

# CUDB System Split Partial Recovery Procedure

## OPERATION INSTRUCTION

**Copyright**

© Ericsson AB 2016–2018. All rights reserved. No part of this document may be reproduced in any form without the written permission of the copyright owner.

**Disclaimer**

The contents of this document are subject to revision without notice due to continued progress in methodology, design and manufacturing. Ericsson shall have no liability for any error or damage of any kind resulting from the use of this document.

**Trademark List**

All trademarks mentioned herein are the property of their respective owners. These are shown in the document Trademark Information.



# Contents

<b>1</b>	<b>Overview</b>	<b>1</b>
1.1	Description	1
1.2	Prerequisites	1
1.3	Typographic Conventions	3
<b>2</b>	<b>Emergency Procedure</b>	<b>4</b>
2.1	Enabling Provisioning	4
2.2	Enabling Traffic	6
<b>3</b>	<b>Recovery Procedure</b>	<b>7</b>
	<b>Glossary</b>	<b>9</b>
	<b>Reference List</b>	<b>10</b>





# 1 Overview

The purpose of this document is to provide system administrators clear operating instructions for recovering from a site failure or IP network loss, and to make traffic or provisioning available until the CUDB site or IP network becomes operational again.

## 1.1 Description

This document describes the Partial Recovery procedure that can be used as an emergency work-around solution for CUDB system deployments in the case of a system split.

## 1.2 Prerequisites

The procedure can be applied in the following situations:

- Symmetrical split situation.
  - Majority partition is split in half by simultaneous site failure. A site is considered down if all the nodes in the site are down. If surviving partition is not hosting the master instance for PLDB, update operations from any Provisioning user are rejected. For more information, refer to [CUDB High Availability, Reference \[1\]](#).
  - Majority partition is split in half by IP backbone failure. One of the partitions is not hosting the master instance for PLDB. Update operations from any Provisioning user are rejected in that partition. Apply procedure to recover provisioning traffic through that specific partition.
- Minority situation.
  - More than half of the sites in majority partition fail simultaneously leaving the rest of the sites in minority partition.
  - IP backbone failure simultaneously splits majority partition in multiple partitions that are less than half of the size of the original partition.
  - The system is in symmetrical split, both partitions are further separated by a new failure event, resulting in all partitions being in minority.
  - IP backbone failure leaves one partition in majority and one or more partitions in minority. If all sites in majority partition fail, there are only minority partitions left. If IP backbone connection is reestablished partitions with minority status remain in minority.

**Note:** Do not execute the procedures of this document in case of a majority situation.



To detect the situations detailed above, execute following command on one node of each live partition:

```
cudbSetPartitionStatus --printPartitionStatus
```

Also check the following alarms in the CUDB system:

— Minority situation:

The [Control, Remote Site Unreachable](#), Reference [2], alarms are raised, one for each unreachable site. These alarms are raised on every site within each surviving partition on the node where the System Monitor (SM) leader is located.

— Symmetrical split situation:

The [Control, Remote Site Unreachable](#), Reference [2], and [Control, Potential Split Brain Detected](#), Reference [3] alarms are raised on every site within each surviving partition on the node where the System Monitor (SM) leader is located.

For more information, refer to [Control, Remote Site Unreachable](#), Reference [2], and [Control, Potential Split Brain Detected](#), Reference [3].

In any other situation, the system is in majority situation, and the Partial Recovery procedure cannot be applied.

---

---

## Attention!

The system must satisfy all the conditions listed above: in case of any differences in the setup, the instructions of this document are not applicable.

---

---

As a result of this procedure, the PLDB and DSG masterships are moved to a surviving CUDB partition which has been selected as Service Partition providing traffic and provisioning.

The partial recovery procedure consists of two procedures:

- An Emergency Procedure, which must be applied when the symmetrical split or minority situation occurs.
- A Recovery Procedure, which must be applied when the split situation ends, so that all sites are operational.

