UC | Facultad de Ciencias

# Diseño de una infraestructura de red de altas prestaciones para los sistemas de almacenamiento y cómputo científico del CPD 3MARES

## (Design of a High-Performance Network Infrastructure for the Storage and Scientific Computing Systems of the 3MARES Datacenter)

Trabajo de Fin de Máster
para acceder al
MÁSTER EN INGENIERÍA INFORMÁTICA

Autor: Daniel Padilla Acebo

Director: Antonio S. Cofiño González

Septiembre - 2024

# Índice General

**Introducción** **1**

1.1 Estructura de la memoria 2

**Contexto del proyecto** **4**

2.1 ¿Qué es el CPD 3MARES? 4

2.1 ¿Para qué se sirve el CPD 3MARES? 6

2.2 ¿Qué es el Servicio de Computación para la Investigación? 8

**Antecedentes y objetivos del proyecto** **10**

3.1 Antecedentes 10

3.2 Objetivos del proyecto 11

**Descripción técnica de los sistemas de cómputo y almacenamiento actuales** **12**

4.1 Clúster de cómputo 12

4.2 Sistema de almacenamiento distribuido 13

**Descripción técnica de los nuevos sistemas de cómputo y almacenamiento** **16**

5.1 Clúster de cómputo 16

5.2 Sistema de Almacenamiento Distribuido 17

**Revisión de las redes de interconexión presentes en el CPD 3MARES** **18**

6.1 Red general del CPD 3MARES 21

6.1.1 Tecnologías 21

6.1.2 Equipamiento 24

6.1.3 Topología 25

6.1.4 Configuración de red 27

6.2 Red de altas prestaciones 30

6.2.1 Equipamiento 30
6.2.2 Topología 30

**Propuesta técnica de mejora de la red de altas prestaciones 33**
7.1 Topología de la red 34
7.2 Interconexión con la Red General del CPD 3MARES 36
7.3 Especificaciones técnicas del nuevo equipamiento de red 37
7.4 Instalación, cableado y configuración necesaria 38

**Conclusiones 40**

**Bibliografía 41**

# Índice de Figuras


**INTRODUCCIÓN**

Vista exterior del CPD 3MARES 5

Vista interior del CPD 3MARES 5

Plano cenital del CPD 3MARES 5

Grupos de investigación usuarios del CPD 3MARES 6

Datos sobre el CPD 3MARES 7

**EQUIPAMIENTO ACTUAL**

Clúster de cómputo Actual 13

Servidores de almacenamiento de metadatos (MDS) 14

Servidores de almacenamiento de datos (OSS) 15

**REDES DE INTERCONEXIÓN DEL CPD**

Link Aggregation Group 22

MultiChassis - Link Aggregation Group 22

Switch Extreme Networks 5420M-24T 25

Switch Extreme Networks 7520-48Y-8C 25

Topología de la red general del CPD 3MARES 26

Switch Mellanox SN2100 30

**PROPUESTA DE MEJORA**

Topología de la propuesta de red de altas prestaciones 34

# Resumen

Este Trabajo de Fin de Máster (TFM) se centra en el rediseño y ampliación de una infraestructura de red para los sistemas de almacenamiento y cómputo científico en el Centro de Procesamiento de Datos (CPD) 3MARES de la Universidad de Cantabria. La iniciativa surge como respuesta directa a la necesidad de integrar y optimizar la red existente para soportar tanto el equipamiento actual como las futuras adquisiciones tecnológicas.

El CPD 3MARES, que desempeña un papel crucial en la investigación y el desarrollo académico, se enfrenta a desafíos técnicos, debido a la heterogeneidad de su equipamiento y las variadas necesidades de sus usuarios, que incluyen desde grupos de investigación internos hasta colaboraciones con empresas externas.

Este proyecto no aborda un análisis técnico de las tecnologías de red, sino que desarrolla una solución de ingeniería con el objetivo de mejorar y expandir la capacidad de la red del CPD para facilitar una interacción más eficiente, robusta y con mejores prestaciones entre los componentes de cómputo y almacenamiento. Además, se busca diseñar una red que no solo resuelva los requerimientos actuales, sino que también ofrezca la flexibilidad necesaria para adaptarse a futuras ampliaciones y exigencias tecnológicas.

Este trabajo subraya la importancia de una infraestructura de red bien planificada y ejecutada, que es esencial para apoyar las demandas de un entorno de investigación heterogéneo y tecnológicamente avanzado como el de la Universidad de Cantabria.

# Abstract

This Master's Thesis (TFM) focuses on the redesign and expansion of a network infrastructure for the storage and scientific computing systems at the Data Processing Center (CPD) 3MARES of the University of Cantabria. The initiative arises as a direct response to the need to integrate and optimize the existing network to support both current equipment and future technological acquisitions.

The CPD 3MARES, which plays a crucial role in research and academic development, faces technical challenges due to the heterogeneity of its equipment and the varied needs of its users, ranging from internal research groups to collaborations with external companies.

This project does not address a technical analysis of network technologies but develops an engineering solution aimed at improving and expanding the CPD's network capacity to facilitate more efficient, robust, and high-performance interactions between computing and storage components. Additionally, the goal is to design a network that not only meets current requirements but also offers the necessary flexibility to adapt to future expansions and technological demands.

This work underscores the importance of a well-planned and executed network infrastructure, which is essential to support the demands of a heterogeneous and technologically advanced research environment such as that of the University of Cantabria.

# Capítulo 1

# Introducción

Como sugiere el título de esta memoria, el proyecto tiene el foco principal en las redes de interconexión entre equipos, concretamente en aquellas redes que consideramos de “altas prestaciones” debido a su alto rendimiento en términos de latencia y/o ancho de banda. Sin embargo, aunque sean este tipo de redes las que se lleven toda la atención a la hora de hablar de las prestaciones de computación y almacenamiento en entornos HPC (*High Performance Computing)*, también es muy importante prestar la debida consideración en el diseño y configuración de todas aquellas redes de interconexión, de prestaciones normales, si se pueden llamar así, que encontramos en un Centro de Procesamiento de Datos. Por supuesto, de estas redes también hablaremos en esta memoria.

Vamos a comenzar por definir aquello que NO es el proyecto desarrollado. NO es un estudio sobre el estado del arte ni tampoco se trata de un análisis técnico de las tecnologías de red en el ámbito del Centro de Datos. El trabajo realizado es un proyecto de ingeniería que responde a una necesidad surgida en el Centro de Procesamiento de Datos de la facultad de ciencias de la Universidad de Cantabria (CPD 3MARES). Esta necesidad no es otra que la expansión e integración de nuevo equipamiento junto con la infraestructura de cómputo y almacenamiento ya existente en el CPD.

Esta situación es muy frecuente en centros de datos con un alto grado de heterogeneidad tanto en las características técnicas como en el ciclo de vida del equipamiento. Causada por la gran variedad de usuarios y sus correspondientes necesidades y financiación. El CPD 3MARES es un claro ejemplo de este tipo de entorno y, por lo tanto, nos enfrentamos a este tipo de eventos de manera regular.

Entonces, al ser una situación tan habitual, es normal que nos planteemos varias cuestiones como: ¿cuál es la dificultad?, ¿qué tiene de particular la situación relevante a este proyecto? Bien, el objetivo de este proyecto no es solo el de incorporar un nuevo lote de servidores, tarea que podría considerarse casi trivial en ciertas ocasiones. La particularidad viene en el

deseo de rediseñar una red de altas prestaciones existente en el CPD, mejorando y ampliando esta infraestructura para que interconecte de manera adecuada tanto el equipamiento actualmente conectado como los nuevos equipos que se van a adquirir.

Como aclaración para todo aquel que lea este documento, en el momento realizar este proyecto y de escribir esta memoria, aún no se ha realizado la compra del nuevo equipamiento, por lo tanto, al referirse en este documento a la infraestructura "actual", se hace alusión a todo aquel equipamiento y configuración ya presente en el CPD antes de la incorporación de los nuevos sistemas de cómputo, almacenamiento y de red.

Como fin a esta introducción se describe la estructura que presenta esta memoria.

## 1.1 Estructura de la memoria

Está dividido en 8 capítulos:

1. **Introducción:**
   Breve descripción del proyecto y sus objetivos

2. **Contexto del proyecto:**
   Presenta toda la información relacionada con el entorno de trabajo donde se ha realizado este proyecto.

3. **Antecedentes y objetivos del proyecto:**
   Expone la situación que motivó el planteamiento y el inicio del proyecto, así como los objetivos que se plantean obtener tras su realización.

4. **Descripción técnica de los sistemas de cómputo y almacenamiento presentes en el CPD 3MARES:**
   Proporciona una descripción general sobre los sistemas de almacenamiento y cálculo existentes, previo a la realización de este proyecto. Solo entrando en los detalles relevantes para el diseño de la nueva red de interconexión.

5. **Descripción técnica de los nuevos sistemas de cómputo y almacenamiento:**
   Al igual que el capítulo anterior, proporciona una descripción general, esta vez del nuevo equipamiento de cómputo y almacenamiento que se va a incorporar en el CPD 3MARES.

6. **Revisión de las redes de interconexión presentes en el CPD 3MARES:**
   Descripción detallada del equipamiento, arquitectura y configuración de las distintas redes físicas y lógicas presentes en el CPD. En este apartado se pretende dar una visión al completo de la topología de interconexión desde un enfoque técnico y visual.

7. **Propuesta técnica de mejora de la red de altas prestaciones:**
   Descripción detallada del proyecto de ampliación y mejora de la red de altas prestaciones objeto de este Trabajo de Fin de Máster.

8. **Conclusiones**

# Capítulo 2

# Contexto del proyecto

El propósito de este apartado es describir en mayor detalle todas las piezas que conforman el contexto que da vida a este proyecto.

A continuación profundizaremos en varios temas mencionados en la introducción.

## 2.1 ¿Qué es el CPD 3MARES?

Ya hemos mencionado este centro varias veces en esta memoria, sin embargo, aún no se ha propuesto una descripción formal y detallada.

