

CUDB Performance Guide

USER GUIDE

Copyright

© Ericsson AB 2016, 2017. 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

1	Introduction	1
1.1	Purpose and Scope	1
1.2	Target Groups	1
1.3	Revision Information	1
1.4	Prerequisites	2
1.5	Typographic Conventions	2
2	Counters in CUDB	3
2.1	Overview	3
2.2	Counter Generation and Publishing	3
2.3	Configuring Counter Output Files Names	5
2.4	CUDB KPIs	5
2.5	Effects of Structure and Configuration on CUDB Counters	12
	Glossary	17
	Reference List	19





1 Introduction

This document describes the performance management solution and Key Performance Indicators (KPIs) provided by Ericsson Centralized User Database (CUDb).

1.1 Purpose and Scope

This document provides an overview of performance management in CUDb, describes available performance data its generation, as well as how it can be collected and used to measure the performance of a CUDb node. A set of KPI counters are also provided to measure CUDb performance.

1.2 Target Groups

This document is intended for CUDb system operators who will be monitoring the performance of CUDb nodes and for solution architects and system integrators who will be integrating CUDb's performance management solution with a management system.

1.3 Revision Information

Rev. A

This document is based on 4/1553-HDA 104 03/9 with revised content and structure.

Rev. B

Other than editorial changes, this document has been revised as follows:

- Section 2.2 on page 3: Updated the third bullet with information about `cudbCounterPublishingPeriod`.
- Section 2.4 on page 5: New section.
- Added information relevant to the KPIs.

Rev. C

Other than editorial changes, this document has been revised as follows:



- Section 2.5.1 on page 12, Section 2.5.2 on page 13, and Section 2.5.3 on page 13: Updated description.

Rev. D

Other than editorial changes, this document has been revised as follows:

- Section 2.3 on page 4: Updated configuration procedure.

1.4 Prerequisites

The reader of this document should have general knowledge of CUDB. Knowledge of LDAP data access mechanisms and CUDB architecture is recommended for proper understanding of the CUDB performance data.

1.5 Typographic Conventions

Typographic conventions can be found in the following document:

- *Typographic Conventions*



2 Counters in CUDB

2.1 Overview

A set of counter groups is provided for each CUDB node, containing performance data for the following:

- Individual Lightweight Directory Access Protocol (LDAP) servers
- Overall CUDB node performance
- Application groups
- Database clusters
- Simple Object Access Protocol (SOAP) notifications

More details about the information provided by CUDB counters can be found in *CUDB Counters List*, Reference [1].

Note: As part of the integration of different application Front Ends (FEs), CUDB also provides the Application Counters Framework. The framework makes it possible for application FEs to have CUDB gather and publish performance management information about their application data stored in CUDB (on behalf of the application FEs). For more information about this framework, refer to *CUDB Application Counters*, Reference [3].

A special set of counters, KPIs, are distributed in the “Overall CUDB node performance” and “Database cluster” counter groups. For more information about KPIs and their purpose, see Section 2.4 on page 5.

2.2 Counter Generation and Publishing

CUDB counters are generated and published independently on each CUDB node, and are available only on that node. They are not replicated to the rest of the CUDB system.

The generation of counter value samples and publishing of counter data are independent processes, with different execution periods:

- Generation period for CUDB’s own counters is 1 minute.
- The publishing period for all counters is set to 15 minutes by default. It is possible to change the publishing frequency to 5 minutes by updating the value of `cudbCounterPublishingPeriod` attribute accordingly. For more information about this attribute, refer to *CUDB Node Configuration Data Model Description*, Reference [4].



Note: If CUDB is integrated with an NMS that provides its own job definitions for jobs that publish CUDB counters, make sure that the publishing frequency for those jobs matches the value of the `cudbCounterPublishingPeriod` attribute. If this is not the case, the specified publishing frequency will be overwritten by the one defined by the `cudbCounterPublishingPeriod` attribute at the first configuration change after the job deployment.

Counters are published in 3GPP XML format and can be found in the following output location:

```
/home/cudb/oam/performanceMgmt/output/
```

The file format is described in *ESA XML Interface for Performance Management*.

