

GDPS Active/Active Sites Update

Steven Cook
IBM Corporation

August 8, 2012
Session Number 11663

Trademarks

The following are trademarks of the International Business Machines Corporation in the United States and/or other countries.

IBM*	ESCON*	Redbooks*
IBM (logo)*	FlashCopy*	Sysplex Timer*
ibm.com*	GDPS*	System p*
AIX*	HyperSw ap	System z*
DB2*	IBM*	Tivoli*
DS6000	IBM logo*	z/OS*
DS8000	Parallel Sysplex*	z/VM*
Dynamic Infrastructure*	POWERS	

* Registered trademarks of IBM Corporation

The following are trademarks or registered trademarks of other companies.

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.

Disclaimer

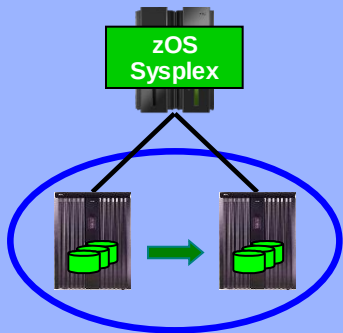
- IBM's statements regarding its plans, directions, and intent are subject to change or withdrawal without notice at IBM's sole discretion.
- Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.
- 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.
- Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including 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 results similar to those stated here

Suite of GDPS service products to meet various business requirements for availability and disaster recovery

Continuous Availability of Data within a Data Center

Single Data Center

Application remain active
Continuous access to data in the event of a storage subsystem outage



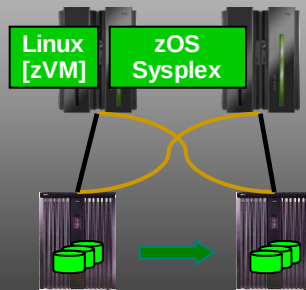
GDPS/PPRC HM

RPO 0 sec & RTO 0 sec

Continuous Availability / Disaster Recovery within a Metropolitan Region

Two Data Centers

Systems remain active
Multi-site workloads can withstand site and/or storage failures



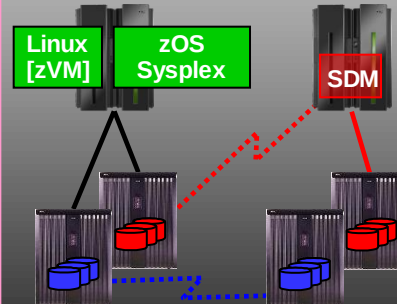
GDPS/PPRC
active/active,
active/standby configs

RPO 0 sec & RTO 1-2 min / <1 hr

Disaster Recovery at Extended Distance

Two Data Centers

Rapid Systems Disaster Recovery with "seconds" of Data Loss
Disaster recovery for out of region interruptions

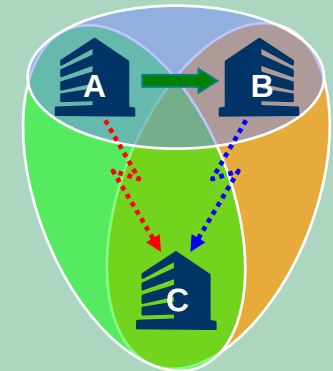


GDPS/GM & GDPS/XRC

RPO few sec & RTO 1hr

Continuous Availability Regionally and Disaster Recovery Extended Distance

Three Data Centers
High availability for site disasters
Disaster recovery for regional disasters



GDPS/MGM & GDPS/MzGM

RPO 0 sec & RTO 1-2 min / <1 hr
RPO few sec & RTO 1 hr

4

Complete your sessions evaluation online at SHARE.org/AnaheimEval

RPO – recovery point objective (data loss)
RTO – recovery time objective (downtime)

Synch replication
Asynch replication



Customer requirements

- **Want to shift focus from a failover model to a nearly-continuous availability model (RTO near zero)**
- **Access data from any site (unlimited distance between sites)**
- **No application changes**
- **Multi-sysplex, multi-platform solution**
 - “Recover my business rather than my platform technology”
- **Ensure successful recovery via automated processes (similar to GDPS technology today)**
- **Provide workload distribution between sites (route around failed sites, dynamically select sites based on ability of site to handle additional workload)**
- **Replace RYO solutions to reduce overall costs**
- **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)

What are GDPS/PPRC customers doing today?

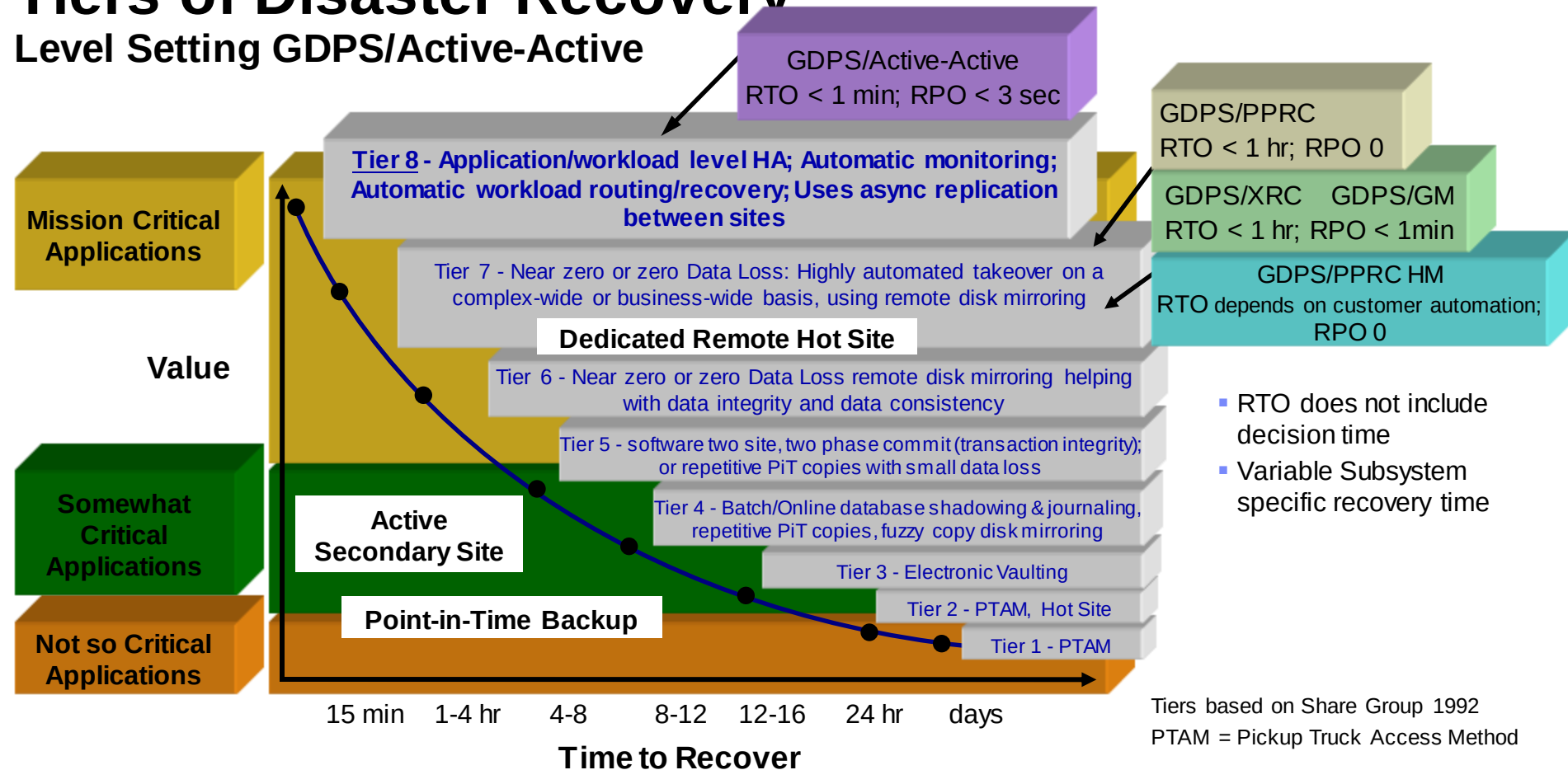
- GDPS/PPRC, based upon a multi-site Parallel Sysplex and synchronous disk replication, is a metro area Continuous Availability (CA) and Disaster Recovery solution (DR)
- GDPS/PPRC supports two configurations:
 - Active/standby or single site workload
 - Active/active or multi-site workload
- **Some customers have deployed GDPS/PPRC active/active configurations**
 - All critical data must be PPRCed and HyperSwap enabled
 - All critical CF structures must be duplexed
 - Applications must be parallel sysplex enabled
 - Signal latency will impact OLTP thru-put and batch duration resulting in the sites being separated by no more than a couple tens of KM (fiber)
- **Issue: the GDPS/PPRC active/active configuration does not provide enough site separation for some enterprises**

What are GDPS/XRC & GDPS/GM customers doing today?