The Selective Replica Check and Data Repair processes are part of the Automatic Handling of Network Isolation function. The aim of this automatic process in CUDB is to attempt to handle and repair data loss that happened due to network split or unexpected PLDB or DSG mastership change. For more information on the



Selective Replica Check and Data Repair processes, refer to CUDB Data Storage Handling, Reference [5].

## 1.3      Typographic Conventions

Typographic conventions can be found in the following document:

— Typographic Conventions



## 2 Emergency Procedure

This section describes the emergency procedure for provisioning recovery and traffic recovery. To enable provisioning, execute procedure from Section 2.1 on page 4, and to enable traffic execute procedure from Section 2.2 on page 6.

### 2.1 Enabling Provisioning

If isolation is caused by multiple sites failure, there is only one surviving partition where provisioning can be enabled. In this case, choose that partition to enable provisioning.

If isolation is caused by IP Backbone failure, which can be combined with site failure, there can be multiple surviving partitions. Choose the surviving partition(s) to enable provisioning.

If Selective Replica Check and Data Repair are enabled, after the backbone problem is fixed and the sites are rejoined, Automatic Handling of Network Isolation fixes data inconsistencies caused by the split and re-establishes replication in the slave replicas, so you can choose multiple surviving partitions to enable provisioning. For more information on Automatic Handling of Network Isolation, refer to [CUDB Automatic Handling of Network Isolation Output Description, Reference \[6\]](#).

If Selective Replica Check and Data Repair are disabled, disable the rest of the sites, so that they do not join the system during the execution of the procedure. The execution of this procedure on the remote nodes leads to data inconsistency and data loss without any rollback possibility. In this case, choose only one surviving partition to enable provisioning.

---

---

### Stop!

Do not select an unstable node in the surviving site. The node selected to execute the below steps in the surviving site must be free of Operating System (OS) alarms and SAF alarms.

---

---

1. For failing CUDB sites, make sure they cannot join the CUDB system in an uncontrolled manner during the execution of the procedure.



---

---

### Attention!

This preparatory step must be performed to avoid any potential conflicts and interference from the failed sites during the execution of the procedure.

---

---

2. If Selective Replica Check and Data Repair are enabled go to Step 3.

Set the PLDB and DS Units to maintenance mode in all nodes of the surviving partitions not selected as the surviving Service Partition. Use the following command to do so:

```
cudbManageStore -a -o maintenance
```

---

---

### Attention!

As the IP backbone is down, this step must be performed locally on-site by the operator (no remote configuration is available). Therefore, operators must be prepared to execute the below commands on remote nodes as well in time. Ericsson takes no responsibility for any data inconsistency if the commands in this step are not executed in time, or at all.

---

---

3. This step can only be executed if the PLDB master instance is not in the Service Partition.

Execute the **cudbTakeAllMasters** command on only one node of the selected Service Partition(s). For more information refer to [CUDB Node Commands and Parameters, Reference \[4\]](#).

---

---

### Attention!

This command cannot be undone: once it is executed, the PLDB and DSG masterships are moved to the surviving partition. Killing the command during waiting does not stop the process of taking the mastership.

---

---

After the above command has been executed, provisioning must be available.



**Note:** After successful command execution, provisioning continues. However, the system might not be geographically redundant, so the missing partitions and the system backbone must be restored immediately. The system recovers geographical redundancy after the recovery procedure is executed, as described in Section 3 on page 7.

## 2.2 Enabling Traffic

This procedure is used to allow traffic in the DSGs, and it can only be applied for partitions in minority situation.

Execution of this procedure does not prevent choosing this partition later as a Service Partition to recover the provisioning. For more information, see Section 2.1 on page 4.

If the `automaticServiceContinuity` parameter is enabled, procedure is done automatically. For more information, refer to [CUDB High Availability, Reference \[1\]](#).

---

---

### Attention!

If the IP backbone is down, this step must be performed locally on-site by the operator. No remote configuration is available.

---

---

1. Execute the `cudbServiceContinuity` command on only one node of the surviving partition. For more information, refer to [CUDB Node Commands and Parameters, Reference \[4\]](#).

