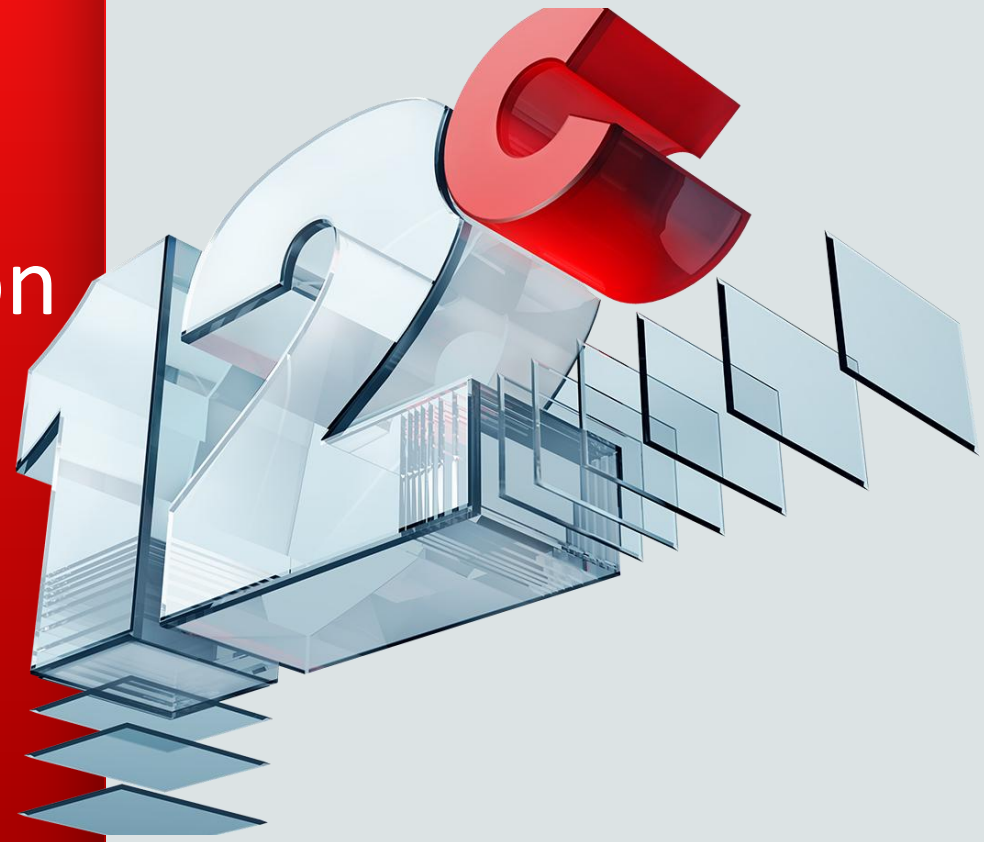

ORACLE®

Safe Harbor Statement

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

ORACLE

Headache-free Split Brain Resolution



Ian Cookson

Product Manager for Oracle Clusterware

Program Agenda

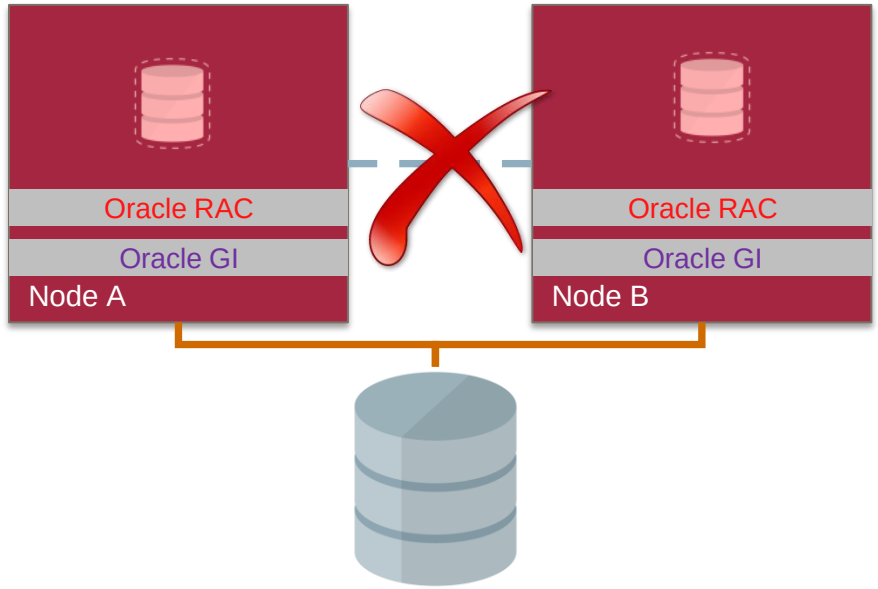
1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

Program Agenda

1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

Split Brain – What Does It Mean for Oracle Clusterware?

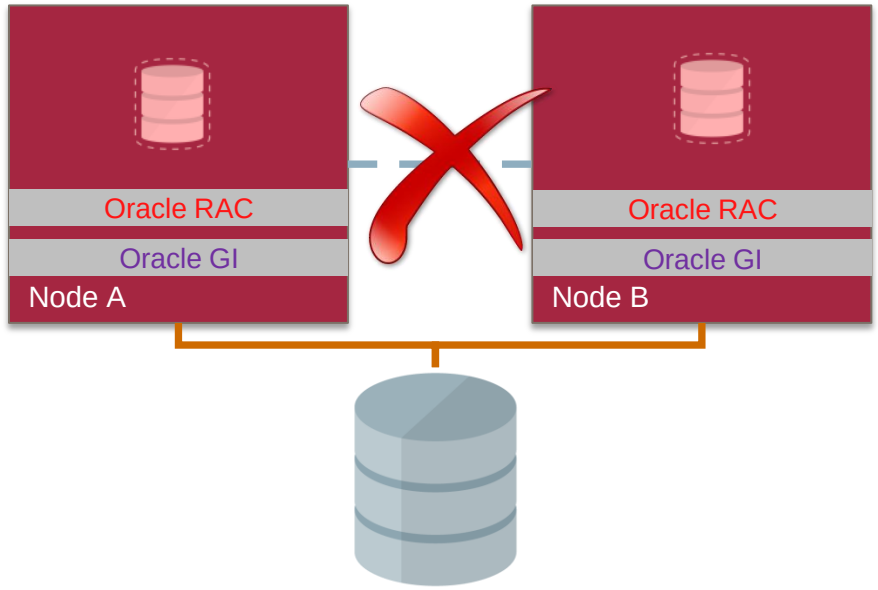
"a condition in which Oracle Clusterware believes that there is a communication failure between nodes"



Split Brain – What's Happening?