- GDPS/XRC and GDPS/GM, based upon asynchronous disk replication, are unlimited distance DR solutions
- The current GDPS asynchronous replication products require the failed site's workload to be restarted in the recovery site and this typically will take 30-60 min
 - Power fail consistency
 - Transaction consistency
- There are no identified extensions to the existing GDPS asynchronous replication products that will allow the RTO to be substantially reduced
- **Issue: GDPS/XRC and GDPS/GM will not achieve an RTO of seconds being requested by some enterprises**

Tiers of Disaster Recovery

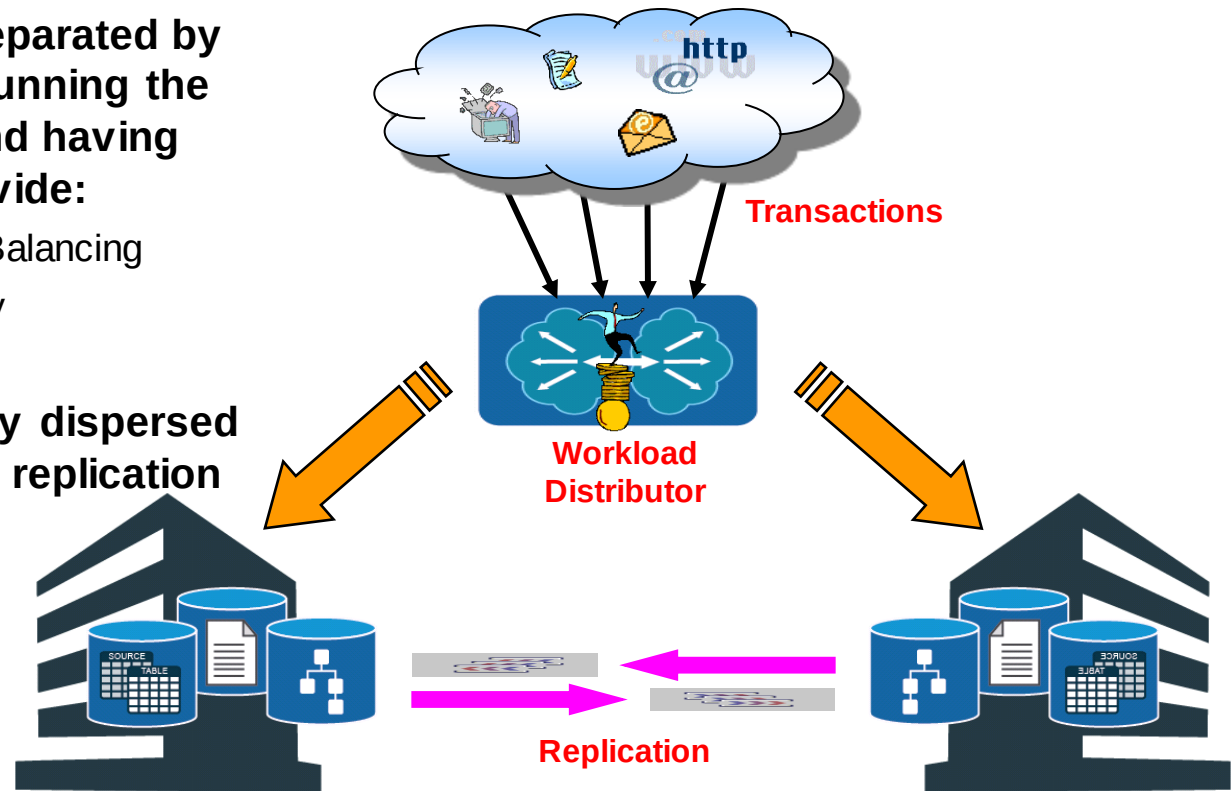
Level Setting GDPS/Active-Active



Failover models can only achieve so much in improving RTO

Active/Active concept

- **Two or more sites, separated by unlimited distances, running the same applications and having the same data to provide:**
 - Cross-site Workload Balancing
 - Continuous Availability
 - Disaster Recovery
- **Data at geographically dispersed sites kept in sync via replication**



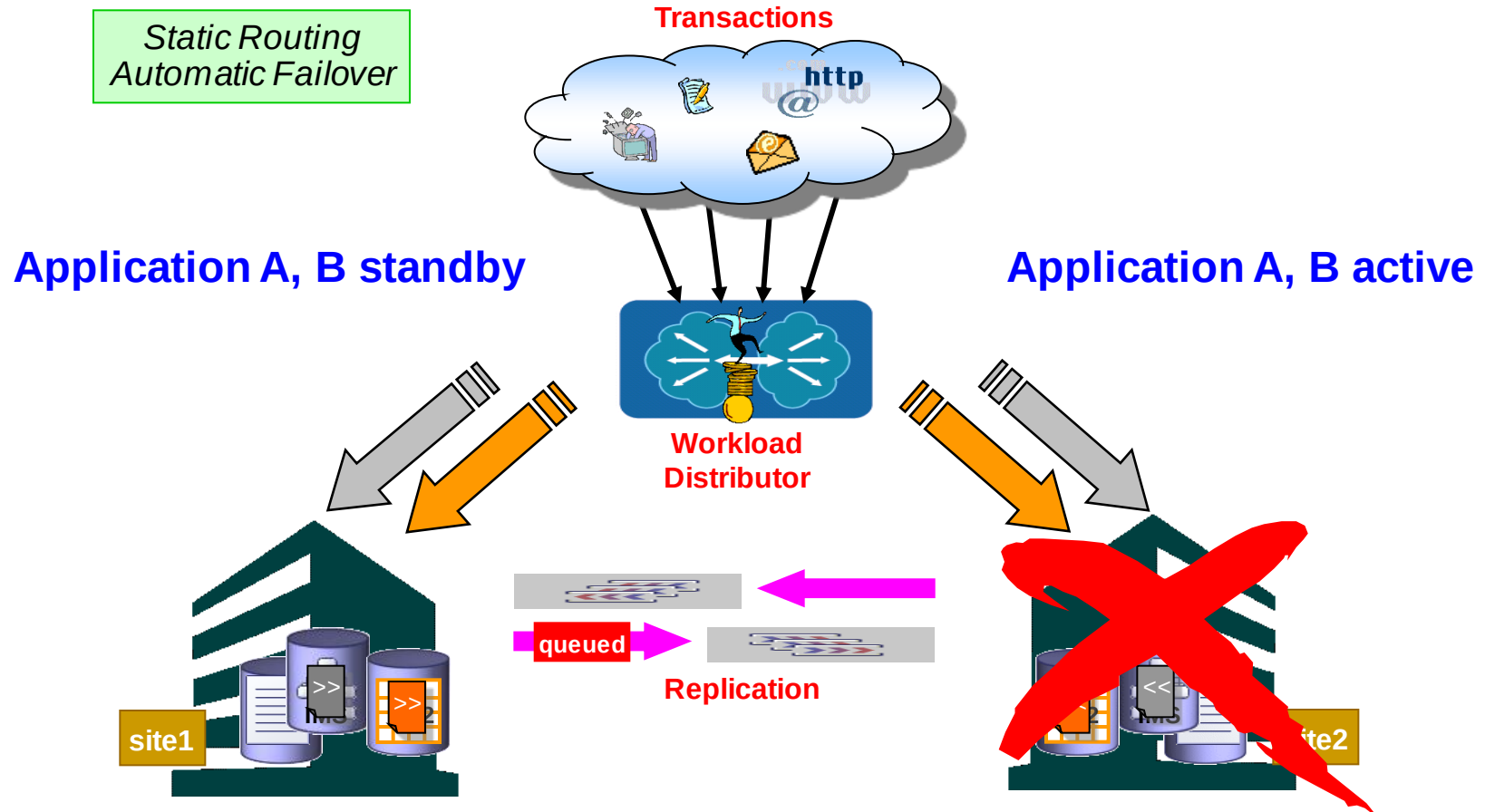
Workloads are managed by a client and routed to one of many replicas, depending upon workload weight and latency constraints; extends workload balancing to SYSPLEXs across multiple sites

Monitoring spans the sites and now becomes an essential element of the solution for site health checks, performance tuning, etc

Active/Active Sites Configurations

- Configurations
 - Active/Standby – GA date 30th June 2011
 - Active/Query – statement of direction
 - Active/Active – intended direction
- A configuration is specified on a workload basis
- A workload is the aggregation of these components
 - **Software:** user written applications (eg: COBOL programs) and the middleware run time environment (eg: CICS regions, InfoSphere Replication Server instances and DB2 subsystems)
 - **Data:** related set of objects that must preserve transactional consistency and optionally referential integrity constraints (eg: DB2 Tables, IMS Databases)
 - **Network connectivity:** one or more TCP/IP addresses & ports (eg: 10.10.10.1:80)

Active/Standby configuration



This is a fundamental paradigm shift from a failover model to a continuous availability model

Active/Query configuration (SOD)

Appl B (grey) is in **active/query** configuration

- using same data as Appl A
- active to both site1 & site2, but favor site1
- routed according to [A] latency policy
- policy for query routing: max latency 5, reset latency 3

Transactions



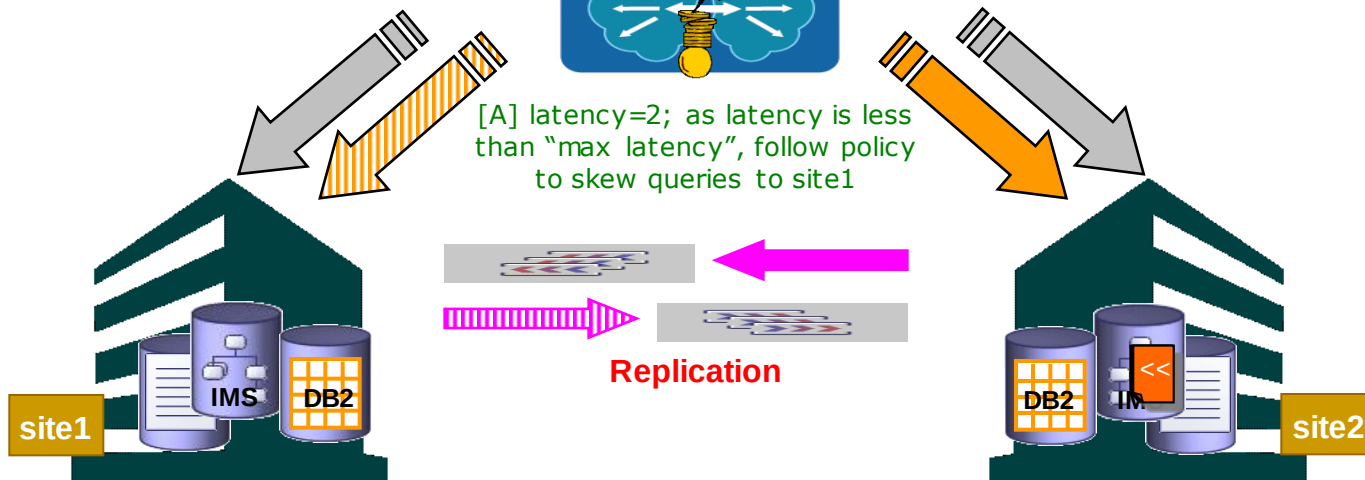
Appl A (gold) is in **active/standby** configuration

- performing updates in active site [site2]

Workload Distributor

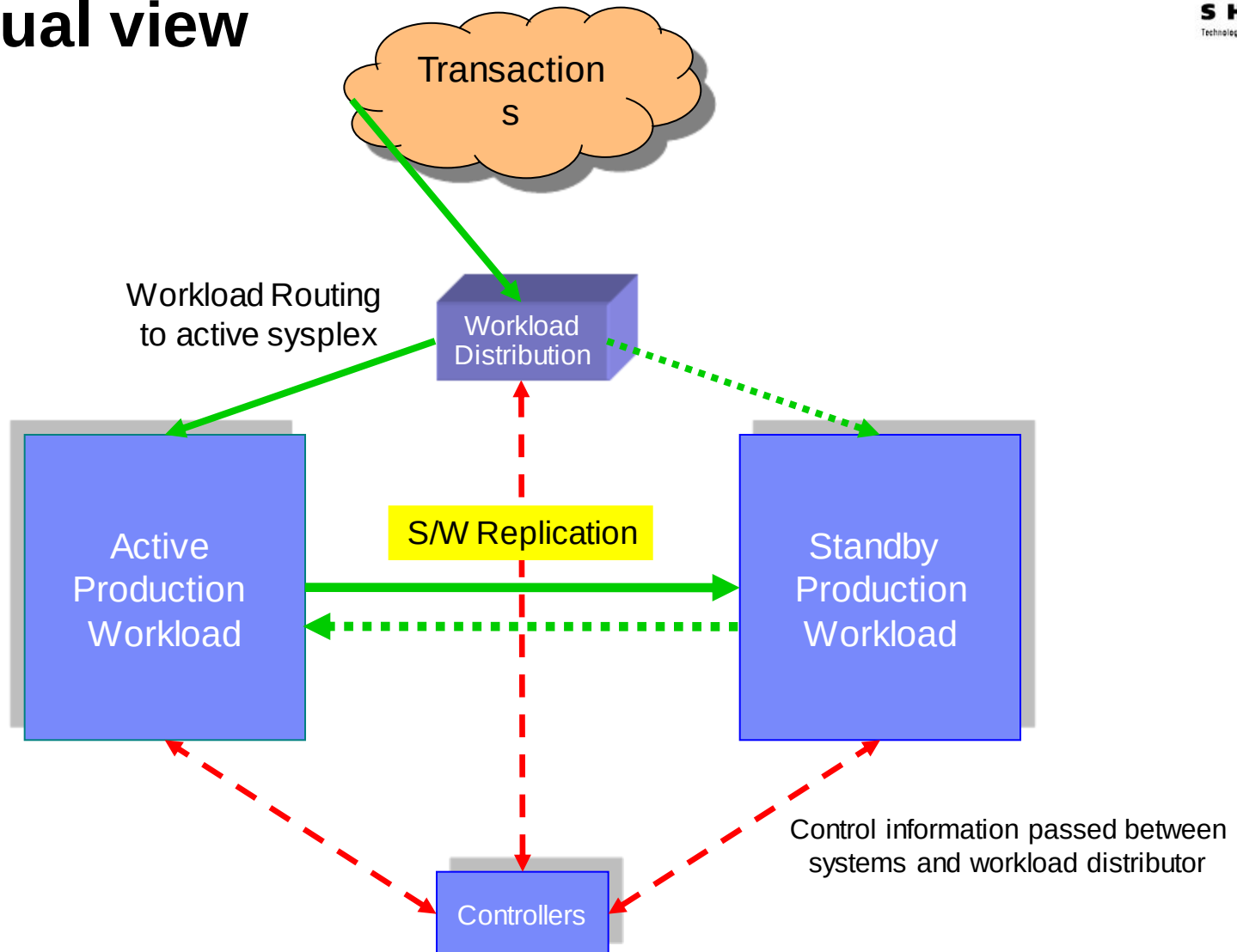


[A] latency=2; as latency is less than "max latency", follow policy to skew queries to site1



Read-only or query transactions to be routed to both sites, while update transactions are routed only to the active site