Para poder ir de lo general a lo concreto, es necesario conocer la definición de Centro de Procesamiento de Datos.

*[1] Un centro de procesamiento de datos es una sala, edificio o instalación física que alberga infraestructura IT para crear, ejecutar y proporcionar aplicaciones y servicios, y para almacenar y administrar los datos asociados con esas aplicaciones y servicios.*

Como hemos mencionado en la introducción, la Facultad de Ciencias de la Universidad de Cantabria alberga un centro de estas características: El Centro de Procesamiento de Datos 3MARES. Este espacio alberga varios *clústeres* de cómputo HPC, sistemas de almacenamiento y virtualización de servicios, dedicados a impulsar la investigación científica y académica.

*Figura 1: Vista exterior del CPD 3MARES*

*Figura 2: Vista interior del CPD 3MARES*

En la *Figura 1* vemos la vista exterior de la sala, situada en uno de los patios interiores de la facultad, en la parte superior se observa parte de las canalizaciones y del equipamiento de refrigeración del CPD. La Figura 2 ofrece una vista dentro de la sala, en ella se ven los armarios o racks distribuidos en una configuración de pasillo caliente y pasillo frío aislados térmicamente para aumentar la eficiencia de la refrigeración.

El CPD 3MARES está compuesto por 30 racks, distribuidos en dos filas (A y B).

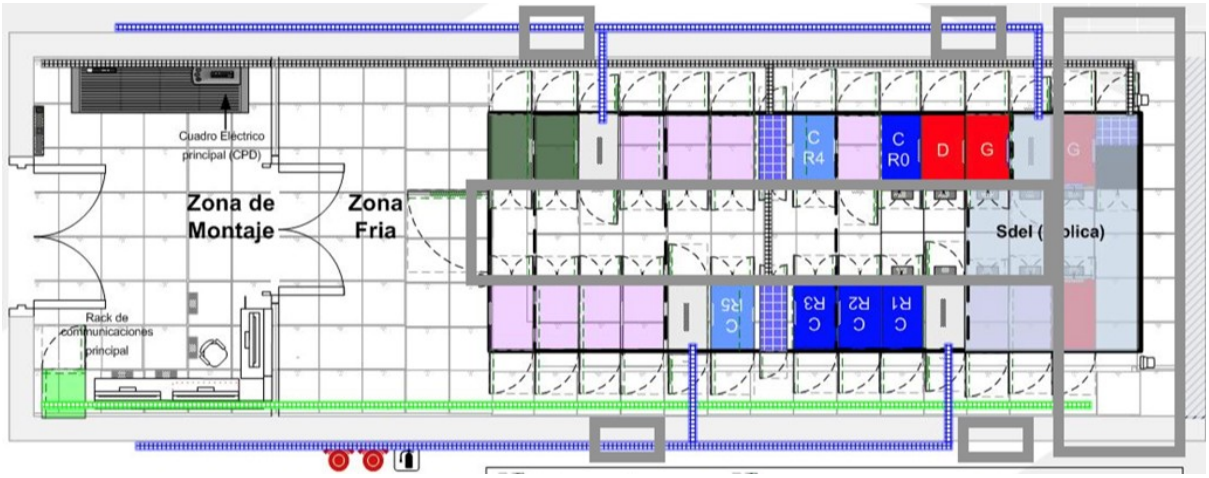


*Figura 3: Plano cenital del CPD 3MARES*

En la *Figura 3* se muestra el plano cenital de la sala, en él están representados los dos espacios físicamente aislados en los que se divide el CPD. Una sala de montaje donde realizar trabajos de mantenimiento apropiados, y separado físicamente, la sala fría donde está instalado el

cubo central con los 30 racks, dispuestos, como hemos comentado, en una distribución de pasillo central caliente.

Tras esta vista general del CPD 3MARES, se espera reflejar mejor la escala y complejidad a todos los niveles que requiere una instalación técnica de estas características.

## 2.1 ¿Para qué se sirve el CPD 3MARES?

Una instalación de este tipo requiere de trabajos continuos de mantenimiento, mejora, ampliación, actualización, etc. Todo esto, por supuesto, conlleva unos gastos considerables, tanto en recursos físicos, como, en dedicación del personal encargado del CPD. Por lo tanto, la pregunta clave que debemos hacernos es la siguiente, ¿Cómo se justifica la creación y el mantenimiento de esta infraestructura?

El CPD 3MARES tiene a día de hoy dos funciones diferenciadas, un 20% de los *racks* están reservados para el Servicio de Informática de la Universidad de Cantabria (SdeI), para la prestación de servicios comunes, y su respaldo (correo, web institucional, ... ). El otro 80% restante de los *racks* está destinado al alojamiento de infraestructura IT perteneciente a distintos grupos de investigación de la Universidad de Cantabria y contratos de investigación con empresas externas.


*Figura 4: Grupos de investigación usuarios del CPD 3MARES*

En la *Figura 4* se muestran algunos de los grupos de investigación y empresas privadas que hacen uso de la infraestructura de cómputo y servicios, alojados y mantenidos en el CPD 3MARES.

Hablando en términos absolutos, el CPD 3Mares aloja actualmente más de un centenar de servidores, varios sistemas de almacenamiento, servicios de virtualización, clusters de cómputo, etc. En la Figura 5 se muestran con un mayor grado de detalle algunos de los “números” del CPD.


*Figura 5: Datos sobre el CPD 3MARES*

Si nos fijamos bien en los datos de la *Figura 5*, quizá nos llame la atención la capacidad total para aproximadamente 1200 servidores, en especial cuando solo hay unos 170 instalados actualmente. Esto es debido principalmente a dos motivos:

Por un lado, cuando se diseñó y construyó el CPD 3MARES (año 2010), tanto la densidad como la eficiencia de computación de los servidores era mucho menor que la de hoy en día. Esto quiere decir que los servidores actuales ocupan menos espacio, consumen menos energía y producen menos calor en relación con su potencia de cómputo.

Es por esto que cuando se realizaron los cálculos para el diseño del CPD se sobredimensiona considerablemente a consecuencia de las mejoras en la eficiencia producidas en la última década y media.

El otro motivo es la financiación y la demanda por parte de los investigadores de la infraestructura de computación. Ambas métricas han ido creciendo con los años, pero se estima que aumenten aún más en los próximos años.

Como conclusión, el CPD 3MARES es una infraestructura de alojamiento, que proporciona a los investigadores de la Universidad de Cantabria un entorno y espacio de IT, gestionados en común para alojar y utilizar un entorno de computación científico de altas prestaciones. Esto tiene la consecuencia de ofrecer un servicio centralizado más eficiente en cuanto a recursos, tanto humanos, físicos y/o energéticos, y una independencia y autonomía de otras infraestructuras externas.

## 2.2 ¿Qué es el Servicio de Computación para la Investigación?

Originalmente, antes de la creación del Servicio de Computación para la Investigación (SCI), el CPD 3MARES se presentaba a los distintos investigadores y grupos de investigación como un servicio de alojamiento de servidores. Esto significa que, previo a la creación del SCI, solo se ofrecían los servicios básicos de alimentación eléctrica, refrigeración y espacio de alojamiento físico de servidores (*racks*). Por lo que las tareas de instalación, configuración e integración del equipamiento que se quisiera alojar en la instalación debía ser realizada por cuenta propia del investigador o grupo de investigación responsable de gestionar los equipos.

Es por ello, que los propios investigadores se plantean la necesidad de usar infraestructura de cómputo HPC, y para ello tengan que conocer y realizar el proceso de instalación y configuración requerido. Esta situación presentaba principalmente tres alternativas posibles ante el problema planteado:

1. Renunciar al uso de infraestructura de cómputo científico.
2. Realizar todo el proceso de manera autónoma.
3. Contratar personal con conocimientos técnicos para realizar esta tarea.

Como se puede ver, todas las alternativas son poco eficientes, e inabordables por los grupos de investigación. La primera supone con casi total seguridad un detrimento considerable de los resultados científicos de un grupo de investigación, debido a la dedicación de recursos humanos necesaria. La segunda, podría llamarse "*ruleta rusa".* Y la tercera, la más empleada en muchos casos, es una solución muy ineficiente.

Por este motivo, en el año 2020 se creó el Servicio de Computación para la Investigación. Un grupo de técnicos encargadas de realizar todas aquellas tareas necesarias para llevar a cabo el mantenimiento y mejora continua de toda la infraestructura presente en el CPD 3MARES:

- Mantenimiento físico de la sala.
- Mantenimiento físico de los servidores.
- Instalación y configuración de redes de interconexión.
- Instalación y actualización de sistemas operativos.
- Instalación, configuración y actualización de servicios.
- Instalación de aplicaciones.
- Monitorización de los sistemas (servidores, redes, frío, alimentación).
- Soporte a usuarios y resolución de incidencias.

Este Trabajo de Fin de Máster ha surgido a partir del trabajo realizado como miembro del Servicio de Computación para la Investigación, donde he llevado a cabo mi actividad profesional en los últimos años.

Tras exponer el contexto donde ha surgido y se ha desarrollado este proyecto, pasamos a exponer las razones por las cuales se justifica la necesidad y utilidad del trabajo realizado.

# Capítulo 3

# Antecedentes y objetivos del proyecto

Se ha descrito el entorno de trabajo, el CPD 3MARES, y también el grupo de trabajo, el Servicio de Computación para la Investigación (SCI). Y recordamos el objetivo principal de este proyecto, que es ampliar y rediseñar la red de altas prestaciones del CPD 3MARES para integrar nuevo equipamiento IT.

En este capítulo vamos a explicar brevemente los antecedentes que motivaron el planteamiento y el inicio del proyecto, así como los objetivos y beneficios que debemos obtener tras su realización.

## 3.1 Antecedentes

La infraestructura de red de altas prestaciones existente en el CPD 3MARES fue diseñada y adquirida para comunicar un pequeño clúster de cómputo (8 nodos) con un sistema de almacenamiento distribuido (14 nodos) que se instalaron en el CPD en el año 2019. Cinco años más tarde, se requiere llevar a cabo la ampliación de ambos sistemas. Esta ampliación, que se describe en detalle en el Capítulo 5, incluye 18 nuevos nodos de cómputo y 4 nuevos servidores de almacenamiento. Lo que supone duplicar el número de equipos interconectados actualmente por la red de altas prestaciones en el CPD3 MARES.

Debido a las características técnicas de la red de interconexión de altas prestaciones, descritas en el Capítulo 6, la interconexión de los nuevos servidores de cómputo y almacenamiento demanda una revisión completa de la red tanto en términos del número absoluto de equipos de interconexión (*switches*) como en la topología, configuración e interconexión de los propios *switches* y de los servidores.

La revisión técnica sobre la red de interconexión de altas prestaciones para poder albergar los nuevos sistemas de cómputo y almacenamiento se detalla posteriormente. De momento, en este apartado solo necesitamos conocer la conclusión obtenida a partir del análisis realizado, que no es otra que la necesidad de la ampliación y rediseño que motivaron el comienzo de este trabajo.

## 3.2 Objetivos del proyecto

Por último, es necesario definir en detalle los objetivos de este proyecto:

- Proporcionar conectividad entre la nueva red de altas prestaciones y la red física general del CPD 3MARES.

- Aumentar la escalabilidad para futuras ampliaciones. La red debe ser capaz de crecer y adaptarse a nuevas demandas sin comprometer el rendimiento ni la estabilidad.

- Eliminar puntos únicos de fallo. Es necesario identificar y eliminar cualquier punto único de fallo en la red, garantizando así una mayor disponibilidad y resiliencia en caso de fallos en el hardware o software.

- Añadir conexiones redundantes para el equipamiento que lo requiera. Para aumentar la robustez de la infraestructura, es necesario añadir redundancia en todos los niveles posibles, esto incluye cablear enlaces redundantes entre equipos finales y *switches*.

Con los objetivos claramente definidos, podemos continuar con el análisis del equipamiento de cómputo, almacenamiento y red de interconexión. Este análisis es crucial para seleccionar los componentes adecuados y diseñar una solución técnica que cumpla con las necesidades actuales y futuras del CPD 3MARES.

# Capítulo 4

# Descripción técnica de los sistemas de cómputo y almacenamiento actuales

Como se ha mencionado en capítulos anteriores de esta Memoria, el CPD 3MARES alberga actualmente más de un centenar de servidores físicos. Este equipamiento, perteneciente a los distintos grupos de investigación, presenta características muy heterogéneas debido principalmente a los siguientes motivos:

- Antigüedad del equipamiento.
- Financiación disponible por el grupo de investigación en el momento de la compra.
- Necesidades y casos de uso particulares.

Es por esto que el CPD 3MARES es un entorno de computación muy heterogéneo. Por suerte, no es necesario conocer en detalle todo este equipamiento. En los próximos apartados se describe lo relevante para el desarrollo de este proyecto. Esto es, toda la infraestructura de red presente en el CPD y también los sistemas actuales de cómputo y almacenamiento distribuido que están interconectados a través de la red de altas prestaciones.

Comenzamos describiendo el equipamiento de cálculo y almacenamiento. Recordamos que el objetivo de este proyecto es la mejora y ampliación de la red de interconexión de altas prestaciones, por esto, vamos a presentar una descripción simplificada del equipamiento relativo a los servidores o nodos.

## 4.1 *Clúster* de cómputo

El CPD cuenta con un *clúster* de cómputo tipo HPC interconectado mediante la red de altas prestaciones. El clúster hace uso de esta red para realizar cálculos pararlelos entre nodos, además de el procesado de datos contenidos en el sistema de almacenamiento distribuido, también interconectado a la red de altas prestaciones.

8 **nodos de cálculo** [2] con las siguientes características:

- 2 procesadores *Intel(R) Xeon(R) Gold (16C/32T)*
- 256GB RAM
- Interfaz de red *ethernet* 100 Gbps (red de altas prestaciones).
- Interfaz de red *ethernet* 10 Gbps (red general).
- [17,18] Interfaz de gestión y administración IPMI (*Intelligent Platform Management Interface*).

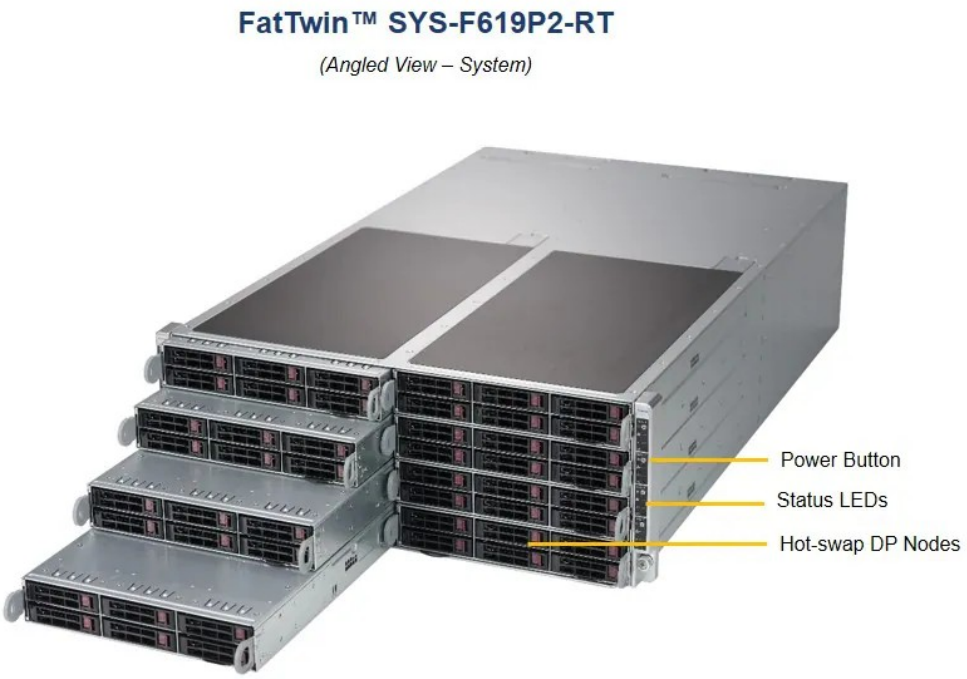


*Figura 6: Clúster de cómputo actual*

## 4.2 Sistema de almacenamiento distribuido

El CPD ofrece un servicio de almacenamiento distribuido basado en [19] *Lustre*, que aprovecha la red de altas prestaciones para las comunicaciones entre los servidores de almacenamiento.

El sistema está compuesto por 14 servidores, divididos en dos grupos en función del rol que desempeñan en el sistema de ficheros:

- MDS[1]: Servidores dedicados al almacenamiento, tratamiento y acceso a los **metadatos** del sistema de ficheros.

1El grupo MDS, en esta Memoria, se refiere tanto al rol MDS como MGS de Lustre

- OSS: Servidores dedicados al almacenamiento, tratamiento y acceso a los **datos** del sistema de ficheros.

La arquitectura del sistema de almacenamiento distribuido del CPD 3MARES está compuesto por 2 servidores de metadatos (MDS) y 12 servidores de datos (OSS) con las siguientes especificaciones técnicas.

**Servidor de metadatos (MDS)** [3]**:**

- 2 procesadores *Intel(R) Xeon(R) Silver (8C/16T)*.
- 384GB RAM.
- Interfaz de red ethernet 100 Gbps (red de altas prestaciones).
- Interfaz de red ethernet 10 Gbps (red general).
- Interfaz de gestión y administración IPMI.

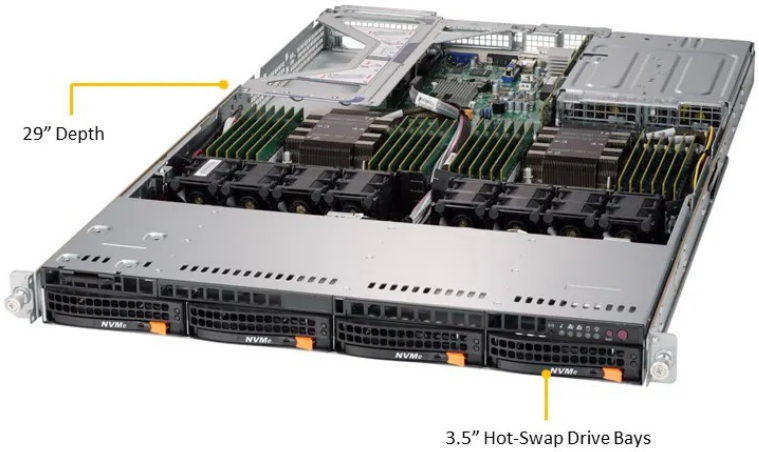


*Figura 7: Servidores de almacenamiento de metadatos (MDS)*

**Servidor de datos (OSS)** [4]**:**

- 2 procesadores *Intel(R) Xeon(R) Silver (16C/32T)*.
- 384GB RAM.
- Interfaz de red ethernet 100 Gbps (red de altas prestaciones).
- Interfaz de red ethernet 10 Gbps (red general).
- Interfaz de gestión y administración IPMI.


*Figura 8: Servidores de almacenamiento de datos (OSS)*

En resumen, en este capítulo hemos descrito los servidores de cómputo y almacenamiento interconectados mediante la red de altas prestaciones actualmente. Los detalles proporcionados son muy escuetos de manera intencional, ya que no es el alcance de esta memoria, describir en profundidad las características técnicas y tecnologías del equipamiento hardware, ni del software del sistema de ficheros *Lustre.*

*En el siguiente capítulo* encontramos una descripción similar del nuevo equipamiento que está por llegar al CPD 3MARES, que también hará uso de la red de altas prestaciones.