1. 'Private Interconnect Failure'
2. Node A believes it is 'the cluster'

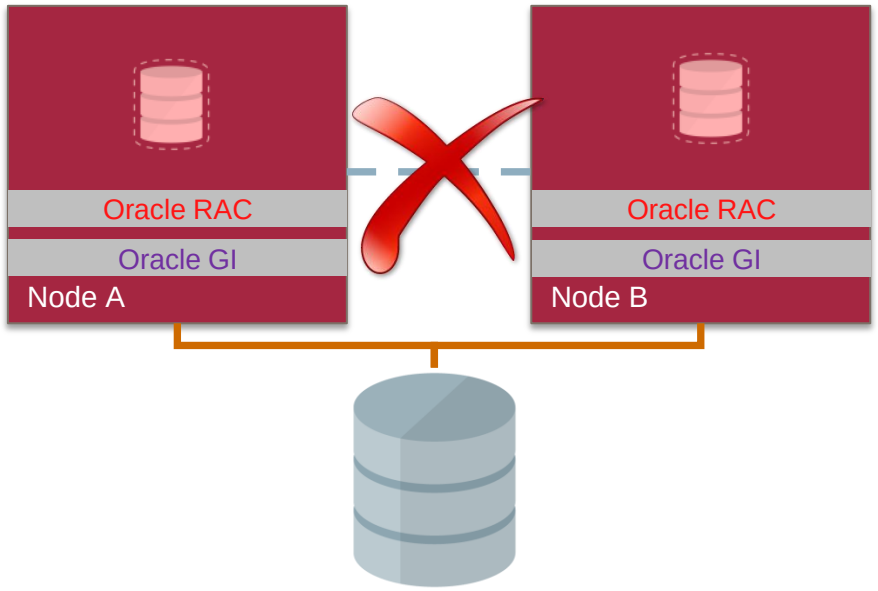
   Node B believes it is 'the cluster'
3. Now what?



Integrity of the shared data is paramount!

Split Brain – How to Resolve It?

- Status of Cluster Nodes?
  - Is a Node dead? or unresponsive?
  - Is it just a network issue?
- Surviving cluster cohorts?
  - Two-node cluster is simple…
- Priorities…?



Integrity of the shared data is paramount!

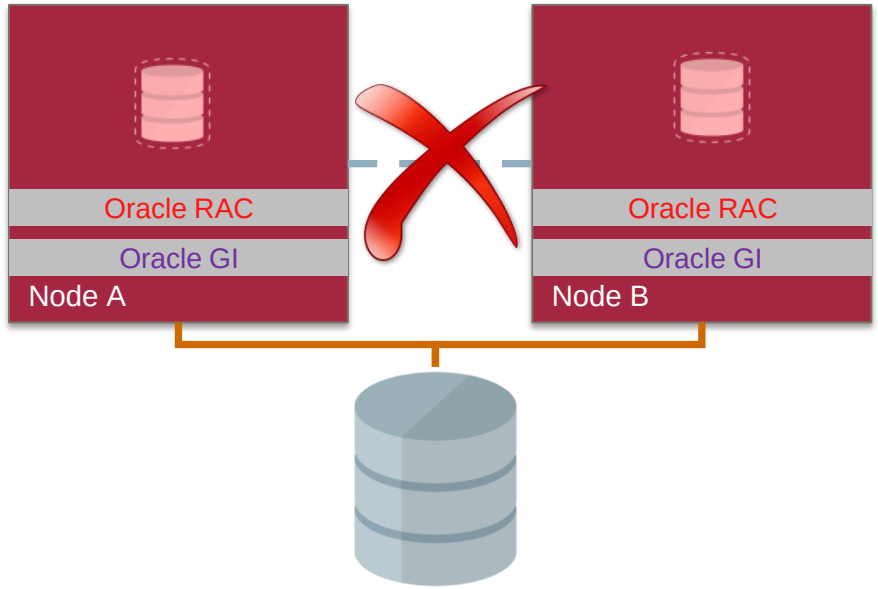
Program Agenda

1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

Clusterware Concepts, Part 1

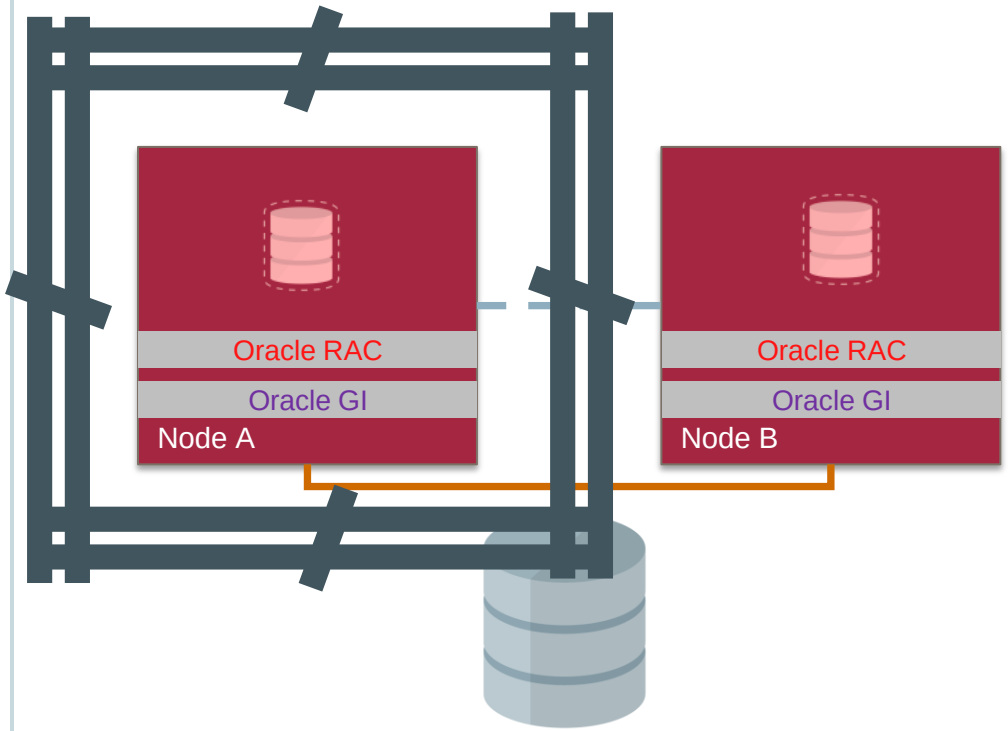
Clusterware terms:

- Fencing
- Rebootless Node Fencing
- Node Eviction
- misscount
- disktimeout



Clusterware Concepts - Fencing

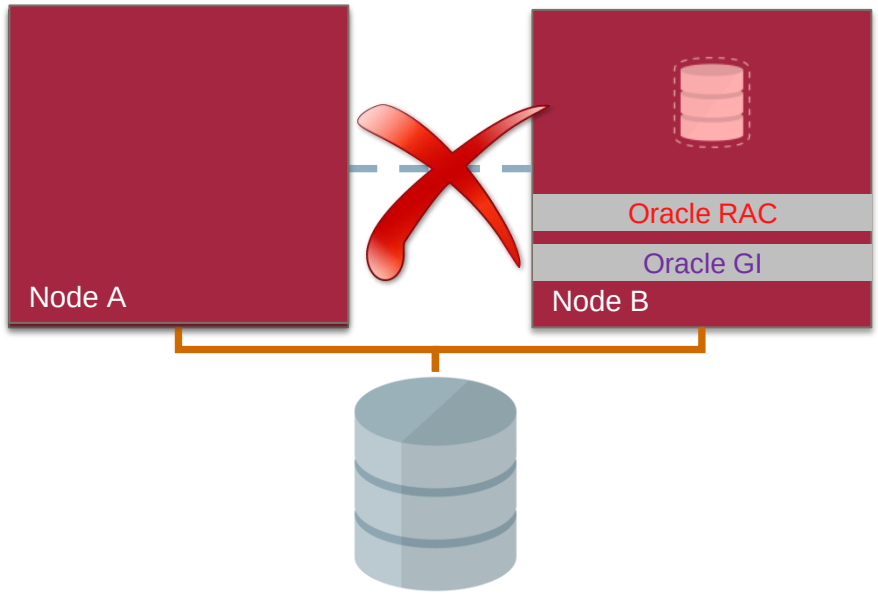
Fencing – conceptually ‘fencing’ the node off from shared cluster resources



