Department
of Excellence
2018 - 2022

EMbeDS
Economics and Management
in the era of Data Science

2020

EMbeDS HPC

DANIELE LICARI

Economics and Management in the era of Data Science (EMbeDS)
Piazza Martiri della Libertà, 33
56127 Pisa
Tel. Centralino: 050883111
www.embeds.santannapisa.it

EMbeDS HPC

L'architettura computazionale EMbeDS per il supporto alle attività di ricerca della Scuola Superiore Sant'Anna è dotata di **2 server HPC DELL POWEREDGE R740** (2 CPU Intel Xeon Gold 6252 da 24 core, 384 GB RAM e n 2 GPU Nvidia Volta V100) e un totale di **40 terabyte di spazio disco** su storage area network (SAN).

Il cluster viene gestito attraverso una piattaforma di virtualizzazione VMware vSphere 6.7U3 che ne garantisce flessibilità, scalabilità e agilità su differenti progetti di ricerca.

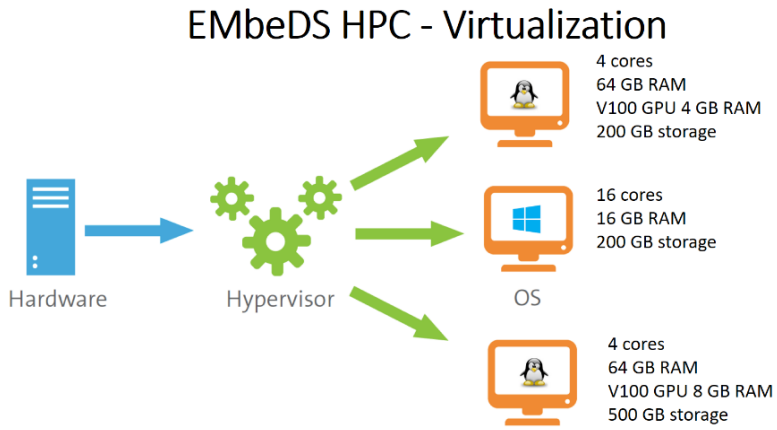

Figure 1. Per virtualizzazione si intende l'astrazione di risorse IT fisiche come hardware, software, memoria e componenti di rete. Il fine è quello di fornire queste risorse a livello virtuale e distribuirle in modo flessibile a seconda delle esigenze tra i diversi gruppi

Il sistema IT EMbeDS integra la soluzione NVIDIA GRID vGPU che consente la condivisione di una scheda GPU NVIDIA Tesla su più macchine virtuali (VM) creando più dispositivi logici vGPU, ognuno dei quali può essere assegnato a una macchina virtuale.

Il sistema IT EMbeDS è stato progetto per favorire l'implementazione e **la sperimentazione di codice calcolo parallelo** su CPU/GPU, strumenti per il Machine Learning (ML), Deep learning e Database.

Connettività EMbeDS HPC

È stata realizzata una connessione ad alta velocita (Fibre Channel) tra lo Storage Area Network (SAN), con 40 TB di spazio disco, e i due server EMbeDS. Le macchine virtuali sono connesse alla rete interna della Scuola Superiore Sant'Anna denominata VDI. **L'accesso esterno** alla rete VDI potrà essere eseguito **attraverso** il protocollo sicuro **VPN** che permette l'estensione a livello geografico della rete interna (https://intranet.santannapisa.it/it/pagine/virtual-private-network-vpn).

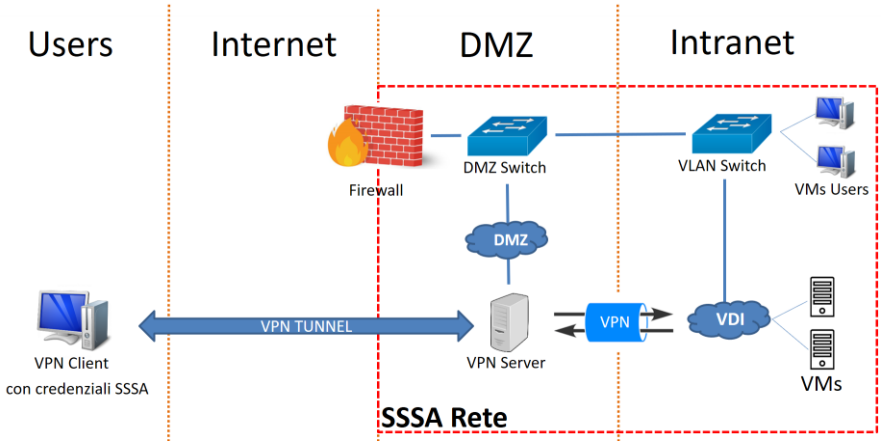

Figure 2. Le macchine virtuali EMbeDS possono essere raggiunte dalla rete esterna attraverso il protocollo sicuro VPN.

Suddivisione dello spazio disco (40 TB)

10 Terabyte di spazio disco sono stati destinati alla gestione dei dischi virtuali delle macchine virtuali (VMFS Datastore), i restanti 30 TB "montanti" in Network File System (NFS) come cartelle all'interno delle macchine virtuali (VM). Il Network File System è un file system che consente ai computer di utilizzare la rete per accedere ai dischi rigidi remoti come fossero dischi locali.

Ogni macchina virtuale ha **3 livelli di isolamento dello spazio disco**, le cartelle condivise in NFS:

1. /home, contiene le directory dati utente (solo su VM Linux)
2. /scratch, può contenere file temporanei o dati da condividere tra gli utenti della singola VM
3. /public, dati condivi tra più VM

Ci aspettiamo un numero limitato di richieste per VM Window, quindi le home degli utenti per sistemi Windows saranno allocate direttamente nel disco virtuale della VM (nei 10 TB di datastore VMFS).

Macchine Virtuali EMbeDS

Potranno essere richieste due tipologie di macchina virtuale:

- **VM vCPU**: 8vCPU, 64 GB RAM
- **VM vGPU:** 8vCPU, 24 GB RAM, vGPU 4GB RAM

Ulteriori risorse di elaborazione **potranno essere allocate dinamicamente** a ciascuna macchina virtuale in base alle esigenze dei singoli gruppi di ricerca (previa autorizzazione Responsabile HPC o Managing Board).

L'accesso alle VM può essere eseguito **attraverso** protocollo **SSH o Desktop Remoto**.

Sistemi O.S. supportati

Le VM potranno essere richieste con i seguenti sistemi operativi installati:

- **Windows 10 Pro 64 bit**
- **Ubuntu 18.04 LTS 64 bit**

Le macchine virtuali Linux avranno pre-installato il sistema di container Docker che semplifica il deploy di applicazioni GPU NVIDIA (https://ngc.nvidia.com/catalog). La maggior parte delle librerie e strumenti per artificial intelligence (AI), deep learning, o high-performance computing (HPC) sono sviluppate per sistemi Linux (forniscono un eccellente supporto hardware).

Virtual GPU

Tutte **le vGPU create sulla stessa GPU fisica devono utilizzare lo stesso profilo (con memoria partizionata in ugual misura).** A ciascuna vGPU viene dato l'accesso temporizzato alla totalità delle risorse di calcolo di una GPU. **I profili sono raggruppati in serie diverse a seconda delle diverse classi di workload** per le quali sono ottimizzati.

- *Serie Q*: per professionisti creativi e tecnici che richiedono le prestazioni e le caratteristiche della tecnologia Quadro
- *Serie C*: carichi di lavoro dei server ad alta intensità di calcolo, come l'intelligenza artificiale (AI), l'apprendimento profondo o il calcolo ad alte prestazioni (HPC)

Le macchine virtuali guest di **Windows** non **supportano** la serie **C**,ma solo i tipi **vGPU NVIDIA serie Q**. Quindi la serie Q è quella più indicata per i sistemi Windows e **la serie C per macchine virtuali guest Linux** (che supporta anche la serie Q).

