Oracle on System z Linux- High Availability Options

Session ID 252

Sam Amsavelu

IBM

Trademarks

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

AIX*	FICON*	IMS	Power7*	Redbooks*	WebSphere*
BladeCenter*	IBM*	InfiniBand	PowerHA	RMF	zEnterprise*
CICS*	IBM (logo)*	Lotus*	Power Systems	System x*	z/OS*
Cognos*	GDPS*	MQSeries*	PowerVM	System z*	z/VM*
DataPower*	Geographically Dispersed Parallel Sysplex	Parallel Sysplex*	PR/SM	System z10*	z/VSE*
DB2*	HiperSockets	POWER*	PureSystems	Tivoli*	
DS8000*	HyperSwap	POWER4*	Rational*		

* 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.

IT Infrastructure Library is a registered trademark of the Central Computer and Telecommunications Agency which is now part of the Office of Government Commerce.

Intel, Intel logo, Intel Inside, Intel Inside logo, Intel Centrino, Intel Centrino logo, Celeron, Intel Xeon, Intel Speed Step, Itanium, and Pentium are trademarks or registered trademarks of Intel Corporation or its subsidiaries in the United States and other countries.

Linux is a registered trademark of Linus Torvalds 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.

Windows Server and the Windows logo are trademarks of the Microsoft group of countries.

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.

UNIX is a registered trademark of The Open Group in the United States and other countries.

Java and all Java based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates.

Oracle and Java are registered trademarks of Oracle and/or its affiliates

Cell Broadband Engine is a trademark of Sony Computer Entertainment, Inc. in the United States, other countries, or both and is used under license therefrom.

Linear Tape-Open, LTO, the LTO Logo, Ultrium, and the Ultrium logo are trademarks of HP, IBM Corp. and Quantum in the U.S. and other countries.

* Other product and service names might be trademarks of IBM or other companies.

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.

Notice Regarding Specialty Engines (e.g., zIIPs, zAAPs and IFLs):

Any information contained in this document regarding Specialty Engines ("SEs") and SE eligible workloads provides only general descriptions of the types and portions of workloads that are eligible for execution on Specialty Engines (e.g., zIIPs, zAAPs, and IFLs). IBM authorizes customers to use IBM SE only to execute the processing of Eligible Workloads of specific Programs expressly authorized by IBM as specified in the “Authorized Use Table for IBM Machines” provided at

www.ibm.com/systems/support/machine_warranties/machine_code/aut.html (“AUT”).

No other workload processing is authorized for execution on an SE.

IBM offers SEs at a lower price than General Processors/Central Processors because customers are authorized to use SEs only to process certain types and/or amounts of workloads as specified by IBM in the AUT.

