IMS DB Replication for Enhanced System Resiliency

Greg Vance, IMS Senior Software Engineer

Disclaimer

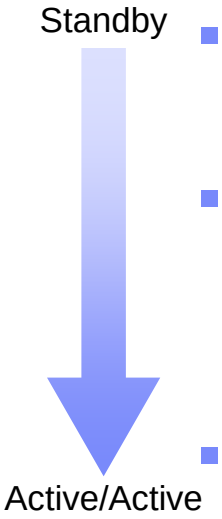
THE INFORMATION CONTAINED IN THIS PRESENTATION IS PROVIDED FOR INFORMATIONAL PURPOSES ONLY. WHILE EFFORTS WERE MADE TO VERIFY THE COMPLETENESS AND ACCURACY OF THE INFORMATION CONTAINED IN THIS PRESENTATION, IT IS PROVIDED "AS IS" WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED. IN ADDITION, THIS INFORMATION IS BASED ON IBM'S CURRENT PRODUCT PLANS AND STRATEGY, WHICH ARE SUBJECT TO CHANGE BY IBM WITHOUT NOTICE. IBM SHALL NOT BE RESPONSIBLE FOR ANY DAMAGES ARISING OUT OF THE USE OF, OR OTHERWISE RELATED TO, THIS PRESENTATION OR ANY OTHER DOCUMENTATION. NOTHING CONTAINED IN THIS PRESENTATION IS INTENDED TO, NOR SHALL HAVE THE EFFECT OF, CREATING ANY WARRANTIES OR REPRESENTATIONS FROM IBM (OR ITS SUPPLIERS OR LICENSORS), OR ALTERING THE TERMS AND CONDITIONS OF ANY AGREEMENT OR LICENSE GOVERNING THE USE OF IBM PRODUCTS AND/OR SOFTWARE.

Agenda

- Background
- GDPS Active/Active Sites solution
- InfoSphere IMS Replication for z/OS

How Much Interruption can your Business Tolerate?

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 unplanned)

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

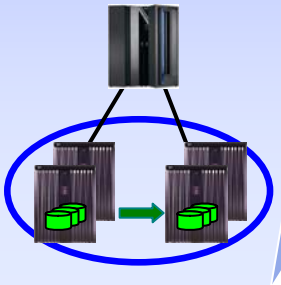
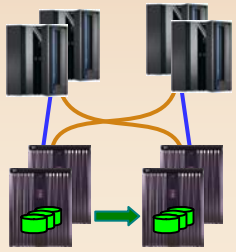
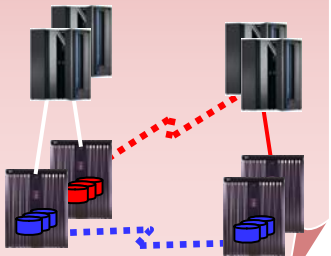
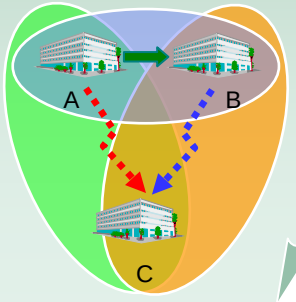
The cost of 1 hour of downtime during core business hours

Industry Sector	Loss per Hour
Financial	$8,213,470
Telecommunications	$4,611,604
Information Technology	$3,316,058
Insurance	$2,582,382
Pharmaceuticals	$2,058,710
Energy	$1,468,798
Transportation	$1,463,128
Banking	$1,145,129
Chemicals	$1,071,404
Consumer Products	$989,795

Source: Robert Frances Group 2006, "Picking up the value of PKI: Leveraging z/OS for Improving Manageability, Reliability, and Total Cost of Ownership of PKI and Digital Certificates."

What are Customers doing Today?

RPO – Recovery Point Objective
RTO – Recovery Time Objective

Continuous Availability of Data within a Data Center	Continuous Availability w/ Disaster Recovery within a Metropolitan Region	Disaster Recovery at Extended Distance	Regional Continuous Availability w/ Disaster Recovery @ Extended Distance
GDPS/HyperSwap Mgr RPO = 0 / RTO = 0	**GDPS/PPRC** RPO = 0 / RTO <1hr (>20 km) RPO = 0 / RTO = 0 (<20 km)	**GDPS/GM & GDPS/XRC** RPO secs / RTO <1 hr	**GDPS/MGM & GDPS/MzGM**
Single Data Center **Applications remain active** **Continuous access to data in the event of a storage subsystem outage**	Two Data Centers **Systems remain active** **Multi-site workloads can withstand site and/or storage failures**	Two Data Centers **Rapid Systems Disaster Recovery with "seconds" of Data Loss** **Disaster recovery for out of region interruptions**	Three Data Centers **High availability for site disasters** **Disaster recovery for regional disasters**
			

Components

Tivoli – NetView, SAz System z, DS8K, PPRC GDPS control code, Services	**Tivoli** – NV, SAz, SA-MP, AppMan System z, DS8K, VTS, PPRC GDPS control code, Services	**Tivoli** – NV, SAz System z, DS8K, Global Mirror, XRC GDPS control code, Services	**Tivoli** – NV, SAz System z, DS8K, MGM, MzGM GDPS control code, Services

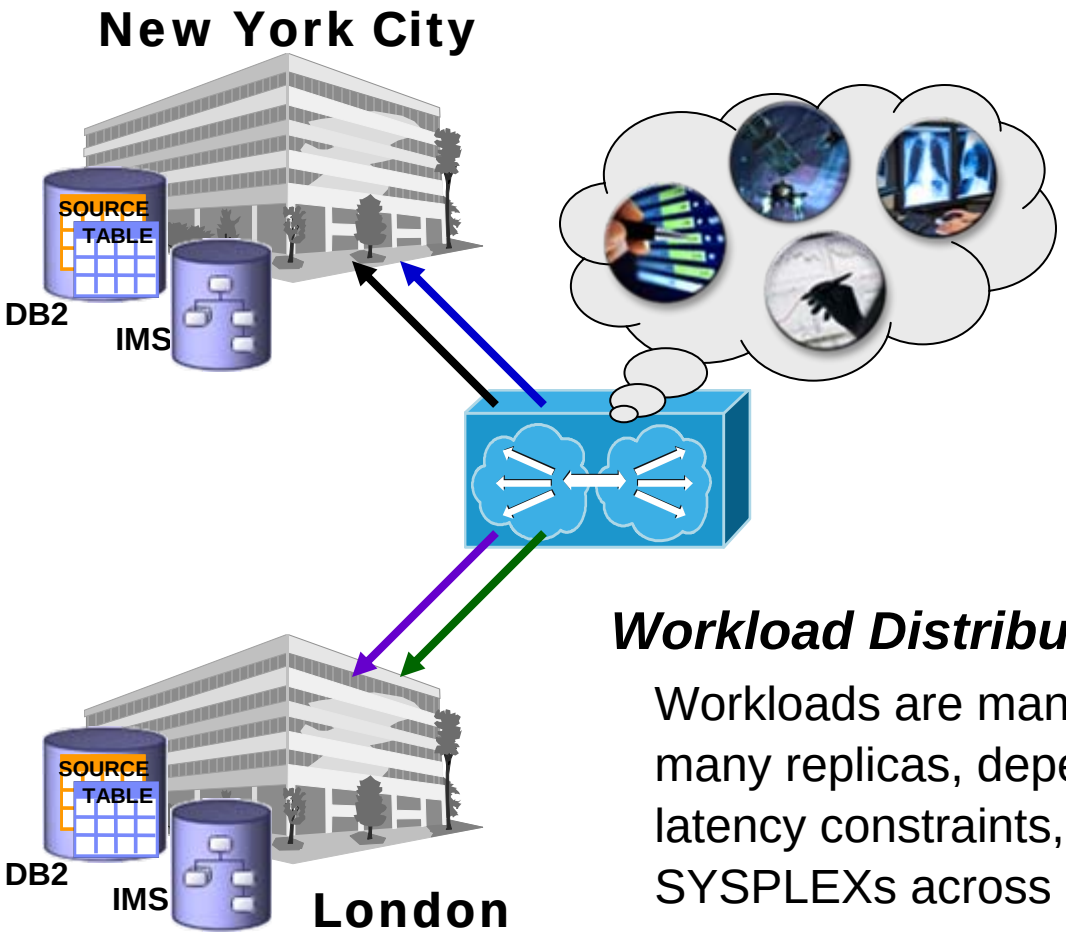
Customer Requirements Behind GDPS/Active-Active

- **Shift from a failover model to a nearly-continuous availability model**
  RTO near zero, and RPO near zero
- **Multi-sysplex, multi-platform solution**
  - Access data from any site across unlimited distance between sites
  - Provide workload distribution between sites
    - Route around failed sites
    - Dynamically select sites based on ability to handle workload
- **Provide application level granularity**
  "Recover my business rather than my platform technology"
  - Current solutions employ an all-or-nothing approach
    e.g. complete disk mirroring, requiring extra network capacity
    - Some workloads may require immediate access from every site.
    - Less critical workloads may only need to update other sites every 24 hours.
- **Ensure successful recovery via automated processes**
  S***uitable for less-skilled operators, like other GDPS technologies***
  - No application changes
  - Replace RYO solutions

Agenda

- Background
- GDPS Active/Active Sites solution
- InfoSphere IMS Replication for z/OS

Active/Active Sites Concept



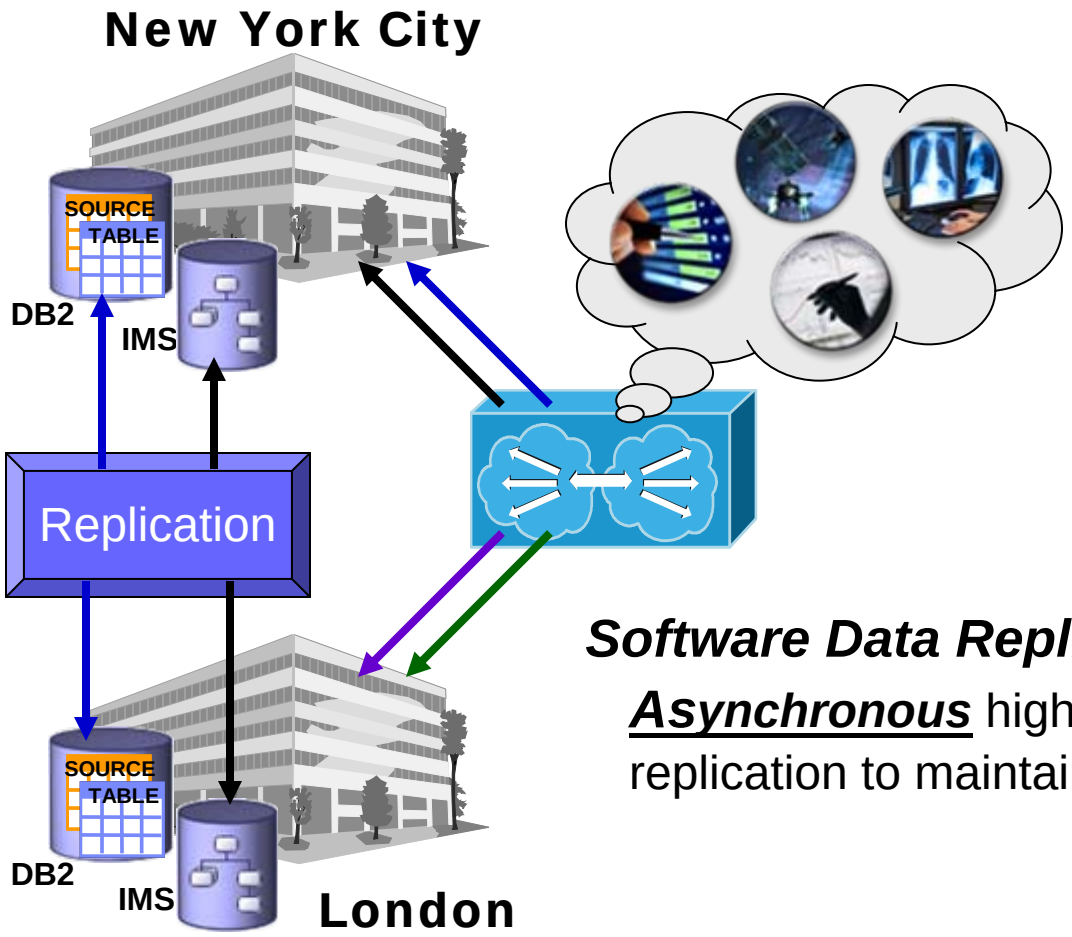
Two sites,

- separated by ***unlimited*** distances,
- running the same applications
- having the same data to provide:
  - Cross-site Workload Balancing
  - ***Continuous*** Availability
  - Disaster Recovery

Workload Distributor:

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

Active/Active Sites Concept



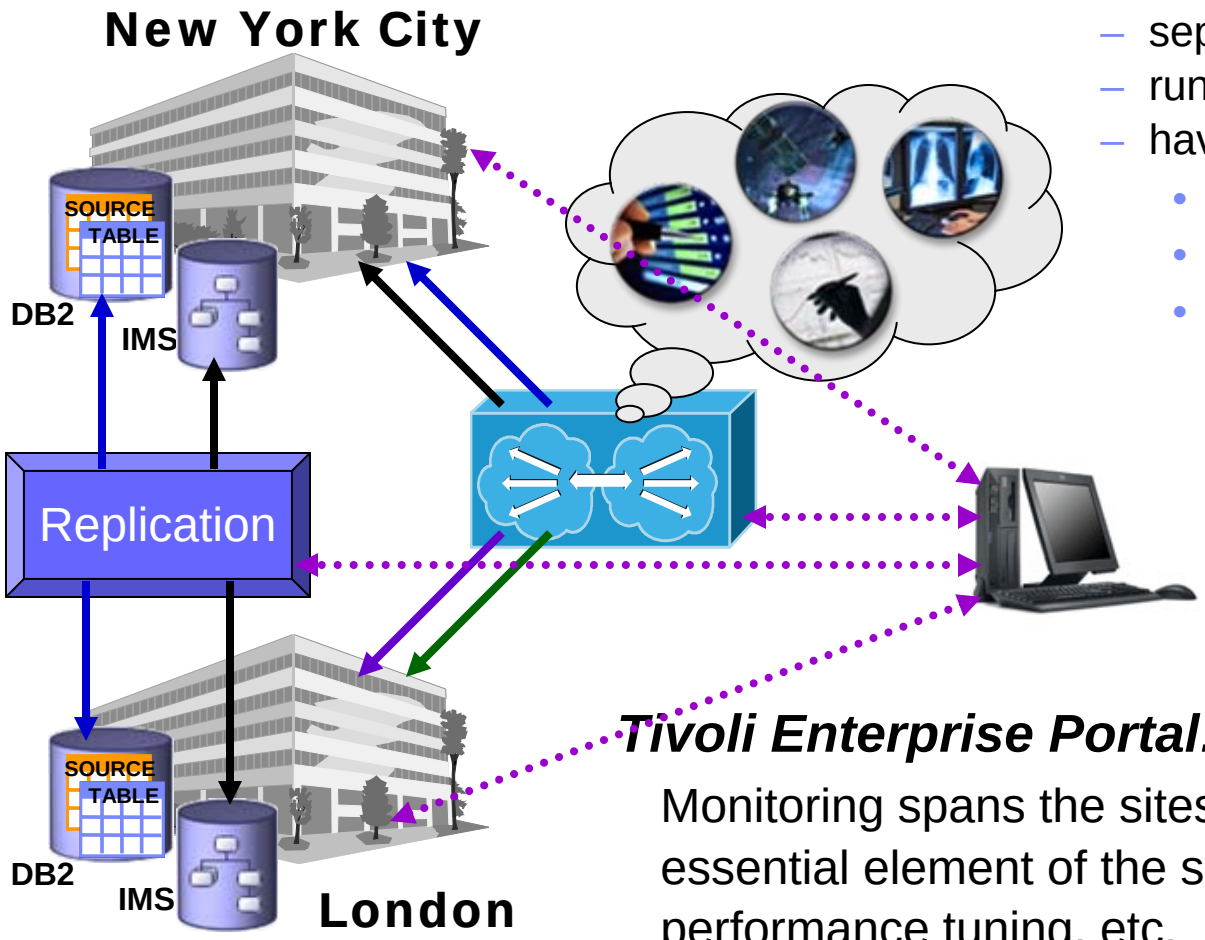
Two sites,

- separated by ***unlimited*** distances,
- running the same applications
- having the same data to provide:
  - Cross-site Workload Balancing
  - ***Continuous*** Availability
  - Disaster Recovery

Software Data Replication:

Asynchronous high throughput low latency data replication to maintain replicas at the dispersed sites.

Active/Active Sites Concept



Two or more sites,

- separated by ***unlimited*** distances,
- running the same applications
- having the same data to provide:
  - Cross-site Workload Balancing
  - ***Continuous*** Availability
  - Disaster Recovery

Tivoli Enterprise Portal:

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

GDPS/Active-Active Sites Configurations

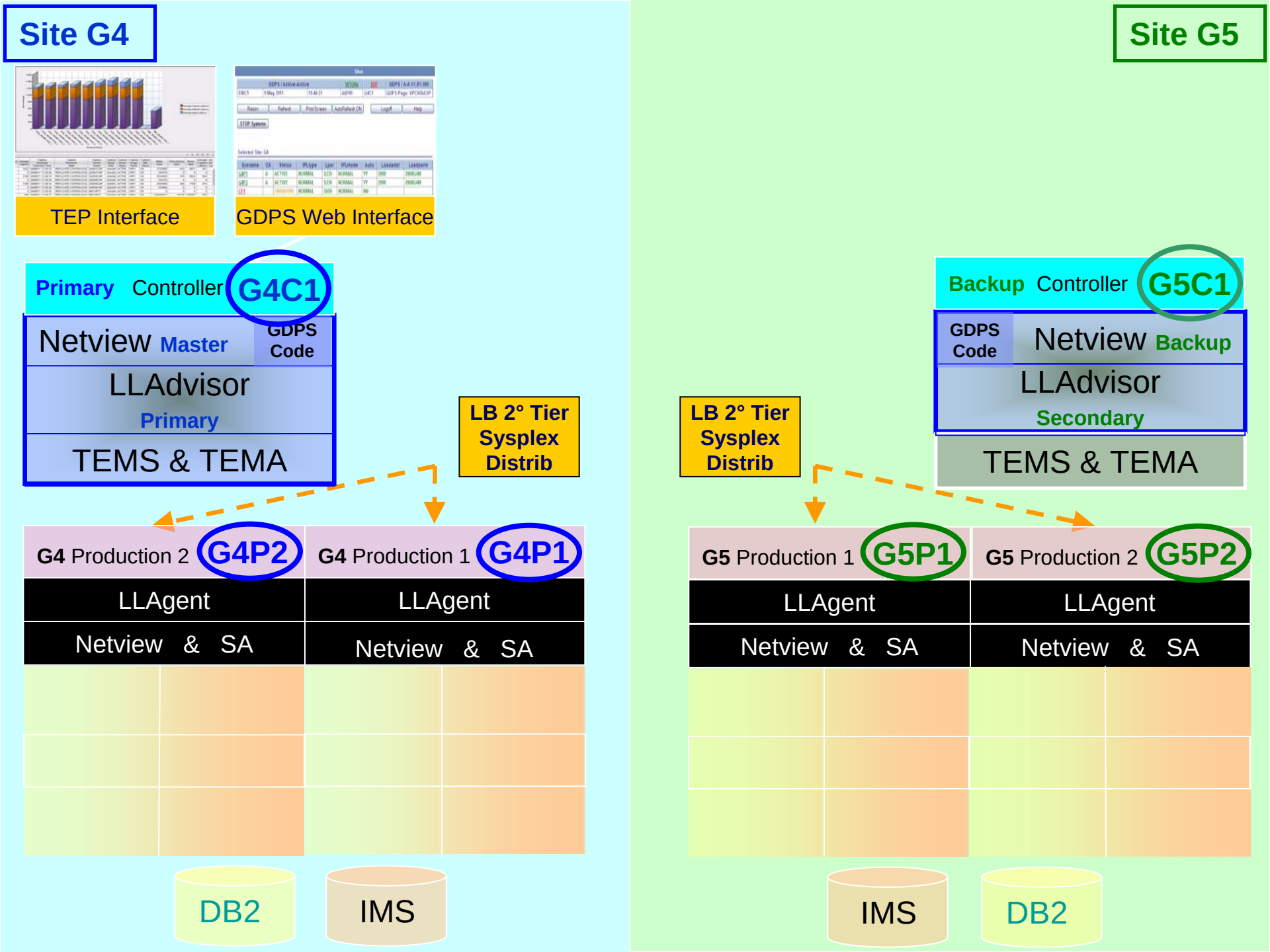
- Configurations
  1. **Active-Standby – Delivered**
  2. *Active-Query – Highest priority for next enhancement*
  3. *…*
- A configuration is specified on a workload basis
- A workload is the aggregation of these components
  - ***Software:*** user written applications (e.g., COBOL program) and the middleware run time environment (e.g., IMS, DB2, CICS…)
  - ***Data:*** related set of objects that must preserve transactional consistency and optionally referential integrity constraints (e.g., IMS Database)
  - ***Network connectivity*****:** one or more TCP/IP addresses & ports (e.g., 10.10.10.1:80)

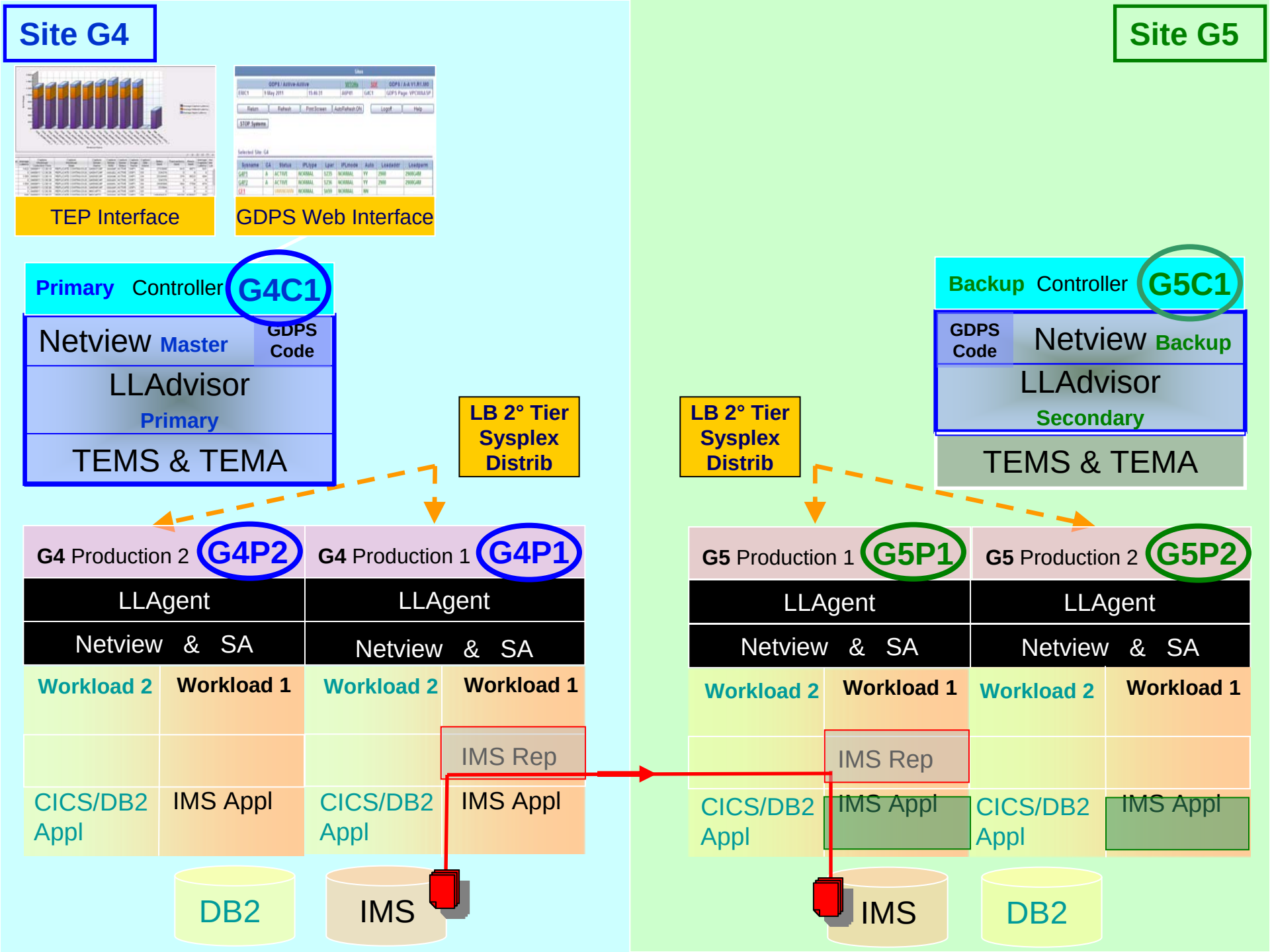
Active/Active Sites functions

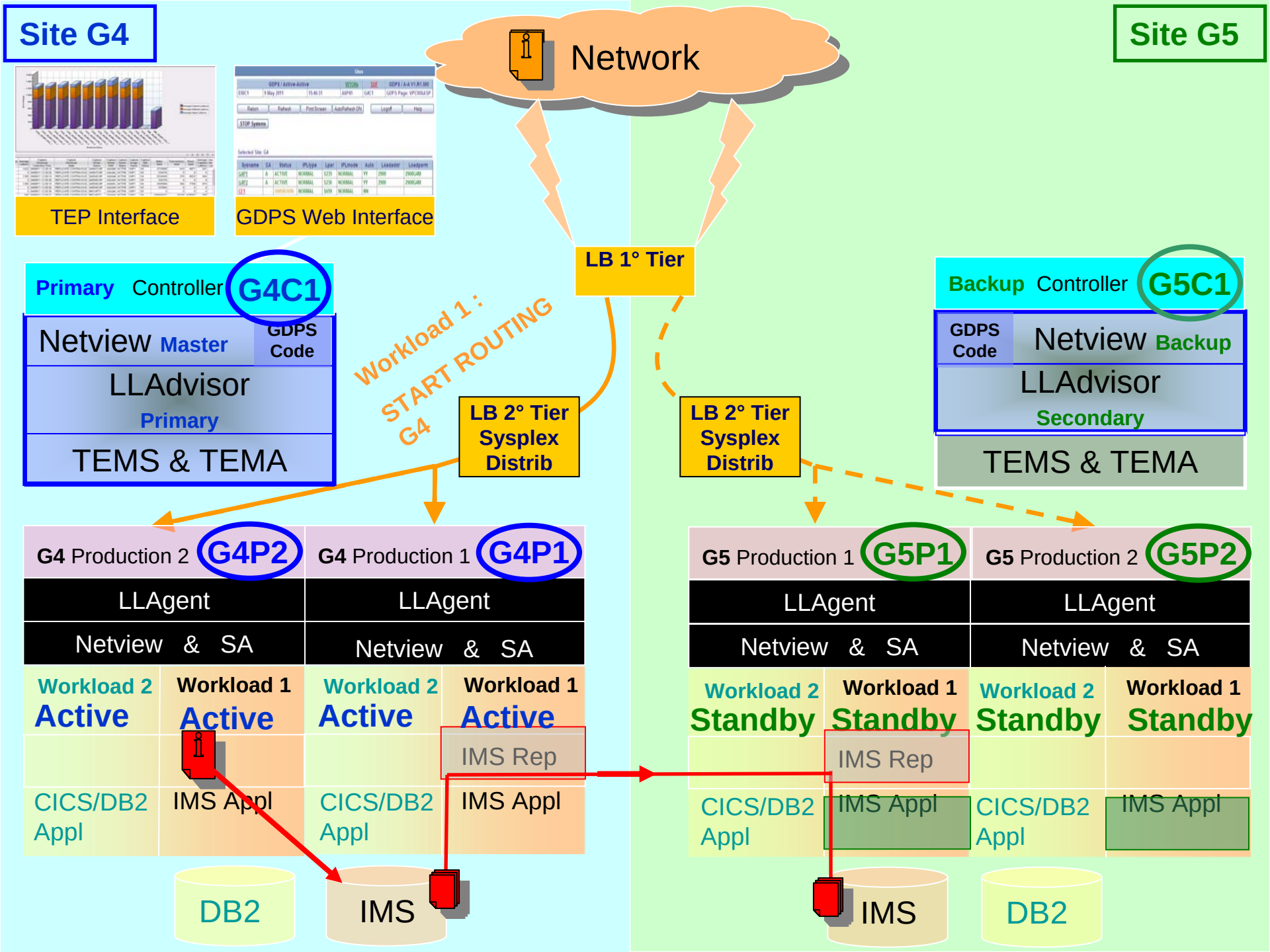
- Start/stop a controller – start and stop an A/A Sites controller
- Start/stop a site – start and stop individual sysplexes (each sysplex maps to a site)
- Stop/start a workload – start and stop individual workloads
- Monitoring – monitor the A/A Sites configuration and, if any conditions that will potentially impact a workload and/or site switch, generate an alert
- Planned workload switch – switch the workload site to the other site initiated by operator action
- Unplanned workload switch – switch failed workload to the other site, either automatically or based upon operator prompt, after the workload failure detection interval
- Planned site switch – switch all workloads executing to the other site initiated by operator action
- Unplanned site switch – switch the failed site's workloads to the other site, either automatically or based upon operator prompt, after the site failure detection interval

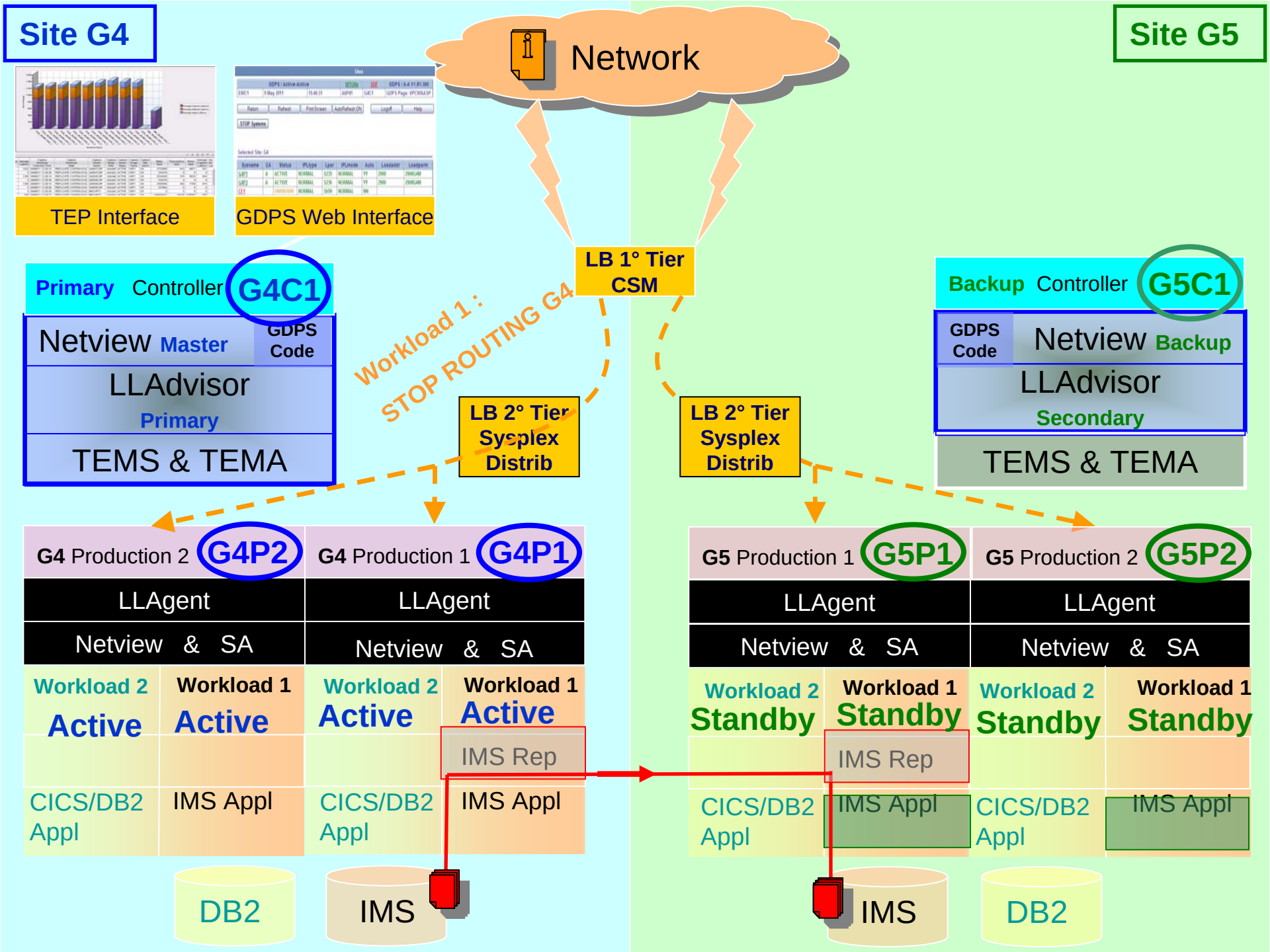
GDPS/Active-Active Components

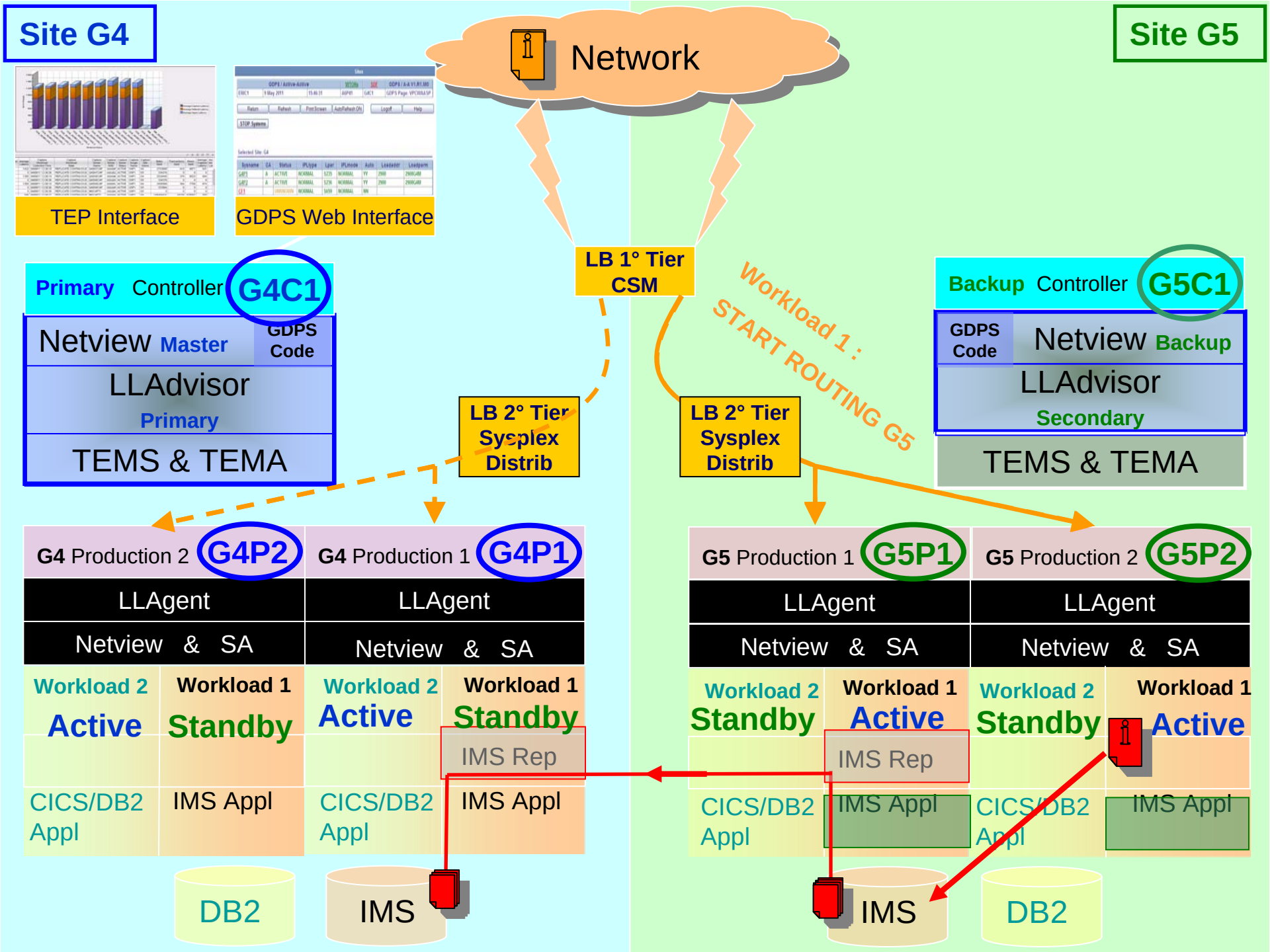
- Operating system
  - z/OS V1R11
- Applications/Middleware
  - DB2 for z/OS V9
  - IMS 10
  - WS MQ V7.0
- Replication
  - InfoSphere Replication Server (DB2) V10
  - InfoSphere IMS Replication for z/OS V10.1
- Management and monitoring
  - GDPS/Active-Active V1.1
  - NetView for z/OS V6.1
  - System Automation for z/OS V3.3
  - IBM Multi-site Workload Lifeline V1.1
  - IBM Tivoli Monitoring V6.2.2
  - Optional OMEGAMON products (required only if the customer wants to monitor the behavior of the respective products/resources that they deal with (IMS, DB2, storage, etc.)


Site G4
Site G5
TEP Interface
GDPS Web Interface
Primary Controller
G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
LB 2° Tier Sysplex Distrib
G4 Production 2
G4P2
G4 Production 1
G4P1
LLAgent
Netview & SA
DB2
IMS
Backup Controller
G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
LB 2° Tier Sysplex Distrib
G5 Production 1
G5P1
G5 Production 2
G5P2
LLAgent
Netview & SA
IMS
DB2


Site G4
Site G5
TEP Interface
GDPS Web Interface
Primary Controller G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
LB 2° Tier Sysplex Distrib
LB 2° Tier Sysplex Distrib
Backup Controller G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
G4 Production 2 G4P2
G4 Production 1 G4P1
G5 Production 1 G5P1
G5 Production 2 G5P2
LLAgent
Netview & SA
Workload 2
Workload 1
IMS Rep
CICS/DB2 Appl
IMS Appl
DB2
IMS


Site G4
Site G5
Network
TEP Interface
GDPS Web Interface
LB 1° Tier
Primary Controller G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
Workload 1 :
START ROUTING
G4
LB 2° Tier
Sysplex
Distrib
LB 2° Tier
Sysplex
Distrib
Backup Controller G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
G4 Production 2 G4P2
G4 Production 1 G4P1
G5 Production 1 G5P1
G5 Production 2 G5P2
LLAgent
Netview & SA
Workload 2
Active
Workload 1
Active
IMS Rep
CICS/DB2
Appl
IMS Appl
Workload 2
Standby
Workload 1
Standby
DB2
IMS

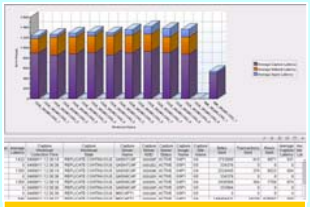

Site G4
Site G5
Network
TEP Interface
GDPS Web Interface
LB 1° Tier
CSM
Primary Controller G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
Workload 1 :
STOP ROUTING G4
LB 2° Tier
Sysplex
Distrib
LB 2° Tier
Sysplex
Distrib
Backup Controller G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
G4 Production 2 G4P2
G4 Production 1 G4P1
G5 Production 1 G5P1
G5 Production 2 G5P2
LLAgent
Netview & SA
Workload 2
Active
Workload 1
Active
IMS Rep
CICS/DB2
Appl
IMS Appl
Workload 2
Standby
Workload 1
Standby
DB2
IMS


Site G4
Site G5
Network
TEP Interface
GDPS Web Interface
LB 1° Tier CSM
Workload 1 : START ROUTING G5
Primary Controller G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
Backup Controller G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
LB 2° Tier Sysplex Distrib
LB 2° Tier Sysplex Distrib
G4 Production 2 G4P2
G4 Production 1 G4P1
G5 Production 1 G5P1
G5 Production 2 G5P2
LLAgent
Netview & SA
Workload 2
Workload 1
Active
Standby
IMS Rep
CICS/DB2 Appl
IMS Appl
DB2
IMS

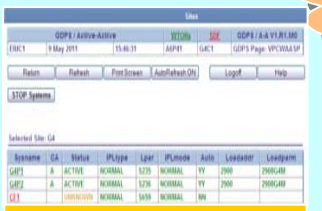
Site G4

Site G5

Network



TEP Interface



GDPS Web Interface

LB 1° Tier CSM

Primary Controller G4C1

Netview Master

GDPS Code

LLAdvisor Primary

TEMS & TEMA

Backup Controller G5C1

GDPS Code

Netview Backup

LLAdvisor Secondary

TEMS & TEMA

LB 2° Tier Sysplex Distrib

LB 2° Tier Sysplex Distrib

G4 Production 2 G4P2

G4 Production 1 G4P1

LLAgent	LLAgent
Netview & SA	Netview & SA
Workload 2 Active / Workload 1 Active	Workload 2 Active / Workload 1 Active
	IMS Rep
CICS/DB2 Appl / IMS Appl	CICS/DB2 Appl / IMS Appl

DB2

IMS

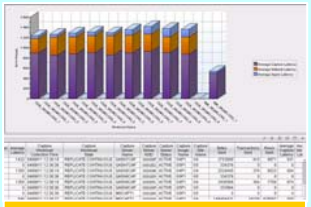
IMS

DB2

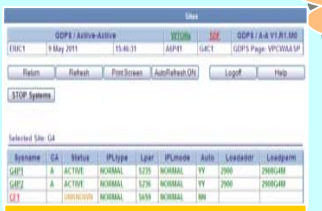
Site G4

Site G5

Network



TEP Interface



GDPS Web Interface

LB 1° Tier CSM

Workload 1 : STOP ROUTING G5

Primary Controller G4C1

Netview Master

GDPS Code

LLAdvisor Primary

TEMS & TEMA

LB 2° Tier Sysplex Distrib

LB 2° Tier Sysplex Distrib

Backup Controller G5C1

GDPS Code

Netview Backup

LLAdvisor Secondary

TEMS & TEMA

G4 Production 2 G4P2	G4 Production 1 G4P1
LLAgent	LLAgent
Netview & SA	Netview & SA
Workload 2 Active / Workload 1 Active	Workload 2 Active / Workload 1 Active
	IMS Rep
CICS/DB2 Appl / IMS Appl	CICS/DB2 Appl / IMS Appl

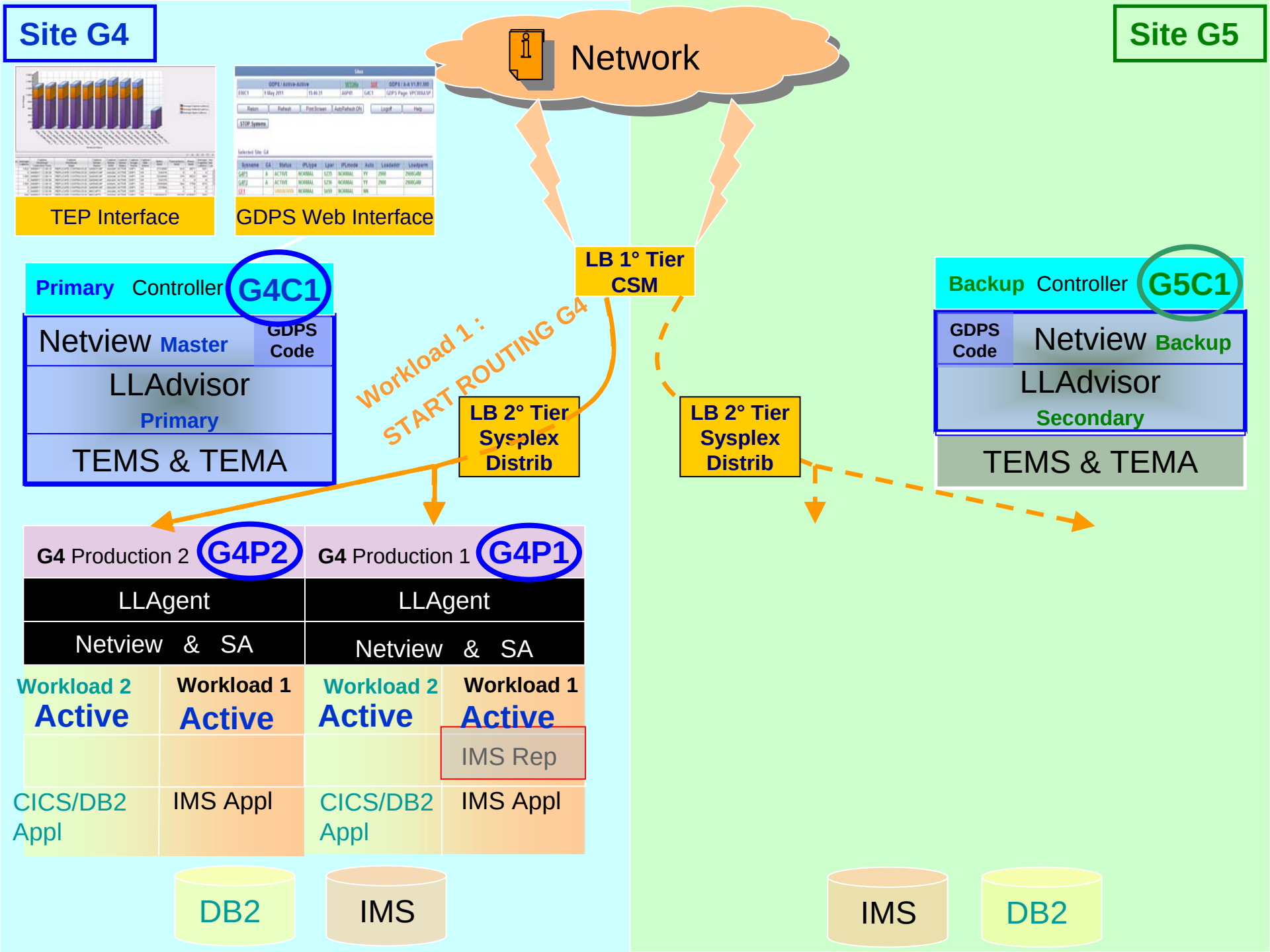
DB2

IMS

Automatic switch or not ?

IMS

DB2


Site G4
Site G5
Network
TEP Interface
GDPS Web Interface
LB 1° Tier CSM
Primary Controller G4C1
Netview Master
GDPS Code
LLAdvisor
Primary
TEMS & TEMA
Workload 1 : START ROUTING G4
LB 2° Tier Sysplex Distrib
LB 2° Tier Sysplex Distrib
Backup Controller G5C1
GDPS Code
Netview Backup
LLAdvisor
Secondary
TEMS & TEMA
G4 Production 2 G4P2
G4 Production 1 G4P1
LLAgent
LLAgent
Netview & SA
Netview & SA
Workload 2 Active
Workload 1 Active
Workload 2 Active
Workload 1 Active
IMS Rep
CICS/DB2 Appl
IMS Appl
CICS/DB2 Appl
IMS Appl
DB2
IMS
IMS
DB2

Active-Standby Configuration Testing

Scenario	Active/Standby switch time	Traditional GDPS product restart time
Planned workload switch	20 seconds	Not possible
Unplanned workload switch	120 seconds*	Not possible
Planned site switch (10 workloads)	20 seconds	About 1-2 hours
Unplanned workload switch (10 workloads)	107 seconds*	About 1 hour

* Failure Detection Interval of 60 seconds
* Internal IBM testing. Your results may vary

GDPS/Active-Active
Positioning

GDPS/Active-Active is for mission critical workloads that have stringent recovery objectives that can not be achieved using existing GDPS solutions.

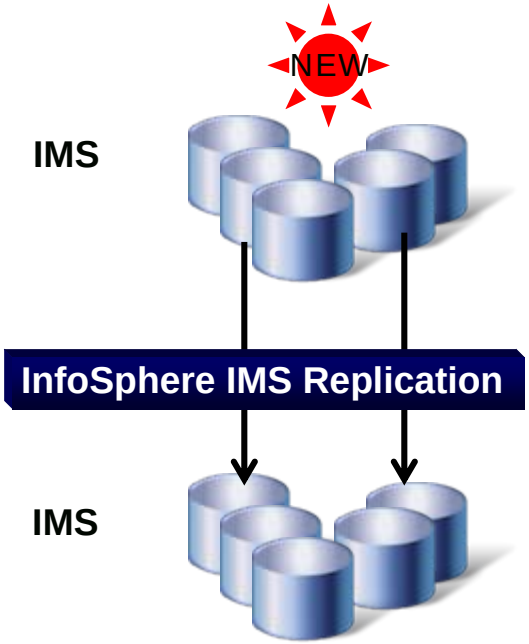
- RTO measured in seconds for unplanned outages
- RPO measured in seconds for unplanned outages
- Non-disruptive site switch of workloads for planned outages
- At any distance
- NOT intended to substitute for local availability solutions (eg: Parallel SYSPLEX enabled applications)

Agenda

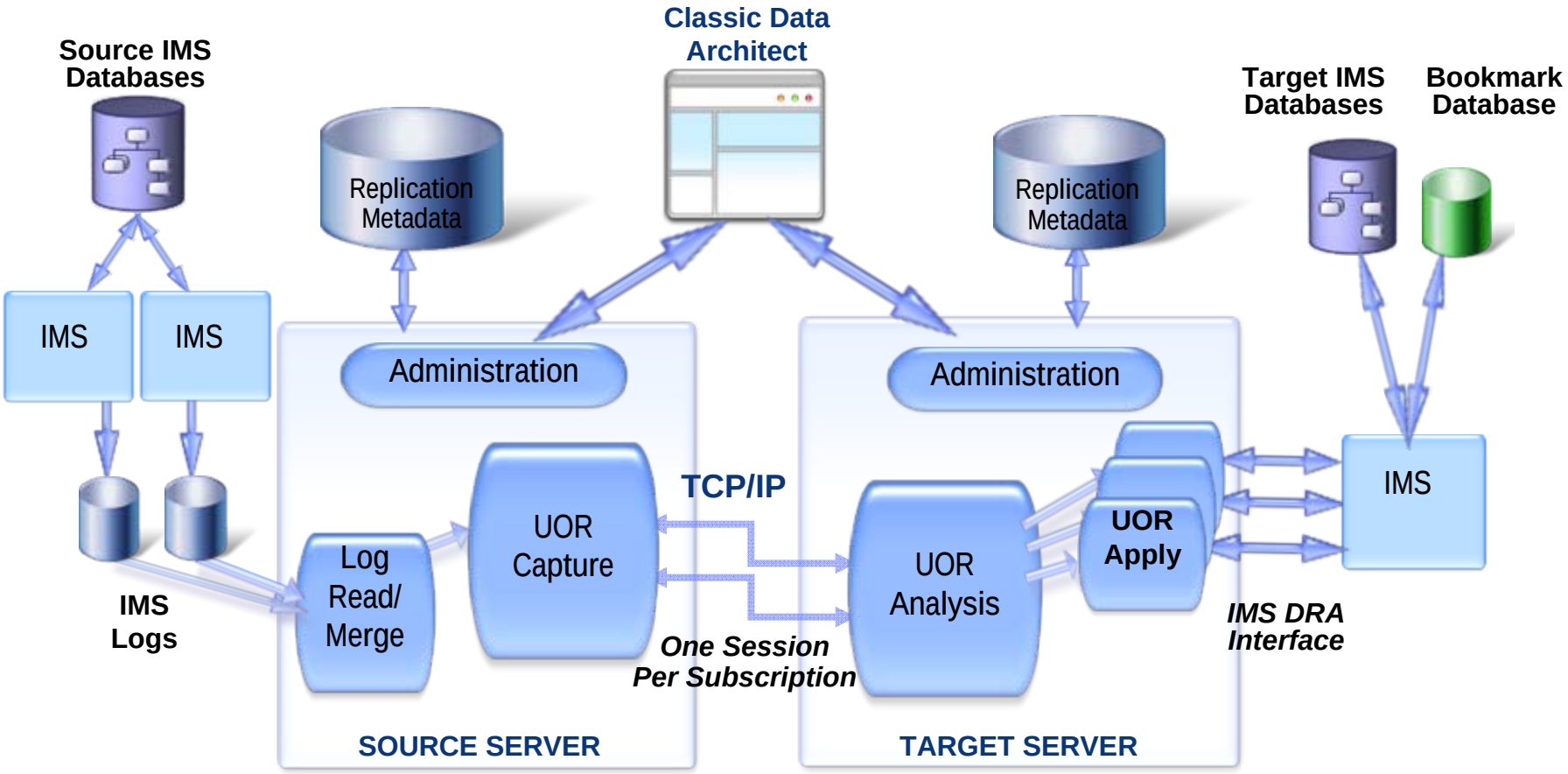
- Background
- GDPS Active/Active Sites solution
- InfoSphere IMS Replication for z/OS

IMS Software-Based Data Mirroring
InfoSphere IMS Replication

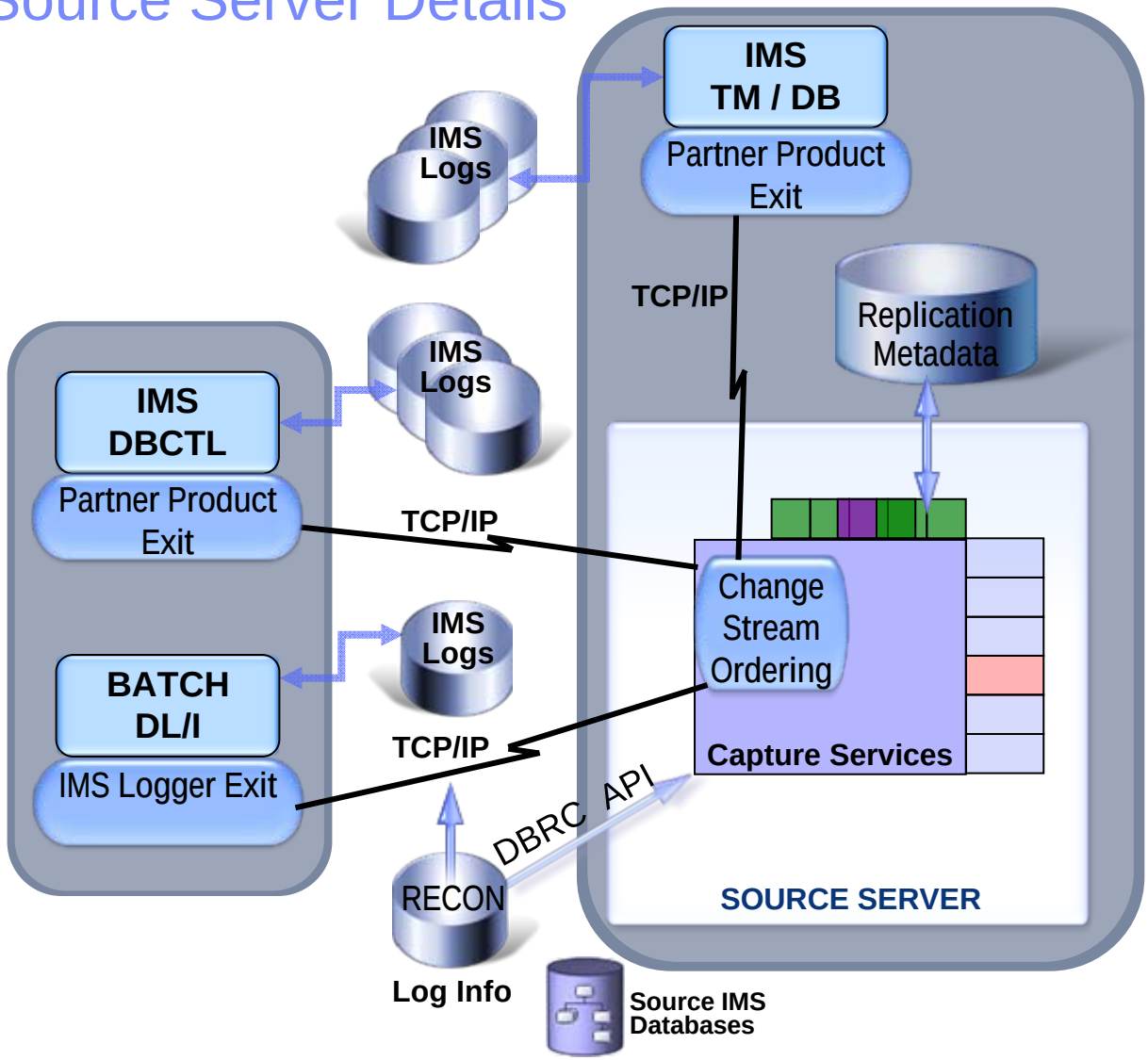
- Unidirectional Replication of IMS data
  - All or nothing at DB level
  - Transaction consistency
  - External initial load of target DB
  - Basic replication monitoring
- IMS “Capture”
  - DB/TM, DBCTL, Batch DL/I, XRF, FDBR
  - Capture x’99’ log records
    - Increase in log volume due to change data capture records
- IMS “Apply”
  - Uses IMS Database Resource Adapter interface
  - Parallel Apply
- Classic Data Architect
  - Administration (some administration can be done via z/OS console commands)
  - Basic replication monitoring



IMS Replication Architecture

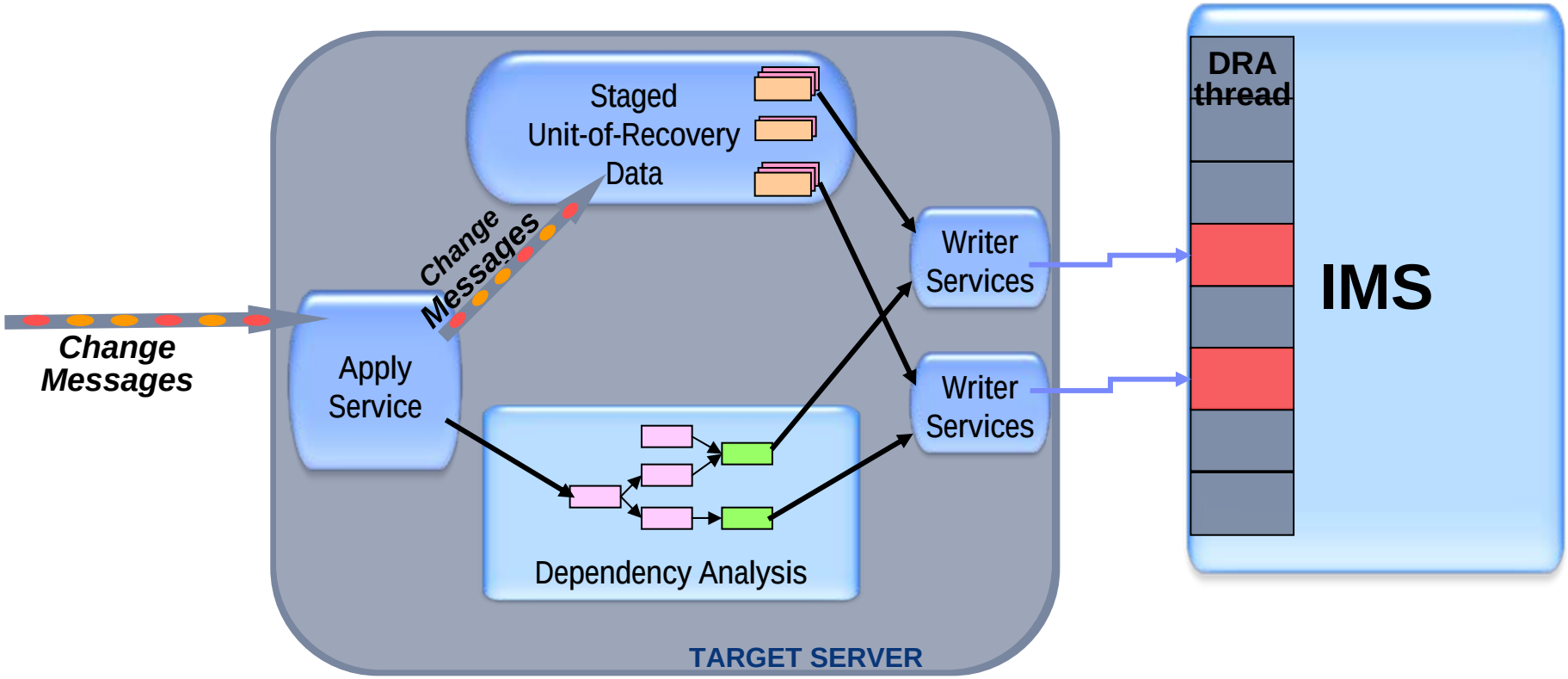


Source Server Details



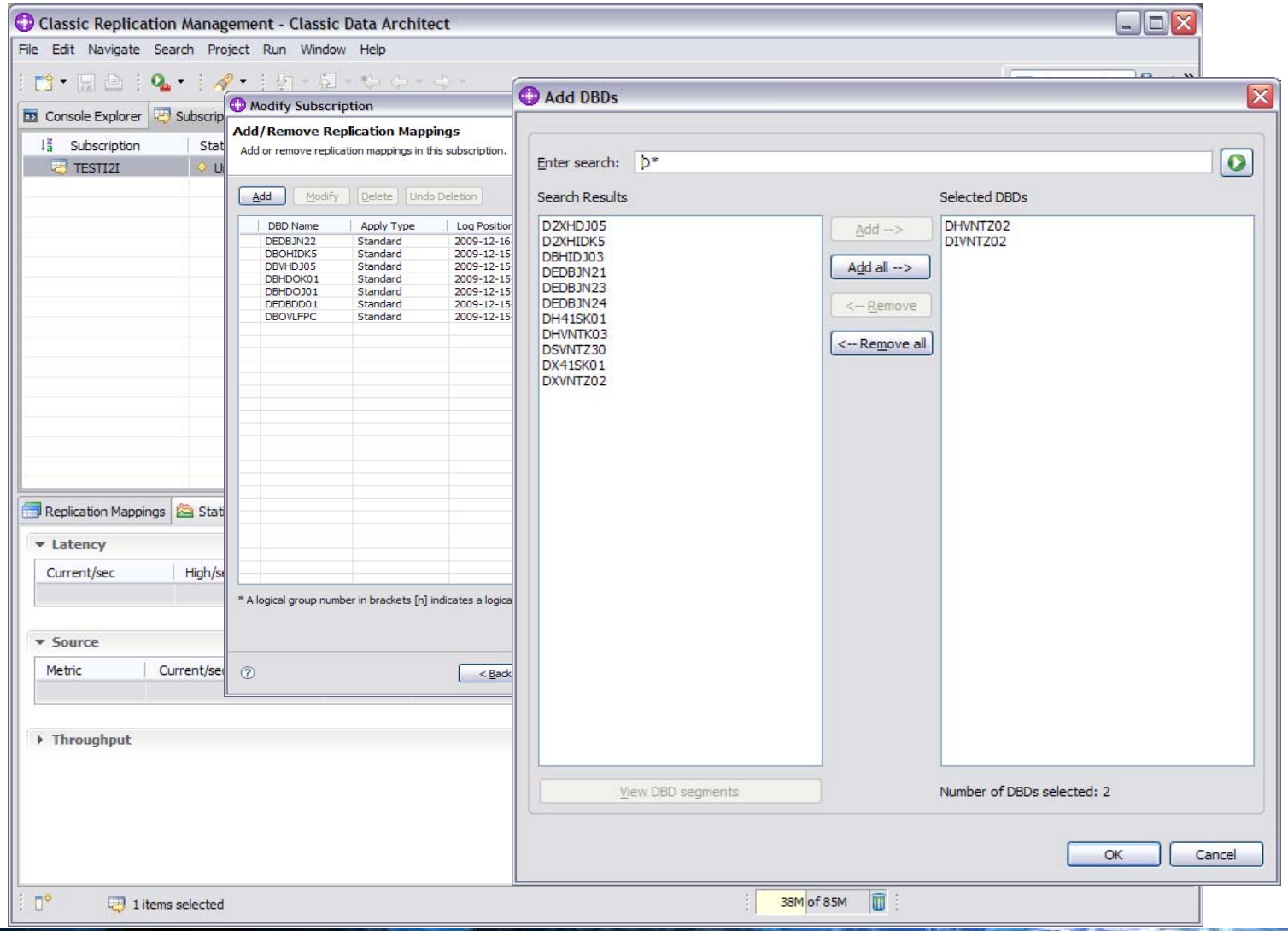
- User exits to notify server of new IMS instance
- Merge Waits for Batch DL/I to complete
- Idle IMS regions can slow processing

Target Server Details

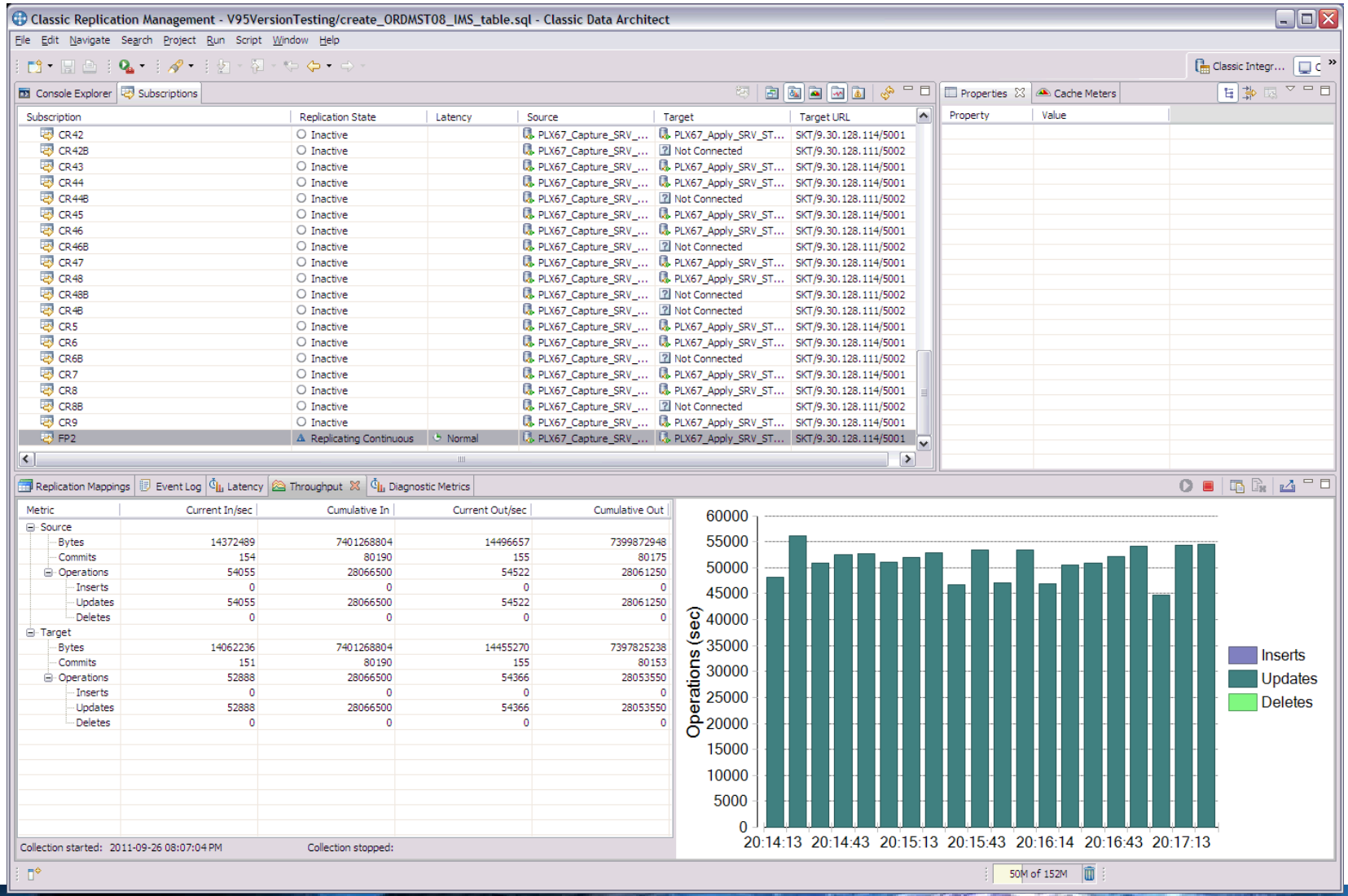


- Parallelism based on dependency analysis within a subscription
- Database and root key used for analysis

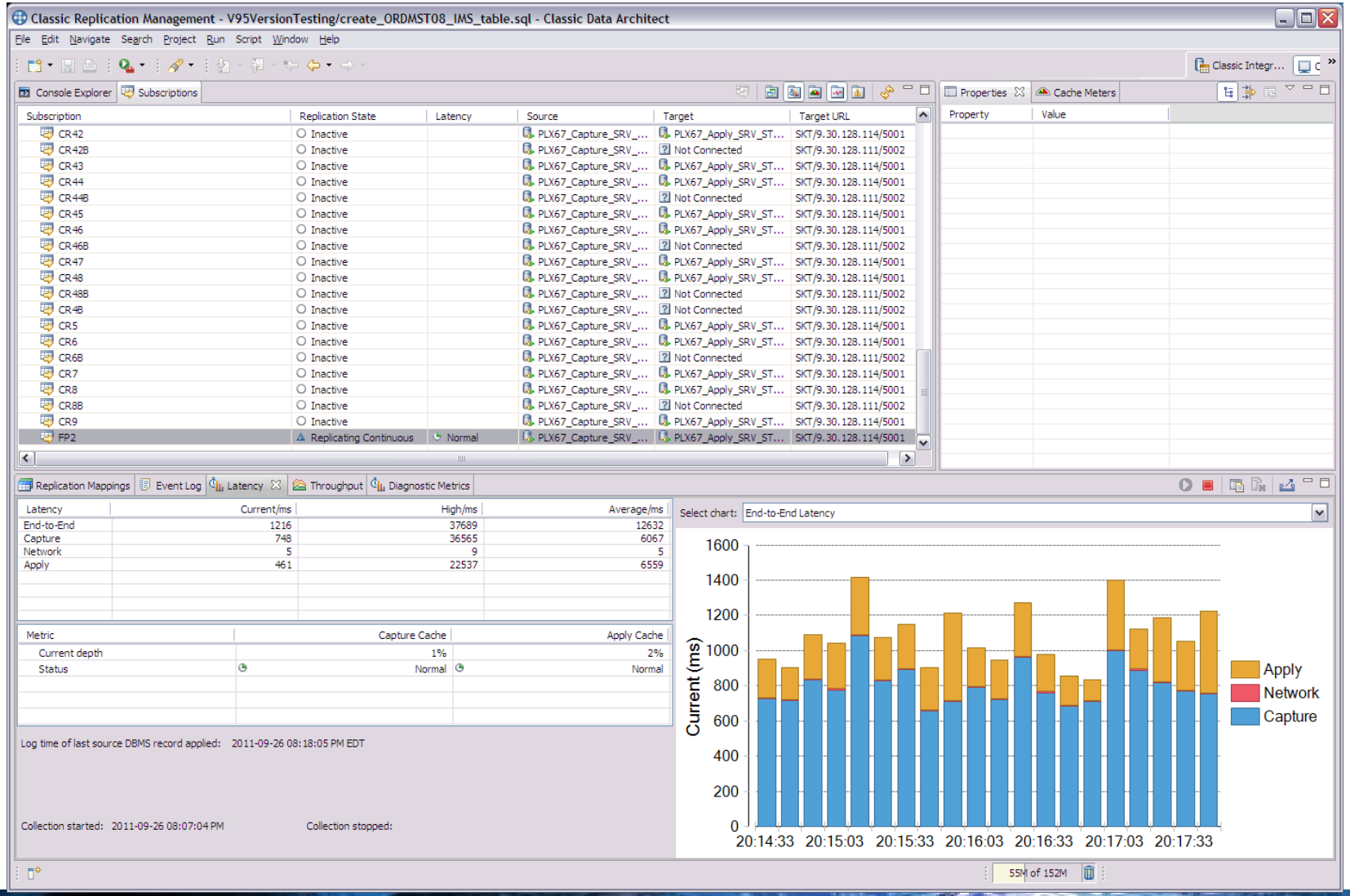
Classic Data Architect – Replication Management



Classic Data Architect - Monitoring Throughput



Classic Data Architect - Monitoring Latency



Adaptive Apply

- Adaptive apply error handling is the default behavior
  - Can be set to standard apply, which does not tolerate conflicts
- If a conflict is detected, the action will be to ignore update
- Conflicts are:
  - Before image mismatch
  - Unable to locate segment to process update
- All conflicts are logged in the event log
    - Manual resolution will be required

Current Restrictions

- All segments for a DB must have change capture logging enabled and will be replicated
  - Must augment the DBD with the EXIT=(…,LOG) specification
  - IMS change capture restrictions
- Workload Restrictions
  - All logically related DBs must be in the same subscription
  - Workload with logically related DBs will be serialized
  - UORs with unkeyed or non-unique keyed segments will be serialized
- External load of target DB
  - Must be a static image copy

Performance considerations

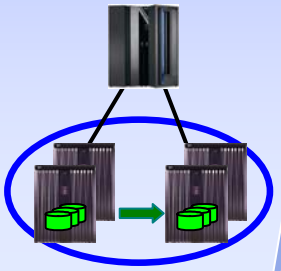
- Transactional consistency vs. Parallelism
  - All updates for a given UR processed as a single transaction during apply
  - All transactions involving the same ‘resource’ will be serially processed in commit order
  - Running transactions in parallel can have application consistency implications
- Increase in log data
- Multiple source IMSs to 1 Apply Target implications
- Internally achieved 53K updates per second
  - ~116,000 updates per second when deploying two apply servers
  - sustained <2sec latency
  - your results may vary

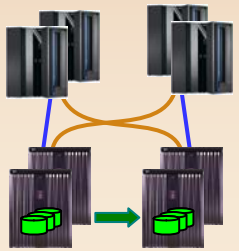
Summary

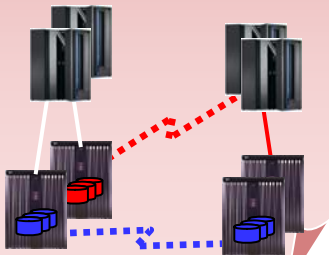
- Asynchronous Replication
  - Allows for unlimited distance support
- Low Latency through parallelism
  - Allows for almost immediate data availability and low RTO
- Transaction Consistency
  - Access with integrity on target system and low RTO
- Subscription independence
  - Switch can be at a workload level vs. system level

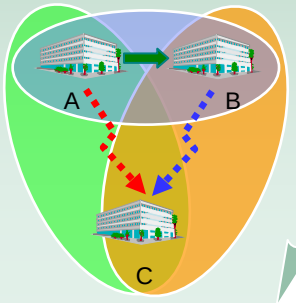
The GDPS " Family "

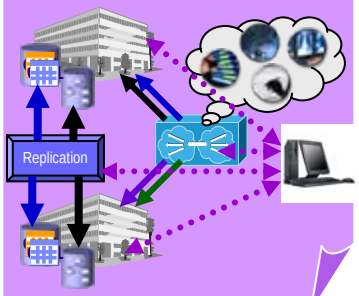
Continuous Availability of Data within a Data Center	Continuous Availability w/ Disaster Recovery within a Metropolitan Region	Disaster Recovery at Extended Distance	Regional Continuous Availability w/ Disaster Recovery @ Extended Distance	Continuous Availability, Disaster Recovery, and Cross-site Workload Balancing at Extended Distance
GDPS/HyperSwap Mgr RPO = 0 / RTO = 0	**GDPS/PPRC** RPO = 0 / RTO <1hr (>20 km) RPO = 0 / RTO = 0 (<20 km)	**GDPS/GM & GDPS/XRC** RPO secs / RTO <1 hr	**GDPS/MGM & GDPS/MzGM**	**GDPS Active-Active Sites** **RPO seconds / RTO seconds**
Single Data Center **Applications remain active** **Continuous access to data in the event of a storage subsystem outage**	Two Data Centers **Systems remain active** **Multi-site workloads can withstand site and/or storage failures**	Two Data Centers **Rapid Systems Disaster Recovery with "seconds" of Data Loss** **Disaster recovery for out of region interruptions**	Three Data Centers **High availability for site disasters** **Disaster recovery for regional disasters**	Two or More Data Centers **All sites active**











Components

Tivoli – NetView, SAz System z, DS8K, PPRC GDPS control code, Services	**Tivoli** – NV, SAz, SA-MP, AppMan System z, DS8K, VTS, PPRC GDPS control code, Services	**Tivoli** – NV, SAz System z, DS8K, Global Mirror, XRC GDPS control code, Services	**Tivoli** – NV, SAz System z, DS8K, MGM, MzGM GDPS control code, Services	**Tivoli** – SA, NetView Multi-site Workload Lifelife DB2 &IMS replication System z, DS8K, Global Copy GDPS control code, Services

Thank You for Joining Us today!

Go to www.ibm.com/software/systemz/events/calendar to:

- Replay this teleconference
- Replay previously broadcast teleconferences
- Register for upcoming events