# Capítulo 5

# Descripción técnica de los nuevos sistemas de cómputo y almacenamiento

En el contexto del CPD 3MARES (Universidad de Cantabria), la adquisición de infraestructuras de estas características se realiza por medio de un concurso abierto y público.

En el momento de redactar esta Memoria, el contrato aún no ha sido adjudicado y, por lo tanto, se desconocen las características técnicas de modelos concretos del equipamiento.

Es por ello, que en este Capítulo se detallan los requisitos técnicos necesarios, para el equipamiento de los nuevos sistemas de cómputo y almacenamiento.

## 5.1 *Clúster* de cómputo

La ampliación del *clúster* de cómputo de alto rendimiento (HPC) del CPD 3MARES se compone de 18 nodos con las siguientes características:

- Dos procesadores de arquitectura x86-64 de la generación I*ntel Xeon Ice Lake* o *AMD Epyc Rome*, con una frecuencia de 2.0 GHz y 32 cores por procesador.
- 256GB de RAM tipo RDIMM a 3200 MHz, optimizando todos los canales de memoria disponibles.
- Dos dispositivos de almacenamiento tipo SSD de 480 GB configurados en RAID 1, diseñados para soportar altas tasas de transferencia y durabilidad (DWPD).
- Interfaz de red Ethernet de 25 Gbps [8] SFP28 (red de altas prestaciones).
- Interfaz de red Ethernet de 10 Gbps SFP28 (red general).

- Interfaz de red de administración compatible con IPMI 2.0 para control remoto.

## 5.2 Sistema de Almacenamiento Distribuido

El nuevo sistema de almacenamiento distribuido estará compuesto por 4 servidores con las siguientes características:

- 2 procesadores de arquitectura x86-64 de la generación *Intel Xeon Ice Lake* o *AMD Epyc Rome*, con una frecuencia de 2.0 GHz y 32 *cores* por procesador.
- 256GB de RAM tipo RDIMM a 3200 MHz, optimizando todos los canales de memoria disponibles.
- 2 dispositivos de almacenamiento tipo SSD de 480 GB configurados en RAID 1, diseñados para soportar altas tasas de transferencia y durabilidad (DWPD).
- 2 Interfaces de red Ethernet de 100 Gbps SFP28 y una conexión redundante de 10 Gbps SFP28.
- Interfaz de red Ethernet de 10 Gbps SFP28 (red general).
- Interfaz de red de administración compatible con IPMI 2.0 para control remoto.

En el capítulo a continuación, explicaremos con detalle las redes de interconexión y todas aquellas tecnologías relevantes para la ampliación y mejora de la red de altas prestaciones, que interconecta directamente el equipamiento descrito en el anterior y presente capítulo.

# Capítulo 6

# Revisión de las redes de interconexión presentes en el CPD 3MARES

En el CPD 3MARES existen dos redes físicas de interconexión: una red general y una red de altas prestaciones. Antes de describir los detalles de cada una de estas redes, es importante comprender los motivos por los que se emplean varias redes de interconexión entre los mismos equipos dentro del CPD.

## Rendimiento

La red de altas prestaciones está diseñada específicamente para manejar grandes volúmenes de datos con baja latencia y alto ancho de banda, lo cual es esencial para las tareas de cómputo intensivo y transferencia de datos entre los sistemas de almacenamiento y procesamiento. Si se utilizara una única red para todas las comunicaciones del CPD, aumentaría el riesgo de sobrecargarla, degradando el rendimiento de la red, perjudicando especialmente a las aplicaciones más demandantes en términos de ancho de banda y/o latencia.

Emplear varias redes, con la configuración adecuada, permite disgregar el tráfico y asegurar que las comunicaciones más demandantes (cómputo paralelo y entrada/salida) dispongan de todos los recursos (de la red de altas prestaciones) en exclusividad. Por otro lado, el resto de servicios del clúster, no ven su funcionamiento y/o rendimiento perjudicado debido a sobrecarga en la red por parte de aquellas aplicaciones mas demantantes puesto que estas usan la red específica de alto rendimiento.

## Seguridad

La presencia de varias redes también nos proporciona beneficios en cuanto a la seguridad. Al emplear redes separadas, logramos que la red de altas prestaciones esté menos expuesta a posibles ataques o fallos que podrían afectar a la red general. Esto supone una ventaja especialmente

importante, ya que esta red se encarga de las comunicaciones de entrada y salida con el sistema de almacenamiento.

## Tolerancia a fallos

La presencia de varias redes independientes aumenta significativamente la tolerancia a fallos del CPD. Si una de ellas experimenta un problema o falla, las otras pueden continuar operando sin interrupciones.

Además, en caso de fallo de una de las dos redes, aplicando la configuración apropiada, la red funcional podría asumir temporalmente algunas comunicaciones y servicios de la red en fallo, si así se desea en caso de que sea necesario. Esta característica mitiga el impacto cuando se produce algún fallo en una de las redes.

Por estos motivos, el CPD 3MARES está equipado con **dos redes físicas de interconexión** independientes y aisladas entre sí en este caso.

Como se menciona en la introducción de esta memoria, la red de altas prestaciones (LAN-HPC), abarca un número reducido de equipos en comparación con el tamaño total del CPD. A su vez, esta red de propósito específico, está destinada únicamente a las siguientes funciones:

- comunicaciones para realizar **cálculos paralelos** en el clúster de cómputo.
- comunicaciones de **entrada/salida** entre los nodos de cómputo con los servidores de almacenamiento.
- comunicaciones internas del **sistema de ficheros distribuido** entre los servidores de almacenamiento.

En contraposición, la red general (LAN-MNG) interconecta todos los equipos presentes en el CPD 3MARES y proporciona todos aquellos servicios básicos en un entorno de estas características:

- **Monitorización** en tiempo real del estado de los equipos y sistemas.
- Envío de **alertas** y notificaciones automáticas ante fallos, problemas de rendimiento, o cualquier otra anomalía detectada en los sistemas.

- Gestión remota **IPMI**.
- Gestor de **trabajos en batch** para el clúster de cómputo, organizando y priorizando las tareas para optimizar el uso del hardware disponible.
- Acceso remoto a través de **SSH** y otros métodos de acceso remoto que permiten a los administradores y usuarios gestionar y controlar los equipos desde ubicaciones fuera del CPD.
- Servicios de **autenticación y autorización** (i.e. FreeIPA).
- Distribución de software mediante **repositorios** locales accesibles mediante protocolos **HTTP o FTP**.
- Sistemas de **backup y recuperación** para la realización de copias de seguridad regulares y restauración de datos en caso de pérdida o corrupción.
- Acceso a recursos compartidos de almacenamiento en red a nivel de bloque (**iSCSI**) y a nivel de sistema de ficheros (**NFS**).
- **Servicios web** utilizados para albergar recursos y aplicaciones dentro del CPD.

En resumen, la separación de la red general y la red de altas prestaciones en el CPD 3MARES permite optimizar el rendimiento, mejorar la seguridad y aumentar la tolerancia a fallos. La red de altas prestaciones está dedicada a las tareas más exigentes, como el cómputo paralelo y el acceso al sistema de ficheros distribuido, garantizando que estas operaciones dispongan de los recursos necesarios sin interferencias. Mientras tanto, la red general da soporte a los servicios generales y de administración del CPD, asegurando que ambos flujos de tráfico se gestionen de manera óptima y segura.

En la siguiente sección se describen en detalle las tecnologías, arquitecturas y configuraciones que definen las redes presentes en el CPD 3MARES.

# 6.1 Red general del CPD 3MARES

Es la red principal del CPD 3MARES que interconecta todos los equipos entre sí. Como hemos visto en el *Capítulo 5*, esta red proporciona la conectividad necesaria para desplegar todos los servicios y funcionalidades presentes en el CPD.

## 6.1.1 Tecnologías

Antes de describir el equipamiento y la topología es necesario presentar varias tecnologías de red que tienen un papel fundamental en su comportamiento, y, por lo tanto, también en su diseño.

### *Link Aggregation Groups [12,13,14]*

La primera de estas tecnologías es la conocida como LAG (*Link Aggregation Group*). LAG permite combinar múltiples enlaces de red físicos en un solo enlace lógico. Al agrupar varios cables o puertos de red en un LAG, se puede aumentar el ancho de banda disponible y proporcionar redundancia. Esto significa que, si uno de los enlaces físicos falla, el tráfico se redirige automáticamente a través de los enlaces restantes, manteniendo la conectividad sin interrupciones.

Para facilitar la agregación dinámica de enlaces físicos, la agregación de enlaces requiere la implementación del protocolo LACP (*Link Aggregation Control Protocol*). El protocolo LACP es parte del estándar IEEE 802.3ad, se utiliza comúnmente para gestionar LAGs. LACP coordina la configuración automática de los enlaces agregados, asegurando que todos los enlaces dentro del grupo estén funcionando correctamente y optimizando el balanceo de carga entre ellos.

En conclusión, la tecnología LAG permite establecer enlaces múltiples que se comporten como un solo enlace lógico. Es una solución efectiva para mejorar la capacidad de red y la tolerancia a fallos, haciendo que sea una opción popular en entornos donde la alta disponibilidad y el rendimiento son cruciales.

La limitación del protocolo LAG, es que se limita a la agregación de enlaces entre un elemento de red (switch) y un nodo (servidor).

De la necesidad de agrupar enlaces de red físicos entre múltiples *switches* surge la tecnología MC-LAG (*Multi Chassis-LAG*), que permite agrupar varios *switches* para que actúen como un solo elemento lógico.

En un entorno MC-LAG, el protocolo LACP coordina los enlaces entre múltiples *switches*, garantizando que el tráfico se distribuya de manera equilibrada y que se puedan realizar ajustes dinámicos si un enlace falla o se añade uno nuevo.

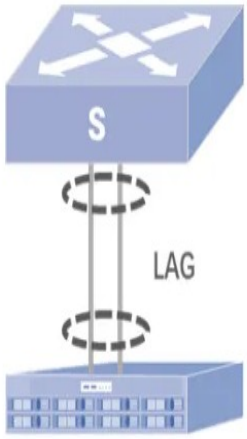


*Figura 9: Link Aggregation Group*

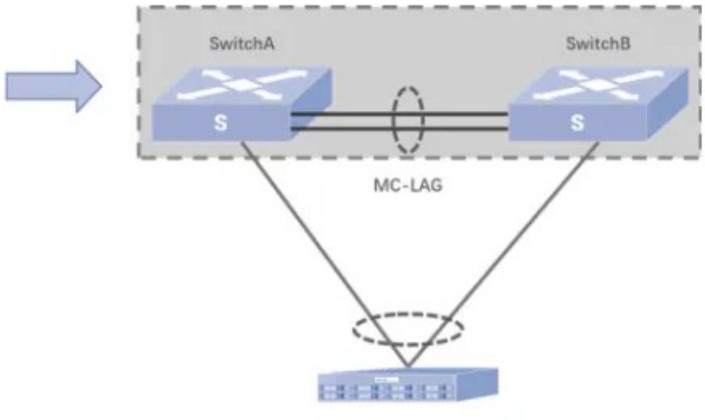


*Figura 10: MultiChassis - Link Aggregation Group*

Como vemos en la *Figura 9*, el protocolo LACP permite crear agrupaciones LAG entre un *switch* y un servidor. En la *Figura 10* podemos ver la diferencia con el mecanismo MC-LAG, que permite agrupar varios switches en un único elemento lógico.

Por último, es importante destacar que MC-LAG no es un protocolo estándar, lo que significa que su implementación puede variar dependiendo del fabricante de los dispositivos de red. Cada proveedor puede tener su propia versión de MC-LAG con características y configuraciones específicas.

## Redes virtuales de área local [15]

En comparación con MC-LAG, las redes virtuales de área local, o *Virtual Local Área Network* (VLAN), es una tecnología mucho más conocida y utilizada en gran variedad de redes de interconexión entre dispositivos. Esto puede incluir desde redes WLAN (*Wireless Local-Area Network)* domésticas hasta centros de computación de alto rendimiento, como es en nuestro caso. En este apartado, vamos a tratar de explicar a alto nivel que son las VLANs y por qué nos beneficiamos de su uso en el entorno del CPD.

En el CPD 3MARES, la red física general o LAN-MNG es la encargada de proporcionar todos los servicios mencionados con anterioridad en este capítulo, desde la monitorización hasta el acceso remoto. Esto presenta varios problemas en cuanto a la seguridad, rendimiento, y facilidad de administración de los equipos del CPD. Por suerte, existen varias funcionalidades y protocolos de red que podemos configurar en los *switches* para solucionar o mitigar las desventajas de esta situación tan común.

Lo que queremos en estos casos es dividir todo ese tráfico que comparte una misma infraestructura de red en diferentes redes, y las VLANs nos permiten hacer justamente eso. El modo de conseguirlo es mediante el etiquetado (por parte de los swtiches) de los paquetes en función de la VLAN o VLANs configuradas en el puerto de origen, con esa etiqueta, los switches reenvian los paquetes únicamente hacia los puertos configurados con la VLAN o VLANs correspondientes. Creando la ilusión de tener redes físicamente independientes dentro de un mismo switch. Aplicando esta configuración en varios switches podemos extender esta funcionalidad a toda la red física general del CPD 3MARES.

Al segmentar el tráfico en redes lógicas separadas, podemos mejorar la seguridad al aislar flujos de datos críticos, optimizar el rendimiento al reducir la congestión, y hacer que la red sea más fácil de administrar. En resumen, las VLANs son la herramienta que necesitamos para mantener nuestra red organizada, eficiente y segura.

Por lo tanto, el uso de VLANs en un entorno como el del CPD3 3MARES es realmente importante por los motivos:

- **Facilidad de administración:**

  Las VLANs permiten una administración más sencilla de la red, ya que facilitan la organización y segmentación lógica de los dispositivos conectados. Esto simplifica las tareas de configuración, monitorización y mantenimiento, permitiendo a los administradores gestionar la red de manera más eficiente sin necesidad de realizar cambios físicos en la infraestructura.

- **Reducción del tráfico en la red:**

  Las VLANs ayudan a reducir el tráfico general en la red al segmentar el tráfico de datos en grupos específicos de dispositivos. Esto asegura que solo los datos necesarios circulen entre los dispositivos de una misma VLAN, disminuyendo la carga total de la red y mejorando su eficiencia.

- **Aplicación de políticas de seguridad:**

  El uso de VLANs permite la implementación y el refuerzo de políticas de seguridad más estrictas. Al segmentar la red, se puede controlar mejor el acceso a recursos específicos y aislar el tráfico de datos sensibles, lo que contribuye a proteger la información crítica y a minimizar el riesgo de acceso no autorizado.

Tras describir estas tecnologías podemos realizar la descripción del equipamiento y la topología de la red de interconexión general del CPD 3MARES.

## 6.1.2 Equipamiento

Cada rack está equipado con dos switches [16] *ToR (Top of Rack)*:

- [5] 1 switch modelo Extreme Networks 5420M-24T con las siguientes características:
  - 24 puertos 1000BASE-T (1Gbps).
  - 4 puertos uplink de tipo SFP28, que soportan velocidades de 1/10/25Gb Gbps.

- [6] 1 switch modelo Extreme Networks 7520-48Y-8C con las siguientes características:
  - 48 puertos tipo SFP28 y que soporta velocidades de 1/10/25 Gbps.
  - 8 puertos troncales de tipo QSFP28, que soportan velocidades de 40/100 Gbps.

*Figura 12: Switch* Extreme Networks 5420M-24T

*Figura 11: Switch Extreme Networks 7520-48Y-8C*

Ambos switches presentan especificaciones técnicas muy distintas tanto en el número como en las velocidades de los puertos. Otra diferencia significativa entre ambos *switches* ToR es el tipo de los puertos de agregación. El *switch* más pequeño y de menos prestaciones emplea puertos 1000BASE-T mientras que el otro *switch*, de mayores prestaciones, emplea puertos de tipo SFP28.

El tipo de puerto no sólo influye en las velocidades soportadas, sino también en el cableado de interconexión entre el switch y los equipos del CPD. El *switch* 5420M-24T emplea cables de cobre de par trenzado frente a los cables de fibra óptica usados por el 7520-48Y-8C.

La interconexión entre los diferentes *racks* está realizada mediante un conmutador de distribución central (situado en el *rack* B8). Este switch lógico está formado por dos equipos idénticos (Extreme Networks 7520-48Y-8C), similares a los switches ToR. Ambos switches de distribución están interconectados mediante dos enlaces [11] DAC (Direct Attach Copper) y forman, mediante la tecnología MC-LAG, un único elemento lógico.

A continuación se describe la topología de la red física general del CPD.

### 6.1.3 Topología

Como sabemos, es la red principal del CPD 3MARES que interconecta todos los equipos entre sí. Su propósito es el de proporcionar conectividad y garantizar todas las comunicaciones necesarias para el funcionamiento adecuado de aplicaciones y servicios dentro del CPD.

En cada Rack, los dos *switches* ToR están interconectados mediante dos enlaces DAC formando un único elemento lógico mediante el uso de un protocolo tipo MC-LAG. Conectados  mediante 2 enlaces de fibra óptica al conmutador central  de distribución de todas las redes físicas, alojado en el rack B8.

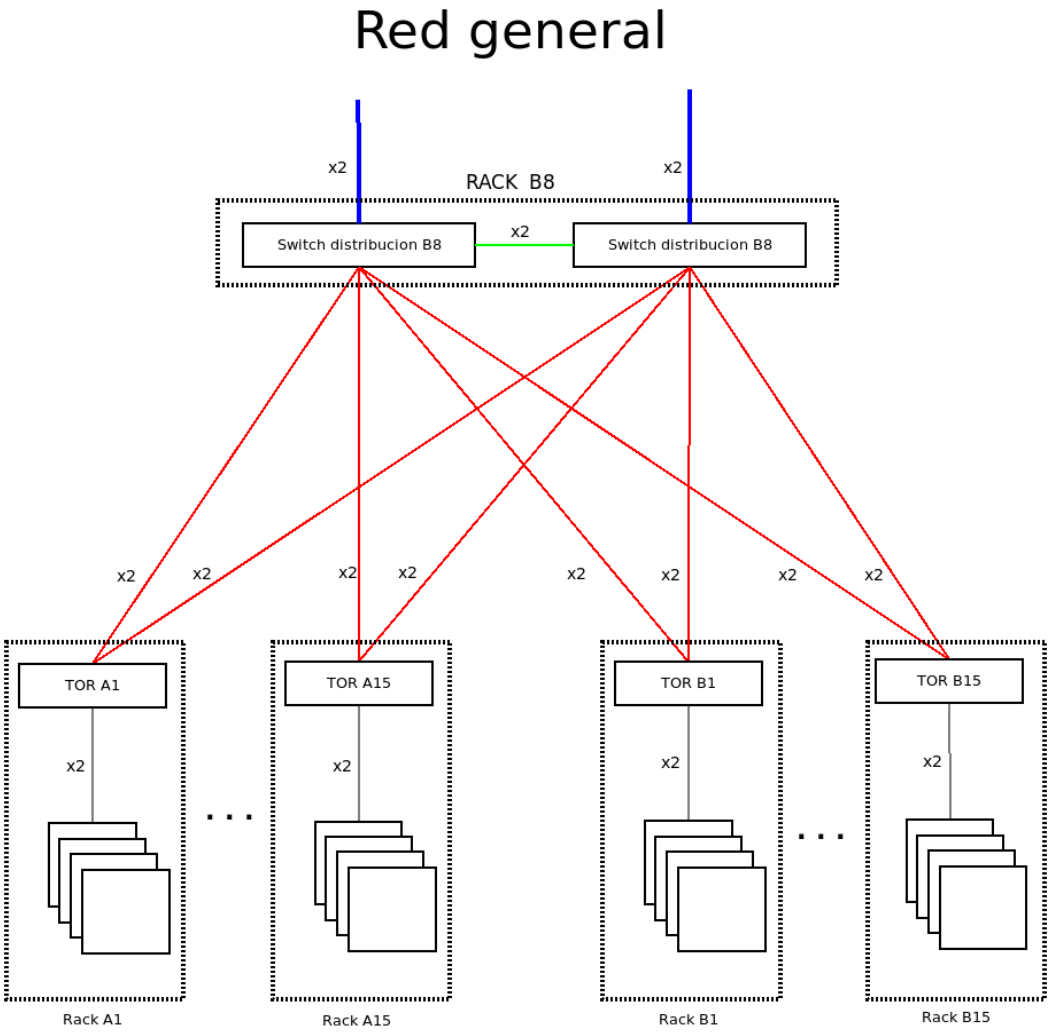


*Figura 13: Topología de la red general del CPD 3MARES*

En la *Figura 16* se muestran los 2 switches de distribución en vez de agruparlos en un único elemento, como hemos hecho con los ToR. Esto es para reflejar de manera más clara el cableado de alta disponibilidad entre los ToR y la distribución del CPD.