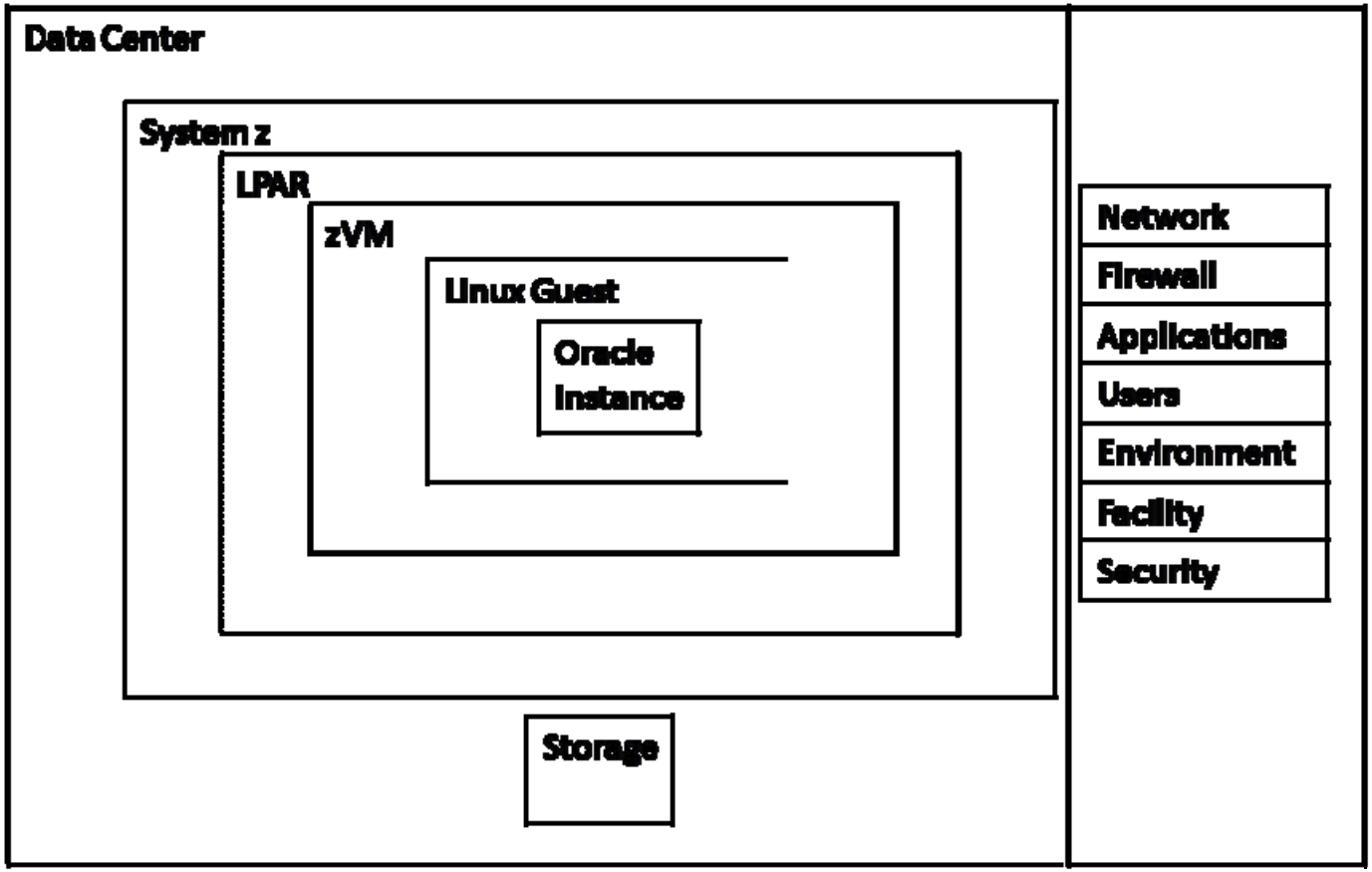
Agenda

- What is High Availability?
- Oracle technologies for HA
- zVM technologies for HA
- Disaster recovery considerations
- Questions

What is High Availability?

- Key component of business resiliency
- High availability solutions always involve redundancy
- Basic rule of high availability is to identify and remove single points of failure in the architecture
- Reliability of data and performance also HA
- Measured by the perception of an end user
- It is not continuous availability (Five 9s)
- Designed to minimize the loss of service

Typical Oracle zVM architecture



Planned outages

- Software maintenance and upgrades
- Hardware refreshes
- Data center relocations
- Building maintenances
- Largest share of time that a database is rendered unavailable is due to planned maintenance activities
- Can be coordinated with end users in advance to minimize the user impact

Unplanned outages

- Datacenter
  - Natural disasters, power failures
- Hardware Components
  - CPU, memory, storage, cable
- System Software Components
  - zVM, Linux
- Application Components
  - Application logic, bottleneck
- Oracle
  - Instances, components, **performance**

Fundamental measurements

- Service Level Agreement (SLA)
  - SLA is the agreed upon process time as well as downtime for an application with the end user.
- Recovery Time Objective (RTO)
  - RTO is the maximum period of time for which a disturbance to an application or a process can be tolerated. RTO is inclusive of issue identification, response and issue resolve time.
- Recovery Point Objective (RPO)
  - RPO is the maximum amount of time data loss can be tolerated. For some, it may be zero data loss like in a stock trading application, and in others it may be a couple of hours worth of data.

