



Driving towards continuously available applications on System z

David Petersen

IBM Distinguished Engineer

Systems and Technology Group





Trademarks

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

FSCON* IBM* Redbooks* IBM (logo)* Sysplex Timer* FlashCopy* GDPS* System p* lbm.com* AIX* HyperSwap System z* DB2* IBM* Tivoli* DS6000 z/OS* IBM logo* DS8000 Parallel Sysplex* z/VM*

Dvnamic Infrastructure* POWER5

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.

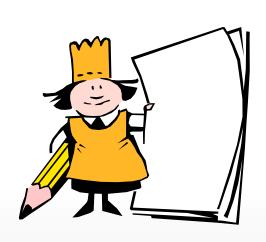
Information about non-IBM products is obtained from the manufacturers of those products or their published announcements. IBM has not tested those products and cannot confirm the performance, compatibility, or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products 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.

^{*} Registered trademarks of IBM Corporation



Agenda

- Level set (short GDPS recap)
- Requirements
- Concepts
- Configurations
- Sample Scenarios
- Summary



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

Continuous

Availability of Data

within a Data Center

RPO=0
[RTO secs]
for disk only

Single Data Center

Applications remain active

Continuous access to data in the event of a storage outage

Continuous
Availability with
DR within
Metropolitan Region

GDPS/PPRC RPO=0 RTO mins / RTO<1h (<20km) (>20km)

Two Data Centers

Systems remain active

Multi-site workloads can withstand site and/or storage failures **Disaster Recovery Extended Distance**

GDPS/GM & GDPS/XRC RPO secs, RTO<1h

Two Data Centers

Rapid Systems
D/R w/ "seconds"
of data loss

Disaster Recovery for out of region interruptions

CA Regionally and Disaster Recovery Extended Distance

GDPS/MGM & GDPS/MzGM RPO=0,RTO mins/<1h & RPO secs, RTO<1h

Three Data Centers

High availability for site disasters

Disaster recovery for regional disasters



Interagency Paper on Sound Practices to Strengthen the Resilience of the U.S. Financial System [Docket No. R-1128] (April 7, 2003)

- 1. Identify clearing and settlement activities in support of critical financial markets
- 2. Determine appropriate recovery and resumption objectives for clearing and settlement activities in support of critical markets
 - ...core clearing and settlement organizations should develop the capacity to
 recover and resume clearing and settlement activities within the business day
 on which the disruption occurs with the overall goal of achieving recovery and
 resumption within two hours after an event.
- 3. Maintain sufficient geographically dispersed resources to meet recovery and resumption objectives.
 - Back-up arrangements should be as far away from the primary site as necessary to avoid being subject to the same set of risks as the primary location.
 - The effectiveness of back-up arrangements in recovering from a wide-scale disruption should be confirmed through testing.
- 4. Routinely use or test recovery and resumption arrangements.
 - One of the lessons learned from September 11 is that testing of business recovery arrangements should be expanded.





How Much Interruption can your Business Tolerate?

Standby

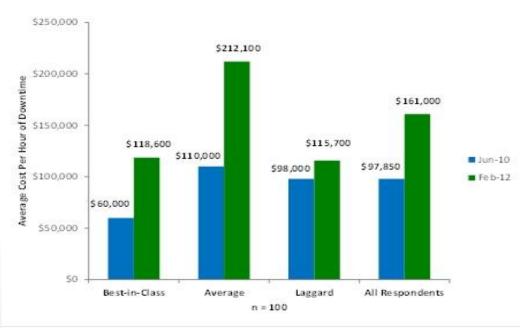
Ensuring Business Continuity:

- 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)

Active/Active

Global Enterprises that operate across timezones no longer have any 'off-hours' window. Continuous Availability is required.