Este diseño se conoce como [10] S*pine-Leaf*. Esta topología, comúnmente utilizada en CPDs, se caracteriza por su organización en dos capas: una *spine* (columna vertebral) y otra *leaf* (hoja).

***Spine***
Consiste en un conjunto de *switches* de alto rendimiento situado en la capa superior de la jerarquía. Todos los switches *leaf* se conectan a cada switch *spine*, creando una red completa y redundante.

***Leaf***
Estos switches están conectados a los dispositivos finales, como servidores de cómputo y almacenamiento. Cada switch **leaf** se conecta a todos los switches **spine**, lo que asegura que cualquier dispositivo en la red pueda comunicarse con cualquier otro a través de un número predecible de saltos.

Las principales ventajas de la topología *Spine-Leaf* son, el ancho de banda, la baja latencia, la redundancia y la gran escalabilidad. Esto la convierte en una elección ideal para entornos donde se requiere alta disponibilidad y escalabilidad, como en entornos modernos como el CPD 3MARES.

## 6.1.4 Configuración de red

En esta red física de interconexión hay configuradas cuatro redes lógicas, VLANs, que separan el tráfico de red en función de su tipología y procedencia:

- VLAN de **Monitorización** empleada para las comunicaciones de los servicios de monitorización y alertas del equipamiento de alimentación y distribución eléctrica del CPD.
- VLAN de **Gestión Remota** empleada para el acceso y las comunicaciones mediante el protocolo IPMI de todo el equipamiento IT del CPD.
- VLAN **interna** empleada para todas las comunicaciones de aplicaciones y servicios entre todos los servidores del CPD.

- VLAN **externa** empleada para la conexión con redes externas y permitir el acceso con redes externas al CPD 3MARES.

Todos los switches pertenecientes a la red física general del CPD 3MARES, tanto los switches ToR (Top of Rack) como los de distribución, están configurados adecuadamente para soportar las VLANs mencionadas. Cada switch ha sido configurado para manejar las VLANs de Monitorización, Gestión Remota, Acceso Interno y Acceso Externo, garantizando que cada tipo de tráfico se mantenga aislado y dirigido a los recursos correspondientes.

A continuación, se detallan las redes IP configuradas dentro de cada una de las VLANs, que permiten gestionar el tráfico de datos en cada segmento de la red. Cada VLAN tiene designada su propia red IP, facilitando la administración de la red al permitir un control más granular y organizado de las distintas comunicaciones y servicios.

## Red de monitorización eléctrica

Red privada para la monitorización que interconecta las interfaces ethernet de todos los dispositivos de alimentación y distribución eléctrica como SAIs (Sistemas de Alimentación Ininterrumpida) y PDUs *(Power Distribution Unit)*.

## Red de gestión remota

Red privada que interconecta los dispositivos IPMI o equivalentes para la telemetría y gestión remota de todo el equipamiento del CPD3M.

## Red interna

Red privada que interconecta todos los elementos del clúster de cómputo, permitiendo el despliegue de nodos, acceso SSH para la administración de servidores, tráfico de control de SLURM, gestión de usuarios, servicios internos del clúster, etc.

## Red externa

Red pública destinada a la conexión con redes WAN externas y/o ṕublicas.

En resumen, todas las redes IP mencionadas emplean la red física general con la diferencia de que las redes de monitorización y gestión remota están conectadas y configuradas mediante VLANs para utilizar el switch Extreme Networks 5420M-24T y las redes de acceso interno y externo para usar el switch Extreme Networks 7520-48Y-8C de la misma manera.

Finalmente, procedemos a continuación con la revisión de la red de altas prestaciones actualmente presente en el CPD 3MARES.

## 6.2 Red de altas prestaciones

Como mencionamos en la introducción de este capítulo, esta red física de propósito específico interconecta el sistema de almacenamiento distribuido con un pequeño clúster de cómputo.

### 6.2.1 Equipamiento

La red está formada por dos switches de tecnología Ethernet modelo [7] Mellanox/Nvidia SN2100 con las siguientes especificaciones:

- ➔ 16 puertos tipo QSFP28 y que soporta velocidades de 100 Gbps.


*Figura 14: Switch Mellanox SN2100*

### 6.2.2 Topología

Con un simple vistazo del esquema de la topología, presentada en la *Figura 15*, se aprecia la simplicidad de esta red. Analizando los detalles, vemos como cada servidor de almacenamiento y nodo de cómputo está conectado a un único switch SN2100. Sin enlaces redundantes. En cuanto a la interconexión entre los dos switches, se realiza mediante dos enlaces de fibra óptica.

Originalmente, se optó por esta topología debido a dos motivos principalmente.

**Financiación**

Como no podía ser de otra manera, el presupuesto siempre juega un papel fundamental en el contexto que nos encontramos. La solución escogida se corresponde con la implementación más barata, prescindiendo de enlaces redundantes entre servidores y switches y usando una topología de un solo nivel, minimizando el número de conmutadores necesarios.

**Propósito y alcance de la red**

No todos los motivos tienen que ver con el dinero, esta topología, aunque sencilla y sin redundancia, cumplía con las necesidades de una red, que originalmente, interconectaba un pequeño número de equipos de cómputo con un sistema de almacenamiento distribuido. La alta disponibilidad y resiliencia ante fallos no era un requisito de la red en el momento de su diseño y adquisición.

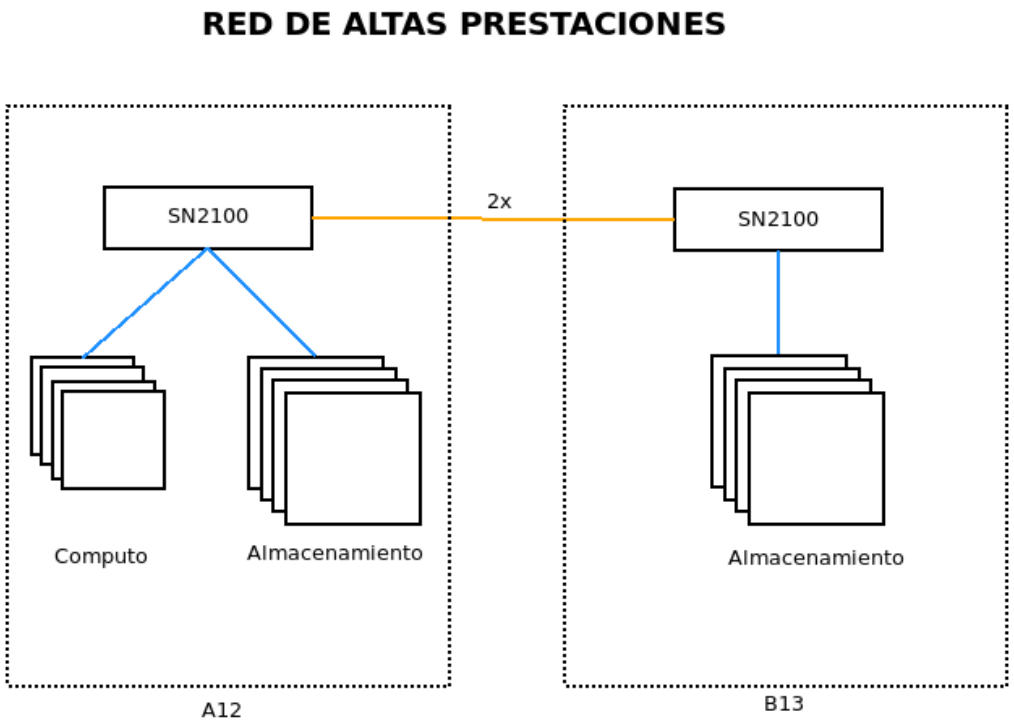


*Figura 15: Topología de la red de altas prestaciones del CPD 3MARES*

No ha sido hasta el comienzo de este proyecto, motivado por la ampliación de la infraestructura de cómputo y almacenamiento, que deseamos proporcionar al CPD de una red de altas prestaciones con una serie de mejoras mencionadas a lo largo de esta memoria.

**Escalabilidad:**

Una de las principales limitaciones de la topología actual es su falta de escalabilidad. Dado que cada servidor de almacenamiento y nodo de cómputo está conectado a un único switch SN2100, la capacidad para expandir la red de manera eficiente es limitada. A medida que crece la demanda de recursos y se añaden más dispositivos, la red puede experimentar cuellos de botella, ya que la adición de nuevos switches o la reconfiguración de la topología existente podría requerir intervenciones complejas y costosas. Esta falta de escalabilidad puede dificultar la capacidad del CPD 3MARES para adaptarse a futuras necesidades de crecimiento sin un rediseño significativo de la infraestructura de red.

