PDUG

Kraków, 28-29 May 2025

Db2 Always On Architecture

The journey to keeping the lights on

Dale McInnis

IBM Canada Ltd.
dmcinnis@ca.ibm.com

POLISH DB2 USERS GROUP

WHO AM I

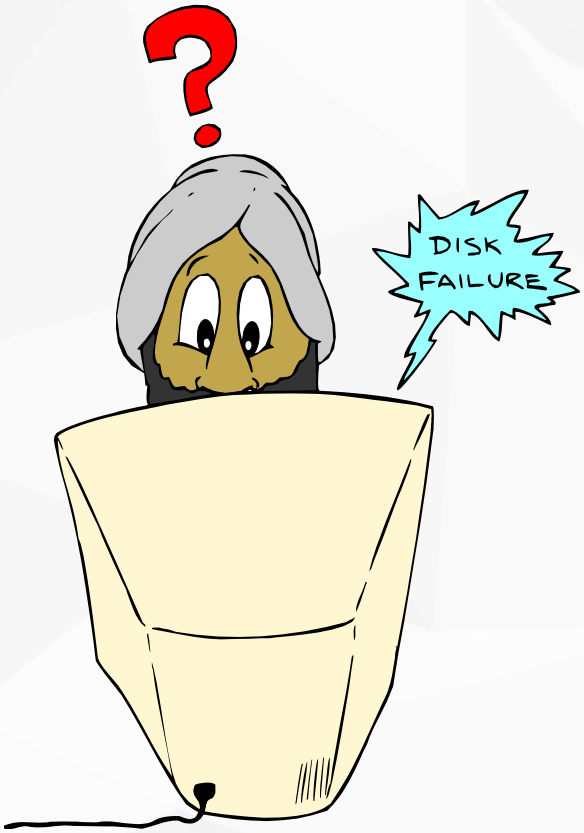
- Certified Technical Specialist (level 3) on the America's Data and AI Technical Sales team
- 37 years with IBM, 25 years at the Db2 LUW High Availability Architect
- Vice Chair for Canadian Information Processing Society (CIPS)– Ontario Canada
- Information Systems Professional (I.S.P) Certified
- Information Technology Certified Professional (ITCP)
- IBM Worldwide Performance and Availability Community co-leader
- IBM Academy of Technology Member / Open Innovation Community
- IBM Academy PREVAIL Conference Lead
- 11 Patents in database resiliency
- Published author
- IDUG Speaker Hall of Fame member
- Ph.D. Student at the Ontario Technical University

PDUG

Agenda

- Why “always on”
- What are the Db2 options?
- Combining technologies for the best resiliency
- The Journey to Always On

What is High Availability?



- GOAL: Minimize the affect of a resource which is temporarily unavailable
- Short term in nature
- Localized to a single site
- May be the result of a planned or unplanned outage
- Achieved by utilizing redundant components

What is Disaster Recovery?

- GOAL: Minimize the affect of a system wide or site wide resource unavailability
- Longer term in nature
- Affecting one or more critical resources
- Always unplanned

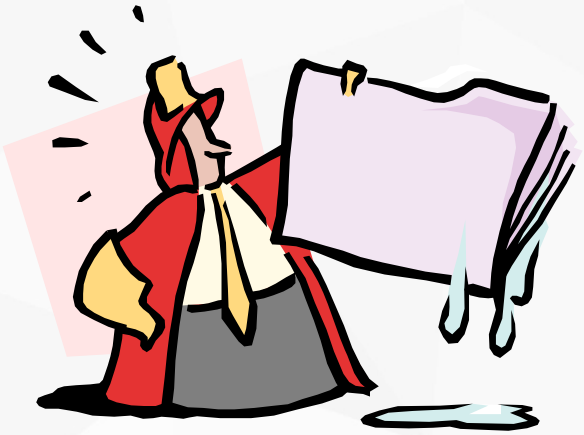
What is RTO?

PDUG

- The **Recovery Time Objective** (RTO) is the duration of time and a service level within which a business process must be restored after a disaster (or disruption) in order to avoid unacceptable consequences associated with a break in business continuity
- It should be noted that the RTO attaches to the business process and not the resources required to support the process.

What is RPO?

PDUG



Recovery Point Objective (RPO) describes the acceptable amount of data loss measured in time.
The Recovery Point Objective (RPO) is the point in time to which you must recover data as defined by your organization. This is generally a definition of what an organization determines is an "acceptable loss" in a disaster situation

The impact of *Downtime* to your Business

88% of organizations require that their databases deliver a minimum of 99.99% or better uptime for their most mission critical applications *

* ITIC Paper
ITIC 2022 Global Server Hardware, Server OS Security Report

Negative Press

Damaged Reputation

Lost Revenue

Loss of market share to competitors

pureScale

Availability	Downtime per Year
99.999%	5 minutes
99.99%	52 minutes
99.9%	8 hours, 45 minutes
99%	3.65 days
95%	18.25 days

Reduced IT / Employee Productivity

Customer dissatisfact-ion

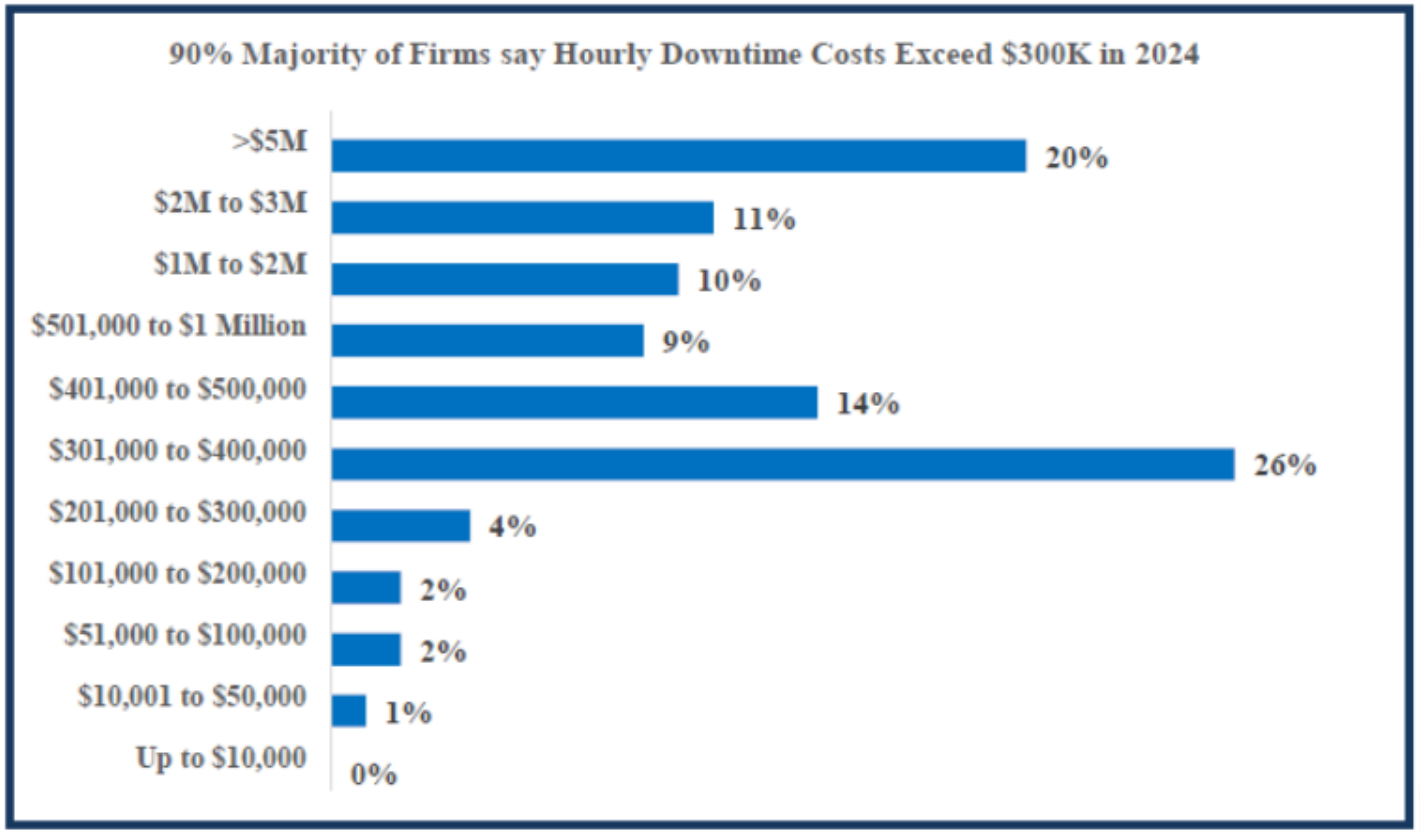
Breach of SLA / Contract

Reduced Company / Shareholder Value

Reduced End-User Productivity

Cost of downtime

Exhibit 2. Nine-in-10 Companies Report Hourly Downtime Costs Top $300K.

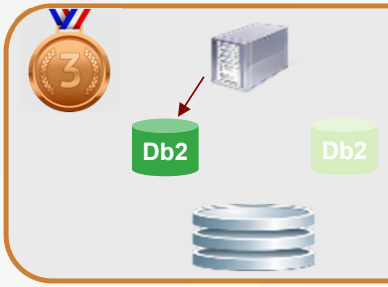


Source: ITIC 2024 Hourly Cost of Downtime Report

Agenda

- Why “always on”
- What are the Db2 options?
- Combining technologies for the best resiliency
- The Journey to Always On

DB2 HA OPTIONS : 24/7/365 ZERO DATA LOSS FOR OLTP

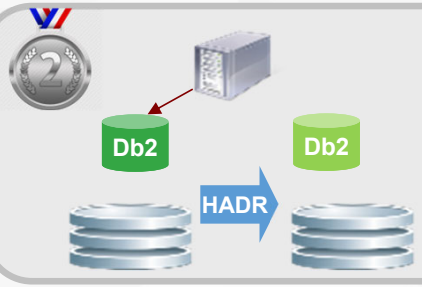


Integrated Clustering

- Active/passive
- Hot/cold, with failover typically in minutes
- Easy to setup

99% Availability

- Db2 ships with integrated TSA failover software
- TSA replaced with Pacemaker on Linux
  - AIX to follow
- No additional licensing required

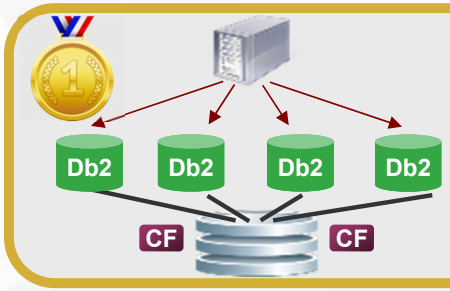


HADR

- Active/passive or active/active (with Reads on Standby)
- Hot/warm or hot/hot (with RoS), with failover typically less than one minute
- Easy to setup

99.99% Availability

- Db2 ships with integrated TSA
- Minimal licensing (full licensing required if standby is active)
- Perform system and database updates with miminal interruption



pureScale

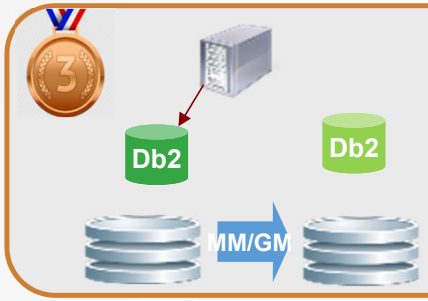
- Active/active
- Hot/hot, with automatic and online failover
- Integrated solution includes CFs, clustering, and shared data access

99.99+% Availability

- Perform system and database updates in rolling online fashion
- Also works with HADR (single target)
- Geographically Dispersed Cluster for multi-site deployment

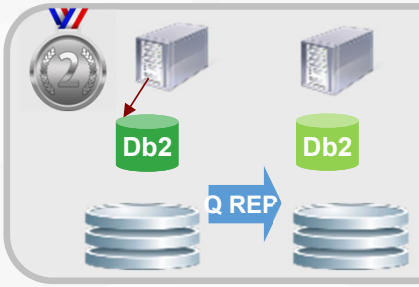
DB2 DISASTER RECOVERY OPTIONS

PDUG



Log Shipping / Storage Based Replication

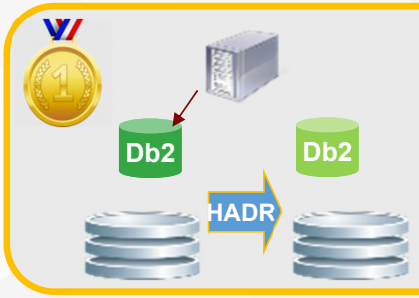
- Active/passive
- Hot/cold, with failover typically in minutes
- Asynchronous
- Complete DB replication only



Logical Replication

- Active/active (updates require conflict resolution / avoidance
- Hot/Hot (Instant failover)
- Asynchronous

- Added flexibility
  - Subsetting
  - Different versions
  - Different topology
  - Multiple standby
  - Time delay
- DDL considerations



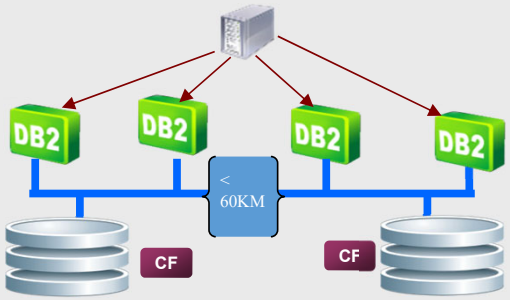
HADR

- Active/passive or active/active (with Reads on Standby)
- Hot/warm or hot/hot (with RoS), with failover typically less than one minute
- Easy to setup

- Complete DB Replication
- Minimal licensing (full licensing required if standby is active)
- Time Delay
- Perform system and database updates minimal interruption

Situational Platinum

GDPC

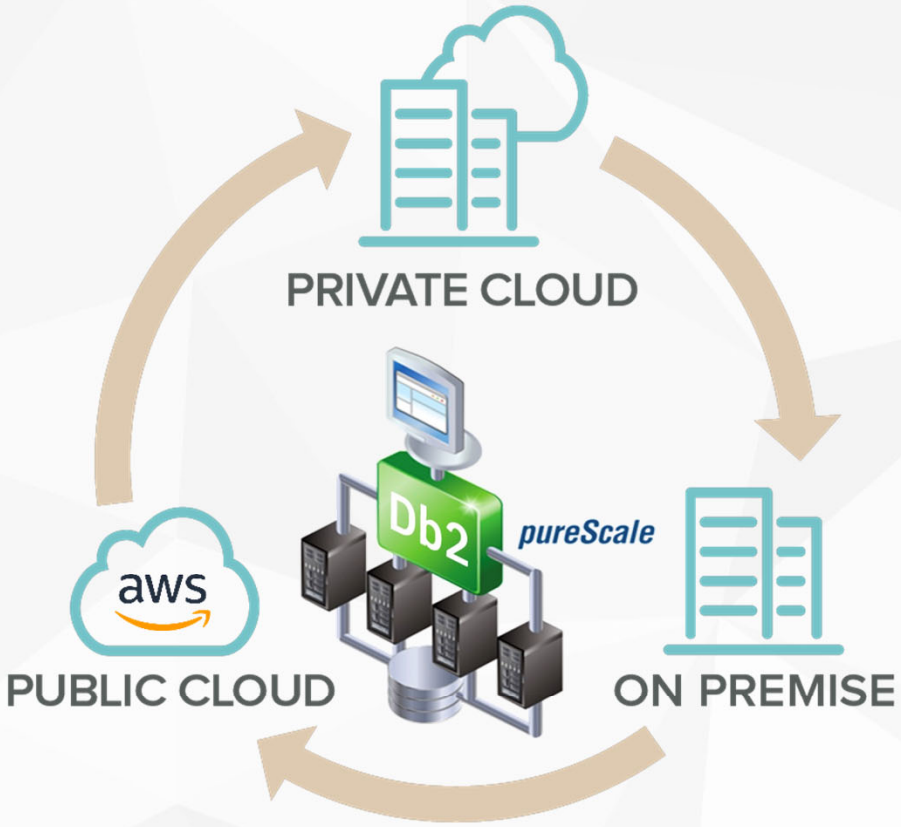


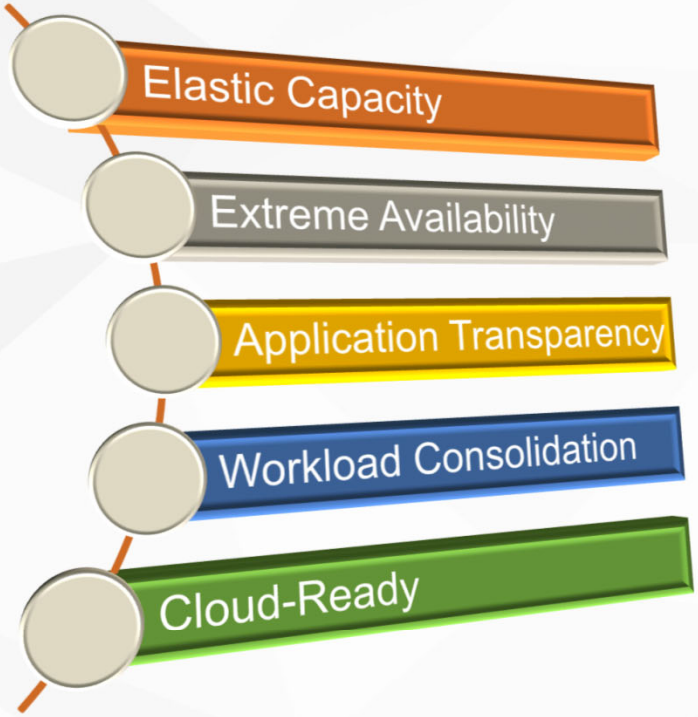
- Active / active (fully coherent)
- Hot / hot (**online** failover)
- Synchronous
- Complete DB replication
- Continuous testing of DR site
- Distance limitations (actually network latency)
- Only available through lab services

Agenda

- Why “always on”
- What are the Db2 options?
  - PureScale
- Combining technologies for the best resiliency
- The Journey to Always On

Db2 LUW's five 9s solution is *pureScale*





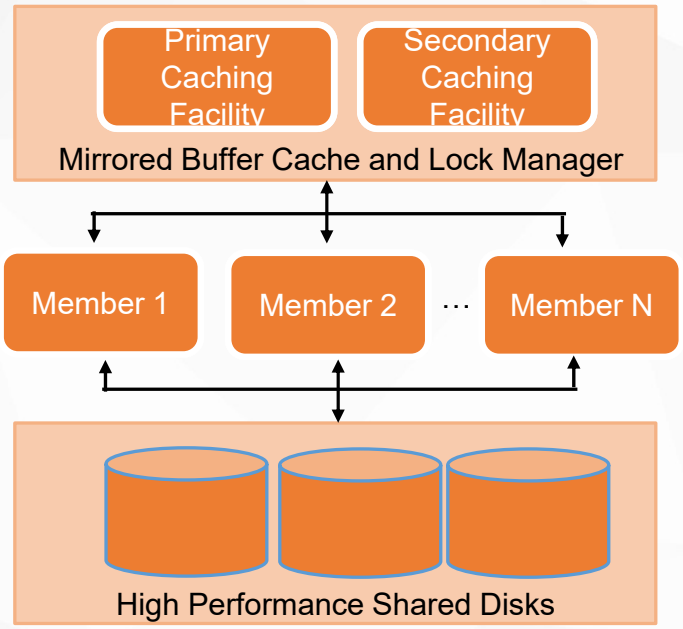
IBM DB2 PURESCALE PRIMER

PDUG

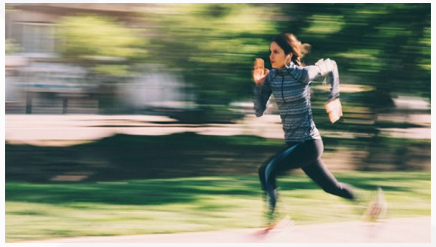
» Db2 pureScale is an active-active database cluster solution for *transactional workloads*. It reduces the risk and cost of business growth.

