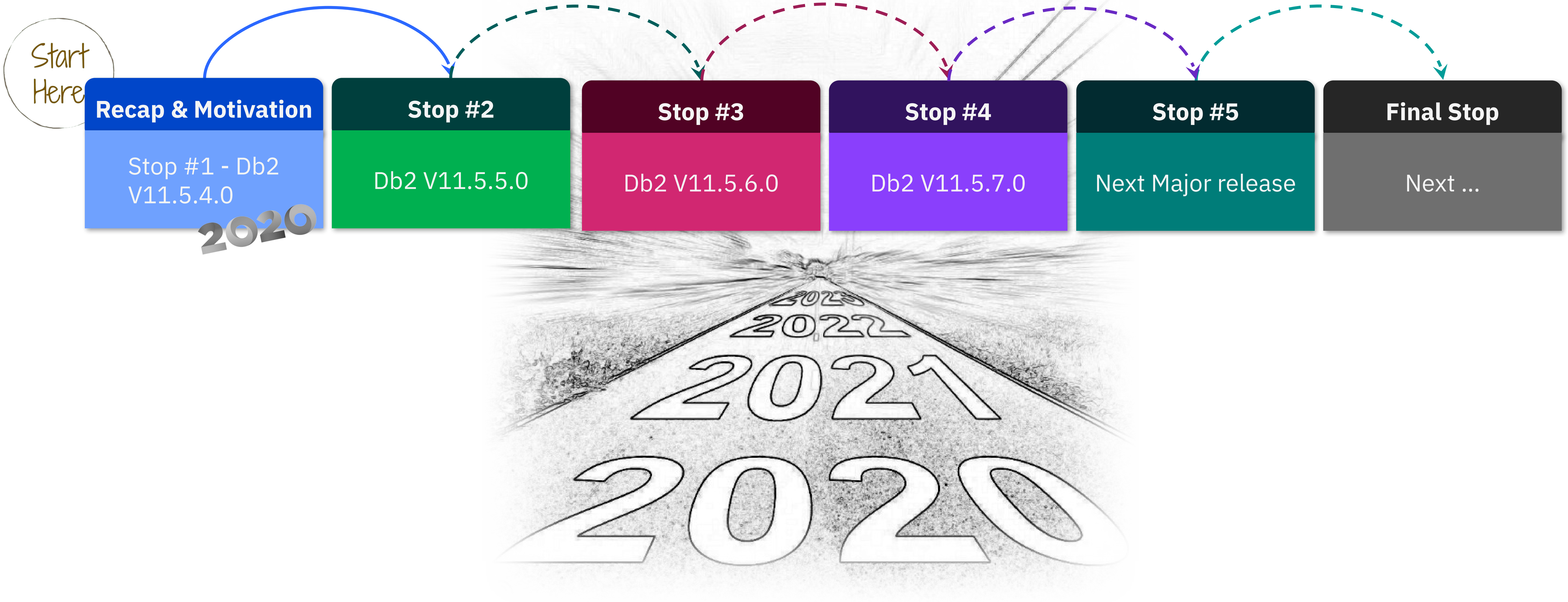
The Next Stop(s) in Db2 Pacemaker HA Solution Journey

Speaker

Toby Haynes

Senior Technical Manager for
Db2 pureScale Development

The Next Stop(s) in Db2 Pacemaker HA Solution Journey - AGENDA



Start Here

Recap & Motivation

Stop #1 - Db2 V11.5.4.0

Stop #2

Db2 V11.5.5.0

Stop #3

Db2 V11.5.6.0

Stop #4

Db2 V11.5.7.0

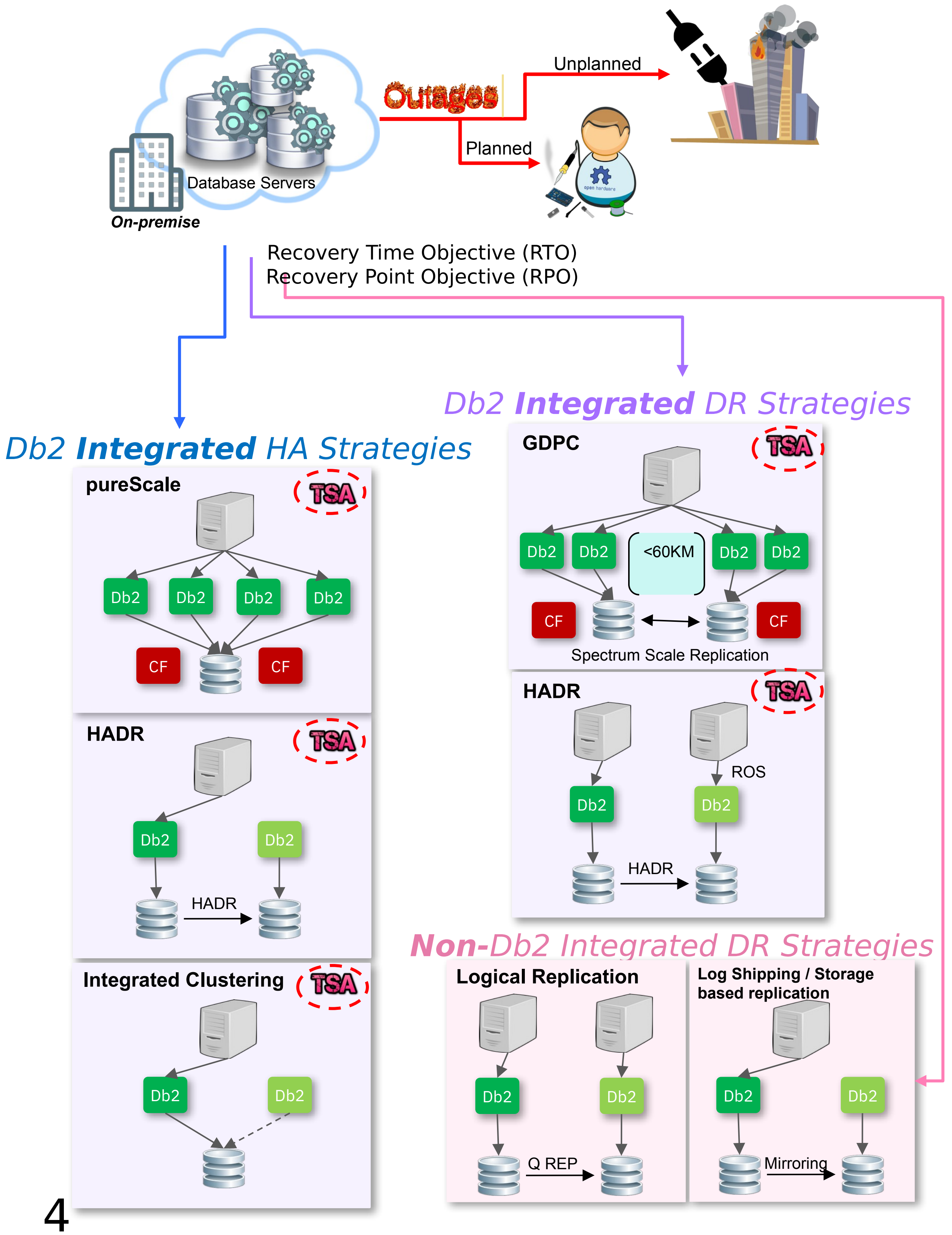
Stop #5

Next Major release

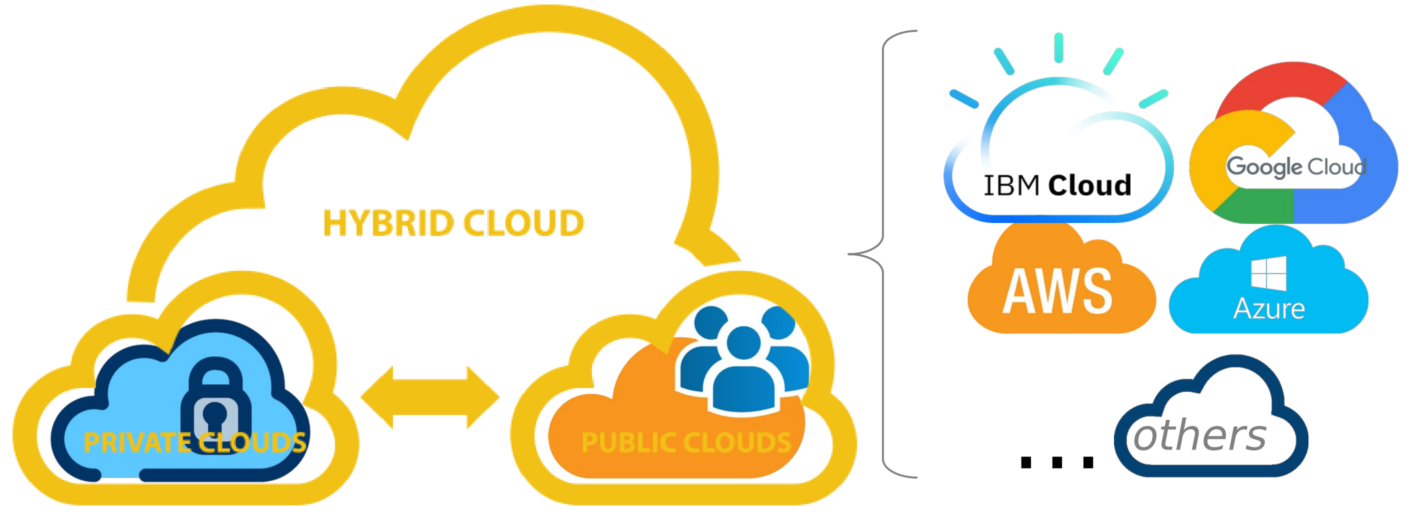
Final Stop

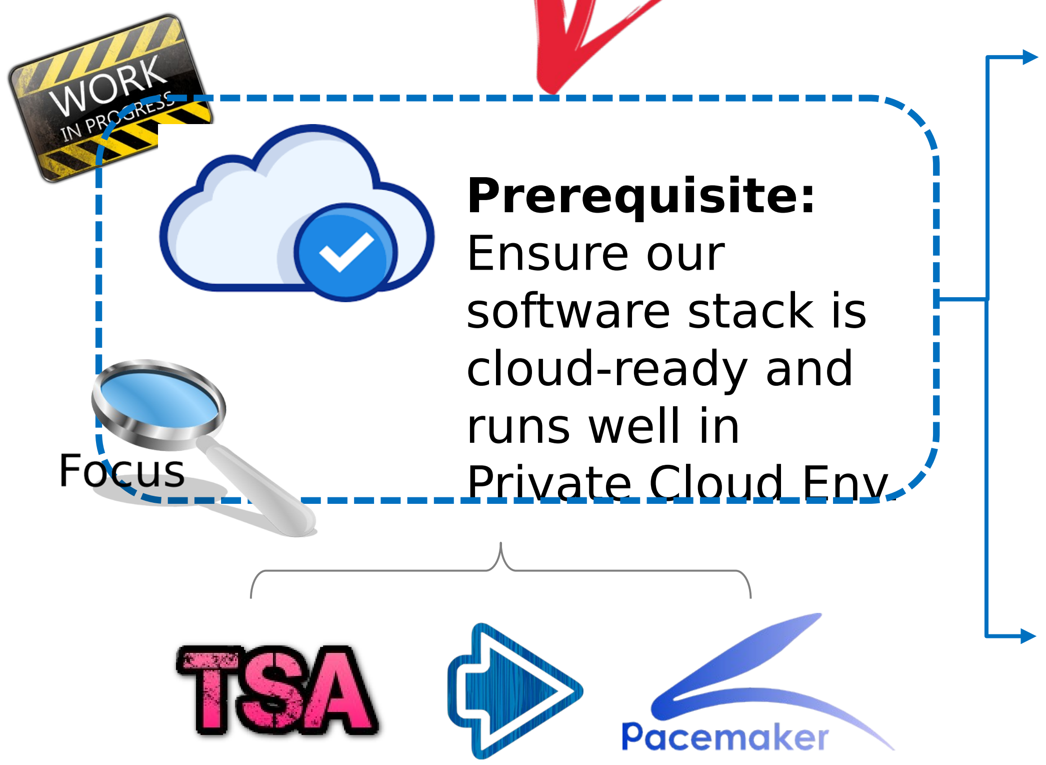
Next ...

Recap & Motivation



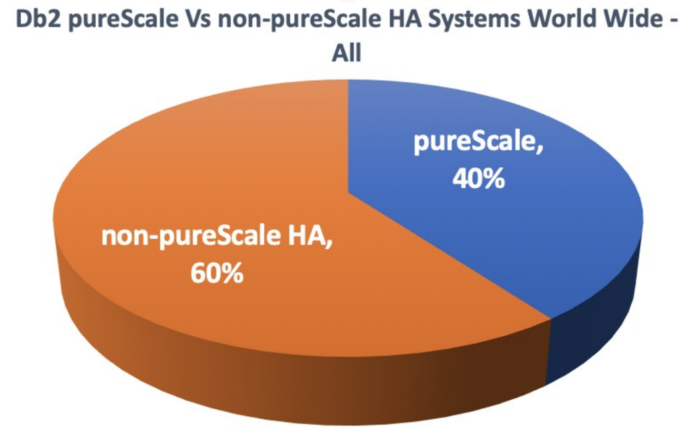
- surging customer requests to deploy on cloud
- *Align with IBM hybrid cloud strategy*





Prerequisite:
Ensure our software stack is cloud-ready and runs well in Private Cloud Env

IMPACT



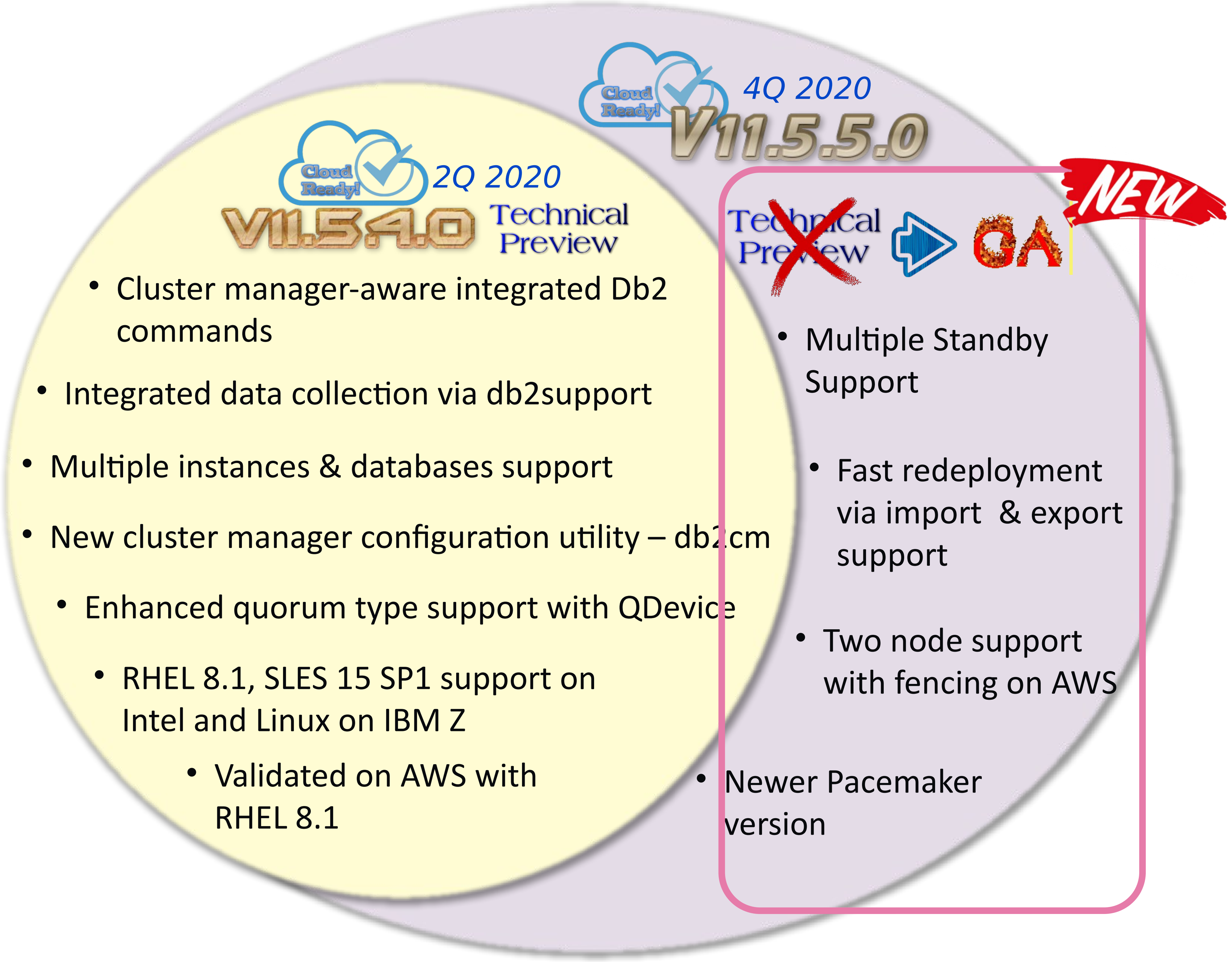
COMPLEXITY

- From *non-pureScale* (HADR, etc.) to *pureScale*

2021 Key Topics:

- From technical Preview to GA
- Key enhancements & focus areas
- What's next ?

