Stay ahead with Pacemaker, the new Db2 cluster manager for automated failover

think 2021

Gerry Sommerville
Db2 Developer
IBM Canada

Andreas Christian
Data & AI Specialist
IBM Germany
linkedin.com/in/andreas-christian

Stefan Hummel
Data & AI Specialist
IBM Germany
linkedin.com/in/stefanhummel

IBM

Please note

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.

Notices and disclaimers

U.S. Government Users Restricted Rights — use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

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. The performance data and client examples cited are presented for illustrative purposes only. Actual performance results may vary depending on specific configurations and operating conditions.

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.

Notices and disclaimers continued

It is the customer's responsibility to ensure 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.

Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products about this publication and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. IBM does not warrant the quality of any third-party products, or the ability of any such third-party products to interoperate with IBM's products. **IBM expressly disclaims all warranties, expressed or implied, including but not limited to, the implied warranties of merchantability and fitness for a purpose.**

The provision of the information contained herein is not intended to, and does not, grant any right or license under any IBM patents, copyrights, trademarks or other intellectual property right.

IBM, the IBM logo, and ibm.com are trademarks of International Business Machines Corporation, registered in many jurisdictions worldwide. Other product and service names might be trademarks of IBM or other companies. A current list of IBM trademarks is available on the Web at "Copyright and trademark information" at: www.ibm.com/legal/copytrade.shtml.

Service Offering Pacemaker

IBM Expert Labs DACH - Data and AI

Experienced Db2 specialists help you to run Pacemaker quickly!

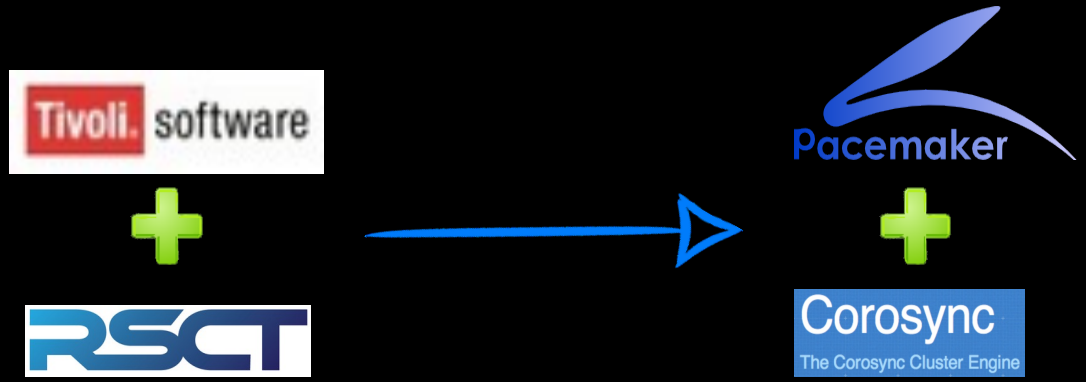
- Offering for new customers

  Setup of Pacemaker as Cluster Manager for HADR environments
  - Analysis of *Requirements* and *Restrictions*
    - *Software Build, Versions*
  - Installation and setup of Pacemaker Software (Qdevice)
  - Validation of the cluster (Testcases like Reboot, user takeover,...)