» **IBM® Db2® pureScale® leverages the parallel sysplex architecture and data sharing technology to provide mainframe-class continuous availability (5 9s)[1]**

1. **Always on - No unplanned downtime**
2. **Extreme scalability – Unlimited capacity**
3. **Application transparency – No application changes, no data redistribution, and no performance tuning required**



[1] **An IBM c**ase study on PUMA with Db2 pureScale



Business Challenge: Flexible scaling, microservices, and continuous availability of Db2 to interact with an increasingly digital world.

Solution: Add Db2 pureScale feature to support flexible scaling. **Develop new functions as microservices in Red Hat® OpenShift® containers for faster digital innovation.**

1

ALWAYS-ON, TRANSACTIONAL DATA

Maintain highly performant, transactional integrity at scale

Benefits

- Always-on: ensure continuity, security and performance to keep applications and daily operations running smoothly
- Cost-efficiency and manageability: enable investment in new digital services
- Better productivity and allocations: free staffing time for value-added activities
- Resiliency: reliably process rapidly changing, diverse and unpredictable workloads

Core Differentiator

- pureScale® leverages the IBM® Db2® parallel sysplex architecture, providing mainframe-class continuous availability (5 9s) that runs anywhere, whenever you need it.

- With AWS inherent redundancy in the network component, the pureScale cluster is even more resilient to certain types of failure compared to on-premises deployment, and with no additional cost to maintain the high-resilient infrastructure.

2
EXTREME SCALABILITY

Scale with near-linear efficiency and high predictability

Benefits

- Scale what you need: On AWS, several pre-configured cluster sizes are available to suit your desired throughput requirement
- Speed: deploying is completely automated and completed within minutes
- Simplicity: take advantage of the traditional cluster expansion method by simply adding new members to the instance

Core Differentiator

- With the pureScale feature on AWS, redeployment is as simple as a click of a button, creating a new, larger cluster within minutes.
- The simplicity and speed of this process makes this the **recommended path to deploy a new cluster with a larger size**, as opposed to manual reconfiguration on existing deployments

3
APPLICATION TRANSPARENCY

Application transparency allows a seamless transition from clusters in traditional data centers to AWS

Benefits

- Build more: With IBM® Db2® pureScale® there are no application changes to make, data to redistribute, or performance tuning to do.
- Reduce costs: Achieve optimal resource utilization at all times, which helps to keep application response times low, while reducing the risk and cost of application changes.

Core Differentiator

- IBM® Db2® pureScale® provides application transparency through automatic load balancing across all active members, keeping your applications running when you move to AWS.

Whether this is a new deployment on AWS, a move from on-premises pureScale to AWS, or a shift from on-premises or AWS single Db2 instance to pureScale on AWS, applications work just as they did before.

pureScale Heritage, Evolution, and Core Values

pureScale has

decades of technological maturity

- Db2 for LUW first introduced in 1993
- pureScale technology first introduced in 1990 on zOS (Parallel Sysplex)
- Sysplex ported to Db2 LUW as pureScale in 2009

rich enterprise functionality

- SQL support - leverage existing (or widely available) skills
- Fully consistent ACID – including multi-statement transactions
- Fine grained access control – restrict access by row/column

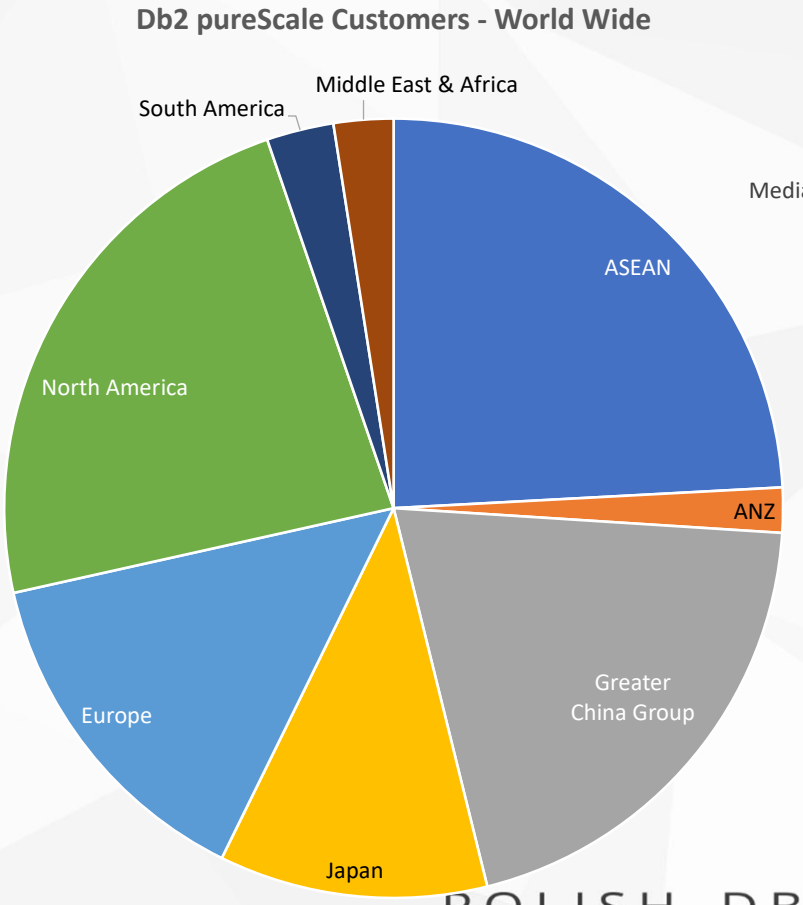
pureScale runs on

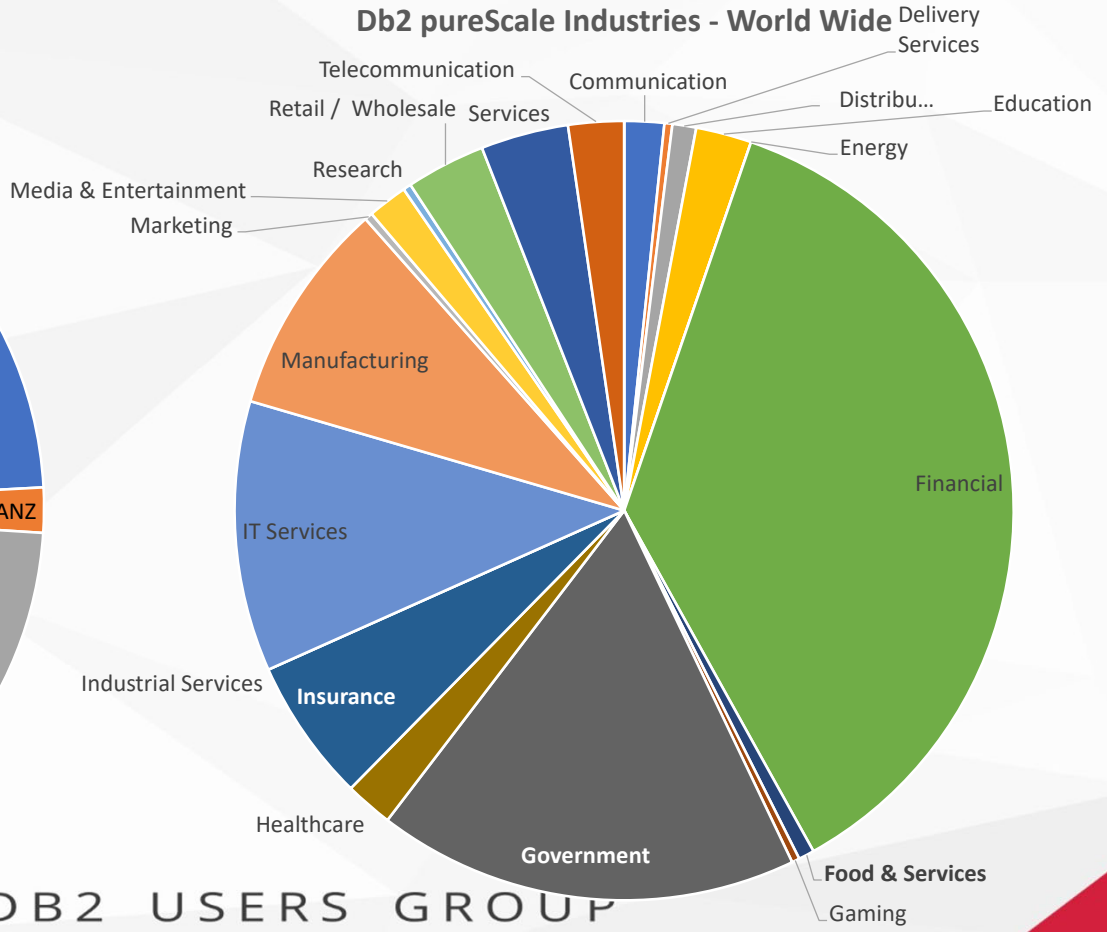
- 2009 • POWER AIX with Infiniband (IB)
- 2012 Jan • POWER AIX with RoCE
- 2012 Dec • x86 RHEL and SLES with IB and RoCE
- 2014 • AIX + Linux with TCP/IP support
- 2017 • POWER Linux
- 2021 • Linux on IBM Z14
- 2022 4Q • Linux on Z15
- 2022 June • Self-Managed service on AWS Marketplace
- 2023 2H • Linux on Z16
- 2025 2H • Linux on Z17 • z/OS HyperSocket support

pureScale is

battled tested in the world's most demanding environments …

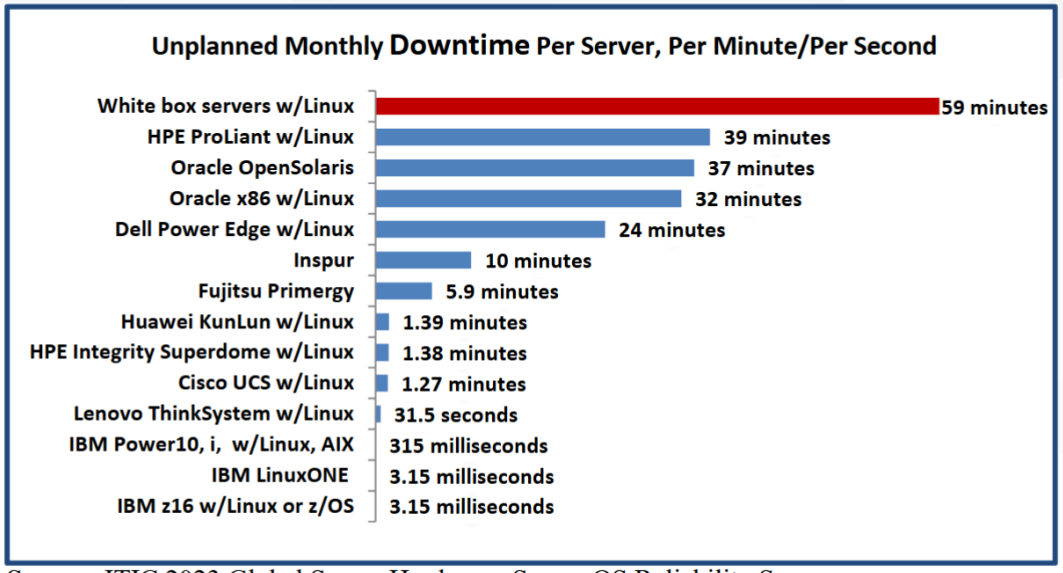
pureScale Successes at Every Geography and Every Industry



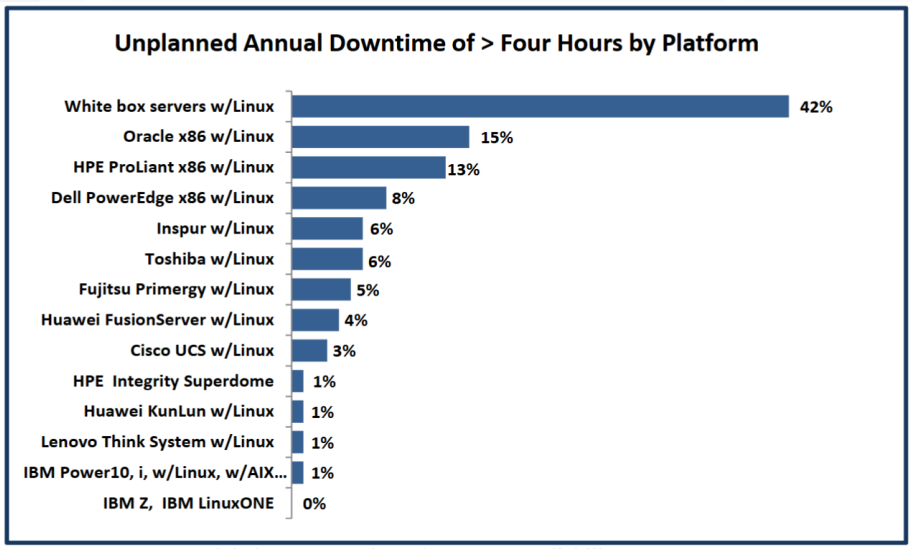


MORE UPTIME FOR APPS, LESS DOWNTIME FOR PEOPLE

- Unplanned outages bring our customers' applications down and generate a lot of distress among the IT staff, impacting productivity
- In addition to unplanned outages and security exposures, we must consider planned outages
- There are many more planned outages needed to replace x86 servers to apply patches and replace H/W components that are about to fail or failed already
- It means more risks, admin work and availability challenges to our clients
- **88% of Power10 corporate customers say their organizations recorded eight nines - 99.999999% - of server reliability**

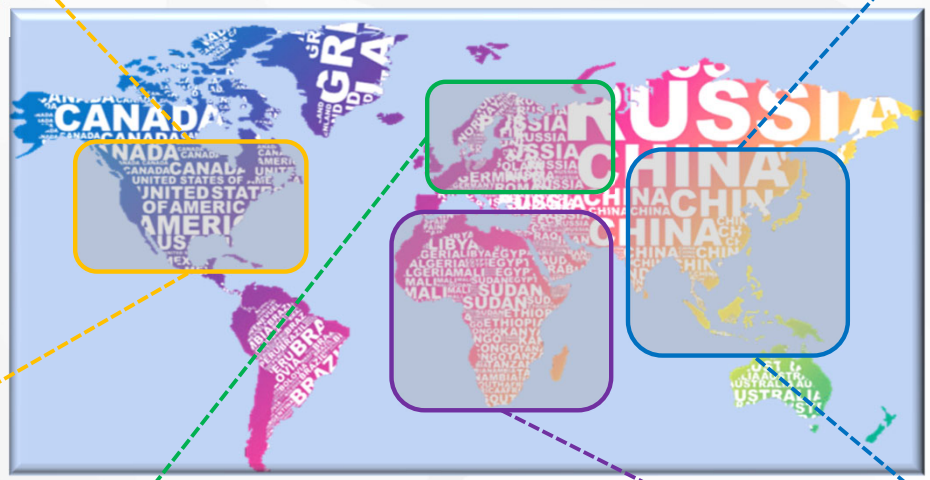


Source: ITIC 2023 Global Server Hardware, Server OS Reliability Survey



Source: ITIC 2023 Global Server Hardware/Server OS Reliability Survey

pureScale References across the world



THRIVENT FINANCIAL

allegiant

BNSF RAILWAY

BlueCross BlueShield

The Clearing House

PENSION FUND OF THE RUSSIAN FEDERATION

Handelsbanken

PUMA

Republic Of Turkey SOCIAL SECURITY INSTITUTION SGK SOSYAL GÜVENLİK KURUMU

PMU

Viasat

बैंक ऑफ महाराष्ट्र Bank of Maharashtra ONE FAMILY ONE BANK (A Government of India Undertaking)

Mobilink

RAKBANK

Central Board of Direct Taxes (CBDT)

LankaClear TRANSACT THE FUTURE

PDUG

新華社 XINHUA NEWS AGENCY

Ministry of Education Republic of China (Taiwan)