Our Journey … Stop #3



- Cluster manager-aware integrated Db2 commands
- Integrated data collection via db2support
- Multiple instances & databases support
- New cluster manager configuration utility – db2cm
- Enhanced quorum type support with QDevice
- RHEL 8.1, SLES 15 SP1 support on Intel and Linux on IBM Z
- Validated on AWS with RHEL 8.1

- Multiple Standby Support
- Fast redeployment via import & export support
- Two node support with fencing on AWS
- Newer Pacemaker version

Note: Roadmap and content subjected to change

Fast re-deployment on same hardware

Backup configuration

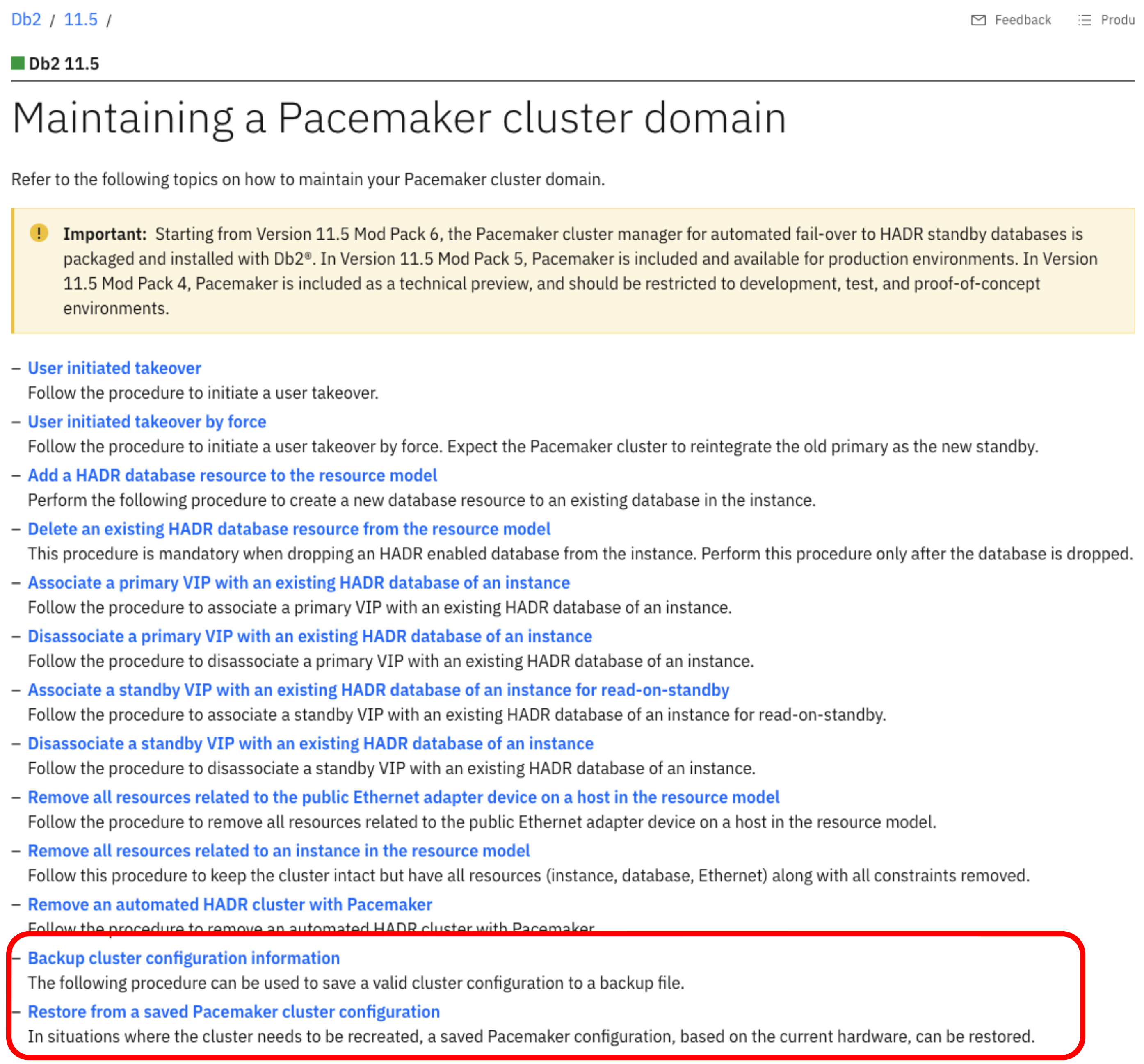
```
[root@jesting1]$ /home/db2inst1/sqllib/adm/db2cm -export /tmp/backup.conf
Exporting configuration to /tmp/backup.conf

[root@jesting1]$ ls -la /tmp/backup.conf
-rw-r--r-- 1 root root 12888 Sep 1 14:22 /tmp/backup.conf
```

Restore configuration (need to clean up existing environment via `db2cm -delete -cluster` first)

```
[root@jesting1]$ /home/db2inst1/sqllib/adm/db2cm -import /tmp/backup.conf
Importing configuration from /tmp/backup.conf
Cluster created successfully.
```

Fast deployment on NEW hardware is possible:

- Requires manual changes to exported file
- Example available in technote off Db2 documentation



Db2 / 11.5 /

Db2 11.5

Maintaining a Pacemaker cluster domain

Refer to the following topics on how to maintain your Pacemaker cluster domain.

Important: Starting from Version 11.5 Mod Pack 6, the Pacemaker cluster manager for automated fail-over to HADR standby databases is packaged and installed with Db2®. In Version 11.5 Mod Pack 5, Pacemaker is included and available for production environments. In Version 11.5 Mod Pack 4, Pacemaker is included as a technical preview, and should be restricted to development, test, and proof-of-concept environments.

- **User initiated takeover**
  Follow the procedure to initiate a user takeover.
- **User initiated takeover by force**
  Follow the procedure to initiate a user takeover by force. Expect the Pacemaker cluster to reintegrate the old primary as the new standby.
- **Add a HADR database resource to the resource model**
  Perform the following procedure to create a new database resource to an existing database in the instance.
- **Delete an existing HADR database resource from the resource model**
  This procedure is mandatory when dropping an HADR enabled database from the instance. Perform this procedure only after the database is dropped.
- **Associate a primary VIP with an existing HADR database of an instance**
  Follow the procedure to associate a primary VIP with an existing HADR database of an instance.
- **Disassociate a primary VIP with an existing HADR database of an instance**
  Follow the procedure to disassociate a primary VIP with an existing HADR database of an instance.
- **Associate a standby VIP with an existing HADR database of an instance for read-on-standby**
  Follow the procedure to associate a standby VIP with an existing HADR database of an instance for read-on-standby.
- **Disassociate a standby VIP with an existing HADR database of an instance**
  Follow the procedure to disassociate a standby VIP with an existing HADR database of an instance.
- **Remove all resources related to the public Ethernet adapter device on a host in the resource model**
  Follow the procedure to remove all resources related to the public Ethernet adapter device on a host in the resource model.
- **Remove all resources related to an instance in the resource model**
  Follow this procedure to keep the cluster intact but have all resources (instance, database, Ethernet) along with all constraints removed.
- **Remove an automated HADR cluster with Pacemaker**
  Follow the procedure to remove an automated HADR cluster with Pacemaker.
- **Backup cluster configuration information**
  The following procedure can be used to save a valid cluster configuration to a backup file.
- **Restore from a saved Pacemaker cluster configuration**
  In situations where the cluster needs to be recreated, a saved Pacemaker configuration, based on the current hardware, can be restored.

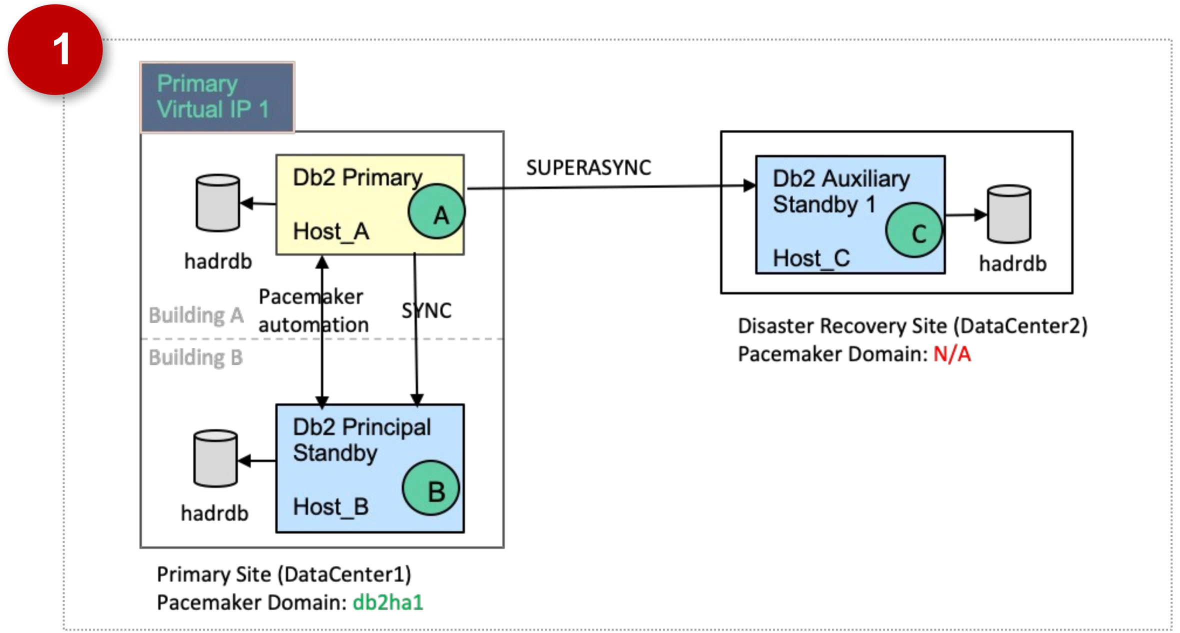
Link to Db2 Doc

Multiple Standby Support

Flexible deployments:

- Up to 3 standbys for each HADR DB: 1 principal standby and up to 2 auxiliary standby.
- Auxiliary standbys can be in 1 or 2 sites that is same or different than the primary
- Automatic failover supported between Principal Primary and Principal Standby
- Manual takeover required from auxiliary standby
- Same support as with TSA today

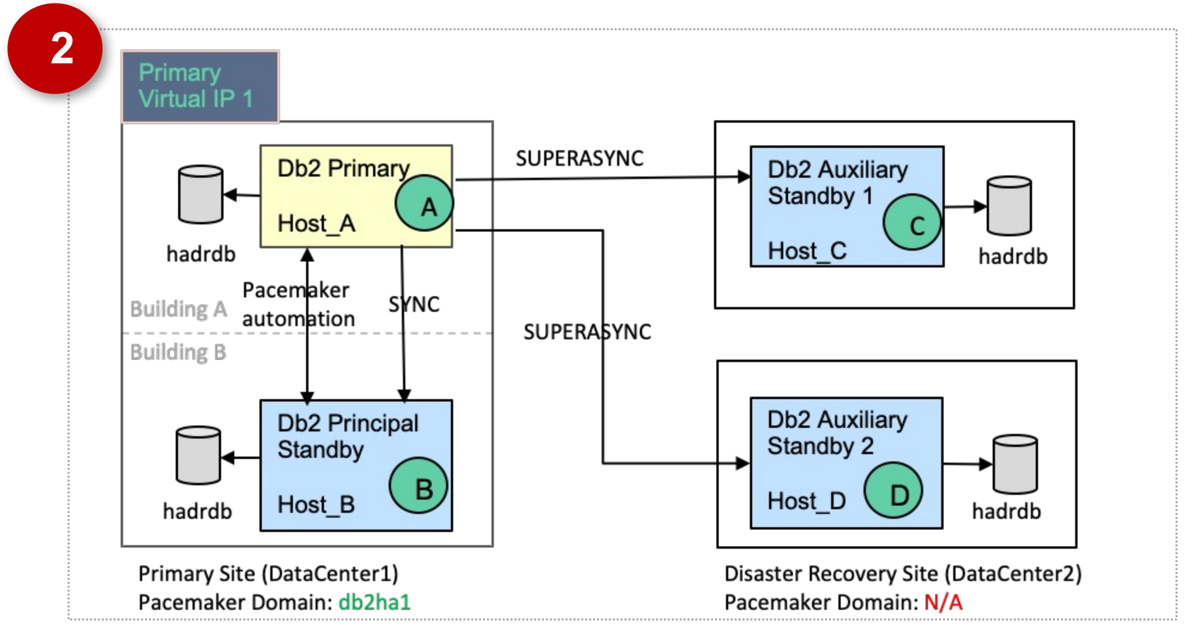
1



Disaster Recovery Site (DataCenter2)
Pacemaker Domain: N/A

Primary Site (DataCenter1)
Pacemaker Domain: db2ha1

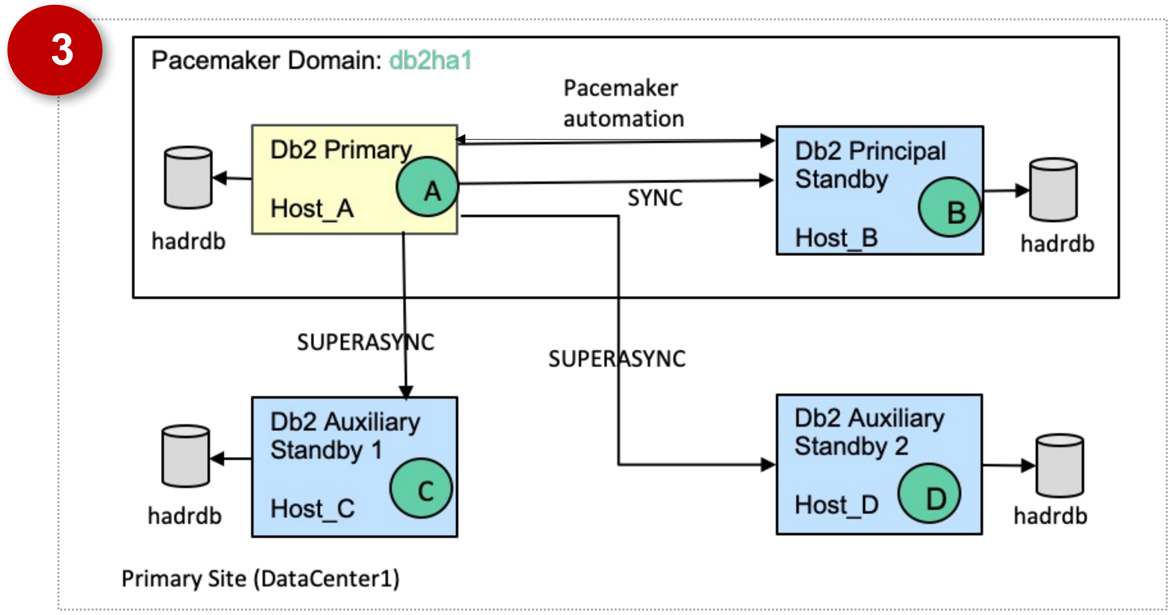
2



Primary Site (DataCenter1)
Pacemaker Domain: db2ha1

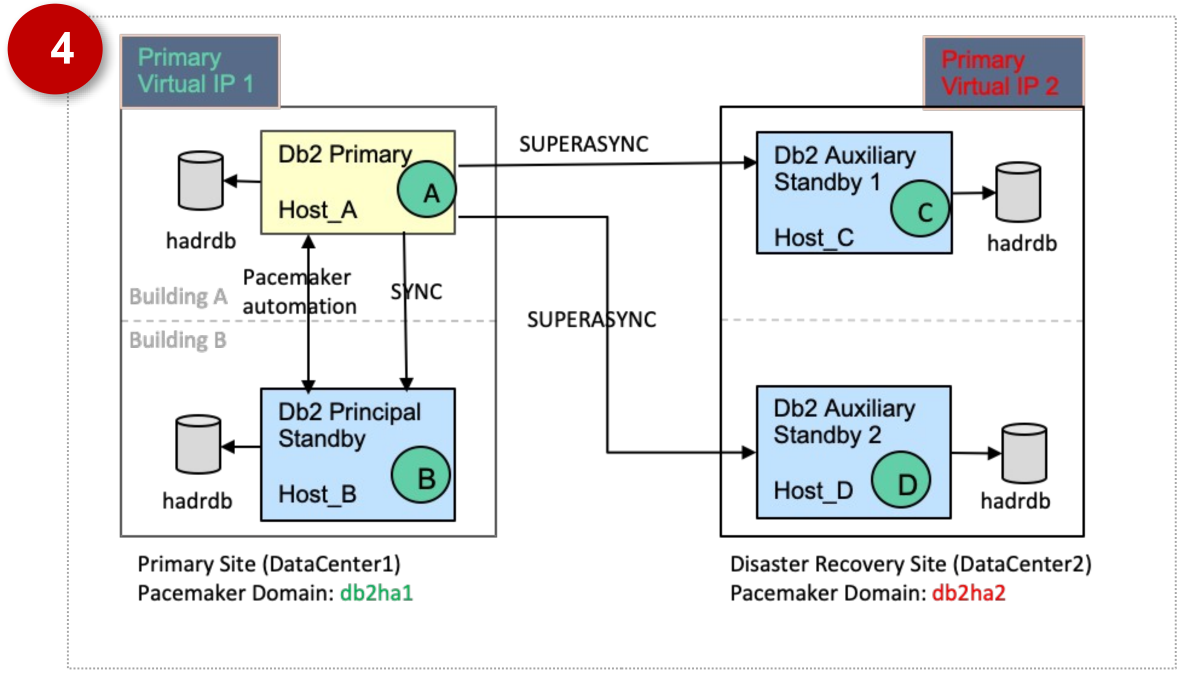
Disaster Recovery Site (DataCenter2)
Pacemaker Domain: N/A

3



Primary Site (DataCenter1)