High level solutions

- Infrastructure resiliency
  - Reliable, redundant, as well as clustered components
- Application resiliency
  - Load balancing, stateless designs
- Data resiliency
  - Physical (storage mirroring
  - Logical (database technologies)

Oracle Databases

- What to be protected
  - Physical (database files)
  - Logical (data corruption)
  - Storage (mirroring, maintenance)
  - Instance (Restart)
  - Server (Infrastructure)
  - Datacenter (DR)
- Build one brick at a time

Oracle technologies for HA

- Backup-recovery
- Flashback
- ASM
- Grid infrastructure
- Oracle RAC One , Oracle RAC
- Application continuity
- Data Guard
- GoldenGate

Backup-Recovery

- Foundation for any robust Oracle highly available environment
- For many customers this is the only option
  - But still not enough considerations provided
- Can be performed at physical or logical level
- Consistent (cold) or inconsistent (hot) backups
- RMAN is ‘the’ utility
- IBM Tivoli, Netbackup vendors provide value added services with RMAN integration

Flashback

- Backup is for physical, media errors etc., whereas Flashback is for logical data errors
- Human errors like an authorized user executing a wrong query and deleting rows in the tables or corrupting some data can be corrected.
- Supports recovery at the row, transaction, table, and the entire database level.
- Extremely fast compared to traditional backup and recovery process as it only restores blocks that have changed

Flashback -cont'd

- Flashback database
  - The Database can be rewound based on SCN, timestamp or restore points.
- Flashback drop
  - The dropped table, and all of its indexes, constraints, and triggers are recovered from the Recycle Bin
- Flashback table
  - The logically corrupted table can be restored to a specific point-in-time.
- Flashback Query
  - Using Oracle Flashback Query, users can query any data at some point-in-time in the past.

ASM

- Integrated database file system and disk manager
- Mirroring the data at file level
  - Optionally ASM supports 2-way mirroring, where each file extent gets one mirrored copy, and 3-way mirroring, where each file extent gets two mirrored copies.
  - Mirrored copy is kept at a disk other than the original copy disk
- Dynamic addition of disks and removal facility
  - Improves the storage availability
- Simplified administrative tasks
  - reduces the complexity of managing thousands of files in a in a large environment

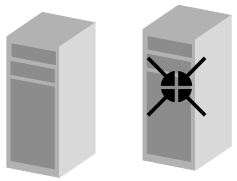
Grid Infrastructure

- Allows clustering of independent servers so that they cooperate as a single system
  - If a clustered server fails, then any managed application can be restarted on the surviving servers.
  - The managed applications can be like Siebel, GoldenGate, WebSphere® or even Oracle databases.

- The applications are protected in active / passive environment
  - Built-in agents to start at the primary node or at other nodes
  - Monitoring frequency, starting, and stopping of the applications and the application dependencies - all can be configured

Grid Infrastructure -cont'd

- HADR applicability



  - Economical
  - Downtime can be from seconds to minutes (restart time)
  - Protection from Computer hardware failures
  - Protection from OS (Linux / zVM) failures
  - Protection from Oracle instance failures
  - Active / Passive implementation, so recovery is **not instantaneous**

Oracle RAC One Node

- Oracle RAC One Node is a single instance of an Oracle RAC database that runs on one node in a cluster
  - In case of sudden failure of the active instance, Oracle RAC One Node will detect the failure, and either restart the failed database or fail it over to another server.

- Uses ‘omotion’ technology to relocate the instance without any downtime and does not need manual intervention
  - During the short period of time when the instance is moved from one node to another, both instances are active. Once all the connections are migrated the first instance goes down.