Depending on counter type, the files contain the following information:

For gauge counters:

- The value of the last generated sample
- Maximum value in the publishing period
- Minimum value in the publishing period

For accumulated counters:

- The value of the last generated sample
- Delta value, compared with the value of the first sample of the publishing period

Note: Drops of counter values for certain accumulated counters may happen in case of an LDAP FE restart. In that case, the delta value is not valid and waiting for the next counter publishing is necessary to get a valid delta value.

Attention!

Files are kept in the specified location for one day.

Counter users collect CUDB counter values by copying the generated files from the output location. It is recommended to retrieve output files with the `cudbadmin` user through SFTP protocol. Refer to *CUDB Users and Passwords*, Reference [9] for more information on user credentials.



2.3 Configuring Counter Output Files Names

The filenames of these counter output files are based on the following format:

A<date>.<starttime>-<stoptime>-<jobname>_<networkElementName>.xml

The variables in the above file name are the following:

<date>	The date of the measurement in format YYYYMMDD.
<starttime>	The start time of the measurement in format HHMM.
<stoptime>	The stop time of the measurement in format HHMM.
<jobname>	The job name of the measurement.
<networkElementName>⁽¹⁾	A string used as unique identity representing the node that runs the ESA.

(1) ESA refers to this variable as *uniqueId*.

networkElementName can be configured.

Refer to *ESA Performance Management*, Reference [11] for a complete description of the file names.

The **<networkElementName>** parameter is set through CUDB configuration CLI, by setting the value of the **networkElementName** configuration attribute. For more information, refer to the *Class CudbLocalNode* table of *CUDB Node Configuration Data Model Description*, Reference [4].

Refer to the *Object Model Modification Procedure in CUDB Node Configuration Data Model Description*, Reference [4] for more information on all the steps required to modify the object model (for example, on using the **applyConfig** administrative operation to activate the changes).

2.4 CUDB KPIs

CUDB KPIs are a special set of CUDB counters, for CUDB systems deployed on native BSP 8100, that help the users evaluate and quantify the usage of the processing and memory capacity of certain CUDB resources.

The different types of KPIs per CUDB node type are as follows:

- for LDAP-FE:
 - **kpiLdapFeLoad**: Processing capacity used by LDAP requests, in %.
 - **kpiRatioDroppedLdap**: Ratio of LDAP requests dropped due to the overload of the local LDAP-FE over the total number of LDAP requests received during the same period, in $0/_{000}$.



- for PLDB:
 - `kpiClusterLoad`: Used processing capacity of the PLDB cluster, in %.
 - `kpiRatioDroppedCluster`: Ratio of LDAP requests dropped due to the overload of PLDB, including requests dropped by LDAP-FE and by the local cluster, over the total number of LDAP requests that attempted to access PLDB during the same period, in $0/_{000}$.
 - `memoryUsage`: Used database memory pages over total database memory pages, in %.
- for DS_n:
 - `kpiClusterLoad`: Used processing capacity of the DS _n cluster, in %.
 - `kpiRatioDroppedCluster`: Ratio of dropped LDAP requests due to overload of the DS _n, including requests dropped by LDAP-FE and by the local cluster, over the total number of LDAP requests that attempted to access DS _n during the same period, in $0/_{000}$.
 - `memoryUsage`: Used database memory pages over total database memory pages, in %.

Like other CUDB counters, KPIs are generated every minute. The value of the counters can be published every 5 or 15 minutes, but in both cases, for load related and drop ratio indicators, the counter is a rolling average of the values collected during the previous 15 minute monitoring period, updated each minute. Memory usage KPI counters are not averaged. The KPI values are available in 3GPP file format as the rest of the CUDB counters.

Refer to *CUDB Counters List*, Reference [1] for further information on KPIs.

2.4.1

Guidelines for CUDB KPIs

The KPIs are distinct for resources located in each CUDB node.

Note: Replicas located in different CUDB nodes can have different KPIs values.

To see how structure and configuration of a CUDB system can affect different KPIs, see Section 2.5 on page 12.

KPI guidelines are given for normal CUDB operation, which is described as follows:

- No network incidents, such as site isolation
- No ongoing maintenance operations



- No irregular high load event or incident, such as a network overload event, ongoing massive provisioning, hardware failure, or resource degradation

Note: No KPI related guideline is given for any scenario other than normal CUDB operation.

While an important index of CUDB performance, the processing load itself may not be the limiting factor in all the cases. Specific combinations of traffic from different applications towards a particular CUDB node or the way the network was deployed, including network latencies, may also limit the overall throughput of a CUDB system, while processing load remains nominal.