4



Primary Site (DataCenter1)
Pacemaker Domain: db2ha1

Disaster Recovery Site (DataCenter2)
Pacemaker Domain: db2ha2

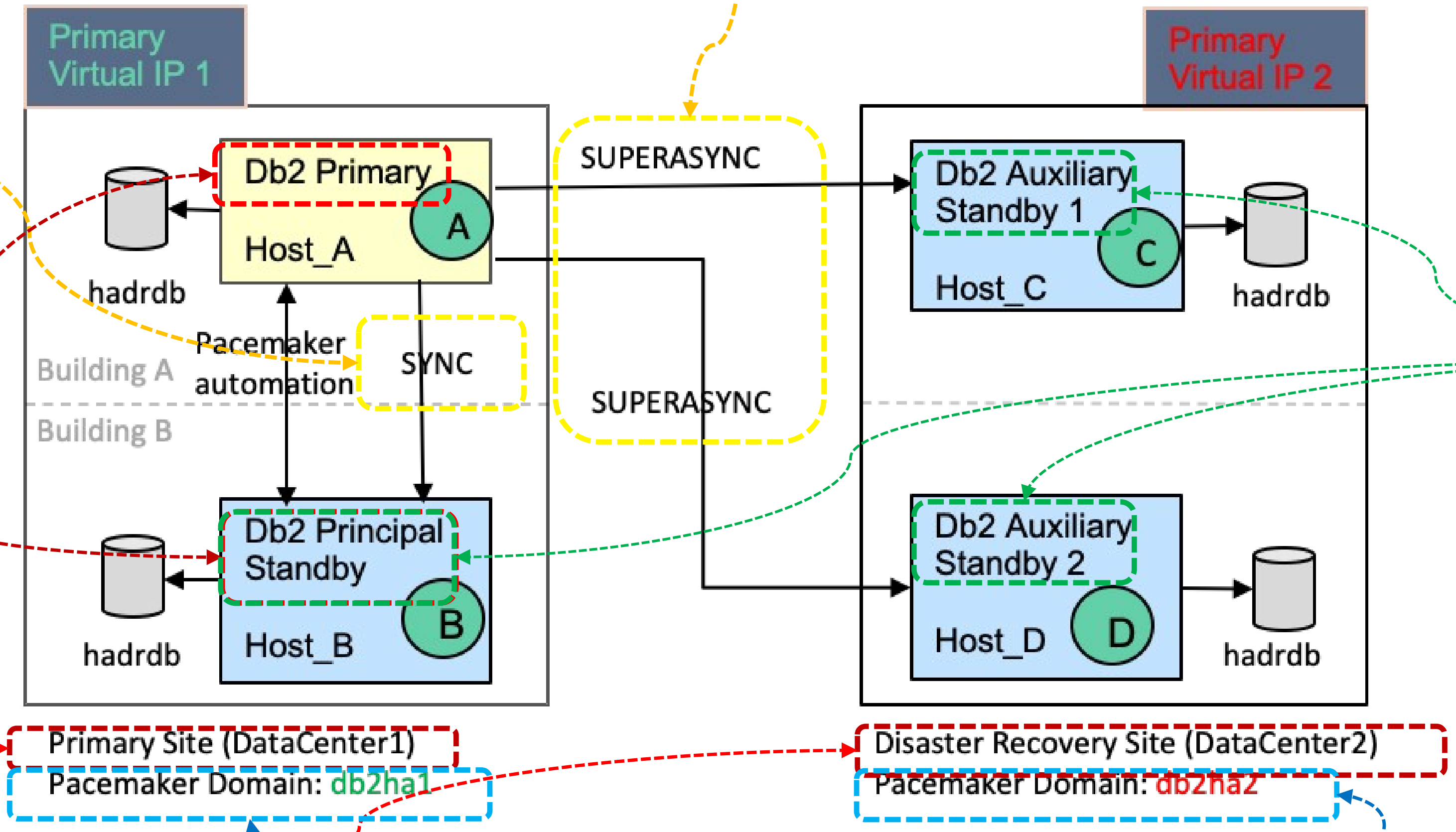
Best Practice Configuration

Detail on Best Practice 2-sites 3 Standbys setup

SUPERASYNC is the effective mode for all auxiliary standbys

Principal Standby can use any either SYNC or NEARSYNC modes

Principal Primary and Principal Standby always reside in the same domain



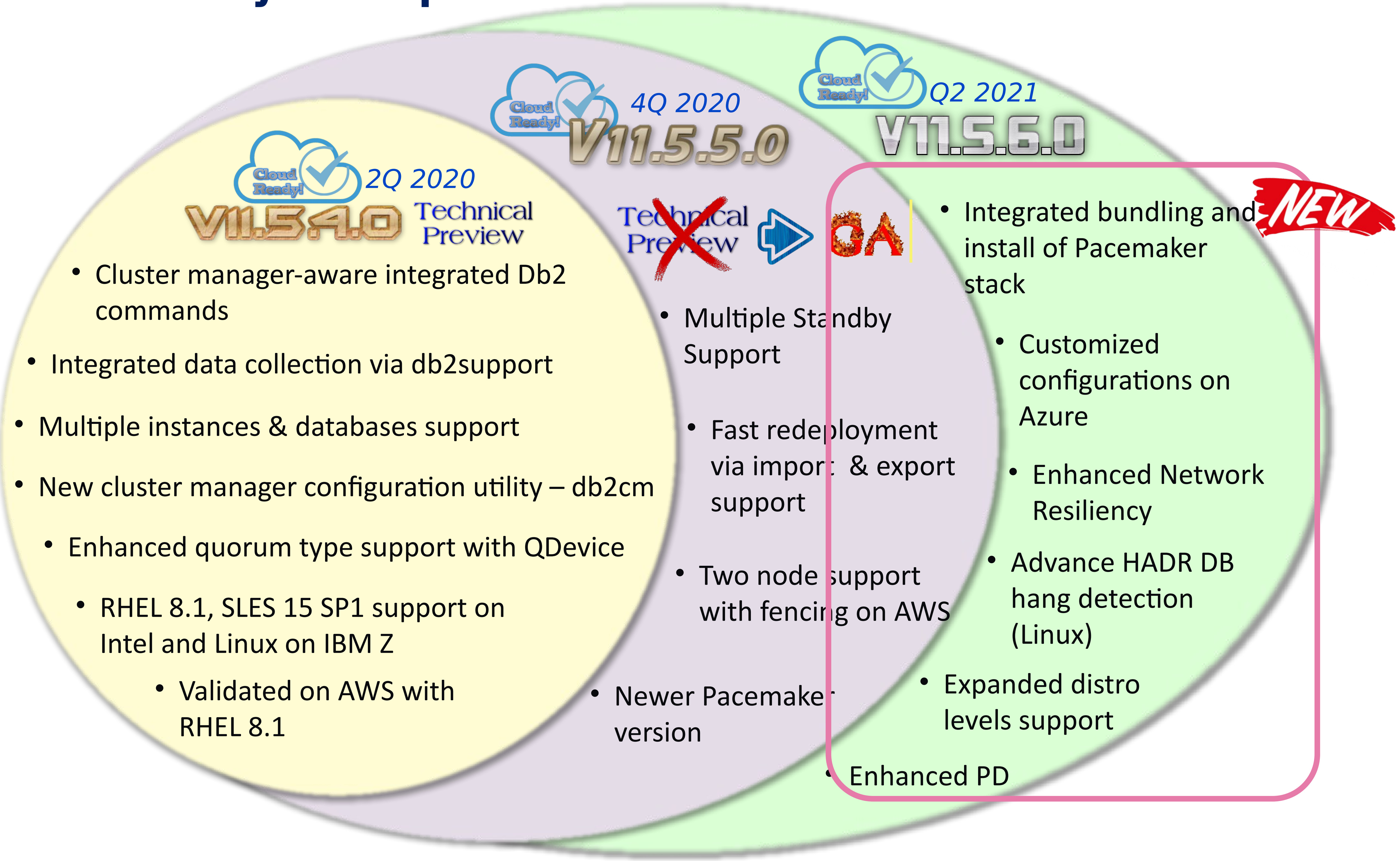
- Reads on standby is supported on all 3 standbys (DB2_HADR_ROS=ON)
- Time-delayed log replay supported on all 3 standbys (db cfg - hadr_replay_delay)

Two sites with two hosts in each

Two disjoint Pacemaker domains with automated failover enabled within each, but not across

Allows DR site to completely replace primary site with automation enabled by default when a manual takeover is issued on any of the auxiliary standbys.

Our Journey … Stop #3



2Q 2020 V11.5.4.0 Technical Preview

- Cluster manager-aware integrated Db2 commands
- Integrated data collection via db2support
- Multiple instances & databases support
- New cluster manager configuration utility – db2cm
- Enhanced quorum type support with QDevice
- RHEL 8.1, SLES 15 SP1 support on Intel and Linux on IBM Z
- Validated on AWS with RHEL 8.1

4Q 2020 V11.5.5.0

~~Technical Preview~~ GA

- Multiple Standby Support
- Fast redeployment via import & export support
- Two node support with fencing on AWS
- Newer Pacemaker version

Q2 2021 V11.5.6.0

NEW

- Integrated bundling and install of Pacemaker stack
- Customized configurations on Azure
- Enhanced Network Resiliency
- Advance HADR DB hang detection (Linux)
- Expanded distro levels support
- Enhanced PD

Note: Roadmap and content subjected to change

Pacemaker Stack + Db2 Software – ALL in ONE

from

V11.5.4.0 + V11.5.5.0:

- Separate download of Pacemaker software stack
  - available via the IBM hosted – Market Registration Site (MRS)
- Separate installation
  - With guided procedures in Db2 documentation.

TO

V11.5.6.0:

- Integrated bundling of Pacemaker software stack with Db2
- Integrated installation via command line utility – *db2_install* and *installfixpack*
  - silent install and GUI to follow in future release
- MRS only hosts cloud specific RPMs – e.g. cloud vendor specific fencing agent
  - this may change in future

Integrated Pacemaker Install

Single command to install Db2 **and** Pacemaker

no change to existing syntax

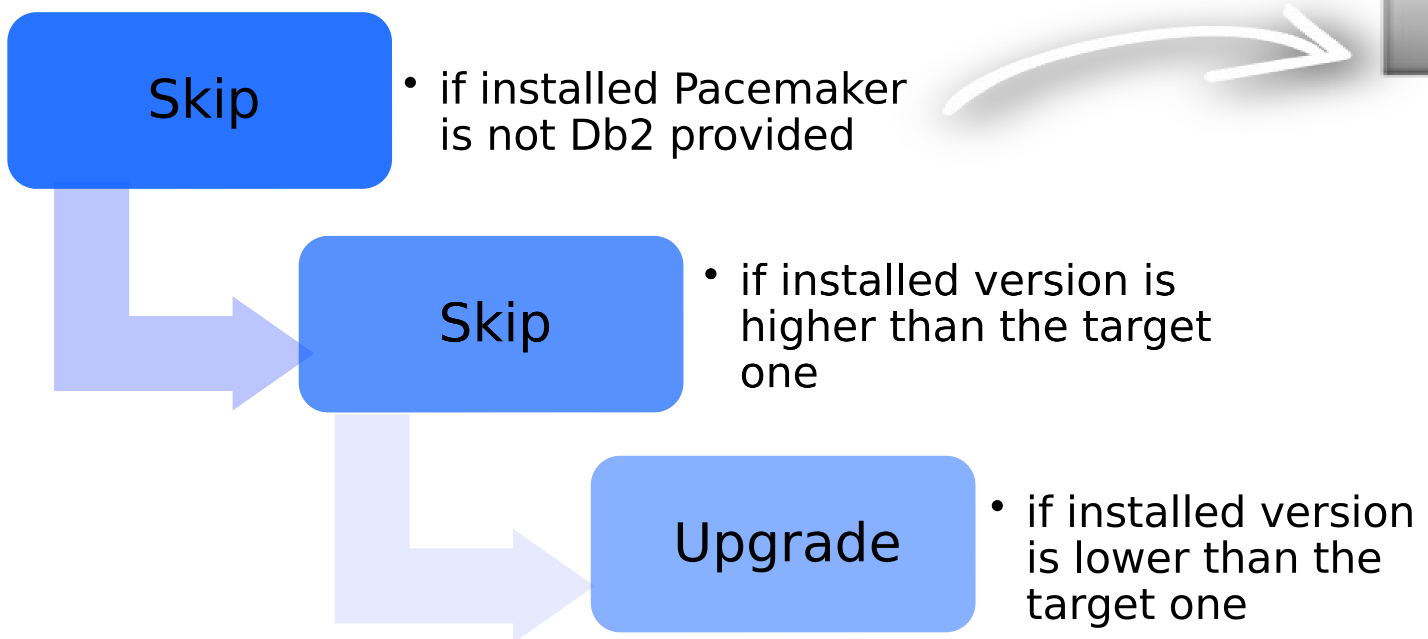
- New install: **db2_install** -y -b -/opt/ibm/db2/V11.5 -p SERVER
- Update: **installFixPack** -y -b /opt/ibm/db2/V11.5 -p /opt/ibm/db2/V11.5.6

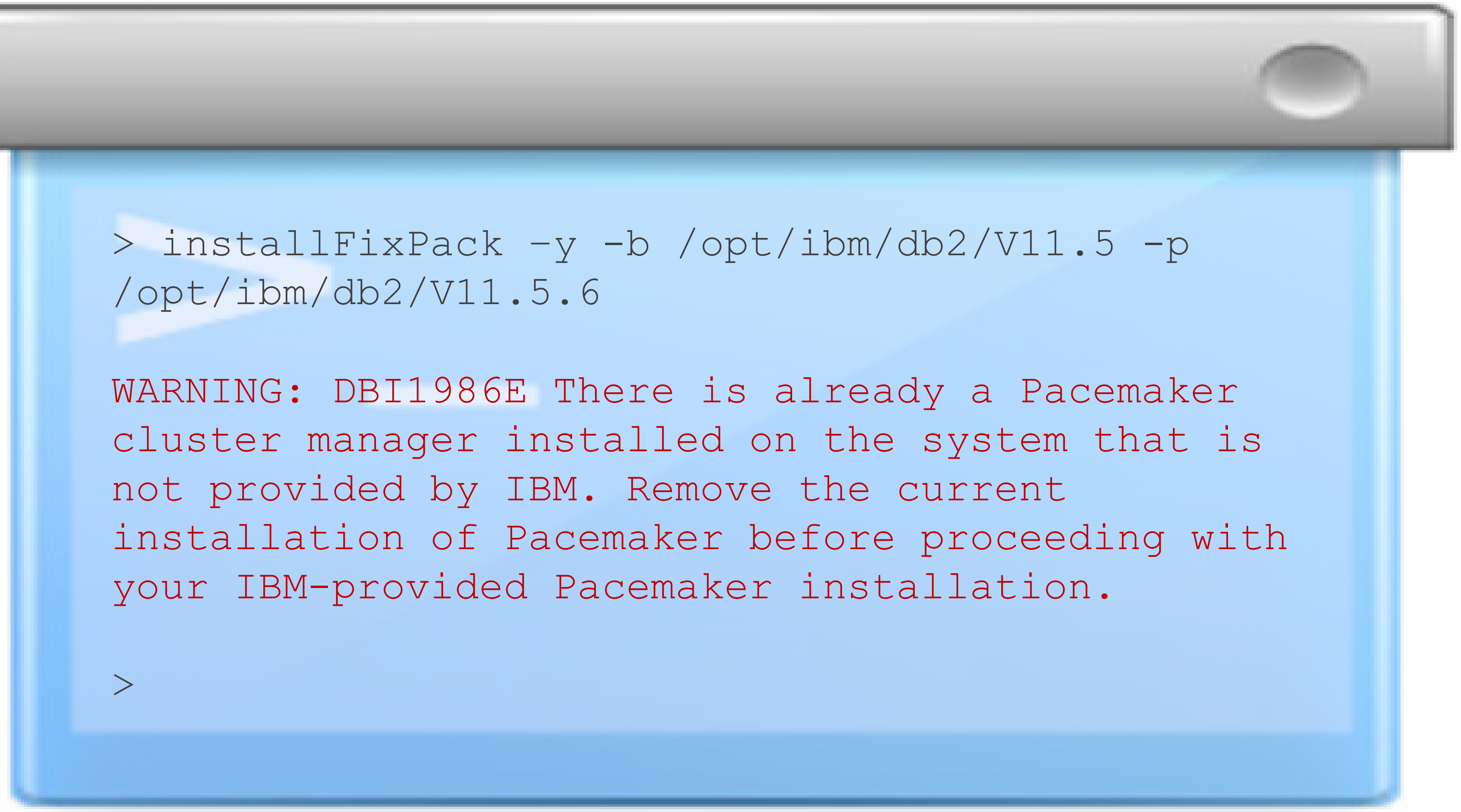
Skip and install Pacemaker later!

- Skip: db2_install -p server -b /opt/ibm/db2/V11.5 **-NOPCMK**
- Install later: Db2_install_image>/universal/db2/<platform>/pcmk/**db2installPCMK**

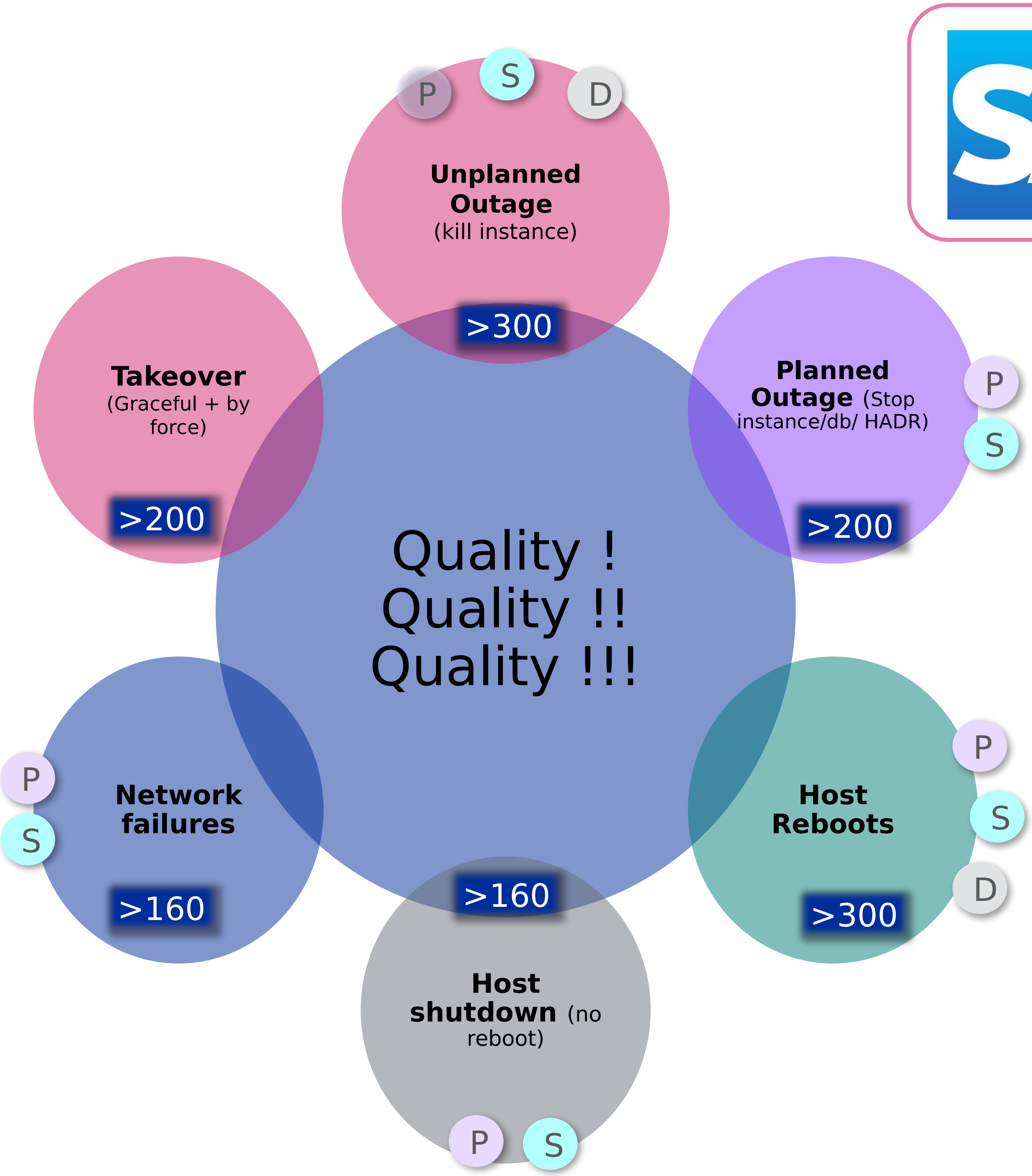
```
Task #33 start
Description: TSA
Estimated time 300 second(s)
Task #33 end

Task #34
start Description: Pacemaker
Estimated time 300 second(s)
Task #34 end
.
.
.
The execution completed Successfully
```

Handle Pacemaker Upgrade automatically:


Skip
if installed Pacemaker is not Db2 provided
Skip
if installed version is higher than the target one
Upgrade
if installed version is lower than the target one



```
> installFixPack -y -b /opt/ibm/db2/V11.5 -p
/opt/ibm/db2/V11.5.6

WARNING: DBI1986E There is already a Pacemaker
cluster manager installed on the system that is
not provided by IBM. Remove the current
installation of Pacemaker before proceeding with
your IBM-provided Pacemaker installation.

>
```



Quality !
Quality !!
Quality !!!