MUFG

臺灣土地銀行 LAND BANK OF TAIWAN

Suning 苏宁

中国民生银行 CHINA MINSHENG BANK

住信SBIネット銀行

交通銀行 BANK OF COMMUNICATIONS

SBI Sumishin Net Bank

中信銀行 CHINA CITIC BANK

SAMSUNG

SAMSUNG C&T

NISSAY

국세청 National Tax Service

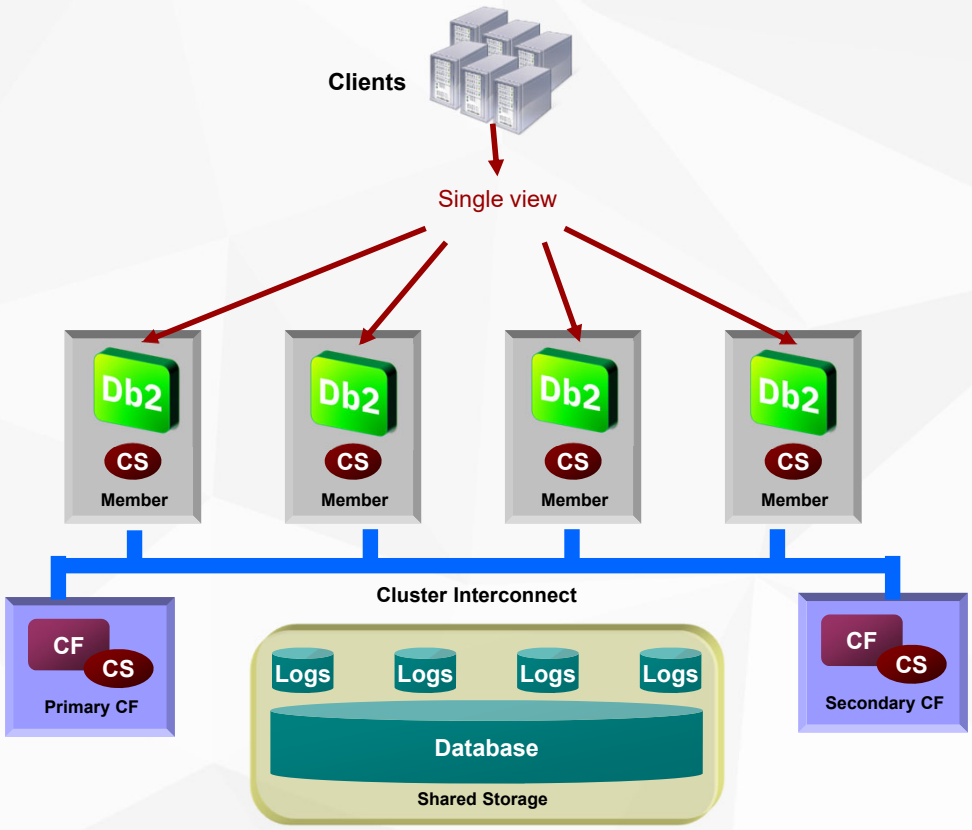
POLISH DE[illegible]UP

Going Forward: Continuous Investments & Commitment to all our customers

All form factors: On-premise and Cloud

All architectures: POWER, Intel, Z

FEATURES OVERVIEW

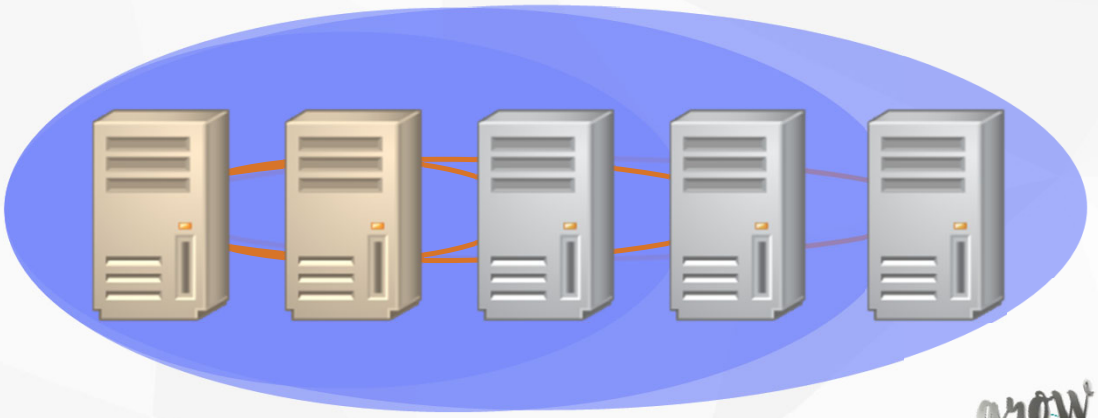


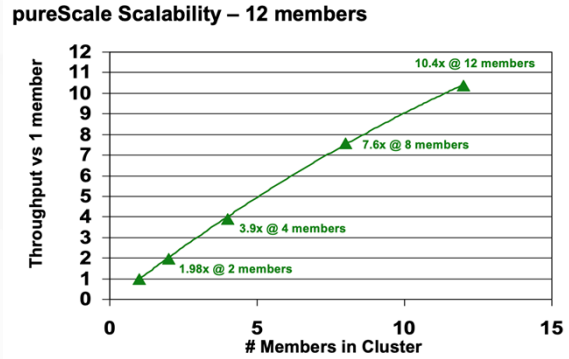
- Automatic Workload Balancing
- Cluster of Db2 nodes running on bare-metals or VMs (POWER, x86, Z, AWS EC2)
- CF — Caching Facility (CF) - Centralized global lock and page cache management
- CS — Cluster Services - Integrated failure detection and recovery automation
- High speed interconnect (RDMA & TCP/IP)
- Shared data

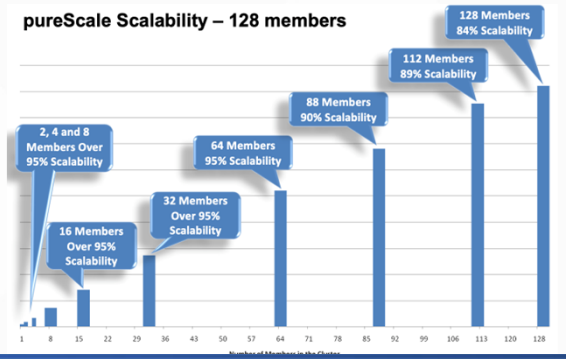
EXTREME CAPACITY AND APPLICATION TRANSPARENCY

PDUG

» **Db2 pureScale is designed to grow to whatever capacity your business requires**







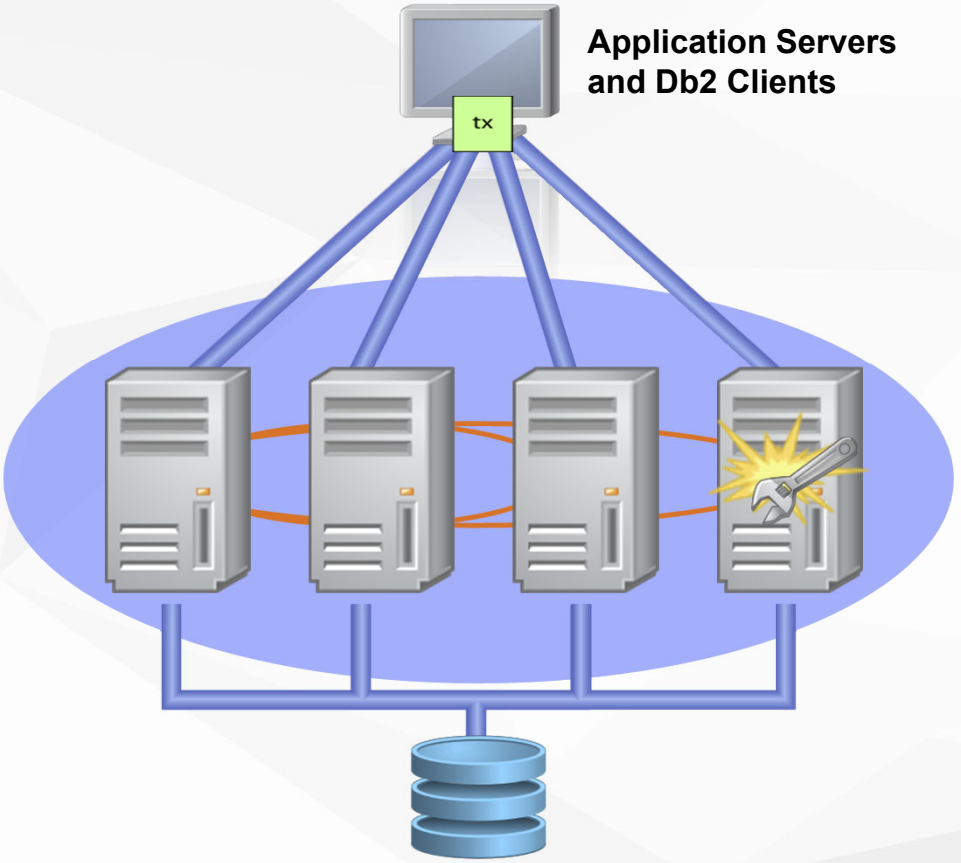
» **Take advantage of extra capacity instantly**

- » No need to modify your application code
- » No need to tune your database infrastructure

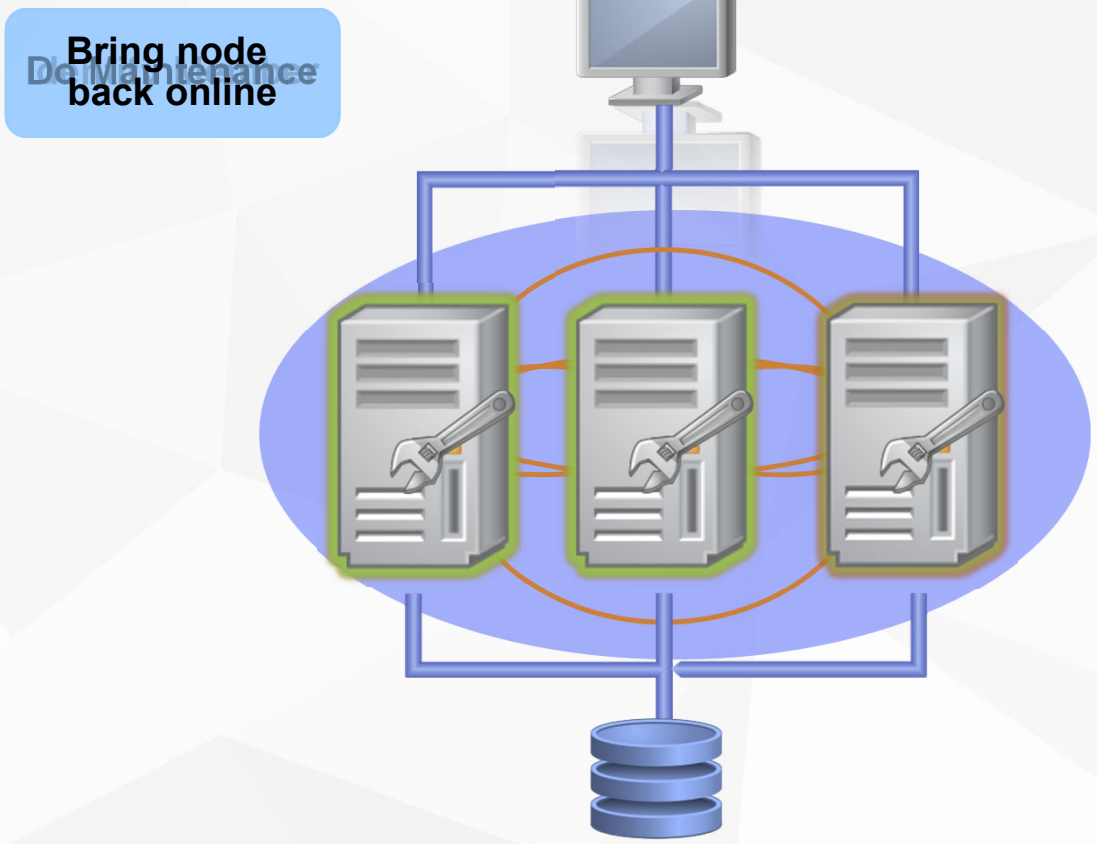
RECOVER INSTANTANEOUSLY FROM NODE FAILURE (*UNPLANNED* OUTAGES)

PDUG

- **Protect from infrastructure related outages**
  - Redistribute workload to surviving nodes immediately
  - Completely redundant architecture
  - Recover in-flight transactions on failing node in seconds (including detection of the problem)



MINIMIZE THE IMPACT OF *PLANNED* OUTAGES



- **Keep your cluster up during various maintenance operations:**
  - Db2 mod/fix pack maintenance – "Rolling Update"
  - OS fixes
  - Hardware updates
  - Administration

- **Supports multiple concurrent hosts maintenance** *(as long as quorum is maintained)*
- **Smart and Fast fix pack or mod pack update**
  - Single command on any host to update a cluster of ANY size (maximum concurrency guaranteed!)

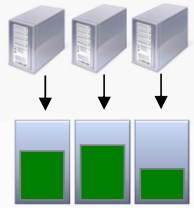
PURESCALE CLIENTS

PDUG

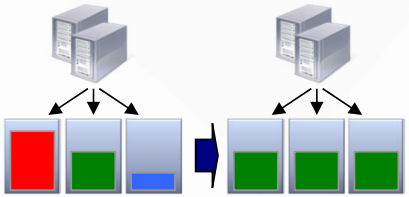
- Db2 front end for customers to submit work (queries, updates, etc.) to server via command line processor (CLP), Java/C apps, etc.

» **Q: Which member(s) will my application run on ?**

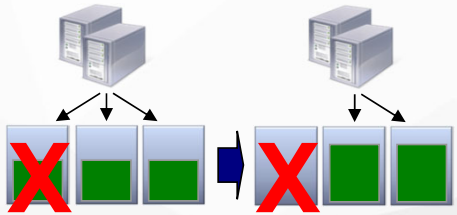
» **Client Affinity**
  - » Direct different groups of clients or workloads to specific members
  - » Consolidate separate workloads/applications on same database infrastructure
  - » Define list of members for failover purposes



» **Workload Balancing (WLB)**
  - » Application requests are balanced across all members or subsets of members
  - » Takes server load of members into consideration

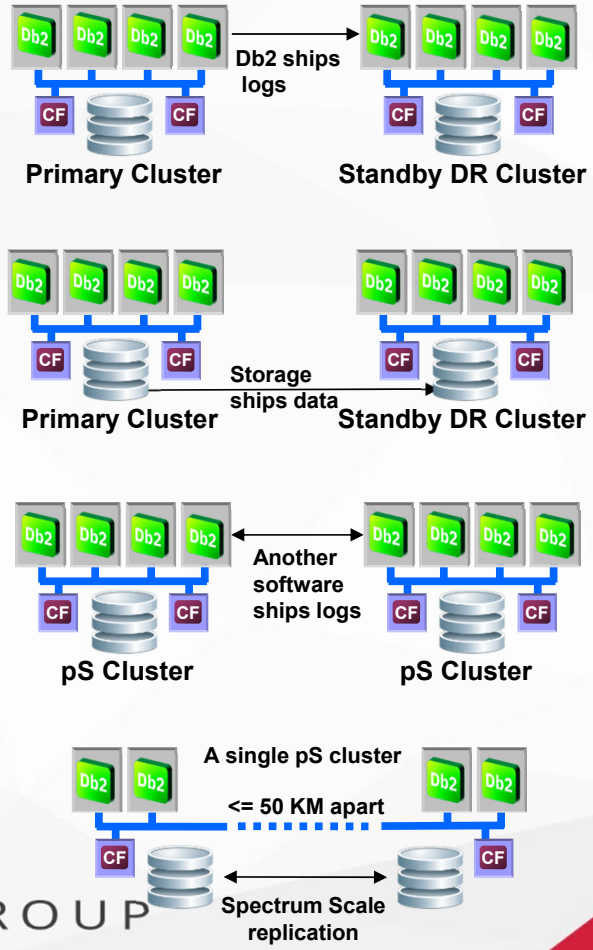


» **Automatic Client Reroute (ACR)**
  - » Client automatically connected to healthy member in case of member failure
  - » May be seamless in that no error messages returned to client
  - » Application may have to re-execute the transaction



DISASTER RECOVERY (DR) OPTIONS

- 1. HADR
  - Db2 is responsible to ship data across
  - Replication between a database in a primary pS cluster and a standby pS cluster with a matching member topology
  - Active/passive DR
- 2. Storage Replication
  - The storage layer is responsible to ship data across
  - Synchronous or asynchronous disk-based replication between a database in a primary pS cluster and a standby pS cluster with a matching member topology
  - Active/passive DR
- 3. Q Replication / InfoSphere Change Data Capture (CDC)
  - Another software is responsible to ship data across
  - Logical replication between two databases
  - Can be active/active DR
- 4. Geographically Dispersed pureScale Cluster (GDPC)
  - Data is replicated on each I/O synchronously
  - Single pS cluster "stretched" over two sites with half of members/CFs at each site
  - Active/active DR



DB2 12 SUPPORTED OS AND PLATFORMS

Distribution	Version	Db2 Server	Db2 Client	HA	pureScale
RHEL	9.4	x86-64 System z Power 9, 10 LE	x86-64, x86-32 System z (64 bit) Power 9, 10 LE (64 bit)	x86-64 System z Power 9, 10 LE (Pacemaker)	x86-64 System z Power 9, 10 LE (Pacemaker)
SUSE	15 SP6	x86-64 System z Power 9, 10 LE	x86-64, x86-32 System z Power 9, 10 LE	x86-64 System z Power 9, 10 LE (Pacemaker)	x86-64 (Pacemaker)
Ubuntu	22.04 LTS	x86-64 System z Power 9, 10 LE	x86-64, x86-32 System z Power 9, 10 LE	N/A	N/A
AIX	7.3 TI2	x86-64 System z Power 9, 10 LE	64bit Power 9, 10 LE	64bit Power 9, 10 LE (TSA)	64bit Power 9, 10 LE (TSA)
Windows Desktop	11	x86-64	x86-64, x86-32	N/A	N/A
Windows Server	2022	x86-64	x86-64, x86-32	N/A	N/A
Mac	Sonoma M1+	N/A	64Bit DSDRIVER only	N/A	N/A
UBI OpenShift	9	x86-64 System z Power LE	N/A	x86-64 System z Power LE	N/A

Db2 **12.1**

PDUG

	Db2 **Community**	Db2 **Starter**	Db2 **Standard**	Db2 **Advanced**	Db2 **Warehouse**
Pricing	Free	$$$$$	$$$$$$	$$$$$	$$$$$$
Use case	Free entry point to help kick start new apps and services	Core capabilities for new apps and services	Production-ready, business-critical applications	Fully-featured for mission-critical workloads	Production-ready, analytical workloads
License metrics	– (Usage restricted to non-prod)	VPC: Prod and Non-prod	• VPC: Prod and Non-prod • AU: Non-prod only	• VPC: Prod and Non-prod • AU: Non-prod only • AU (SAP only): Prod and Non-prod	VPC: Prod and Non-prod
License types	Community	Perpetual, Monthly, and Subscription	Perpetual, Monthly, and Subscription	Perpetual, Monthly, and Subscription	Perpetual, Monthly, and Subscription
Resource limits	4 CPU cores 8 GB Memory	4 CPU cores 16 GB Memory	16 CPU cores 128 GB Memory	Unlimited	Unlimited
Deployment	Single-node	Single-node	Single or multi-node for HA/DR	Single or multi-node	Multi-node
Workload type	Transactional	Transactional	Transactional & Analytics	Transactional & Analytics	Analytics
Support	Community Support	Enterprise Support	Enterprise Support	Enterprise Support	Enterprise Support
Db2 Native Encryption & Audit	◉	–	◉	◉	◉
Db2 HADR (High Availability & Disaster Recovery)	◉	–	◉	◉	–
Db2 Federation	◉	–	◉	◉	◉
Db2 Backup & Restore (Point-in-Time Recovery)	◉	–	◉	◉	◉
Db2 pureScale	◉	–	–	◉	–
Db2 Database Partitioning Feature (DPF) for Data Warehousing	◉	–	–	◉	◉

POLISH DB2 USERS GROUP

DB2 12 PURESCALE UPDATES

PDUG

- Replacement of TSA with Pacemaker technology for cluster management
- Recovery through drop member operations
  - » Prereq to support online drop member in a future release
- Improved database activation resiliency
- Improved handling of a database failure that does not result in node failure
  - » i.e., Force Database Shutdown scenarios – in these failures, the clients will be failed over to alternate nodes until the database is activated, after which the node will become available again to handle workloads.
- Db2 pureScale HADR support for enterprise-grade end-to-end SSL encryption
- Verbs support for RDMA Everywhere in pureScale (SLES on Intel & Z, and AIX)
  - » RDMA operations work directly on a remote node's memory, bypassing its CPU, and include RDMA reads, writes, and atomic operations

DB2 12 PURESCALE AWS UPDATES

PDUG

- Self-Managed pS non-RDMA on AWS with Pacemaker
- Self-Managed pS EFA (Elastic Fabric Adapter) on AWS with Pacemaker*
  - » Performance comparable to RDMA - suitable for heavy duty OLTP workload
  - » Fast I/O fencing on AWS (NVMe reservation)

* Technical Preview

DB2 12.1.1 PURESCALE DELIVERABLES

PDUG

- Online Index Reorg
- Self-Managed pS EFA on AWS with Pacemaker
- Self-Managed pS on Azure with Pacemaker
- More reorg and AMT enhancements
- Fully scalable online (out/in)

Any statements regarding IBM's future direction, intent or product plans are subject to change or withdrawal without notice.