Therefore, in addition to processing load, further indicators are necessary to measure CUDB performance. The drop ratio KPIs can provide an early warning if different limits of the system are close to being exceeded and because of that not all resulting received traffic is successfully handled.

The drop ratio KPIs have a very high granularity level to alert the user to cases the indicator is not zero, even if the rejection rate is still very low. Accordingly, one-hundredth of a percent ($0.01\% = 0.1\text{‰}$) is used as the unit of measurement for the drop ratio KPIs. For example, the `kpiRatioDropped` value at $5^{0/000}$ indicates a rejection rate of 0.05 %.

Another indicator of CUDB performance is the level of database cluster occupancy, provided by the `memoryUsage` for each database cluster counter.

Prior to checking the value of a KPI counter, ensure that there are no active alarms in the system affecting the CUDB node or CUDB resource that the specific KPI counter is associated with. Refer to *CUDB Health Check*, Reference [5] for instructions on how to check the list of active alarms.

An overview of the procedure for following up on the value of a KPI counter associated with a CUDB resource is shown in Figure 1.

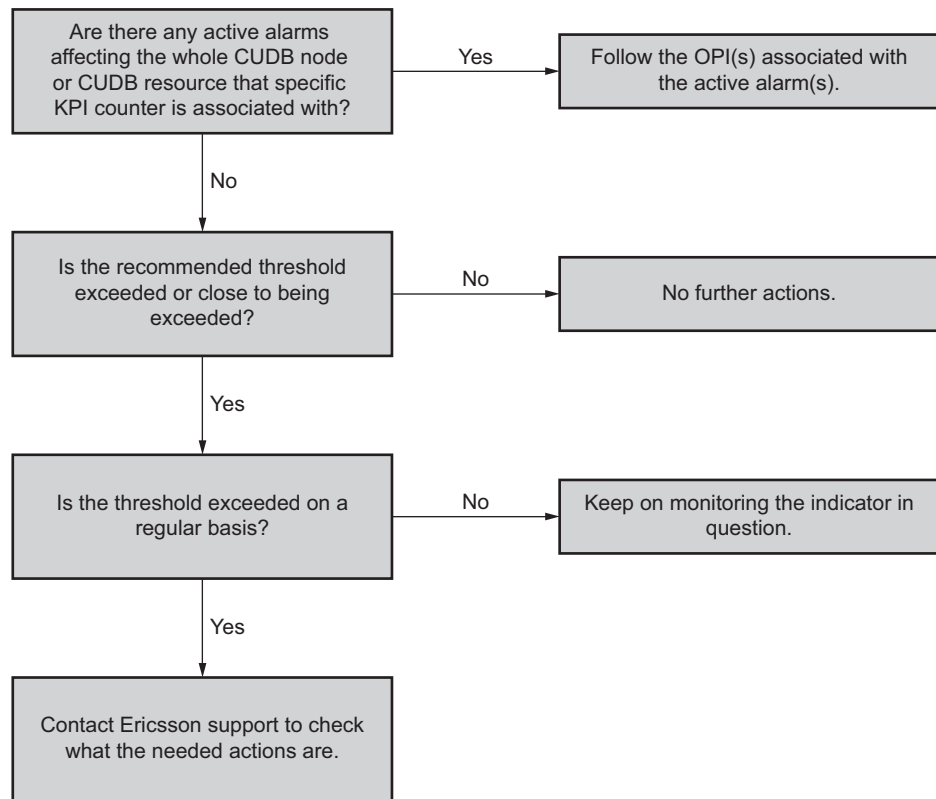


Figure 1 KPI Counter Value Evaluation



2.4.1.1

LDAP KPIs

Table 1

KPI counter name	Threshold	Additional information and recommendation
kpiLdapFeLoad	70 – 80 %	<p>Traffic rejection starts at high CPU usage, above 70% – 80%.</p> <p>If this threshold is reached occasionally, continue monitoring the indicator.</p> <ul style="list-style-type: none"> • If threshold for kpiLdapFeLoad is exceeded on a regular basis, check if kpiLdapFeLoad is at similarly high levels in other CUDB nodes. If that is the case, contact Ericsson support and evaluate the need to expand the system. • Otherwise, if the kpiLdapFeLoad is lower in other CUDB node(s), revise the connectivity map between application FEs and CUDB nodes and contact Ericsson support, if needed.
kpiRatioDroppedLdap	0	<p>If this KPI counter exceeds zero on a regular basis, check the kpiRatioDroppedLdap in other CUDB nodes and, with Ericsson support, evaluate the need to expand the system or to revise the connectivity map between application FEs and CUDB nodes.</p>