- Offering for TSA customers

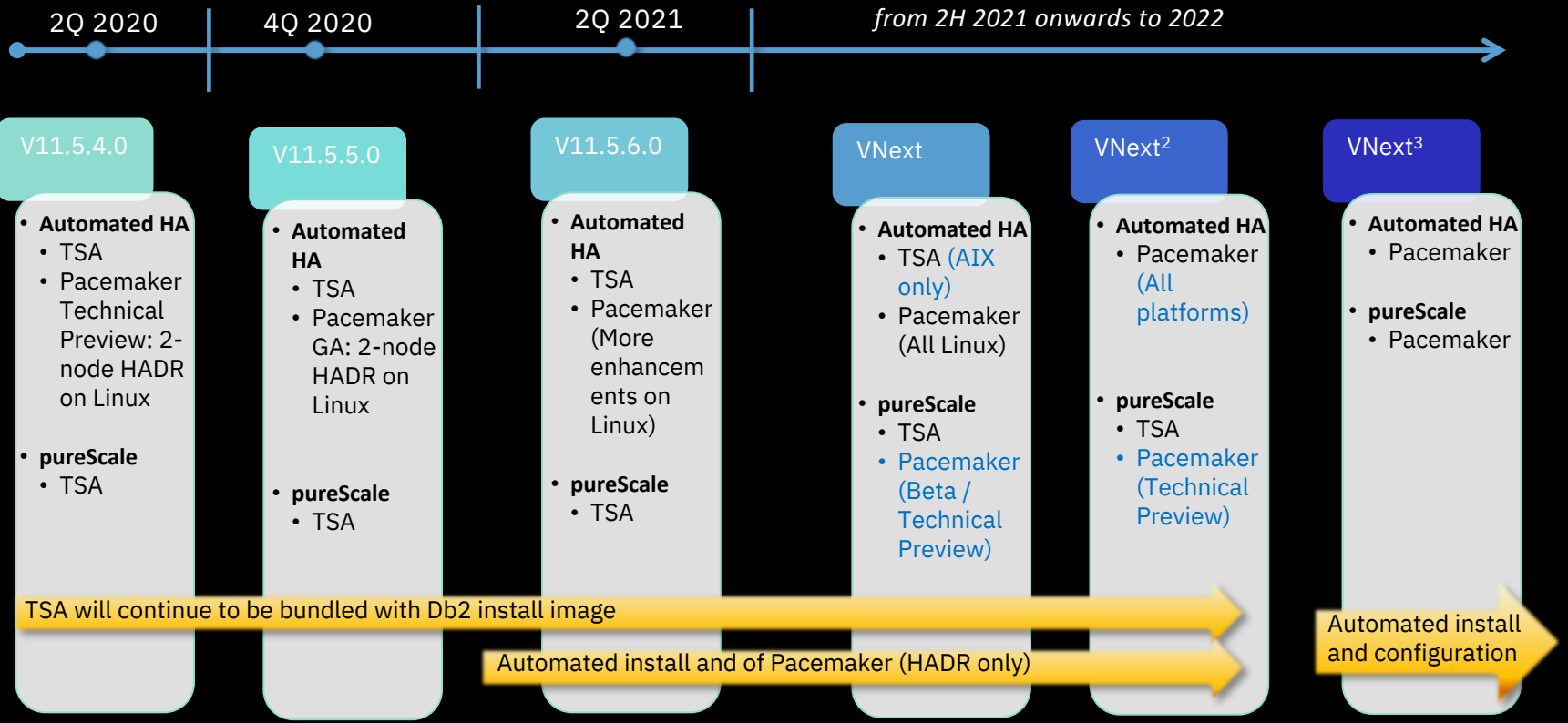
  Migration of existing TSA Cluster to Pacemaker
  - Installation Pacemaker Software
  - Migration
    - Backup of existing TSA configurations
    - TSA Cluster Cleanup
    - Creation of Pacemaker Cluster and Ressources
  - Validation of Cluster (Testcases like Reboot, user takeover,...)

Why Pacemaker?



- Modernized stack
  - Cloud ready
  - Open source
    Allow for future port to AIX
- Simpler…
  - Architecture
  - Diagnostics
  - Support model
- Better performance

Sneak peek at the road map



Note: Roadmap subjected to change

Supported platforms

Note:
There is no plan to support older OS/Db2 releases than what is currently supported

Architecture / Platforms / OS Version	TSA	Pacemaker
Intel / RHEL / 7.x	V11.5.4.0	No Plan
Intel / RHEL / 8.1	V11.5.4.0	V11.5.4.0+
Intel / SLES / 12 SPx	V11.5.4.0	No Plan
Intel / SLES / 15 SPy	No	V11.5.4.0+
Linux on IBM Z / RHEL 8.1	V11.5.4.0	V11.5.4.0+
Linux on IBM Z / SLES 15 SP1	No	V11.5.4.0+
POWER 8 RHEL 7.x	Yes	No Plan
POWER 8 & 9 / RHEL 8.1	No	In roadmap
POWER / AIX / 7.2 TL4	V11.5.4.0	In roadmap

Environments	TSA Support Statement	Pacemaker Support Statement
On-premise DC	Yes, see table above	Yes, see table above
Non-containerized Private Cloud	No	Yes, see table above
Non-containerized Public Cloud	No	Yes, see table above, validated on AWS on Intel RHEL
Container	No	Not yet.

