

CUDB Deployment Guide

CONFIGURATION DIRECTION

Copyright

© Ericsson AB 2016. 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	Scope	1
1.2	Revision Information	1
1.3	Target Groups	1
1.4	Typographic Conventions	2
2	Overview	2
3	Deployment Considerations	3
3.1	IP Transport Network Considerations	3
3.2	Geographical Redundancy Considerations	3
3.3	CUDB Site and Node Number Considerations	4
3.4	CUDB Node Distance Considerations	6
3.5	Virtual Deployment Considerations	7
4	CUDB Deployment Examples	8
4.1	Example Networks Dimensioned with 5 DSGs	8
4.2	Example Networks Dimensioned with 15 DSGs	11
4.3	Example Networks Dimensioned with 30 DSGs	14
	Glossary	19
	Reference List	21





1 Introduction

This document describes reference network configurations to help designing the deployment of an Ericsson Centralized User Database (CUDB) system on operator networks.

1.1 Scope

The document covers and describes the following:

- Considerations of CUDB deployments.
- Good practice recommendations and CUDB deployment strategies.
- Examples of CUDB deployment configurations.

1.2 Revision Information

Rev. A

This document is based on 2/1553-HDA 104 03/9 with the following changes:

- Terminology updates throughout the document because of virtualized deployment support.
- Section 1.3 on page 1, Section 2 on page 2, and Section 4 on page 7: Added virtualization information.
- Section 3.2 on page 3: Updated Triple Geographical Redundancy requirements. Added new consideration on virtual deployments.
- Section 3.5 on page 7: Added as new section on virtual deployment considerations.
- Section 4 on page 7: Updated deployment examples.
- Section 4.1.1 on page 8: Removed obsolete information from the captions of Table 1 and Figure 3.

1.3 Target Groups

This document is intended for Ericsson technical sales support personnel who help customers with designing the appropriate infrastructure deployment for a CUDB system.



1.4 Typographic Conventions

Typographic conventions can be found in the following document:

- *Typographic Conventions*

2 Overview

Ericsson evolved its telecom subscriber database used by its mobile circuit switch and packet switch networks, along with its IP Multimedia System (IMS) solution. The development resulted in the physical-logical decoupling of the subscriber data, the application service, and the business logic.

This allows the separate dimensioning of processing and storage resources, as well as the application of the most optimal technology and platform for each layer. The physical-logical decoupling also allows the unification of the specific user data located in the Home Location Register / Authentication Center (HLR/AUC) and the Home Subscriber Server / Subscription Locator Function (HSS/SLF) under a common subscriber profile. This profile is stored in a common database (the CUDB), dramatically reducing the complexity of subscriber data management.

All subscriber-related data is stored and managed in the Data Layer of CUDB, while the logic of the services triggered and executed in the classic HLR/AUC monolithic architecture is performed on the HLR Front End (FE) node. Similarly, the logic of the services provided by the classic HSS/SLF Monolithic Architecture is executed on the HSS/SLF FE node, while the subscriber data is stored in CUDB. The same strategy is applied to other application FEs.

The CUDB system consists of a set of CUDB nodes cooperating to provide CUDB service in the operator network. A CUDB node provides access to the CUDB system services. CUDB nodes can be placed in different locations for distribution and redundancy purposes.

CUDB stores subscriber profiles and provisioned (static) or non-provisioned (dynamic) service data associated to subscribers. As a subscriber-centric database, CUDB also holds the service profiles for the supported applications.

CUDB supports applications such as HLR/AUC, Mobile Number Portability (MNP), HSS/SLF, and Ericsson Authentication, Authorization, and Accounting (AAA). Other applications can also be supported through integration services.

Refer to the *CUDB Technical Product Description*, Reference [1] for more information about CUDB system functions and architecture.



3 Deployment Considerations

The following sections outline some considerations to take into account when deploying a CUDB system.

3.1 IP Transport Network Considerations

All CUDB nodes are connected through a backbone infrastructure (but the transport network is not part of the CUDB system). The backbone is equivalent in all the CUDB nodes in the system, so all CUDB nodes in a CUDB system have the same connectivity.

The CUDB system has been designed on the assumption that the transport network is reliable between all the CUDB nodes.

The network requirements depend on which applications store data in CUDB.

Depending on the use of the links, the following security approaches are available:

- Traffic from application FEs can be encrypted using SSL/TLS or SASL.
- Traffic between CUDB nodes (concerning Lightweight Directory Access Protocol (LDAP) proxy and replication) can be encrypted using SSL/TLS.
- CUDB also implements integrity check and symmetric authentication of messages based on electronic signature for monitoring traffic.