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





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 2010³
- Some industries fine for downtime and inability to meet regulatory compliance
- Downtime ranges from 300–1,200 hours per year, depending on industry¹

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)
- Multi-sysplex, multi-platform solution
 - "Recover my business rather than my platform technology"
- Ensure successful recovery via automated processes (similar to GDPS technology today)
 - Can be handled by less-skilled operators
- 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)



From High Availability to Continuous Availability

GDPS/PPRC	GDPS/XRC or GDPS/GM	GDPS/Active-Active
Failover Model	Failover Model	Near CA model
Recovery Time ≈ 2 min	Recovery Time < 1 hour	Recovery time < 1 minute
Distance < 20 km	Unlimited distance	Unlimited distance

- GDPS/Active-Active is for mission critical workloads that have stringent recovery objectives that can not be achieved using existing GDPS solutions
 - RTO approaching zero, measured in seconds for unplanned outages
 - RPO approaching zero, measured in seconds for unplanned outages
 - Non-disruptive site switch of workloads for planned outages
 - At any distance
- Active-Active is NOT intended to substitute for local availability solution such as Parallel Sysplex



Terminology

Active/Active Sites

 This is the overall concept of the shift from a failover model to a continuous availability model

GDPS active/active continuous availability

- This is the formal name of the overall solution under which IBM will deliver capabilities over a period of time
- While IBM currently provides the GDPS active-standby configuration, our future road map includes additional configurations that can lead to full active-active function

GDPS/Active-Active

 The name of the GDPS product which provides, along with the other products that make up the solution, the capabilities mentioned in this presentation such as workload, replication and routing management and so on. This can be shortened to GDPS/A-A



Active/Active Sites concept

 Two or more sites, separated by <u>unlimited</u> distances, running the same applications and having the same data to provide:

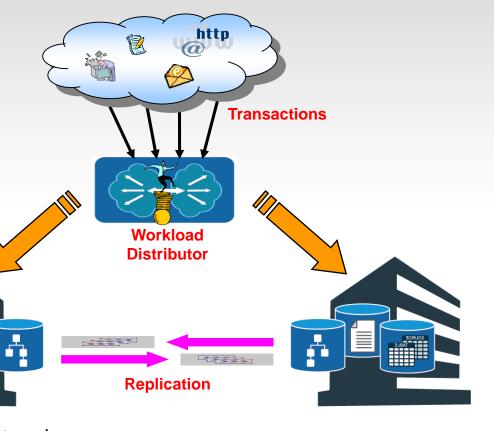
- Cross-site Workload Balancing

Continuous Availability

Disaster Recovery

 Data at geographically dispersed sites kept in sync via s/w

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

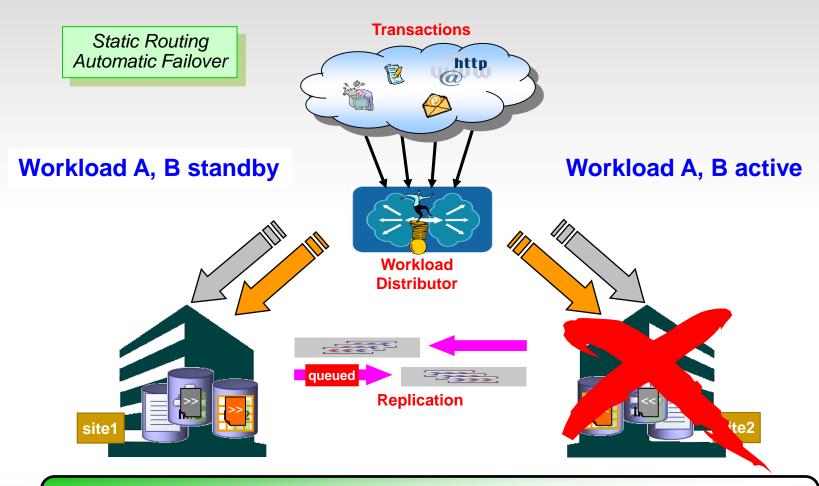


Active/Active Sites Configurations

- Configurations
 - 1. Active/Standby GA date 30th June 2011
 - 2. Active/Query GA date 31st Oct 2013
 - 3. 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

Appl B (grey) is in active/query configuration

- using same data as Appl A but read only
- active to both site1 & site2, but favor site1
- Appl B query routing according to Appl A latency policy