Oracle RAC One -cont'd

- HADR applicability
  - Protection from Computer hardware failures
  - Protection from OS (Linux / zVM) failures
  - Protection from Oracle instance failures
  - Protection from storage failures (when ASM is used)
  - In planned outages it is possible to have continuous availability of Oracle instances

Oracle RAC

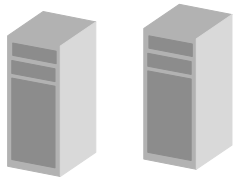
- Oracle RAC technology allows multiple Oracle instances running across multiple nodes to access the same database providing a single logical instance view

- With Oracle RAC all nodes are active and it enables the continuous availability of Oracle instance

- Oracle Extended RAC is an architecture where the nodes in the cluster are separated into different data centers

Oracle RAC -cont'd

- HADR applicability
  - Protection from Computer hardware failures
  - Protection from OS (Linux / zVM) failures
  - Protection from Oracle instance failures
  - Protection from storage failures (when ASM is used)
  - Active / Active configuration and hence continuous availability
  - Fast application notification (FAN) with integrated Oracle client failover
  - Complex, expensive solution



Oracle Application failover

- When a database outage occurs the applications may encounter errors or hangs.

- Oracle's high availability features address these by providing APIs to automate client failover

  - Fast Application Notification (FAN)
  - Fast Connection Failover (FCF)
  - Transparent Application Failover (TAF)

Oracle Application -cont'd

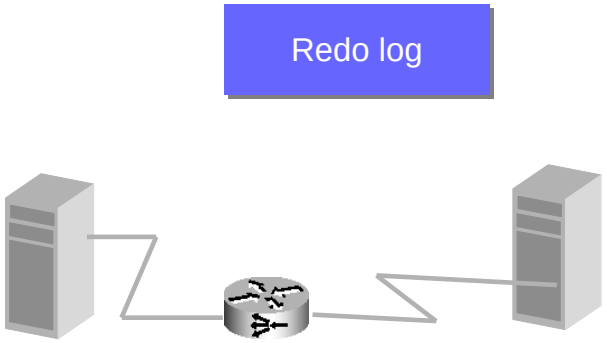
- HADR applicability
  - A server failure, Linux crash or other faults can causes the crash of an individual Oracle instance in an Oracle RAC database
  - Relocate the database services to new or surviving instances
  - Notify the clients that a failure has occurred
  - Redirect the clients to the relocated or a surviving instance

Oracle Data Guard

- Configuration consists of one primary database and one or more (up to 30) standby databases.
- Maintains standby databases as transactionally consistent copies of the primary database
- If the primary database becomes unavailable, Oracle Data Guard can switch any standby database to the primary role, minimizing the downtime associated with the outage

Oracle Data Guard --cont'd

- A standby database can be
  - A physical standby database (exact)
  - Oracle Active Data Guard
  - A transient logical standby database
    - A snapshot standby database
      - (For testing, cloning)
    - A logical standby database
      - (From redo log SQL created and SQLApply)



Oracle Data Guard --cont'd

- HADR applicability
  - SWITCH OVER
  - Addresses both High Availability and DR requirements
  - Data Guard technology complements with Oracle RAC
  - Compared to RAC, Oracle Data Guard architecture has one or more synchronized standby databases. That ensure protection of data from failures, disasters, errors, and corruptions

Oracle GoldenGate

- Oracle GoldenGate is an asynchronous, log-based, real-time data replication technology

- Moves data across heterogeneous database, hardware, and operating system environments

- Supports multi master replication, hub-and-spoke deployment, and data transformation

- Can be deployed for data distribution and data integration

Oracle GoldenGate--cont'd

- HADR applicability
  - Maintains transactional integrity
  - Resilient against interruptions and failures
  - Heterogeneous replication, transformations, multiple topologies
  - All sites fully active (read/write)

Infrastructure Resiliency

- System z continues to build on decades of design of both hardware and software to keep the system up and running to provide the highest availability, and is unmatched with 99.999% reliability.

- zVM powerful hypervisor offers the following
  - Mature technology (41 years)
  - Software hypervisor integrated with hardware
  - Effective sharing of
    - CPU, memory, and I/O resources
    - Virtual network - virtual switches and routers

What can go wrong with zVM?