For further networking information, refer to *CUDB Node Network Description*, Reference [2] and *CUDB Security and Privacy Management*, Reference [3].

3.2 Geographical Redundancy Considerations

CUDB deployments can be configured to be geographically redundant. The feature allows storing copies of the same data in different CUDB nodes in different sites, providing a high redundancy and resiliency level. This feature is available in the following two configurations:

- **Double Geographical Redundancy:** This redundancy configuration (also known as 1+1 redundancy) allows two geographical replicas per partition, each one hosted in a different node located in a different site.
- **Triple Geographical Redundancy:** This redundancy configuration (also known as 1+1+1 redundancy) allows three replicas per partition, each one hosted in a different node, located in a different site.

Note:

- Triple Geographical Redundancy can only be used if the Advanced Network Protection Value Package is available.
- Standalone configuration without geographical redundancy is supported for Customer Trial, Customer Test, and Ericsson Internal systems.
- During the installation of a virtualized CUDB system, do not to deploy nodes belonging to different CUDB sites in the same cloud infrastructure zone.

Geographical redundancy is one of the key features ensuring the high availability of the CUDB system: higher number of data replicas result in higher availability of the system.

For details, refer to *CUDB Multiple Geographical Areas*, Reference [4].

3.3 CUDB Site and Node Number Considerations

CUDB sites are physical locations where the CUDB nodes of a CUDB system are deployed. The maximum number of supported sites is six but the recommended configurations are 2 and 3 sites. The preferred one is to set up three CUDB sites to reduce the impact of node and IP backbone failures, and to increase the reliability of the entire system.

To ensure the simple network design of the CUDB system, it is recommended to deploy the system with symmetrical CUDB nodes (that is, with nodes storing a similar set of Data Store Groups, or DSGs, in terms of DSG allocation) across the CUDB sites of the system. Depending on the system deployment, this may result in nodes containing a substantial amount of free space for additional Data Store Units (DS), enabling further expansion. However, if system expansion is not considered, the number of CUDB nodes can be reduced by deploying the system with asymmetrical CUDB nodes (that is, with nodes whose DSG allocation differs from each other).

Figure 1 shows an example deployment consisting of symmetrical nodes, configured with double geographical redundancy:

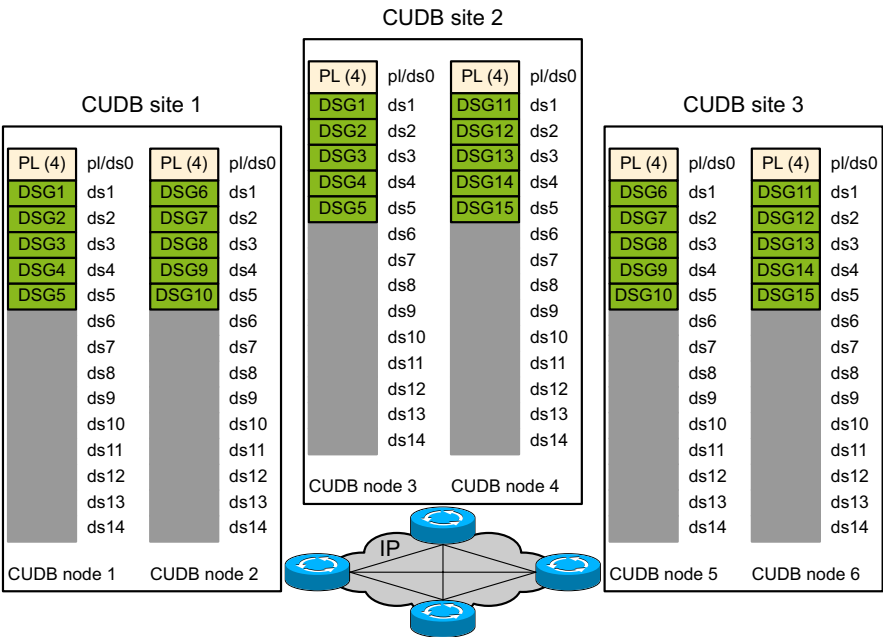


Figure 1 Example CUDB Deployment with Symmetrical CUDB Nodes

The system shown on Figure 1 is deployed with double geographical redundancy, and the nodes are distributed across three CUDB sites. In accordance with the deployment concept described above, the system consists of symmetrical CUDB nodes: Node 1 is symmetrical to Node 3, Node 2 holds a similar set of DSs as Node 5, and so on.