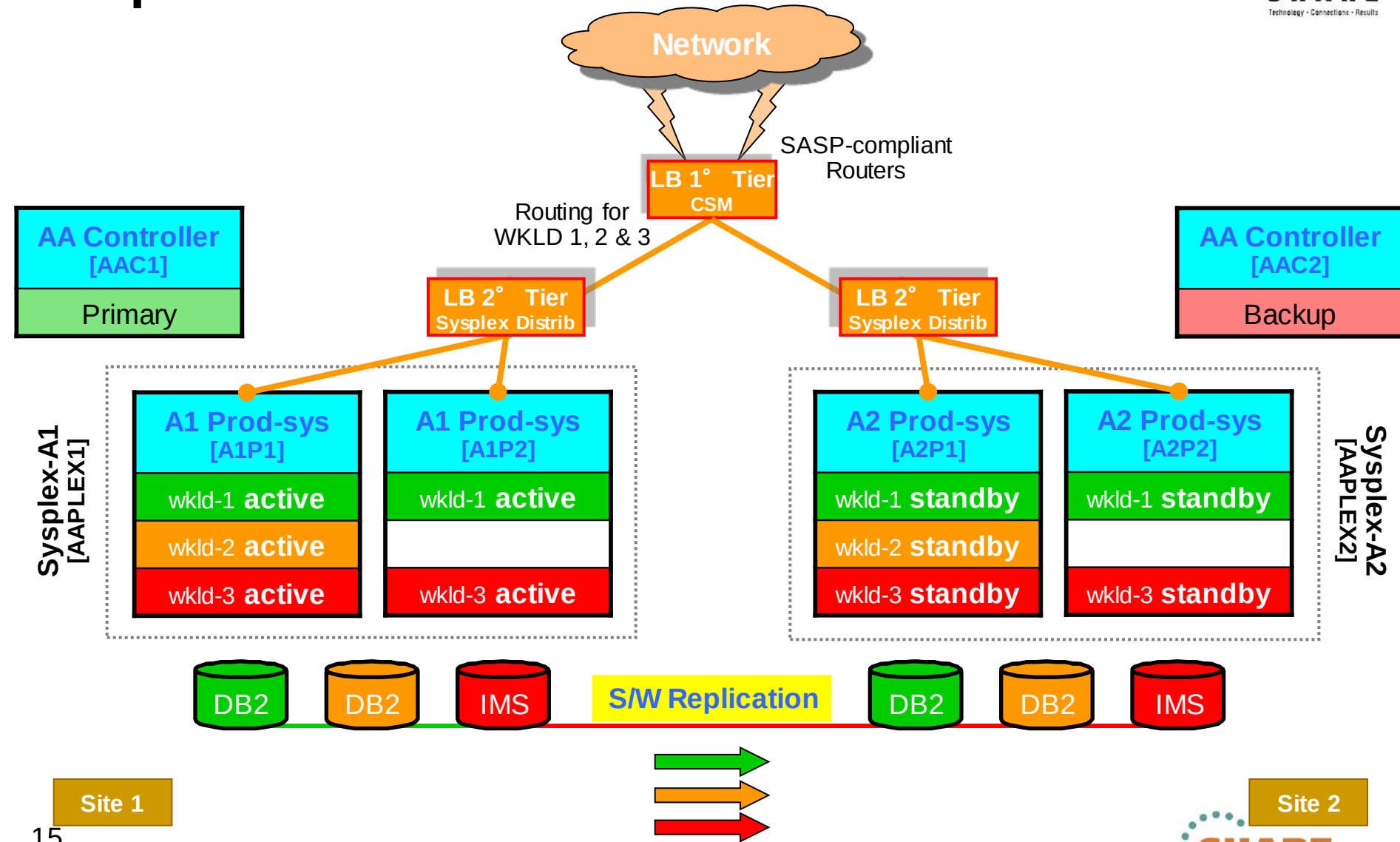
Conceptual view



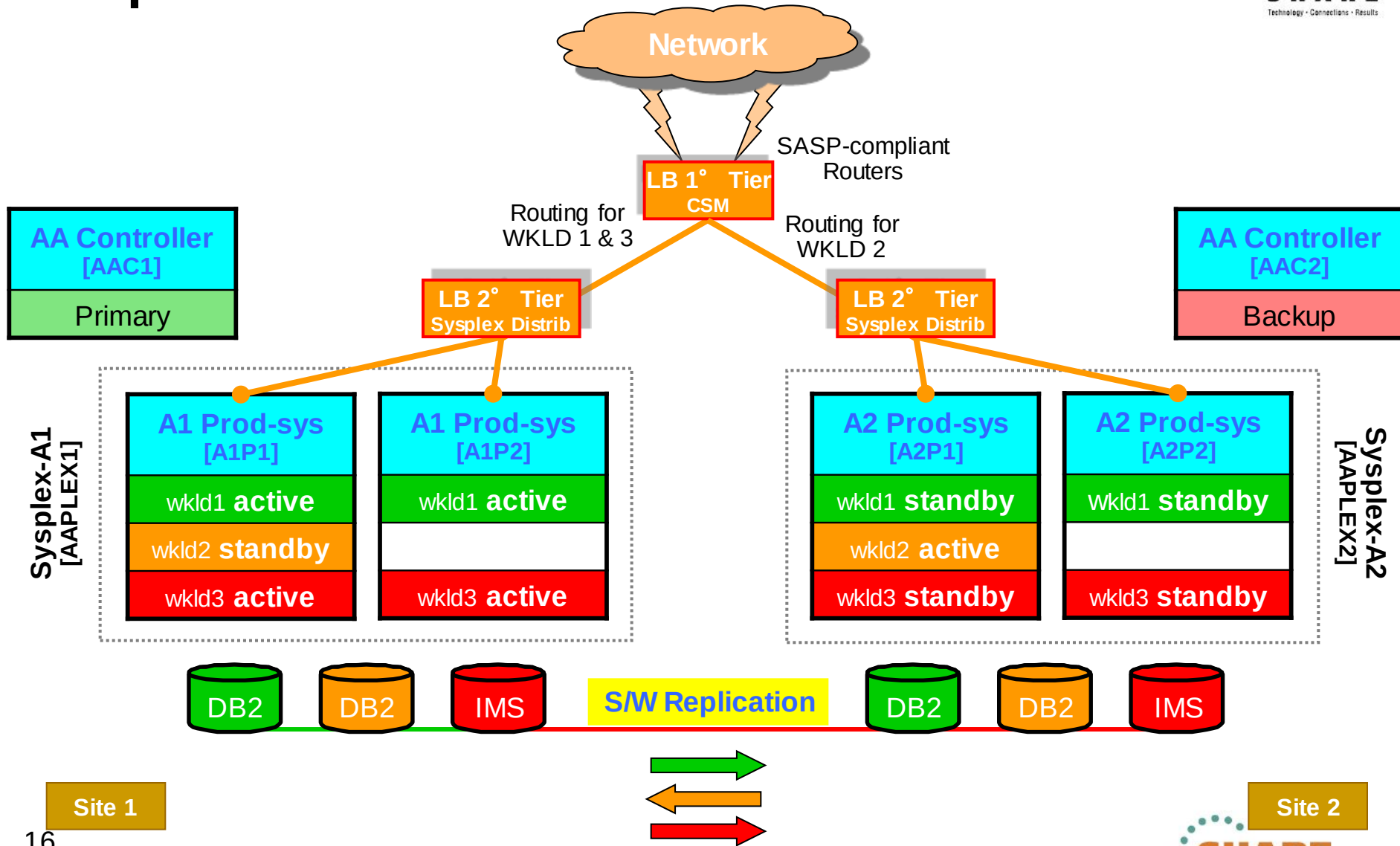
What is a GDPS/Active-Active environment?

- **Two Production Sysplex environments (also referred to as sites) in different locations**
 - One active, one standby – for each defined workload
 - Software-based replication between the two sysplexes/sites
 - IMS and DB2 data is supported
 - VSAM data expected to be supported in the future
- **Two Controller Systems**
 - Primary/Backup
 - Typically one in each of the production locations, but there is no requirement that they are co-located in this way
- **Workload balancing/routing switches**
 - Must be Server/Application State Protocol compliant (SASP)
 - RFC4678 describes SASP
 - **What switches/routers are SASP-compliant?** ... the following are those we know about
 - Cisco Catalyst 6500 Series Switch Content Switching Module
 - F5 Big IP Switch
 - Citrix NetScaler Appliance
 - Radware Alteon Application Switch (bought Nortel appliance line)

Sample scenario – all workloads active in one site



Sample scenario – both sites active for individual workloads



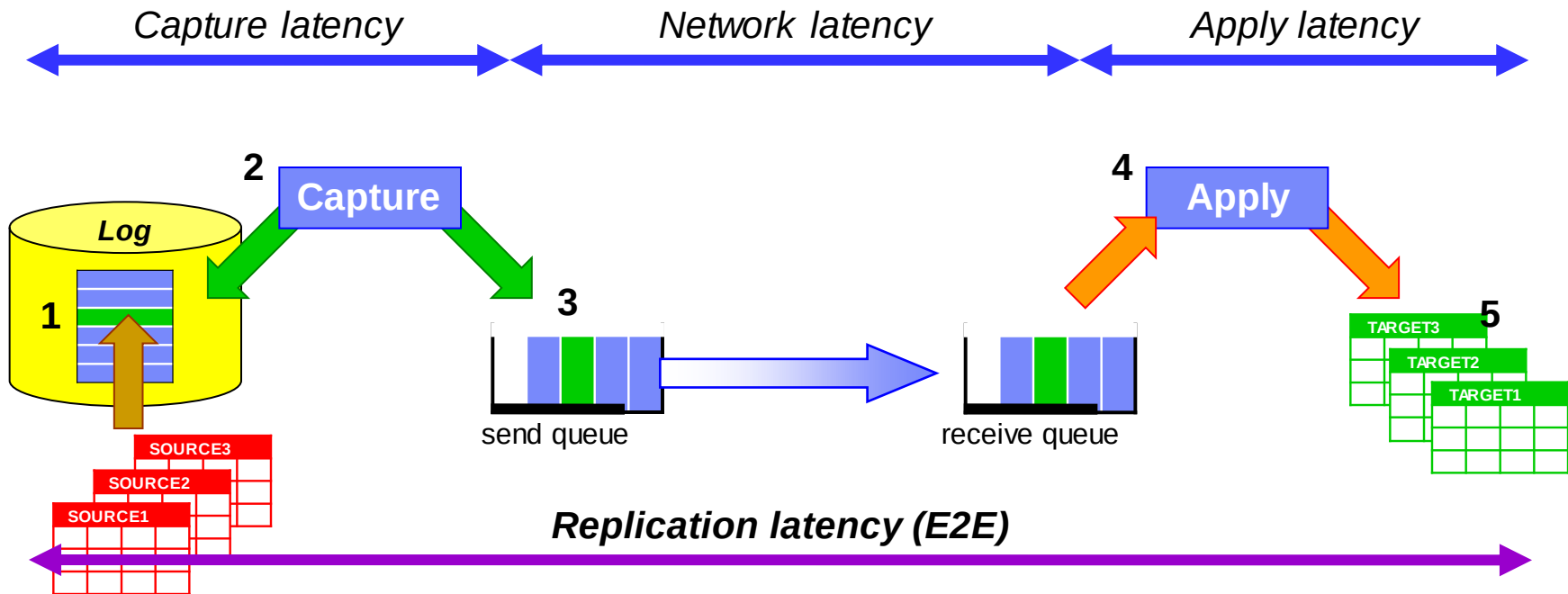
What S/W makes up a GDPS/Active-Active environment?