Quorum Support

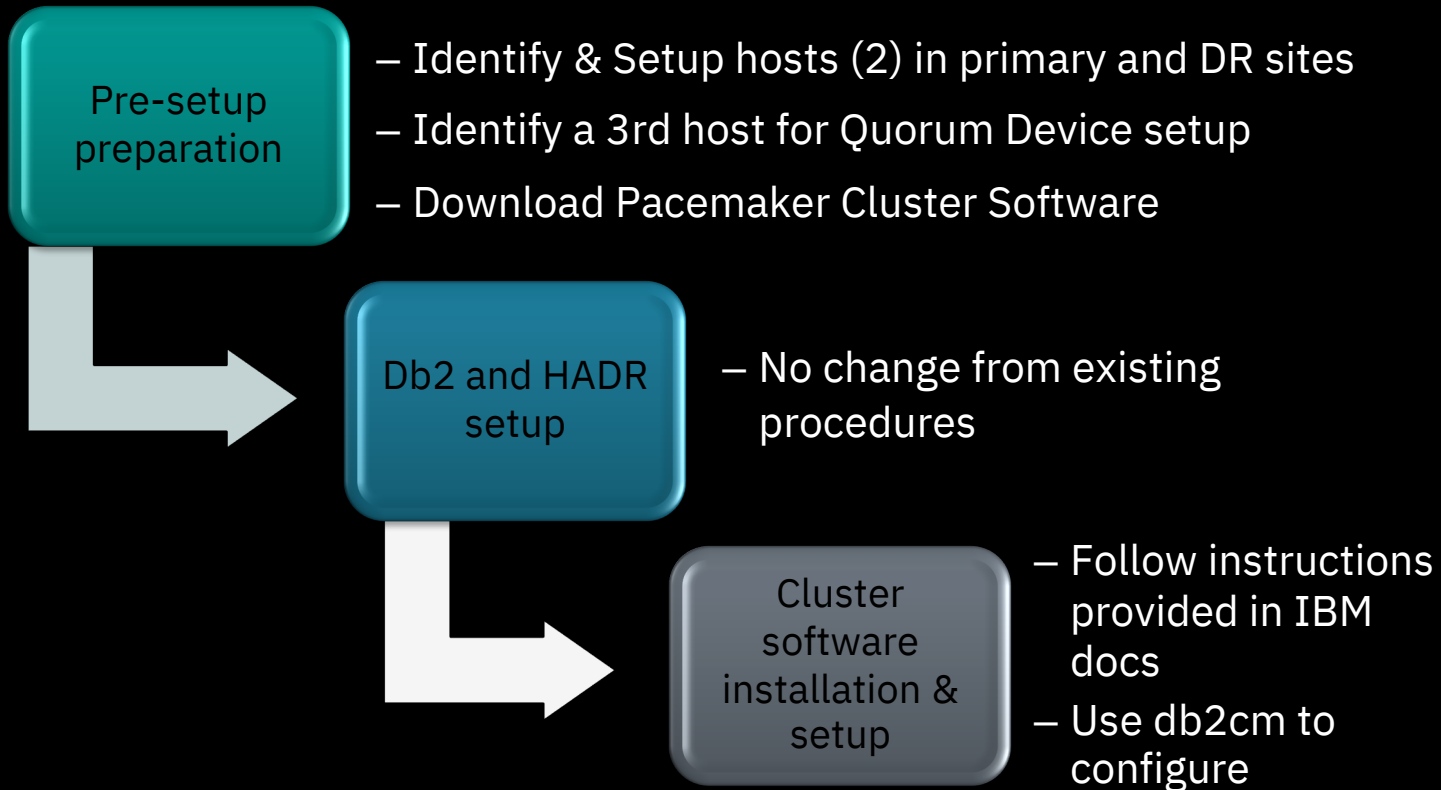
- No IP/Disk tiebreaker support in Pacemaker

- Pacemaker recommends using Qdevice for reliable quorum
  - Qdevice requires a 3rd light weight host to run an arbitrator daemon.
  - No need to install Db2 or full Pacemaker stack on the 3rd host.
  - Small memory footprint.

- A single Qdevice host can provide quorum support for multiple clusters.

- Qdevice is the recommended quorum solution.