**Redundancia y resiliencia ante fallos:**

Otra desventaja crítica es la ausencia de redundancia en las conexiones de los nodos de cómputo y almacenamiento. En la configuración actual, cada dispositivo está conectado a un único switch, lo que significa que un fallo en un switch o en un enlace podría causar una interrupción significativa en el servicio para los dispositivos conectados a ese switch. Esta falta de redundancia expone a la red a riesgos mayores de disponibilidad, lo que puede afectar negativamente a las operaciones críticas del CPD 3MARES.

# Capítulo 7

# Propuesta técnica de mejora de la red de altas prestaciones

Finalmente, tras el trabajo realizado de recopilación, procesado, análisis y exposición de todos los componentes en general, y de la red en particular, que tienen relación con el objetivo final de este trabajo. Podemos definir el proyecto de ampliación y mejora de la red de altas prestaciones del CPD 3MARES.

Antes, es necesario recordar los objetivos que definimos para la nueva infraestructura de red de altas prestaciones:

**Ampliar el número de equipos interconectados:**

El objetivo es aumentar la capacidad de la red para interconectar más nodos de cómputo y dispositivos de almacenamiento. Esto permitirá que el CPD 3MARES pueda soportar un mayor número de aplicaciones y servicios, además de facilitar la expansión futura y el crecimiento de la infraestructura sin comprometer el rendimiento.

**Mejorar la tolerancia ante fallos:**

La nueva infraestructura de red debe ser más resiliente, capaz de soportar fallos en enlaces o dispositivos sin interrumpir las operaciones. Esto se logrará mediante la implementación de tecnologías redundantes y configuraciones que aseguren la continuidad del servicio, en caso de fallo en alguno de los *switches*.

**Hacer un uso más eficiente de los recursos:**

Optimizar el uso del ancho de banda y otros recursos de red es crucial para maximizar el rendimiento. La nueva red se diseñará para equilibrar la carga de trabajo entre los enlaces disponibles, minimizando la latencia y evitando cuellos de botella.

**Proporcionar conectividad entre esta red y la red general del CPD:**

Es necesario proporcionar la interconexión entre la red de altas prestaciones y la red general del CPD. Esto permitirá que los servicios y aplicaciones puedan comunicarse de manera efectiva, manteniendo la

flexibilidad y la seguridad necesarias para un entorno complejo como el del CPD 3MARES.

## 7.1 Topología de la red

La topología de la nueva red de altas prestaciones sigue el modelo *Spine-Leaf*, al igual que la red general del CPD 3MARES, aprovechando su escalabilidad y eficiencia, características esenciales para un entorno de alto rendimiento.

En esta arquitectura, la red se organiza en dos capas: la capa **Spine**, que actúa como la columna vertebral de la red, y la capa **Leaf**, que conecta directamente a los dispositivos finales, como servidores de cómputo y sistemas de almacenamiento.

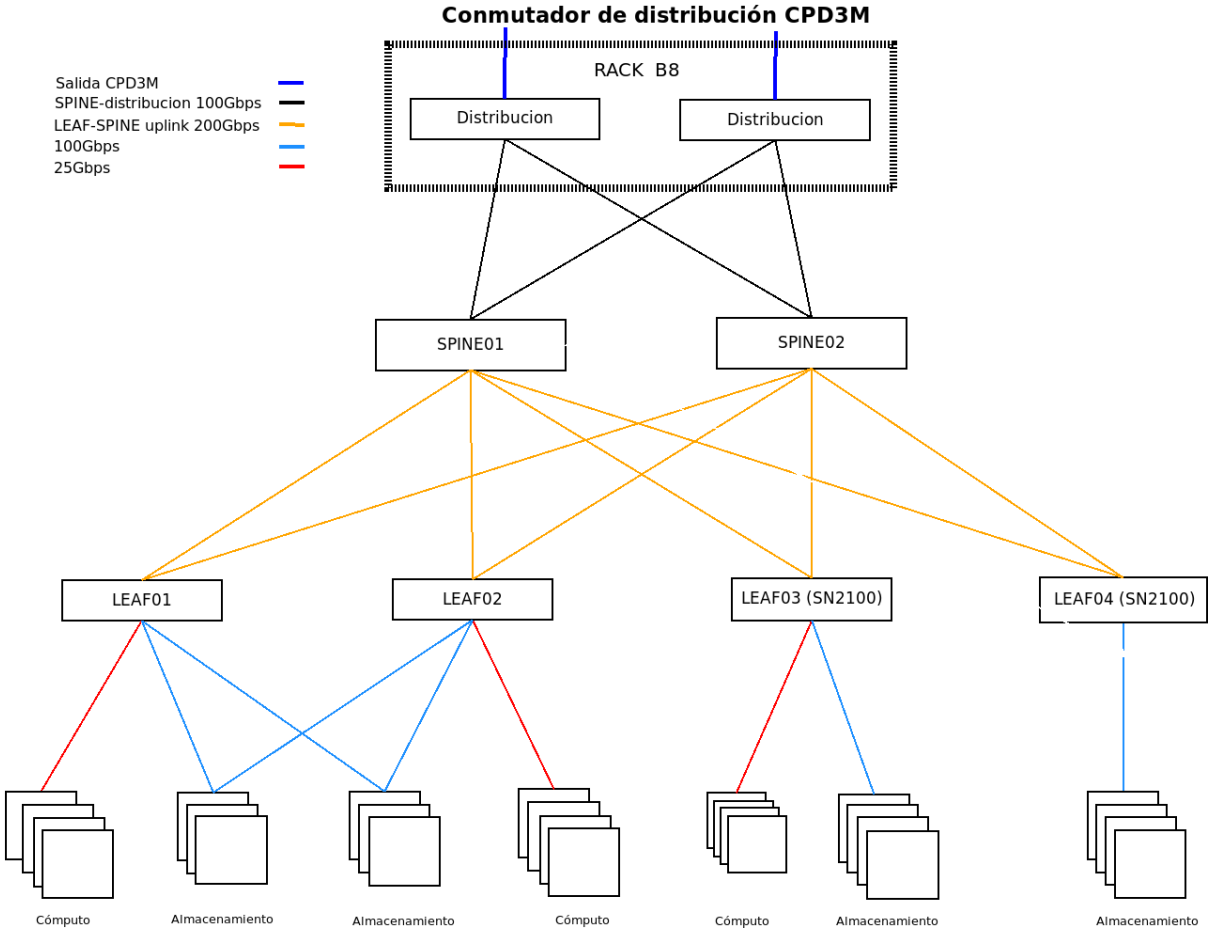


*Figura 16: Topología de la propuesta de red de altas prestaciones*

## Capa *Spine*

Está compuesta por dos *switches* idénticos (spine01 y spine02) situados en el rack B8. Estos están conectados hacia abajo en la jerarquía con los *switches leaf,* y hacia arriba con los switches de distribución del CPD 3MARES. Ambos switches spine01 y spine02 estarán interconectados entre sí mediante enlaces DAC y configurados para formar un único elemento lógico mediante un protocolo tipo MC-LAG.

## Capa *Leaf*

Esta capa está compuesta por cuatro switches *Leaf* (leaf01-leaf04), situados en los racks donde se alojan los servidores de cómputo y de almacenamiento. Estos switches son los encargados de conectar directamente los dispositivos finales a la red, proporcionando una conectividad de alto rendimiento. Cada switch *Leaf* está conectado a ambos switches *Spine*, evitando cuellos de botella, maximizando el uso del ancho de banda disponible y eliminando puntos únicos de fallo en la red.

Como puede verse en la *Figura 16*, la capa Leaf está compuesta por cuatro switches, dos de ellos (leaf03 y leaf04) son los Mellanox sn2100 que forman la red de altas prestaciones actual, reutilizados y combinados con dos nuevos switches (leaf01 y leaf02) de prestaciones y características similares.

Por último, antes de dar por finalizado este apartado relativo a la topología de la nueva red de altas prestaciones, vamos a describir los detalles necesarios para realizar el cableado entre los switches y los servidores.

## Interconexión de los equipos de cómputo y almacenamiento

Como se muestra en la *Figura 16*, el equipamiento que conforma el nuevo sistema de almacenamiento se conecta de manera redundante a dos switches *Leaf* diferentes. Cada equipo tiene un enlace a un switch de altas prestaciones dentro de su mismo rack y un enlace adicional a un switch *Leaf* en un rack vecino. Esta decisión de diseño está justificada por la necesidad de garantizar alta disponibilidad y tolerancia a fallos.

Al conectar los equipos de esta forma, se asegura que, en caso de fallo de uno de los switches o enlaces, el equipo siga teniendo conectividad a través del otro enlace.

En cuanto a los equipos de cómputo y el sistema de almacenamiento actual, se conectarán al switch Leaf de su propio rack con un enlace no redundante. Esta decisión se toma en base a no ocupar puertos en los switches Leaf de manera innecesaria, o mejor dicho, cuando este uso no esté justificado por las necesidades de las aplicaciones y servicios que se ejecutarán en los servidores.

A continuación se describe la interconexión entre la nueva red de altas prestaciones con la red general del CPD.

## 7.2 Interconexión con la Red General del CPD 3MARES

Uno de los objetivos de la nueva infraestructura de red de altas prestaciones es su integración con la red general del CPD 3MARES. A diferencia de la configuración anterior, donde la red de altas prestaciones estaba físicamente aislada, la nueva topología propuesta permite una interconexión entre ambas redes a través de los *switches* de distribución del CPD.

Esta integración se logra mediante la configuración adecuada de los switches de distribución, utilizando las técnicas de enrutamiento routing y VLANs necesarias. Gracias a esta configuración, se puede establecer una comunicación entre las dos redes físicas y sus respectivas subredes lógicas. Esto proporciona una flexibilidad adicional, ya que permite interconectar completamente los recursos de la red de altas prestaciones con los de la red general, cuando sea necesario.

En conclusión, con esta nueva arquitectura de red, el CPD 3MARES se beneficiará de una infraestructura más unificada y versátil. Permitirá que los equipos conectados a la red de altas prestaciones interactúen con los servicios y dispositivos de la red general sin las limitaciones que imponía la separación anterior.