Figure 2 shows a deployment without symmetrical CUDB nodes, but also configured with double geographical redundancy:

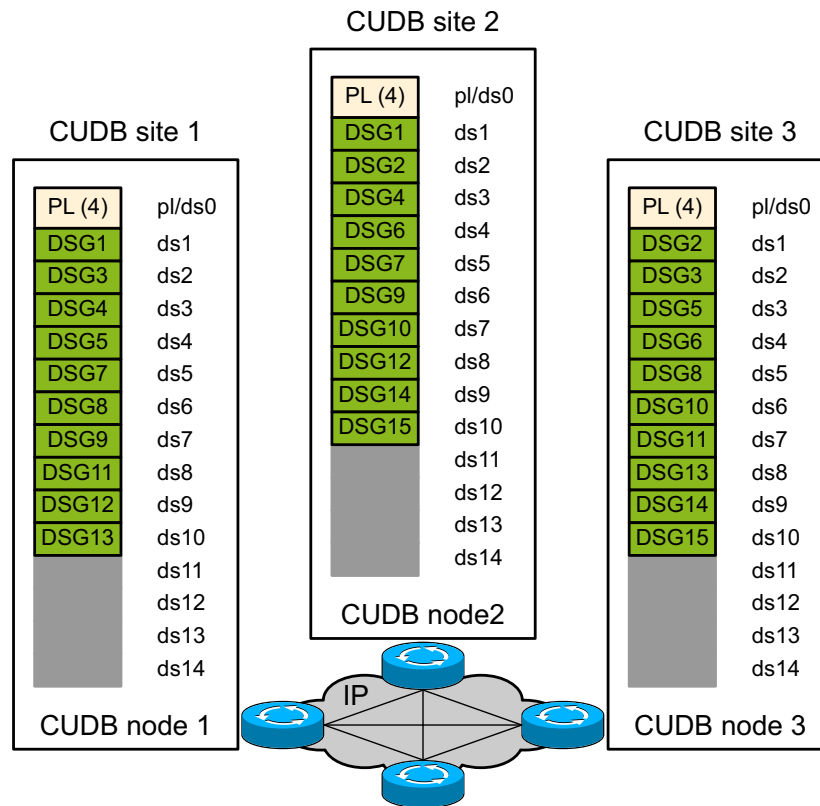


Figure 2 Example CUDB Deployment with Asymmetrical CUDB Nodes

3.4 CUDB Node Distance Considerations

CUDB systems are often deployed to cover wide areas, either with complex deployments consisting of a significant number of CUDB nodes, or with smaller systems consisting of less nodes, separated by long distances.

The distance between CUDB nodes greatly affects network latency, which impacts CUDB internal interfaces (that is, the replication traffic between CUDB nodes, the LDAP proxy traffic, and the supervision protocol as well).

3.4.1 Network Latency Impacts on LDAP Proxy Traffic

CUDB is a geographically distributed system that provides local single logical points of access for traffic and provisioning. Each CUDB node provides access to the whole subscriber base, regardless of the data distribution across CUDB nodes. In case a CUDB node receives an LDAP query, the system checks if the master of the requested data is allocated in the local CUDB node. If not, then the LDAP query is sent to the CUDB node where the master is allocated. Such queries are considered a proxy operation, and their latency is higher than the latency of local LDAP queries.



High latency or bad quality in the transport network could dramatically affect CUDB performance, especially LDAP proxy traffic.

3.5 Virtual Deployment Considerations

BC servers of the site must be deployed in a way that a single hardware failure in a cloud system does not result in the loss of the majority of the BC servers in the site.

Warning!

If the nodes belonging to a site are to be deployed in the same cloud system and in the same availability zone, extra caution must be taken during the creation of host aggregates for the System Controller (SC) Virtual Machines (VMs).

When creating host aggregates, consider the BC server distribution in the site configuration. Refer to the corresponding table in the “Server Resiliency” section of *CUDB High Availability*, Reference [5] for more information, and consider the following:

- If the site contains one or two CUDB nodes, no special measures are necessary.
- If the site contains between 3 and 5 nodes, those nodes host one or two BC servers in the SCs. The host aggregates created for the deployment of SC VMs cannot share compute hosts: such deployment can introduce a single point of failure in the BC cluster of the site.
- If the site contains more than 5 nodes, the information in the previous bullet is valid for the first 5 nodes.

Note: The considerations in this section are not valid for a virtualized CUDB system deployment on Ericsson's Cloud Execution Environment (CEE), on BSP 8100 hardware with GEP5 blades, using vCUDb_16CPU_47GB flavors for SCs and PLs (It is not possible to collocate VMs on the compute hosts).