- Planned downtime activities
  - System z hardware upgrades requiring Power On Reset (POR).
  - LPAR configuration changes requiring reboot of the LPAR
  - z/VM maintenance activities

- Unplanned outages
  - The System z hardware failures
  - The network / connectivity failures
  - Disk subsystem I/O channels failures
  - The LPAR microcode could fail
  - zVM failures

zVM Technologies

- Virtualization (CPU, memory, I/O, network)
  - Dynamically add CPU, memory resources at the Linux
- Ability to define the ‘share’ of CPU resources
- Hipersockets (used as interconnect)
- VSWITCH under z/VM and OSA Channel Bonding
  - Network redundancy
- ECKD DASD devices accessed over FICON channels, redundant multi-pathing is provided

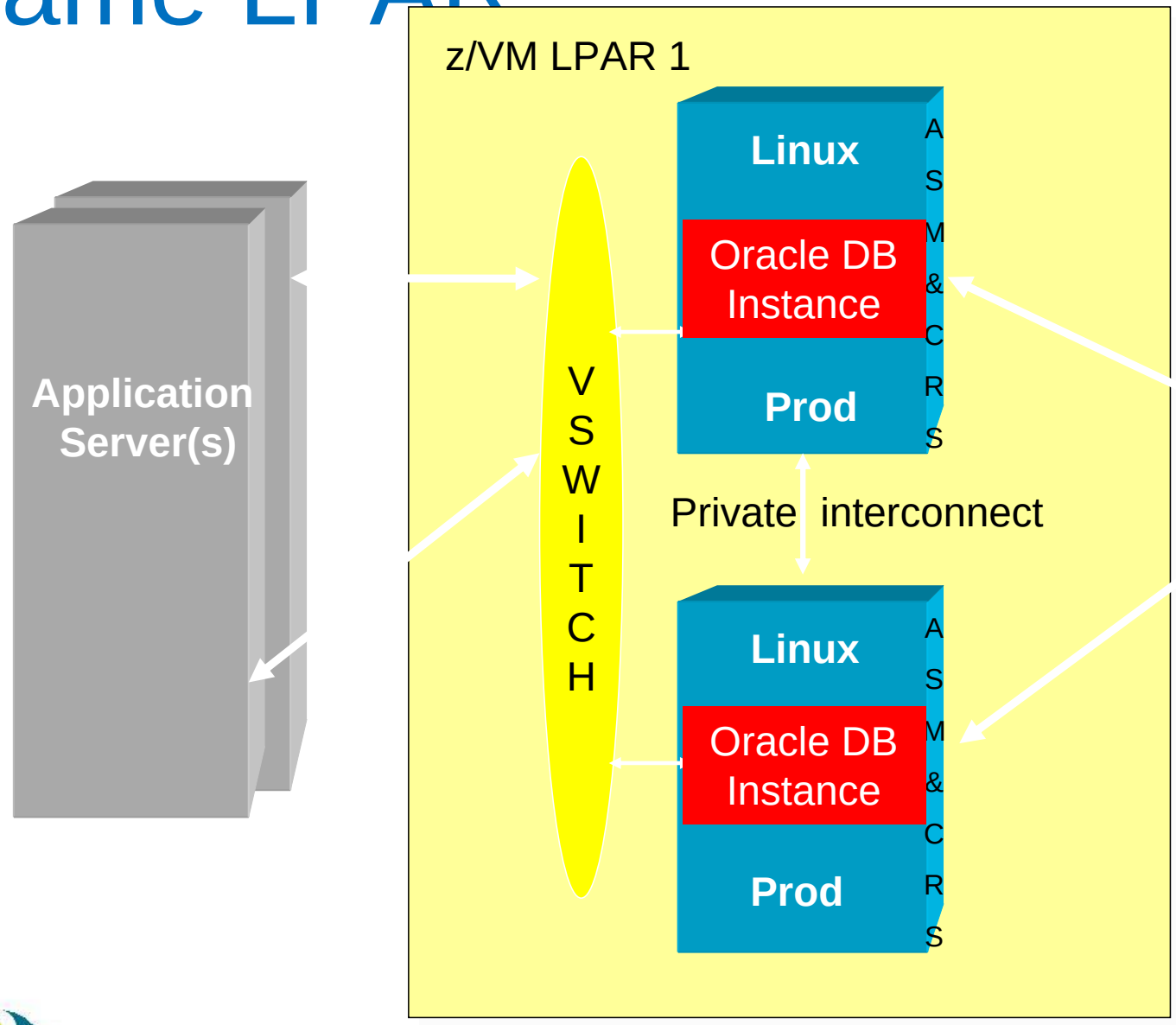
zVM SSI

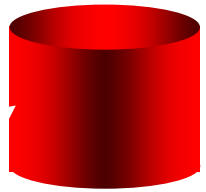
- New clustering from zVM 6.2
- The idea is to relocate ‘live’ any work load from one SSI member to other member
-

Customer patterns on z

- A clustering solution implemented
  - Across LPARs in the same System z Box
  - Across two System z
  - Active / Passive
  - Active / Active