High level flow of new installation & Setup

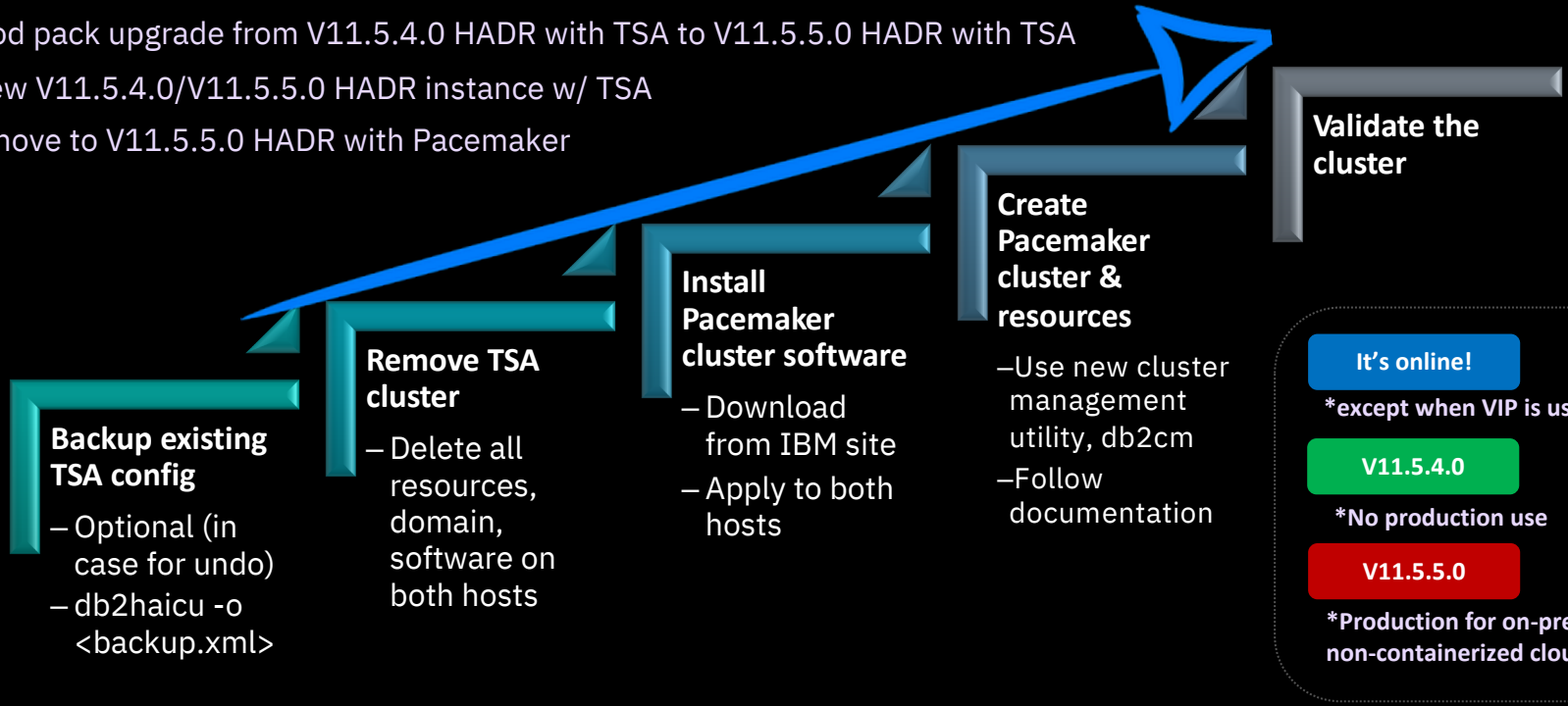


Convert from TSA to Pacemaker

Typical "in-release" conversion scenario, start with one of the following:

- Mod pack upgrade from V11.5.0.0 HADR with TSA to V11.5.4.0 HADR with TSA
- Mod pack upgrade from V11.5.4.0 HADR with TSA to V11.5.5.0 HADR with TSA
- New V11.5.4.0/V11.5.5.0 HADR instance w/ TSA

Then move to V11.5.5.0 HADR with Pacemaker



Resource Model

The resource model specifies the healthy state of the cluster, and how Pacemaker will react in order to restore the cluster to that state in case of failure. You can view the configured resource model using the "crm config show" command.