2.4.1.2

PLDB KPIs

Table 2 PLDB KPIs

KPI counter name	Threshold	Additional information and recommendation
kpiClusterLoad	70 – 80 %	<p>Traffic rejections starts at a high CPU usage.</p> <p>If the threshold is occasionally reached, continue monitoring the indicator.</p> <p>If the threshold is exceeded on a regular basis, check if kpiClusterLoad is at similarly high levels in other PLDB replicas. If that is the case, contact Ericsson support and evaluate the need to expand the system.</p> <p>Otherwise, if the kpiClusterLoad is lower in other PLDB replica(s), revise the connectivity map between application FEs and CUDB nodes and contact Ericsson support if needed.</p>
kpiRatioDroppedCluster	0	<p>If the value of this KPI counter exceeds zero on a regular basis, check the kpiRatioDroppedCluster in other PLDB replicas and, with Ericsson support, evaluate the need to expand the system or revise the connectivity map between application FEs and CUDB nodes.</p>
memoryUsage	75	<p>If this threshold has been reached and the “Storage Engine, Memory Usage Too High In PLDB, Warning” alarm has not been addressed yet, refer to the instructions specified in <i>Storage Engine, Memory Usage Too High In PLDB, Warning, Reference</i> [7].</p> <p>If the actions described in the OPI do not lower the counter below the threshold, contact Ericsson support and evaluate the need to perform a PLDB expansion.</p>



2.4.1.3

DS KPIs

Table 3 DS KPIs

KPI counter name	Threshold	Additional information and recommendation
kpiClusterLoad	40 – 50 %	<p>Traffic rejection starts at high CPU usage, above 70%- 80%, but due to high availability within a DS, the recommended KPI threshold is 40% – 50%. If one of the database processes fails within a DSG replica, the surviving process keeps providing the database service without noticeable traffic impact. At the same time, under normal operation, with no process failure, the DSGs are prepared to cope with high load.</p> <p>If the threshold is reached occasionally, continue monitoring the indicator.</p> <p>If kpiClusterLoad threshold is exceeded on a regular basis in a DSG, check if kpiClusterLoad is at similarly high levels in other master DSG replicas. If that is the case, contact Ericsson support and evaluate the need to expand the system with additional DSGs.</p> <p>Otherwise, if the threshold is exceeded in just one or a few DSGs, consider reallocating data from the highly occupied DSGs towards DSG(s) with lower occupancy levels. Refer to <i>CUDB Multiple Geographical Areas</i>, Reference [6] for additional information.</p>



KPI counter name	Threshold	Additional information and recommendation
kpiRatioDroppedCluster	0	If this KPI counter exceeds zero on a regular basis, check the kpiRatioDroppedCluster in other master DSGs replicas and, with Ericsson support, evaluate the need for reallocation or for expansion with additional DSG(s).
memoryUsage	75	<p>If this threshold has been reached and the “Storage Engine, Memory Usage Too High In DS, Warning Threshold Reached” alarm has not been addressed yet, refer to the instructions specified in <i>Storage Engine, Memory Usage Too High In DS, Warning Threshold Reached</i>, Reference [8].</p> <p>If the actions described in the OPI do not lower the KPI below the threshold, contact Ericsson support to evaluate the need to perform a DSG expansion.</p>

2.5 Effects of Structure and Configuration on CUDB Counters

In order to properly understand and interpret counter values, important aspects of CUDB data access, architecture, and features need to be taken into account. The relationship of the previous factors with CUDB counters is described in the following sections, as well as some general considerations.

2.5.1 Master Distribution

Depending on the configured combinations of `readModeInDS` and `readModeInPL` configuration parameters, master DS replicas may receive higher amounts of traffic compared to slave replicas within the same DSG.

