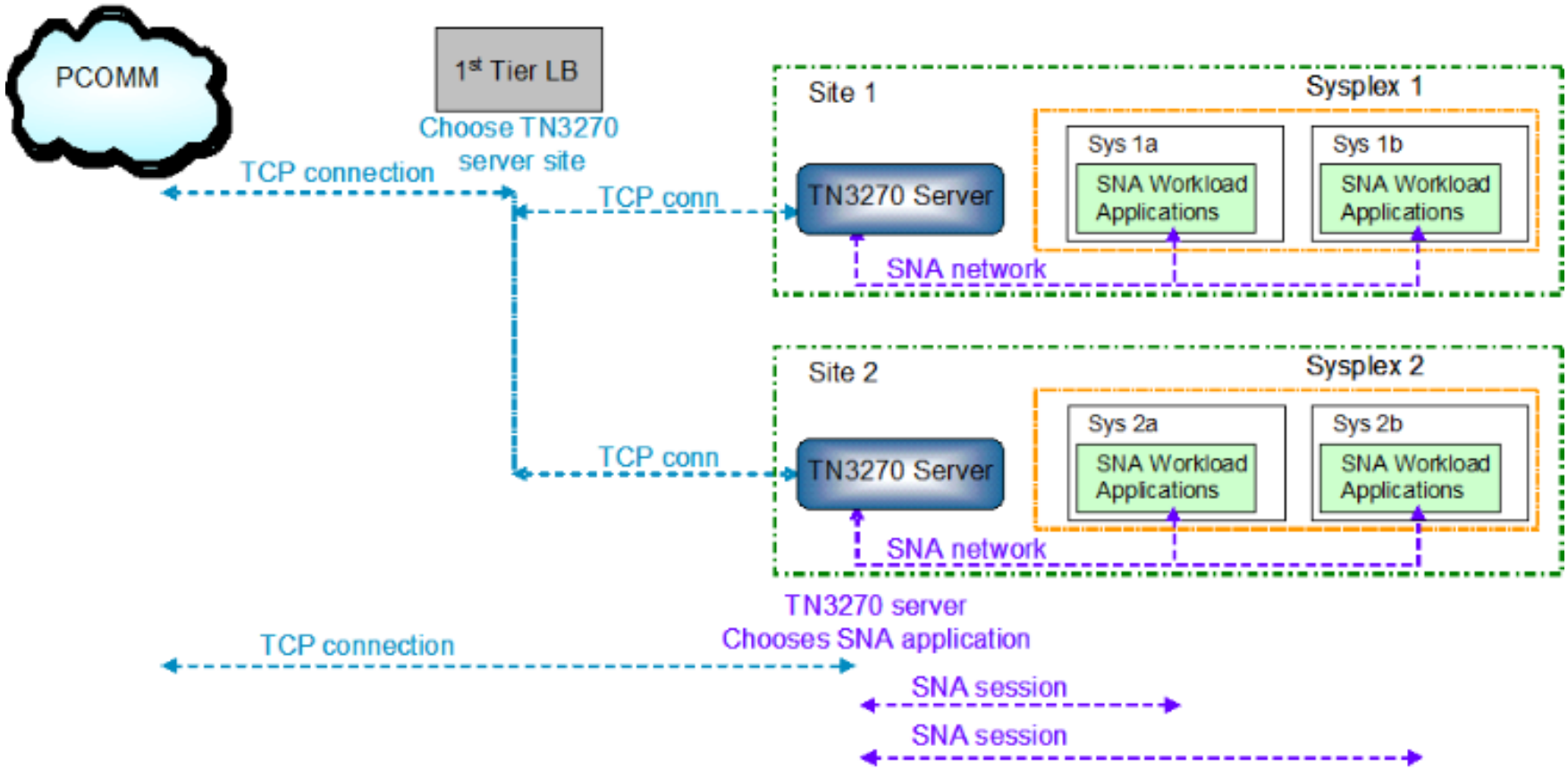
Linux on System z environment



SNA-based workloads

- Provide support for workloads with SNA applications that use IP connectivity from the clients
  - Excludes direct client to data center connections over Enterprise Extender
- Lifeline Agent monitors the SNA application
  - Provides similar LPAR health information as TCP applications
- Support for different IP/SNA environments
  - TN3270 access to SNA applications
  - Remote API access to SNA applications
  - Message Broker access to SNA applications
  - Customer-written gateway access to SNA applications

SNA-based workload environment



Multi-site Workload Lifeline

Disaster Recovery and Continuous Availability

Do I need continuous availability for my workloads?