- GDPS/Active-Active
- IBM Tivoli NetView for z/OS
 - IBM Tivoli NetView for z/OS Enterprise Management Agent (NetView agent)
- IBM Tivoli Monitoring
- System Automation for z/OS
- IBM Multi-site Workload Lifeline for z/OS
- Middleware – DB2, IMS, CICS...
- Replication Software
 - IBM InfoSphere Replication Server for z/OS (DB2)
 - IBM InfoSphere IMS Replication for z/OS

Integration of a number of software products

ts

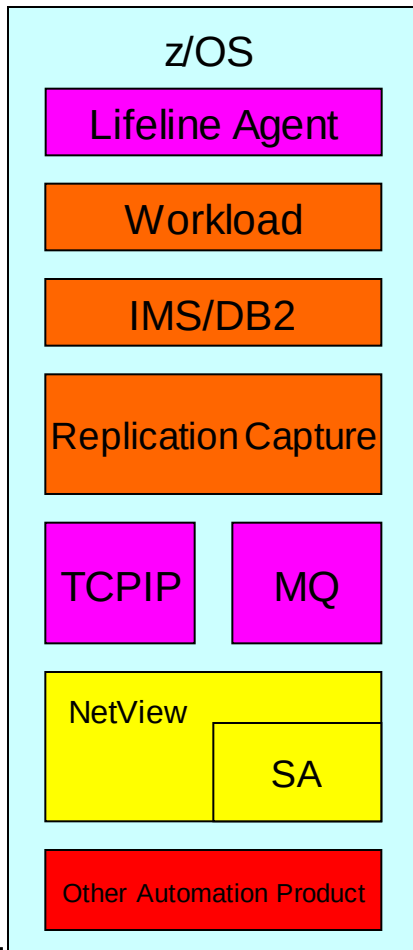
S/W replication technique (for example DB2)



1. Transaction committed
2. Capture read the DB updates from the log
3. Capture put the updates on the send-queue
4. Apply received the updates from the receive-queue
5. Apply copied the DB updates to the target databases

Architectural building blocks

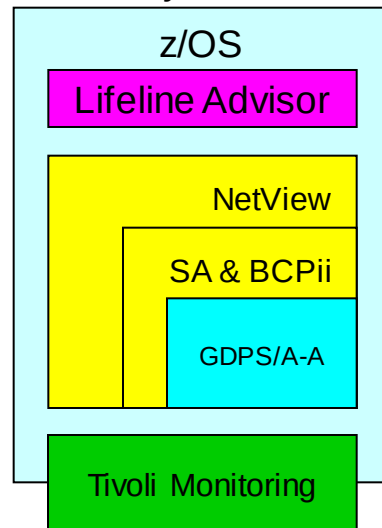
Active Production



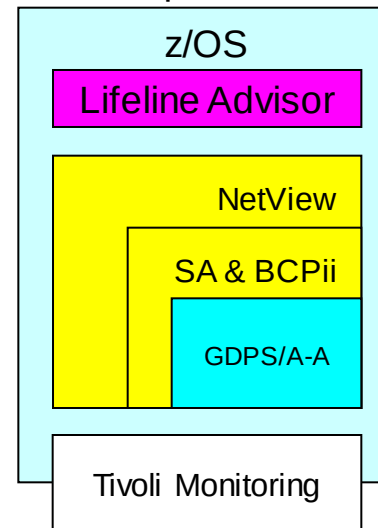
WAN & SASP-compliant Routers
used for workload distribution

SE/HMC LAN

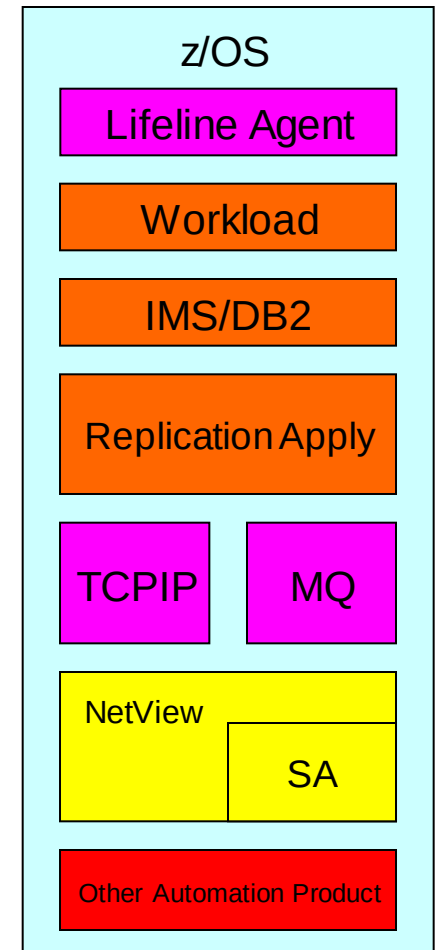
Primary Controller



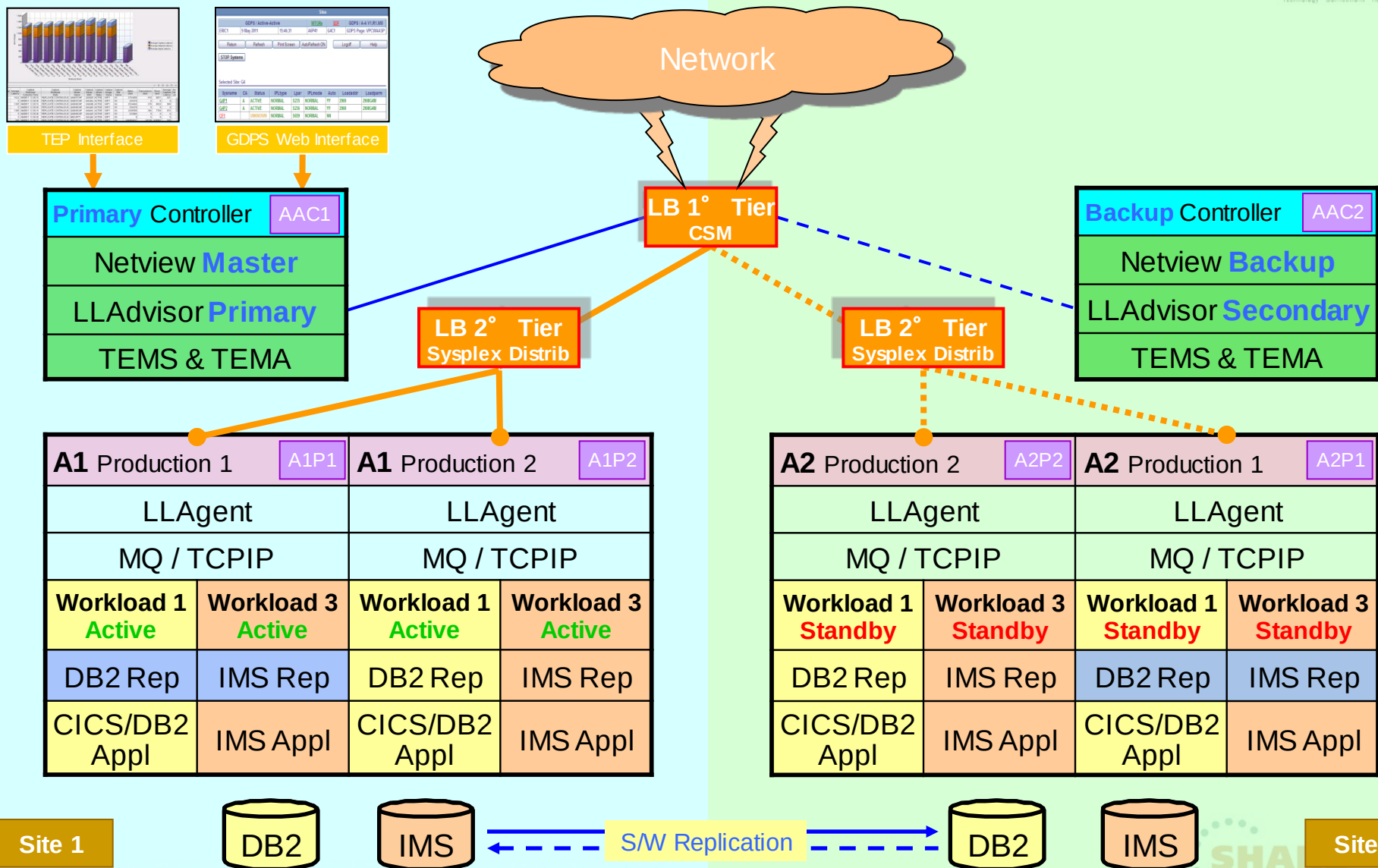
Backup Controller



Standby Production



GDPS/A-A configuration



GDPS/Active-Active (the product)