The resource model is composed of several parts

Resource agent scripts

Scripts developed by Db2 which implement specific actions

Located under the directory /usr/lib/ocf/resource.d/heartbeat

- db2ethmon
  - Ethernet adapter monitor
- db2inst
  - db2sysc start, stop, monitor
- db2hadr
  - HADR database start, stop, monitor and takeover.
- IPAddr2
  - Open-source agent used for virtual IP

Resources

Services made highly available by the cluster.

Utilizes a resource agent and a given configuration to start, stop, monitor, (etc) different services.

Pacemaker will call the resource agent script specifying which action to perform based on the resource configuration.

Constraints

Rules external to the resource definition that determine where and when a resource can run.

3 types of constraints:

- Location constraints
  - Define where a resource can run
- Co-location constraints
  - Define a dependency on another resource.
- Order constraints
  - Define an order in which the resources must start.

db2cm -list

```
[root@db2tea1 ~]# db2cm -list
```

Domain

```
Domain information:
Domain name                 = hadom
Pacemaker version           = 2.0.2-1.db2pcmk.el8
Corosync version            = 3.0.3
Current domain leader       = db2tea1
Number of nodes             = 2
Number of resources         = 6
```

Cluster membership

```
Node information:
Name name            State
---------------      --------
db2tea1              Online
kedge1               Online
```

Resources

```
Resource Information:

Resource Name       = db2_db2inst1_db2inst1_SAMPLE
  Resource Type                 = HADR
    DB Name                     = SAMPLE
    Managed                     = true
    HADR Primary Instance       = db2inst1
    HADR Primary Node           = db2tea1
    HADR Primary State          = Online
    HADR Standby Instance       = db2inst1
    HADR Standby Node           = kedge1
    HADR Standby State          = Online

Resource Name       = db2_db2tea1_db2inst1_0
  State                         = Online
  Managed                       = true
  Resource Type                 = Instance
    Node                        = db2tea1
    Instance Name               = db2inst1

Resource Name       = db2_db2tea1_
  State                         = Online
  Managed                       = true
  Resource Type                 = Network Interface
    Node                        = db2tea1
    Interface Name              = eth1

Resource Name       = db2_kedge1_db2inst1_0
  State                         = Online
  Managed                       = true
  Resource Type                 = Instance
    Node                        = kedge1
    Instance Name               = db2inst1

Resource Name       = db2_kedge1_eth1
  State                         = Online
  Managed                       = true
  Resource Type                 = Network Interface
    Node                        = kedge1
    Interface Name              = eth1
```

db2cm –list (cont'd)