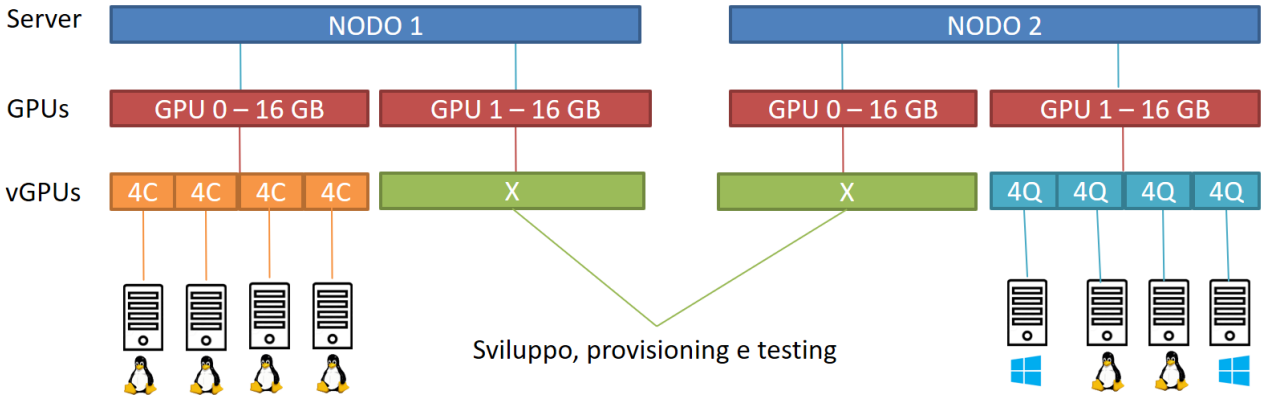

Figure 3. Esempio suddivisione vGPU

Le VM che necessitano di maggiore RAM vGPU potranno essere spostate su GPU partizionate con un profilo più alto (es. 8 o 16 GB RAM).

Sperimentazione

Sono state predisposte due macchine virtuali standard (8vCPU, 24 GB RAM, vGPU 4GB RAM) per gli istituti di Economia e Management al fine di favorire l'accessibilità, la sperimentazione e l'uso delle tecnologie vGPU EMbeDS. Gli amministratori dei due sistemi computazionali potranno in completa autonomia approvare le richieste provenienti dagli afferenti dei due istituti e creare le credenziali di accesso alle risorse.

Gli utenti che accedono alle due VM avranno modo di:

- testare le risorse computazionali EMbeDS al fine di valutare se presentare un progetto secondo le modalità descritte nel documento in allegato (vedi paragrafo successivo);
- sperimentare codice calcolo parallelo su CPU/GPU, strumenti per il Machine Learning (ML) e Deep learning.

La gestione della VM per Economia è affidata al **Dott. Andrea Vandin e per Management alla Dott.ssa Valentina Lorenzoni**.

Richiesta Risorse

La richiesta di una macchina virtuale EMbeDS può essere fatta esclusivamente da **membri affiliati ad EMbeDS** (vedi paragrafo successivo). Il richiedente dovrà inviare un **documento** (ALLEGATO A). **contenente una descrizione** puntuale del progetto (finalità, durata, risorse), il nominativo del **responsabile scientifico** e un **amministratore di sistema** (ALLEGATO B).

La richiesta dovrà essere approvata dal responsabile IT EMbeDS (Dott. Daniele Licari) **o dal Managing Board** (vedi paragrafo Managing Board).

Affiliazione EMbeDS

Procedura di affiliazione/afferenza di soggetti interni alla Scuola:

- richiesta indirizzata alla Coordinatrice Scientifica di EMbeDS e valutata in sede di Managing Board;
- passaggio nelle Giunte degli Istituti di Economia e di Management.

Procedura di affiliazione/afferenza di soggetti esterni alla Scuola (in quanto integrata con le procedure previste dal Regolamento Affiliazioni):

- richiesta indirizzata alla Coordinatrice Scientifica di EMbeDS (includente l'indicazione di quale dei due Istituti dovrebbe veicolare l'affiliazione) e valutata in sede di Managing Board;
- passaggio nella Giunta dell'Istituto che veicolerebbe l'affiliazione;
- passaggio in Senato Accademico per approvazione finale.

Ruoli e Responsabilità

Il responsabile scientifico di un progetto

Il responsabile scientifico dovrà fornire le informazioni utili a delineare in modo dettagliato il progetto: dovranno essere esplicitati gli obiettivi di ricerca ed il prevedibile sviluppo del progetto proposto. Dovrà essere specificato se e in che misura siano coinvolti nel progetto soggetti o enti esterni al Dipartimento, con la descrizione del relativo ruolo nelle diverse fasi progettuali.
Qualora vengano trattati dati personali, in forza del DR 200/2019 il responsabile scientifico è designato responsabile interno del trattamento dei dati e dovrà relazionarsi con il responsabile della protezione dei dati della Scuola (Data Protection Officer email: dpo@santannapisa.it) **prima** di attivare il flusso, descrivendo, in particolare, le categorie di dati processate, gli interessati, le finalità e i mezzi del trattamento, la *governance* del flusso (titolare del trattamento dati, responsabile, contitolare se il trattamento coinvolge soggetti esterni alla Scuola), le misure di sicurezza, l'esito della valutazione del rischio, laddove pertinente.