After the above command has been executed, traffic must be available in all DSGs.

2. (Optional) To recover the provisioning in this partition, execute the specific emergency procedure. For more information, see Section 2.1 on page 4.



## 3 Recovery Procedure

When the IP Backbone or failed sites are working again, follow the steps below to restore the original redundancy configuration:

**Note:** If the Selective Replica Check and Data Repair are enabled, skip to Step 5.

1. For sites not recovering from site failure go to step 2. Before connecting the CUDB nodes in the failed sites to the CUDB system, set the PLDB and all the DS Units in each node to maintenance mode executing the following command in each of the nodes of the failed sites:

```
cudbManageStore -a -o maintenance
```

2. Use the following command to check which node runs the SM leader:

```
cudbSystemStatus -B
```

3. Check that the [Control, Remote Site Unreachable, Reference \[2\]](#), alarm is not raised. To do so, login to the node where the SM leader is located, and check for the alarms raised in the node with the following command:

```
cudbSystemStatus -a
```

If the [Control, Remote Site Unreachable, Reference \[2\]](#), alarm is raised, then the original 1+1 double geographical redundant or 1+1+1 triple geographical redundant CUDB system is not recovered yet. Check that the nodes in the failed sites are properly connected to the CUDB system.

To set all the PLDB and DS Units to ready, execute the following command in all the nodes that were not part of selected Service Partition:

```
cudbManageStore -a -o ready
```

4. At this point in time, the PLDB and the DS Units in the nodes of the failed sites become slaves of the master PLDB and master DS Units hosted in the surviving site and start to gather the differences from the masters. Check the replication status of the nodes by executing the following command:

```
cudbSystemStatus -R
```

5. If a slave database cluster is unable to sync with the current master, check if [Storage Engine, Unable to Synchronize Cluster in DS, Major, Reference \[7\]](#) or [Storage Engine, Unable to Synchronize Cluster in PLDB, Major, Reference \[8\]](#) is raised. In those cases, execute the following command to restore the slave DS Unit or PLDB where the replication is not running:

```
cudbUnitDataBackupAndRestore
```

For more information, refer to [CUDB Backup and Restore Procedures, Reference \[9\]](#).



**Note:** If Self-Ordered Backup and Restore function is enabled, it does not start if all PLDB replicas in the site are slaves that are unable to synchronize with the current master. In case the Automatic Handling of Network Isolation or the Self-Ordered Backup and Restore functions or both are enabled, alarms related to these functions appear before are raised before the *Storage Engine, Unable to Synchronize Cluster in DS, Major, Reference [7]* or *Storage Engine, Unable to Synchronize Cluster in PLDB, Major, Reference [8]* alarms. If Self-Ordered Backup and Restore process restores the replication automatically, these alarms are not raised at all.

Once the replication is working, if needed, move the DSG mastership to the desired node by using the **cudbDsgMastershipChange** command, taking into account if Automatic Mastership Change function is active.

For more information about the commands used in this section, refer to *CUDB Node Commands and Parameters, Reference [4]*. For more information on the Automatic Handling of Network Isolation procedure, refer to *CUDB High Availability, Reference [1]*.



## Glossary

For the terms, definitions, acronyms and abbreviations used in this document, refer to [CUDB Glossary of Terms and Acronyms](#), Reference [10].



## Reference List

### Ericsson Documents

- [1] CUDB High Availability
- [2] Control, Remote Site Unreachable
- [3] Control, Potential Split Brain Detected
- [4] CUDB Node Commands and Parameters
- [5] CUDB Data Storage Handling
- [6] CUDB Automatic Handling of Network Isolation Output Description
- [7] Storage Engine, Unable to Synchronize Cluster in DS, Major
- [8] Storage Engine, Unable to Synchronize Cluster in PLDB, Major
- [9] CUDB Backup and Restore Procedures
- [10] CUDB Glossary of Terms and Acronyms