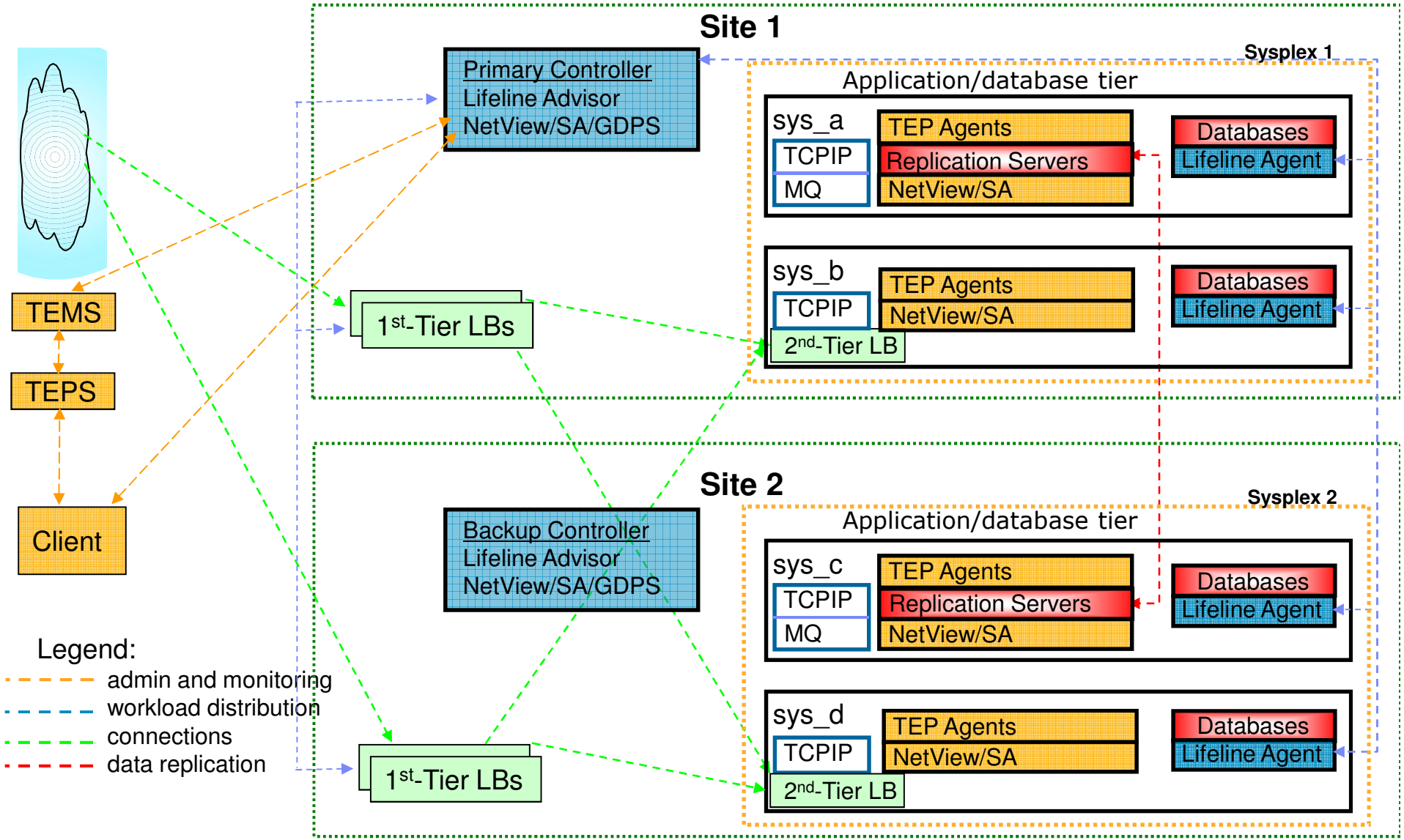
⇨ ***Steps for achieving continuous availability***

- Database replication
- Multi-site Workload Lifeline
- ***GDPS Active-Active Sites***