Same LPAR



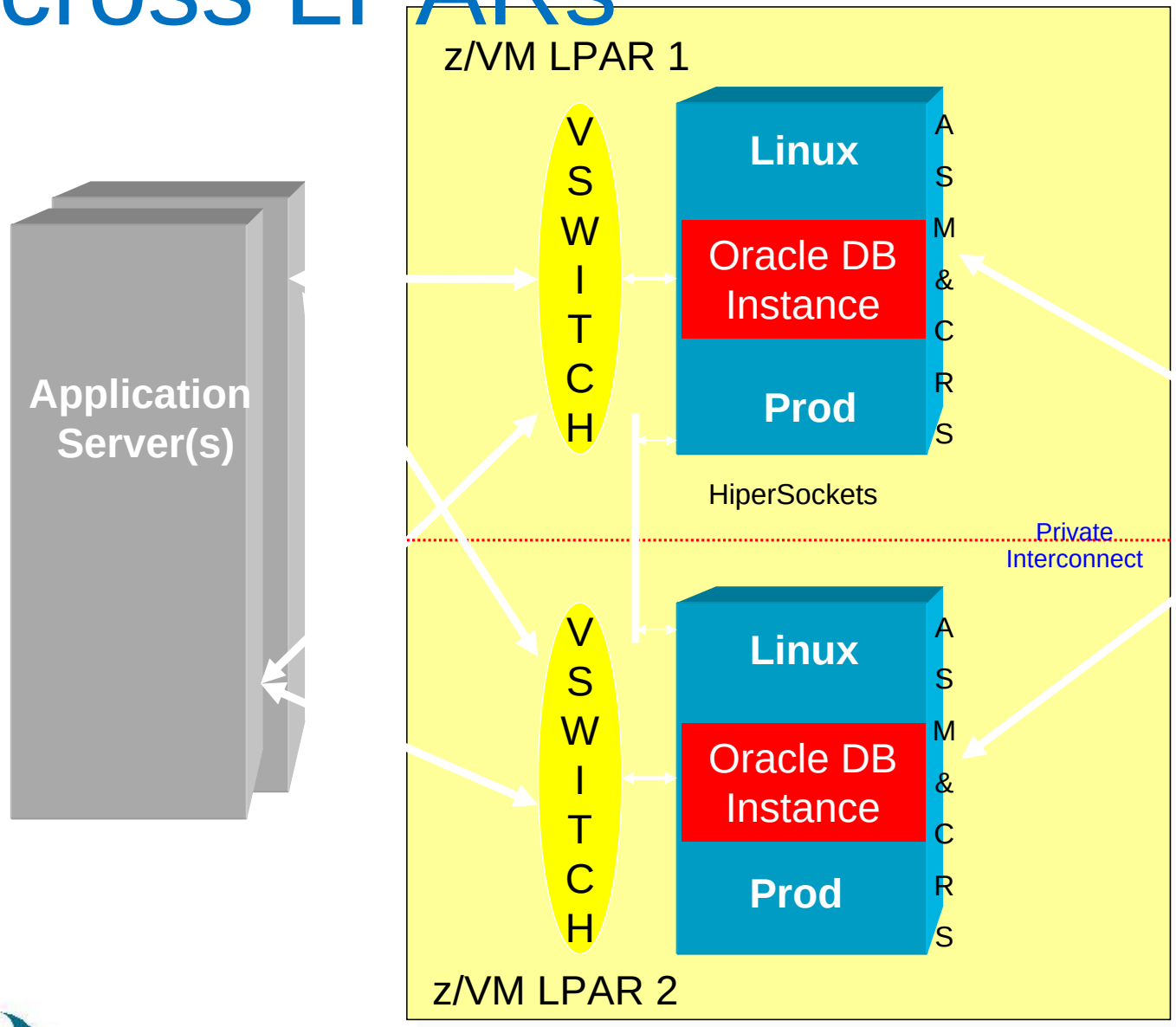


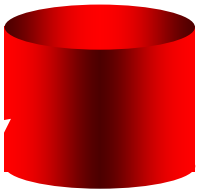
Oracle Database

- **All IFLs and memory can be shared across nodes.**

Interconnect - The private network communication link that is used to synchronize the memory cache of the nodes in the cluster. Can be a VSWITCH, Linux channel bonding or Oracle HAIP.

