

# IBM MQ V9.3 アップデート・セミナー

---

## 第2章：新機能・変更点

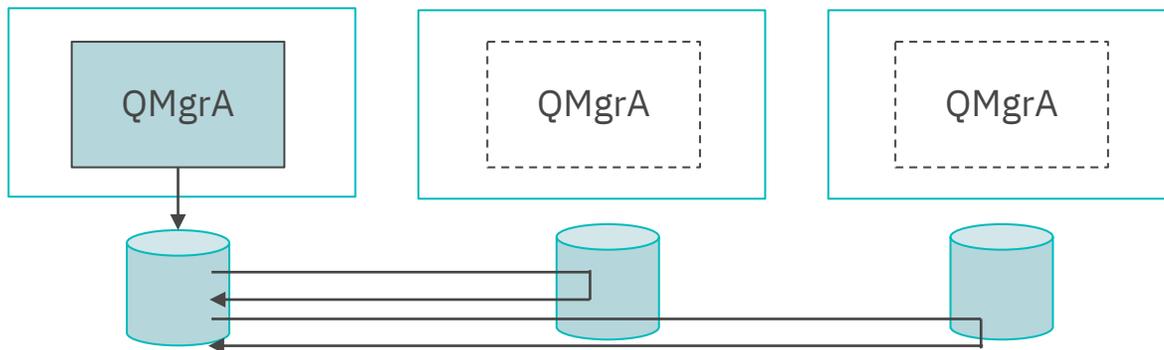
### 4. Replicated Data Queue Manager (RDQM)

## ■ RDQM(Replicated Data Queue Manager)

- ◆ DISKレプリカによるキュー・マネージャーの高可用性ソリューション
  - MQオブジェクトの作成/削除/属性更新、PUT/GETなどのDISK更新を伴う操作がすべて他ノードにレプリカされる
- ◆ HA RDQM(高可用性)とDR RDQM(災対)をサポート
- ◆ Linuxでのみ提供(LinuxのDRBDを使って実装)
- ◆ Advancedライセンスが必要

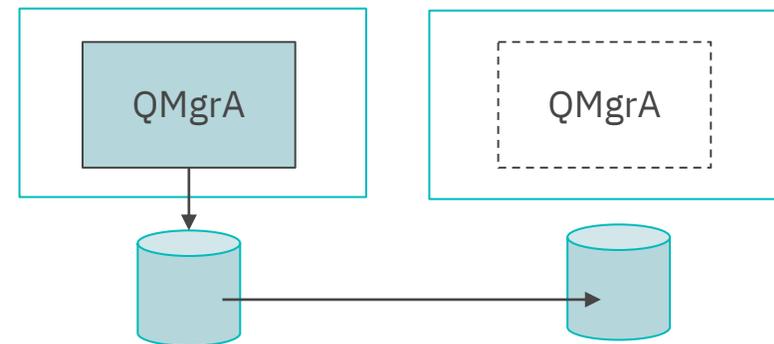
## ■ HA RDQM

- ◆ 3ノードで構成(2ノード、4ノードは不可)
- ◆ DISKは同期レプリカ
- ◆ ノード障害時、キュー・マネージャーは自動的にテークオーバー
  - Pacemakerがキュー・マネージャーの監視とテークオーバーを行う
  - 仮想IPの引き継ぎもサポート