Unplanned Outage (kill instance) >300 — P S D

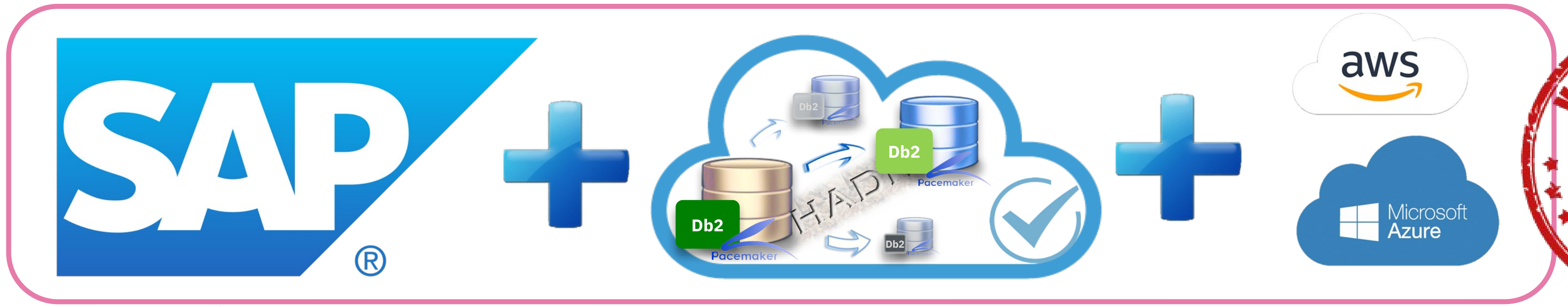
Planned Outage (Stop instance/db/ HADR) >200 — P S

Host Reboots >300 — P S D

Host shutdown (no reboot) >160 — P S

Network failures >160 — P S

Takeover (Graceful + by force) >200



VERIFIED

- >1200 executions of successful scenario
- Covers SLES and RHEL on x86
- Covers standard quorum (QDevice) as on-prem and cloud fencing agent
- Run with SAP Workload

Cloud Exploration: Motivation and Results

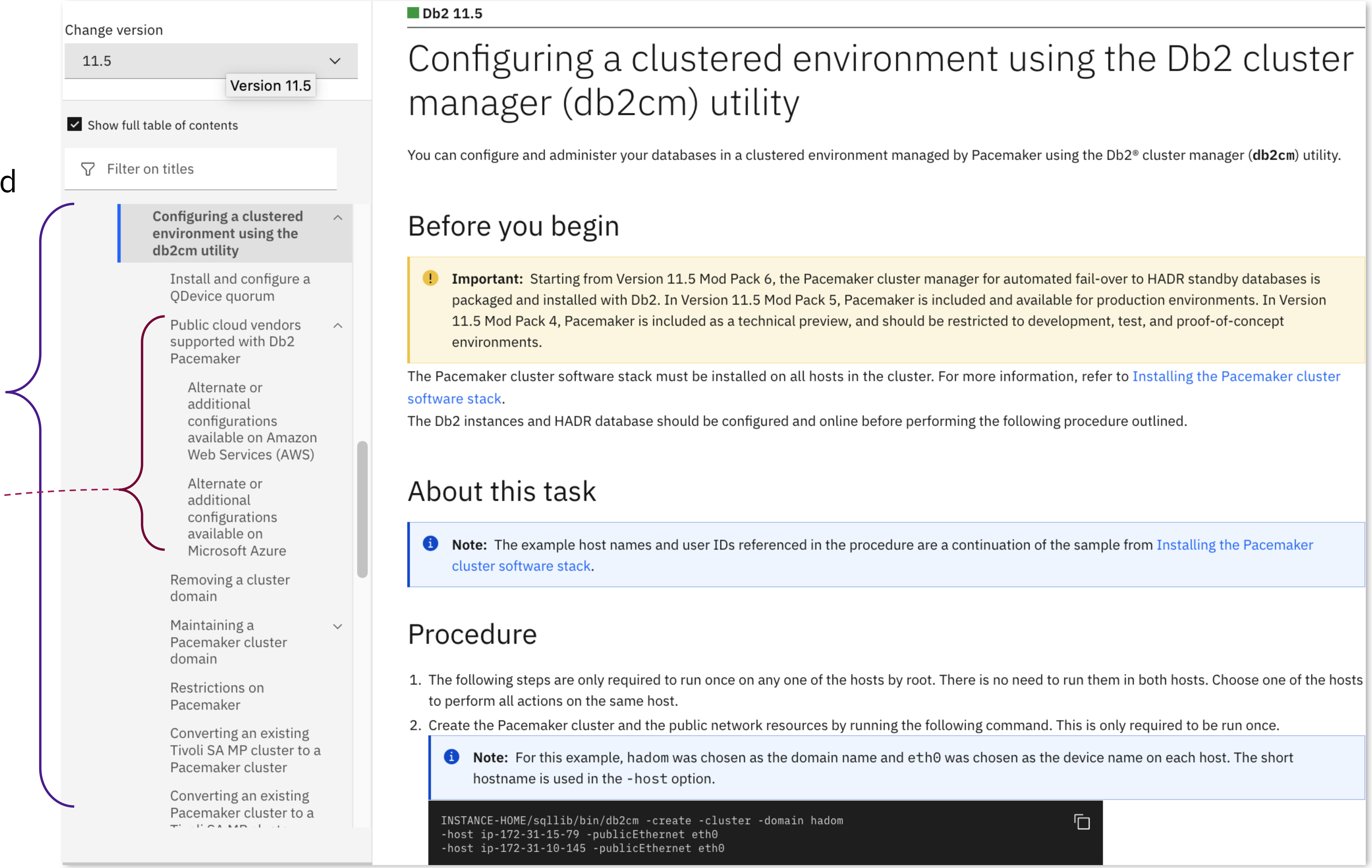
Solution goal:

- Ensure all Db2 LUW HA solutions can be deployed anywhere

Instructions applicable to deployments on all form factors (on-premises and cloud)

Cloud specific section

- Augment overall configuration to run optimally on cloud
- Focus on Quorum alternatives and Virtual IP setup


Change version
11.5
Version 11.5
☑ Show full table of contents
Filter on titles
Configuring a clustered environment using the db2cm utility
Install and configure a QDevice quorum
Public cloud vendors supported with Db2 Pacemaker
Alternate or additional configurations available on Amazon Web Services (AWS)
Alternate or additional configurations available on Microsoft Azure
Removing a cluster domain
Maintaining a Pacemaker cluster domain
Restrictions on Pacemaker
Converting an existing Tivoli SA MP cluster to a Pacemaker cluster
Converting an existing Pacemaker cluster to a

🟩 Db2 11.5

Configuring a clustered environment using the Db2 cluster manager (db2cm) utility

You can configure and administer your databases in a clustered environment managed by Pacemaker using the Db2® cluster manager (**db2cm**) utility.

Before you begin

Important: Starting from Version 11.5 Mod Pack 6, the Pacemaker cluster manager for automated fail-over to HADR standby databases is packaged and installed with Db2. In Version 11.5 Mod Pack 5, Pacemaker is included and available for production environments. In Version 11.5 Mod Pack 4, Pacemaker is included as a technical preview, and should be restricted to development, test, and proof-of-concept environments.

The Pacemaker cluster software stack must be installed on all hosts in the cluster. For more information, refer to Installing the Pacemaker cluster software stack.

The Db2 instances and HADR database should be configured and online before performing the following procedure outlined.

About this task

Note: The example host names and user IDs referenced in the procedure are a continuation of the sample from Installing the Pacemaker cluster software stack.

Procedure

1. The following steps are only required to run once on any one of the hosts by root. There is no need to run them in both hosts. Choose one of the hosts to perform all actions on the same host.
2. Create the Pacemaker cluster and the public network resources by running the following command. This is only required to be run once.

   Note: For this example, hadom was chosen as the domain name and eth0 was chosen as the device name on each host. The short hostname is used in the -host option.

```
INSTANCE-HOME/sqllib/bin/db2cm -create -cluster -domain hadom
-host ip-172-31-15-79 -publicEthernet eth0
-host ip-172-31-10-145 -publicEthernet eth0
```

Link to Db2 doc

Azure Exploration #1: Alternate quorum mechanism on Azure via Fencing

End-to-end setup overview

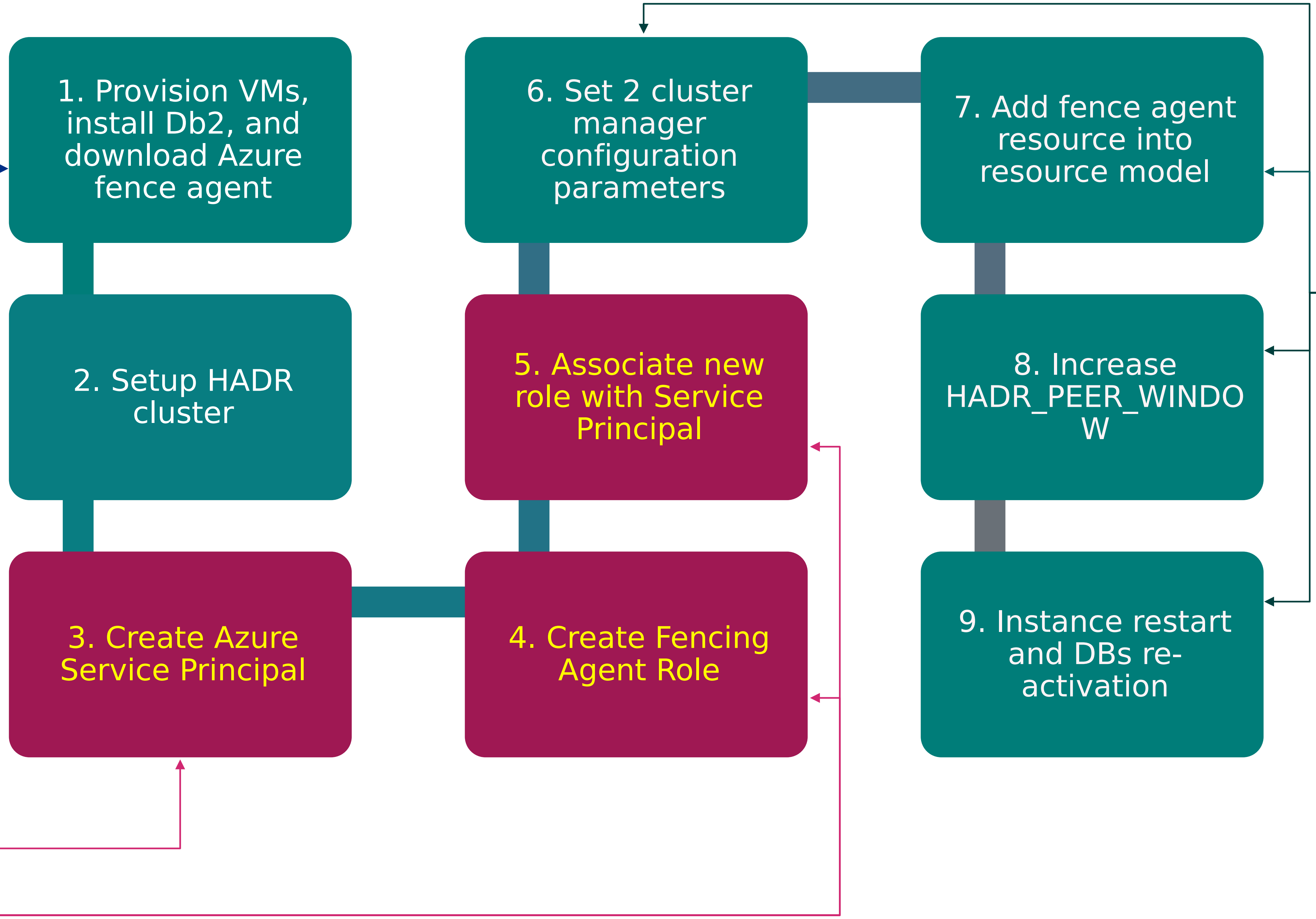
Fence agent available in MRS

- Db2_Azure_fence_agents_4.7.1-3_noarch.tar.gz

Azure VM & Account Configurations

- Create ID (use as username for service principal)
- Create a custom role to execute fence agent on your Azure VMs.
- Assign the role to the service principle created (role) for both VMs

This effectively allows Azure to execute the fence agent installed to power on/off your VMs (as part of fencing)



1. Provision VMs, install Db2, and download Azure fence agent

2. Setup HADR cluster

3. Create Azure Service Principal

4. Create Fencing Agent Role

5. Associate new role with Service Principal

6. Set 2 cluster manager configuration parameters

7. Add fence agent resource into resource model

8. Increase HADR_PEER_WINDOW

9. Instance restart and DBs re-activation

Pacemaker config changes

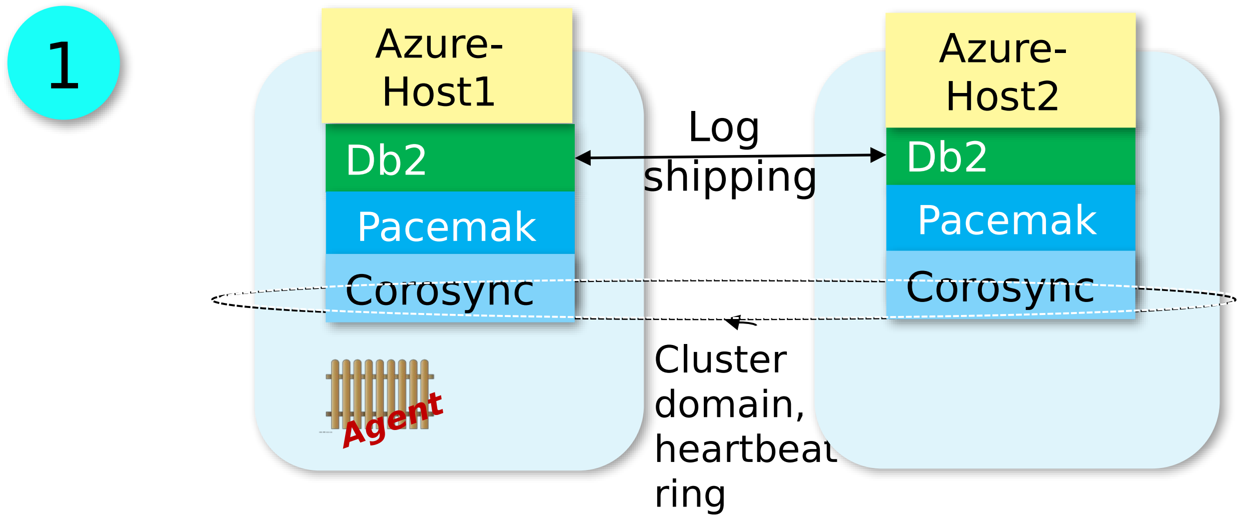
- set *wait_for_all* to 0 – allow 1 host to be online without majority in a 2-hosts setup
- set heartbeat loss toleration to 30 seconds (due to Azure non-reboot maintenance limitation)

Db2 DB config param:

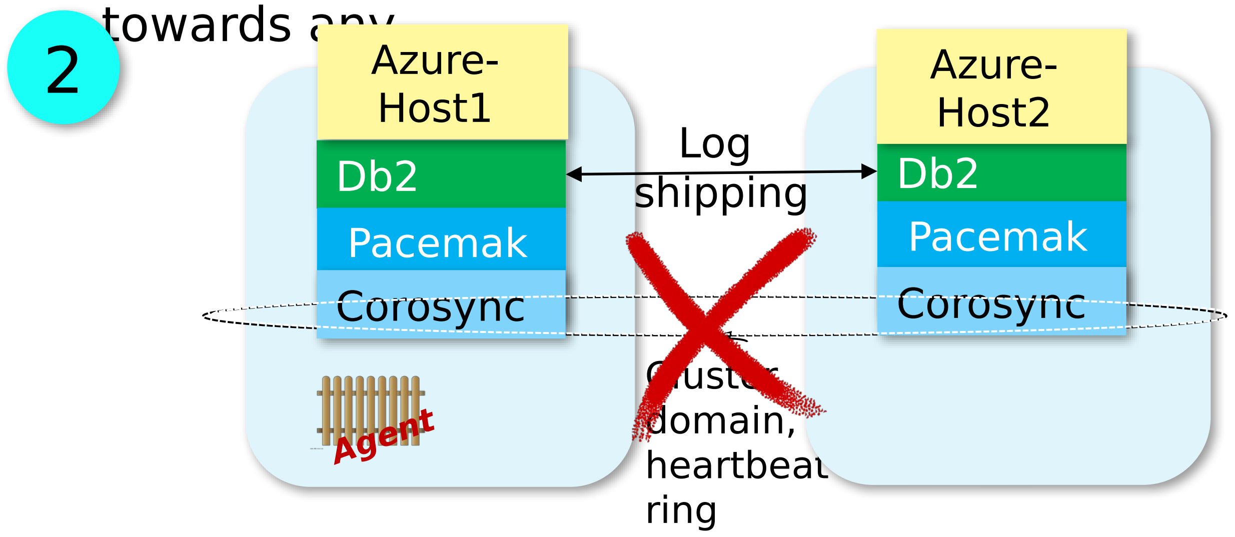
- Set HADR_PEER_WINDOW to >=300 seconds due to longer fencing time required