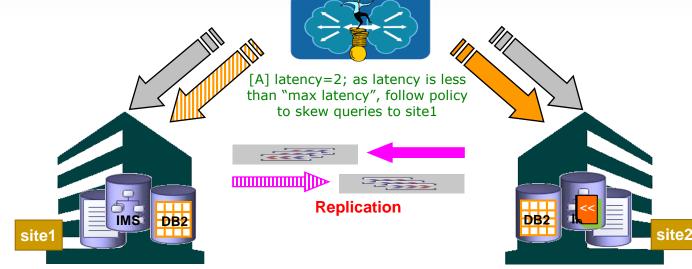
Transactions



Appl A (gold) is in active/standby configuration

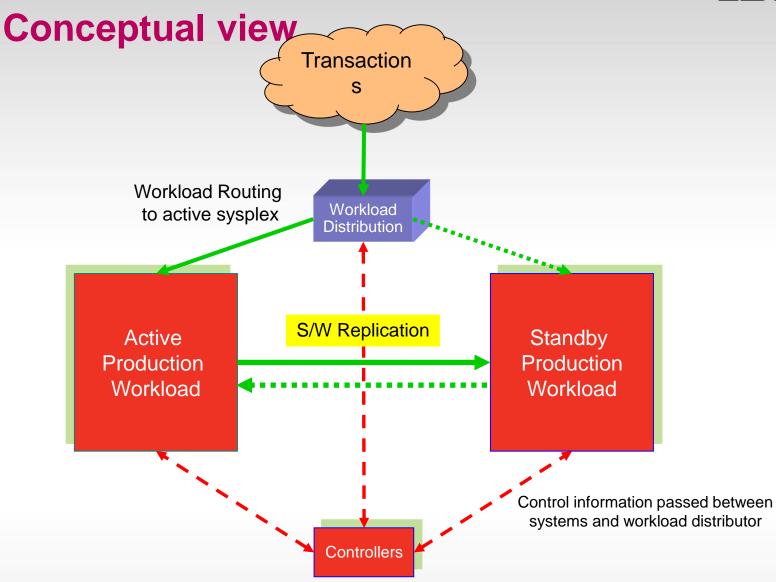
performing updates in active site [site2]





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







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 supported in the future (SoD)

Two Controller Systems

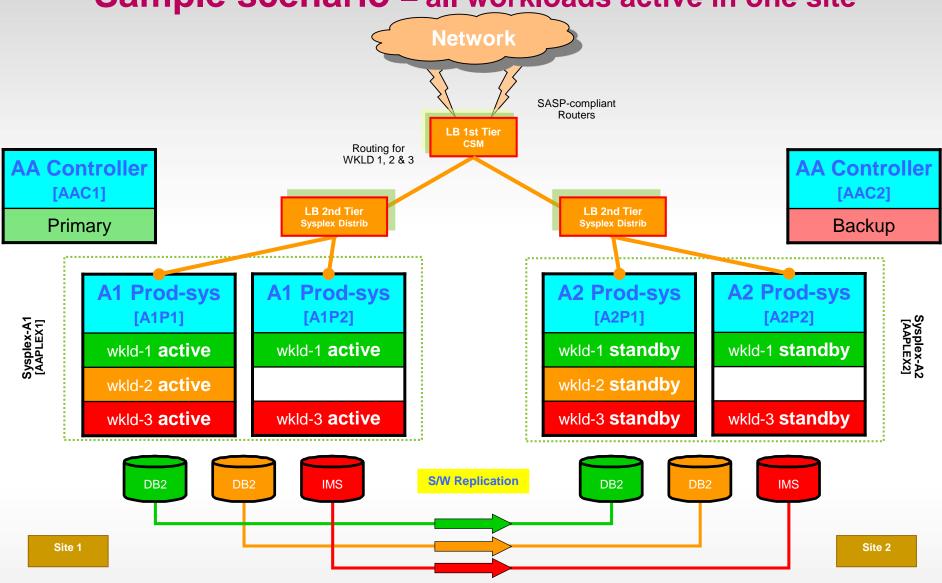
- Primary/Backup
- Typically one in each of the production locations, but there is no requirement that they are co-located in this way

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 **Network** SASP-compliant Routers **LB 1st Tier** Routing for WKLD 1 & 3 Routing for **AA Controller AA Controller** WKLĎ 2 [AAC1] [AAC2] LB 2nd Tier **LB 2nd Tier Primary Sysplex Distrib** Sysplex Distrib Backup A2 Prod-sys A1 Prod-sys **Prod-sys** A2 Prod-sys [A1P2] [A2P1] [A1P1] [A2P2] Sysplex-A2 [AAPLEX2] Sysplex-A1 [AAPLEX1] wkld1 active wkld1 active wkld1 standby wkld1 standby wkld2 active wkld2 **standby** wkld3 active wkld3 standby wkld3 active wkld3 standby S/W Replication DB2 DB₂ **IMS** IMS Site 1 Site 2



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

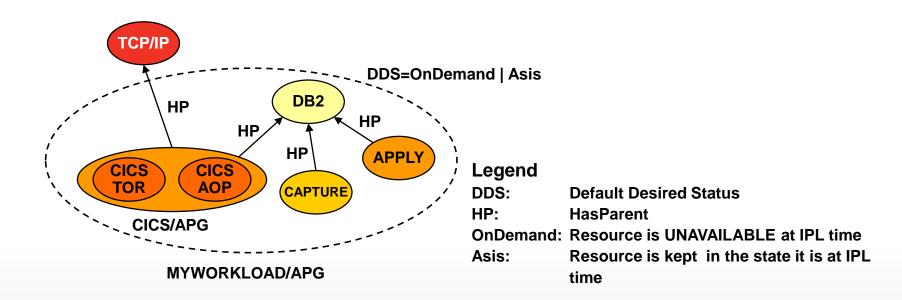
- GDPS/Active-Active
- IBM Tivoli NetView for z/OS
 - IBM Tivoli NetView for z/OS Enterprise Management Agent (NetView agent)
- IBM Tivoli Monitoring
- System Automation for z/OS
- Multi-site Workload Lifeline for z/OS (SA z/OS)
- Middleware DB2, IMS, CICS...
- Replication Software
 - IBM InfoSphere Replication Server for z/OS (DB2)
 - IBM InfoSphere IMS Replication for z/OS
- Optionally the Tivoli OMEGAMON XE suite of monitoring products

Integration of a number of software products



Software – deeper insight

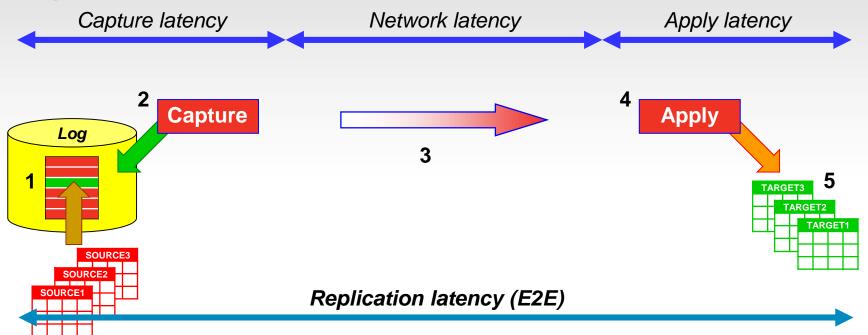
- All components of a Workload should be defined in SA z/OS* as
 - One or more Application Groups (APG)
 - Individual Applications (APL)
- The Workload itself is defined as an Application Group
- SA z/OS keeps track of the individual members of the Workload's APG and reports a "compound" status to the A/A Controller



^{*} Note that although SA is required on all systems, you can be using an alternative automation product to manage your workloads.



S/W replication (DB2, IMS,etc.) – deeper insight

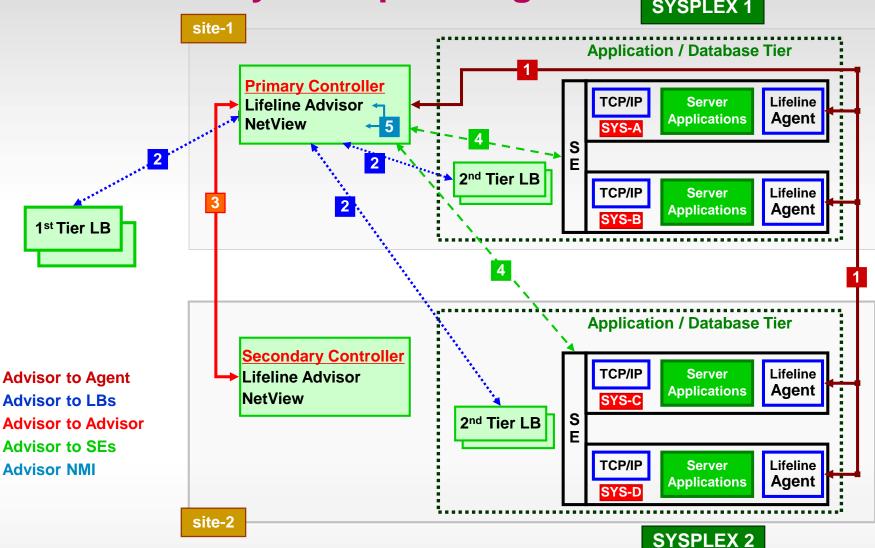


- 1. Transaction committed
- 2. Capture reads DBMS update transactions from the log
- 3. Capture transmits committed transactions to Apply
- 4. Apply receives the transactions from Capture
- 5. Apply executes the transactions at the target databases



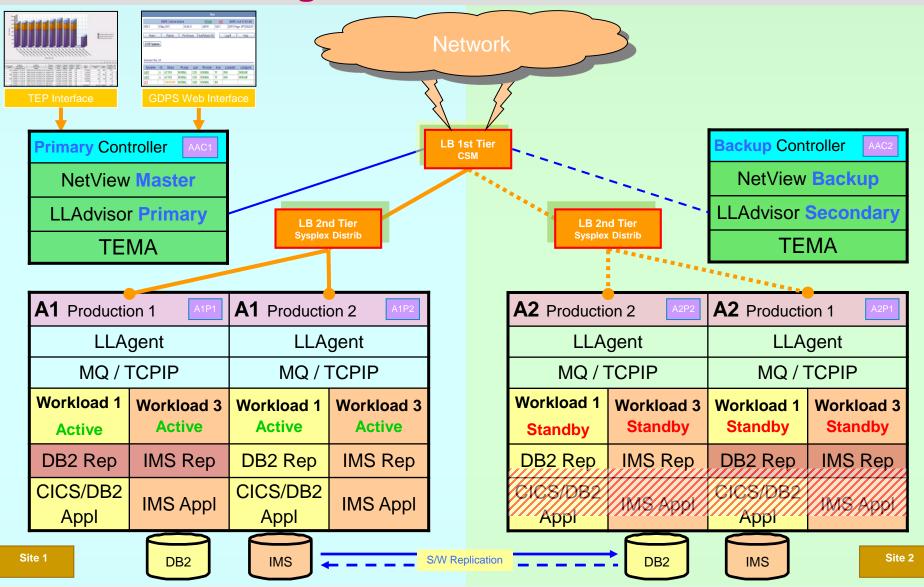
Connectivity – deeper insight

SYSPLEX 1





GDPS/A-A configuration



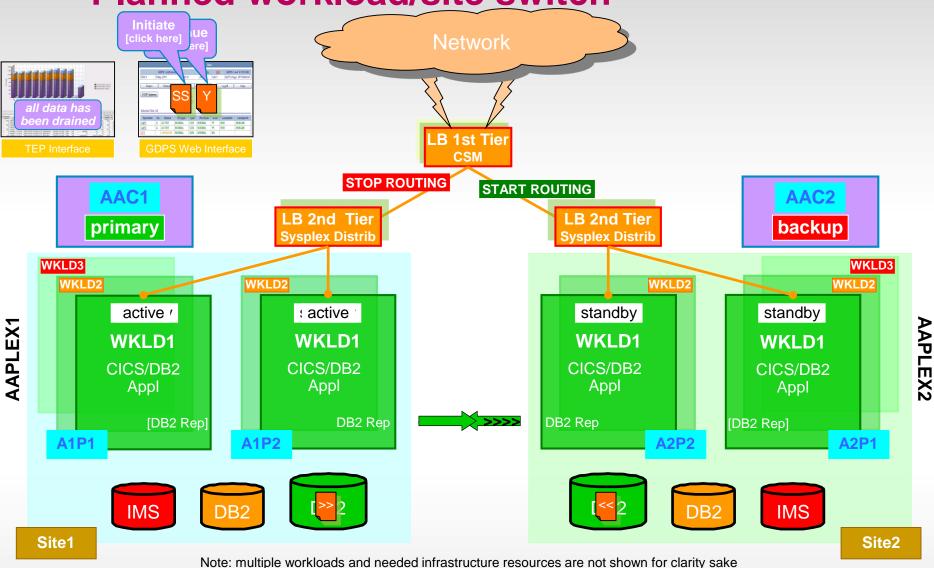


GDPS/Active-Active (the product)

- 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



Planned workload/site switch





Planned workload/site switch (cont)

COMM = 'Switch all workloads to SITE2'

ROUTING = 'STOP WORKLOAD=ALL SITE=AAPLEX1'

ASSIST = 'CHECK ALL WORKLOAD UPDATES REPLICATED'

REPLICATION = 'START WORKLOAD=ALL SITE=AAPLEX2'

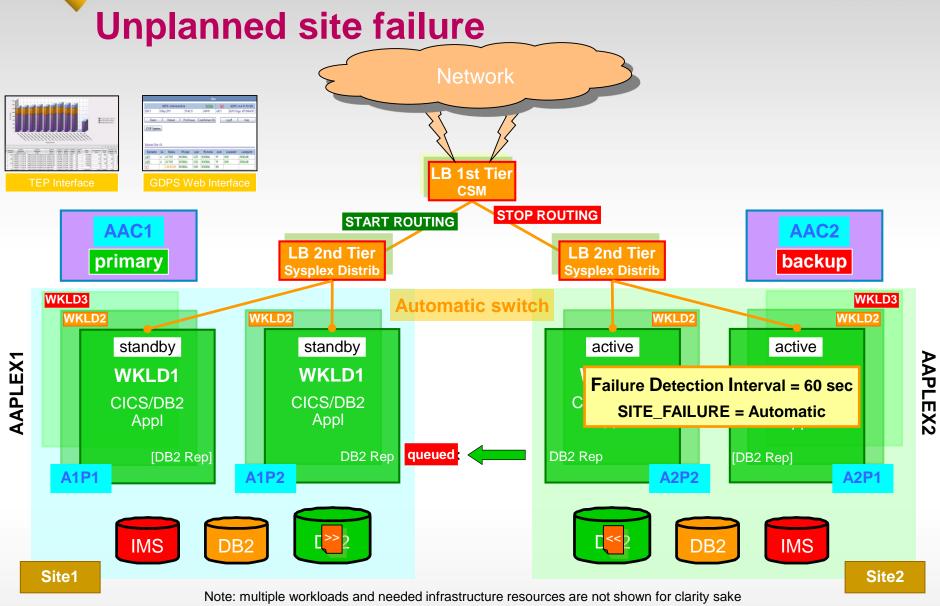
ROUTING = 'START WORKLOAD=ALL SITE=AAPLEX2'

Note: Replication is expected to be active in both directions at all times

- 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 replication from Site2 to Site1 (reverse direction), if not already active
- Start routing transactions for workloads previously active in Site1 to Site2

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









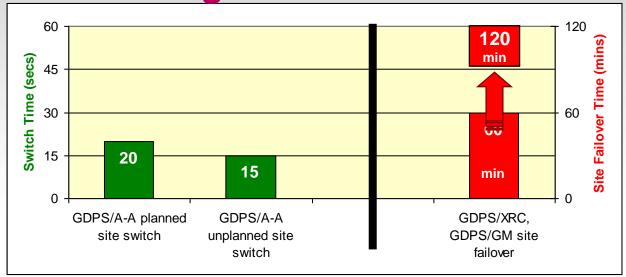
Go Home scenario

After an unplanned workload/site outage	After a planned workload/site outage		
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.	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.		
Assume the data is still available on the disk subsystems			
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.