Clusterware Concepts – Fencing Actions

Two approaches to implement node fencing in Clusterware:

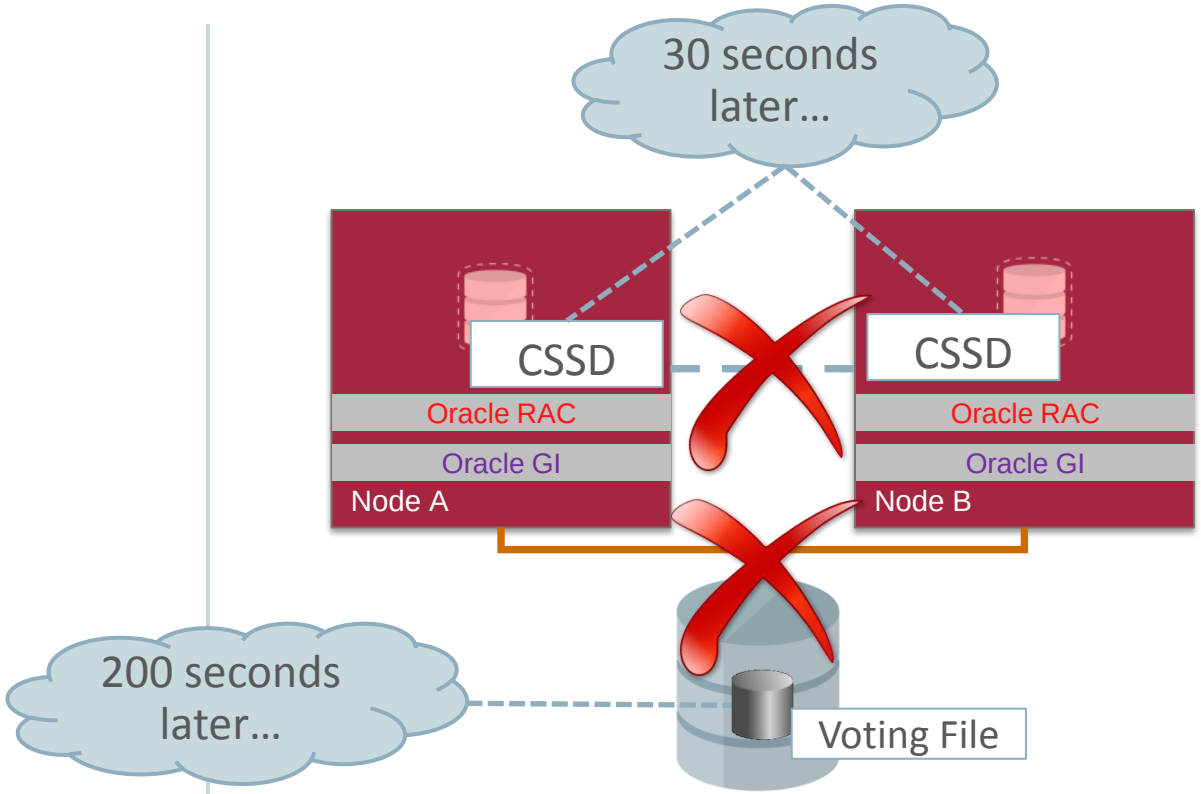
- Rebootless Node Fencing
- Node Eviction (reboot)



Clusterware Concepts – When Does Fencing Occur?

When is a node fenced?

- *misscount* is exceeded
- *disktimeout* is exceeded

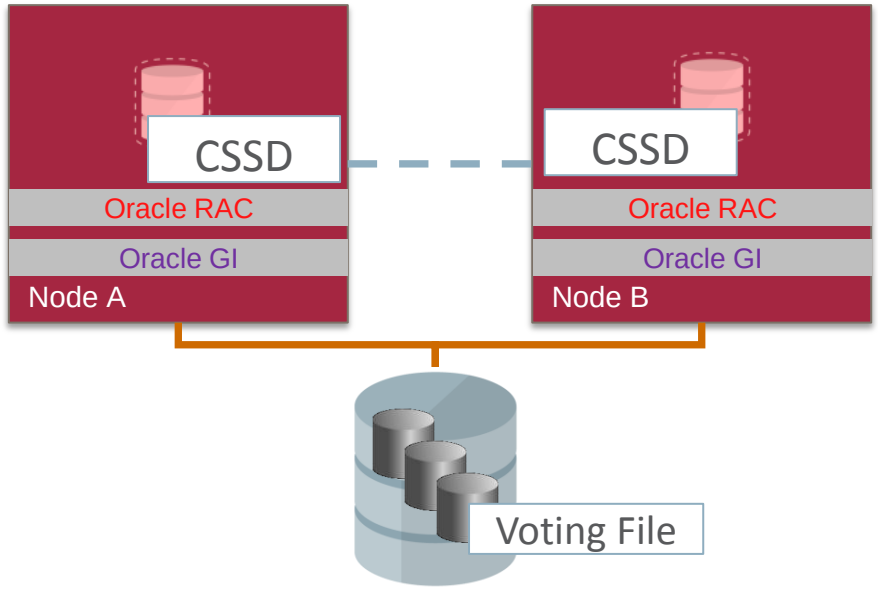


Integrity of the shared data is paramount!

Clusterware Concepts – What is the Voting File?

Mechanism by which Clusterware tests that it can:

- read/write to shared storage
- verifies cluster participation



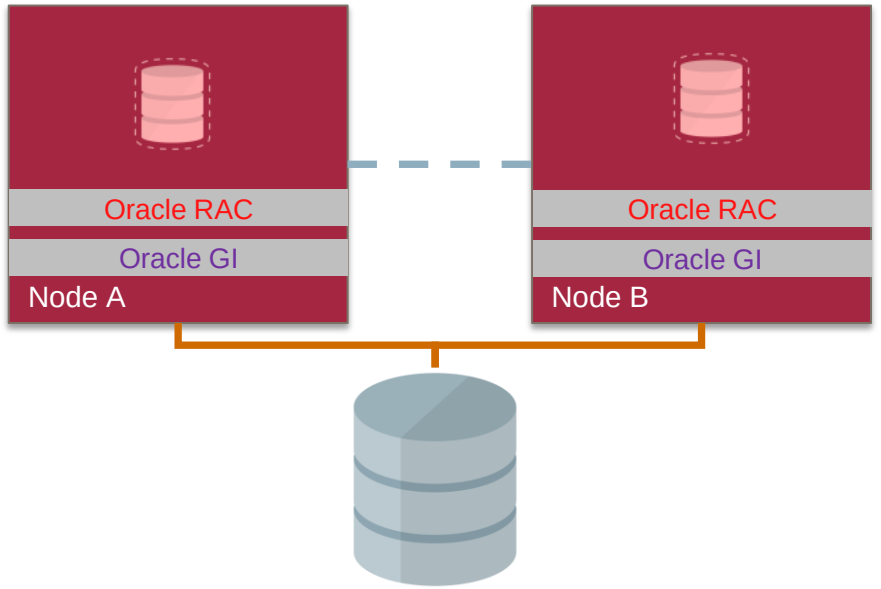
Integrity of the shared data is paramount!