Full instructions : link

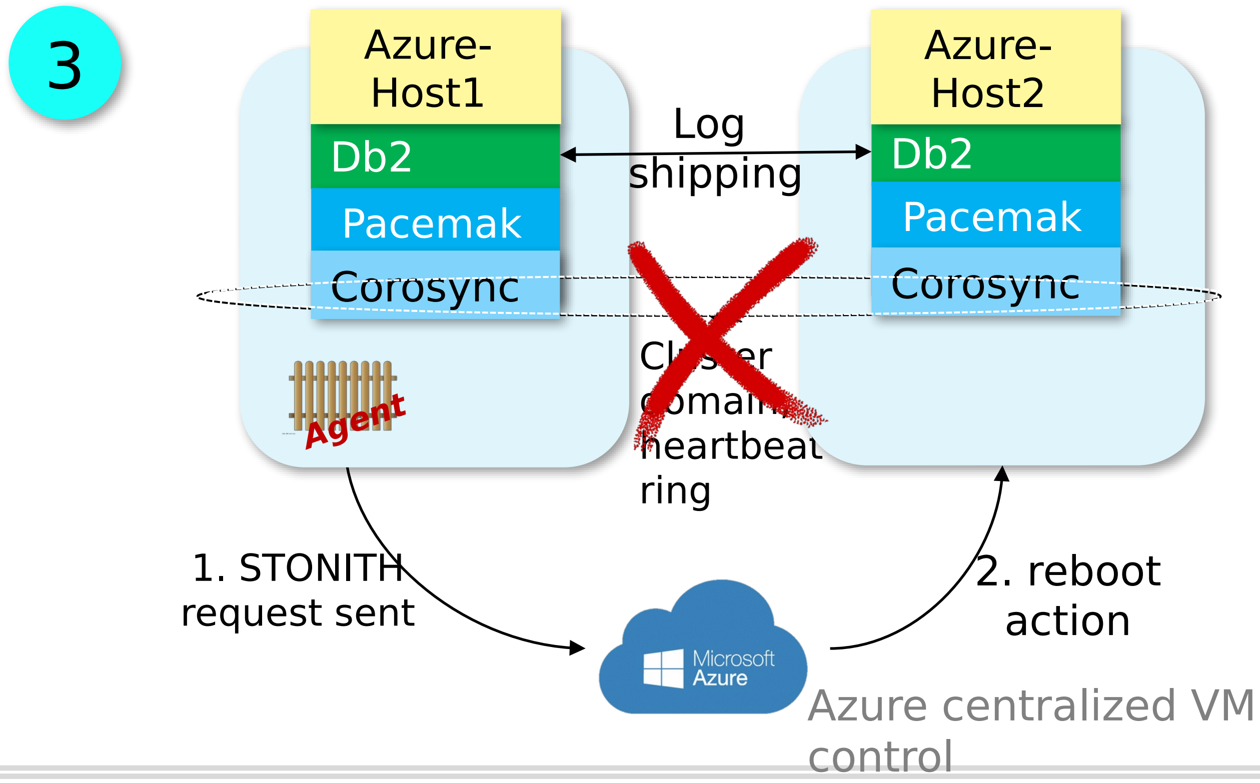
Azure Exploration #1: Fencing Internal Workings

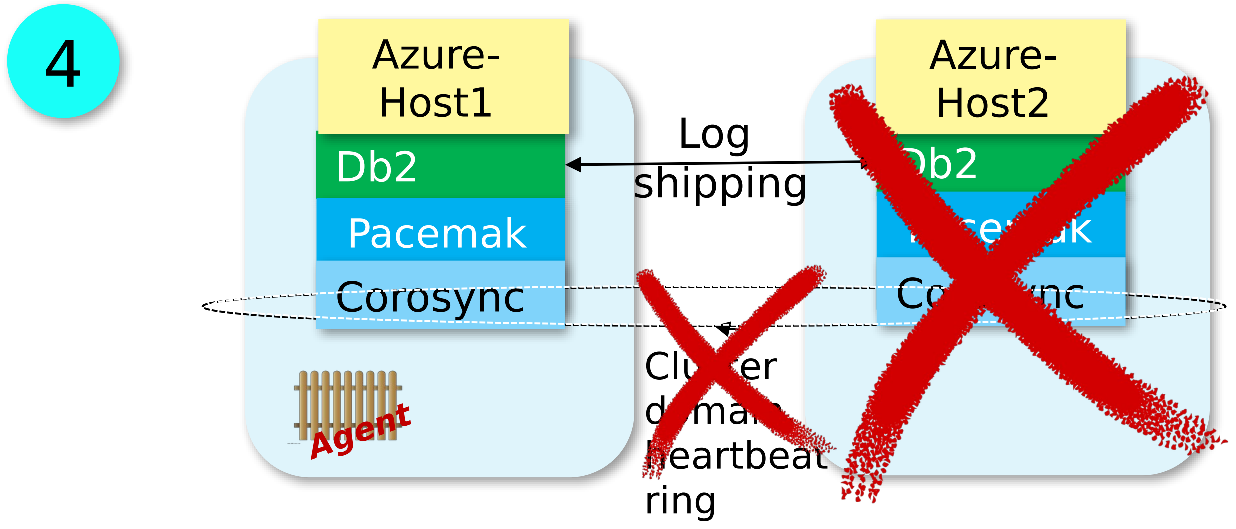


- Fencing agent setup and activated in resource model.
- symmetric cluster setting allow this resource to be online on any one host (not BOTH) without any bias towards any.



- All heartbeat rings are broken – two hosts lost communication





- The fencing action as a result of the reboot action may take some time to complete due to Azure infrastructure
- In-house lab testing show it can be up to 6x the time needed with 3rd VM with Qdevice quorum

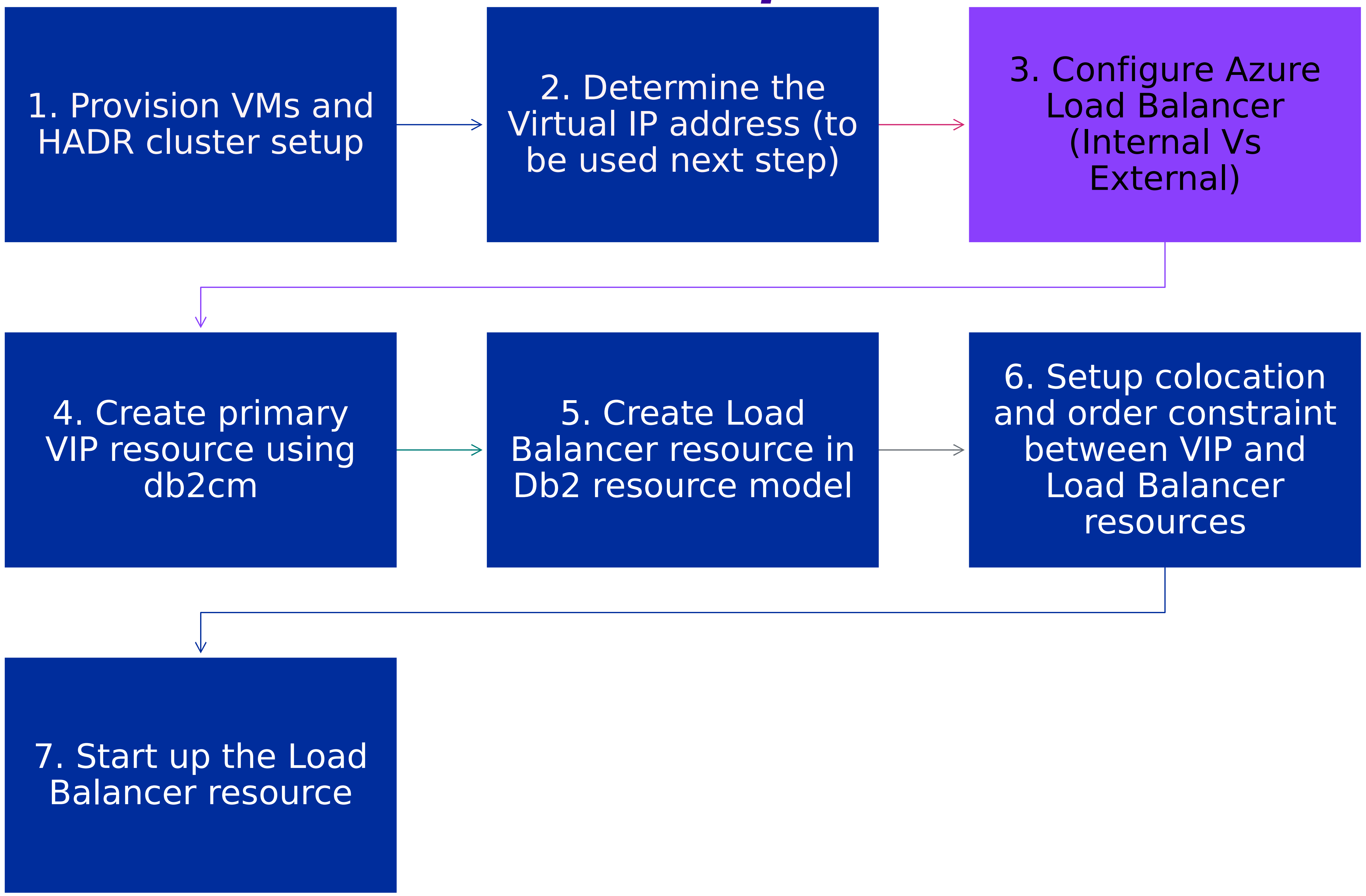
Qdevice Vs Fencing ?

RULES

Based your decision on the need for faster recovery from primary host failure Vs on-going cost of maintaining a small 3rd VM.

Azure Exploration #2: Virtual IP setup with Azure Load Balancer

End-to-end setup overview



1. Provision VMs and HADR cluster setup

2. Determine the Virtual IP address (to be used next step)

3. Configure Azure Load Balancer (Internal Vs External)

4. Create primary VIP resource using db2cm

5. Create Load Balancer resource in Db2 resource model

6. Setup colocation and order constraint between VIP and Load Balancer resources

7. Start up the Load Balancer resource

Configuration in your Azure account

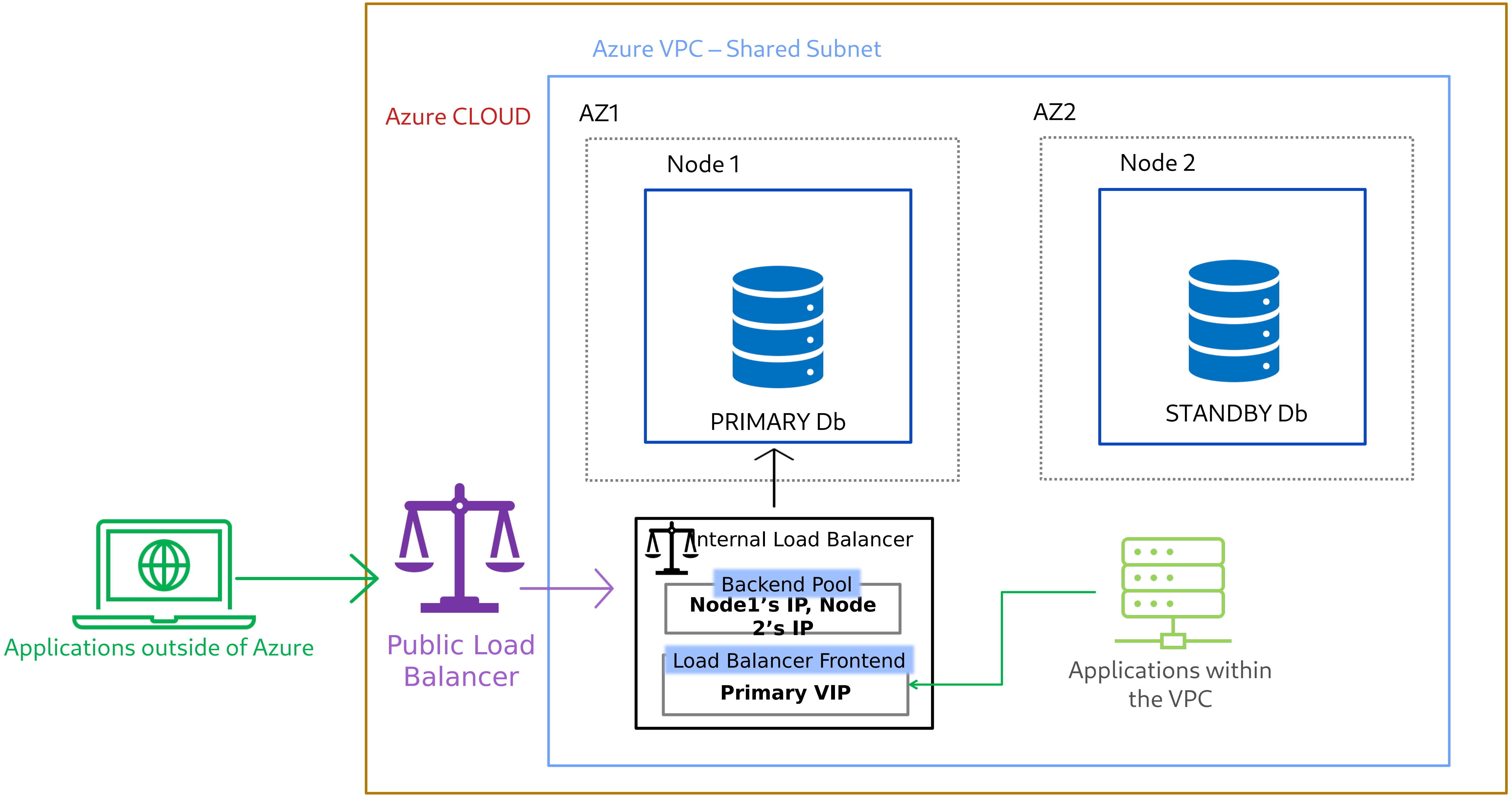
- Internal Load Balancer for app traffic from within same VPC
- External Load Balancer for app traffic outside of VPC

Db2 resource model changes

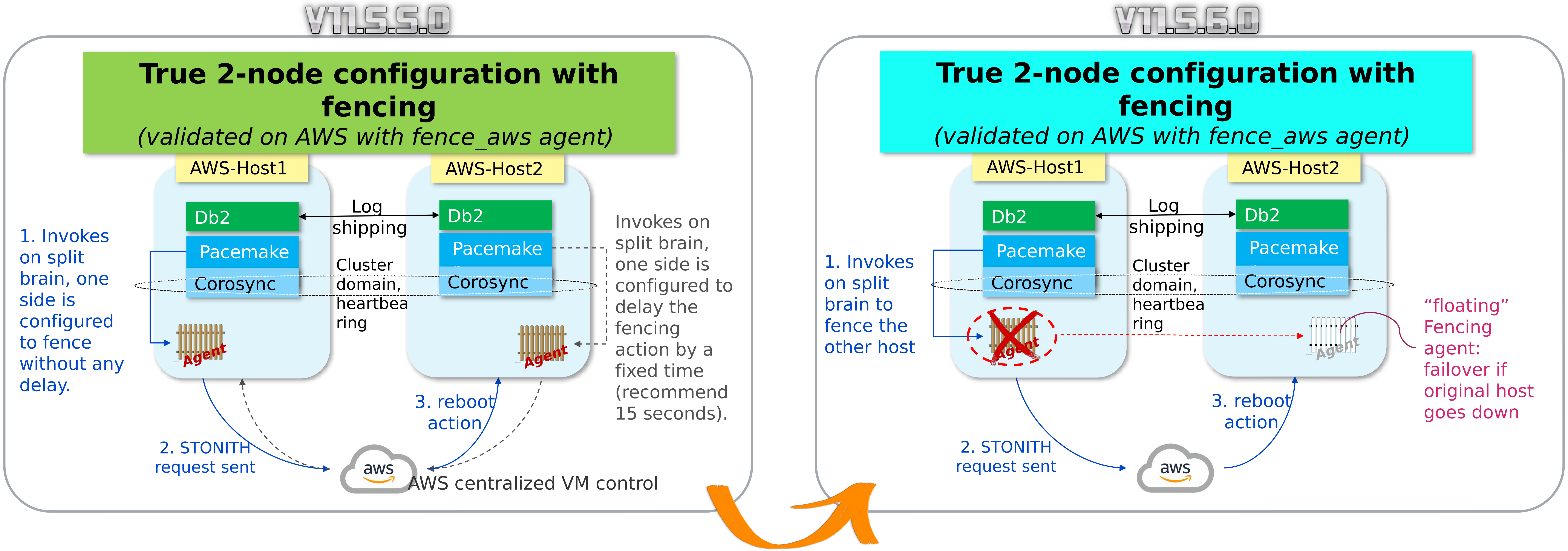
- incorporate the DB specific Load Balancer into the resource so that it floats with the corresponding VIP
- set heartbeat loss toleration to 30 seconds (due to Azure non-reboot maintenance limitation)

Full instructions : link

Azure Exploration #2: Azure with Load Balancer Topology



AWS Fencing Setup Optimization: From 2 fencing agents to 1


V11.5.5.0
True 2-node configuration with fencing
(validated on AWS with fence_aws agent)
AWS-Host1
AWS-Host2
Db2
Pacemake
Corosync
Log shipping
Cluster domain, heartbea ring
Agent
Agent
3. reboot action
2. STONITH request sent
aws
AWS centralized VM control
V11.5.6.0
True 2-node configuration with fencing
(validated on AWS with fence_aws agent)
AWS-Host1
AWS-Host2
Db2
Pacemake
Corosync
Log shipping
Cluster domain, heartbea ring
Agent
Agent
3. reboot action
2. STONITH request sent
aws

1. Invokes on split brain, one side is configured to fence without any delay.

Invokes on split brain, one side is configured to delay the fencing action by a fixed time (recommend 15 seconds).

1. Invokes on split brain to fence the other host

“floating” Fencing agent: failover if original host goes down

Instead of setting up the fencing agents as 2 separate independent resources, setup only 1 and allow the resource to failover to the other host naturally on host failure.