L'amministratore di sistema VM EMbeDS

L'amministratore di sistema si impegna a mettere in atto misure tecniche per garantire un livello di sicurezza adeguato agli standard ICT della Scuola Superiore Sant'Anna[1],[2] (Vedi Appendice A). È una figura essenziale per la sicurezza delle banche dati e la corretta gestione e manutenzione delle risorse EMbeDS a lui assegnate.

L'amministratore di una macchina virtuale EMbeDS avrà credenziali con privilegi di amministratore e sarà responsabile per:

- Gestione degli utenti nel sistema (messa a punto e il mantenimento degli account);
- Verifica del buon funzionamento della VM e delle periferiche;
- Organizzare la riparazione in occasione di un guasto software del sistema;
- Monitoraggio delle prestazioni del sistema;
- Installare SOLO il software necessario alle finalità del progetto;
- Creare un backup dei dati critici e la relativa politica per il loro recupero;
- Monitorare la connessione di rete (in strutture complesse esiste l'amministratore di rete);
- Aggiornamento del sistema operativo e software applicativo;
- Attuare le politiche per l'utilizzo del sistema informatico e della rete;
- Protezione dei dati e lo stato di messa in sicurezza dei server (Firewall, IPS)

Il Responsabile IT EMbeDS

Il responsabile IT EMbeDS è responsabile per:

- Consulenza tecnico-scientifica su progetti EMbeBS
- Supporto servizi EMbeDS (DB, repository, containers)
- Gestione e Aggiornamento Server EMbeDS e sistema di virtualizzazione

[1] https://intranet.santannapisa.it/system/files/2018-04/Attuazione%20delle%20misure%20di%20sicurezza%20ICT%20rev05%20firmato.pdf
[2] https://www.sans.org/security-resources/policies/server-security/pdf/server-security-policy

- Registro dei Progetti EMbeDS
- Rilascio risorse EMbeDS
- Creazione l'account di amministratore di sistema
- Monitoraggio risorse
- Gestione NFS server

Managing Board

Il Managing Board EMbeDS è costituito da:

- Prof.ssa Francesca Chiaromonte (coordinatrice scientifica)
- Prof. Alessandro Nuvolari
- Prof. Andrea Piccaluga
- Prof. Fabio Iraldo
- Prof.ssa Sabina Nuti
- Direttore Generale

APPENDICE A: Misure di sicurezza dei VM EMbeDS

Misura	**Descrizione**	**Modalità**
Patching	Utilizzare le versioni supportate delle applicazioni e dei sistemi operativi e applicare gli aggiornamenti ad alto rischio entro 20 giorni dal loro rilascio da parte della sua azienda, tutti gli altri aggiornamenti entro 90 giorni	Obbligatorio
Gestione delle vulnerabilità	Eseguire una scansione delle vulnerabilità ogni 3 mesi e occuparsi delle vulnerabilità ad alto rischio entro 10 giorni	Obbligatorio
Firewall	Attivare un firewall, consentendo solo i servizi necessari	Obbligatorio
Gestione degli accessi e delle credenziali	Principio del minor privilegio (PoLP); valutare periodicamente l'accesso di ogni utente, le regole di complessità della password e rimuovere gli utenti non utilizzati	Obbligatorio
Protezione dei dati personali	Attuare i requisiti per la protezione dei dati personali (regolamenti nazionali ed europei, GDPR)	Obbligatorio
Protezione da Malware	Installare un antivirus/antimalware e tenerlo aggiornato	Obbligatorio

Sviluppo dei software	Includere aspetti di privacy e Sicurezza by default e by design	Obbligatorio
Cancellazione	Rimuovere definitivamente tutti i dati e le informazioni non necessarie dai server	Consigliato
Backup	Eseguire un backup almeno una volta alla settimana per quanto riguarda i dati e le impostazioni. Crittografare i dati trasmessi e memorizzati.	Consigliato
Crittografia delle comunicazioni	Assicuratevi che tutte le informazioni siano criptate	Consigliato
2-factor authentication	Implementare un sistema di accesso a 2 fattori per gli utenti amministrativi	Consigliato

ALLEGATO A: Modulo Richiesta Risorse EMbeDS su progetti di ricerca

Il responsabile scientifico dovrà fornire le informazioni utili a delineare in modo dettagliato il progetto: dovranno essere esplicitati gli obiettivi di ricerca ed il prevedibile sviluppo del progetto proposto. Dovrà essere specificato se e in che misura siano coinvolti nel progetto soggetti o enti esterni al Dipartimento, con la descrizione del relativo ruolo nelle diverse fasi progettuali.