- **Automation code** is an extension on many of the techniques tried and tested in other GDPS products and with many client environments for management of their mainframe CA & DR requirements
- **Control code** only runs on Controller systems
- **Workload management** - start/stop components of a workload in a given Sysplex
- **Replication management** - start/stop replication for a given workload between sites
- **Routing management** - start/stop routing of transactions to a site
- **System and Server management** - STOP (graceful shutdown) of a system, LOAD, RESET, ACTIVATE, DEACTIVATE the LPAR for a system, and capacity on demand actions such as CBU/OOCOD
- **Monitoring** the environment and **alerting** for unexpected situations
- **Planned/Unplanned situation management and control** - planned or unplanned site or workload switches; automatic actions such as automatic workload switch (policy dependent)
- **Powerful scripting capability** for complex/compound scenario automation

Pre-requisite products

- **IBM Multi-site Workload Lifeline v1.1**
 - Advisor – runs on the Controllers & provides information to the external load balancers on where to send transactions and information to GDPS on the health of the environment
 - There is one primary and one secondary advisor
 - Agent – runs on all production images with active/active workloads defined and provide information to the Lifeline Advisor on the health of that system
- **IBM Tivoli NetView for z/OS v6.1**
 - Runs on all systems and provides automation and monitoring functions. The NetView Enterprise Master normally runs on the Primary Controller
- **IBM Tivoli Monitoring v6.2.2 FP3**
 - Can run on the Controllers, on zLinux, or distributed servers – provides monitoring infrastructure and portal plus alerting/situation management via Tivoli Enterprise Portal, Tivoli Enterprise Portal Server and Tivoli Enterprise Monitoring Server

Pre-requisite products...

- **IBM InfoSphere Replication Server for z/OS v10.1**
 - Runs on production images where required to capture (active) and apply (standby) data updates for DB2 data. Relies on MQ as the data transport mechanism (QREP)
- **IBM InfoSphere IMS Replication for z/OS v10.1**
 - Runs on production images where required to capture (active) and apply (standby) data updates for IMS data. Relies on TCPIP as the data transport mechanism
- **System Automation for z/OS v3.3**
 - Runs on all images. Provides a number of critical functions:
 - BCPii
 - Remote communications capability to enable GDPS to manage sysplexes from outside the sysplex
 - System Automation infrastructure for workload and server management
- **Optionally the OMEGAMON suite of monitoring tools to provide additional insight**

Pre-requisite software matrix

Pre-requisite software [version/release level]		GDPS Controller	A-A Systems	non A-A Systems
Operating Systems				
	z/OS 1.11 or higher	YES	YES	YES
Application Middleware				
	DB2 for z/OS V9 or higher	NO	YES ¹⁾	as required
	IMS V11	NO	YES ¹⁾	as required
	Websphere MQ V7	NO	MQ is only required for DB2 data replication	as required
Replication				
	InfoSphere Replication Server for z/OS V10.1	NO	YES ¹⁾	as required ²⁾
	InfoSphere IMS Replication for z/OS V10.1	NO	YES ¹⁾	as required ²⁾
Management and Monitoring				
	GDPS/A-A V1.1	YES	NO	NO
	Tivoli NetView for z/OS V6.1	YES	YES	YES
	Tivoli System Automation for z/OS V3.3 + SPE APARs	YES	YES	YES
	Multi-site Workload Lifeline Version for z/OS 1.1	YES	YES	NO
24	Tivoli Monitoring V6.2.2 Fix Pack 3	YES	YES	NO

Pre-requisite software matrix (cont)

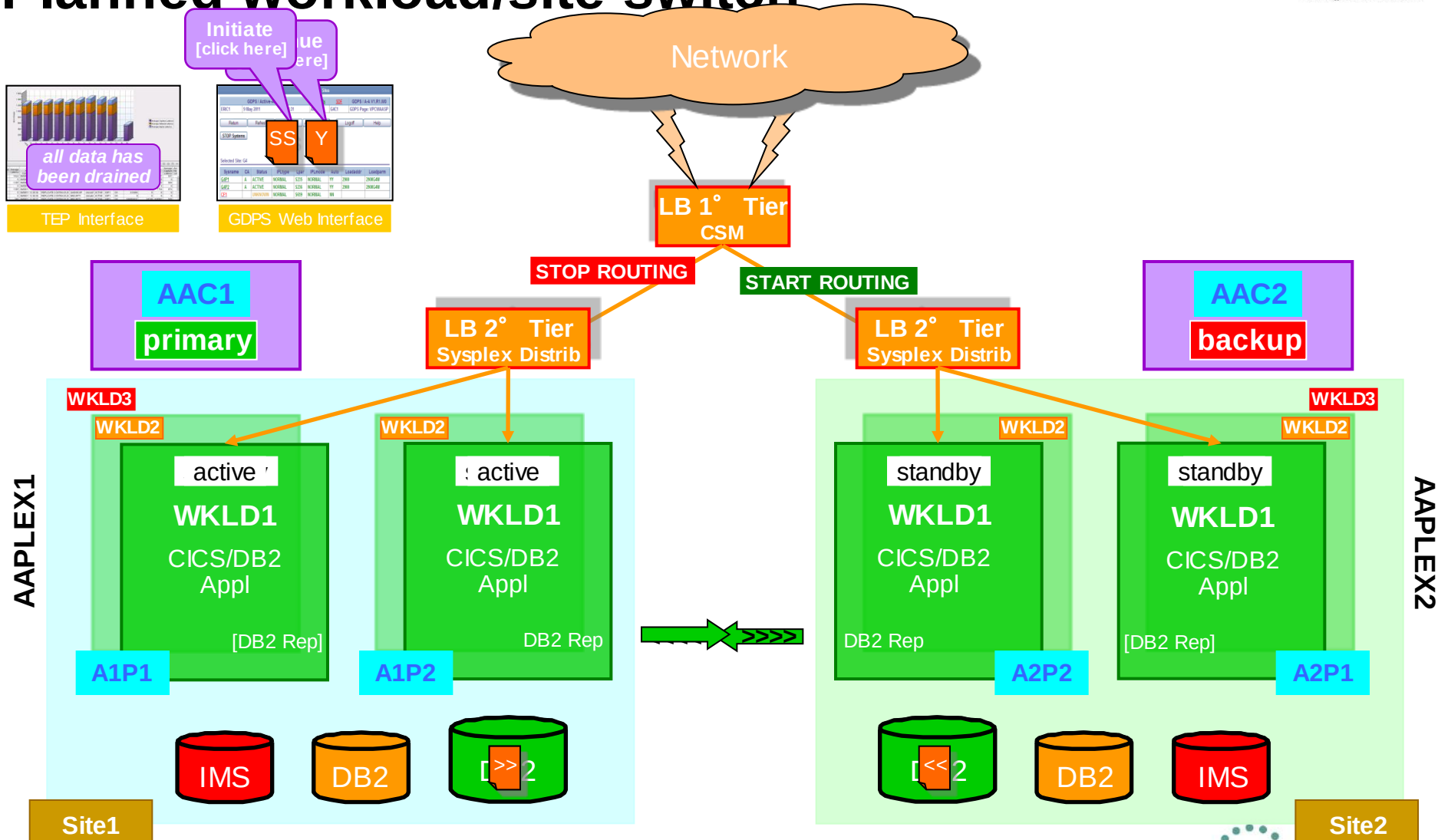
Pre-requisite software [version/release level]		GDPS Controller	A-A Systems	non A-A Systems
Optional Monitoring Products				
	IBM Tivoli OMEGAMON XE on z/OS V4.2.0	YES	YES	as required
	IBM Tivoli OMEGAMON XE for Mainframe Networks V4.2.0	YES	YES	as required
	IBM Tivoli OMEGAMON XE for Storage V4.2.0	YES	YES	as required
	IBM Tivoli OMEGAMON XE for DB2 Performance Expert (or Performance Monitor) on z/OS v4.2.0	NO	YES ¹⁾	as required
	IBM Tivoli OMEGAMON XE on CICS for z/OS v4.2.0	NO	YES ¹⁾	as required
	IBM Tivoli OMEGAMON XE on IMS v4.2.0	NO	YES ¹⁾	as required
	IBM Tivoli OMEGAMON XE for Messaging v7.0	NO	YES ¹⁾	as required

¹⁾ workload dependent

Note: Details of cross product dependencies are listed in the PSP information for GDPS/Active-Active which can be found by selecting the **Upgrade:GDPS** and **Subset:AAV1R1** at the following URL:

<http://www14.software.ibm.com/webapp/set2/psearch/search?domain=psp&new=y>

Planned workload/site switch



20

Note: multiple workloads and needed infrastructure resources are not shown for clarity sake

Complete your sessions evaluation online at SHARE.org/AnaheimEval

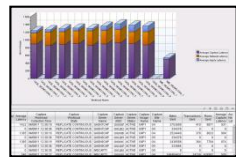
Planned workload/site switch (cont)

```
COMM = 'Switch all workloads to SITE2'  
ROUTING = 'STOP WORKLOAD=ALL SITE=AAPLEX1'  
ASSIST = 'CHECK ALL WORKLOAD UPDATES REPLICATED'  
ROUTING = 'START WORKLOAD=ALL SITE=AAPLEX2'
```