GDPS/A-A – testing results*



Configuration

- 9 * CICS/DB2 and 1 * IMS workload
- Distance between sites 300 miles (≈500 km)
- Site failure detection interval is 60 seconds

Planned site switch

- Operations initiated switch of the workloads in a site to the other site took 20 seconds
- Current GDPS and disk replication will take ≈1-2 hours

Unplanned site switch

- Automatic switch of failed site workloads to the surviving site took 15 seconds
- Current GDPS & disk replication will take about ≈1 hour

^{*} IBM laboratory results; actual results may vary



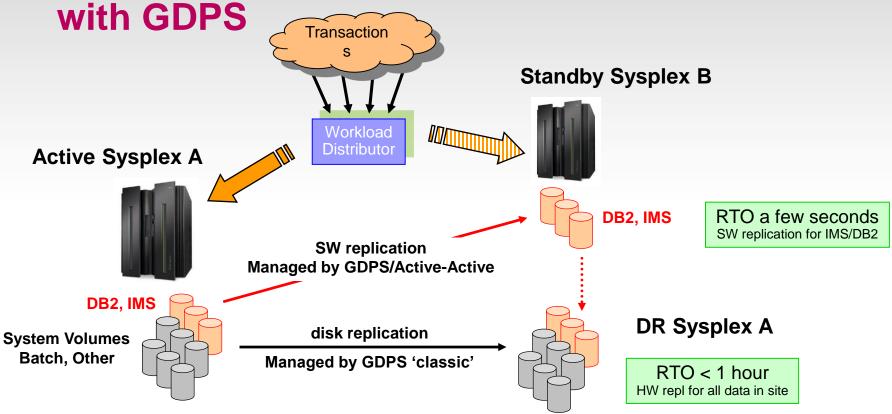
Deployment of GDPS/Active-Active

- 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

Continuous Availability of Data within a Data Center

> **GDPS/PPRC HM** RPO=0 [RTO secs] for disk only

Single Data Center

Applications remain active

Continuous access to data in the event of a storage outage

Continuous Availability with DR within **Metropolitan Region**

GDPS/PPRC RPO=0 RTO mins / RTO<1h (<20km) (>20km)

Two Data Centers

Systems remain active

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

Disaster Recovery Extended Distance

GDPS/GM & GDPS/XRC RPO secs, RTO<1h

Two Data Centers

Rapid Systems D/R w/ "seconds" of data loss

Disaster Recovery for out of region interruptions

CA Regionally and Disaster Recovery Extended Distance

GDPS/MGM & GDPS/MzGM RPO=0,RTO mins/<1h RPO secs, RTO secs & RPO secs, RTO<1h

Three Data Centers

High availability for site disasters

Disaster recovery for regional disasters

CA, DR, & Crosssite Workload Balancing **Extended Distance**

GDPS/Active-Active

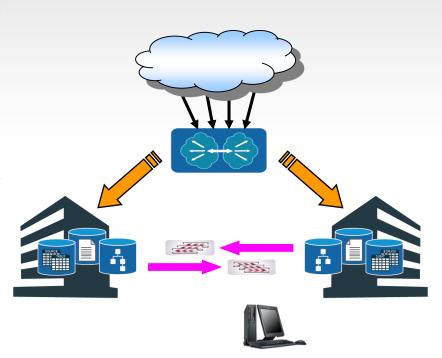
Two Active Data Centers

Automatic workload switch in seconds; seconds of data loss



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



Dank u Dutch

Merci **French**

Спасибо

Russian

Gracias

Arabic

감사합니다

děkuji Czech

Spanish

धन्यवाद

Hindi

תודה רבו^{Korea}

Hebrew

Teşekkür ederim

Turkish

Obrigado

Brazilian Portuguese

Dankon

Esperanto

Thank You

Chinese

Tak

ありがとうございます

Japanese

Trugarez **Breton**

Danke

German

Danish

Grazie

Italian

நன்றி

Tack så mycket **Swedish**

ขอบคุณ Thai

go raibh maith agat

Gaelic





Backup Charts



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 or higher

- 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 1.12 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 V1.2	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 V1.1	YES	YES	NO
Tivoli Monitoring V6.2.2 Fix Pack 3	YES	YES	NO

¹⁾ workload dependent

²⁾ can use Replication Server instances, but not the same instances as the A-A workloads



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

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:AAV1R2 at the following URL:

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





Thank You