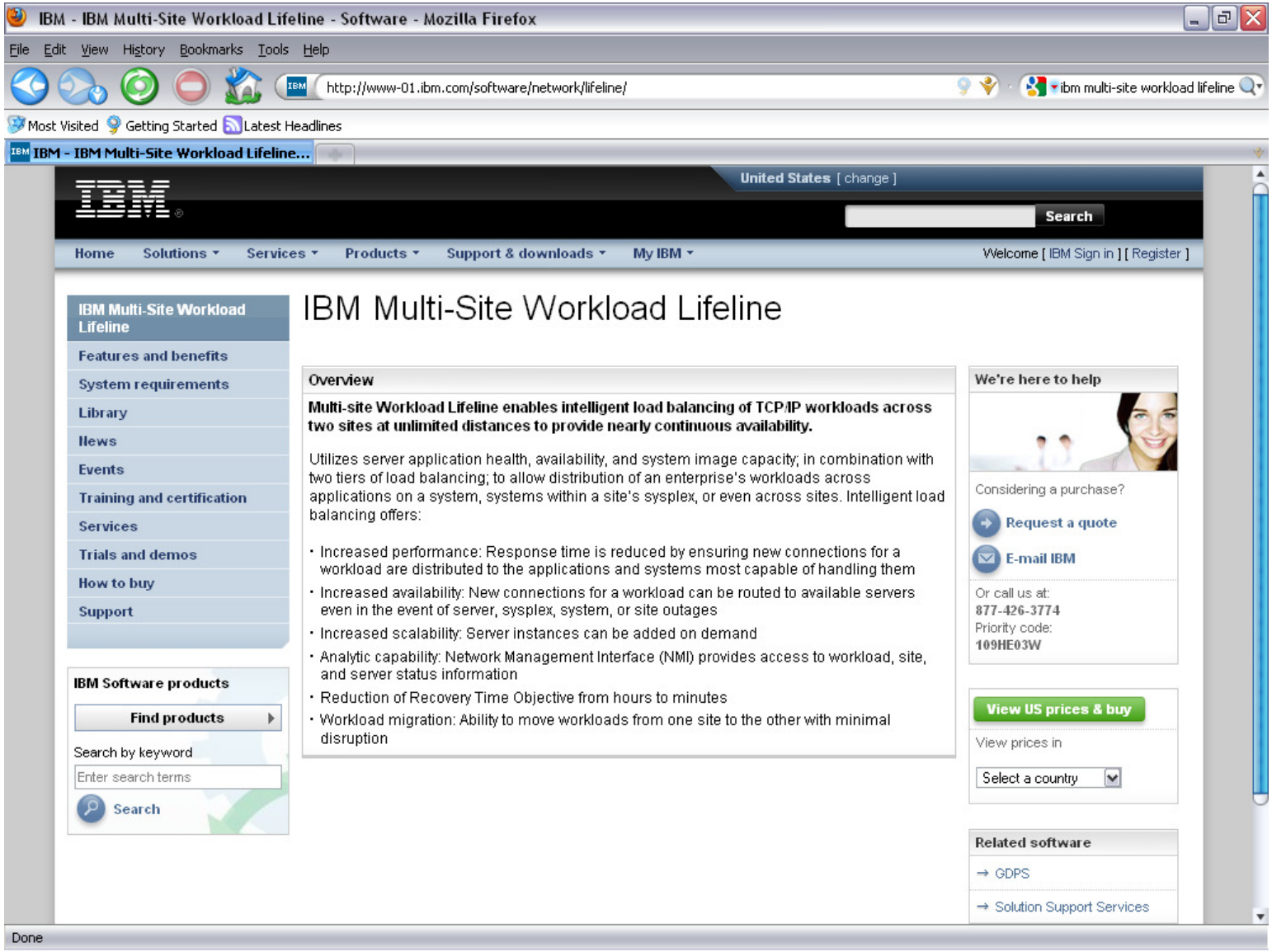
GDPS Active-Active Sites

- The complete solution for providing continuous availability at extended distances for your business critical workloads
- In addition to the database replication product and Lifeline, the following products are required
  - GDPS/Active-Active
  - Tivoli NetView Monitoring for z/OS (includes NetView)
  - System Automation for z/OS
  - Optionally, the Tivoli OMEGAMON XE suite of monitoring products
- GDPS Active-Active Sites provides
  - Automation for managing and switching workloads for planned and unplanned outages
  - Detailed monitoring of the components making up the GDPS Active-Actice Sites solution

GDPS Active/Active Sites Structure



For more information… www-01.ibm.com/software/network/lifeline



Please fill out your session evaluation

Enabling Continuous Availability with IBM Multi-site Workload Lifeline

Session # 17086

QR Code: