



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Architettura di rete del data center di primo livello per il progetto FOSSR

Fabrizio Marangio, Angelo Esposito, Mario Sicuranza

RT-ICAR-NA-2023-03

Data: dicembre 2023



Consiglio Nazionale delle Ricerche, Istituto di Calcolo e Reti ad Alte Prestazioni (ICAR) – Sede di Napoli, Via P. Castellino 111, I-80131 Napoli, Tel: +39-0816139508, Fax: +39-0816139531, e-mail: napoli@icar.cnr.it, URL: www.na.icar.cnr.it



*Consiglio Nazionale delle Ricerche
Istituto di Calcolo e Reti ad Alte Prestazioni*

Architettura di rete del data center di primo livello per il progetto FOSSR

Fabrizio Marangio, Angelo Esposito, Mario Sicuranza

Rapporto Tecnico N: RT-ICAR-NA-2023-03

Data: Dicembre 2023

I rapporti tecnici dell'ICAR-CNR sono pubblicati dall'Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche. Tali rapporti, approntati sotto l'esclusiva responsabilità scientifica degli autori, descrivono attività di ricerca del personale e dei collaboratori dell'ICAR, in alcuni casi in un formato preliminare prima della pubblicazione definitiva in altra sede.

Architettura di rete del data center di primo livello per il progetto FOSSR

Fabrizio Marangio, Angelo Esposito, Mario Sicuranza

Istituto di Calcolo e Reti ad Alte Prestazioni del Consiglio Nazionale delle Ricerche
Via Pietro Castellino, 111 – 80131 Napoli, Italia

E-mail: {fabrizio.marangio, angelo.esposito, mario.sicuranza}@icar.cnr.it

Abstract

Il presente documento tecnico mira a descrivere l'architettura di rete definita per l'istituzione di un nuovo data center da costituirsi nell'ambito del progetto FOSSR¹ (Fostering Open Science in Social Science Research) che prevede la realizzazione di una rete di quattro nodi dislocati. Il data center, che verrà collocato presso la sala CED ICAR-CNR sede di Napoli, sarà il più grande tra quattro nodi distribuiti in diverse sedi del CNR dislocate su territorio nazionale che insieme costituiranno un'infrastruttura cloud e di calcolo dedicata alle scienze sociali.

Keywords: Networking, PNRR, architettura, data center, centro elaborazione dati.

1 Introduzione

Il Piano Nazionale di Ripresa e Resilienza (PNRR) [1] rappresenta un ambizioso programma di interventi e cambiamenti avviato dal Governo italiano nel 2021 con l'obiettivo di stimolare e rafforzare l'economia del Paese dopo le sfide imposte dalla crisi pandemica. Le risorse stanziare per il PNRR sono pari a 191,5 miliardi di euro, ripartite in sei missioni:

- Digitalizzazione, innovazione, competitività e cultura - 40,32 miliardi
- Rivoluzione verde e transizione ecologica - 59,47 miliardi
- Infrastrutture per una mobilità sostenibile - 25,40 miliardi
- Istruzione e ricerca - 30,88 miliardi
- Inclusione e coesione - 19,81 miliardi
- Salute - 15,63 miliardi

Il Governo italiano ha dato il via a un *Piano Nazionale Complementare (PNC)* [2] con un fondo di 30,6 miliardi di euro per sostenere ulteriori interventi. In totale, gli investimenti pianificati, compresi quelli del PNRR e del Fondo Complementare, ammontano a 222,1 miliardi di euro.

Nell'ambito della missione 4 "Istruzione e ricerca" [3], il Ministero si impegna a rafforzare le competenze presenti nel Paese e ad affrontare le sfide tecnologiche e ambientali. L'obiettivo è incrementare gli investimenti nella ricerca e nello sviluppo, promuovendo una sinergia più efficace tra il settore privato e gli enti pubblici. Le misure delineate nel Piano puntano a facilitare un dialogo più proficuo tra il mondo della ricerca e quello produttivo, con l'obiettivo di potenziare l'innovazione

¹ FOSSR web site <https://www.fossr.eu/>

nelle piccole e medie imprese e di favorire la loro partecipazione in filiere strategiche, cruciali per la competitività nazionale ed europea. Questo coinvolge settori chiave, dalle IPCEI (batterie, microelettronica, idrogeno) ai Centri di trasferimento tecnologico, creando un terreno fertile per l'avanzamento tecnologico e la crescita sostenibile.

Nell'ambito della missione 4 si configura il progetto “FOSSR” (*Fostering Open Science in Social Science Research*) [4]. L'obiettivo di FOSSR è la creazione di un *Italian Open Science Cloud* per le Scienze Sociali, che fornirà strumenti innovativi e servizi per indagare le questioni legate al cambiamento economico e sociale delle società contemporanee:

- *Analisi demografiche*: studio dei fenomeni che si riferiscono alla popolazione e in particolare alla sua determinazione statica e all'analisi della sua evoluzione;
- *Struttura dell'economia*: aziende innovative e in crescita rapida, processi e risultati dell'innovazione, nuove modalità di produzione della conoscenza;
- *Problematiche sociali*: invecchiamento, distribuzione della ricchezza, disuguaglianze, istruzione, migrazione, comportamenti sociali, modelli di simulazione sociale, progettazione, implementazione e valutazione delle politiche pubbliche (ad esempio, politiche di ricerca e innovazione, politiche sanitarie).

Questi cambiamenti sono temi chiave affrontati dal PNRR e per affrontarli in maniera adeguata, è necessario fare riferimento all'enorme mole di dati raccolta e processata dalle *Infrastrutture di Ricerca (RI)* che mettono a disposizione dati aperti e in linea con i principi *FAIR* [5] [6]. Questo fa sì che possano essere ottenute informazioni e indicatori di alta qualità atti a migliorare le capacità decisionali basate su evidenze in diversi settori politici.

L'investimento principale sarà la creazione di un'adeguata infrastruttura IT e di una rete di data center per fornire ai ricercatori risorse (strumenti, dati e servizi) per sostenere il flusso di lavoro dei progetti collaborativi proposti.

Il *Work Package 7 (WP7)* [7] del progetto FOSSR è incentrato sulla progettazione e la realizzazione di una rete di data center distribuiti situati in 5 aree strategiche (Napoli, Palermo, Catania, Torino e Pisa). La rete finale includerà due nodi di primo livello di dimensioni medio-grandi (Napoli e Palermo) e due nodi di secondo livello di piccole dimensioni (Torino e Catania). Inoltre, la rete prevede nella sede del CNR di Pisa l'installazione di hardware ulteriore atto a costituire un'infrastruttura IT dedicata alla realizzazione di un *Virtual Research Environment (Ambiente Virtuale di Ricerca - VRE)*.

I due nodi di primo livello dovranno fornire accesso a risorse hardware e software per l'elaborazione ad alte prestazioni e la memorizzazione di big data mediante tecnologie di cloud computing. Queste strutture consentiranno alla comunità di ricerca di analizzare e memorizzare grafi di conoscenza per le scienze sociali mediante gli algoritmi più avanzati di intelligenza artificiale (IA) e deep learning (DL) e di condividere i dati processati, le applicazioni e i servizi sviluppati con la comunità scientifica e ad enti governativi per finalità di controllo e governo. La presenza di due nodi di primo livello fornirà all'infrastruttura cloud ridondanza per consentire un'alta disponibilità delle sue funzionalità e la possibilità di ripristinare i dati in seguito ad emergenze grazie alla replicazione degli stessi effettuata tra i vari siti fornendo un disaster recovery dislocato localmente. Verrà inoltre consentito lo sviluppo di applicazioni scalabili, in grado di gestire carichi e picchi di traffico elevati, con architettura multi-regione per la resilienza alle interruzioni locali.

Il nodo di primo livello da costituirsi presso la sede di Napoli dell'istituto ICAR del CNR sarà il nodo

col maggior numero di risorse hardware sia in termini numerici che di potenza di calcolo. Il progetto del data center prevede 8 rack con 127 server, ognuno dei quali dotato di due processori multi-core e 1 TB di RAM, consentendo diversi tipi di flussi di lavoro di elaborazione: (data-intensive, cpu-intensive, elaborazione parallela e virtualizzazione). Alcuni ospiteranno GPU di ultima generazione per eseguire applicazioni e gestire carichi di lavoro massicci come algoritmi di Intelligenza Artificiale (IA). Il cluster di server di calcolo sarà collegato a due sistemi di archiviazione parallela separati: uno dedicato all'archiviazione a blocchi con almeno 400 TB composti da unità NVMe e SAS HDD e un secondo dedicato all'archiviazione ad oggetti con almeno 1,2 PB con unità SSD e SAS HDD. I server e i sistemi di archiviazione dovranno essere interconnessi tramite tecnologie di rete ad alta velocità e bassa latenza specializzate per i sistemi di calcolo ad alte prestazioni.

Obiettivo di questo rapporto tecnico è la descrizione dell'architettura di rete che andrà ad interconnettere tali dispositivi in modo da garantire le caratteristiche di velocità e alta affidabilità previste e che si intendono garantire.

2 Topologie e Architetture per Data Center

In un contesto in cui la società è fortemente permeata dalla presenza e dal flusso incessante di dati, i data center stanno acquisendo sempre più importanza, diventando il fulcro su cui si basa l'intera infrastruttura tecnologica globale. Non sono semplici entità statiche, ma dinamiche e adattabili, capaci di supportare le esigenze sempre più complesse di elaborazione, memorizzazione e distribuzione di informazioni. Questi centri tecnologici sono progettati con l'obiettivo di fornire una piattaforma versatile, robusta e scalabile. La loro versatilità è evidente nel supporto a una vasta gamma di applicazioni, dalla gestione di dati sensibili e operazioni critiche a soluzioni di cloud computing che compongono l'infrastruttura digitale di molte aziende.

Al giorno d'oggi, esistono tre principali configurazioni per i data center [8], ognuna delle quali presenta vantaggi e svantaggi. Alcuni data center di grandi dimensioni spesso implementano due o addirittura tutte e tre queste configurazioni nella stessa struttura. Esse sono:

- **Centralizzata:** è una configurazione appropriata per i data center di dimensioni più contenute. Ciascun server è collegato in modo efficiente agli switch principali, che sono centralizzati nell'area principale di distribuzione. Questo garantisce un utilizzo molto efficiente degli switch e semplifica la gestione e l'aggiunta di componenti. Tale configurazione ha problemi di scalabilità che rendono difficile l'espansione del data center.
- **A zone:** in cui gli switch possono essere distribuiti tra le posizioni di fine fila (end-of-row, EoR) o in posizione mezzana (middle-of-row, MoR) e possono generalmente supportare più rack server. Questa soluzione è molto scalabile, ripetibile e prevedibile. L'architettura a zone è di solito il design più conveniente, fornendo il massimo livello di utilizzo degli switch e delle porte riducendo al minimo i costi di cablaggio.
- **Top of Rack (ToR):** questo tipo di configurazione di solito consiste in due o più switch posizionati nella parte superiore del rack in ogni armadio server. Tutti i server nel rack sono collegati a entrambi gli switch per garantire la ridondanza. Gli switch top-of-rack hanno collegamenti ascendenti allo strato successivo di commutazione. Il top-of-rack semplifica significativamente la gestione dei cavi e riduce al minimo i requisiti di contenimento del cavo. Questo approccio fornisce anche uno switching veloce da porta a porta per i server all'interno del rack e un sovradimensionamento prevedibile del collegamento ascendente.

Le configurazioni mostrate si riferiscono alla disposizione fisica e logica dei componenti della rete all'interno di un data center. Sono, cioè, delle mappe che illustrano come i vari elementi possono essere collegati e comunicano tra loro. È necessario però focalizzarsi sull'organizzazione complessiva del data center includendo aspetti come la progettazione degli strati di rete (accesso, aggregazione e nucleo), andando a considerare come i componenti lavorano insieme per il raggiungimento degli obiettivi. Esistono ormai varie topologie di rete per data center ma esse si basano principalmente sull'evoluzione delle seguenti topologie principali [9]:

- **Three-Tier:** è un'architettura di rete che suddivide le risorse di rete in tre livelli, ciascuno con un ruolo specifico (Figura 1):
 - *Livello di accesso:* questo livello è responsabile della connessione dei dispositivi di end-user alla rete.
 - *Livello di distribuzione:* questo livello si pone come un ponte tra il livello di accesso e il core layer, fornendo routing, policy enforcement e controllo del traffico.
 - *Livello core:* questo livello rappresenta il backbone della rete ed è responsabile della distribuzione del traffico di rete tra i diversi carichi di lavoro.

Un'architettura di questo tipo presenta vantaggi in termini di *efficienza* andando a ridurre la congestione reindirizzando in maniera appropriata il traffico, *sicurezza* andando a isolare le risorse sensibili e *scalabilità*. Quest'ultima caratteristica è però limitata dall'utilizzo del protocollo STP che supporta fino a cento switch. In caso di data center per cui è prevista una forte espansione, tale architettura potrebbe non essere la scelta migliore. Le architetture a tre livelli gestiscono il traffico *nord-sud* e utilizzano il protocollo *Spanning Tree Protocol (STP)*. Tale protocollo soffre di una problematica dovuta ai cosiddetti *Broadcast Storm*.

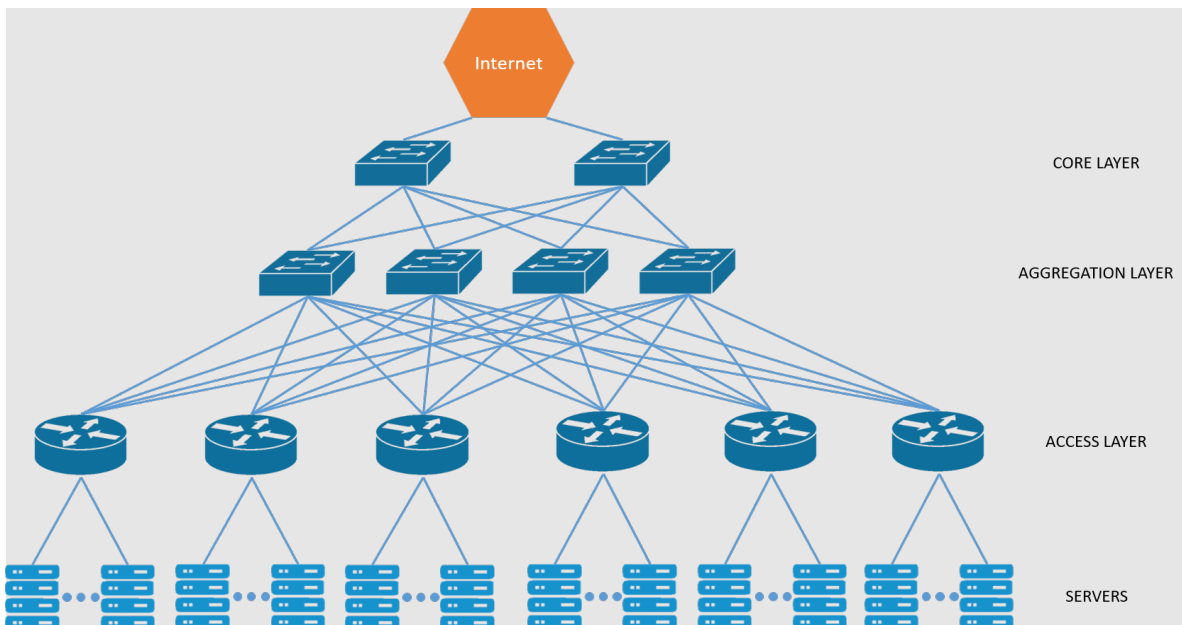


Figura 1 - Three Tier Topology

- **Spine-Leaf:** è un'architettura a due strati composta dal livello Spine e dal livello Leaf (Figura 2):

- *Spine*: sono collegati a tutti gli switch Leaf in una topologia a maglia. Sono in grado di gestire il traffico di rete di livello 3 e sono così chiamati in quanto, connettendosi a tutti gli switch Leaf, costituiscono la “spina dorsale” della rete.
- *Leaf*: prendono il ruolo degli switch di accesso essendo collegati direttamente ai server e sono solitamente ridondati. Possono gestire sia il traffico di rete di livello 2 che di livello 3. Indipendentemente dallo switch Leaf a cui è collegato un server, deve attraversare lo stesso numero di dispositivi ogni volta che si connette a un altro server.

Questa topologia riduce al minimo la latenza e i colli di bottiglia, perché ogni payload deve passare solamente da uno switch Spine e da un switch Leaf per raggiungere il suo punto finale. Gli switch Spine hanno un'elevata densità di porte e costituiscono il nucleo dell'architettura. Questo tipo di rete è spesso utilizzata in ambienti in cui è necessaria una connettività affidabile e scalabile, come in data center di grandi dimensioni o distribuiti su più sedi. I vantaggi sono sostanzialmente tre. In primis, percorsi ridondanti tra i dispositivi forniscono una maggiore *resilienza* in caso di guasti o congestione. È possibile aggiungere o rimuovere dispositivi senza modificare la topologia della rete (*scalabilità*). Il traffico di rete può essere distribuito su più percorsi (*efficienza*). Inoltre, risulta essere una topologia molto efficiente dal punto di vista energetico, riducendo l'impatto ambientale. L'architettura Spine-Leaf è pensata per il traffico *est-ovest* e non utilizza il protocollo STP. Sebbene STP preveda percorsi ridondanti tra gli switch, è possibile attivarne solo uno alla volta, il che significa che i percorsi possono essere sovrabbondanti. Il mancato utilizzo di STP riduce al minimo la latenza e i potenziali colli di bottiglia dell'architettura. Le caratteristiche descritte rendono quindi le reti Spine-Leaf particolarmente adatte per data center volti al *cloud computing*.

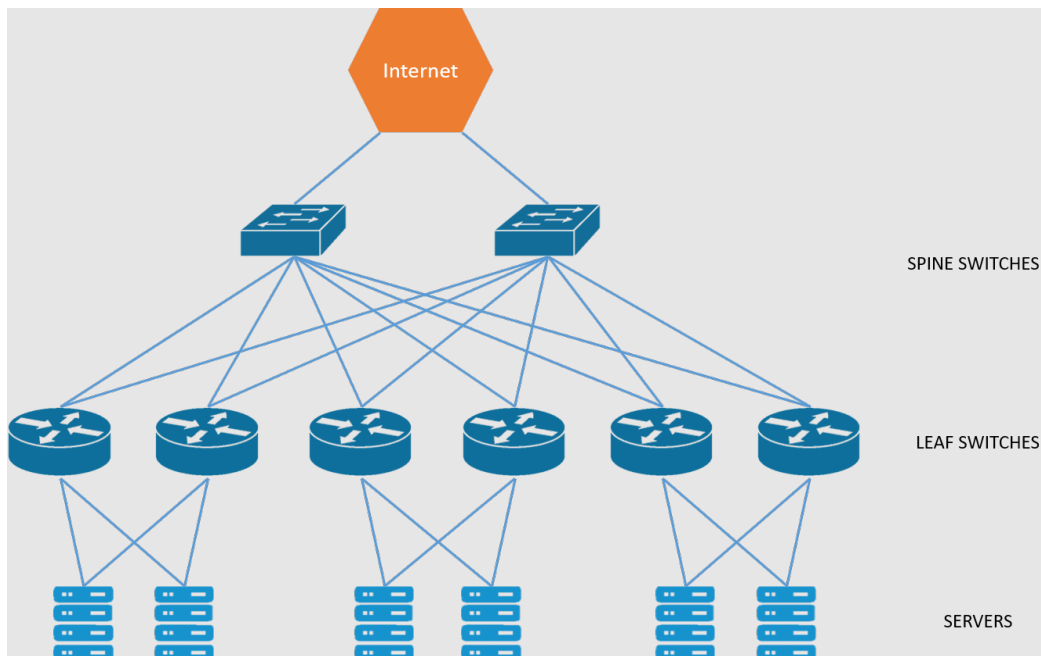


Figura 2 - Spine-Leaf Architecture

3 Architettura di rete per il nodo di primo livello di Napoli

Come precedentemente descritto, il data center progettato per la sede di Napoli del CNR-ICAR sarà il più grande tra i data center costituenti l'Open Cloud del progetto FOSSR. Il data center sarà dotato di componenti hardware in grado di gestire il sistema, garantendone disponibilità e funzionalità, fornendo al contempo un'infrastruttura informatica adeguata al funzionamento di servizi e strumenti avanzati per l'analisi dei dati utilizzando algoritmi di IA come quelli basati su Deep Learning (DL) e Machine Learning (ML).

L'Area 1 di Napoli del CNR, di cui l'Istituto ICAR fa parte, costituisce un fulcro importante della ricerca scientifica e tecnologica in Italia. La sua connessione alla rete GARR rappresenta un passo significativo verso l'integrazione e lo sviluppo delle risorse digitali nel panorama accademico e scientifico del paese. La rete GARR, una delle più avanzate in Italia, è un pilastro fondamentale per la comunità dell'istruzione, della ricerca e della cultura. La sua infrastruttura in fibra ottica si estende per circa 15.000 km, comprendendo sia backbone che collegamenti di accesso. Questa rete è il frutto del progetto GARR-T Progress, il quale ha svolto un ruolo essenziale nell'evoluzione della connettività nelle quattro regioni di convergenza del sud Italia: Calabria, Campania, Puglia e Sicilia. Grazie agli sforzi del progetto GARR-T Progress, la capacità della rete ha raggiunto livelli molto alti. Attualmente, ogni sezione del backbone può gestire fino a 200 Gbps, permettendo un flusso di dati robusto e affidabile. Inoltre, i collegamenti di accesso sono stati progettati con una grande scalabilità, consentendo loro di adattarsi agilmente alle crescenti esigenze degli utenti e delle istituzioni accademiche e di ricerca. La capacità massima disponibile per un singolo collegamento di sito utente è di 200 Gbps, dimostrando l'impegno di GARR nel fornire risorse di connettività all'avanguardia per supportare la ricerca, l'innovazione e lo sviluppo tecnologico in Italia.

La rete dell'Area di Ricerca di Napoli 1 e quindi dell'Istituto ICAR è collegata attualmente al GARR a una velocità di 1 Gbps, ma come da piano dei fabbisogni del CNR è previsto l'upgrade di connettività a 10Gbps per le aree di Ricerca del CNR. Questo aumento della larghezza di banda è motivato dalla crescente complessità delle attività di ricerca, che richiedono lo scambio rapido e affidabile di dati di grandi dimensioni, simulazioni complesse e collaborazioni internazionali in tempo reale. Pertanto, per supportare questo ampliamento è in corso un upgrade hardware presso il centro stella dell'area di ricerca di Napoli [10]. La nuova infrastruttura di rete dell'area sarà basata su tre switch Juniper EX4300-32F [11], che costituiranno il nuovo centro stella. Questi switch forniscono una base affidabile per la gestione del traffico di rete e assicurano una connettività stabile e sicura tra i diversi nodi della rete. L'attuale configurazione del data center dell'istituto ICAR è mostrata in Figura 3 e consta dei seguenti dispositivi:

- **ICOS-IRET - 16 nodi PowerEdge MX740c DELL:** questo cluster è collegato a uno switch DELL MX9116N connesso a sua volta tramite una rete a 100 Gb con due switch 2 Sonicwall NSA 4700 in alta disponibilità (HA) collegati a loro volta agli switch Juniper con una rete a 10 Gb.
- **ISILON - Cluster DELL EMC:** questo cluster è collegato a 2 switch DELL Powerswitch Z9264F-ON connessi a loro volta con uno switch DELL MX9116N con una rete a 25 Gb. Lo switch è collegato a switch Juniper con una rete a 10 Gb.
- **VIVARA - Cluster DELL composto da 3 nodi PowerEdge MX740c:** questo cluster è collegato allo stesso switch DELL MX9116N del punto precedente con una rete a 10 Gb.
- **YODA - Cluster DELL composto da 19 nodi M620 e un cluster IBM-BEN da 73 nodi Power 9:** questi cluster sono collegati a due switch Juniper EX2200-24P-4G con una rete a 1 Gb, interconnessi agli switch Juniper, collegati a GARR, con una rete a 1 Gb.

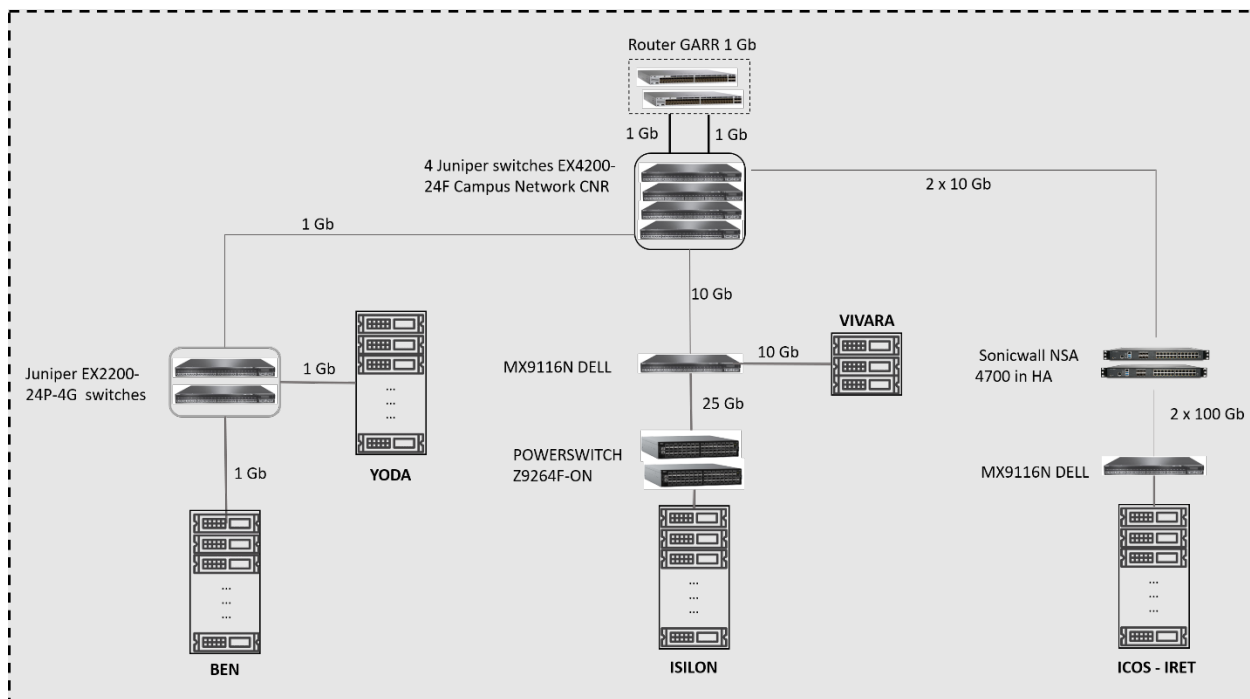


Figura 3 - Attuale configurazione ICAR Napoli

L'infrastruttura FOSSR sarà collegata alla rete tramite una topologia di rete Spine-Leaf (descritta nel paragrafo 2). Come visto precedentemente, questo tipo di architettura viene utilizzato sempre più frequentemente per vari motivi. In primo luogo, c'è la necessità di gestire un volume di traffico di rete sempre crescente ed è inoltre necessario considerare l'impatto ambientale della propria architettura di rete. Il data center prevede connessioni ad alta velocità e affidabilità e dovrà tener in conto di una architettura scalabile nel tempo. Inoltre, l'utilizzo di tecniche di virtualizzazione previste nel progetto in termini di container e macchine virtuali, aumenta le comunicazioni server-to-server e quindi sfrutta appieno le potenzialità dell'architettura Spine-Leaf che permette di diminuire notevolmente il sovraccarico dovuto al traffico elevato, evitando eventuali colli di bottiglia.

Per una migliore comprensione dell'architettura di rete, è necessario descrivere i dispositivi che si andranno a interconnettere. Il data center sarà composto da un totale di 127 server così suddivisi:

- **Nodi di calcolo:** 100 server *DELL PowerEdge R7525* (General Purpose) con interfaccia di rete *DELL Intel Ethernet XXV710 Dual Port @25GbE SFP28 Adapter, PCIe Full Height* e transceiver *DELL SFP28 @25GbE, 85C*.
- **Nodi di management e storage:** 6 server *DELL PowerEdge R7525 (nVME)* con interfaccia di rete *DELL Intel Ethernet XXV710 Dual Port @25GbE SFP28 Adapter, PCIe Full Height* e transceiver *DELL SFP28 @25GbE, 85C*.
- **Nodi Ceph:** 11 server *DELL PowerEdge R7525 (vSAN Ready)* aventi le stesse interfacce di rete e transceiver dei nodi precedenti più un'interfaccia di rete *PCI-Express Broadcom® 57508 Dual Port 100GbE QSFP Adapter, PCIe Full Height*.
- **Nodi GPU Ready:** 10 server *DELL PowerEdge R7525 (GPU Double-Wide Ready)* con interfaccia di rete *DELL Intel Ethernet XXV710 Dual Port @10/25GbE SFP28 Adapter, PCIe Low Profile*.

Tali dispositivi saranno installati in otto rack configurando la rete nel seguente modo:

- Rete di gestione IBN (In-Band management Network):** su questa rete transiterà il traffico utile per la gestione dell'infrastruttura e l'interconnessione ad alta velocità dei nodi di archiviazione Ceph della partizione Cloud. Questa rete dovrà essere implementata in una topologia a Spine-Leaf completamente ridondante con velocità di accesso ai nodi di almeno 25 Gbps. L'interconnessione tra i livelli Spine e Leaf dovrà essere realizzata tramite link ridondanti con velocità di almeno 100 Gbps e tale che il rapporto di sovrascrittura non sia superiore a 3:1. Per una corretta gestione dei cavi, verranno utilizzati due switch Leaf per ogni rack con un numero sufficiente di porte da 25 Gbps, per l'interconnessione con i nodi, e porte da 100 Gbps per l'interconnessione con le Spine. Oltre alle porte necessarie per le interconnessioni, è necessario prevedere un numero di porte aggiuntive pari ad almeno il 10% delle porte occupate per garantire la futura scalabilità dell'infrastruttura. Saranno necessari due switch Spine, ciascuno dotato di un numero sufficiente di porte 100GbE per interconnettere in modo ridondante tutti gli switch Leaf rispettando il livello di sovrascrittura prescritto, tutti i nodi Ceph tramite 2 connessioni a 100GbSR4 MPO per nodo e per realizzare l'interconnessione alla rete esistente tramite almeno 4 link a 25Gb SR. Gli switch Spine e Leaf dovranno essere nativamente in grado di realizzare una topologia priva di SPOF attraverso l'implementazione di meccanismi MC-LAG in grado di rendere trasparenti ai nodi, a parte una possibile riduzione della larghezza di banda della connessione, qualsiasi guasto o riavvio di qualsiasi apparecchiatura appartenente alla rete. L'interconnessione ai nodi sarà effettuata tramite link in fibra a 25Gb SR. In Figura 4 è mostrato uno schema indicativo della rete IBN.

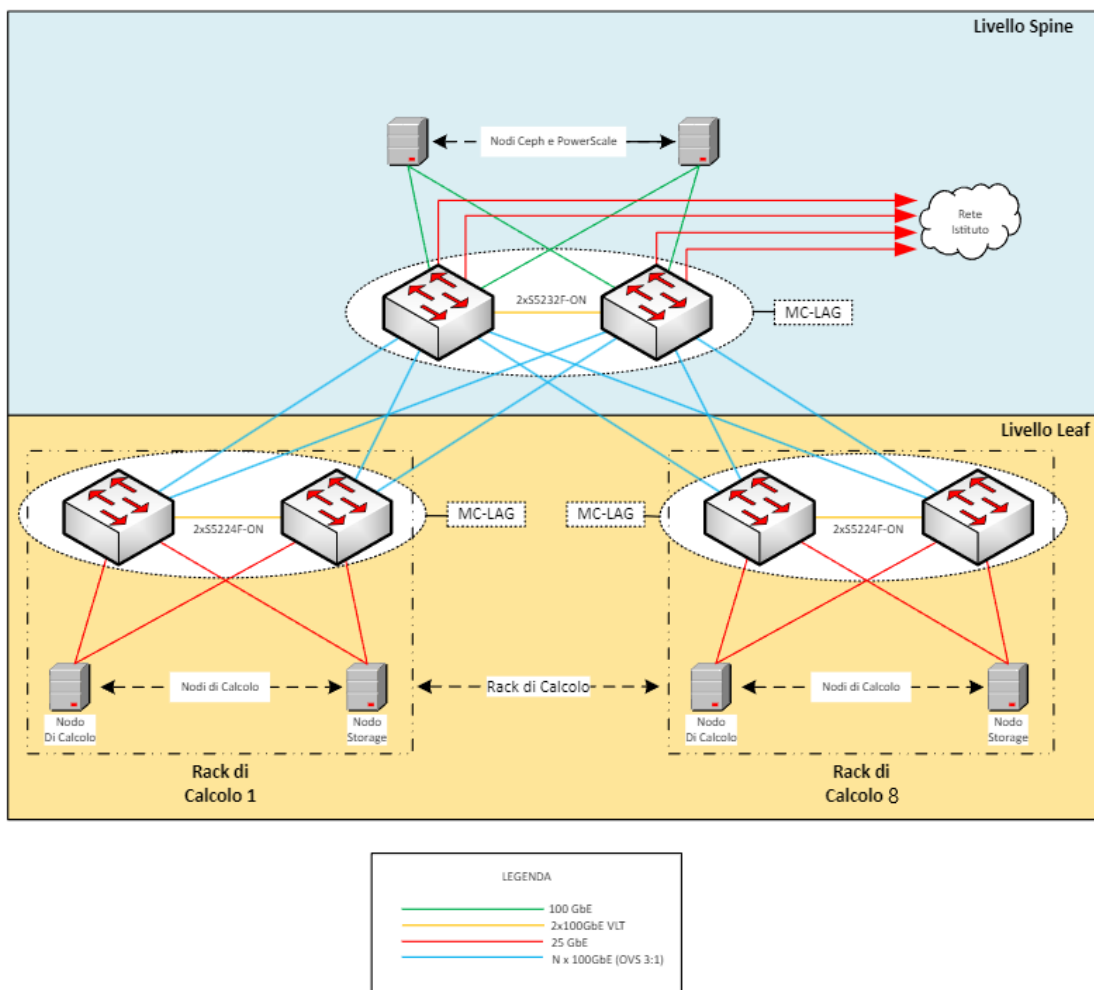


Figura 4 - Schema Rete IBN

- **Rete di gestione fuori banda OOBND (Out-Of-Band management Network):** questa rete interconetterà i controller di gestione della scheda base di tutti i nodi e le interfacce di gestione OOB di tutti gli altri componenti dell'infrastruttura (sistemi di archiviazione, apparecchiature di rete, PDUs, ecc.). Tale rete dovrà essere implementata in topologia Spine-Leaf e ridondante al livello Spine. La velocità di accesso potrà essere a 1 Gbps in tecnologia Base-T (rame). D'altra parte, le interconnessioni tra Spine e Leaf dovranno essere implementate in modalità ridondante mediante collegamenti in fibra ottica con una velocità di almeno 10 Gbps. Pertanto, per ogni rack del nuovo Data Center dovranno essere utilizzati almeno uno switch 1/10 Gbps (per un totale di 8 switch) con un numero sufficiente di porte da 1 Gbps per interconnettere tutti i nodi nel rack più un'ulteriore quota di porte libere pari ad almeno il 10% delle porte occupate. Dovrebbe essere prevista anche una coppia di switch Spine, ciascuno dotato di un numero sufficiente di porte 10 Gbps per interconnettere tutti gli switch Leaf in modo ridondante e per realizzare l'interconnessione alla rete dell'istituto tramite almeno 2 connessioni 10 GbSR.
- **Rete di interconnessione a bassa latenza (Low Latency Network - LLN):** Questa rete interconetterà tutti i nodi di calcolo della partizione HPC, realizzata con tecnologia e hardware RockPort. Tradizionalmente, le reti InfiniBand sono state preferite per la loro interconnettività a bassa latenza ed elevata larghezza di banda, che le rende una scelta popolare negli ambienti di calcolo ad alte prestazioni (HPC). Tuttavia, il passaggio alla tecnologia e all'hardware RockPort rappresenta un cambiamento strategico guidato dall'evoluzione tecnologica e da considerazioni di compatibilità. Le reti InfiniBand, rinomate per le loro caratteristiche di prestazioni superiori, sono state a lungo la spina dorsale dei cluster HPC, facilitando una comunicazione efficiente tra i nodi di calcolo. Nonostante la loro efficacia, l'adozione della tecnologia RockPort rappresenta un'alternativa interessante.

Le reti RockPort, pur essendo relativamente recenti nel campo dell'HPC, offrono metriche di prestazione comparabili e vantano una maggiore compatibilità con le architetture hardware contemporanee. La migrazione da InfiniBand a RockPort è in linea con gli obiettivi generali del progetto, ovvero l'ottimizzazione dell'utilizzo delle risorse e la semplificazione dei flussi di lavoro operativi. Sfruttando la tecnologia e l'hardware RockPort, l'architettura di rete è pronta a soddisfare i futuri requisiti di scalabilità e le innovazioni tecnologiche, garantendo una competitività e un'adattabilità sostenute nel panorama dinamico del calcolo ad alte prestazioni. Inoltre, la perfetta integrazione delle reti RockPort con l'infrastruttura hardware acquistata nell'ambito della convenzione sottolinea la sinergia strategica tra investimenti tecnologici e obiettivi del progetto. Questo allineamento armonioso massimizza l'efficienza nell'utilizzo delle risorse e attenua le potenziali complessità operative associate a componenti hardware e di rete eterogenei.

In sintesi, il passaggio alla tecnologia RockPort rappresenta un'iniziativa proattiva per migliorare la resilienza, le prestazioni e la scalabilità dell'infrastruttura di rete all'interno della partizione HPC; sfruttando le capacità intrinseche della tecnologia e dell'hardware RockPort, il progetto cerca di creare una solida base favorevole all'innovazione sostenuta e all'eccellenza computazionale.

In Figura 5 è mostrata la rete relativa al cloud FOSSR. Come descritto sopra, la rete conterrà due switch Spine collegati a due switch Leaf per nodo (mostrati nella parte superiore dell'immagine). I collegamenti tra gli switch e con i server saranno di 25 Gbps. Verrà inoltre creata una rete parallela che verrà utilizzata esclusivamente dagli amministratori per la gestione dell'infrastruttura (mostrata nella parte inferiore dell'immagine). A differenza della rete principale, avremo uno switch Leaf per

nodo e due connettori. La larghezza di banda in questo caso sarà di 1 Gbps. La Figura 6 riassume lo stato della rete dopo l'integrazione della rete FOSSR col sistema preesistente.

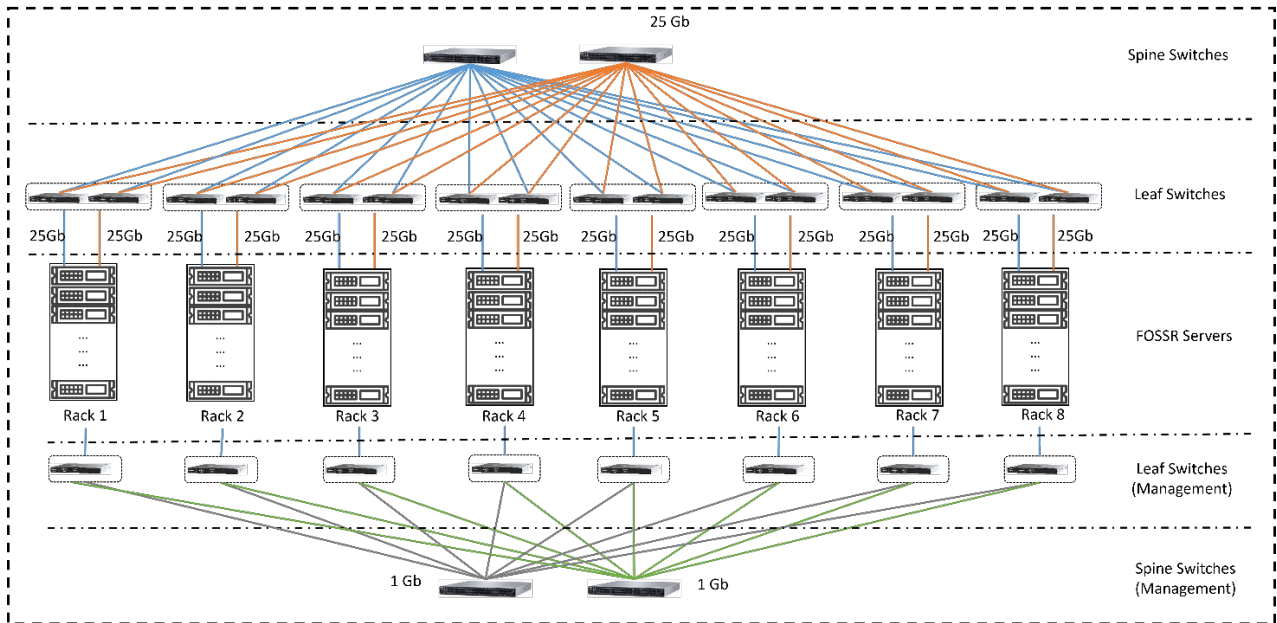


Figura 5 - FOSSR Network

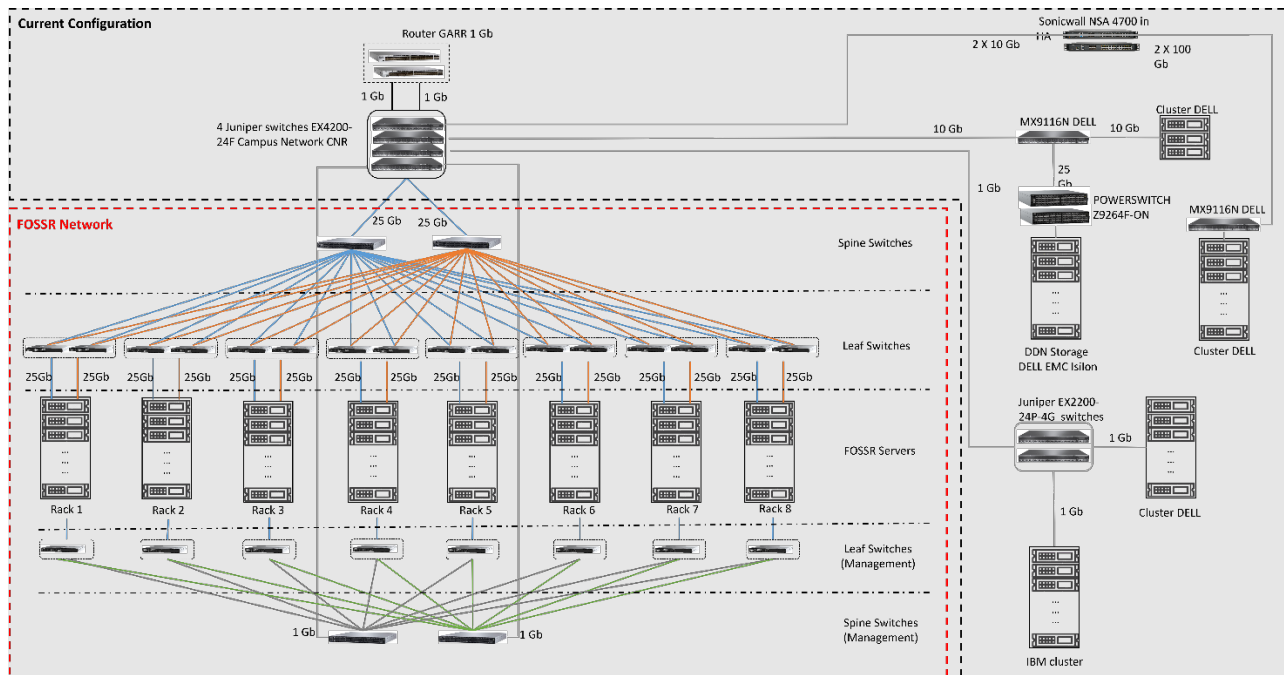


Figura 6 - Rete Finale Risultante

4 Conclusioni

Il rapporto tecnico appena presentato descrive le scelte effettuate per la progettazione dell'infrastruttura di rete del nuovo data center dell'istituto ICAR che verrà costituito nell'ambito del progetto FOSSR. Tale progetto punta alla creazione di un'infrastruttura cloud per le scienze sociali, distribuita su quattro data center dislocati su territorio nazionale. Il data center che verrà costituito presso la sede dell'istituto ICAR di Napoli sarà il maggiore dei quattro, sia in termini di dimensioni che di potenza di calcolo. Dopo aver descritto le principali configurazioni di rete maggiormente utilizzate nei data center, il documento espone le scelte di progettazione per il nodo di Napoli. L'architettura di rete sarà quindi basata su un modello Spine-Leaf e sarà composto da due reti distinte. Una IBN che gestirà il traffico di rete proveniente dall'esterno e una OoBN per il traffico di rete relativo alla gestione da parte degli amministratori. La rete risultante andrà a interconnettere i dispositivi garantendo connettività ad alta velocità e bassa latenza.

5 Acknowledgement

Il lavoro svolto e descritto in questo rapporto tecnico è stato realizzato nell'ambito del progetto PNRR FOSSR, <https://www.fossr.eu/>

6 Riferimenti

- [1] G. Italiano, «PNRR,» [Online]. Available: <https://www.italiadomani.gov.it/content/sogei-ng/it/it/home.html>. [Consultato il giorno 12 2023].
- [2] G. Italiano, «PNC,» [Online]. Available: <https://www.italiadomani.gov.it/it/il-piano/Risorse/piano-complementare/piano-nazionale-complementare.html>. [Consultato il giorno 12 2023].
- [3] G. Italiano, «Missione 4 "Istruzione e Ricerca",» [Online]. Available: <https://www.mur.gov.it/it/pnrr/missione-istruzione-e-ricerca>. [Consultato il giorno 12 2023].
- [4] C. N. d. Ricerche, «Progetto FOSSR,» [Online]. Available: <http://www.fossr.eu/>. [Consultato il giorno 12 2023].
- [5] GoFair, «Fair Principles,» [Online]. Available: <https://www.go-fair.org/go-fair-initiative/>. [Consultato il giorno 12 2023].
- [6] A. J. e. Al., «FAIR Principles: Interpretations and Implementation Considerations,» *Data Intelligence*, vol. 2, n. 1-2, pp. 10-29, 2020.
- [7] P. FOSSR, «Network of Data Center,» [Online]. Available: <https://www.fossr.eu/cosa-facciamo/rete-di-data-center/>.
- [8] Commonslope, «Data center topologies and architectures,» [Online]. Available: <https://www.commslope.com/globalassets/digizuite/2391-data-center-best-practices-ebook-ch3-co-110101-en.pdf>. [Consultato il giorno 12 2023].
- [9] F. Europa, «What is Data Center Architecture,» [Online]. Available: <https://community.fs.com/article/what-is-data-center-architecture.html>. [Consultato il giorno 12 2023].
- [10] G. L. Angelo Esposito, «Evoluzione dell'Infrastruttura di Rete dell'ICAR-CNR Sededi

Napoli,» [Online]. Available: <https://intranet.icar.cnr.it/wp-content/uploads/2023/12/RT-ICAR-NA-2023-02-V6.pdf>. [Consultato il giorno 12 2023].

[11] Juniper, «EX4300 Line of Ethernet Switches Datasheet,» [Online]. Available: <https://www.juniper.net/us/en/products/switches/ex-series/ex4300-line-of-ethernet-switches-datasheet.html>. [Consultato il giorno 12 2023].