Clusterware Concepts – Common Causes for Fencing?

Resource starvation (memory/cpu)

I/O path disruption

Outage on private interconnect

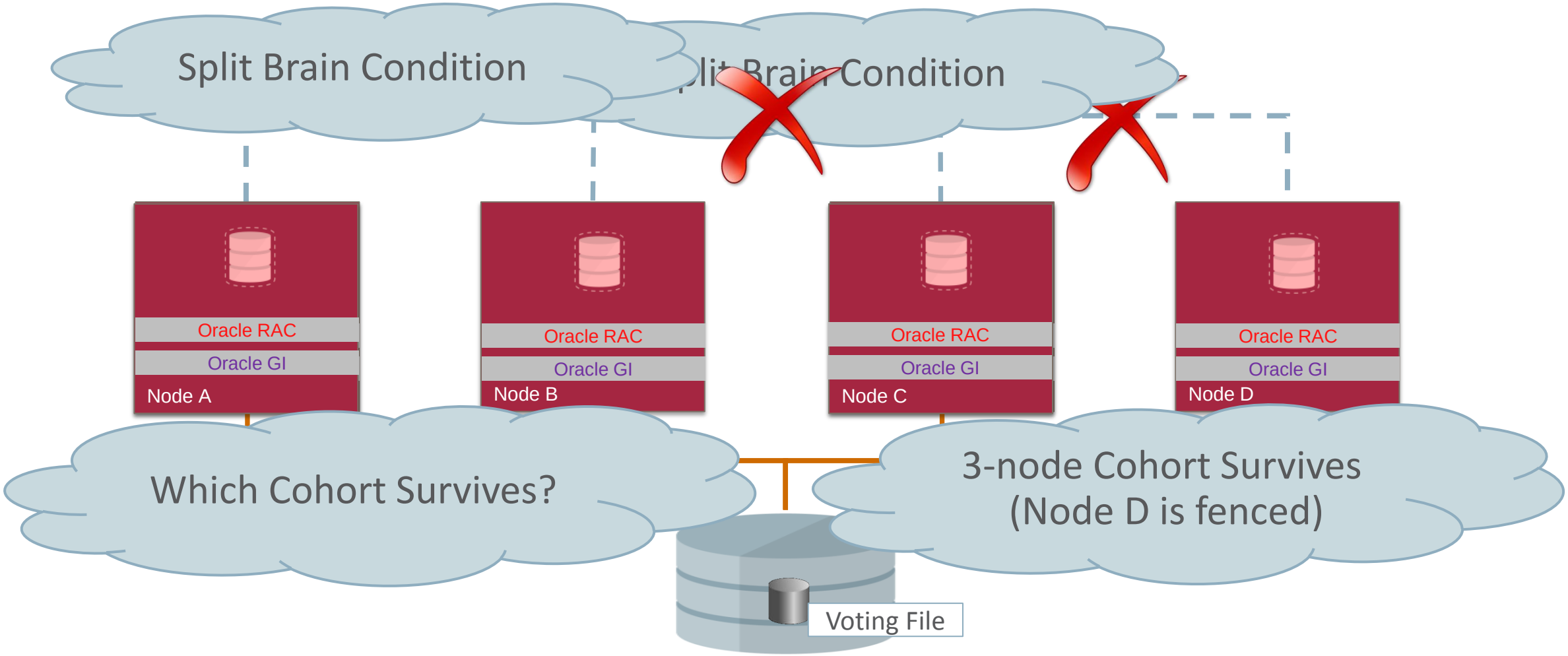


Integrity of the shared data is paramount!

Program Agenda

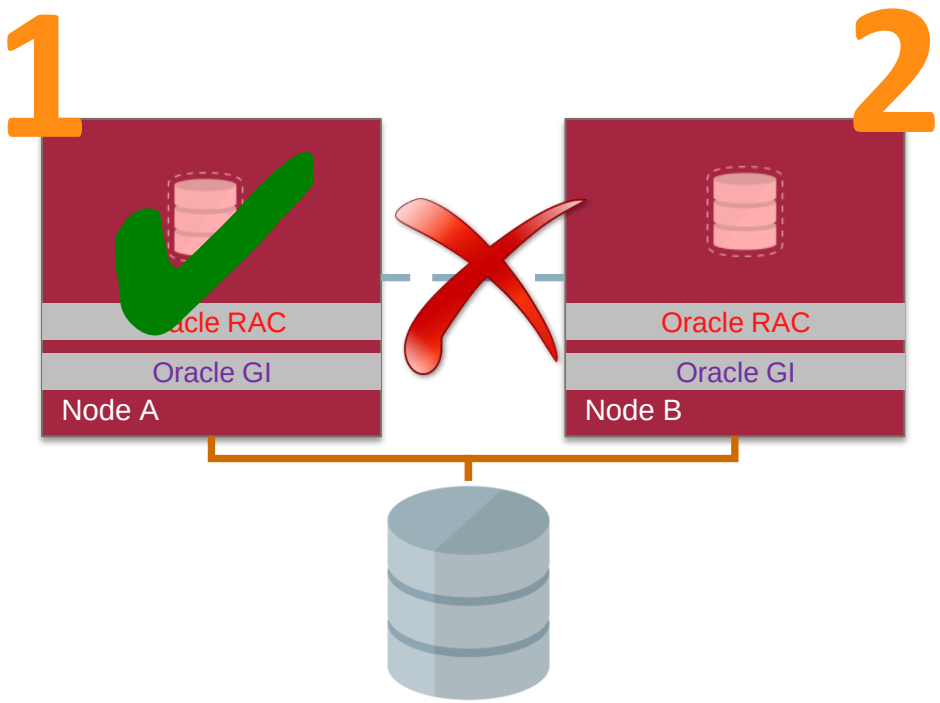
1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