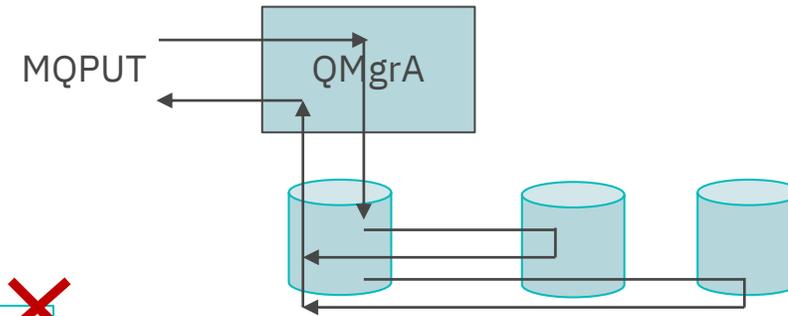
## ■ DR RDQM

- ◆ 2ノードで構成
- ◆ DISKは非同期レプリカ
  - ※オプションで同期レプリカも可能
- ◆ 災対側への切り替えは手動

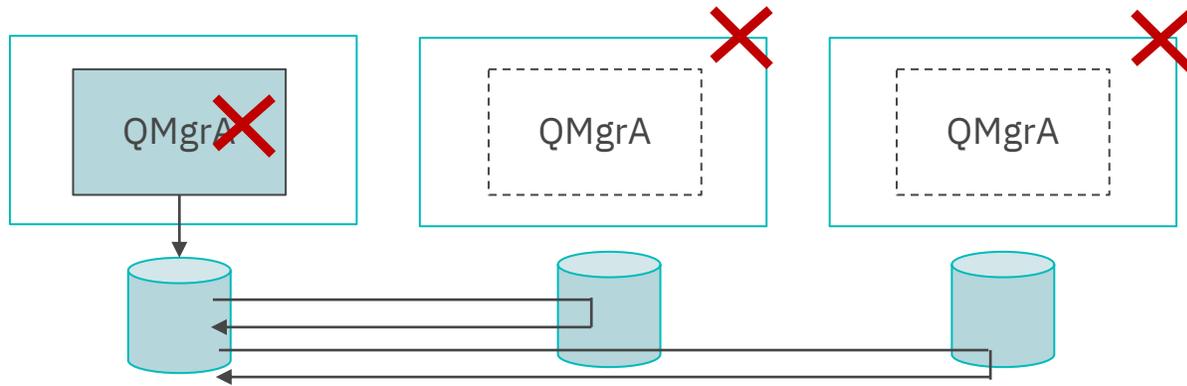


# HA RDQMの補足

- MQIの完了がレポートDISKの書き込み後  
→ 共有DISKでのHAよりもパフォーマンスは落ちる