This will be reflected in the following counter values:

- `intendedLdapRequests, Dsn`
- `processedLdapRequests, Dsn`
- `kpiClusterLoad, Dsn`

Master PLDB replicas may receive higher amounts of traffic compared to the slave replicas during provisioning. This will be reflected in the following counter values:

- `intendedLdapRequests, Plldb`



- `processedLdapRequests, Pldb`
- `kpiClusterLoad, Pldb`

If a node hosts multiple master replicas, the values of the following counters may be higher compared to nodes with fewer master replicas:

- `ldapTpsAtFrontEndn`
- `receivedLdapReqsTotal`
- `processedLdapReqsLocalNode`
- `notificationsSent`
- `kpiLdapFeLoad`

For more information on `readModeInDS` and `readModeInPL`, refer to *CUDB Node Configuration Data Model Description*, Reference [4] and *CUDB LDAP Data Access*, Reference [2].

2.5.2 Distribution of Subscriber Profiles

Depending on the reading mode configuration of LDAP users, higher memory occupation in a DSG may result in its master replica receiving more traffic. In terms of CUDB counters, this means that DSG master replicas with higher `memoryUsage, Dsn` counter values may also have higher values than master replicas of other DSGs for the following counters:

- `intendedLdapRequests, DSn`
- `processedLdapRequests, DSn`
- `kpiClusterLoad, DSn`

A higher active/inactive subscriber ratio in a DSG may also result in its master replica receiving more traffic. Such master replicas may have higher values of the same counters as listed above, compared to master replicas of other DSGs in the system.

2.5.3 Application FE Connections

The CUDB nodes that are the primary targets for Application FE connections will receive most of the traffic intended for a CUDB System. Depending on the master distribution in the system and the reading mode configuration of LDAP users, such traffic may either end at the primarily affected nodes or be proxied to other nodes in the system.

If the CUDB nodes connected to Application FEs do not host many master replicas, they may have a high number of proxied requests, resulting in a higher value of `processedLDAPReqsRemoteNodes` than other nodes of the system.



If there are no nodes in the system with a high concentration of master replicas, nodes with Application FE connections will have higher values than other nodes in the system for the following counters:

- `ldapTpsAtFrontEndn`
- `receivedLdapReqsTotal`
- `processedLdapReqsLocalNode`
- `kpiLdapFeLoad`
- `kpiClusterLoad, Pldb`

Otherwise, depending on the reading mode configuration of LDAP users, nodes with a concentration of master replicas may have the highest values for the listed counters.

2.5.4 Network Issues

Increased network latency can result in a higher number of failed proxied requests, such as in the increased value of `nonProcessedLdapReqsRemoteNodes`.

Network issues in communication with Notification end points can result in failed SOAP notifications, such as an increase of `notificationsFailed` counter values.

2.5.5 Overload Protection and Load Regulation

Incidents in the core network or on UDC solution level can cause high traffic and trigger the overload protection and load regulation mechanisms, resulting in an increased value of the dropped requests counters:

- `droppedLdapReqsLocalLdapLayer`
- `droppedLdapReqsLocalClusters`
- `droppedLdapRequests, Pldb`
- `droppedLdapRequests, Dsn`
- `nonProcessedLdapReqsRemoteNodes`
- `droppedAndFailedLdapReqsAppGrpn`
- `droppedAndFailedLdapReqsAppGrpn`
- `kpiRatioDroppedCluster, Pldb`
- `kpiRatioDroppedCluster, Dsn`
- `kpiRatioDroppedLdap`



2.5.6 General Considerations

CUDB maintenance operations can impact local redundancy of a CUDB node or cause high network, storage, and processing load, resulting in an increase of dropped or failed requests as well as the load related and drop ratio KPI counter values.

Infrastructure problems or maintenance can impact the capacity and availability of network, storage, and processing resources, resulting in an increase of dropped or failed requests as well as the load related and drop ratio KPI counter values.





Glossary

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





Reference List

CUDB Documents

- [1] *CUDB Counters List*
- [2] *CUDB LDAP Data Access*
- [3] *CUDB Application Counters*
- [4] *CUDB Node Configuration Data Model Description*
- [5] *CUDB Health Check*
- [6] *CUDB Multiple Geographical Areas*
- [7] *Storage Engine, Memory Usage Too High In PLDB, Warning*
- [8] *Storage Engine, Memory Usage Too High In DS, Warning Threshold Reached*
- [9] *CUDB Users and Passwords, 3/006 51-HDA 104 03/10*
- [10] *CUDB Glossary of Terms and Acronyms*

Other Ericsson Documents

- [11] *ESA Performance Management*