Across LPARs



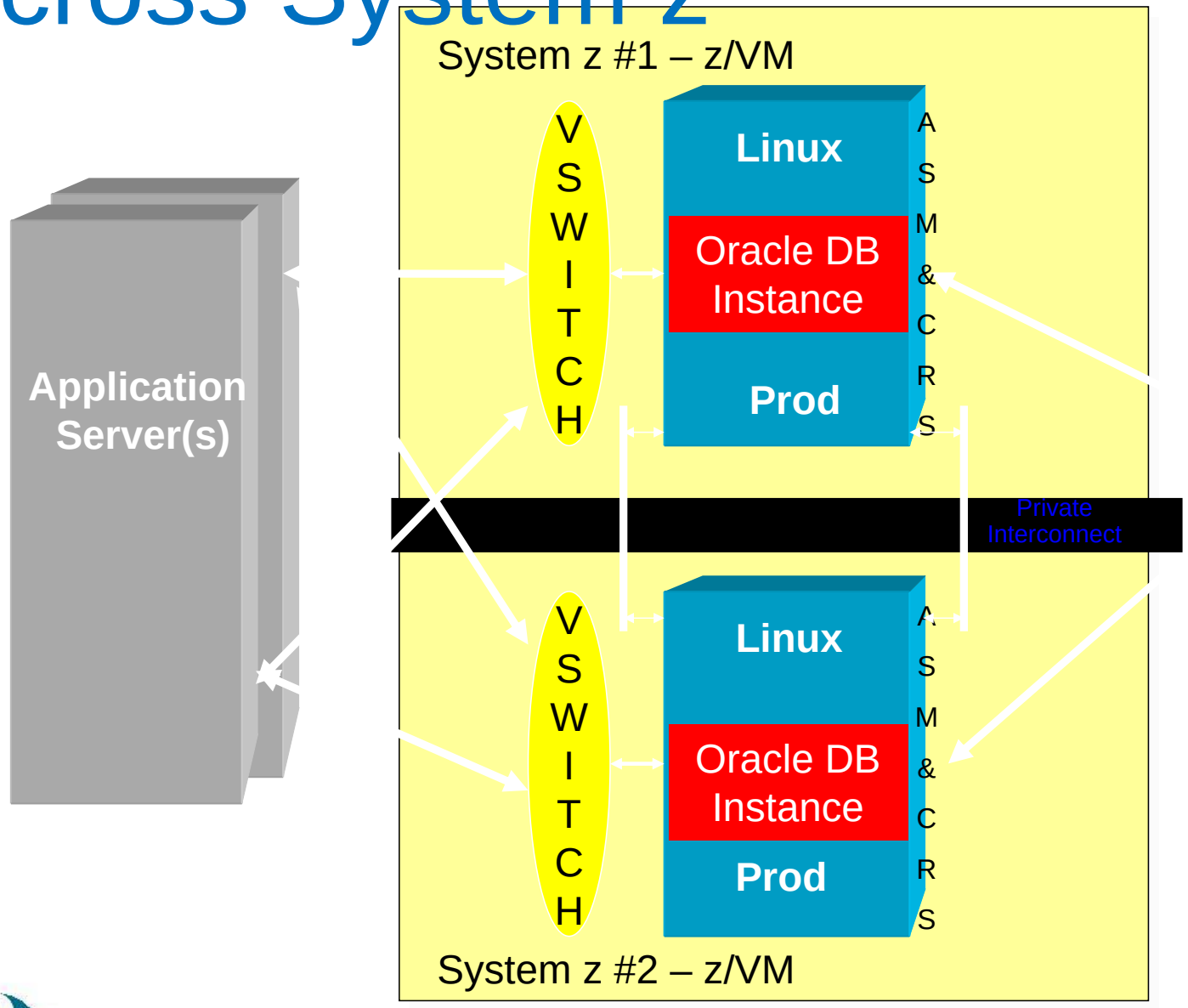


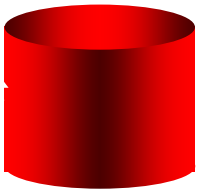
Oracle Database

Comments:

- **IFLs can be shared across LPARs but memory cannot.**
- **HiperSockets is a memory based technology that provides high-speed TCP/IP connectivity between LPARS within a System z.**

Across System z





Oracle Database

Comments:

- **Physically separate System z machines**

Customer patterns on z

- A clustering solution implemented for 300 databases
  - Across two System z (Active - Passive)
  - Active / Passive
  - 5 node cluster
    - 4 nodes on one System z with 15 databases each
    - The other node on another System z – backup node
  - The second set of cluster in the Passive z

Disaster Recovery

- Disaster Recovery solutions are an extension of high availability solutions with the added capability of providing resiliency with geographic dispersion (dual site concept)

- Disaster Recovery configuration that is completely identical across tiers on the production site and standby site is called a symmetric site

- An ideal DR solution should be highly reliable, less complex in design, utilize proven technologies and be less expensive to implement

DR Challenges

- DR solutions are expensive
- The redundant systems are never utilized or under utilized in many situations
- Very hard to test if it really works
- No ROI until a disaster occurs
- Hardware, software maintenance still needed
- Long distance across data centers create data synchronization challenges

DR Technologies

- Many of the System z customers have well established business processes for DR scenarios. Their current DR environments can be extended to include Oracle databases running on the System z Linux environment also