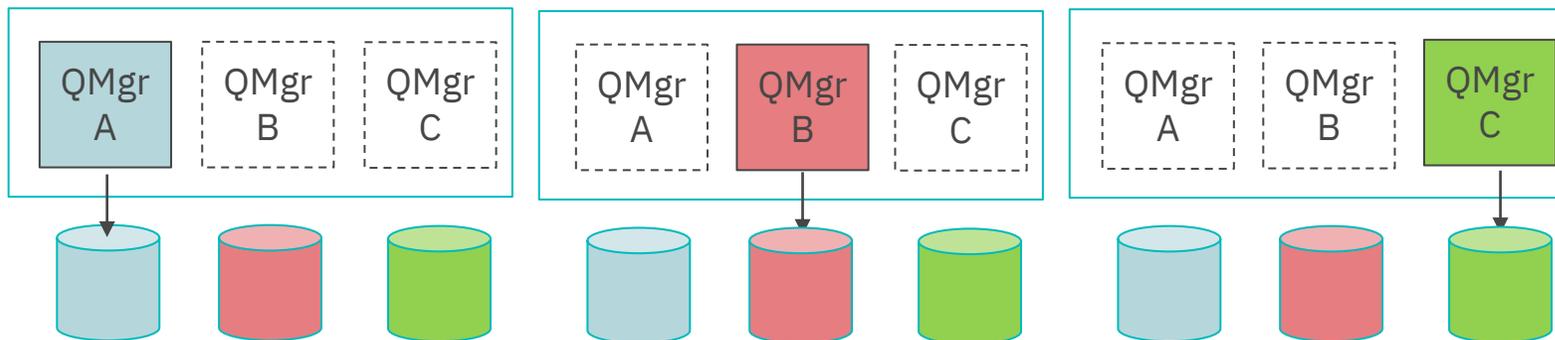


- 2ノード・ダウンでキー・マネージャー停止



- Active-Active構成をサポート

◆ 各々のキー・マネージャーに優先ノードを設定可能

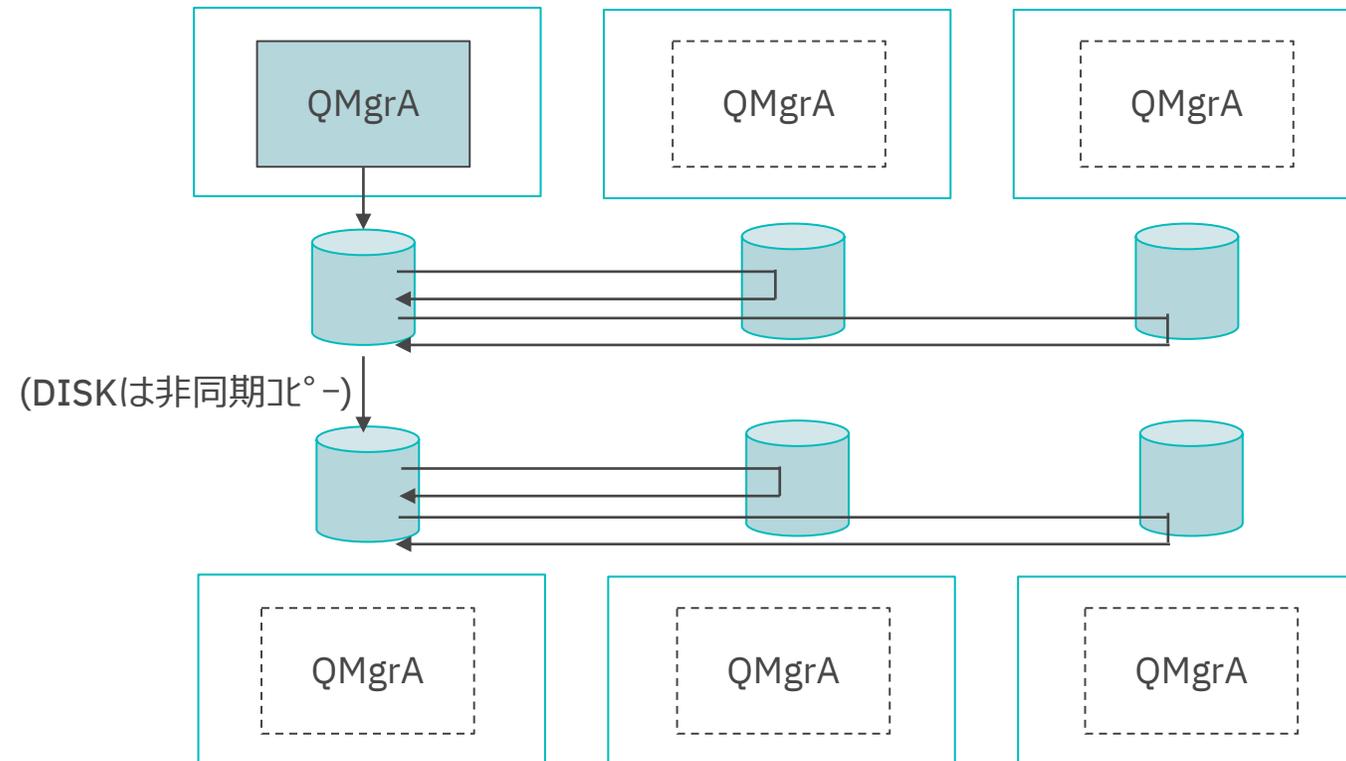


# DR RDQMの補足

## ■ HA RDQMと組み合わせ可能

◆ DR/HA RDQM構成

◆ DR/HA RDQM構成では、DRのDISKは非同期レポ-のみ(同期レポ-はできない)



### ■ 導入準備 ※各ノードで実施

#### ◆ mqm1-ユーザー、mqmグループの作成

- 全ノードでUID/GIDが一致している必要がある
- MQ導入時に自動作成されるため、導入前に同じUID/GIDを指定して作成しておく

### ■ 導入 ※各ノードで実施

#### ◆ 通常のMQファイルセット(MQSeriesRuntime-\*.rpm、MQSeriesServer-\*.rpm、...)に加えて(A)～(D)を導入

※(B)、(C)、(D)はOSバージョンに合うものを導入

(A) <MQ導入ディレクトリ>/Advanced/RDQM/MQSeriesRDQM-\*.rpm

- RHEL7 ... <MQ導入ディレクトリ>/Advanced/RDQM/PreReqs/el7/
- RHEL8 ... <MQ導入ディレクトリ>/Advanced/RDQM/PreReqs/el8/

(B) pacemaker-1/\*.rpm

(C) drbd-utils-9/\*.rpm

(D) kmod-drbd-9/xxx.rpm

※(B)、(C)はディレクトリ配下のrpmをすべて導入

※(D)は、OSカールのバージョンに一致するrpmを導入

(rpmと一緒に置いてあるmodverコマンドが表示したrpmを導入)

※「(B)、(C)、(D)、MQ→(A)」の順序で導入

```
[kmod-drbd-9]# ls
kmod-drbd-9.1.7+ptf.1_3.10.0_1062-1.x86_64.rpm
kmod-drbd-9.1.7+ptf.1_3.10.0_1127-1.x86_64.rpm
kmod-drbd-9.1.7+ptf.1_3.10.0_1160-1.x86_64.rpm
kmod-drbd-9.1.7+ptf.1_3.10.0_957-1.x86_64.rpm
modver
[kmod-drbd-9]# ./modver
kmod-drbd-9.1.7+ptf.1_3.10.0_1160-1.x86_64.rpm
```

### ■ 構成準備 ※各ノードで実施

- ◆ drbdpoolの名前でホリウム・グループを作成
  - ※当ホリウム・グループからキーマネージャごとに論理ホリウムが切り出される
- ◆ IPの作成
  - Pacemaker通信用(基本、代替) <← DR RDQMでは不要
  - データ複製通信用(HA、DR) <← DR/HA RDQMでは2つ必要
  - ※最大でHA RDQMは3つ、DR RDQMは1つ、DR/HA RDQMは4つのIPを使用
  - ※同一のIPを複数の用途で使用可能
  - ※耐障害性を上げるためには、各IPが別個のNICを使用するように作成
- ◆ ホスト名の設定
  - ※「uname -n」で値が返却される必要がある
- ◆ F/Wの穴開け
  - Pacemaker通信用 UDP : 5404-5407
  - HAデータ複製通信用 TCP : 7000-7100
  - DRデータ複製通信用 TCP : 任意<キーマネージャ作成時にポート番号を指定>
- ◆ SE-LinuxをEnforcingに設定している場合は、下記コマンドを実行
  - semanage permissive -a drbd\_t

### ■ 構成準備の補足

◆ MQ管理ユーザーは、mqmグループに加えてhaclientグループへの所属が必要

※haclientグループは、Pacemakerの導入時に自動作成される

◆ 下記のコマンドの実行には、root権限が必要. MQ管理ユーザーで実行する場合は、sudoの利用許可 (+ haclientグループへの所属) が必要

- crtmqm
- dltmqm
- rdqmadm -c/-u

## ■ HA RDQM構成

### ◆ HAグループ (Pacemakerクラスター)の作成 ※各ノードで実施

- /var/mqm/rdqm.iniを編集  
※3ノード分のIP情報を設定(3ノードで同じ内容)
- 下記コマンドを実行  
rdqmadm -c

サンプル1

```
Node:
  HA_Replication=192.168.6.1
Node:
  HA_Replication=192.168.6.2
Node:
  HA_Replication=192.168.6.3
```

サンプル2

```
Node:
  Name=node1
  HA_Primary=192.168.4.1
  HA_Alternate=192.168.5.1
  HA_Replication=192.168.6.1
Node:
  Name=node2
  HA_Primary=192.168.4.2
  HA_Alternate=192.168.5.2
  HA_Replication=192.168.6.2
Node:
  Name=node3
  HA_Primary=192.168.4.3
  HA_Alternate=192.168.5.3
  HA_Replication=192.168.6.3
```

| パラメータ          | 説明   |
|----------------|--|
| Name           | ホスト名(uname -nの値)<br>※RHEL7ではオプション                    |
| HA_Primary     | Pacemaker通信用(基本)のIP<br>※設定なしの場合、HA_ReplicationのIPを使用 |
| HA_Alternative | Pacemaker通信用(代替)のIP<br>※設定なしの場合、HA_PrimaryのIPを使用     |
| HA_Replication | データ複製通信用のIP<br>※必須                                   |

### ◆ キー・マネージャの作成

- いずれかの2ノード

`crtmqm -sxs -fs <ディスク・サイズ> <キー・マネージャ名>`

- 残りの1ノード

`crtmqm -sx -fs <ディスク・サイズ> <キー・マネージャ名>`

※-ld、-mdは指定不可(その他のcrtmqmのオプションは指定可能)

※ディスク・サイズ、キー・マネージャ名は同じ値

※ディスク・サイズはキー・マネージャに割り当てるサイズ(ホリウム・グループから切り出す論理ホリウムのサイズ)

### ◆ 優先ノードの設定

`rdqmadm -p -m <キー・マネージャ名> -n <ホスト名1>[,<ホスト名2>]`

※いずれかのノードで設定

※ホスト名1が第1優先ノード、ホスト名2が第2優先ノード

※キー・マネージャ作成直後は、「-sx」でキー・マネージャを作成したノードが第1優先ノード、第2優先ノードはなし

※第2優先ノードの設定はオプションだが、設定を推奨

### ◆ 仮想IPの作成(オプション)

`rdqmint -m <キー・マネージャ名> -a -f <IPアドレス> -f <物理インターフェイス名>`

### ■ DR RDQM構成

#### ◆ キー・マネージャの作成

- 通常側

```
crtmqm -rr p -rl <通常ノドのIP> -ri <災対ノドのIP>  
-rn <災対ノドのホスト名> -rp <災対ノドのポート番号> -fs <ディスク・サイズ> <キー・マネージャ名>
```

- 災対側

```
crtmqm -rr s -rl <災対ノドのIP> -ri <通常ノドのIP>  
-rn <通常ノドのホスト名> -rp <通常ノドのポート番号> -fs <ディスク・サイズ> <キー・マネージャ名>
```

※「-rt s」を追加指定すると、DISKが同期モードになる

※「-ld、-mdは指定不可」などはHA RDQMと同じ

## ■ DR/HA RDQM構成

### ◆ HAグループ° (Pacemakerクラスター)の作成

- 通常側と災対側のそれぞれで作成 (手順は、HA RDQMと同じ)

### ◆ キー・マネージャーの作成

DR RDQMのキー・マネージャー作成コマンド + -sxs/-sx(HA RDQMのキー・マネージャー作成オプション)

いずれかの2ノードは-sxs、残りの1ノードは-sxを指定

- 通常側

crtmqm **-rr p** -rl <通常ノードのIP> -ri <災対ノードのIP>  
-rp <災対ノードのポート番号> -fs <ディスクサイズ> <QMgr名>

- 災対側

crtmqm **-rr s** -rl <災対ノードのIP> -ri <通常ノードのIP>  
-rp <通常ノードのポート番号> -fs <ディスクサイズ> <QMgr名>

-rnは指定しない

※「-ld、-mdの指定不可」などはHA RDQMと同じ

カマ区切りで3ノードのIPを設定

### ◆ 優先ノード°の設定、仮想IPの作成(オプション)

- 通常側/災対側のそれぞれで実施(手順は、HA RDQMと同じ)

## ■ HA RDQM構成の補足

### ◆ HA RDQMを1つのノードから構成する方法

- MQ管理1-ザ-を他のノードにパスワードなしでSSH接続できるように設定
- 下記をいずれかのノードで実行
  - rdqm.iniを編集
  - rdqmadm -c
  - crtmqm **-sx** -fs <ディスクサイズ> <キー・マネージャ名>

## ■ DR/HA RDQM構成の補足

### ◆ DR/HA RDQMを構成する別の方法

- rdqm.iniにデータ複製通信用(DR)のIPを設定
  - 若干、crtmqmの引数が簡単になる
    - ・ -rlの指定が不要
    - ・ -riの代わりに-rnでDRGroup名を指定

自分側のIP  
(crtmqmの-rlに指定するIP)

相手側のIP  
(crtmqmの-riに指定するIP)

```
Node:
Name=node1
HA_Primary=192.168.4.1
HA_Alternate=192.168.5.1
HA_Replication=192.168.6.1
DR_Replication=192.168.1.1
Node:
Name=node2
HA_Primary=192.168.4.2
HA_Alternate=192.168.5.2
HA_Replication=192.168.6.2
DR_Replication=192.168.1.2
Node:
Name=node3
HA_Primary=192.168.4.3
HA_Alternate=192.168.5.3
HA_Replication=192.168.6.3
DR_Replication=192.168.1.1
DRGroup:
Name=DR1
DR_Replication=192.168.2.1
DR_Replication=192.168.2.2
DR_Replication=192.168.2.3
```

## ■ HA RDQMの運用コマンド

◆ RDQMのキー・マネージャーには1次、2次の区別がある。キー・マネージャーの起動は、1次に対して行う

- 2次は起動できない(strmqmコマンドにエラーが返る)
- 第1優先ノードのキー・マネージャーが1次。残りの2ノードのキー・マネージャーは2次
- 第2優先ノードを設定していれば、第1優先ノードが停止しているとき、第2優先ノードのキー・マネージャーが1次になる(第2優先ノードを設定していないと、どちらのノードのキー・マネージャーが1次になるかは不定)

◆ ノードのサスペンド

`rdqmadm -s [-n <ノード名>]`

- 該当ノードで動いているキー・マネージャーを他ノードに移動(手動テークオーバー)
- 一時点でサスペンドにできるノードは1つだけ(2つ目のノードのサスペンドはエラー)
- -nオプションで他ノードをサスペンドにできる
- 下記で元の状態に戻る

`rdqmadm -r [-n <ノード名>]`

### ◆ リソースの実行抑止の解除

- `rdqmclean -m <キー・マネージャー名>`  
特定のキー・マネージャーに関連した「failed resource actions」を削除
- `rdqmclean -a`  
すべての「failed resource actions」を削除
- 「failed resource actions」  
Pacemakerがリソースの実行に失敗すると、該当ノードでのリソース実行を抑止するために設定
- 障害原因を取り除いた後にコマンドを実行

## ■ DR RDQMの運用コマンド

### ◆ キー・マネージャの1次への格上げ

`rdqmdr -p <キー・マネージャ名>`

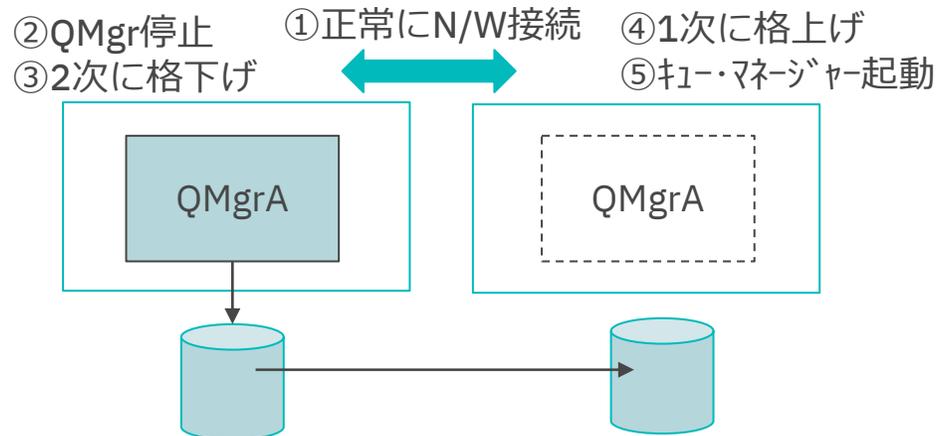
- 2次のキー・マネージャは起動できない  
→ 災対側のキー・マネージャは2次. 災害発生時は、1次に格上してからキー・マネージャを起動
- 通信可能な状態で、相手側でキー・マネージャが1次になっていると、1次に格上げできない.  
通信不可の状態ならば、(強制的に)1次に格上げ可能

### ◆ キー・マネージャの2次への格下げ

`rdqmdr -s <キー・マネージャ名>`

※1次/2次の変更は、キー・マネージャを停止して実施

参考. (制御された) 災対側への切り替え手順

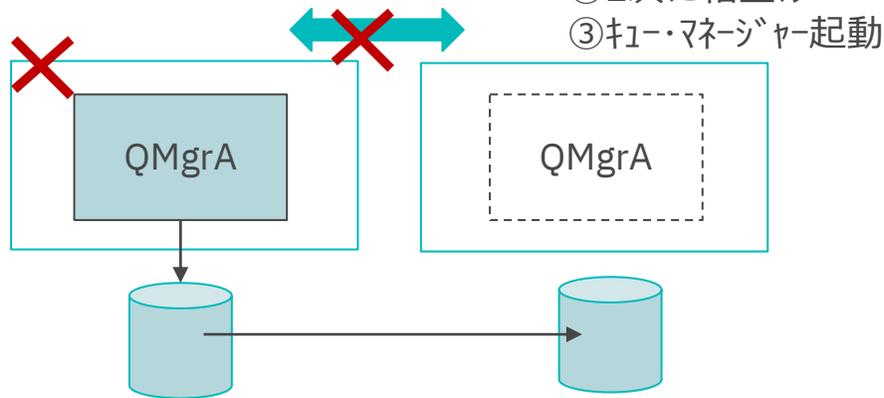


## ◆ 区分データ状態の解消

- 1時点で通常側/災対側の両方のキー・マネージャーが1次になると、区分データ状態になる  
 例えば、下図のような手順で災害訓練を行うと区分データ状態になる  
 ※災対側のキー・マネージャーが1次になっている間、通常側のキー・マネージャーは停止しているが、  
 1時点で通常側/災対側の両方のキー・マネージャーが1次になったと判断される  
 ※訓練終了後にN/W接続を回復すると、区分データ状態になる
- 区分データ状態の解消には、1次/2次を再設定  
 2次側のキー・マネージャーDISKが廃棄され、1次側のDISKをフル・コピー

### 参考. 災害訓練

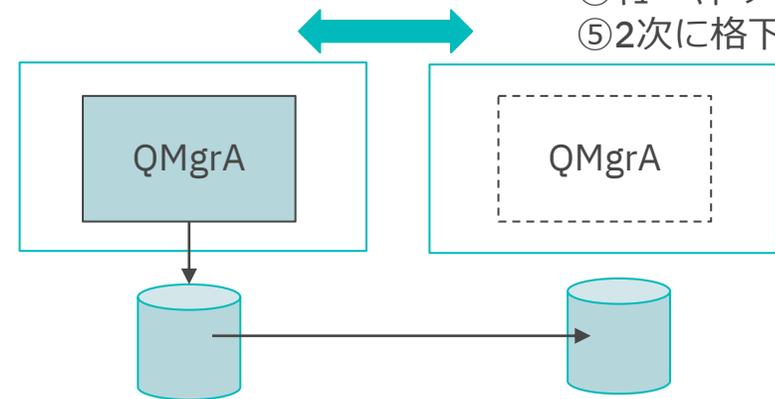
①ノード停止(N/W接続不可)



②1次に格上げ  
③キー・マネージャー起動

訓練終了後

⑥ノード起動(N/W接続の回復)



④キー・マネージャー停止  
⑤2次に格下げ

N/W接続が回復すると、両ノードのキー・マネージャーが区分データ状態になる

### ■ 稼働状況/構成情報の表示

#### ◆ rdqmstatus -n

- ノードの稼働状況

```
Node node1 is online
Node node2 is online
Node node3 is online
```