DB2 12.1.1+ PURESCALE ROADMAP

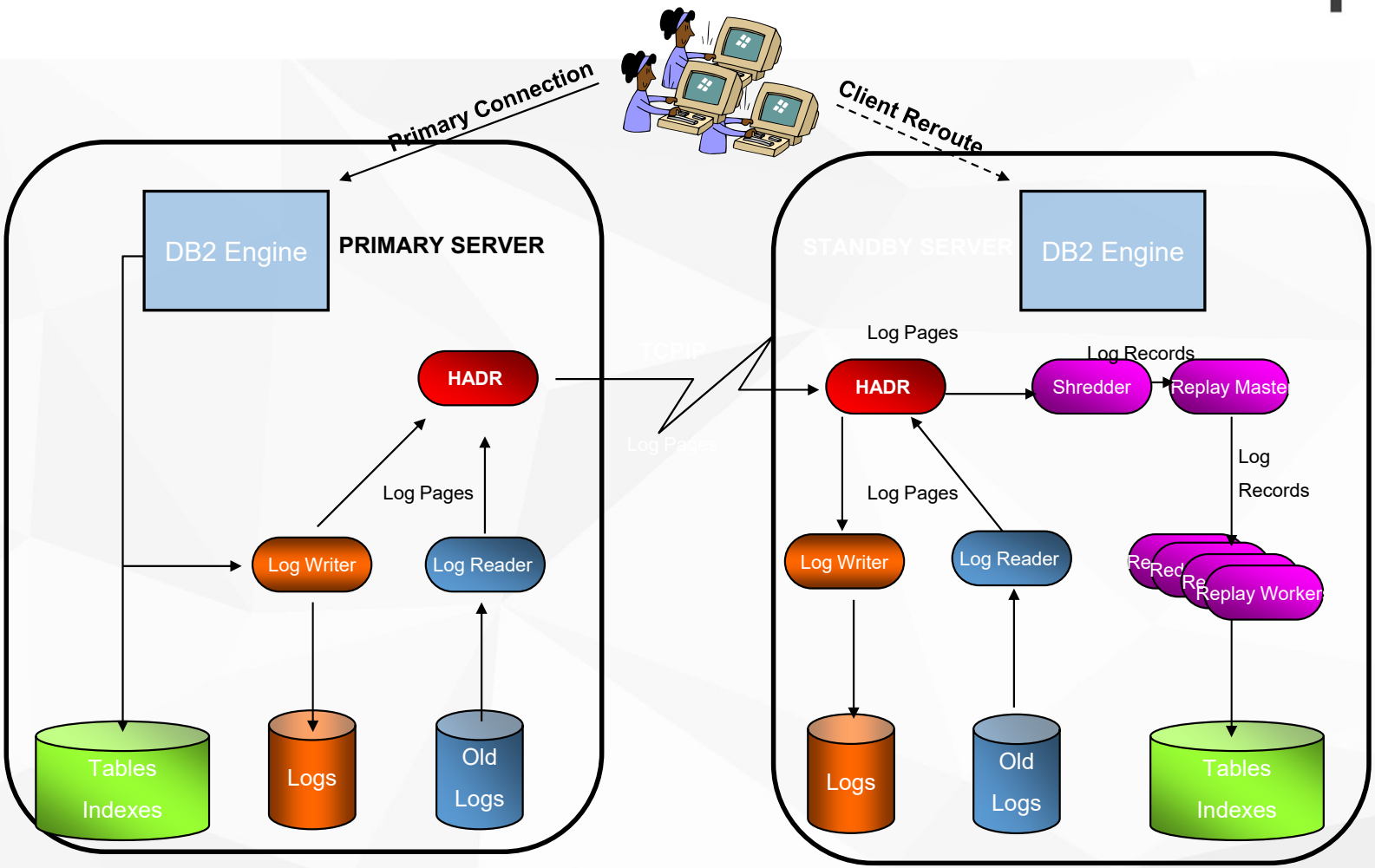
PDUG

- Multi-cloud: Cross-cloud HADR
- Read-on-standby for HADR
- Mixed topologies in HADR configurations
- Online version upgrade to next major release

Agenda

- Why “always on”
- What are the Db2 options?
  - HADR
- Combining technologies for the best resiliency
- The Journey to Always On

HADR IMPLEMENTATION



PDUG

Logged operations are replicated

- Example: DDL, DML, table space and buffer pool creation/deletion.

Not logged operations are not replicated.

- Example: database configuration parameter. not logged initially table, UDF libraries.

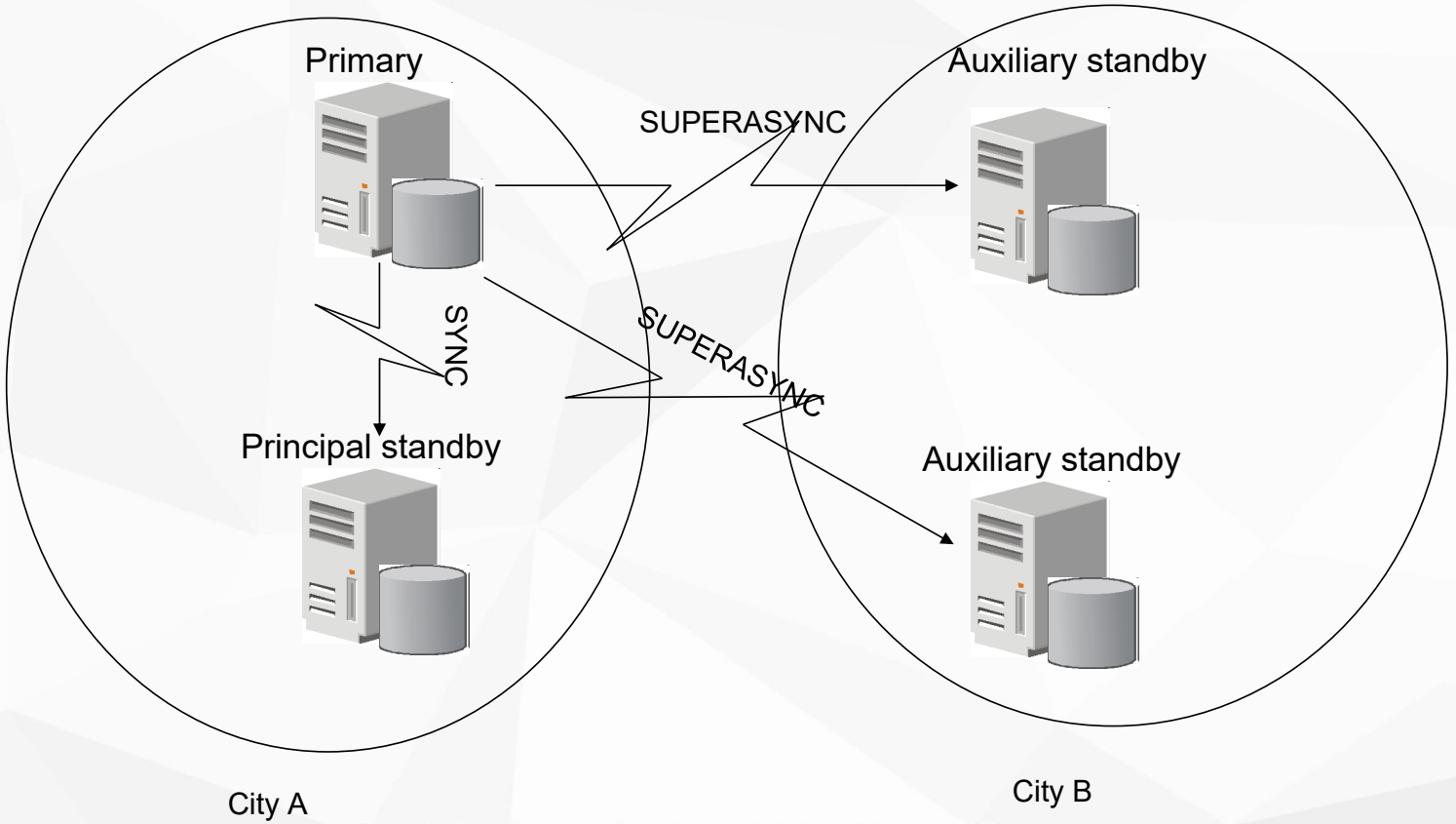
Index pages are not replicated unless LOGINDEXBUILD is enabled

- Ensure logsecond is maxed out as index rebuild is a single transaction

How do I prevent non-logged operations?

- Enable BLOCKNONLOGGED db cfg parameter

Typical HADR Multiple Standby Deployment



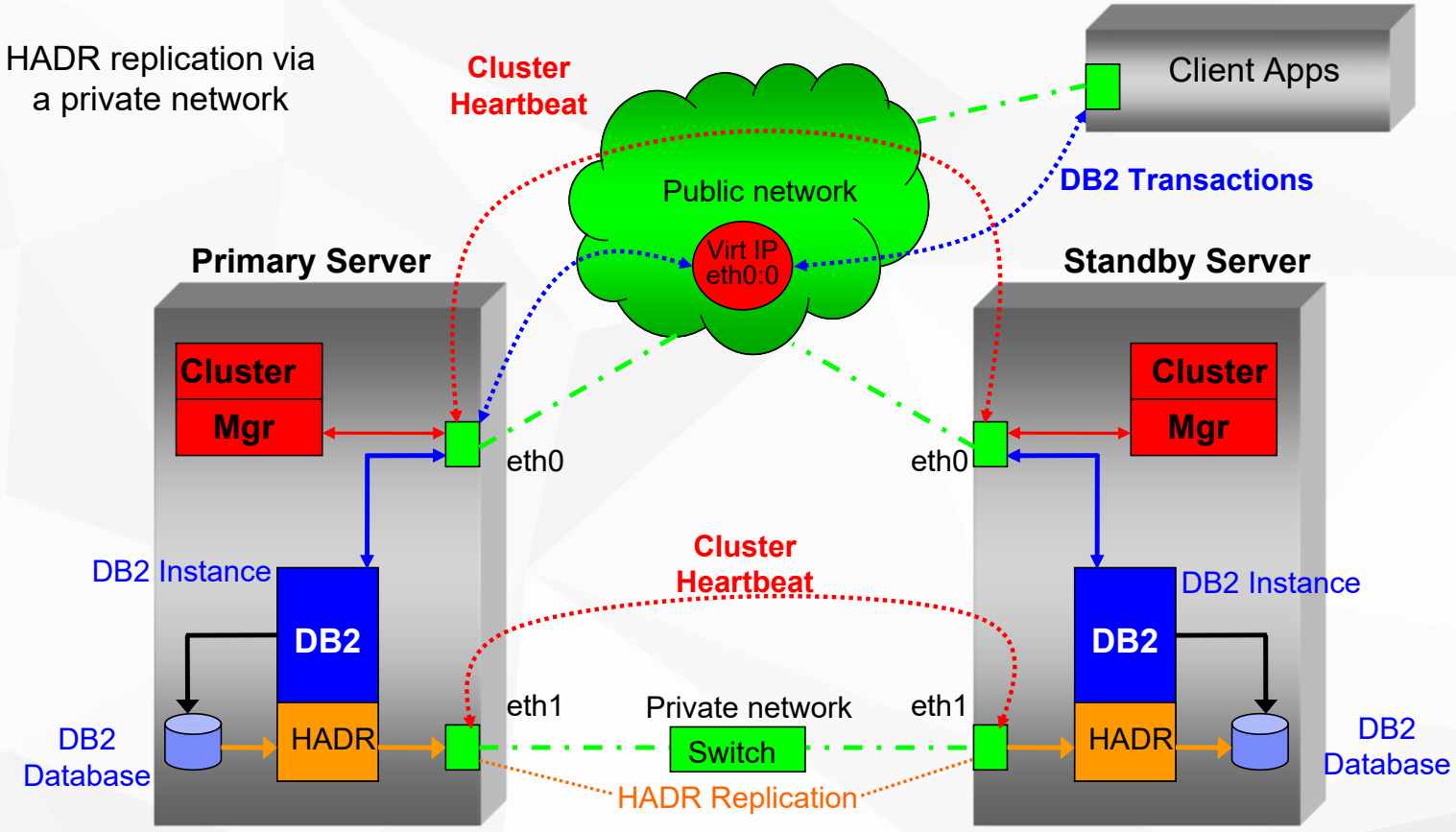
HADR AUTOMATION - MULTIPLE OPTIONS EXISTS – DEPENDING ON THE ENVIRONMENT

PDUG

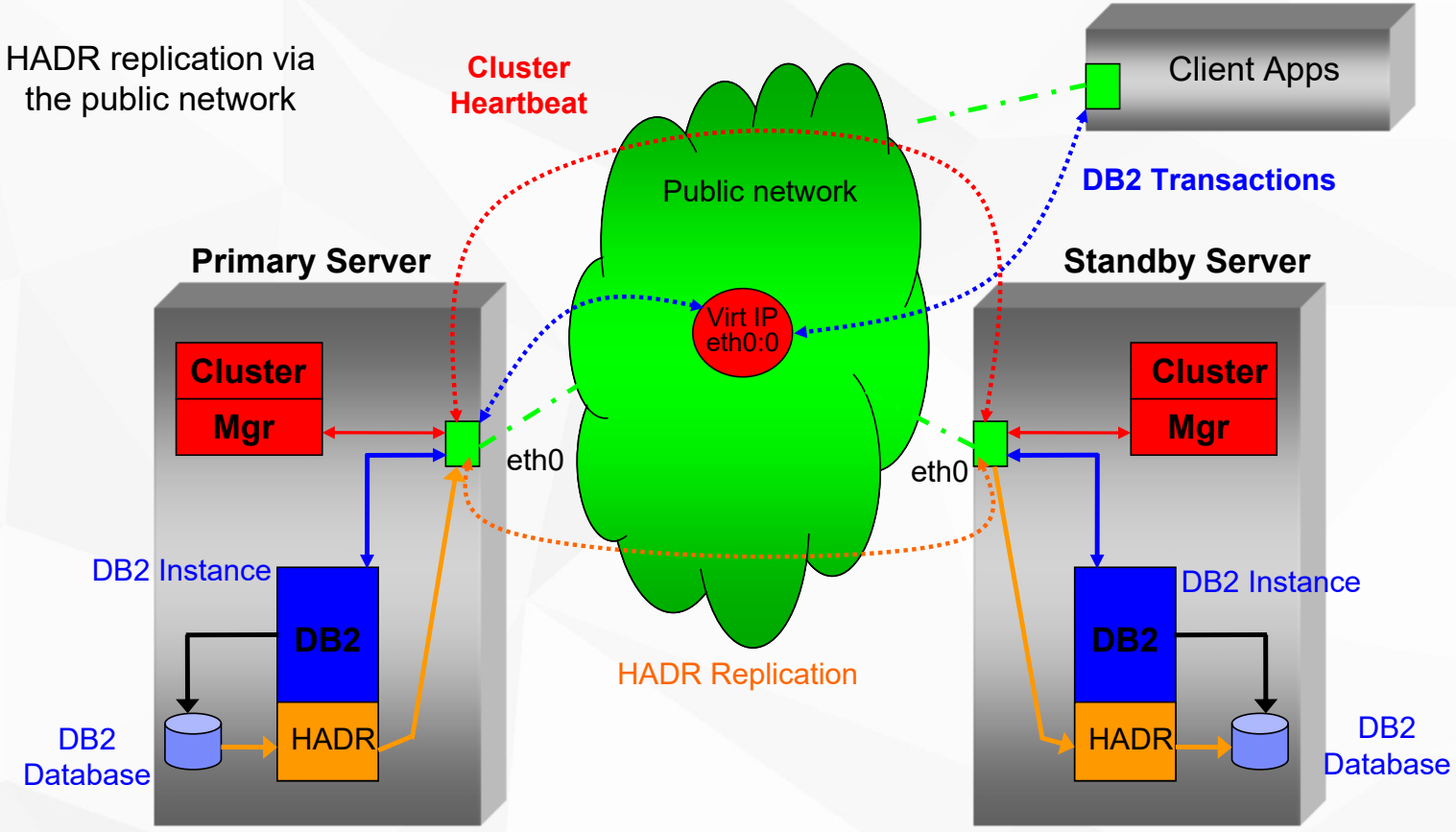
AIX	• TSA or other vendor supplied agents (e.g. VCS)
Linux (PPCLE, X86, z/Linux)	• Pacemaker, TSA or other vendor supplied agents (e.g. VCS)
Cloud Deployment	• Built into Db2U, utilizes etcd + Governor

Strategic direction is Pacemaker and eventually all deployments environments will support Pacemaker

Recommended DB2 HADR environment



Alternate example of a DB2 HADR environment



- **Follow instructions at**

https://www.ibm.com/docs/sl/db2/12.1?topic=feature-integrated-solution-using-pacemaker

- **Setting up two-node Db2 HADR Pacemaker cluster with fencing on AWS**

https://www.ibm.com/support/pages/setting-two-node-db2-hadr-pacemaker-cluster-fencing-aws

- **Setting up a Db2 HADR Pacemaker cluster with Overlay IP as Virtual IP on AWS**

https://www.ibm.com/support/pages/setting-db2-hadr-pacemaker-cluster-overlay-ip-virtual-ip-aws

Mandatory changes with Pacemaker

Quorum

- No disk tiebreaker and IP tiebreaker support.
- Split brain scenario needs to handled differently

Solution

Split brain prevention:

- Use Qdevice (with a 3rd arbitrator host)
- Same as HADR

Fencing Mechanism

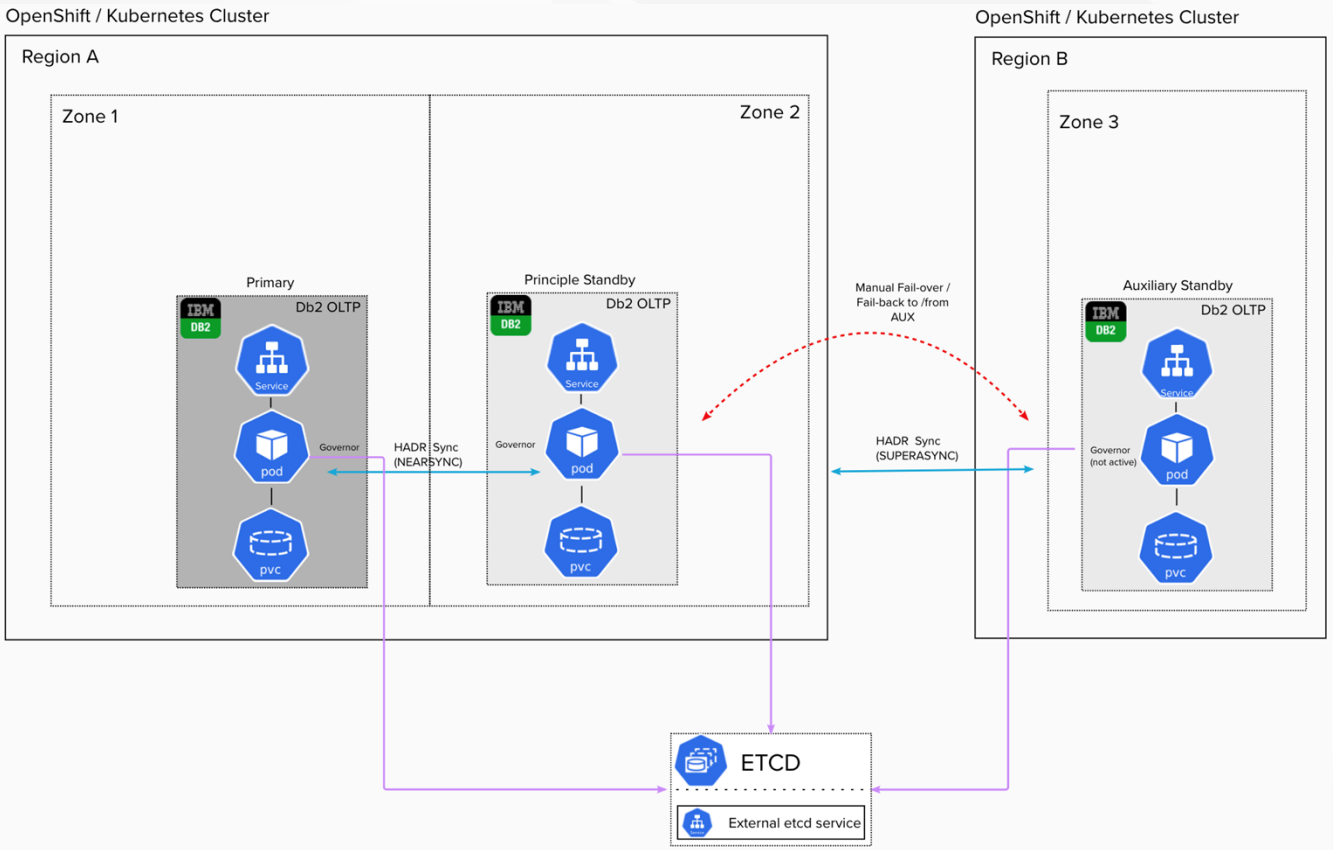
- No 1-1 mapping of RSCT's Critical Resource Protection feature.
- Node fencing (prevents data corruption) needs additional setup

Solution

Node fencing:

- Utilize Software Watchdog (default or separate install)
- Use in combination with Qdevice to trigger a reboot when a node eviction action is deemed necessary.

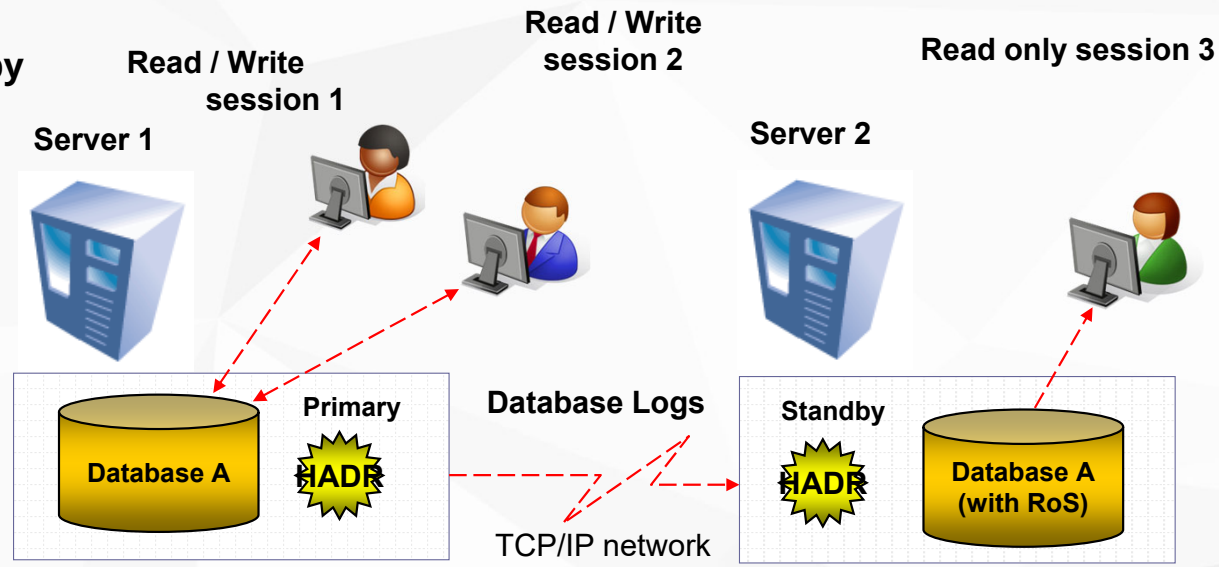
Db2U HADR
Topologies supported



- Intra/Inter clusters
  - 1 site – 1 standby
  - 2 sites – 1 standby
  - 2 sites – 2 standbys
    - Site 1: Primary / Principal standby
    - Site 2: Aux standby
- Semi-automated for > 1 site/cluster/namespace
  - Automated failover to Principal Standby
  - Manual fail-over/back to Auxiliary Standby

HADR READ ON STANDBY (ROS)

- Reads on Standby provides high availability, disaster recovery and allows reporting workloads.
- Improve resource utilization on your HA or DR hardware
- Offload reporting work from your primary, Increase capacity of HADR system
- Maximize Return on Investment and decrease Total Cost of Ownership
- **No CGTT or DGTT support**
- **Requires full license on standby**



Agenda

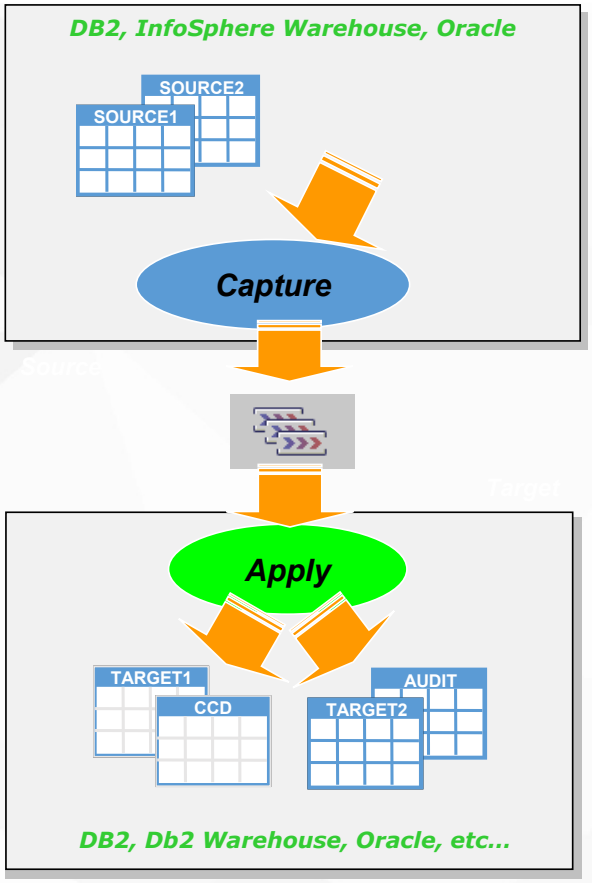
- Why “always on”
- What are the Db2 options?
  - Logical Replication
- Combining technologies for the best resiliency
- The Journey to Always On

LOGICAL REPLICATION - IIDR/IDR

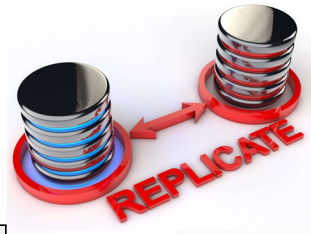
The IIDR product includes 3 technologies that can be used	• Q Repl – Best suited for DB2 to DB2 homogenous replication • CDC – Best suited for heterogenous replication • SQL Repl – Best suited for publish/subscription topologies
Either Q Repl and CDC are recommended	• SQL Repl will require an additional table (CD_TABLE) for every table that will be replicated
If your source DB makes use of Referential Integrity(RI) Constraints then Q Repl is the preferred technology to use	• CDC Best Practices state that the RI constraints on the target be dropped OR fast replay will be disabled • Q Repl allows for RI constraints to be on the target with fast replay enabled

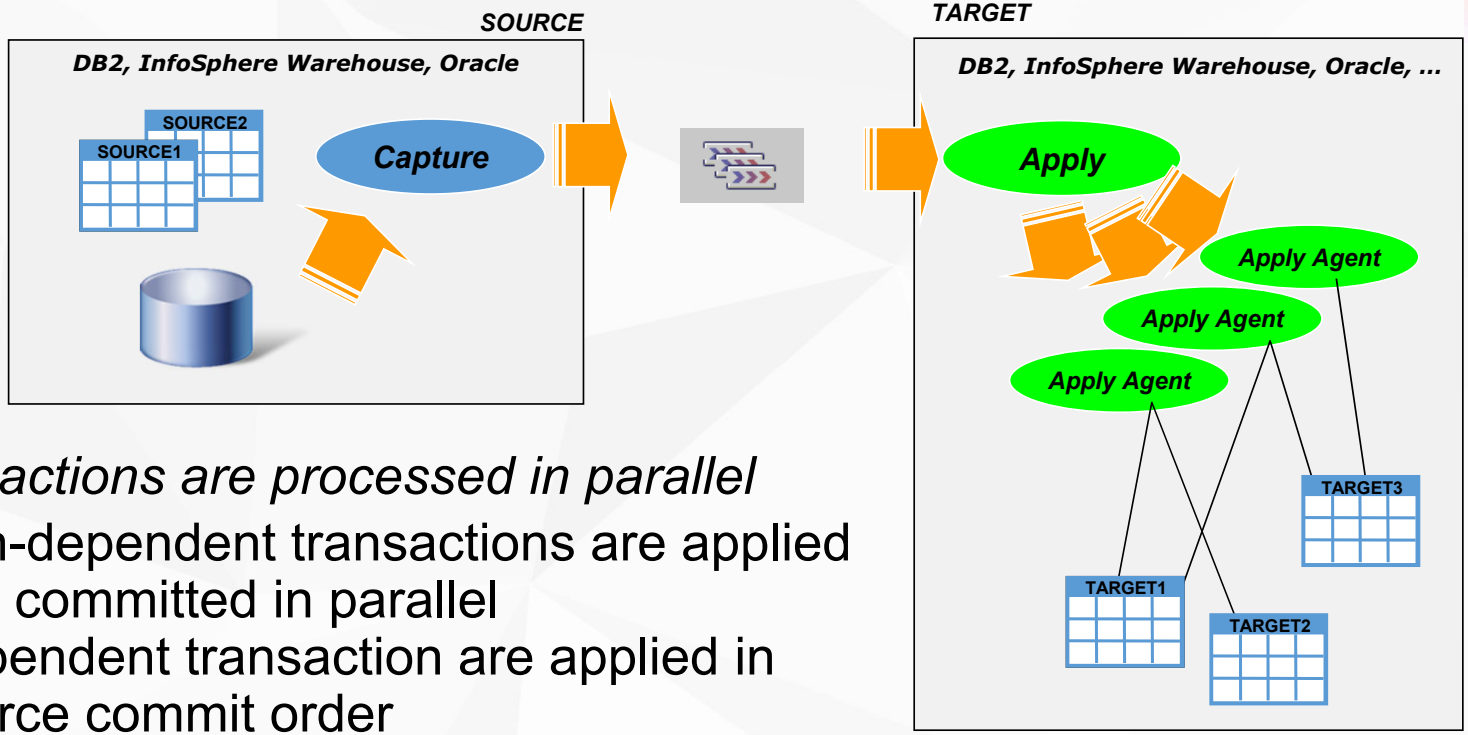
Logical Replication - Q Replication

- *Change data read from database log*
- *Changes shipped directly to target*
  - No staging
- *MQ provides transport persistence*
  - A key to recoverability
- *Data then applied to target tables*
  - Highly parallel … more on next slide
- *DBA friendly*
  - All metadata in tables
  - Statistics and messages in tables



BEST OF BREED PERFORMANCE – PARALLEL APPLY



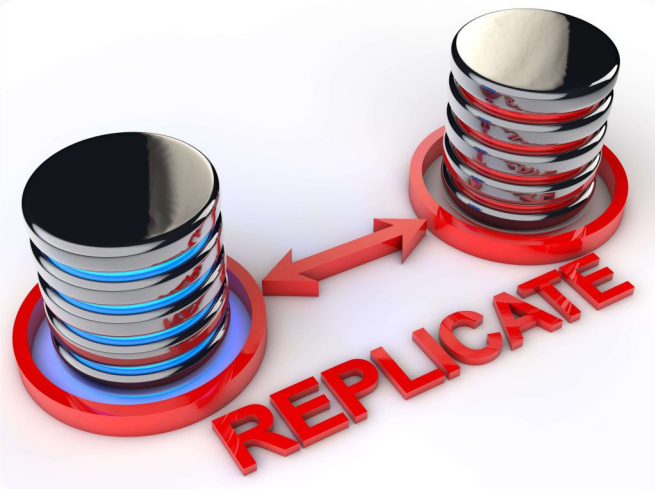


Transactions are processed in parallel

- Non-dependent transactions are applied and committed in parallel
- Dependent transaction are applied in source commit order

REPLICATION CHALLENGES

- Tables with no unique indexes
- Generated Always Columns
- Tables with Identity Columns
- Data Sequences
- Large Objects
- Referential Integrity
- Tables with triggers that update other tables
- DDL Replication



Agenda

- Why “always on”
- What are the Db2 options?
- Combining technologies for the best resiliency
- The Journey to Always On

TRADITIONAL THINKING

IBM Resiliency Services provides standardized High Availability and Disaster Recovery solutions today, where planned and unplanned downtime cause digital service disruption, impacting both financial obligations and brand.

- **High Availability**
  - Protects against infrastructure outages within a data center.
  - ***Does not protect against data center failures***
- **Disaster Recovery**
  - Protects against large-scale infrastructure and data center outages.
  - ***Requires some level of downtime*** *while systems are recovered* – dependent on the solution. Near-zero downtime solutions are most expensive.
  - *Data loss is dependent on solution.* Near-zero data loss solutions are most expensive.
  - ***Requires extensive program management, maintenance, annual testing****.*

Consider moving from Disaster Recovery To Disaster Avoidance

COMBINING HADR AND STORAGE REPLICATION

Provide continuous availability,

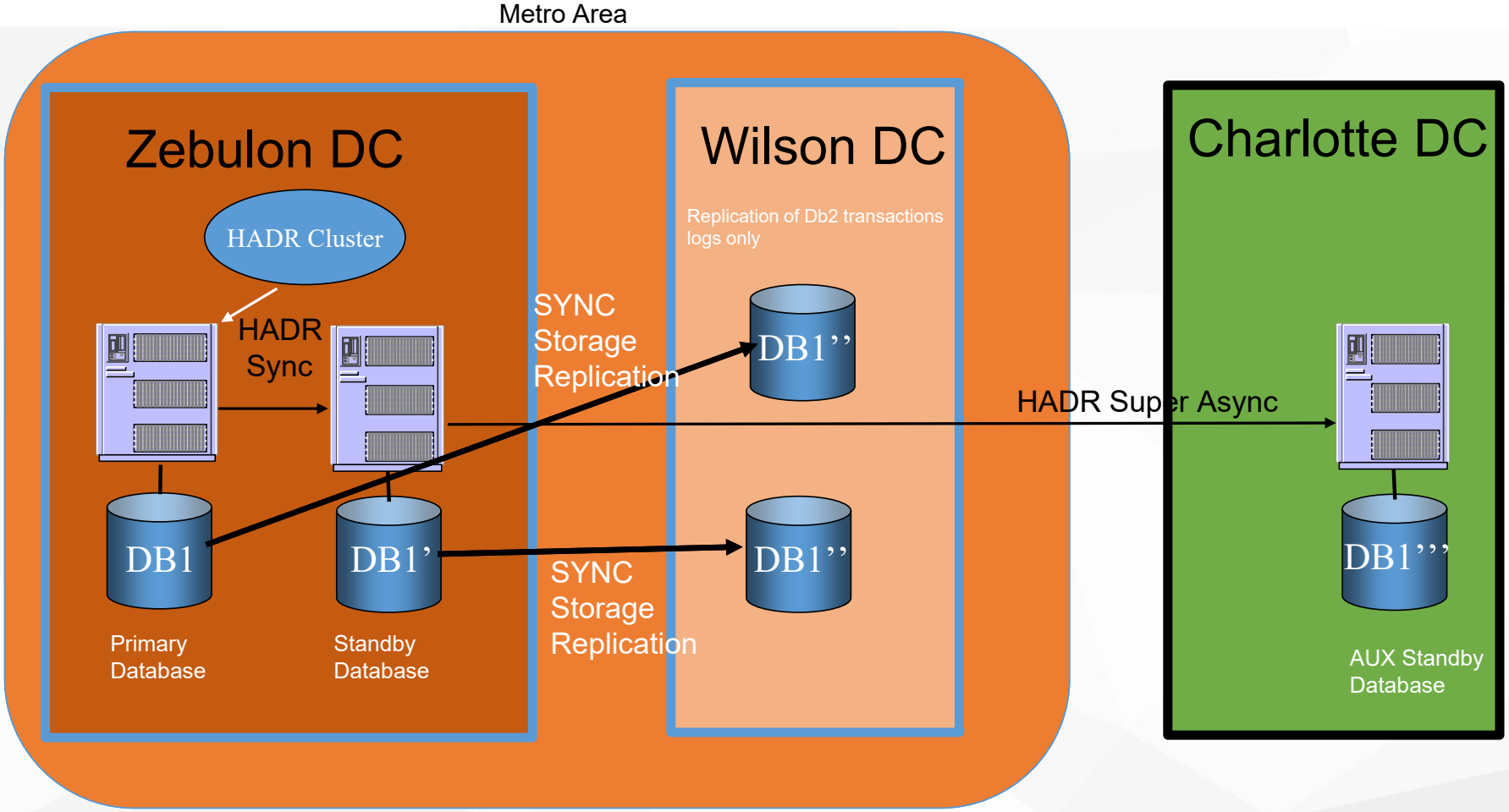
Protect from localized and site wide failures