4 CUDB Deployment Examples

This section describes the basic proposals of CUDB deployments regarding network size (in number of DSGs).

The number of CUDB nodes needed for hosting the amount of subscribers on each reference deployment has been estimated by basic dimensioning.

CUDB is deployed on the BSP 8100 infrastructure with Generic Ericsson Processor version 5 (GEP5) blades or with Generic Ericsson Processor version 3 (GEP3) blades, or virtualized on a cloud infrastructure with the following considerations:

- A CUDB node can have up to 36 blades or virtual machines.
- Combining GEP3 and GEP5 blades in a single BSP 8100 node is not supported.
- Combining BSP 8100 (GEP3) nodes and BSP 8100 (GEP5) nodes in a single system is supported.

The Processing Layer Database (PLDB) consists of 4-16 PLDB blades in BSP 8100 (GEP3), 2-16 PLDB blades in BSP 8100 (GEP5), or 2-16 virtual machines in a virtualized CUDB. The more PLDB blades or virtual machines are defined, the less DSGs can be configured.

Note: In case of BSP 8100 (GEP5), when the "BSP Capturing Unit Option" is used, the number of total PLDB and DS blades are reduced. Refer to the BSP 8100 CPI for more information.

The CUDB configurations described in the following sections are reference deployments only, and therefore must be interpreted as guidelines, not actual CUDB configurations. Make sure that the recommendations on system robustness and resiliency described in Section 3 on page 2 are followed when planning the deployment.

Actual deployments delivered to operators are configured according to the appropriate dimensioning exercises performed for specific cases, though the general recommendations included in this guide apply to actual deployments as well.

4.1 Example Networks Dimensioned with 5 DSGs

This section describes two example CUDB deployments with 5 DSGs.



4.1.1

Example Network with Double Geographical Redundancy

This example describes a reference deployment with double geographical (or 1+1) redundancy, based on the In Service Performance (ISP) requirements.

Table 1 and Figure 3 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has two copies.
- The system consists of two sites.
- The sites consist of symmetrical CUDB nodes.

Table 1 CUDB System Deployment with 5 DSGs and Double Geographical Redundancy

Location	Number of CUDB Nodes
Site 1	1
Site 2	1

Figure 3 depicts the deployment scenario.

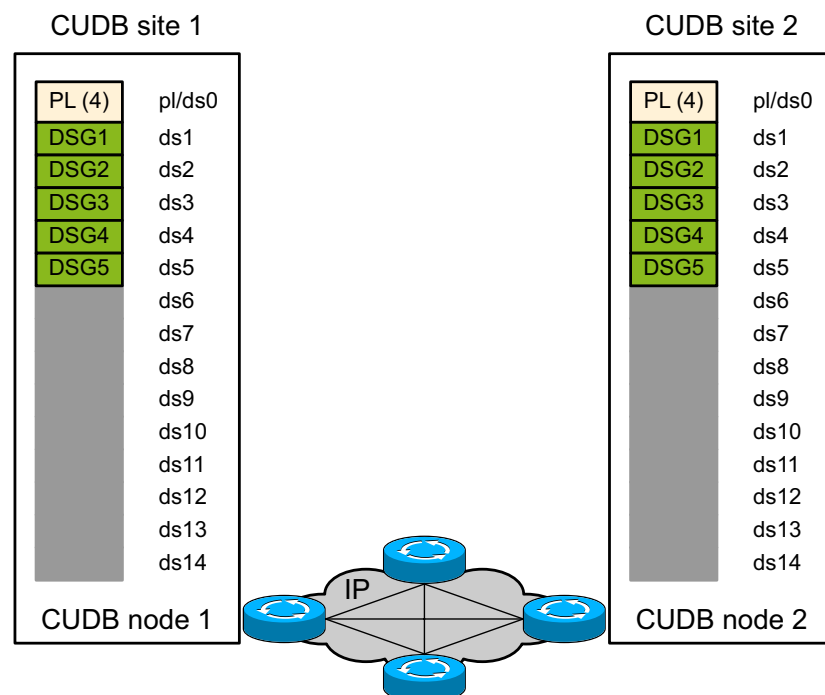


Figure 3 CUDB System Deployment with 5 DSGs and Double Geographical Redundancy



The picture shows that there are nine 'empty' slots (from DS6 to DS14) in all the nodes, as the demanded capacity is already reached with the first five DS Units.

All subscriber partitions from DSG1 to DSG5 are replicated in the second the CUDB node is located in the other site.