- For Oracle databases the major requirement for DR is data resiliency and can be achieved by any of the following technologies
  - Storage array based remote mirroring solutions
  - Extended cluster solutions (Extended RAC)
  - Oracle Data Guard based solutions

Oracle MAA and DR

- Oracle MAA recommends to build a DR solution based on Oracle Data Guard technology for Oracle databases for the following reasons:
  - Automatic and fast switch-over
  - Transactionally consistent data
  - Detection and deletion of data corruptions
  - Application, system vendor or storage agnostic
  - Planned downtime reduction by using database rolling upgrades.

Summary

- We discussed the following items to successfully plan a highly available and disaster recovery (HADR) environment for Oracle Databases running on System z Linux in a virtualized environment.
  - Single instance - including ASM
  - Clustering Active / Passive solution
  - RAC One node
  - Multi-node RAC on one LPAR
  - Multi-node RAC on more than on LPAR on one CEC
  - Multi-node RAC on multiple CECs
  - Use of Data Guard
  - Use of GoldenGate

Questions?
Comments?

THANK YOU

IBM Advanced Technical Skills

Pre-sales technical support for Oracle on System z technologies

samvelu@us.ibm.com

COLLABORATE 13

Flashback -cont'd

- Flashback Versions Query
  - Using Oracle Flashback versions Query, users can retrieve different versions of a row across a specified time interval
- Flashback Transaction Query
  - Flashback Transaction Query shows the changes made by a transaction and also produces the SQL statements necessary to flashback or undo the transaction.
- Flashback Transaction
  - A single transaction, and optionally, all of its dependent transactions, can be flashed back