Split Brain Resolution in Current Releases



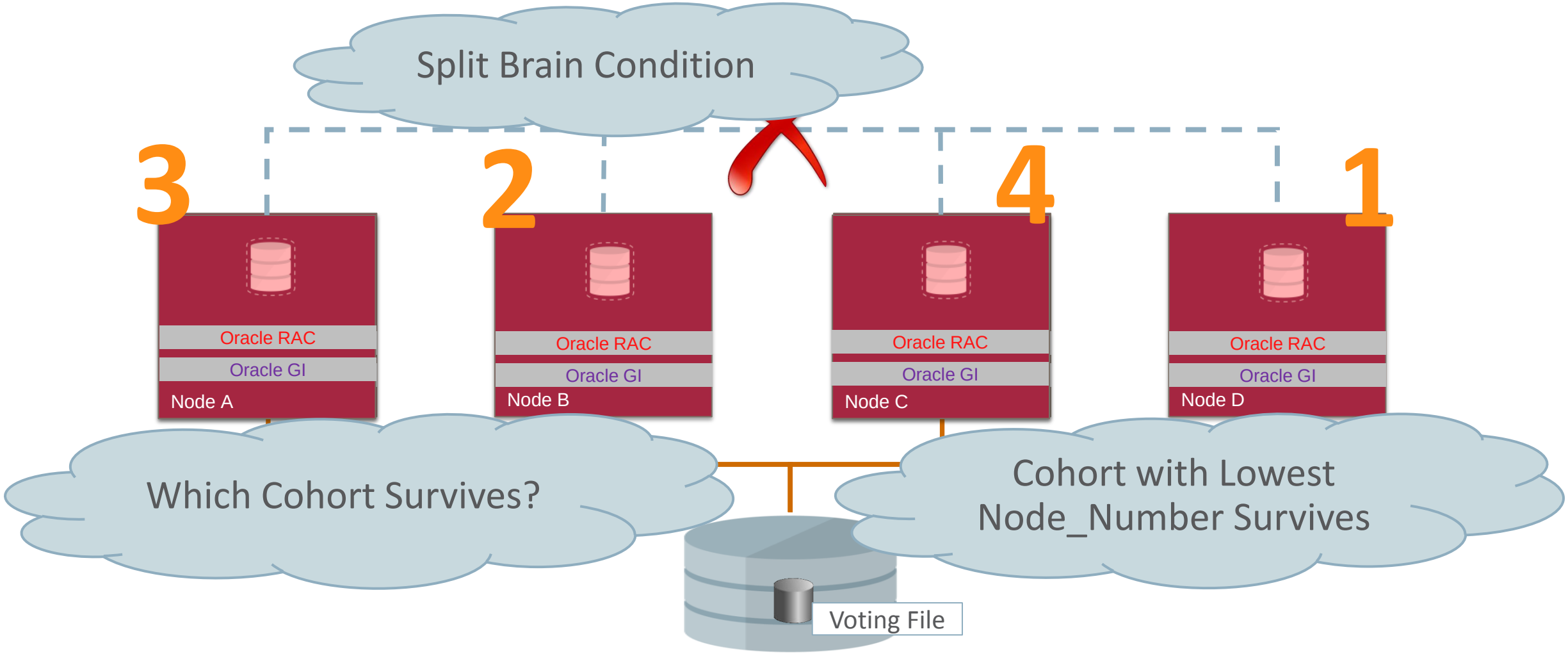
Split Brain Resolution in Two-node Clusters

Which Cluster Cohort Survives?



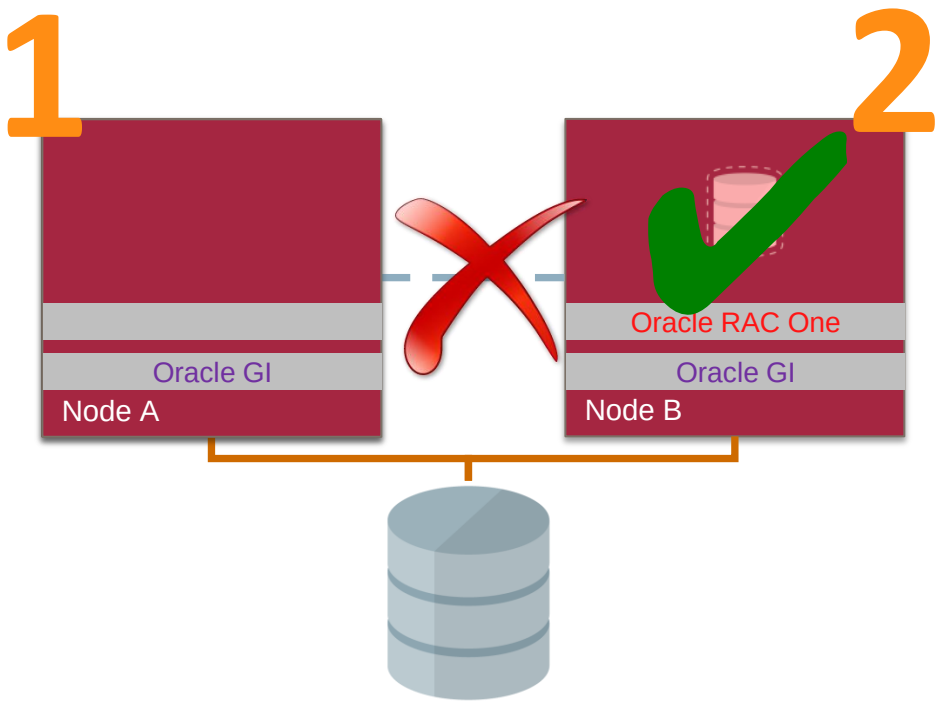
I've Got a Headache…

Split Brain Resolution in Larger Clusters



Split Brain Resolution in Oracle Clusterware 12c Rel 1

Which Cluster Cohort Survives?



I've Still Got a Headache…

Program Agenda

1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

Clusterware Concepts, Part 2

Clusterware terms:

- Cluster Resources
- Singleton Resources
- User-defined Resources
- ora* resources

Clusterware Concepts, Part 2

Clusterware terms:

- Cluster Resources
- Singleton Resources
- User-defined Resources
- ora* resources

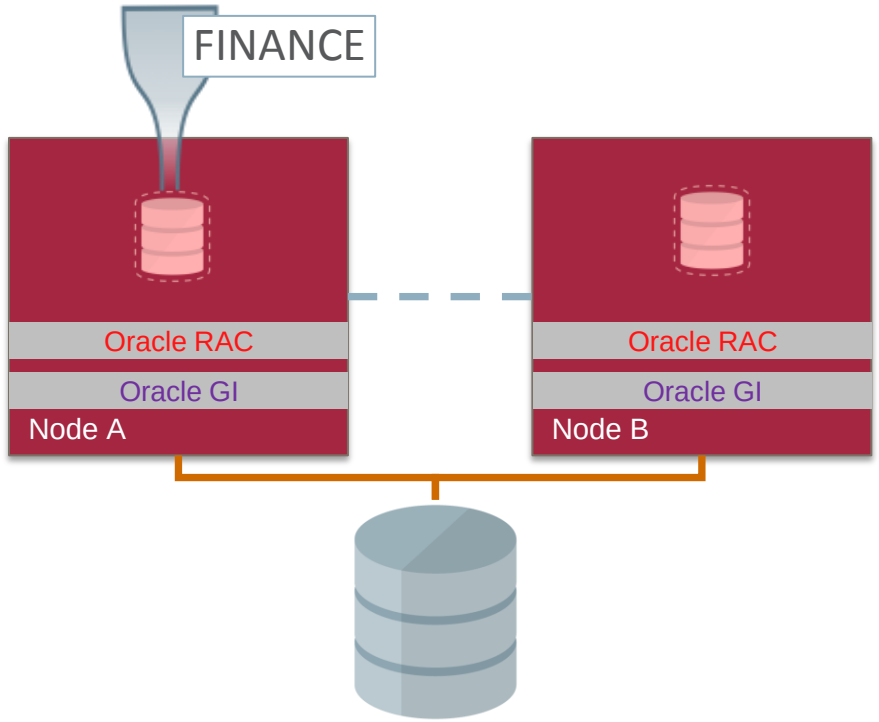
What is a Cluster Resource?