En el apartado a continuación, se definen las características técnicas que deben cumplir los nuevos switches y demás componentes de la nueva red de altas prestaciones.

## 7.3 Especificaciones técnicas del nuevo equipamiento de red

Definir estas especificaciones es imprescindible para asegurar que el equipamiento adquirido satisfaga plenamente las necesidades, objetivos y expectativas del proyecto que hemos descrito a lo largo de esta memoria.

A continuación, se describen las características técnicas requeridas tanto para los switches Spine cómo Leaf:

1. **Matriz de conmutación sin sobresuscripción:**

   Esto garantiza que cada puerto pueda operar a su máxima capacidad, de manera concurrente, asegurando que no haya limitaciones en el rendimiento de la red, especialmente bajo condiciones de alta carga.

2. **Soporte del protocolo IEEE 802.3ad (LACP):**

   El soporte de LACP es necesario para la agregación de enlaces, lo que como hemos explicado anteriormente, permite combinar múltiples conexiones físicas en un solo enlace lógico. Esto añade redundancia, mejorando la disponibilidad y tolerancia a fallos de la red.

3. **Soporte para los protocolos de enrutamiento estándares (OSPFv2, OSPFv3, ISIS, … ):**

   Estos protocolos de enrutamiento son necesarios para la propagación automática de rutas.

4. **Compatibilidad con [9] RoCE (RDMA over Converged Ethernet):**

   *Remote Direct Memory Access (RDMA)* es una tecnología que permite implementar el principio Direct Memory Access (DMA) entre servidores interconectados mediante una red de intercopnexión.

   DMA permite a un computador mover datos entre la memoria principal y los periféricos sin involucrar al procesador, lo que libera a la CPU para realizar otras tareas. RDMA lleva este concepto a un nivel superior, permitiendo que un servidor acceda directamente a la memoria de otro a través de la red, RDMA over Converged Ethernet (RoCE) implementa esta funcionalidad a través de redes Ethernet.

   Es necesario que el nuevo equipamiento de red implemente esta funcionalidad por dos motivos: rendimiento y compatibilidad con el sistema de ficheros distribuido lustre actual.

En cuanto a las características concretas que deben tener los switches de la capa Spine y los de la capa Leaf:

**Spine**
Los switches Spine deben tener puertos con velocidades de 200 Gbps, con un número suficiente de puertos para los requisitos actuales de conectividad y un 50% adicional de puertos libres para futuras ampliaciones. Esto permite que la red pueda escalar sin necesidad de ampliar el equipamiento de esta capa, simplificando futuras ampliaciones en la red de altas prestaciones.

**Leaf**
Los dos nuevos switches Leaf deben soportar velocidades de 100 Gbps y tener suficientes puertos para cumplir con las necesidades de conectividad actuales (sistemas de cómputo y almacenamiento nuevos y actuales).

Por último, para completar la propuesta desarrollada en este proyecto, vamos a describir el proceso de instalación y configuración de todo el nuevo equipamiento de red.

## 7.4 Instalación, cableado y configuración necesaria

Este apartado es, de alguna manera, un corolario de todas las instrucciones relativas a la interconexión física y a la configuración software, que hemos ido mencionando a lo largo de este capítulo.

### Instalación física del equipamiento

Por motivos de distribución de consumo eléctrico, refrigeración, organización lógica y cableado, se tendrán en cuenta las siguientes instrucciones en el momento de instalación del nuevo equipamiento de red de altas prestaciones en el CPD 3MARES:

- Se instalará cada uno de los nuevos switches Leaf en un rack distinto, donde también se repartirá el equipamiento de cómputo y almacenamiento adquirido.
- Se instalarán en el rack B8 los dos switches Spine suministrados. Estos switches compartirán el armario con el conmutador de distribución del CPD3M. De esta manera, el cableado entre las dos redes físicas (general y alto rendimiento) se realiza dentro de un mismo rack, lo que abarata y simplifica el cableado.

## Cableado del equipamiento

Se requiere realizar todo el cableado descrito en el apartado dedicado a la topología de la nueva red, siguiendo el esquema descrito en la *Figura 16*.

Para ello es necesario liberar puertos del switch SN2100 instalado en el rack A12, por estar actualmente todos ocupados. Para liberar estos puertos se requiere el uso de dos cables breakout de puertos 100G a 4 x 25 Gbps, de esta forma, todos los nodos de cómputo se conectarán a una velocidad de 25 Gbps a la red de altas prestaciones.

- Se realizará una conexión desde cada switch SPINE hacia cada uno de los switches de distribución, como se muestra en la Figura 16. Para dichos enlaces se utilizarán los puertos troncales/Uplink de los switches de distribución.

## Configuración software del equipamiento

Para la integración con la red general del CPD, es necesario llevar a cabo la siguiente configuración:

- Realizar la configuración de VLANs necesaria para la comunicación entre las diferentes redes del CPD3M.

Llegado este punto, se ha descrito el proceso y las características del equipamiento, instalación y configuración del hardware y software necesario para llevar a cabo la ampliación y mejora de la red de altas prestaciones del CPD, cumpliendo con los objetivos marcados al inicio de esta memoria.

# Capítulo 8

# Conclusiones

Este trabajo de fin de máster ha surgido como respuesta a la necesidad de mejorar y ampliar la red de altas prestaciones del CPD 3MARES, asegurando que todas las fases del diseño y configuración de la red estuvieran alineadas con las necesidades específicas de este entorno. El crecimiento progresivo en la demanda de recursos de cómputo y almacenamiento ha hecho evidente que la infraestructura actual, aunque funcional, ya no es suficiente para soportar los requisitos operativos y las expectativas de futuro del CPD.

Para abordar esta necesidad, el trabajo comenzó con una recopilación, estudio y análisis detallado de las arquitecturas de las redes de interconexión presentes en el CPD 3MARES. Se analizaron tanto la red física general como la red de altas prestaciones actual, esta segunda en particular, evaluando sus limitaciones en cuanto a la escalabilidad, redundancia y rendimiento de la infraestructura. Este análisis permitió comprender las capacidades de las infraestructuras actuales y definir los objetivos y mejoras que se proponen para la propuesta realizada.

Con base en este análisis, se procedió a la elaboración de la propuesta técnica para la ampliación y mejora de la nueva red de altas prestaciones, resultado de este proyecto. Esta propuesta incluye la determinación de las características del hardware necesario, la topología y las interconexiones entre los distintos componentes, así como la instalación y configuración del nuevo equipamiento.

Quiero recalcar finalmente que, en el momento de la redacción de esta memoria, no ha sido posible realizar la compra mediante concurso público del equipamiento descrito en la propuesta. Sin embargo, sea cual sea su implementación concreta, hemos trabajado y aprendido muchísimo durante todo el ciclo de vida de este proyecto.

# Bibliografía


[1] IBM. "Centros de datos". https://www.ibm.com/es-es/topics/data-centers, 2024.

[2] Supermicro. "Supermicro SYS-F619P2-RTN." https://www.supermicro.com/en/products/system/4U/F619/SYS-F619P2-RTN.cfm, 2024.

[3] Supermicro. "Supermicro SYS-6019U-TN4RT." https://www.supermicro.com/en/products/system/1U/6019/SYS-6019U-TN4RT.cfm, 2024.

[4] Supermicro. "Supermicro SSG-6048R-E1CR60L." https://www.supermicro.com/en/products/system/4U/6048/SSG-6048R-E1CR60L.cfm, 2024.

[5] Extreme Networks. "Universal Switches 5420M." https://www.extremenetworks.com/products/switches/universal-switches/5420, 2024.

[6] Extreme Networks. "Universal Switches 7520." https://www.extremenetworks.com/products/switches/universal-switches/7520, 2024.

[7] NVIDIA. "SN2100 Datasheet." https://network.nvidia.com/files/doc-2020/pb-sn2100.pdf, 2020.

[8] Stordis. "Introduction to Transceivers." https://stordis.com/introduction-to-transceivers/, 2024.

[9] NVIDIA. "RDMA over Converged Ethernet (RoCE)." https://docs.nvidia.com/networking/display/mlnxofedv23070512/rdma+over+converged+ethernet+(roce), 2024.

[10] Hewlett Packard Enterprise. "What is Spine-Leaf Architecture?" https://www.hpe.com/emea_europe/en/what-is/spine-leaf-architecture.html, 2024.

[11] ServeTheHome. "What is a Direct Attach Copper (DAC) Cable?" https://www.servethehome.com/what-is-a-direct-attach-copper-dac-cable/, 2024.

[12] IEEE 802.3ad. "Link Aggregation Protocol (LACP)." IEEE Standards Association, 2010.

[13] La Salle-URL. "Multi-Chassis Link Aggregation (MLAG) Overview." https://blogs.salleurl.edu/en/multi-chassis-link-aggregation-mlag-overview, 2024.

[14] Wikipedia. "Multi-Chassis Link Aggregation Group (MC-LAG)." https://en.wikipedia.org/wiki/Multi-chassis_link_aggregation_group, 2024.

[15] Jain, R. "Virtual Local Area Networks (VLANs)." https://www.cse.wustl.edu/~jain/cis788-97/ftp/virtual_lans/index.html, 1997.

[16] Huawei. "Top of Rack (ToR) Overview." https://support.huawei.com/enterprise/en/doc/EDOC1100023542/4679373c/tor, 2024.

[17] IBM. "IPMI overview." IBM Knowledge Center. https://www.ibm.com/docs/en/power8/8335-GCA?topic=ipmi-overview, 2024.

[18] Intel. "IPMI second-gen interface specification, v2.0 rev 1.1." https://www.intel.la/content/www/xl/es/products/docs/servers/ipmi/ipmi-second-gen-interface-spec-v2-rev1-1.html, 2024.

[19] Lustre. "Architecting Lustre Storage." White Paper, 2024.

[20] Hewlett Packard Enterprise. "High-Performance Computing Systems." https://www.hpe.com/psnow/doc/c01451157, 2024.