Fence

```
Fencing Information:
  Not Configured
```

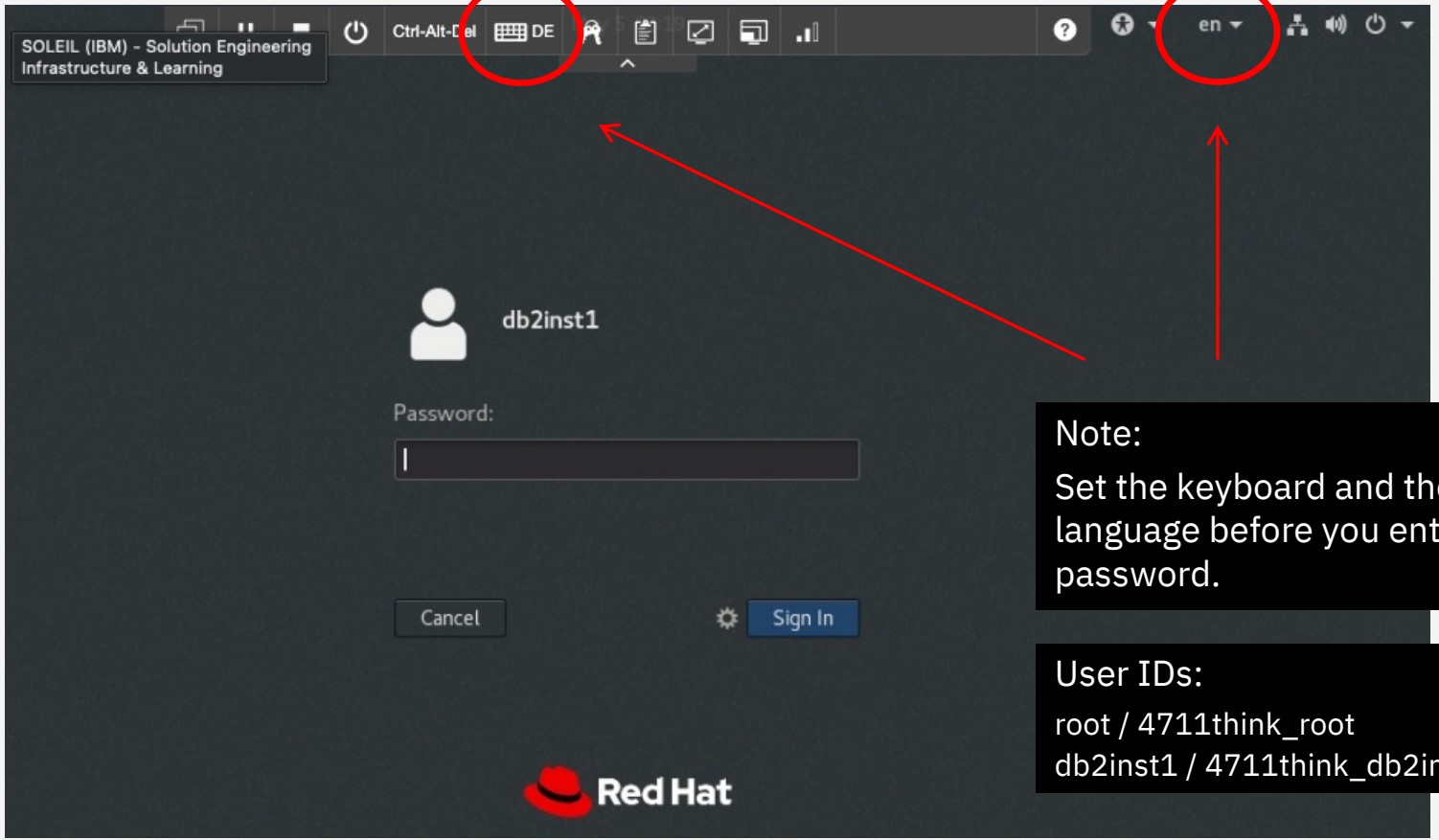
Quorum

```
Quorum Information:
  Qdevice

Qdevice information
-------------------
Model:                  Net
Node ID:                1
Configured node list:
    0 Node ID = 1
    1 Node ID = 2
Membership node list:  1, 2

Qdevice-net information
----------------------
Cluster name:       hadom
QNetd host:         tierce1:5403
Algorithm:          LMS
Tie-breaker:        Node with lowest node ID
State:              Connected
```