- Program, application or script
- NFS-mount
- VIP
- Service
- …

Clusterware Concepts, Part 2

Clusterware terms:

- Cluster Resources
- Singleton Resources
- User-defined Resources
- ora* resources



Clusterware Concepts – User-Defined vs Ora* Resources

[grid@Node A~] crsctl status resource

```
NAME=MYVIP
TYPE=app.appviptypex2.type
TARGET=OFFLINE
STATE=OFFLINE
```

User-Defined Resource

crsctl

```
NAME=ora.LISTENER_SCAN1.lsnr
TYPE=ora.scan_listener.type
TARGET=ONLINE
STATE=ONLINE on Node A
```

srvctl

Ora* Resource

Program Agenda

1. Split Brain – What is it?
2. Clusterware Concepts, Part 1
3. Split Brain Resolution in Current Releases
4. Clusterware Concepts, Part 2
5. Split Brain Resolution in Oracle Clusterware 12c Rel 2

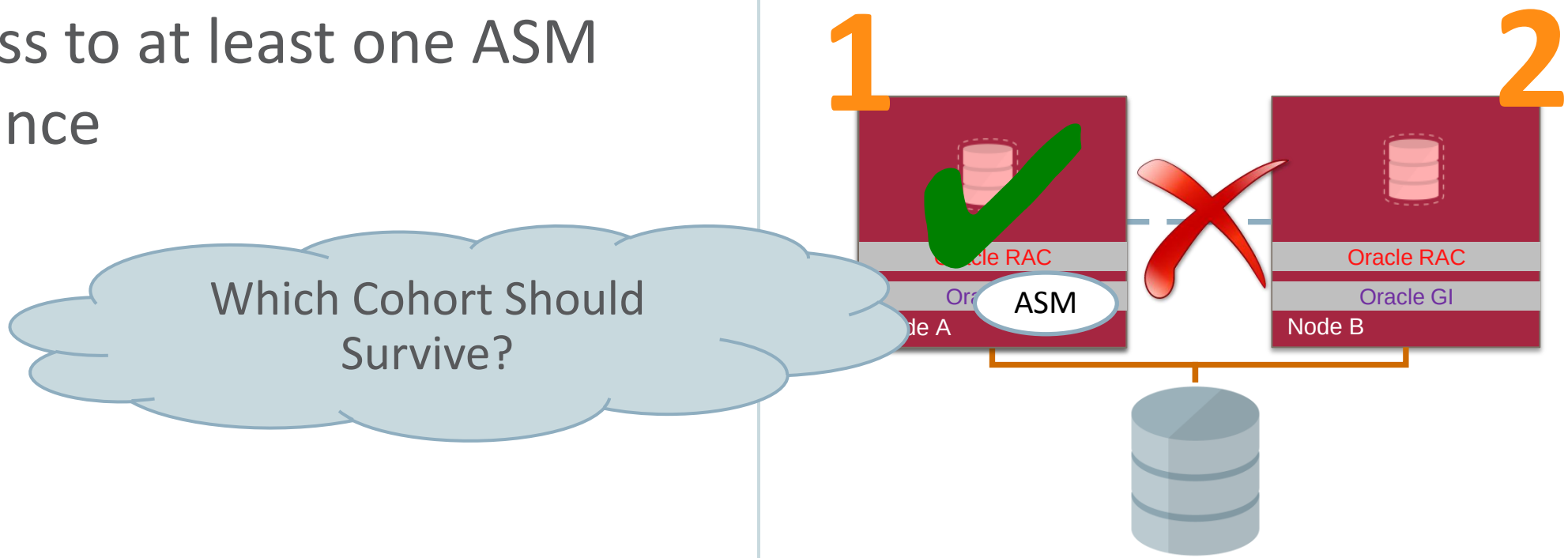
Split Brain Resolution in Oracle Clusterware 12c Rel 2
If Everything Else is Equal…

- Cohorts of equal size
- Cohorts are both viable for doing work
  - ASM instance accessible
  - Public network available

Which Cohort Should be Fenced?

Split Brain Resolution – Access to an ASM Instance

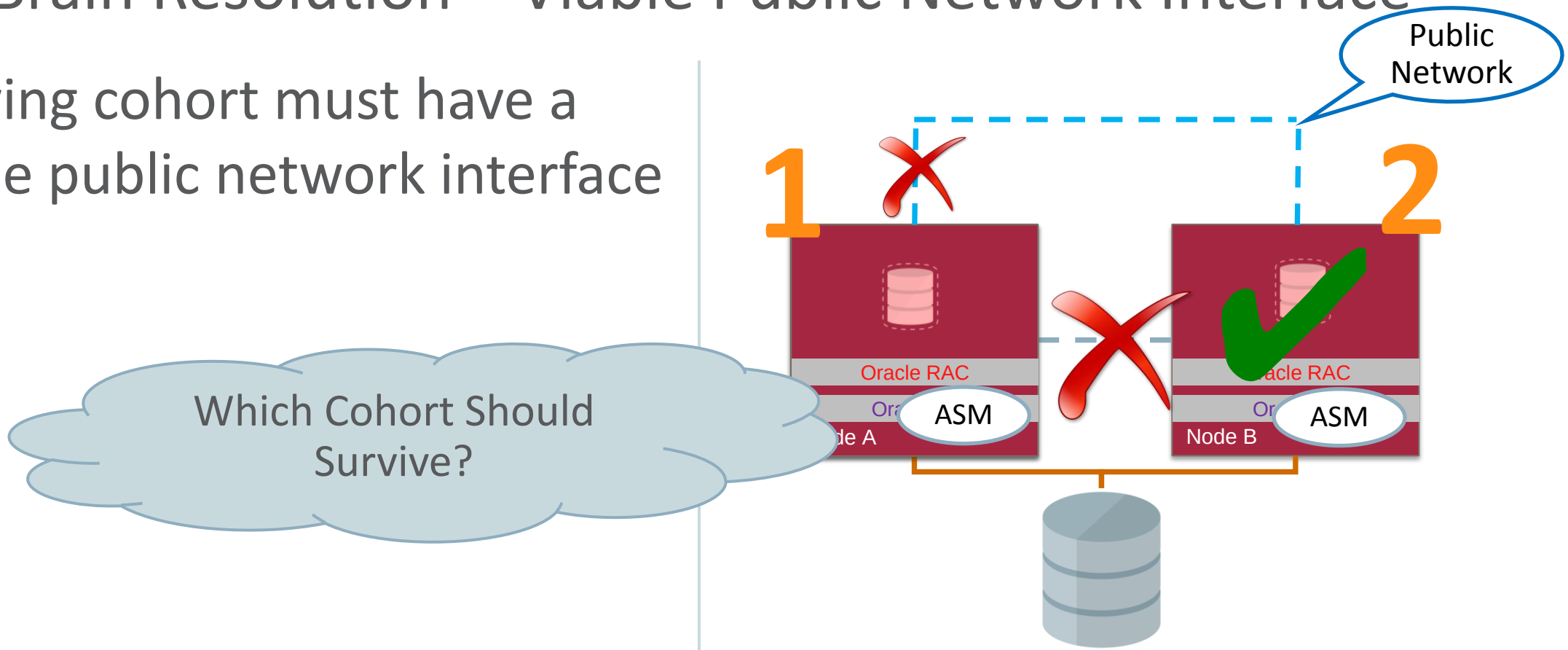
Surviving cohort must have access to at least one ASM instance