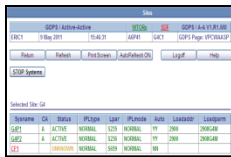
- **Stop routing transactions** to all workloads active to Sysplex AAPLEX1 in Site1
- Wait until all updates on AAPLEX1 are replicated to Sysplex AAPLEX2 in Site2
 - check via the TEP or the Replication Dashboard that all updates have drained from the active to standby site, before stopping replication between the sites
- **Start routing transactions** for workloads previously active in Site1 to Site2
- **Note:** Replication is expected to be active in both directions at all times

**The workloads are now processing transactions in Site2
for all workloads with replication from Site2 to Site1**

Unplanned site failure

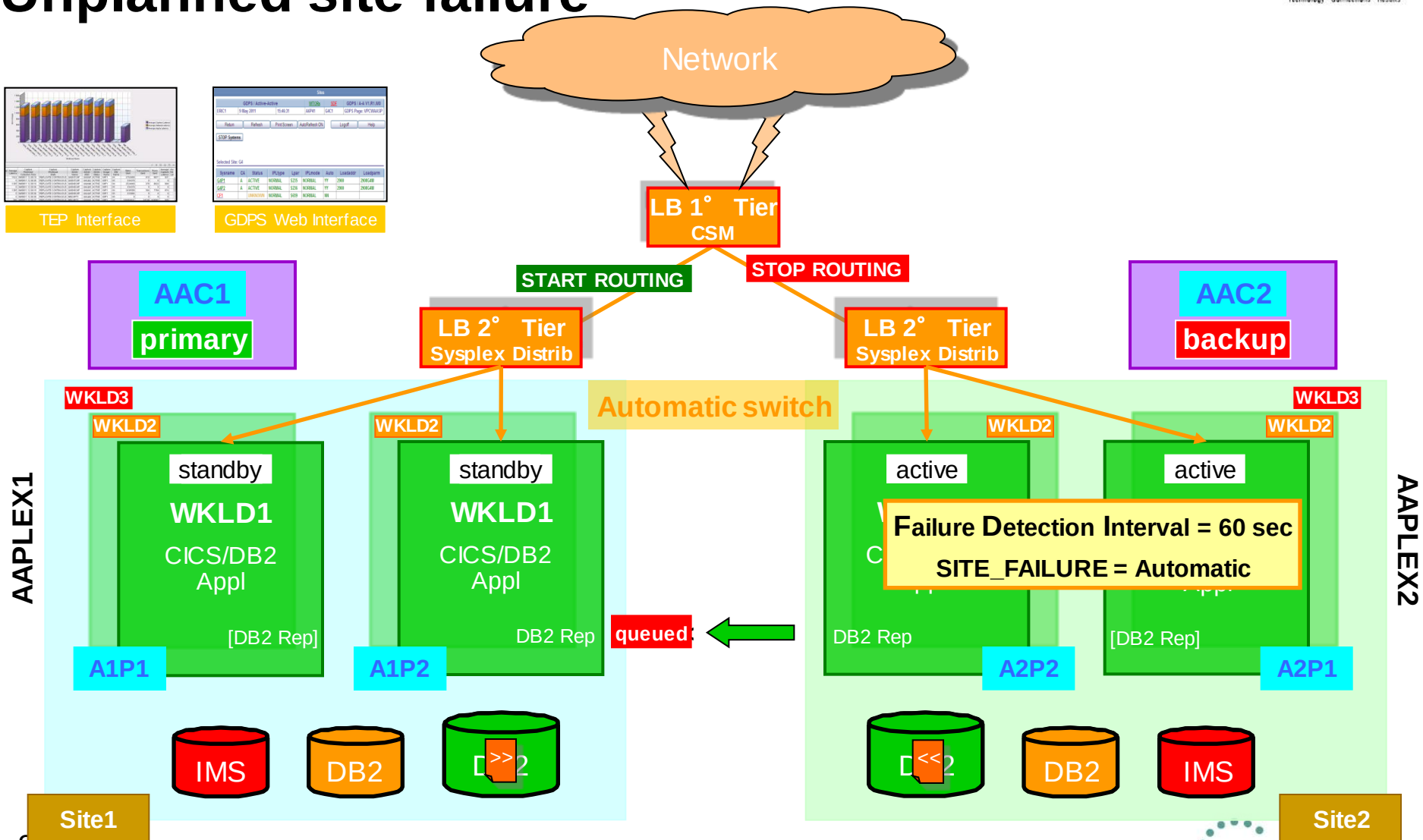


TEP Interface



System	CS	Status	Program	User	Priority	Size	Language	Language
GDPS	A	ACTIVE	NORMAL	CS	NORMAL	17	288	COBOL
GDPS	A	ACTIVE	NORMAL	CS	NORMAL	17	288	COBOL
GDPS	A	ACTIVE	NORMAL	CS	NORMAL	17	288	COBOL

GDPS Web Interface



Note: multiple workloads and needed infrastructure resources are not shown for clarity sake

Complete your sessions evaluation online at SHARE.org/AnaheimEval

Go Home scenario

After an unplanned workload/site outage <i>Note: there is the potential for transactions to have been stranded in the failed site, had completed execution and committed data to the database at the time of the failure, but this data had not been replicated to the standby site.</i> <i>Assume the data is still available on the disk subsystems</i>	After a planned workload/site outage <i>Note: as the process to perform a planned site switch ensures that there are no stranded updates in the active site at the start of the switch, there is no need to start replication in the opposite direction in order to deliver stranded updates.</i>
Start the site or workload that had failed	Start the site or workload that had been stopped
Restart replication from the site brought back online to the currently active site - this delivers any stranded changes resulting from the unplanned outage (*)	
Re-synchronize the recovering site with data from the currently active site , by starting replication in the other direction	Re-synchronize the restarted site or workload with data from the currently active site , by starting replication from the active to now standby site
Re-direct the workload , once the recovered site is operational and can process workloads	Re-direct the workload , once the restarted site is both operational and the data replication has caught up and can now process workloads

(*) attempts to apply the stranded changes to the data in the active site may result in an exception or conflict, as the before image of the update that is stranded will no longer match the updated value in the active site. For IMS replication, the adaptive apply process will discard the update and issue messages to indicate that there has been a conflict and an update has been discarded. For DB2 replication, the update may not be applied, depending on conflict handling policy settings, and additionally an exception record will be inserted into a table.

Testing results*

Configuration:

- **9 * CICS-DB2** workloads + **1 * IMS** workload
- Distance between site 300 miles (≈500kms)
- Setting: Site failure detection interval = 60 seconds

Test1:

Planned site switch

GDPS Active/Active

20 seconds

GDPS/XRC
GDPS/GM

≈ 1/2 hour

Test2:

Unplanned site switch
After a site failure
(Automatic)