4.1.2 Example Network with Triple Geographical Redundancy

This scenario is similar to the example shown in Section 4.1.1 on page 8, with the exception that an additional site and CUDB node is required to achieve the selected redundancy level.

Table 2 and Figure 4 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has three copies.
- The system consists of three sites.
- The sites consist of symmetrical CUDB nodes.

Table 2 CUDB System Deployment with 5 DSGs and Triple Geographical Redundancy

Location	Number of CUDB Nodes
Site 1	1
Site 2	1
Site 3	1

Figure 4 depicts the deployment scenario.

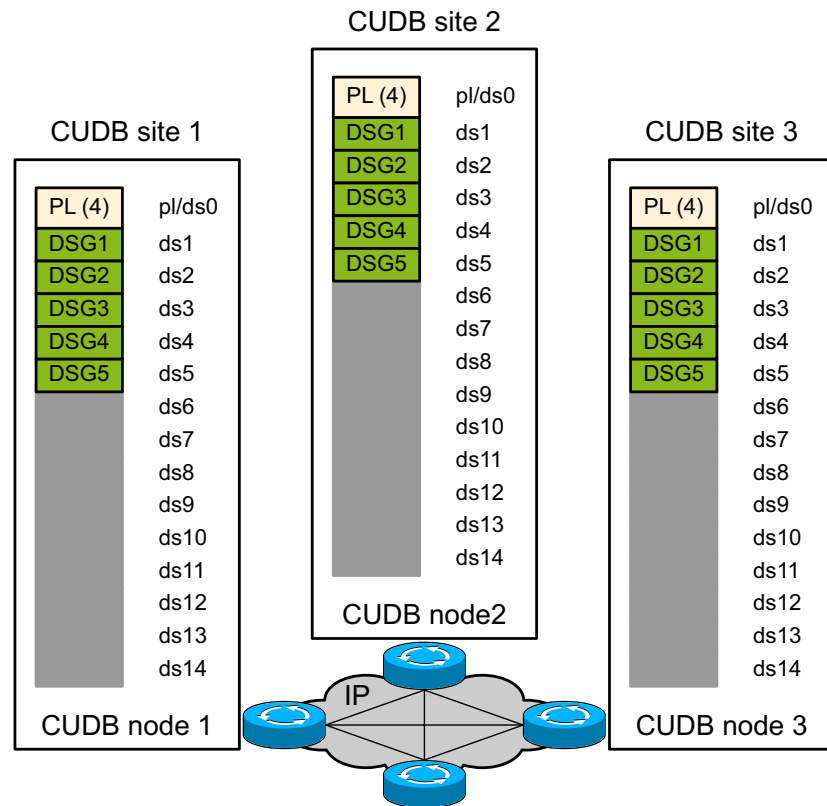


Figure 4 CUDB System Deployment with 5 DSGs and Triple Geographical Redundancy

The picture shows how all CUDB nodes hold a copy of each subscriber partition.

4.2 Example Networks Dimensioned with 15 DSGs

This section describes three example CUDB deployments with 15 DSGs.

4.2.1 Example Networks with Double Geographical Redundancy

This example describes a reference deployment with double geographical (or 1+1) redundancy, based on the ISP requirements.

Although CUDB systems handling 15 DSGs can be deployed with double geographical redundancy, it is recommended to configure the system with triple geographical (1+1+1) redundancy to ensure deployment robustness.

Table 3 and Figure 5 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has two copies.

- The system consists of two sites.
- The sites consist of symmetrical CUDB nodes.

Table 3 *CUDB System Deployment with 15 DSGs, Two Sites and Double Geographical Redundancy*

Location	Number of CUDB Nodes
Site 1	2
Site 2	2

Figure 5 depicts the deployment scenario.

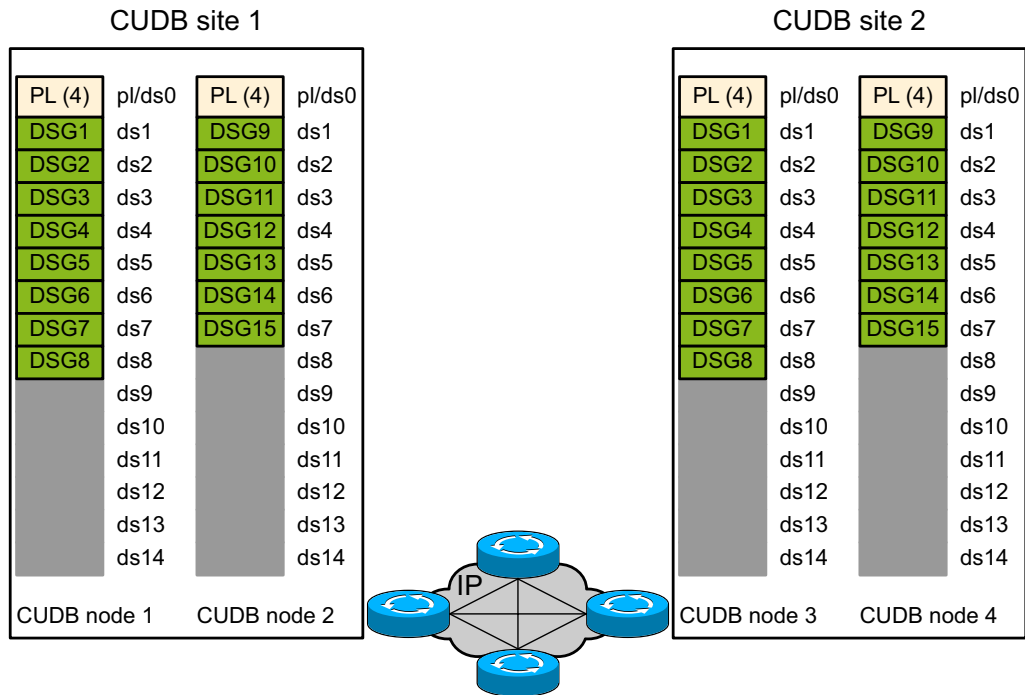


Figure 5 *CUDB System Deployment with 15 DSGs, Two Sites and Double Geographical Redundancy*

Table 4 and Figure 6 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has two copies.
- The system consists of three sites.
- The sites consist of asymmetrical CUDB nodes.



Table 4 CUDB System Deployment with 15 DSGs, Three Sites and Double Geographical Redundancy

Location	Number of CUDB Nodes
Site 1	1
Site 2	1
Site 3	1

Figure 6 depicts the deployment scenario.

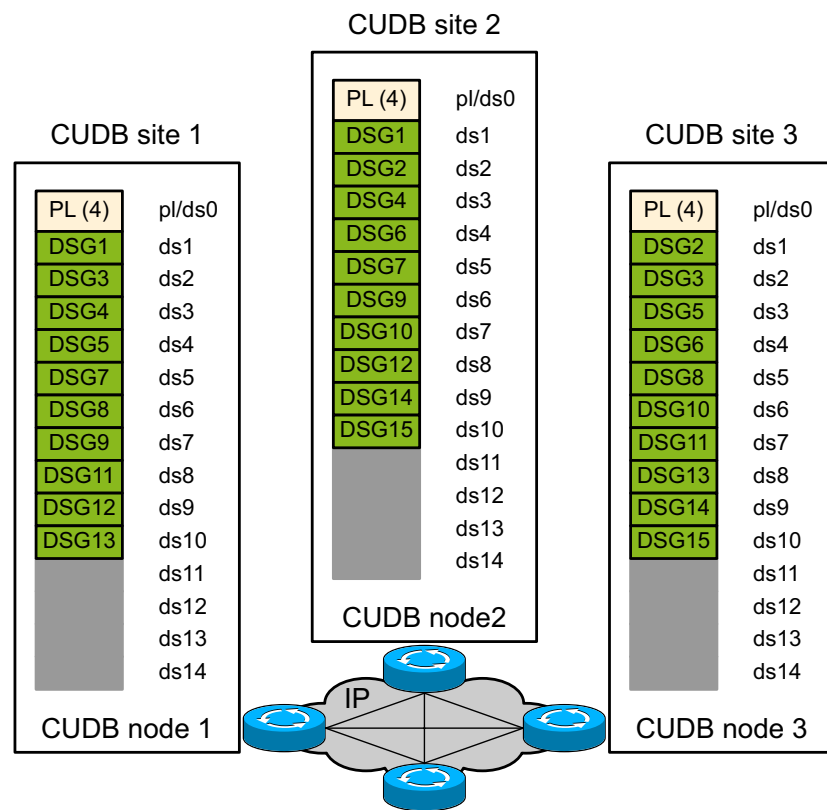


Figure 6 CUDB System Deployment with 15 DSGs, Three Sites and Double Geographical Redundancy

4.2.2

Example Network with Triple Geographical Redundancy

This example describes a reference deployment with triple geographical (or 1+1+1) redundancy.

Table 5 and Figure 7 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has three copies.

- The system consists of three sites.
- The sites consist of symmetrical CUDB nodes.