If Everything Else is Equal…

Split Brain Resolution – Viable Public Network Interface

Surviving cohort must have a viable public network interface



If Everything Else is Equal…

Split Brain Resolution in Oracle Clusterware 12c Rel 2
If Everything Else is Equal…

1. Customer can designate which server(s) and resource(s) are **critical**
2. Clusterware will evaluate cluster resources on **implied workload**
3. Cluster cohort containing the **lowest cluster node number**

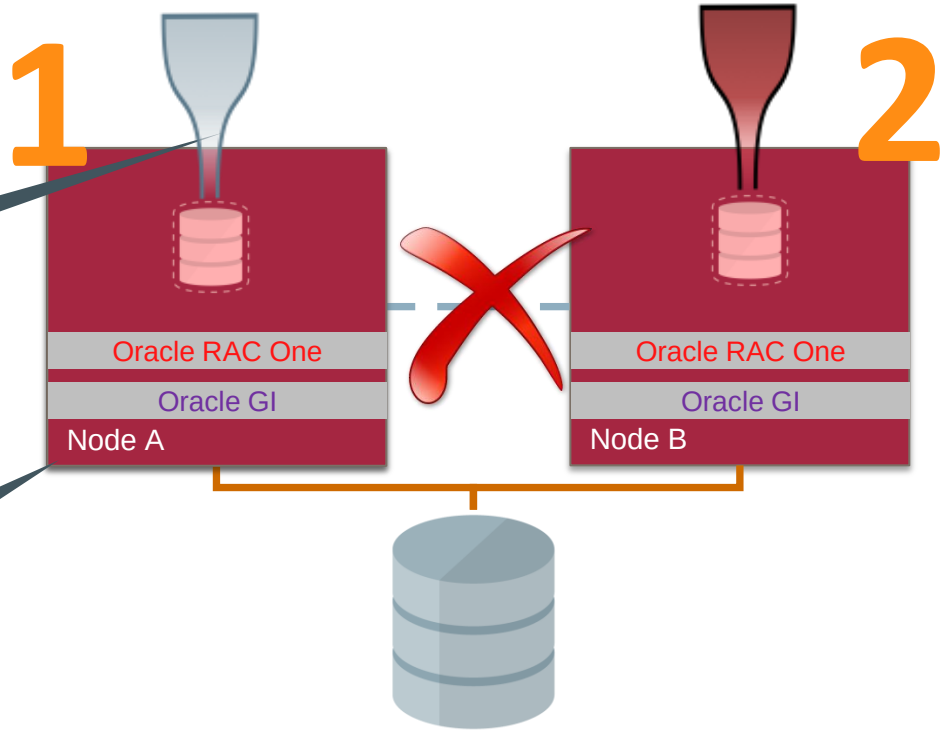
Split Brain Resolution – User Input of What is Critical

User Input for Resolving Split Brains

- css_critical
- for nodes and resources

srvctl modify service … -css_critical {YES|NO}

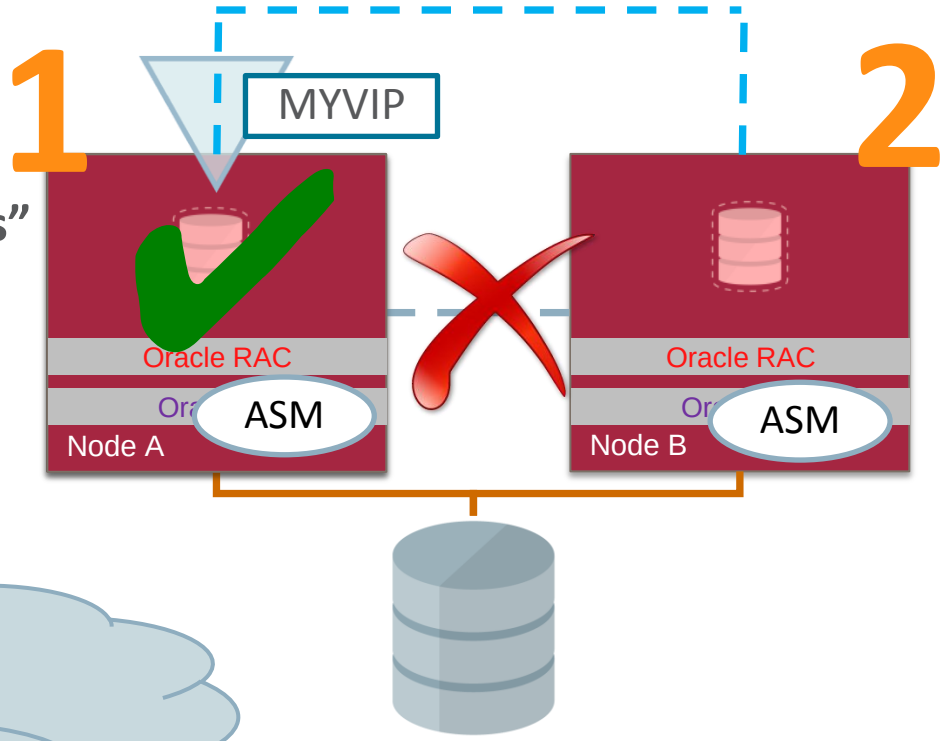
crsctl set server css_critical {YES|NO}



Split Brain Resolution – Workload Generating Resources

[grid@Node A~] crsctl modify resource MYVIP -attr
"USER_WORKLOAD=yes"