Enhanced Problem Determination

- Added millisecond resolution in the Pacemaker log. (similar to /var/log/message)
- Imperative to reconstruct timeline of events in any scenario

Example: /var/log/pacemaker/pacemaker.log

Before

```
Jan 13 10:50:02 talkers1.fyre.ibm.com pacemakerd     [3829732] (qb_ipcs_us_withdraw)    info: withdrawing server sockets
Jan 13 10:50:02 talkers1.fyre.ibm.com pacemakerd     [3829732] (crm_xml_cleanup)  info: Cleaning up memory from libxml2
Jan 13 10:50:02 talkers1.fyre.ibm.com pacemakerd     [3829732] (crm_exit)   info: Exiting pacemakerd | with status 0
```

After

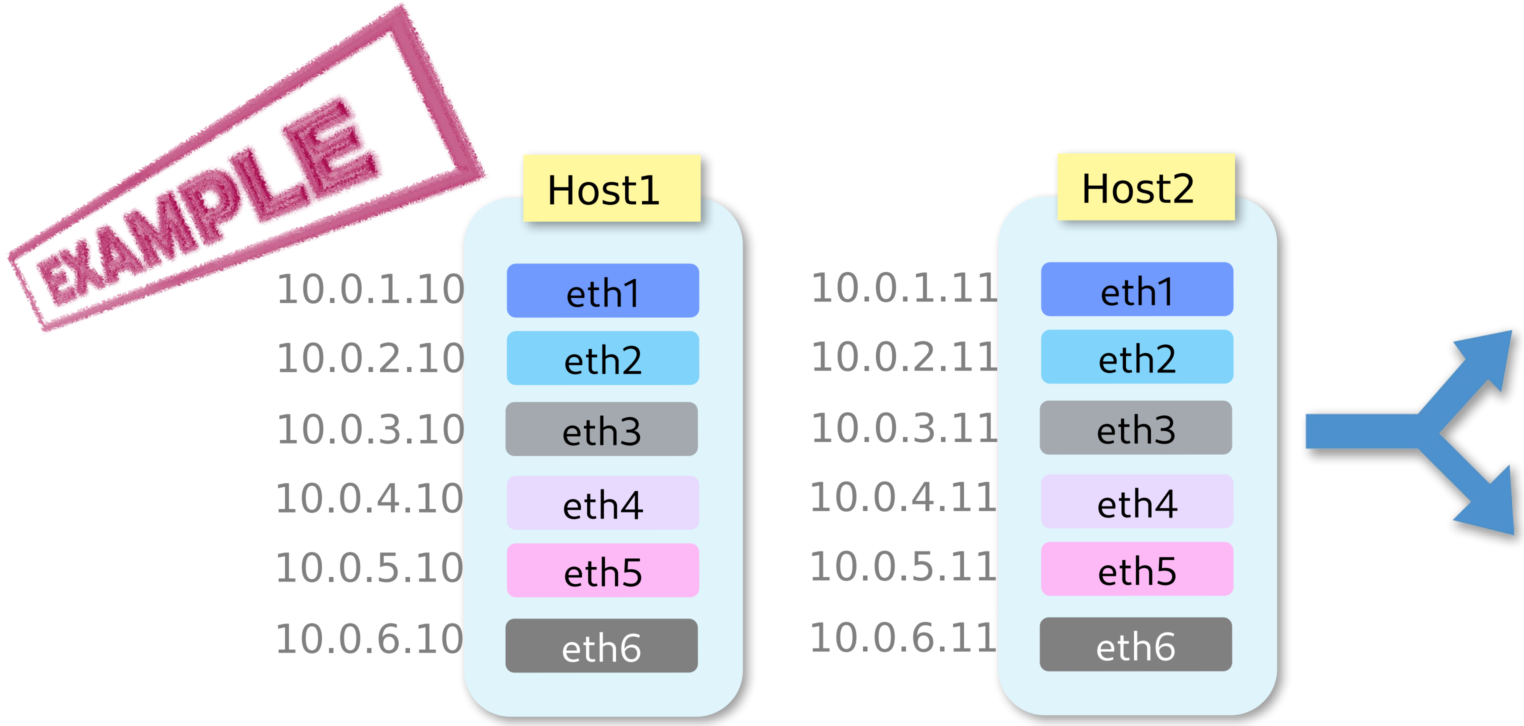
```
Jan 13 11:04:46.555 talkers1.fyre.ibm.com pacemakerd     [1889531] (crm_log_init)    info: Changed active directory to /var/….
Jan 13 11:04:46.564 talkers1.fyre.ibm.com pacemakerd     [1889531] (get_cluster_type)   info: Detected an active 'corosync' cluster
Jan 13 11:04:46.564 talkers1.fyre.ibm.com pacemakerd     [1889531] (mcp_read_config)    info: Reading configure for stack: corosync
```

IBM
DB2
contributes
Pacemaker
open source

- First contribution made from Db2 development to the open-source Pacemaker community

“Db2-aware” Network Resiliency

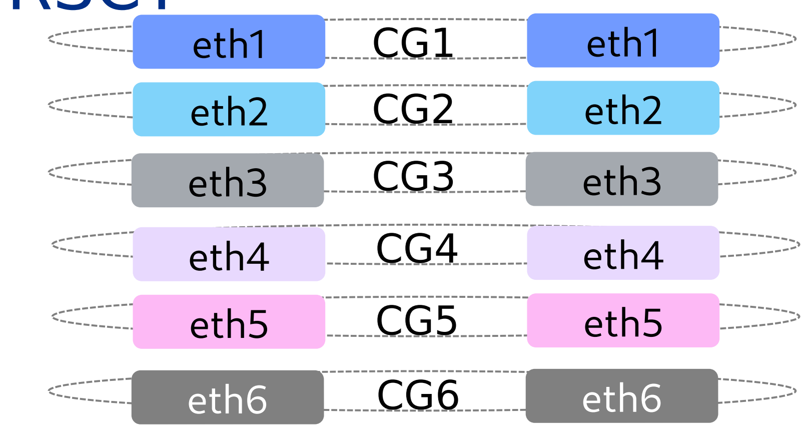
- Cluster Membership – who’s in and who’s out – relies on “Node Liveliness Test"
  - RSCT: Communication Group (a.k.a. CG)
  - Corosync: Heartbeat Ring (a.k.a. HBR)



Description

- Each adapter within a host has unique IP subnet 10.0.x.0
- The two hosts have the same set of IP subnets (6 in total).
- Assumption: each IP in the same subnet can ping each other.

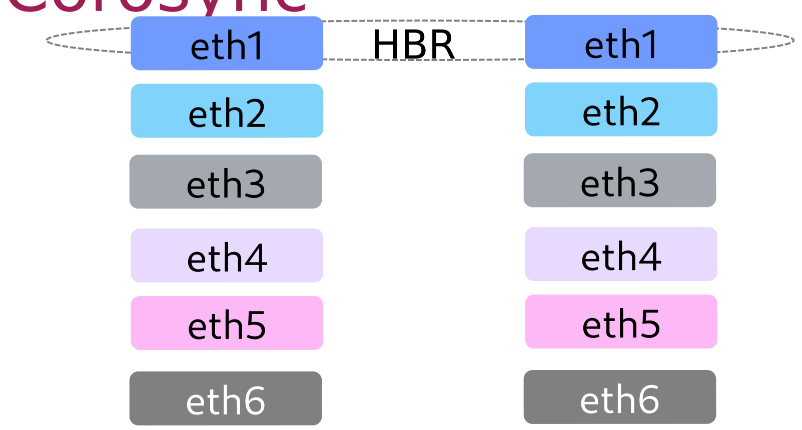
RSCT



RSCT Node liveliness Test

- 6 CGs created due to 6 different unique IP subnets
- A host is only deemed “dead” if **ALL** 6 CGs are broken at the same time and last longer than the preset grace period.

Corosync



Corosync Node liveliness Test

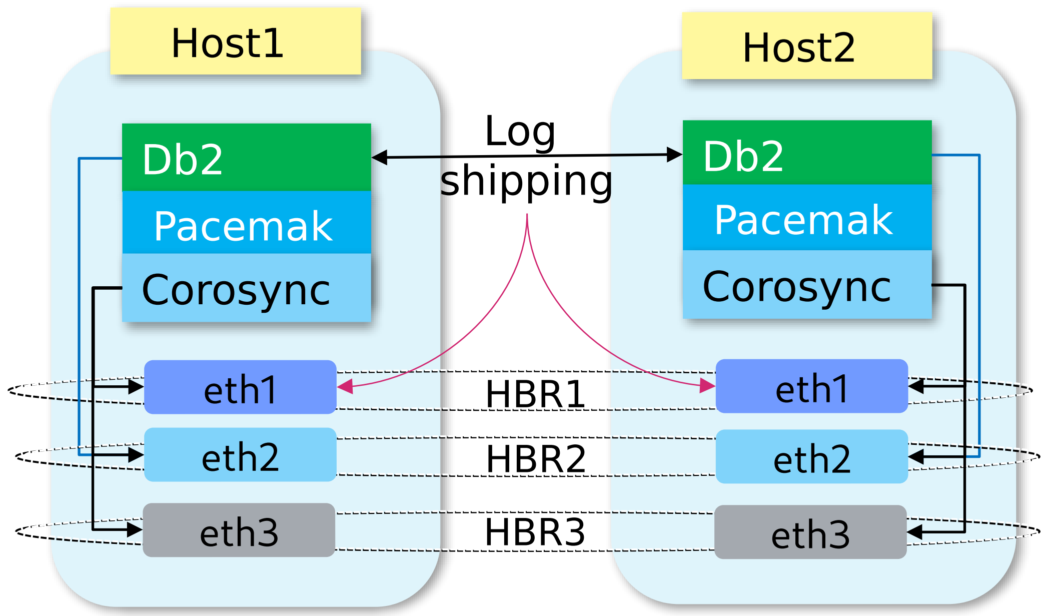
- By default, only hostname’s IP is included
- A host is deemed “dead” if eth1 lost the heartbeat regardless of the state of the other 5 NICs.

Observation: *Neither default logic is ideal for Db2 …*

“Db2-aware” Network Resiliency (cont’d)

- **Problem:** One is overkill, the other is too simplistic

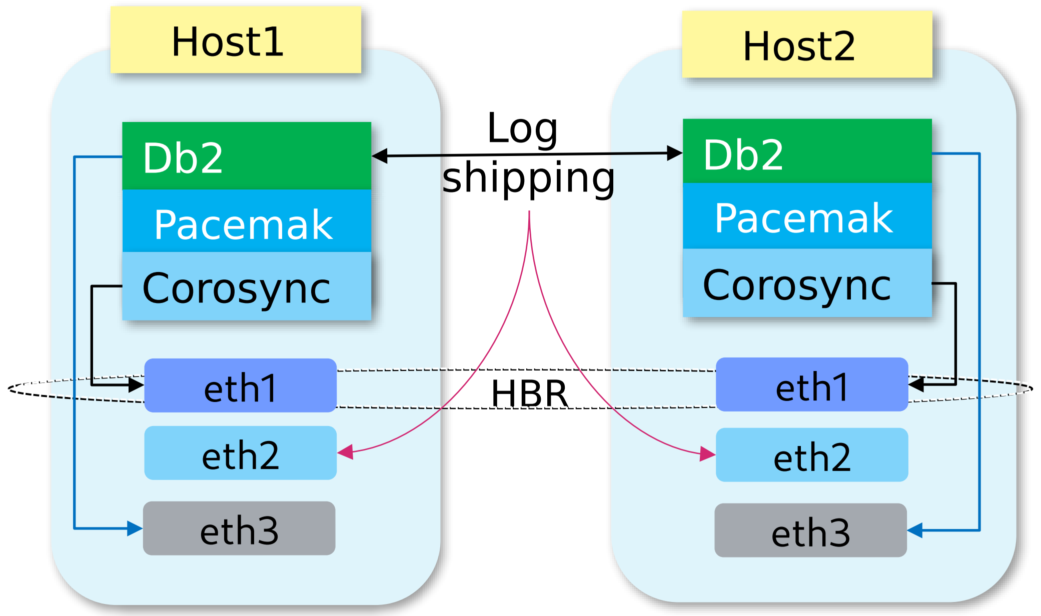
Making Corosync behave in RSCT way



The “Overkill” way:

- HBRs setup on all TCP/IP capable adapters
- HBR3 is not used by Db2 but can delay failure detection and recovery if only HBR1 and HBR2 are down.
- Not tailored to Db2 use case

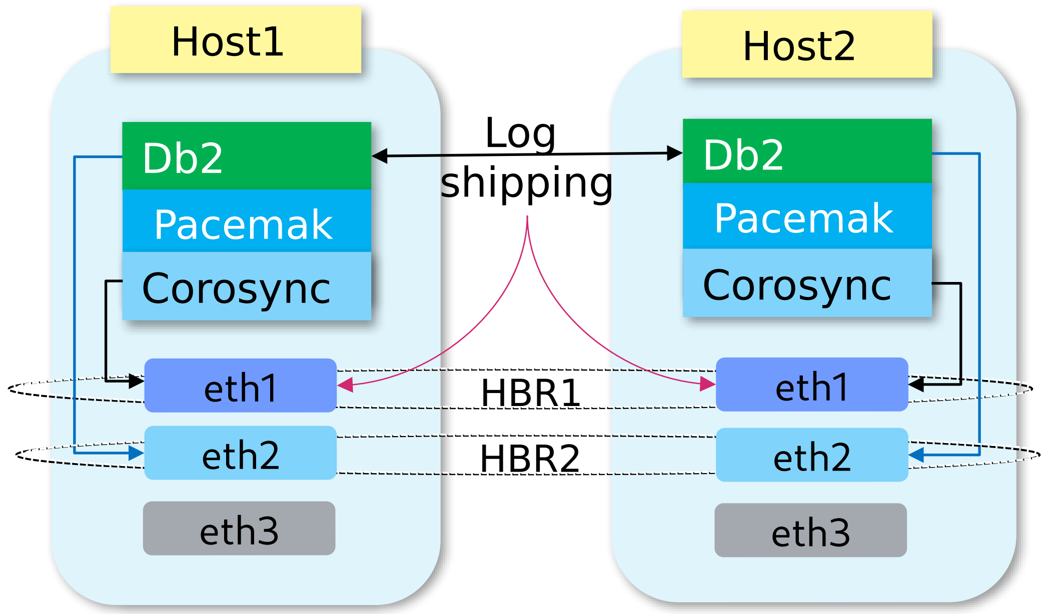
Corosync default behaviour



The simplistic way:

- Single HBR may lead to false positive depending on which IP is used.
- Worst case is when HBR picks a network not used by Db2. That network failed but all Db2’s NICs are functional.

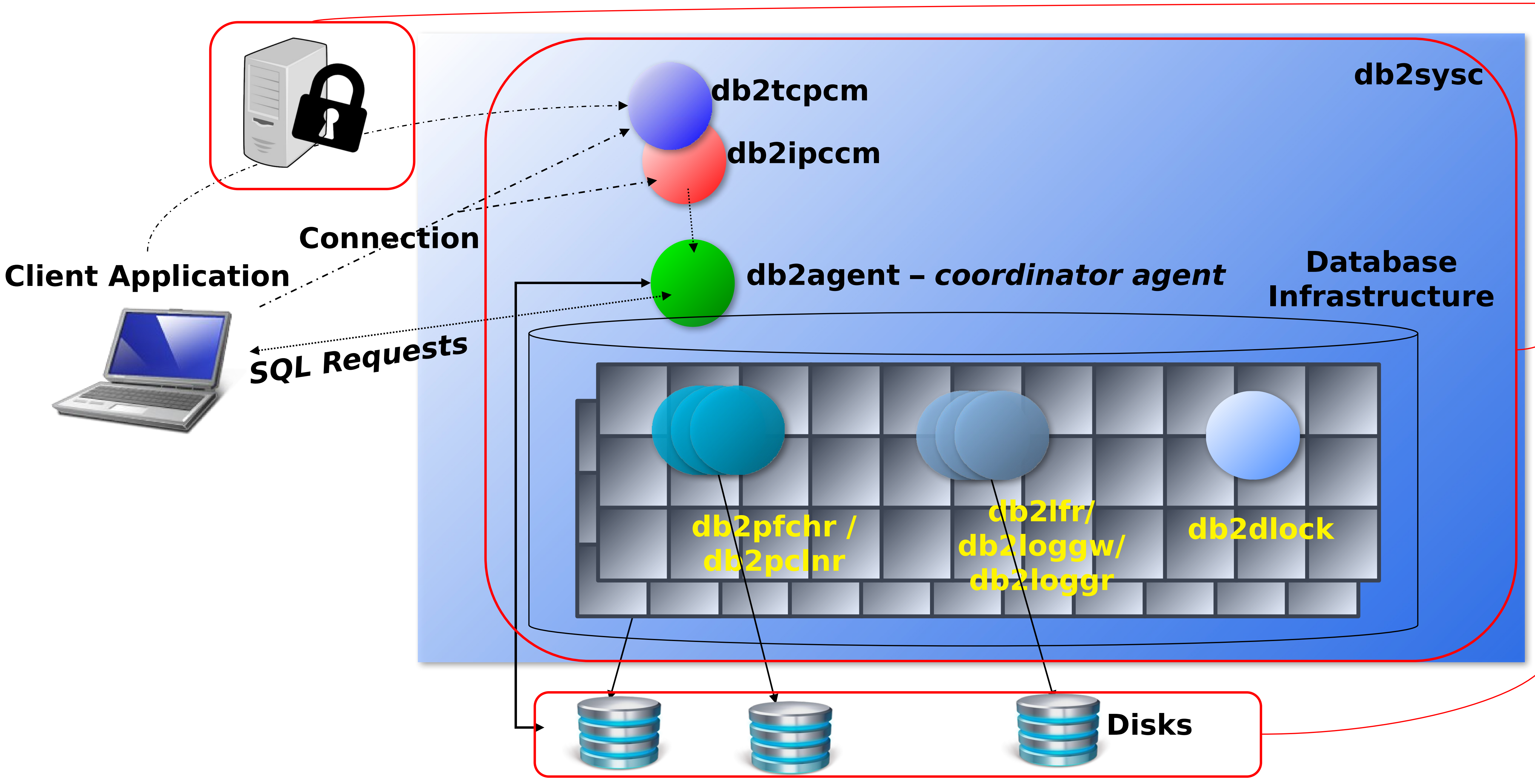
A better approach … config HBR to only include Db2’s relevant NICs



Note

- Only setup HBR on NICs used by Db2.
- In 11.5.6.0, HBR1 is setup by default
- Instructions available to setup additional HBRs with other NICs for each log shipping network
- Future: automatic discovery of Db2 relevant NICs and creation of HBR(s)

Advance HADR DB hang detection on Linux – *Potential of hang*



1. At external authentication server

2. At Db2 level caused by:

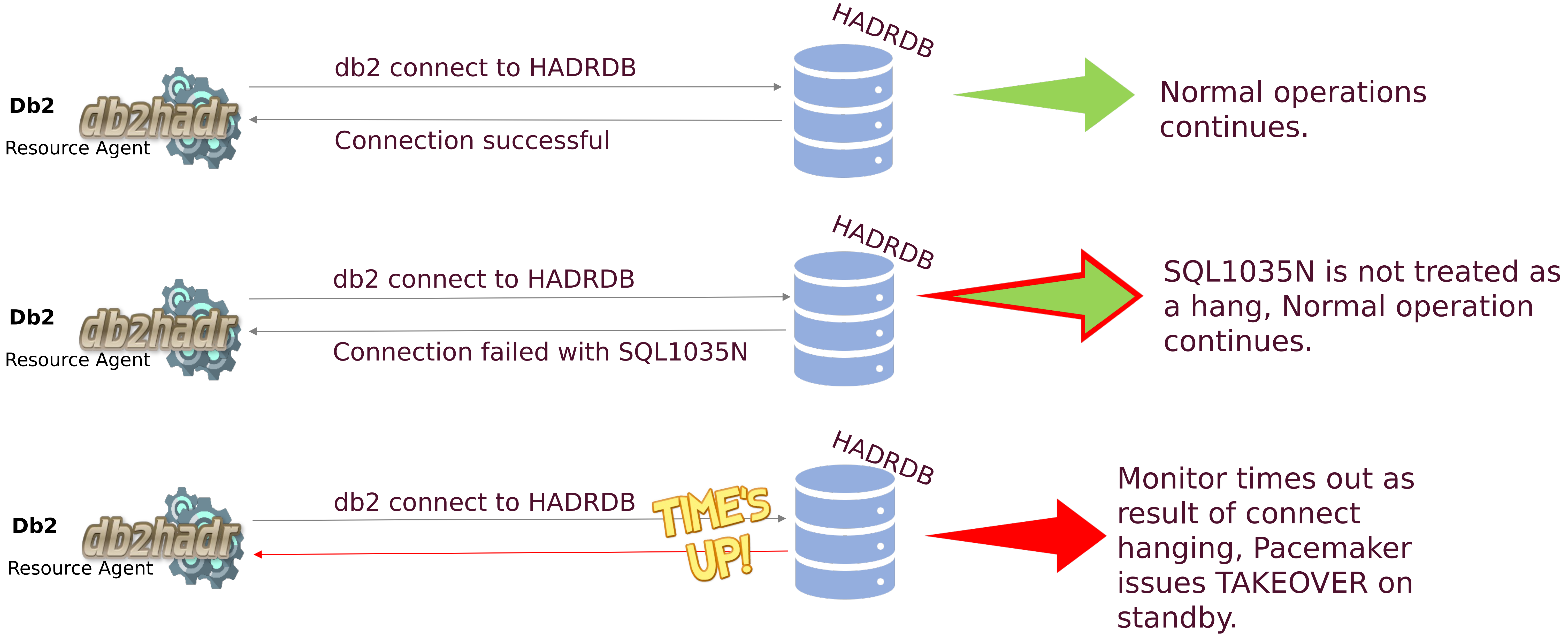
- deadlock within Db2 engine threads
- Db2 operation stuck at kernel calls
- overloaded system / lack of CPU resources

3. At storage I/O layer or network

Approach: Focus on the database connect

Advance HADR DB hang detection on Linux (cont’d)

Database monitoring via the db2hadr resource agent is now capable of detecting hangs while connecting to the primary database.