Zero data loss in the case of a DR

RPO must be 0 whereas RTO can be in minutes

Ability to exploit a “bunker site” near the primary data center

RECOMMENDED CONFIGURATION



COMBINING PURESCALE AND Q REPL

Provide continuous availability,

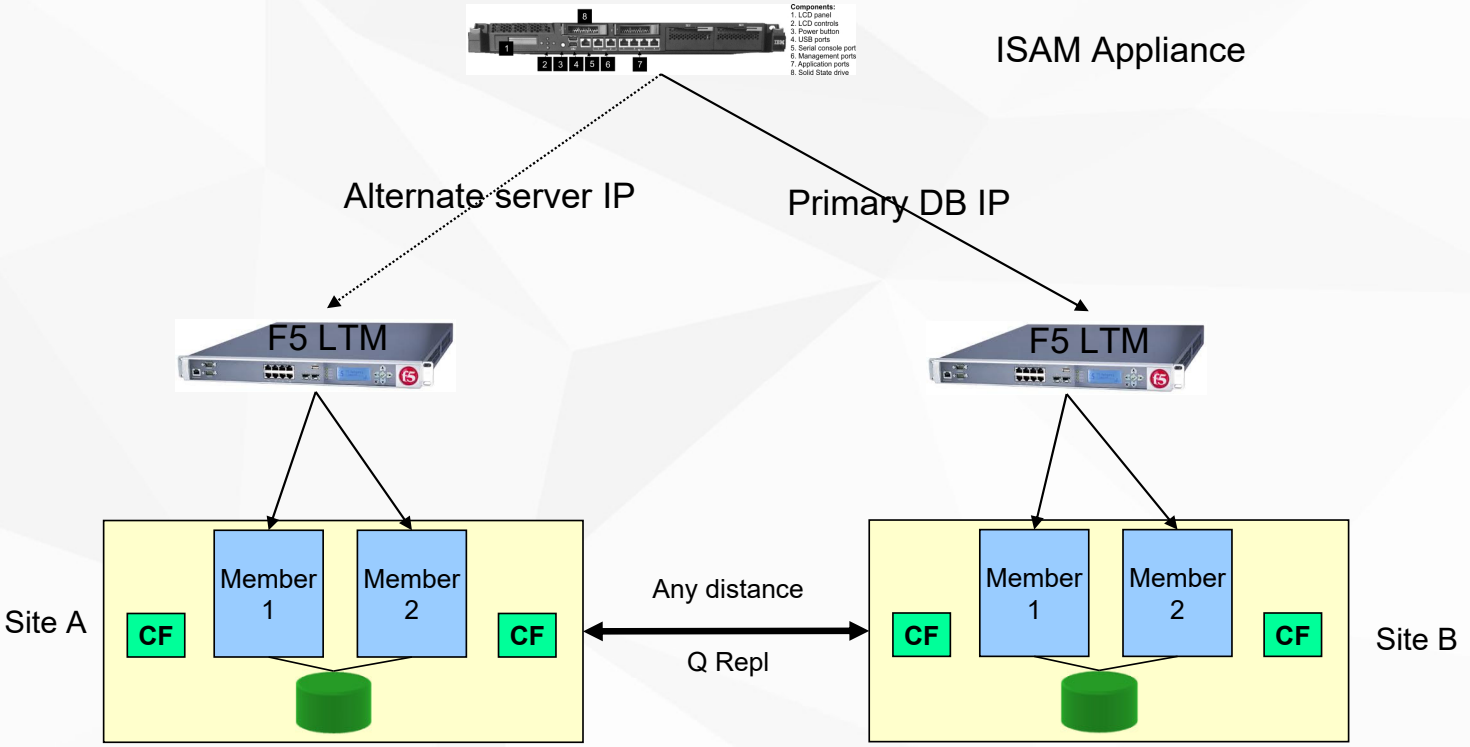
Protect from localized and site wide failures

Workload is read mostly

Very few writes

Application device was limited to have 2 IP addresses registered

ISAM CONFIGURATION



COMBINING PURESCALE AND Q REPL

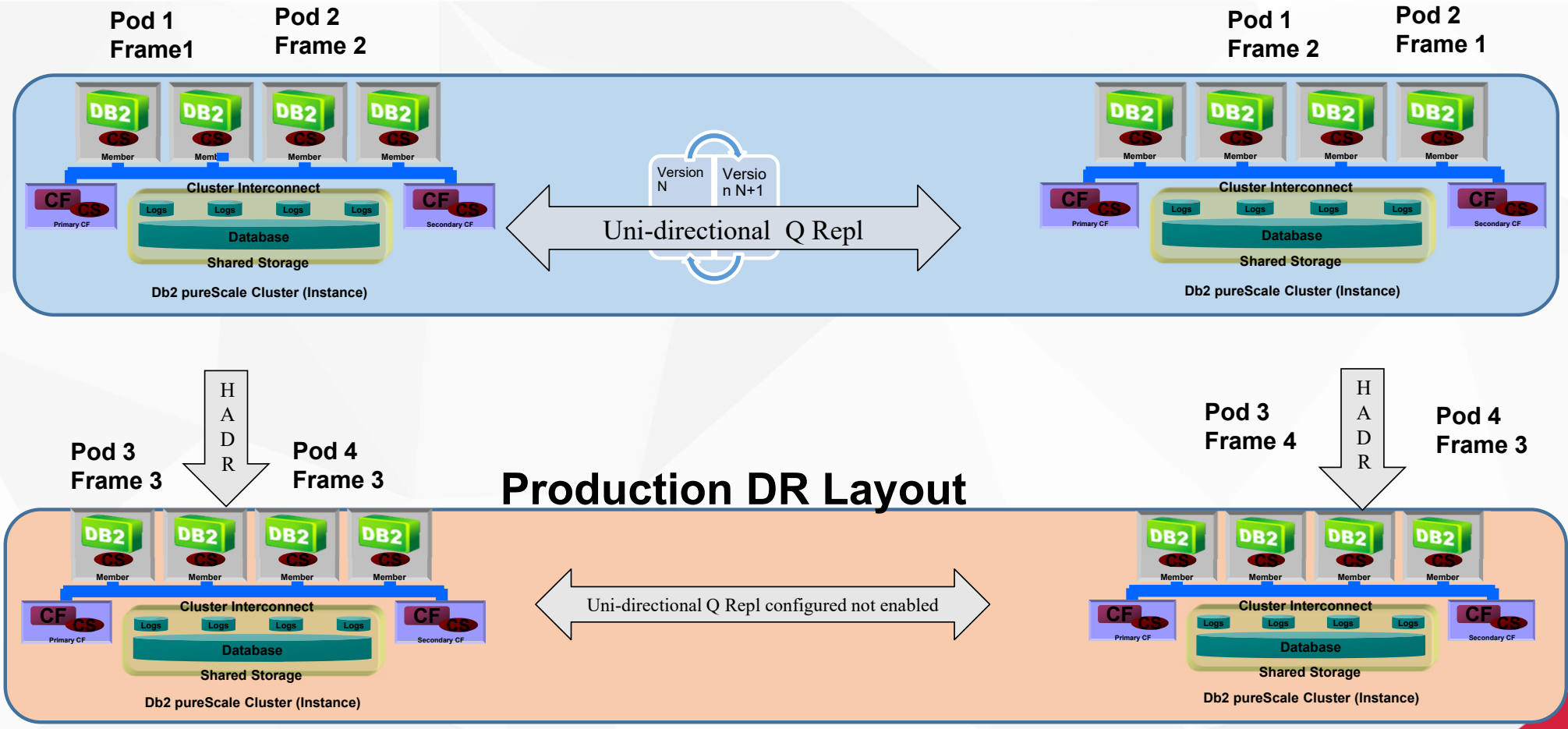
Provide continuous availability,

Protect from localized and site wide failures

Provide ability to roll out upgrades with zero downtime

Application changes including schema changes must be handled

PRODUCTION LAYOUT



COMBINING HADR AND Q REPL

Provide continuous availability,

Protect from localized and site wide failures

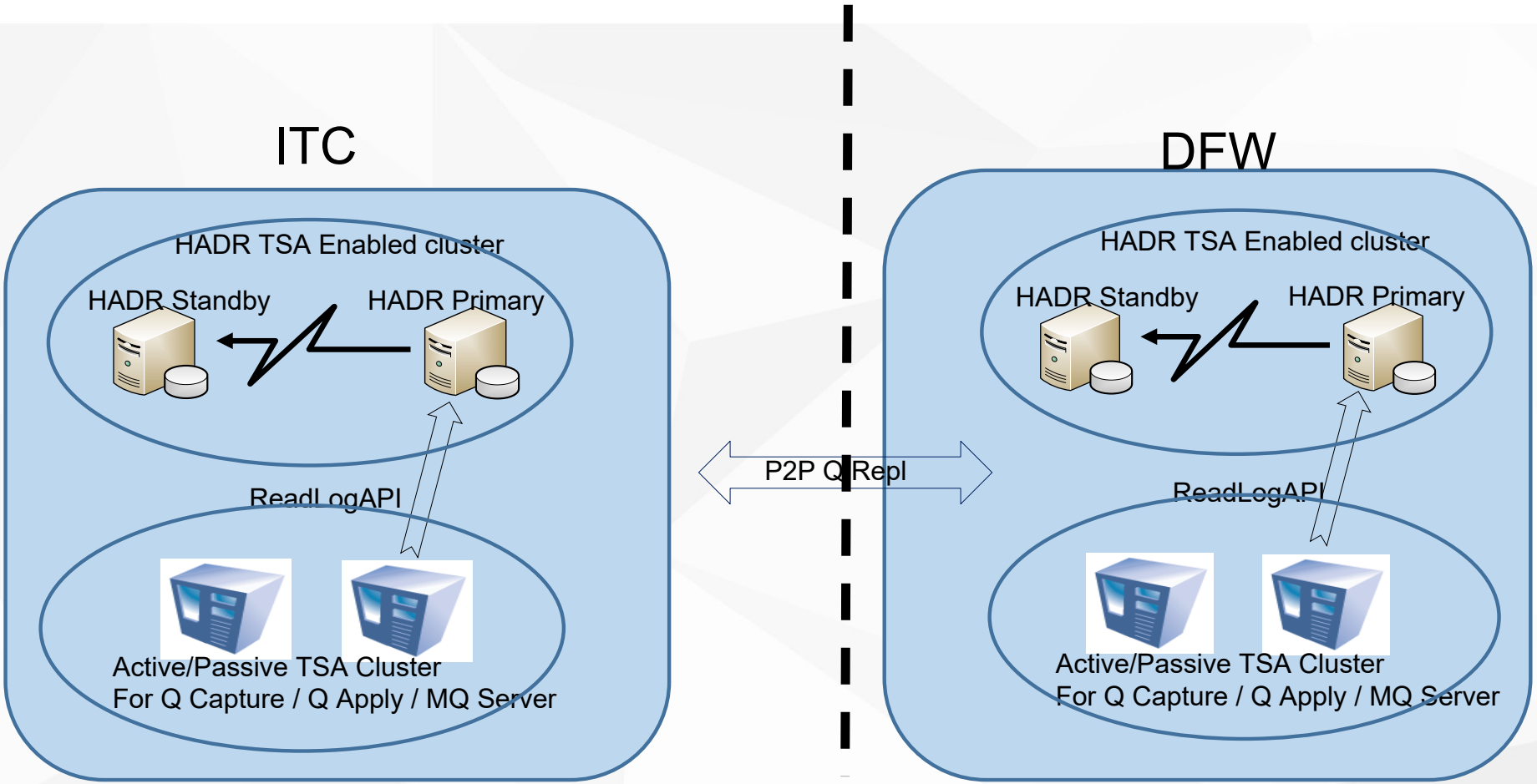
Provide ability to roll out upgrades with zero downtime

Application changes including schema changes must be handled

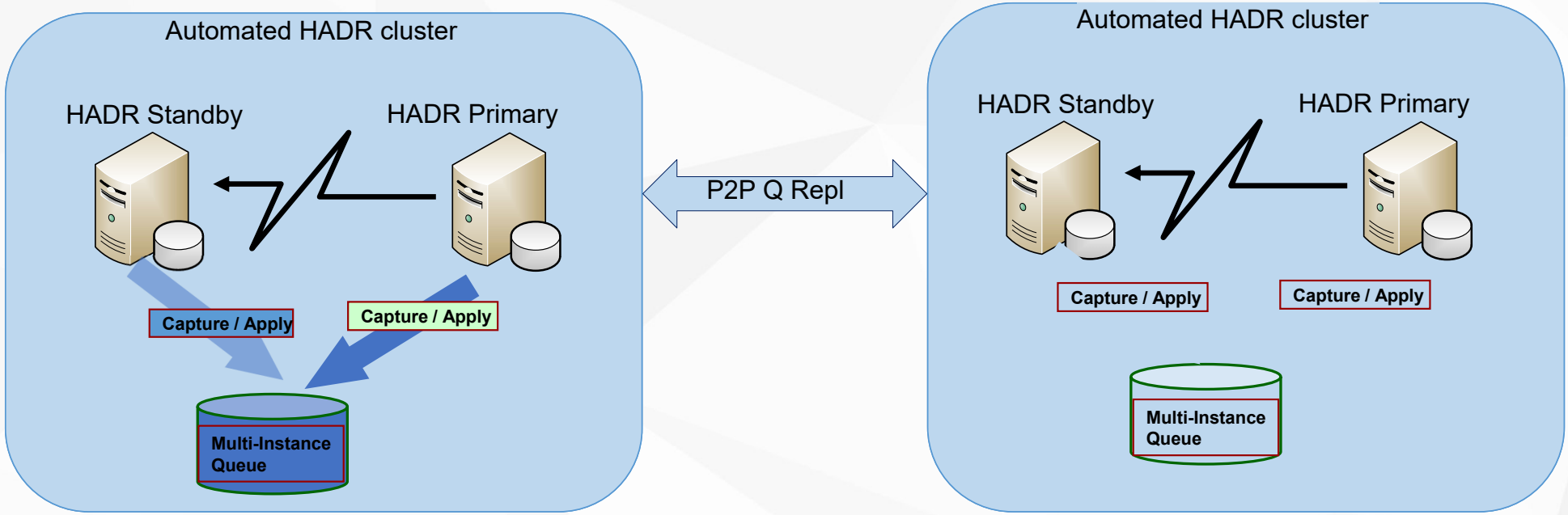
Ability to split workload to eliminate conflicts

Combination of Akamai and F5 routing

ACTIVE-ACTIVE DESIGN



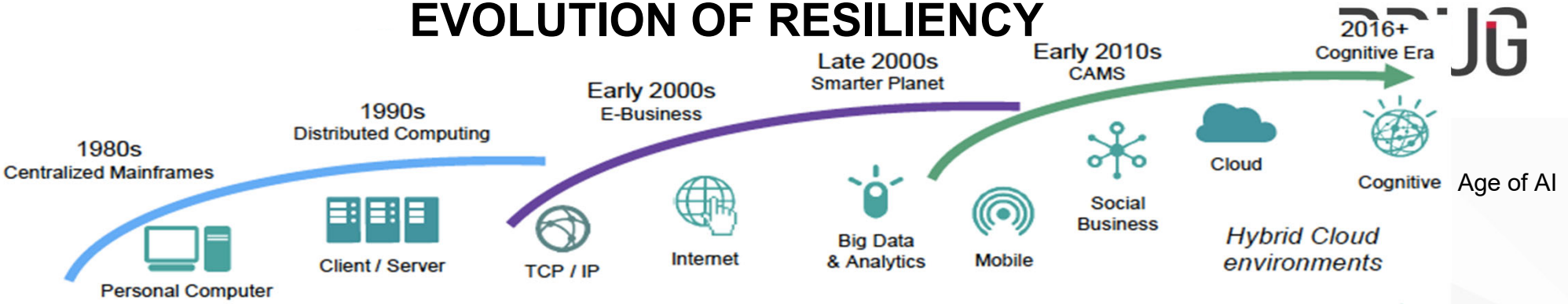
LOCAL ACTIVE-ACTIVE DESIGN MODIFIED



Agenda

- Why “always on”
- What are the Db2 options?
- Combining technologies for the best resiliency
- The Journey to Always On

EVOLUTION OF RESILIENCY



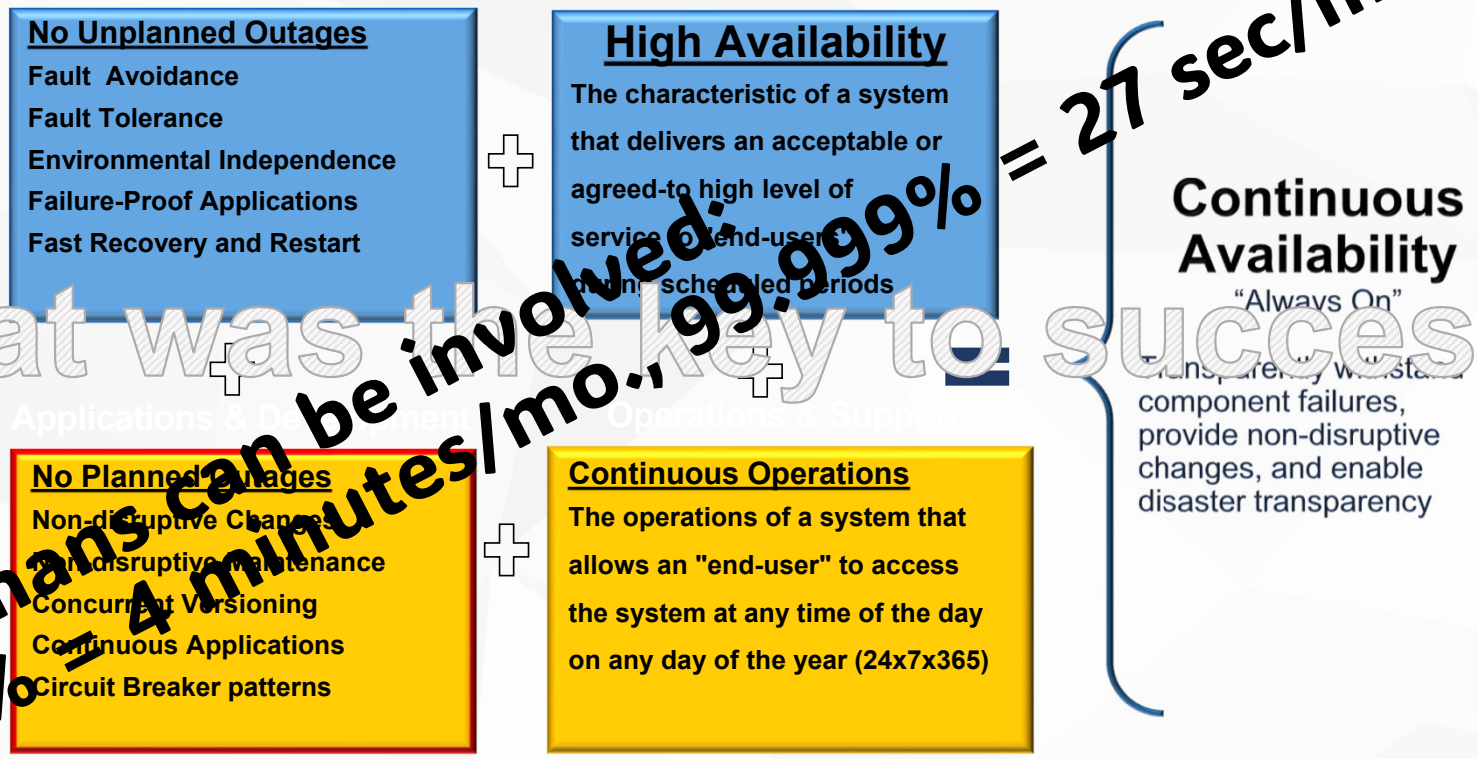
Resiliency has become a business priority triggered by the need for "always on" service and for data protection.
With the growing complexity of hybrid environments, clients are looking for new solutions.