GDPS Active/Active

107 seconds

GDPS/XRC
GDPS/GM

≈ 1/2 hour

Positioning

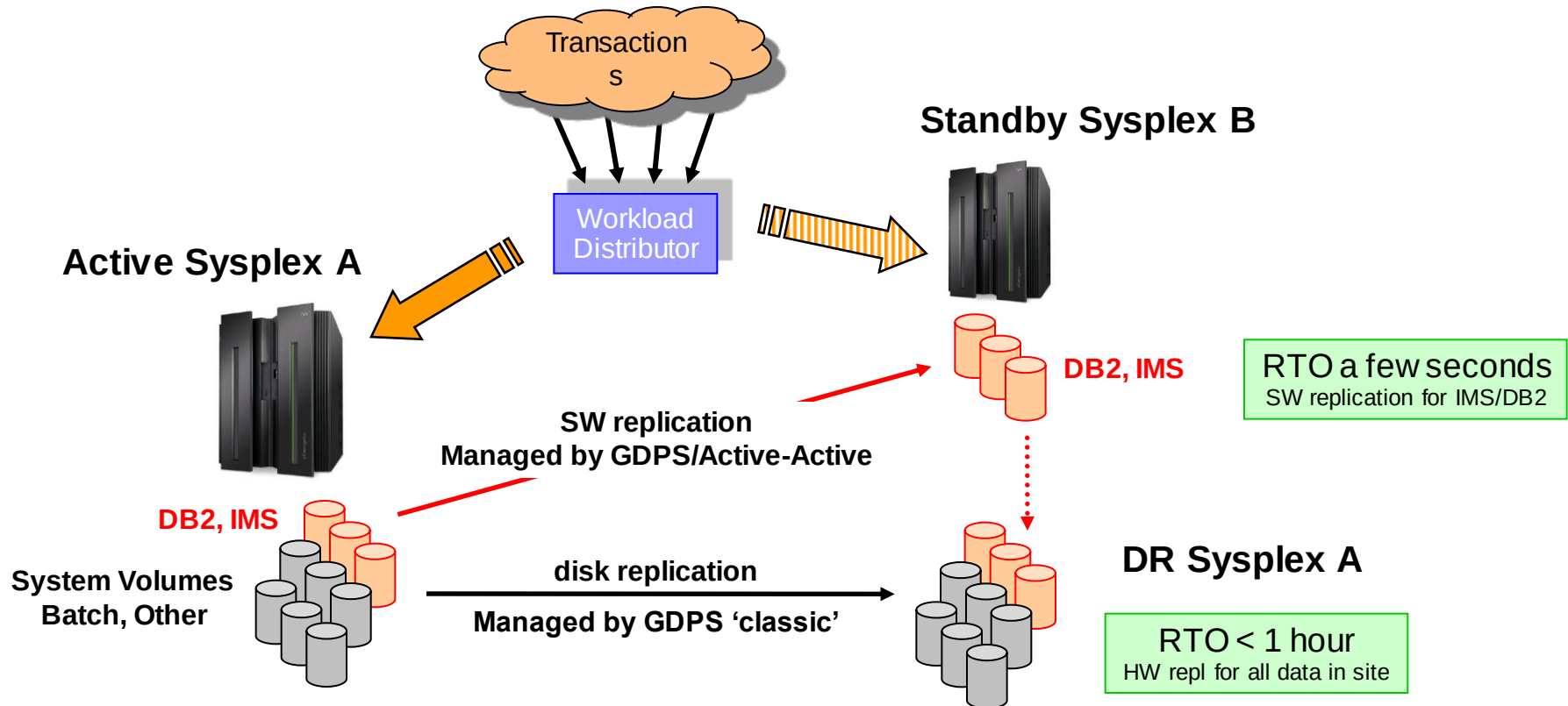
- **GDPS/Active-Active is for mission critical workloads that have stringent recovery objectives that can not be achieved using existing GDPS solutions**
 - RTO measured in seconds for unplanned outages
 - RPO measured in seconds for unplanned outages
 - Non-disruptive site switch of workloads for planned outages
 - At any distance
 - NOT intended to substitute for local availability solutions (e.g., Parallel Sysplex enabled applications)

Deployment of GDPS/Active-Active

- **Option 1 – create new sysplex environments for active/active workloads**
 - Simplifies operations as scope of Active/Active environment is confined to just this or these specific workloads and the Active/Active managed data
- **Option 2 – Active/Active workload and traditional workload co-exist within the same sysplex**
 - Still will need new active sysplex for the second site
 - Increased complexity to manage recovery of Active/Active workload to one place, and remaining systems to a different environment, from within the same sysplex
 - Existing GDPS/PPRC customer will have to understand operational interactions between GDPS/PPRC and GDPS/Active-Active

No single right answer – will depend on client environment and requirements/objectives

Disk Replication and Software Replication with GDPS



Two switch decisions for Sysplex A problems ...

Workload Switch – switch to SW copy (B); once problem is fixed, simply restart SW replication
Site Switch – switch to SW copy (B) and restart DR Sysplex A from the disk copy

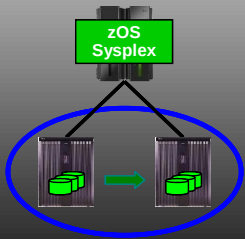
There are multiple GDPS service products under the GDPS solution umbrella to meet various customer requirements for Availability and Disaster Recovery

GDPS/PPRC HM

Continuous Availability of Data within a Data Center

Single Data Center
Applications remain active

Continuous access to data in the event of a storage subsystem outage



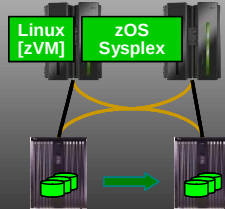
RPO=0 & RTO=0

GDPS/PPRC

Continuous Availability / Disaster Recovery within a Metropolitan Region

Two Data Centers
Systems remain active

Multi-site workloads can withstand site and/or storage failures

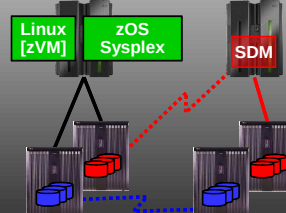


A/S RPO=0 & RTO<1 hr or
A/A RPO=0 & RTO mins

GDPS/GM & GDPS/XRC

Disaster Recovery at Extended Distance

Two Data Centers
Rapid Systems Disaster Recovery with "seconds" of Data Loss
Disaster recovery for out of region interruptions

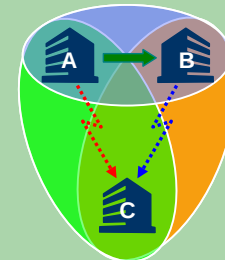


RPO secs & RTO<1 hr

GDPS/MGM & GDPS/MzGM

Continuous Availability Regionally and Disaster Recovery Extended Distance

Three Data Centers
High availability for site disasters
Disaster recovery for regional disasters



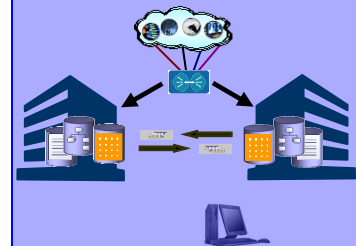
A/S RPO=0 & RTO<1 hr or
A/A RPO=0 & RTO mins
and RPO secs & RTO<1 hr

GDPS/Active-Active

Continuous Availability, Disaster Recovery, and Cross-site Workload Balancing at Extended Distance

Two or More Data Centers

All sites active



RPO secs & RTOsecs

Components

Tivoli – NV, SAz
STG – System z, DS8K, PPRC
GTS – GDPS code, Services

Tivoli – NV, SAz, SA MP, AppMan
STG – System z, DS8K, PPRC, VTS
GTS – GDPS code, Services

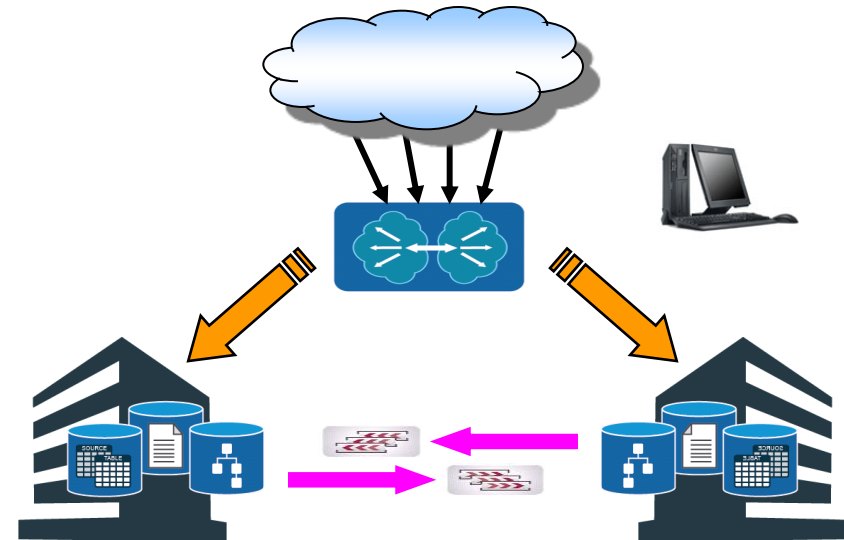
Tivoli – NV, SAz
STG – System z, DS8K, GM, XRC
GTS – GDPS control, Services

Tivoli – NV, SAz
STG – System z, DS8K, MGM, MzGM
GTS – GDPS code, Services

Tivoli – NV, SAz
AIM – Multi-site Workload Lifeline
IM – DB2 & IMS replication
STG – System z, DS8K, GC
GTS – GDPS code, Services

Summary

- Manages availability at a workload level
- Provides a central point of monitoring & control
- Manages replication between sites
- Provides the ability to perform a controlled workload site switch
- Provides near-continuous data and systems availability and helps simplify disaster recovery with an automated, customized solution
- Reduces recovery time and recovery point objectives – measured in seconds
- Facilitates regulatory compliance management with a more effective business continuity plan
- Simplifies system resource management



GDPS/Active-Active is the next generation of GDPS

Related Sessions

Tuesday, 3pm in Salon D – Session 11662

- **GDPS End to End Support (xDR and DCM)**

Wednesday, 8am in Salon H – Session 11663

- **GDPS Active/Active Sites Update**

Friday, 11am in Salon H – Session 11661

- **GDPS 3.9 Update**

Dank u

Dutch

Merci

French

Спасибо

Russian

Gracias

Spanish

شكراً

Arabic

감사합니다

Korean

Tack så mycket

Swedish

धन्यवाद

Hindi

תודה רבה

Hebrew

Obrigado

Brazilian
Portuguese

谢谢

Chinese

Dankon

Esperanto

Thank You

ありがとうございます

Japanese

Trugarez

Breton

Danke

German

Tak

Danish

Grazie

Italian

நன்றி

Tamil

děkuji

Czech

ขอบคุณ

Thai

go raibh maith agat

Gaelic