Table 5 CUDB System Deployment with 15 DSGs and Triple Geographical Redundancy

Location	CUDB Nodes
Site 1	2
Site 2	2
Site 3	2

Figure 7 depicts the deployment scenario.

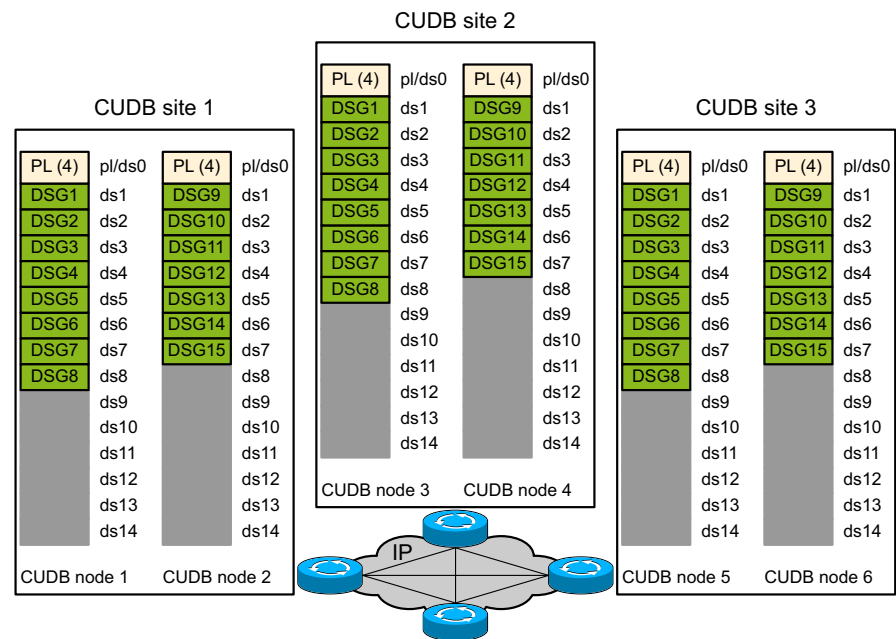


Figure 7 CUDB System Deployment with 15 DSGs and Triple Geographical Redundancy.

4.3 Example Networks Dimensioned with 30 DSGs

This section describes two example CUDB deployments with 30 DSGs.

4.3.1 Example Network with Double Geographical Redundancy

This example describes a reference deployment with double geographical (1+1) redundancy, based on the ISP requirements.



Although CUDB systems handling 30 DSGs can be deployed with double geographical redundancy, it is recommended to configure the system with triple geographical (1+1+1) redundancy to ensure deployment robustness.

Table 6 and Figure 8 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has two copies.
- The system consists of three sites.
- The sites consist of asymmetrical CUDB nodes.

Table 6 CUDB System Deployment with 30 DSGs, Three Sites and Double Geographical Redundancy

Location	Number of CUDB Nodes
Site 1	2
Site 2	2
Site 3	2

Figure 8 depicts the deployment scenario.

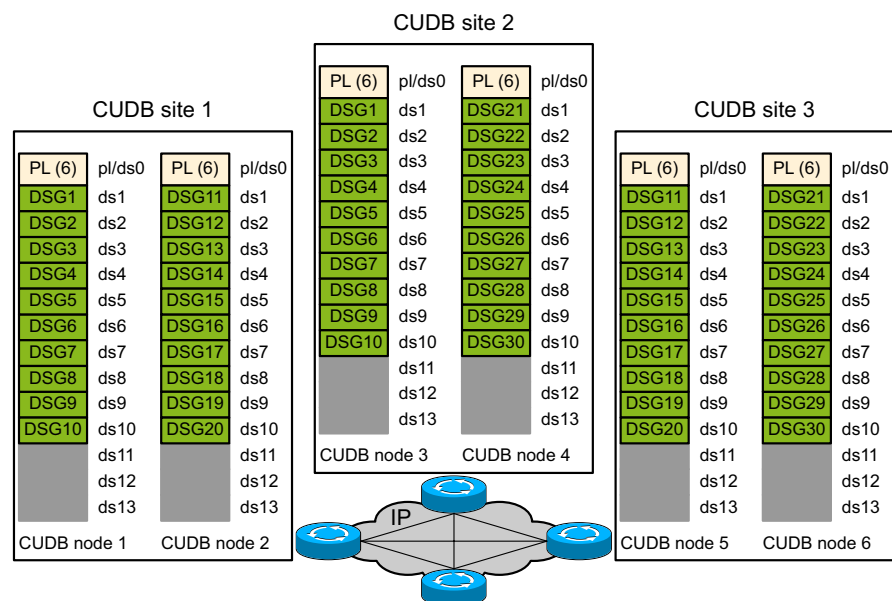


Figure 8 CUDB System Deployment with 30 DSGs, Three Sites and Double Geographical Redundancy.