Lab Environment - Login



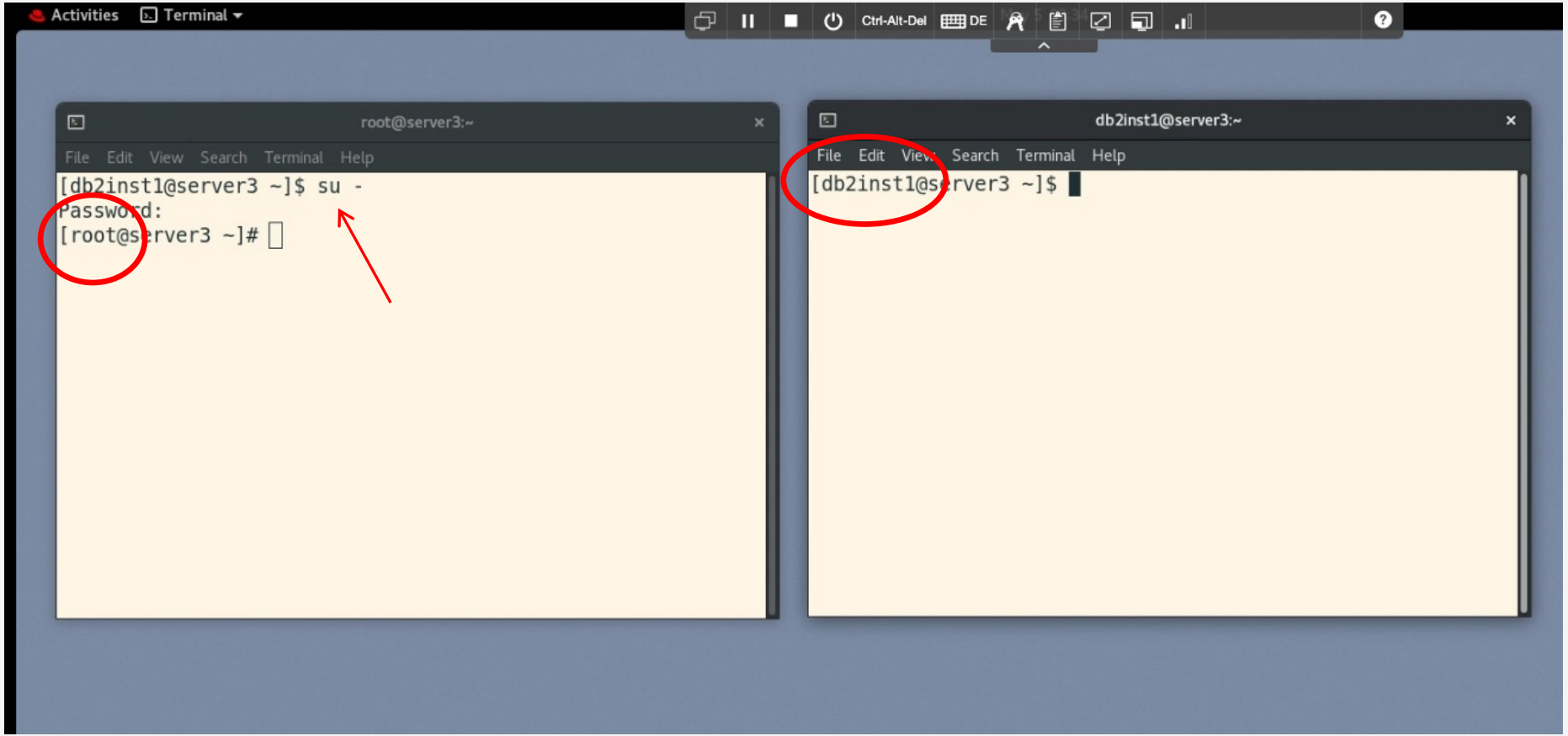
Note:

Set the keyboard and the language before you enter the password.

User IDs:

root / 4711think_root
db2inst1 / 4711think_db2inst1

Lab Environment - Terminals



Thank you

Important Commands

crm status
- Prints the status of the cluster at the time it was run

crm_mon
- Same output as crm status, but continuously updates as the cluster is running.

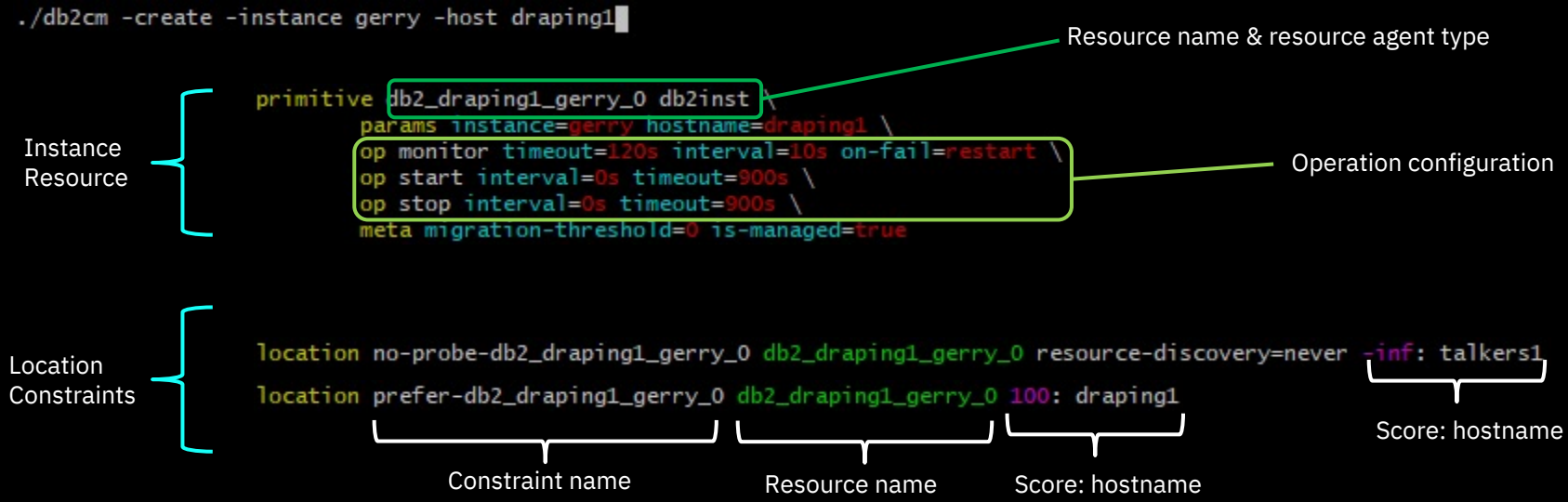
crm config show
- Prints out cluster's configuration including resources, constraints, and more.