Db2 Resource Agent

db2 connect to HADRDB

Connection successful

HADRDB

Normal operations continues.

Db2 Resource Agent

db2 connect to HADRDB

Connection failed with SQL1035N

HADRDB

SQL1035N is not treated as a hang, Normal operation continues.

Db2 Resource Agent

db2 connect to HADRDB

HADRDB

Monitor times out as result of connect hanging, Pacemaker issues TAKEOVER on standby.

Advance HADR DB hang detection on Linux (cont’d)

Enablement

- Off by default, enabled via environment variable. Effective immediately, no instance restart required.
- Add the following to instance user’s $HOME/.profile

```
export DB2_HADR_HANG_DETECTION=CONNECT
```

Users can specify additional SQL codes to be ignored by the monitor

- `export DB2_HADR_HANG_SQL_BYPASS=SQL1040N,SQL1035N,SQL1060N`
- Ignored codes will not result in the monitor returning a failed state (i.e. no TAKEOVER issued)
- Current list of SQL codes ignored by default:

Maximum Applications

Maximum Connections

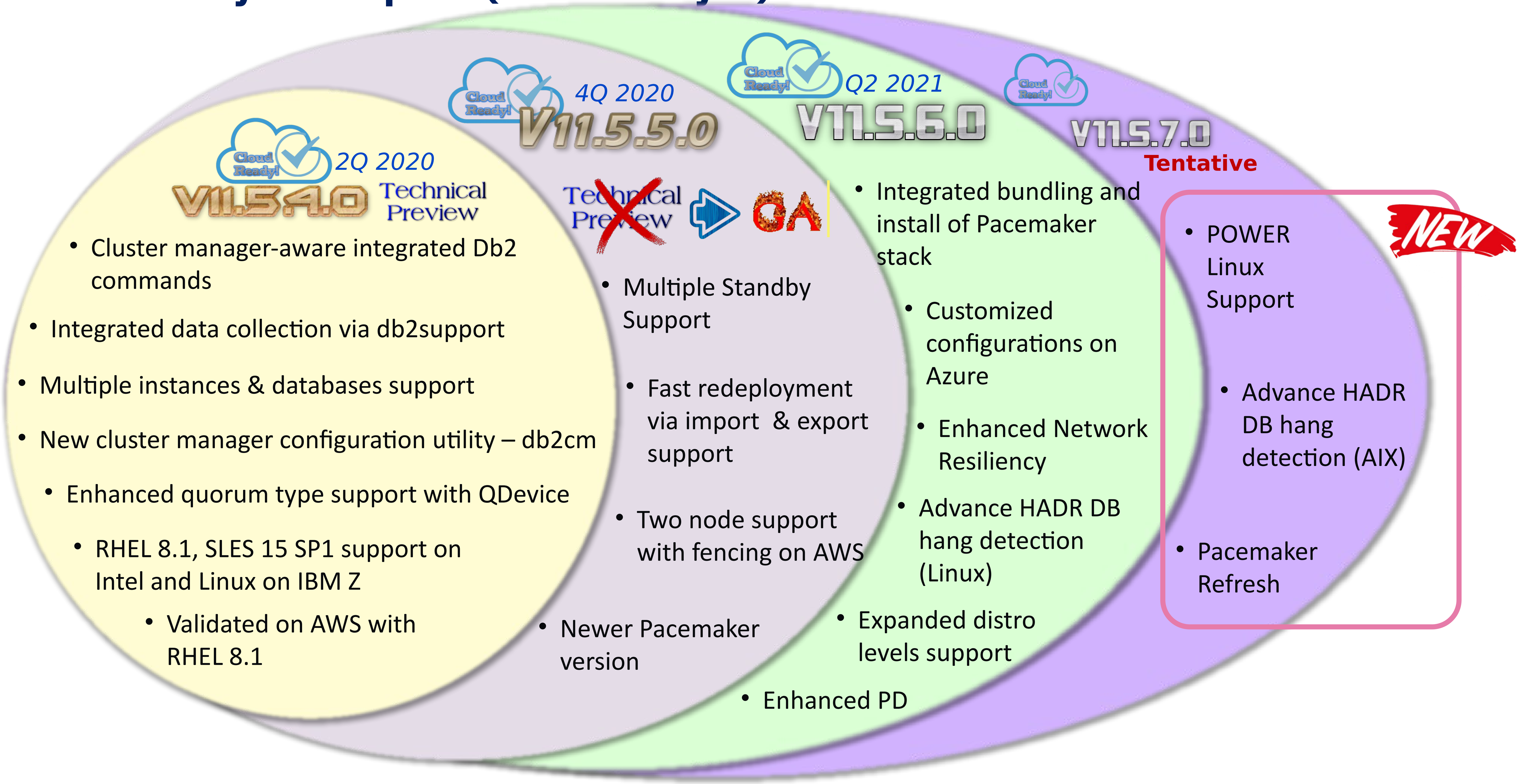
SQL1040N, SQL1226N, SQL1245N, SQL1035N, SQL1060N, SQL20157N…

Supported Platforms Update

Additional platforms supported on-premise and cloud.

- RHEL 8.2 on Intel x86_64
- RHEL 8.2 on IBM Z s390x
- SLES 15 SP2 on Intel x86_64
- SLES 15 SP2 on IBM Z s390x

Our Journey … Stop #4 (not there yet)



Cloud Ready! 2Q 2020

V11.5.4.0 Technical Preview

- Cluster manager-aware integrated Db2 commands
- Integrated data collection via db2support
- Multiple instances & databases support
- New cluster manager configuration utility – db2cm
- Enhanced quorum type support with QDevice
- RHEL 8.1, SLES 15 SP1 support on Intel and Linux on IBM Z
- Validated on AWS with RHEL 8.1

Cloud Ready! 4Q 2020

V11.5.5.0

~~Technical Preview~~ GA

- Multiple Standby Support
- Fast redeployment via import & export support
- Two node support with fencing on AWS
- Newer Pacemaker version

Cloud Ready! Q2 2021

V11.5.6.0

- Integrated bundling and install of Pacemaker stack
- Customized configurations on Azure
- Enhanced Network Resiliency
- Advance HADR DB hang detection (Linux)
- Expanded distro levels support
- Enhanced PD

Cloud Ready!

V11.5.7.0

Tentative

NEW

- POWER Linux Support
- Advance HADR DB hang detection (AIX)
- Pacemaker Refresh

Note: Roadmap and content subjected to change

Expanded Platform, OS levels Coverage, and Change of Support Statement

1 Power™ Linux

Starting from RHEL 8.4 & SLES 15 SP3

2 Change OS level support from <u>specific</u> release to:

- RHEL 8.x and up
- SLES 15 SPy and up

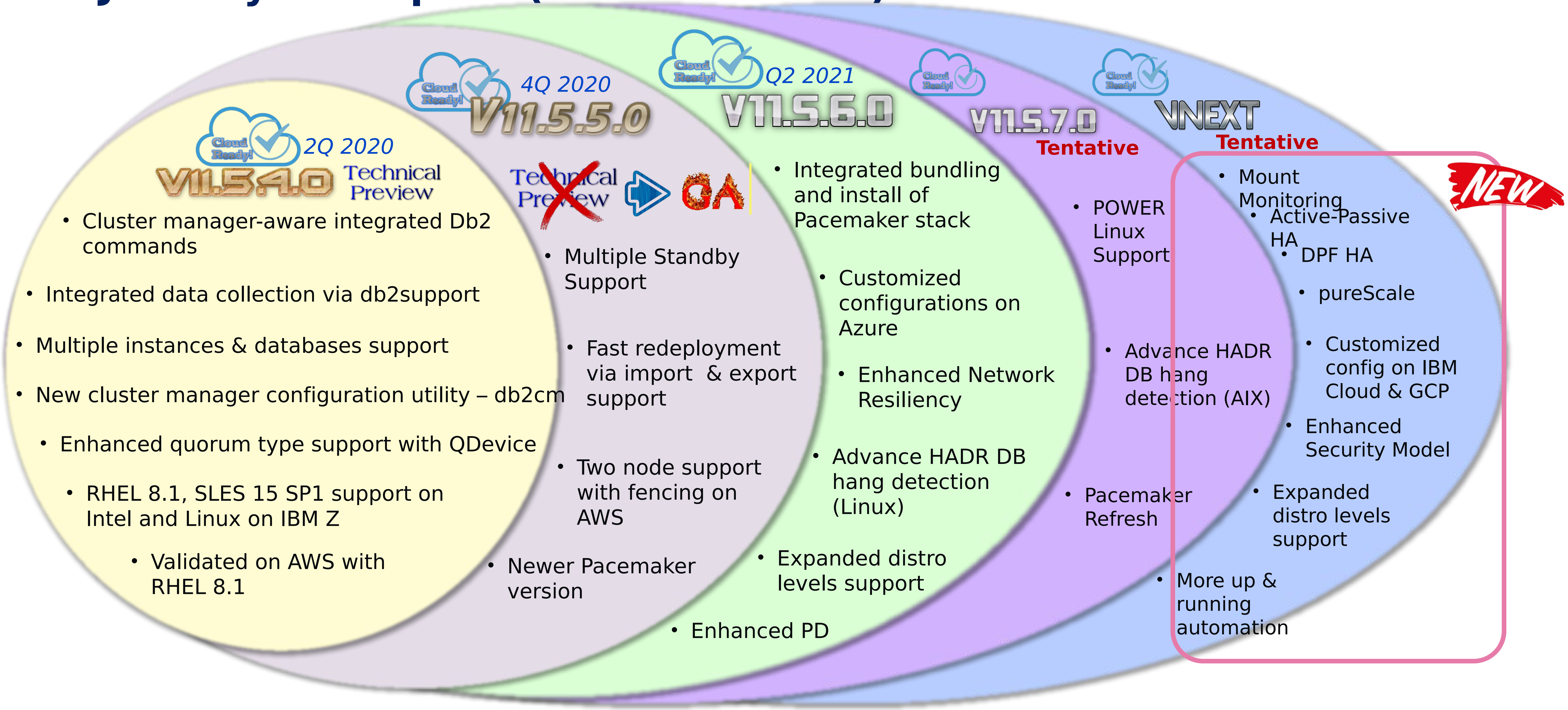
1 Power™ AIX

2

Frequent Pacemaker stack refresh

- at least once per year

Our Journey … Stop #5 (2022 tentative)



2Q 2020

V11.5.4.0 Technical Preview

- Cluster manager-aware integrated Db2 commands
- Integrated data collection via db2support
- Multiple instances & databases support
- New cluster manager configuration utility – db2cm
- Enhanced quorum type support with QDevice
- RHEL 8.1, SLES 15 SP1 support on Intel and Linux on IBM Z
- Validated on AWS with RHEL 8.1

4Q 2020

V11.5.5.0

Technical Preview → GA

- Multiple Standby Support
- Fast redeployment via import & export support
- Two node support with fencing on AWS
- Newer Pacemaker version

Q2 2021

V11.5.6.0

- Integrated bundling and install of Pacemaker stack
- Customized configurations on Azure
- Enhanced Network Resiliency
- Advance HADR DB hang detection (Linux)
- Expanded distro levels support
- Enhanced PD

V11.5.7.0

Tentative

- POWER Linux Support
- Advance HADR DB hang detection (AIX)
- Pacemaker Refresh

VNEXT

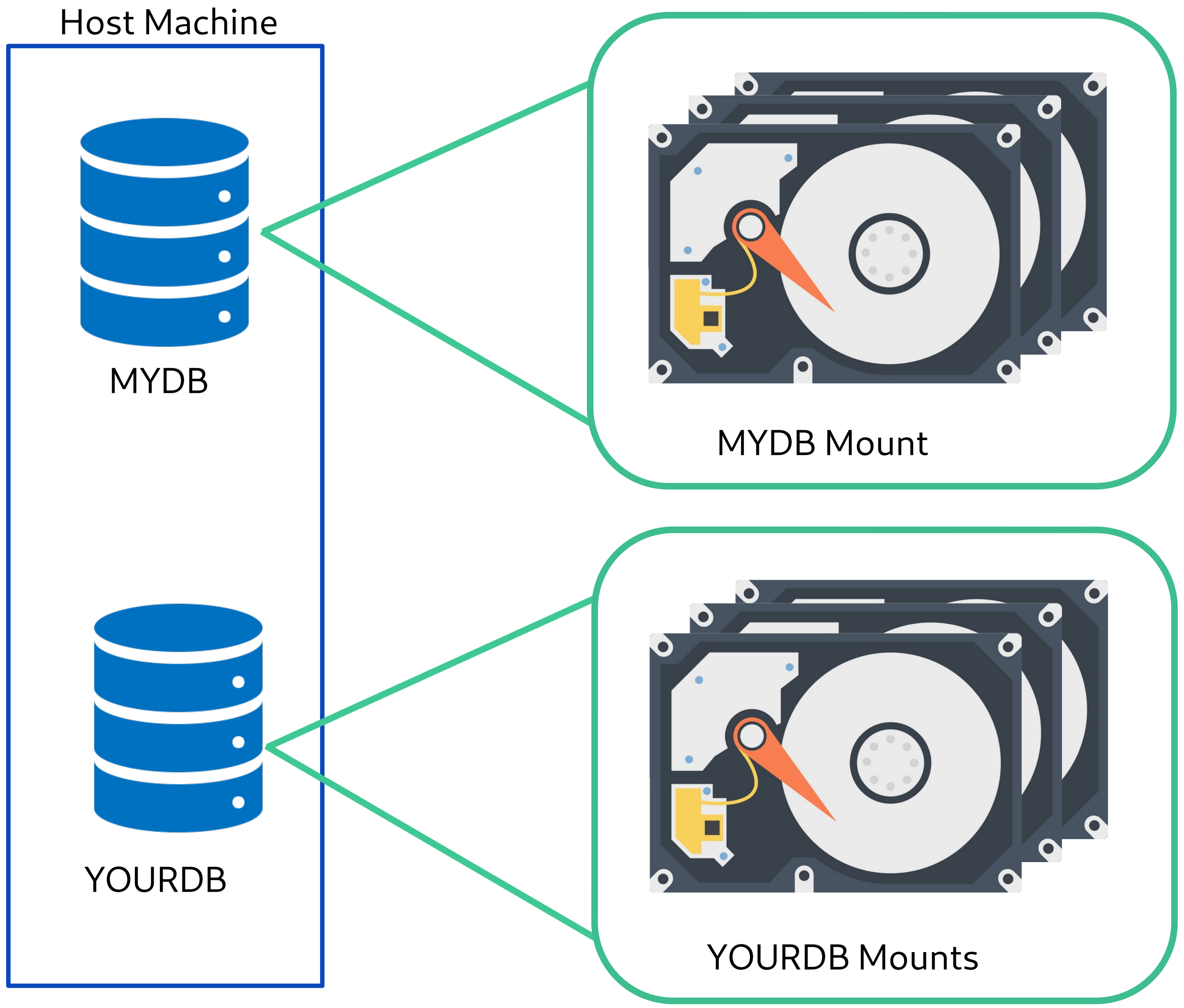
Tentative

- Mount Monitoring
- Active-Passive HA
- DPF HA
- pureScale
- Customized config on IBM Cloud & GCP
- Enhanced Security Model
- Expanded distro levels support
- More up & running automation

Note: Roadmap and content subjected to change

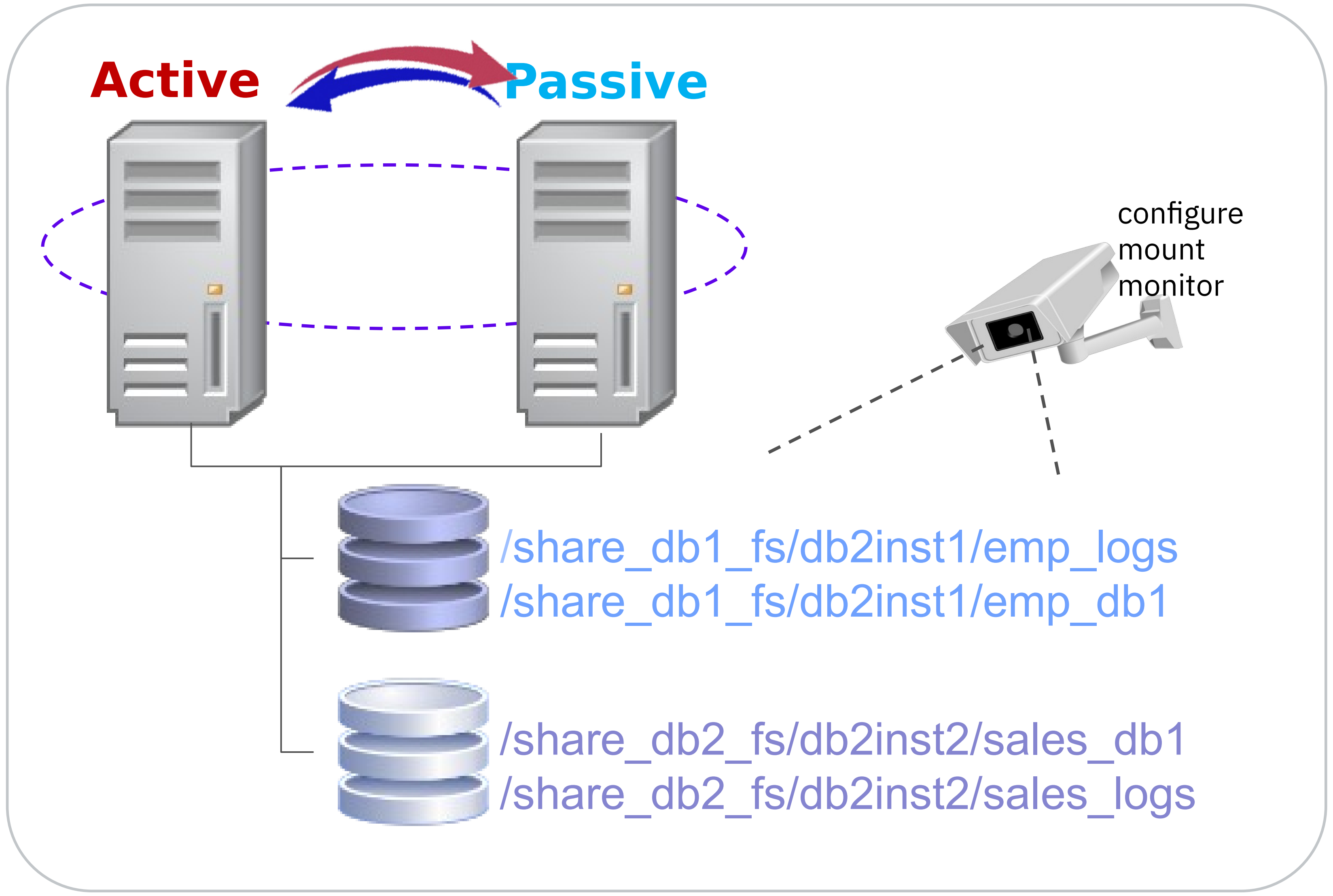
Mount Automation

- Make file-systems highly available
- Adds order constraint between the database and its associated file-systems.
- Ensures the database file-systems are operational before a database is activated.
- Attempt to automatically bring file-systems back online in failure scenarios.
- Used in various topologies.