RTO* Trend	"Wave 1" Early 1989 – 2006	"Wave 2" Early 2007 - 2012	"Wave 3" Early 2012 – 2016+	"The next Wave"
	Weeks - Days	*Days - Minutes*	*Hours - Seconds*	*Minutes – Zero downtime*
	✓ **Infrastructure Recovery**	✓ **Cloud Recovery**	✓ **Resiliency for Business Process, Application, Data, & Compute**	✓ **Software Defined Resiliency**
	▪ Focus on recovery of systems & data ▪ Tape-based backup ▪ IT drives need for backups	▪ Focus on recovery of systems, data & applications ▪ Disk-based backup / replicas ▪ Regulations drive need for Disaster Recovery	▪ Customer expectations for "Always On" drives need for continuous availability ▪ Recovery of Applications up to business process ▪ Continuous replication ▪ Hybrid / Cloud Resiliency ▪ Automation	▪ Dynamic / cognitive orchestration & provisioning ▪ Continuous replication ▪ Backup & Recovery built in by software vendor ▪ Managing disparate backup & recovery environments as one ▪ Preventive Analytics ▪ Resiliency for IoT

Time

*RTO : Recovery Time Objective

Trend to requiring Continuous Availability



What was the key to success?

No Humans can be involved:
99.99% = 4 minutes/mo., 99.999% = 27 sec/mo.

Why Zero Downtime Now?

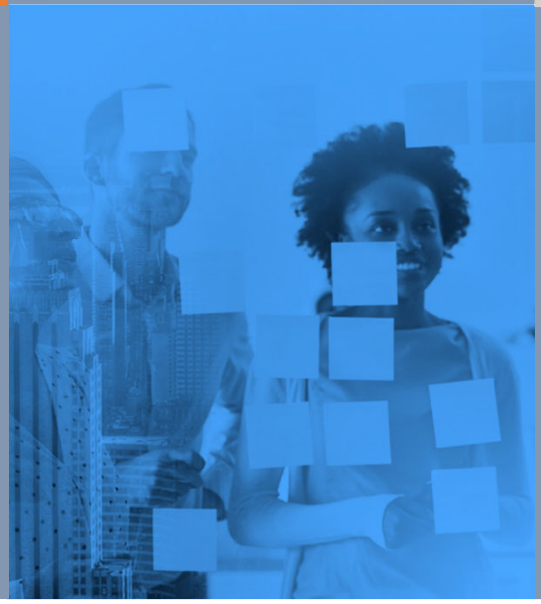
The journey to cloud transformation is underway while business service availability expectations are increasing

Zero Downtime
for planned changes

Platform updates, security updates, application releases, etc. during normal working hours.

DevOps
Continuous Deployment

Zero downtime application releases



99.99%
~4.5 min./mo.
99.999%
~26sec./mo.

3-Active clouds
Multi-region
Auto-mitigation
No planned downtime

Everything breaks, plan on it

Multi-region is a must-have
(unless you like recovery time)

Running the business service active in multiple-cloud availability zones and regions enables auto-mitigation of even disastrous regional failures, such as fires, floods and fools (bad operators or miscreants).

It also allows zero downtime for planned changes by changing only one cloud availability zone and region at a time.

$$ParallelAvailability = 1 - \left(\prod_{i=1}^{N} (1 - ComponentAvailability_i) \right)$$

www.ibm.com – Always-on since June 2001

For ibm.com, $2+M USD per hour attributable to our corporate portal

Always On methods are based on people, process, apps and IT.

The apps provide resiliency!

- **PEOPLE:** Manage end-to-end operations as **one focused team** aligned to the business service vs. technology silos
- **PaaS:** Design a Continuously Available Platform with **patterns** that fit your business application requirements (RPO & RTO)
- **Business Applications**: Mandate Application Architecture **patterns** that **fit the platform**
- **Process & Governance**: Ensure business, development, and operational processes are integrated, agile, and focused on the availability of the service – **know how it works, know how it fails**

What is the most challenging aspect?

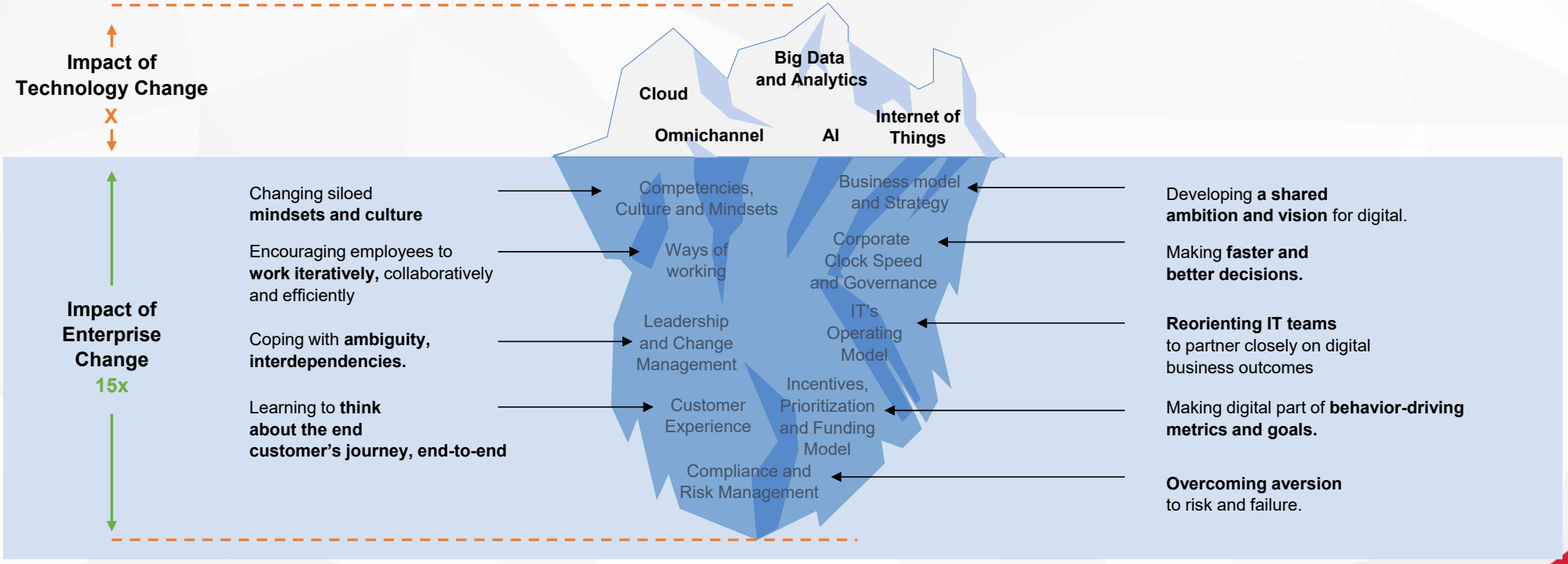
Psst...the IT is the easy part

Technology is the easy part

Changing delivery teams to be aligned to the business services, evolving service management from reactive to proactive, and implementing Site Reliability Engineering practices often are a longer journey than the technology transformation.

Transformation is more than technology

Behavior and Talent-Related Changes and Challenges for Digital Transformation



THINK DIFFERENTLY - DIGITAL FORMS OF ENGAGEMENT ARE DRIVING DEMAND FOR ALWAYS-ON SERVICES

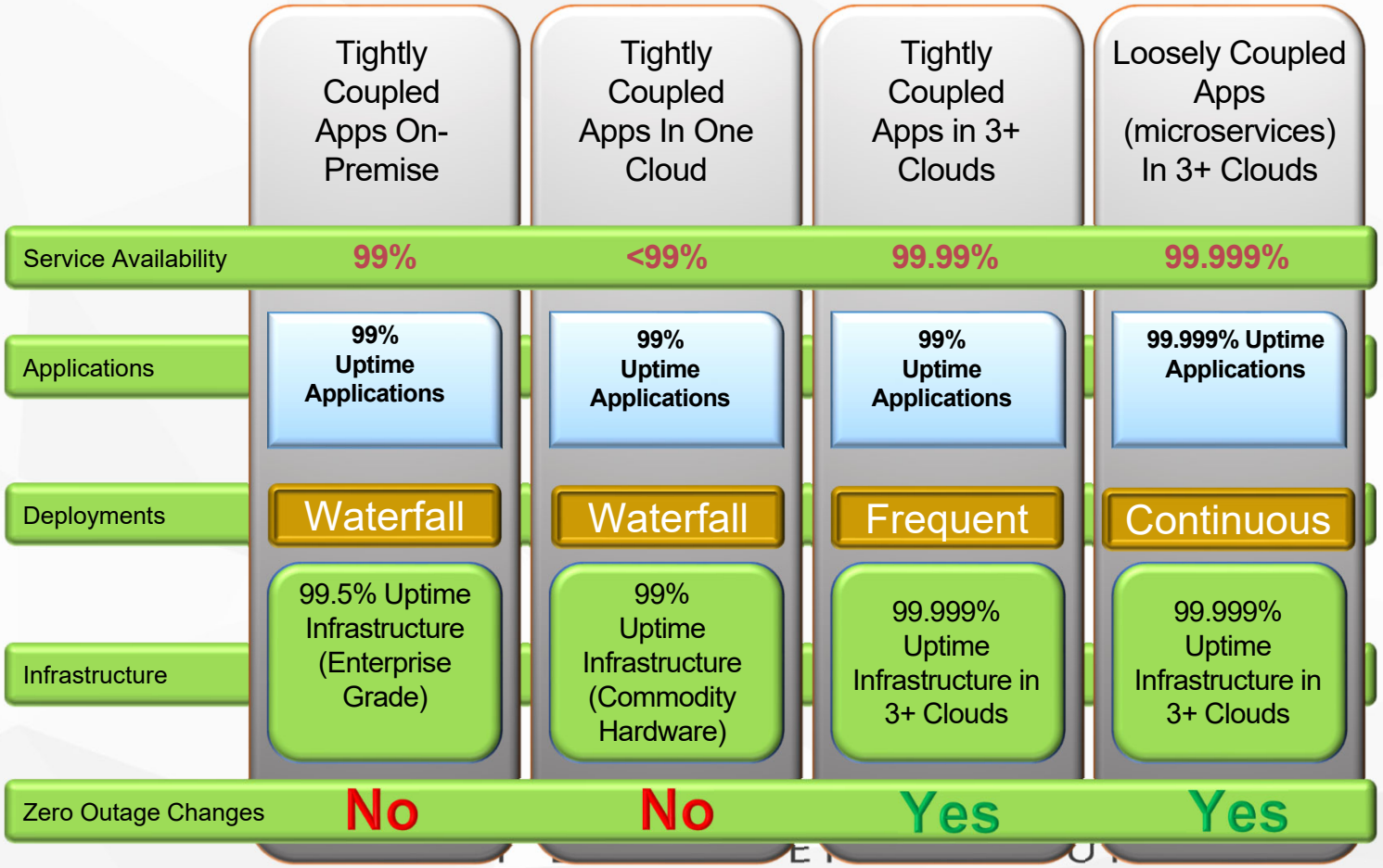
Consider **deploying** cloud enabled and cloud native SoE workloads in **two or more regions**. This pattern enables disaster avoidance rather than disaster recovery, allowing for digital services availability even with an entire region being down. Enterprises also gain greater agility required of dev/ops processes including Continuous Deployments with zero downtime.

ALWAYS-ON

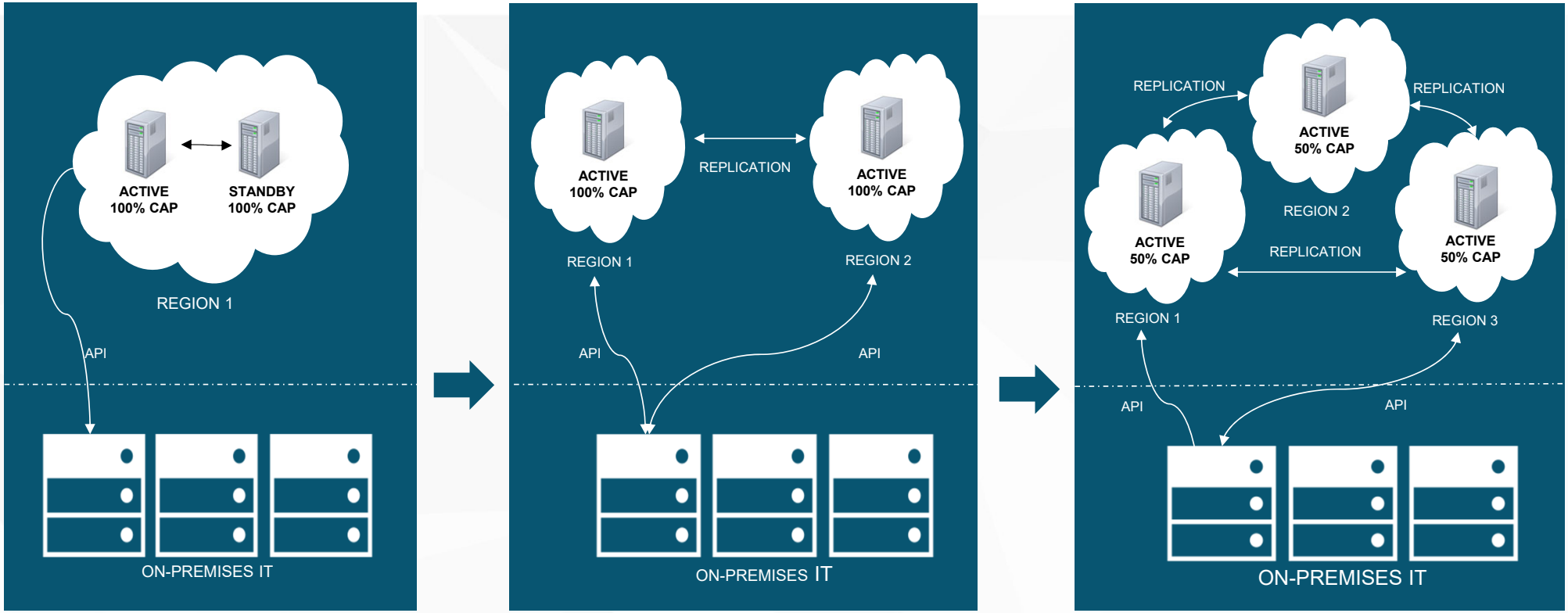
Protects against large-scale infrastructure and data center outages.
Zero downtime. Near-zero data loss.
Reduced program management requirements.

THE CLOUD DOESN'T MAKE AN APPLICATION AGILE OR RESILIENT, THE APP DOES BY RUNNING IN MULTIPLE CLOUD REGIONS

PDUG



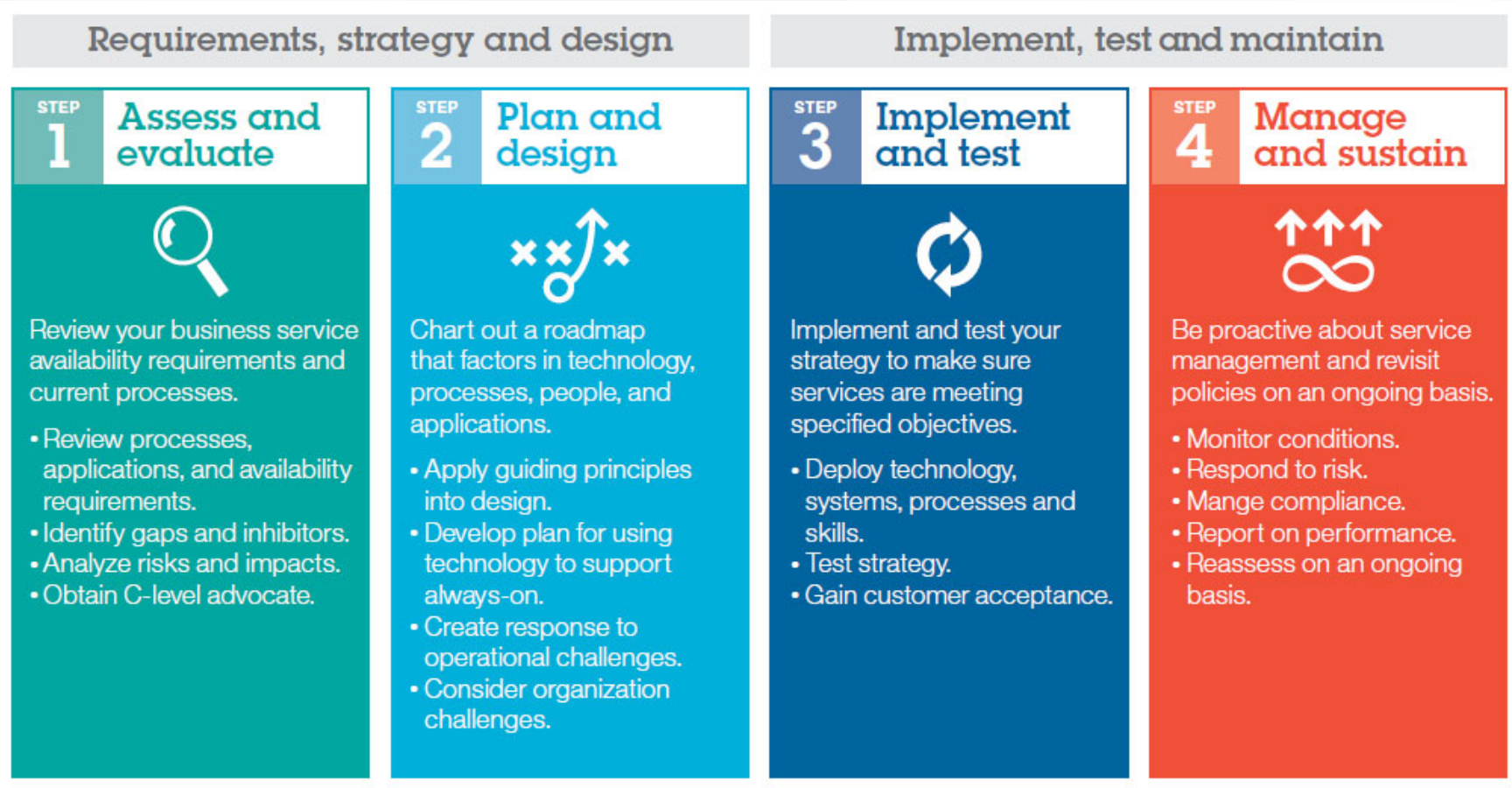
Always On Hybrid Cloud Approach



Production Capacity	200%	200%	150%
Platform Availability	99.5%	99.999%	99.99999%
Failure Impact	100%	50%	33%
Maintenance Windows	Yes	Sometimes	No

The Journey to Always On

Follow these four steps to build an always-on platform aligned to your business goals.