4.3.2

Example Network with Triple Geographical Redundancy

This example describes a reference deployment with triple geographical (1+1+1) redundancy.



Table 7 and Figure 9 show and describe an example deployment with the following configuration settings:

- Each piece of data stored in the system has three copies.
- The system consists of three sites.
- The sites consist of symmetrical CUDB nodes.

Table 7 CUDB System Deployment with 30 DSGs, Three Sites and Triple Geographical Redundancy

Location	Number of CUDB Nodes
Site 1	3
Site 2	3
Site 3	3

Figure 9 depicts the deployment scenario.

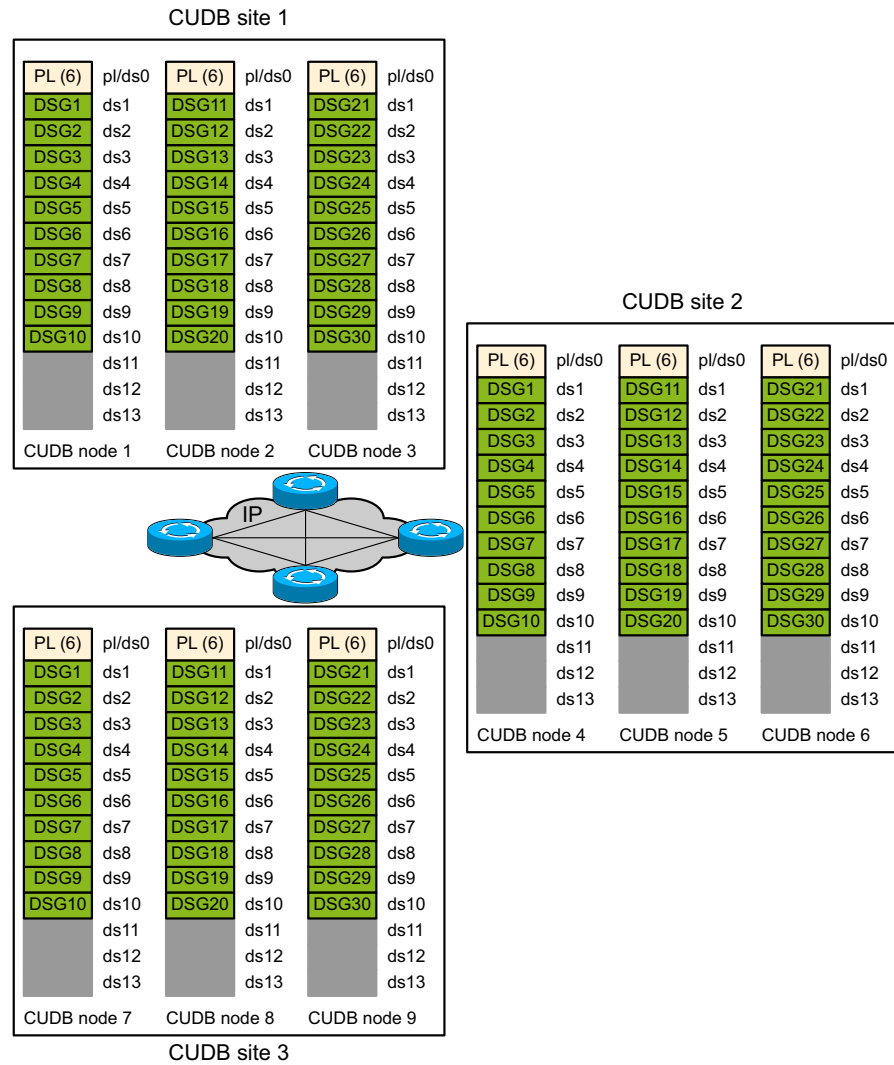


Figure 9 CUDB System Deployment with 30 DSGs, Three Sites and Triple Geographical Redundancy





Glossary

For the terms, definitions, acronyms and abbreviations used in this document, refer to *CUDB Glossary of Terms and Acronyms*, Reference [6].





Reference List

CUDB Documents

- [1] *CUDB Technical Product Description*
- [2] *CUDB Node Network Description*
- [3] *CUDB Security and Privacy Management*
- [4] *CUDB Multiple Geographical Areas*
- [5] *CUDB High Availability*
- [6] *CUDB Glossary of Terms And Acronyms*