[grid@Node A~] crsctl stat res -w **"USER_WORKLOAD == yes"**

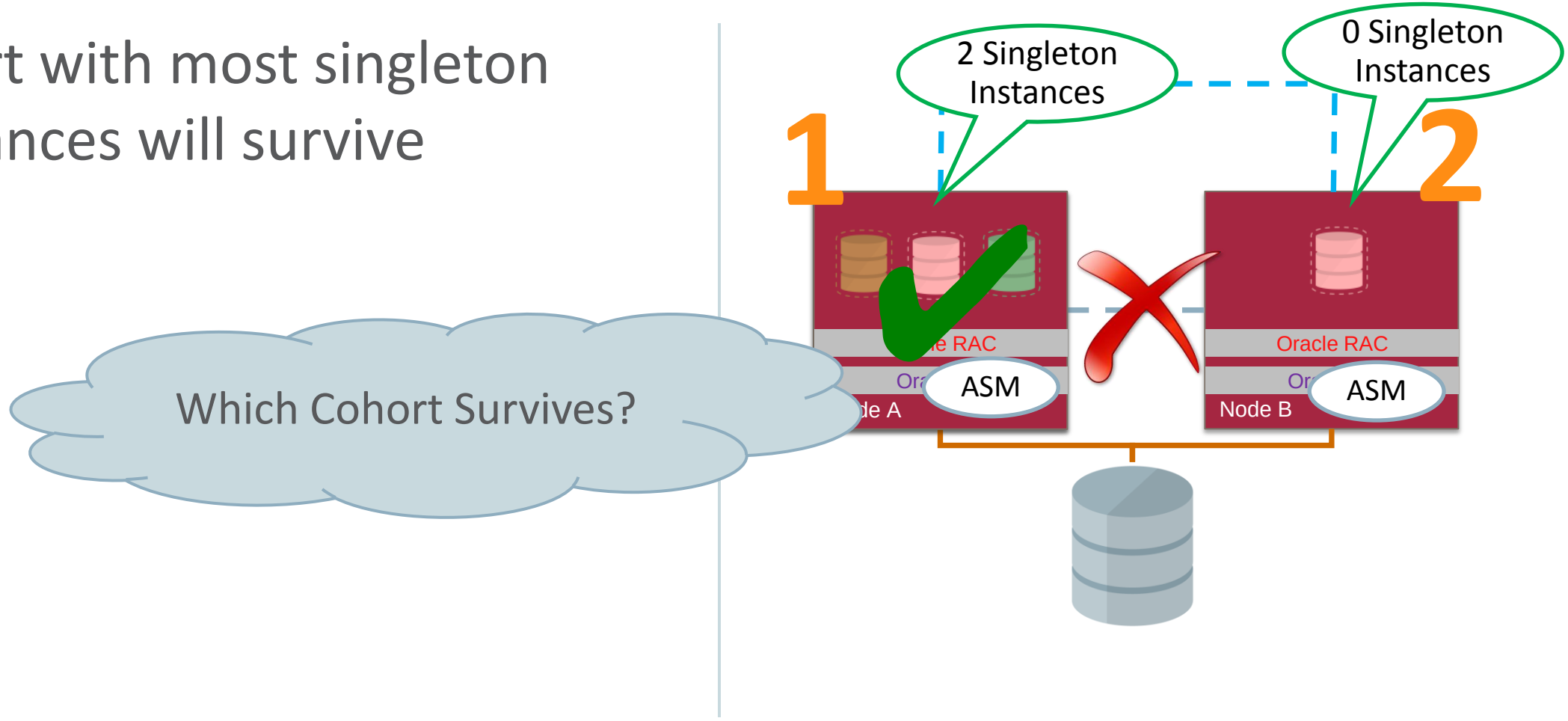


NAME=MYVIP
TYPE=app.appviptypex2.type
TARGET=OFFLINE
STATE=OFFLINE

Which Cohort Survives?

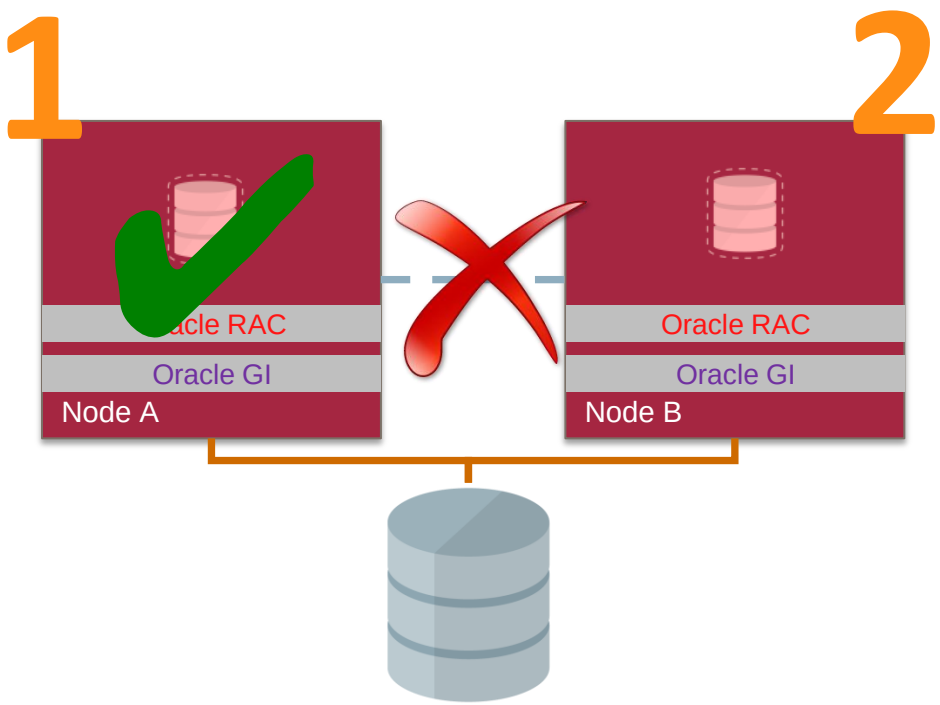
Split Brain Resolution – More Singleton Database Instances

Cohort with most singleton instances will survive



Split Brain Resolution – Default Behaviour (Node Number)

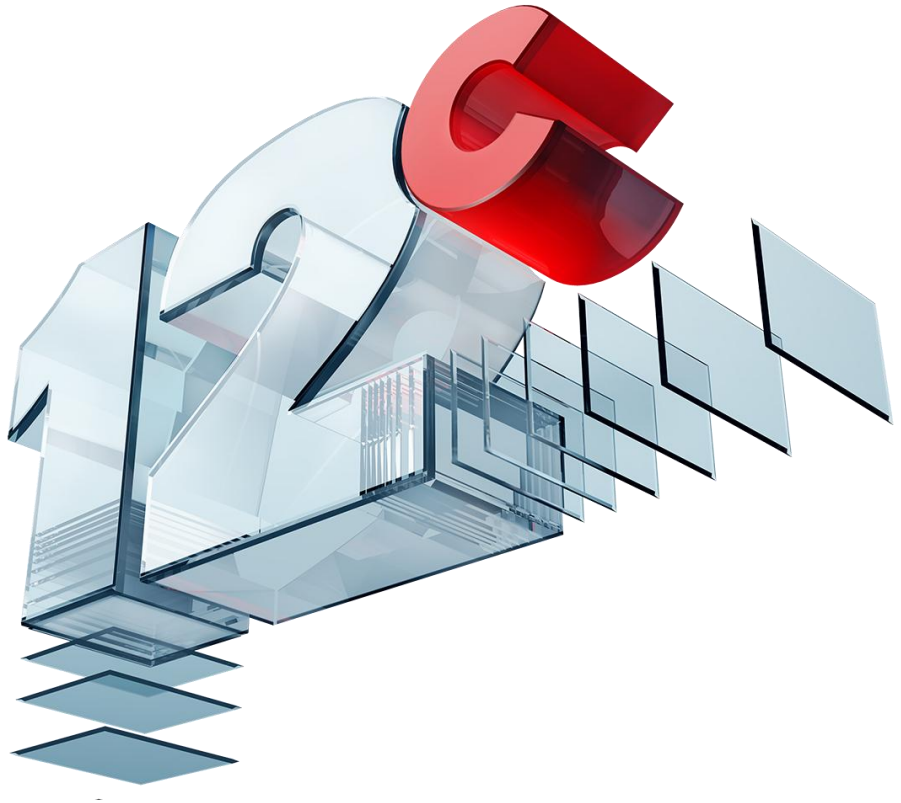
Cohort with lowest node number will survive



I've Got a Headache, again

Headache-free Split Brain Resolution

- Split Brain – What is it?
- Clusterware Concepts
- Split Brain Resolution
  - Prior to Oracle Clusterware 12c
  - Oracle Clusterware 12c Rel 1
  - Oracle Clusterware 12c Rel 2




ORACLE®