crm resource refresh
- Resets resources failure counts. May be asked to run this by db2 support.

db2cm –list
- Db2 command that prints information relating to resource status and cluster configuration.

db2cm

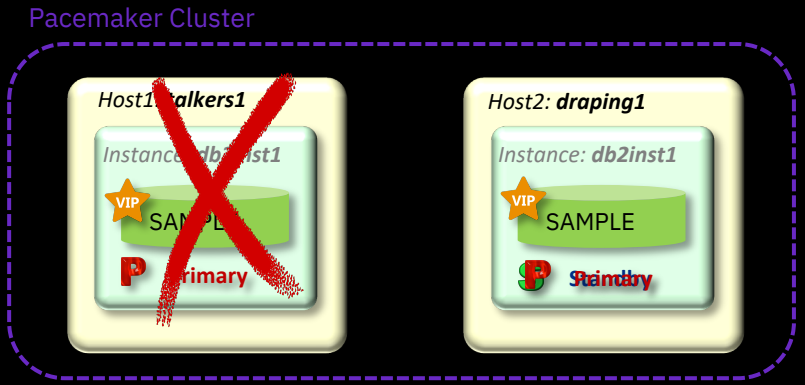
- New command line tool replacing db2haicu
- Configures automation for Db2 ‘services’ (db2 instance, HADR database)



The above information can be displayed using the `crm config show` command.

See the Pacemaker documentation for more information.

Host failure -Automatic Failover



1. Host 1 fails

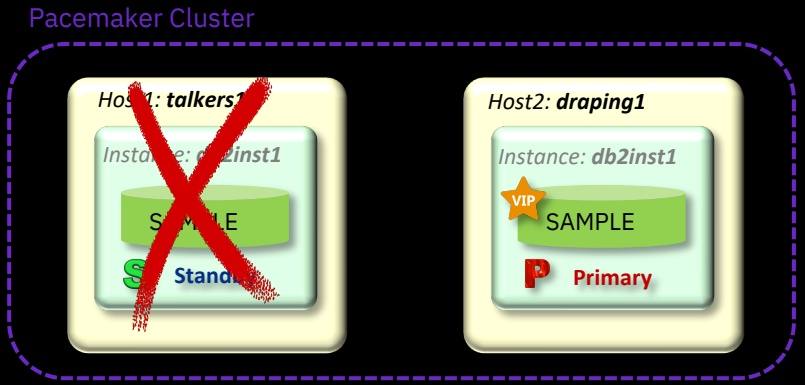
2. Pacemaker detects talkers1 has left the cluster via Corosync and the resources running on that host are now offline.

3. Pacemaker issues a takeover on the standby database via the promote action of the db2hadr resource.
db2 TAKEOVER HADR ON DB SAMPLE

4. Once the TAKEOVER completes successfully, the db2hadr resource agent sets a reintegration flag for the database.

5. The virtual IP starts on host2 which now hosts the primary database.

Host failure (cont'd) –Database reintegration



1. Host1 comes back online and rejoins the cluster, but neither the database nor instance is running.

2. Pacemaker then starts the instance via the start action specified by the db2inst resource.
db2start (via db2gcf)

3. Once the instance is started, the db2inst start action will also attempt to activate all databases asynchronously.

Assuming TAKEOVER was successful, the reintegration flag will have been set. Upon detecting the reintegration flag, the database will be activated via
db2 START HADR ON DB SAMPLE AS STANDBY

Note 1: If the TAKEOVER had not been successful, then the reintegration flag would not have been set. In such a case the database on host 1 would resume the PRIMARY role via db2 ACTIVATE DB SAMPLE.

Note 2: If databases could not be activated as part of instance start, then Pacemaker will activate them individually.