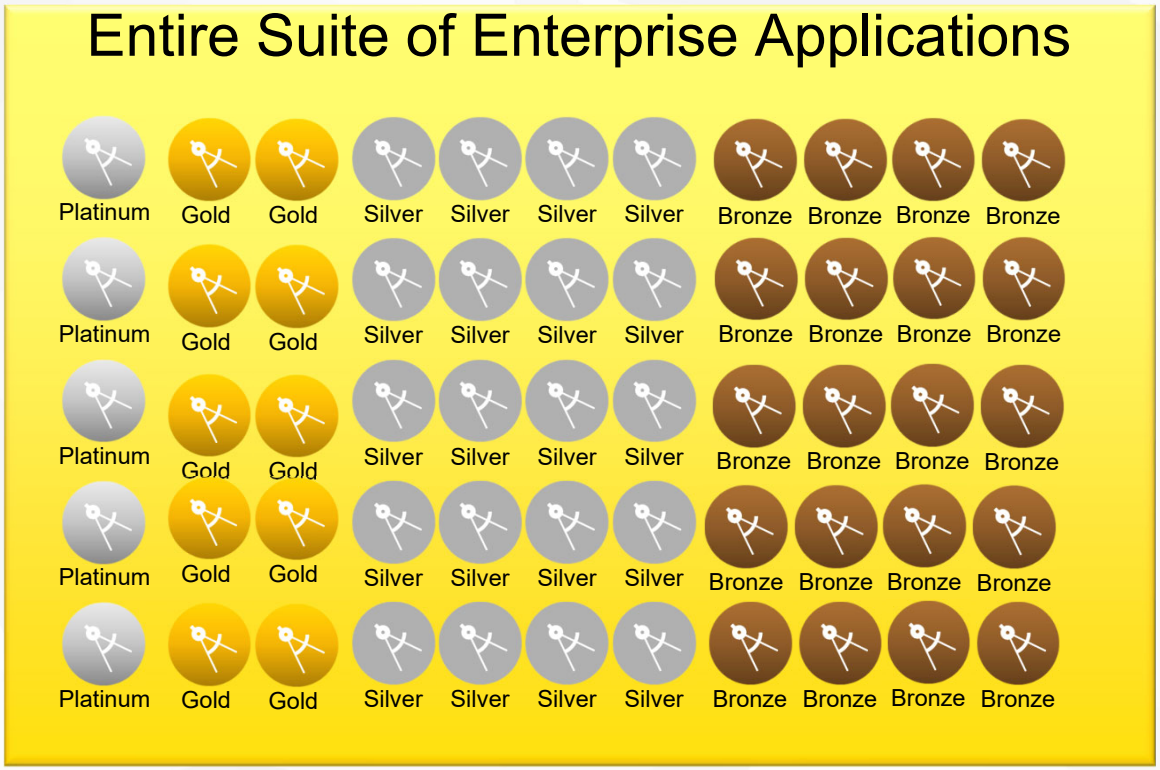


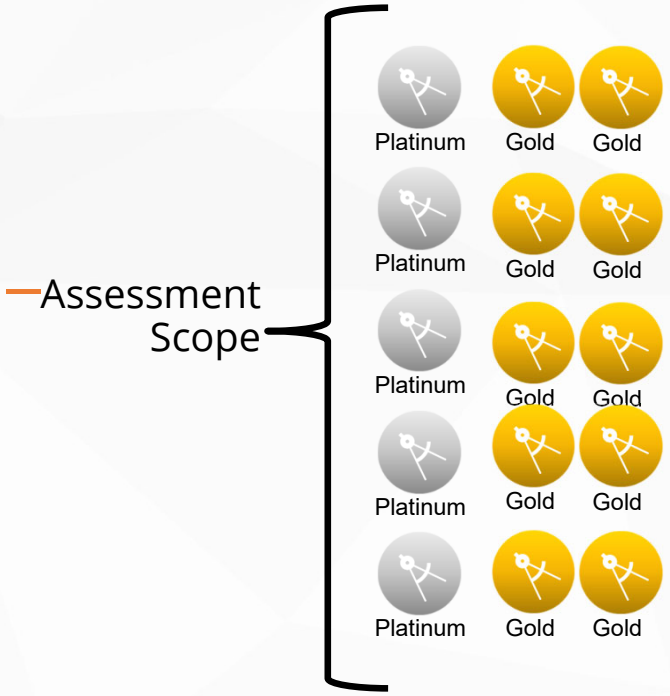
Assess and Evaluate the Business Service Requirements

Business Criteria	Platinum <5% Continuous Availability "Always On"	Gold <5% Near Continuous "Almost Always On"	Silver 20-40% High Availability "Usually On"	Bronze 50-70% Moderate Availability
Business Function	✓ Targeted to applications & business functions that, if unavailable, will result in either financial or legal penalties based on regulatory restrictions	✓ Targeted to business applications and functions that present a potentially broad impact across the internal organization	✓ Targeted to applications that support analysis of business functions	✓ Targeted to non-critical, back-end, offline business functions
Business Impact	✓ Typically assigned to the **5-10%** of applications that drive revenue & profits	✓ During critical processing windows, must be available	✓ Typically backend processes with minimal impact to higher class services	✓ Typically less desirable methods are available to achieve same business function to support tolerance for extended outages
Tolerance for Downtime	✓ Ability to provide continuous availability 24x7x365	✓ Ability to provide constant availability within a defined processing window with availability requirements reduced outside the window	✓ Ability to provide consistent availability within a defined processing window	✓ Availability desired but not mandated with extended outages tolerated by business
Component Failure Impact	✓ Component and regional failures will not cause disruption in service	✓ Component failures should not present a disruption in service	✓ Redundancy at the subcomponent level limits outages based on a single subcomponent failure	✓ Potential outages due to single points of failure inherent within technology & application design
Maintenance and Change Impact	✓ Maintenance & changes required to be concurrent and/or staggered, with no interruption to service	✓ Maintenance & changes required to be concurrent or predefined outage window for change introduction	✓ Maintenance & changes require predefined outage window where changes can be introduced	✓ Maintenance & changes require a liberal outage window where changes can be introduced

Continuous Operations Required!

Only the Platinum and Gold Applications/Business Services are what we focus on for Always On Journey

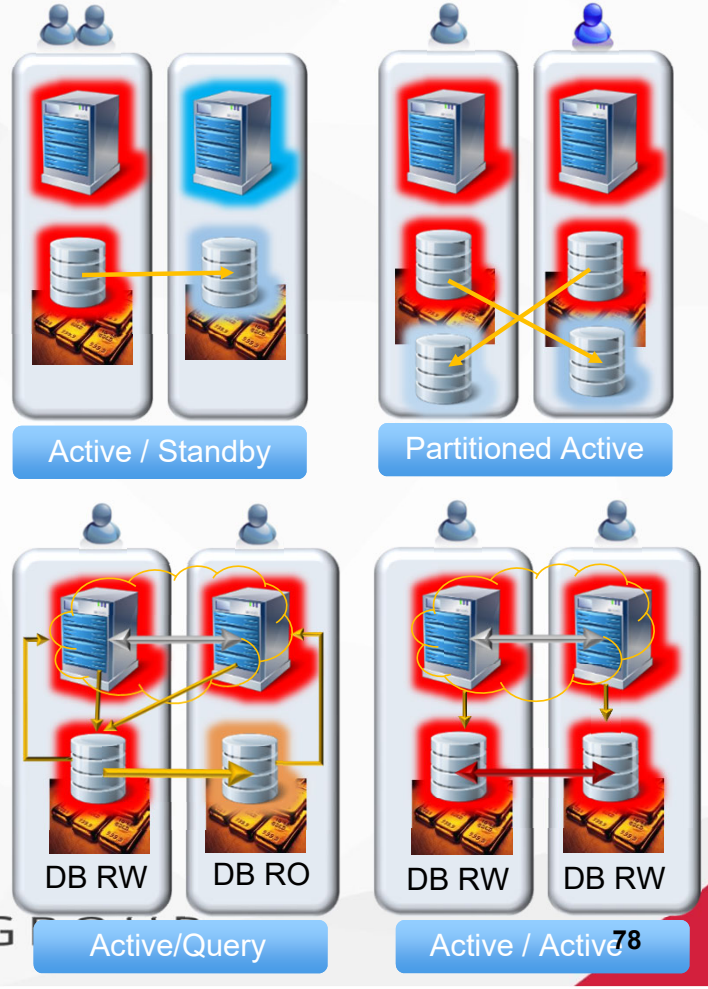




» ALDM, RAD and APOC, are key enabling assessments that when taken into consideration facilitate the application evaluation and planned modernization for resilient applications.

Assess the Application's Resiliency Potential

App Architecture	Resiliency Description
Active / Standby Bronze App	•Traditional DR or warm standby environment • RTO = hours to days • RPO=0?
Partitioned Active (No WAN Clustering, unidirectional DB replication w/failover) Silver App	•Each site application cluster runs independently, as do the DB's. Users are directed to one or the other site. DB's send records to System of Record • RTO=hours • RPO=0?
Active / Query (WAN replication, unidirectional DB replication w/failover) Gold App	•Each site application cluster live, reads performed from local DB, writes performed on primary DB only. • RTO = minutes to hours • **RPO=0 to seconds**
Active / Active (WAN replication & bidirectional DB replication) Platinum App	•All applications uncoupled and databases read writeable • RTO = seconds to minutes • **RPO = 0 to seconds**



Design Challenge: Brewers CAP Theorem – "Pick Two"

PDUG

It is impossible for a distributed computer system to simultaneously provide all three of the following guarantees:

1. **Consistency** - all distributed nodes have a single up-to-date copy of all data at all times
2. **Availability** – every request receives a success/failure response
3. **Partition Tolerance** – system continues to run despite arbitrary message loss or failure of part of the system. e.g. The network stops delivering messages between server sets.


- Consistency


- Availability


- Partition Tolerance

<u>Consider your SoE vs SoR requirements separately:</u>

- Systems of Engagement = Availability + Partition Tolerance
- Financial Data = Consistency + Partition Tolerance
- Most Other Data = Availability + Partition Tolerance

79

ALWAYS ON GUIDING PRINCIPLES (1)

PDUG

1. **Core Principles** – transparently withstand component failures, provide non-disruptive changes, and enable disaster transparency
2. **Think Differently** – legacy architectural practices no longer apply
3. **KISS** – Keep It Simple Stupid, complexity adds obfuscation and prolonged service recovery
4. **Concurrent Versioning** – non-disruptive changes is the ability to run two versions at once
5. **Continuous Operations** – design in platform concurrency to enable non-disruptive changes
6. **Design each "cloud" identically** – best practices should be followed per "cloud", then interconnect
7. **Fail Small** – everything breaks, minimize the impact in design
8. **Virtualize Nearly Everything** – Virtualization provides flexibility and mobility, both essential
9. **Automate Nearly Everything** – avoid human error and inconsistency
10. **Design For Failure – know how it works, know how it breaks** and how to mitigate it's impact
11. **Applications Must be Designed for Failure** – fail gracefully, minimize impact to consumer
12. **Avoid HA Takeover – service parallelism** (clustering) is more reliable and faster
13. **Availability is provided by peer "clouds"** – failure in one "cloud" doesn't impact the others, the fault domain is isolated to each "cloud", service is still functional in the other(s)
14. **Share Nothing** – each cloud must be able to provide the business service independently, perhaps with reduced capacity (contingency planning enables critical functions during capacity reduction)

ALWAYS ON GUIDING PRINCIPLES (2)

PDUG

15. **Availability Zones** – CA, near CA, and HA environments have their own architectural requirements and change windows, keep them separate, share nothing
16. **Add Global Traffic Management** – routes consumers to the best "cloud" to consume the service. Domain Name Service based, closely coupled with SLB and DNS services
17. If application must maintain state across "clouds**", use in memory application grid** – fast & tolerant and sessions must be small to take advantage of this technology, else don't use sessions beyond individual "cloud"
18. **Add Application Level Data Replication** – capture and apply changes to all peers. In order to provide fast failover or transparent service bypass, logical data replication is required to avoid human tasks. Bi-directional peer-to-peer allows writes anywhere, but OoR induces eventual data consistency.
19. **Never stretch a cluster across "clouds"** – extends fault domain beyond individual cloud
20. **Include Out of Region** – must mitigate 3-F's (Fire, Flood, and Fools) outside region, integrate it into your change practices
21. **Don't Forget Security** – the "Fools" can cause unexpected damage
22. **Don't Forget Performance Engineering** – Development must embrace performance engineering. Business must make development and operations aware of any planned media events that may bring "flash mobs" very early. Applications must be efficient. IT must be sufficient.

These guiding principles build upon the many guiding principles common in HA and DR design and are here to guide practitioners beyond core HA design.

Always On New Technologies Deeper Dive

In order to run resilient clouds, we need to introduce:

- **Global Traffic Management**
  - Resolve www.ibm.com to the best responding clouds IP addresses
- **Session Grid**
  - I put these items in my cart and hopped clouds, cart's still full
- **Data replication**
  - Create, Read, Update, Delete data anywhere and everywhere
- **Ops Dashboard**
  - Business service XYZ is spitting errors in cloud 2, bypass it

- ✓ Global Traffic Management
  - ✓ Sends end user to the best cloud using Domain Name Service
- ✓ Session Grid
  - ✓ If Apps not session-less, need session grid to synchronize
- ✓ Application Level bi-directional, multi-master, peer-to-peer async data replication
  - ✓ Synchronize data
- ✓ "Single pane of glass" perspective of all cloud transactions and errors

3-Active “Always On” Method born in IBM in 1998

PDUG

People and Process – The Most Challenging

- Proactive End to End Fully Managed Model
- Technical Leaders Responsible for the Service
- Dedicated SME staff aligned to business services
- Global distributed staff for 24x7x365 coverage
- Virtual Colocation for efficient communications
- Daily Change Management calls, Agile Delivery, Continuous Operations

IT Technology – The Easy Part!

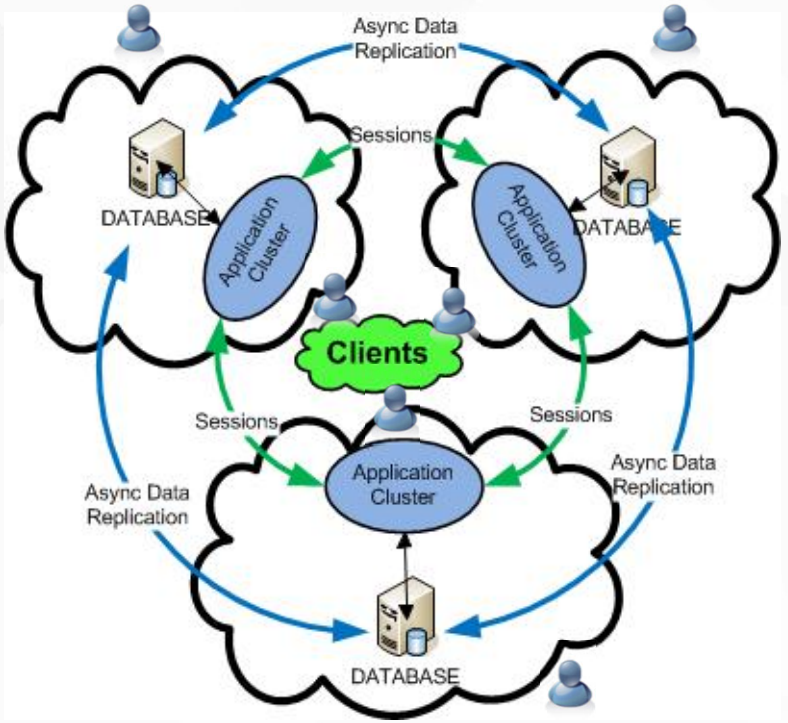
- Platform as a Service designed for Always On
- Identical in all 3 locations – all live
- Automated and end-to-end monitored
- No HA takeover, all components live

Enabling Technologies not in HA Solutions:

- Global Traffic Management
- Application Session Replication Grid
- Bidirectional Peer-to-peer Logical Data Replication

Application Developers Must Think Differently

- Platform mandates Non-Functional Requirements
- Think “Integrate across the WAN”
- Non-destructive updates/releases/schema changes
- Eventual Data Consistency
- Must Generate Unique Indexes/Keys/etc.
- Explicit SQL required for data conflict remediation
- Application may implement “soft locks” to mitigate data conflicts



3-Active Platform as a Service
Capacity 150%

Questions?
Dale McInnis dmcinnis@ca.ibm.com

IBM RedPapers and Redbooks for more Always On Multi-Active Concepts
(Google "IBM Always On Redpapers"):
- "Always On: Assess, Design, Implement and Manage Continuous Availability"
http://www.redbooks.ibm.com/abstracts/redp5109.html?Open
- The Value of Active-Active Sites with Q Replication for IBM DB2 for z/OS An Innovative IBM Client's Experience
http://www.redbooks.ibm.com/abstracts/redp5140.html?Open

Assessment Offerings:
ALDM (Application Discovery & Dependency Mapping) https://w3.gsar.ibm.com/services/gsar/gda/sbb_details.xhtml?id=198
RAD (Resilient Architecture Design) for IT Infrastructure
https://w3-03.ibm.com/tools/cm/iram/assetDetail/generalDetails.faces?guid={E165A380-BF03-D415-D99D-05365A442683}&v=1.0&submission=false
APOC (Application Performance Optimization Consulting)
https://w3-03.ibm.com/tools/cm/iram/assetDetail/generalDetails.faces?guid=E8F68785-C147-47F9-27B7-21CA47102688&v=1.0&submission=false

Application Modernization – Cloud Native Application Design
- Top 9 Rules for Cloud Applications - Kyle Brown, IBM Distinguished Engineer , Bluemix
http://www.ibm.com/developerworks/websphere/techjournal/1404_brown/1404_brown.html

- "Going Cloud Native" – a great article with links on how to modernize services and organizations
http://CloudNative.online

PDUG

POLISH DB2 USERS GROUP

Thank You

Speaker: Dale McInnis
Company: IBM Canada Ltd.
Email Address: dmcinnis@ca.ibm.com

Please fill out your session evaluation before leaving!