#### ◆ rdqmstatus(オプションなし)

- 該当ノードのDRBD構成情報
  - カーネルのバージョン(OS、DRBD)
  - DRDBのバージョン
  - DRDBカーネルの状態
- 構成されているキーマネージャーの一覧
  - キーマネージャーのHA/DR構成情報と稼働状況

```
Node: node1
OS kernel version: 3.10.0-1160.15.2
DRBD OS kernel version: 3.10.0-1160
DRBD version: 9.1.7+ptf.1
DRBD kernel module status: Loaded

Queue manager name: RDQM01
Queue manager status: Running
HA status: Normal
HA current location: This node
HA preferred location: This node,node2
HA blocked location: None

Queue manager name: RDQM02
Queue manager status: Running
DR role: Primary
DR status: Normal
```

## ◆ rdqmstatus -m <キー・マネージャー名>

- 該当キー・マネージャーの詳細なHA/DR構成情報と稼働状況
- 他ノードの稼働状況とDISKの同期状況

```
Node: node1
Queue manager status: Running
CPU: 0.00%
Memory: 180MB
Queue manager file system: 50MB used, 0.2GB allocated [42%]
HA role: Primary
HA status: Mixed
HA control: Enabled
HA current location: This node
HA preferred location: This node,node2
HA blocked location: None
HA floating IP interface: None
HA floating IP address: None
```

```
Node: node2
HA status: Remote unavailable
HA out of sync data: 628KB
HA last in sync: 2022-08-09 15:31:06
```

```
Node: node3
HA status: Normal
```

```
Node: node1
Queue manager status: Running
CPU: 0.00%
Memory: 180MB
Queue manager file system: 50MB used, 0.2GB allocated [42%]
DR role: Primary
DR status: Remote unavailable
DR type: Asynchronous
DR port: 8000
DR local IP address: 192.168.56.104
DR remote IP address: 192.168.56.105
DR out of sync data: 560KB
DR last in sync: 2022-08-09 15:30:58
```

- ◆ `rdqmstatus -m <キー・マネージャ名> -a`
  - 「`rdqmstatus -m <キー・マネージャ名>`」の情報  
+ 該当QMgrの「failed resource actions」

```
Node: node1
Queue manager status: Running
CPU: 0.00%
Memory: 180MB
Queue manager file system: 50MB used, 0.2GB allocated [42%]
HA role: Primary
HA status: Mixed
HA control: Enabled
HA current location: This node
HA preferred location: This node,node2
HA blocked location: node2
HA floating IP interface: None
HA floating IP address: None

:
Node: node3
HA status: Normal

Failed resource action: Start
Resource type: Queue manager
Failure node: node2
Failure time: 2022-08-09 15:25:52
Failure reason: Not installed
Blocked location: node2
```

### ■ rdqmstatusコマンドの拡張(表示項目の追加)

- ◆ カーネルのバージョン(OS、DRBD)、DRDBのバージョン、DRDBカーネルの状態
  - オプションなしのrdqmstatusに表示
  - OSカーネルとDRDBカーネルのバージョンが一致していない場合は、修正要
- ◆ ブロック・チェック
  - 「failed resource actions」のためにリソースの実行が抑止されているノードの一覧
  - オプションなし、-mオプションのrdqmstatusの両方に表示
- ◆ 最後にDISKの同期を行った日時
  - -mオプションのrdqmstatusに表示
  - DISKが同期できないときにのみ表示
- ◆ 「failed resource actions」の一覧
  - -aオプションのrdqmstatusに表示
  - 「rdqmstatus -m <キー・マネージャー名> -a」は、MQv9.3で追加
- ◆ -mオプションのrdqmstatusに表示していた一部の項目をオプションなしのrdqmstatusにも表示
  - 優先ノード、HAステータス、DRステータス

### ■ rdqmcleanコマンドの提供

- ◆ 「failed resource actions」を削除し、リソースの実行抑止を解除