Feature inclusion in future Db2 release subjects to change without notice.

Active-Passive HA – *Existing behaviour with TSA/RSCT*


Active
Passive
configure
mount
monitor
/share_db1_fs/db2inst1/emp_logs
/share_db1_fs/db2inst1/emp_db1
/share_db2_fs/db2inst2/sales_db1
/share_db2_fs/db2inst2/sales_logs

Setup:

- Database on shared file system
- Configure mount monitoring on the DB file system mounts
- Cluster manager ensure the shared FS is only active on one of the hosts at any given time.
- Automated file system mount point failover

With RSCT:

Key to success: RSCT's Critical Resource Protection Feature

- Defined at resource level
- Configurable actions (reboot, shutdown, none, etc) on failure
- Db2 sets action to reboot on all resources in this HA configuration

A mount point failure results in:

- mount monitor detects the failure and marks the corresponding mount resource as failed
- Critical Resource Protection is triggered to reboot the host.
- TSA detects the Db2 instance failure on ACTIVE hosts and fails over to the PASSIVE host.
- The rest of the resource model (mounts, DB, and instance) will be brought online on the passive host automatically

Feature inclusion in future Db2 release subjects to change without notice.

Active-Passive HA – *New behaviour with Pacemaker/Corosync*

With Corosync:

- Lack of disk/IP tiebreaker support means split brain scenario needs to be handled differently
- No 1-1 mapping of RSCT's Critical Resource Protection feature. This means node fencing (prevents data corruption) needs additional setup

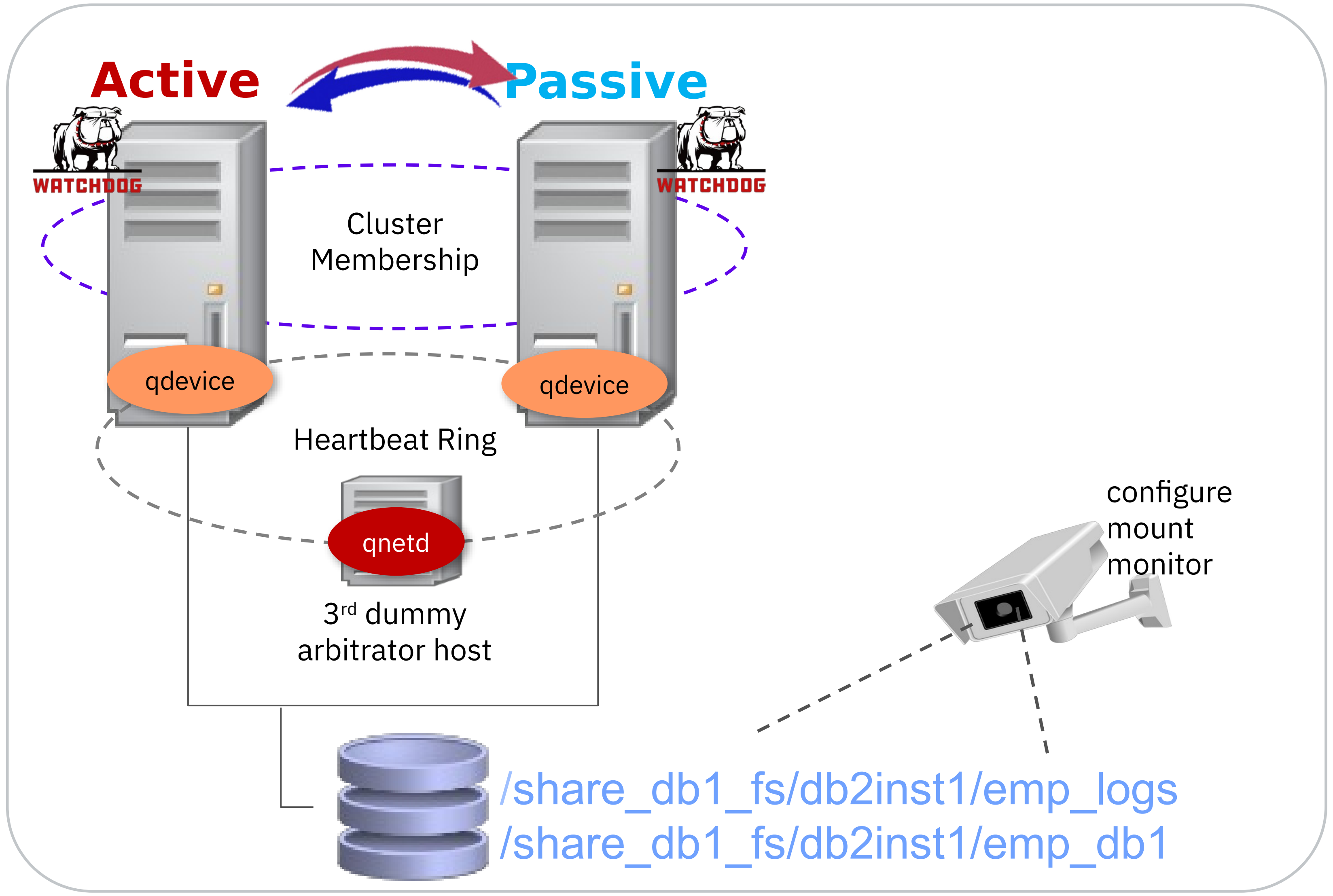
Potential Solutions:

Split brain prevention:

- Use QDevice (with a 3rd arbitrator host)
- Or use Storage-Based Death (SBD) with a shared disk across hosts

Node fencing:

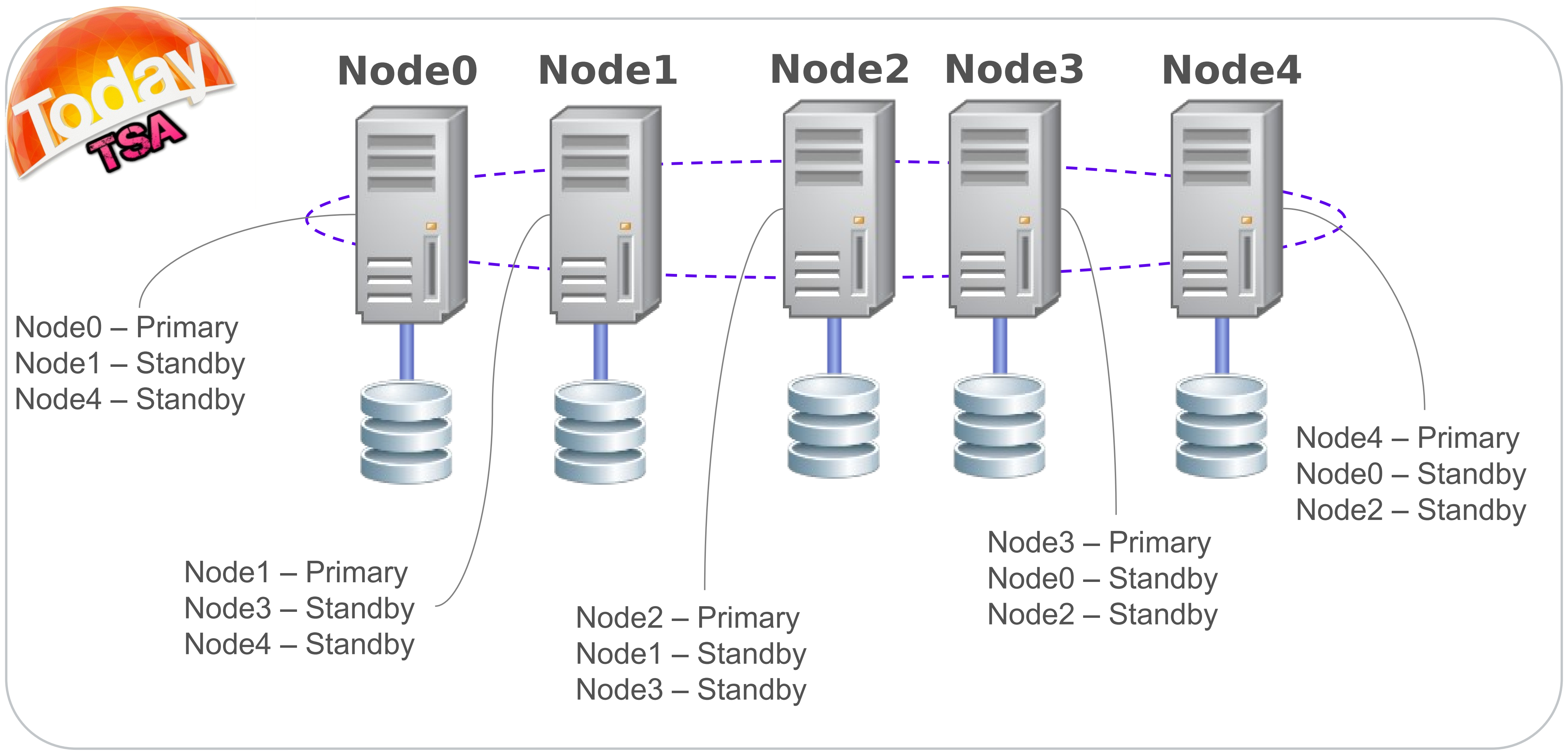
- Utilize Software Watchdog (default or separate install)
- Use in combination with QDevice to trigger a reboot when a node eviction action is deemed necessary.



Sample configuration with Qdevice
(subjected to change)

Feature inclusion in future Db2 release subjects to change without notice.

Database Partition Feature (DPF) HA configuration



Node0 Node1 Node2 Node3 Node4

Node0 – Primary
Node1 – Standby
Node4 – Standby

Node1 – Primary
Node3 – Standby
Node4 – Standby

Node2 – Primary
Node1 – Standby
Node3 – Standby

Node3 – Primary
Node0 – Standby
Node2 – Standby

Node4 – Primary
Node0 – Standby
Node2 – Standby

FUTURE
Pacemaker

- **Goal:**
  - Consolidate & Simplify Configuration by aligning support with most common use case
- **Potential design:**
  - One standby host dedicated for a few partitions
  - Multiple “micro-cluster” with separate automation within the same DPF instance
  - Roving Standby Support

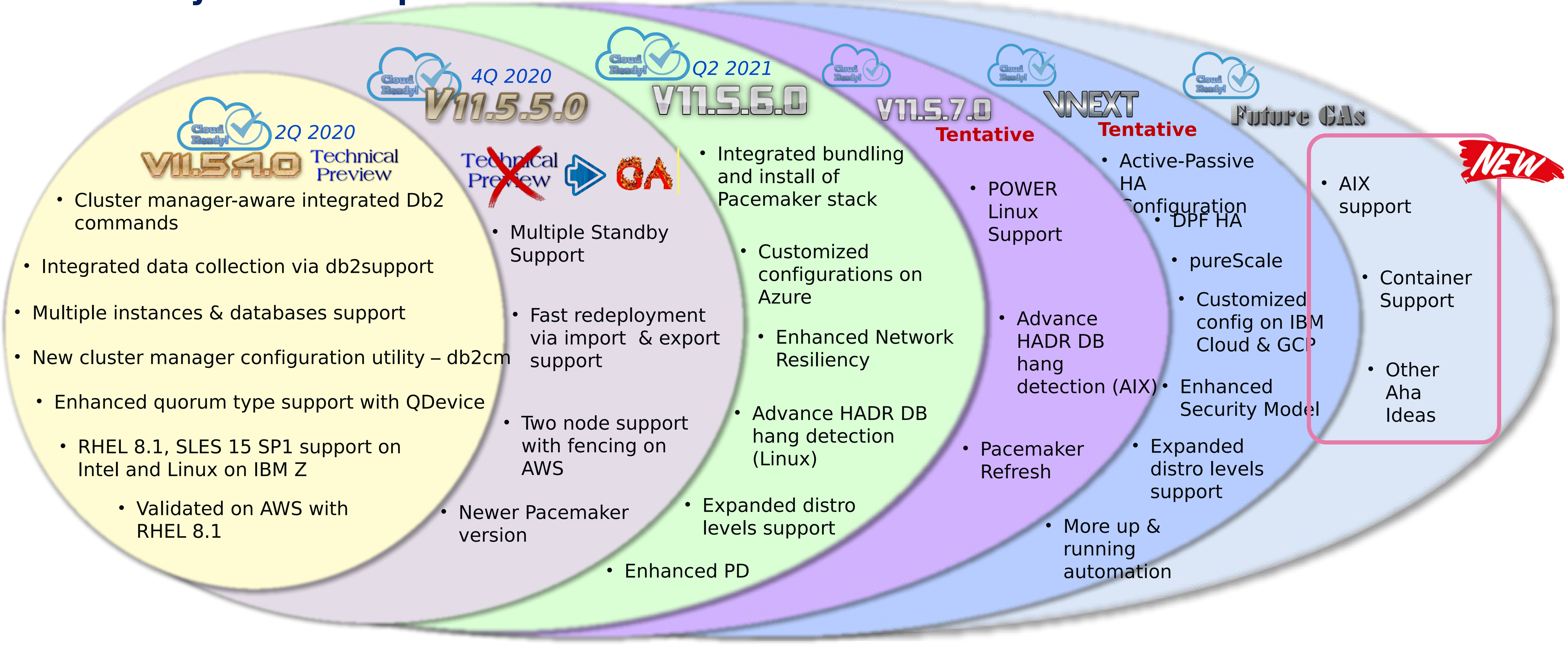
Feature inclusion in future Db2 release subjects to change without notice.

pureScale ... a teaser

- Cloud-Ready !!!!!
- New & Simplified Resource Model
- Different quorum mechanism (fewer shared disk requirement)
- Db2-optimized node-liveliness test
- More accurate RDMA network liveliness test
- Built-in RDMA network performance evaluation and aggregate history
- Smarter unified cluster management utility interface
- Reduced dependency in support infrastructure
- ... and many others

Feature inclusion in future Db2 release subjects to change without notice.

Our Journey … final Stop



V11.5.4.0 2Q 2020 Technical Preview

- Cluster manager-aware integrated Db2 commands
- Integrated data collection via db2support
- Multiple instances & databases support
- New cluster manager configuration utility – db2cm
- Enhanced quorum type support with QDevice
- RHEL 8.1, SLES 15 SP1 support on Intel and Linux on IBM Z
- Validated on AWS with RHEL 8.1

V11.5.5.0 4Q 2020 ~~Technical Preview~~ GA

- Multiple Standby Support
- Fast redeployment via import & export support
- Two node support with fencing on AWS
- Newer Pacemaker version

V11.5.6.0 Q2 2021

- Integrated bundling and install of Pacemaker stack
- Customized configurations on Azure
- Enhanced Network Resiliency
- Advance HADR DB hang detection (Linux)
- Expanded distro levels support
- Enhanced PD

V11.5.7.0 Tentative

- POWER Linux Support
- Advance HADR DB hang detection (AIX)
- Pacemaker Refresh

VNEXT Tentative

- Active-Passive HA Configuration
- DPF HA
- pureScale
- Customized config on IBM Cloud & GCP
- Enhanced Security Model
- Expanded distro levels support
- More up & running automation

Future GAs

NEW

- AIX support
- Container Support
- Other Aha Ideas

Note: Roadmap and content subjected to change

Notice and Disclaimers

- IBM’s statements regarding its plans, directions, and intent are subject to change or withdrawal without notice and at IBM’s sole discretion.
- Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision.
- The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract.
- The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.
- Performance is based on measurements and projections using standard IBM benchmarks in a controlled environment. The actual throughput or performance that any user will experience will vary depending upon many factors, including considerations such as the amount of multiprogramming in the user’s job stream, the I/O configuration, the storage configuration, and the workload processed. Therefore, no assurance can be given that an individual user will achieve results similar to those stated here.

Notice and Disclaimers

-

- **U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.**
- Information in these presentations (including information relating to products that have not yet been announced by IBM) has been reviewed for accuracy as of the date of initial publication and could include unintentional technical or typographical errors. IBM shall have no responsibility to update this information. **This document is distributed "as is" without any warranty, either express or implied. In no event, shall IBM be liable for any damage arising from the use of this information, including but not limited to, loss of data, business interruption, loss of profit or loss of opportunity.** IBM products and services are warranted per the terms and conditions of the agreements under which they are provided.

- IBM products are manufactured from new parts or new and used parts.
  In some cases, a product may not be new and may have been previously installed. Regardless, our warranty terms apply."

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

- Performance data contained herein was generally obtained in a controlled, isolated environments. Customer examples are presented as illustrations of how those customers have used IBM products and the results they may have achieved. Actual performance, cost, savings or other results in other operating environments may vary.

- References in this document to IBM products, programs, or services does not imply that IBM intends to make such products, programs or services available in all countries in which IBM operates or does business.

- Workshops, sessions and associated materials may have been prepared by independent session speakers, and do not necessarily reflect the views of IBM. All materials and discussions are provided for informational purposes only, and are neither intended to, nor shall constitute legal or other guidance or advice to any individual participant or their specific situation.

- It is the customer's responsibility to insure its own compliance with legal requirements and to obtain advice of competent legal counsel as to the identification and interpretation of any relevant laws and regulatory requirements that may affect the customer's business and any actions the customer may need to take to comply with such laws. IBM does not provide legal advice or represent or warrant that its services or products will ensure that the customer follows any law.

Questions: thaynes@ca.ibm.com

The information in this presentation is representative of the presenter and their views and opinions are not necessarily those of IBM.