Scheda tecnica

Soggetto proponente	
Responsabile scientifico della ricerca (nominativo, ruolo rivestito, esperienza pregressa e incarico attualmente rivestito - Allegare curriculum)	
Area tematica	
Titolo della ricerca	
Sede della ricerca	
Descrizione del progetto di ricerca (obiettivi, fasi, metodologie, output)	
Ricercatori (nominativi, incarico attualmente rivestito, rapporto con il soggetto proponente)	
Amministratore di sistema (nominativo)	
Durata	
Trattamento dati personali (*governance* del flusso, le misure di sicurezza, valutazione del rischio).	

CONTENERE IN UN MASSIMO DI DUE PAGINE (CV ESCLUSO)

Data e Luogo

ALLEGATO B:
Atto di nomina di Responsabile del Trattamento in qualità di Amministratore di Sistema
Provvedimento Generale del Garante per la Protezione dei Dati Personali del 27 novembre 2008
(da redigere su carta intestata del soggetto Titolare del Trattamento)

Il sottoscritto XXXX, in qualità di Titolare (o Referente Scientifico) del trattamento di dati personali operati nell'ambito della propria attività professionale, con il presente atto

designa

il Sig.re (nome, cognome e ruolo svolto in azienda se soggetto interno)
o
il Sig.re o la Società (nome, cognome, partita Iva, sede se soggetto esterno)

responsabile del trattamento in qualità di amministratore di sistema

ai sensi e per gli effetti degli articoli da 31 a 36 del D.Lgs. 30 giugno 2003 n. 196 nonché in osservanza del Disciplinare Tecnico in materia di misure minime di sicurezza di cui appendice A) del medesimo documento.

L'Amministratore di Sistema (*system administrator*) viene designato quale figura professionale dedicata alla gestione e alla manutenzione di impianti di elaborazione EMbeDS con cui vengono effettuati trattamenti di dati personali, compresi i sistemi di gestione delle basi di dati (*database administrator*), la sicurezza del sistema e tutti gli applicativi installati.

Il Sig.re o la Società ……., nella qualità di Amministratore di Sistema ha il potere e il dovere di compiere tutto quanto si renderà necessario ai fini del rispetto e della corretta applicazione del D.Lgs. 30 giugno 2003 n. 196, con particolare riferimento al profilo relativo alla sicurezza nella custodia e nel trattamento dei dati personali.

La designazione del Sig.re …….. o della Società ……. avviene in ragione del possesso in capo a quest'ultimo/a dei requisiti di capacità tecniche, professionali e di condotta.
Specificatamente, l'Amministratore di Sistema sarà tenuto a:

1 - classificare analiticamente le banche dati ed impostare/organizzare un sistema complessivo di trattamento dei dati personali comuni e sensibili che riguardi tutte le operazioni richiamate dall'art. 4, comma 1, lett. a) nessuna esclusa, predisponendo e curando ogni relativa fase applicativa nel rispetto della normativa vigente in materia di protezione dei dati personali;
2 – individuare per iscritto il/i soggetto/i incaricato della custodia delle parole chiave per l'accesso al sistema informativo e vigilare sulla sua attività;
3 – individuare per iscritto gli altri soggetti, diversi dal/dagli incaricato/i della custodia delle parole chiave, che possono avere accesso ad informazioni che concernono le medesime;
4 – impostare e gestire un sistema di autenticazione informatica per i trattamenti di dati personali effettuati con strumenti elettronici al GDPR;
5 - adottare un sistema idoneo alla registrazione degli accessi logici (autenticazione informatica) ai sistemi di elaborazione e agli archivi elettronici; le registrazioni (*access log*) devono avere caratteristiche di completezza, inalterabilità e possibilità di verifica della loro integrità adeguate al raggiungimento dello scopo di verifica per cui sono richieste; tali registrazioni devono comprendere i riferimenti temporali e la descrizione dell'evento che le ha generate e devono essere conservate per un congruo periodo, non inferiore a sei mesi;

6 - assicurare e gestire sistemi di salvataggio e di ripristino dei dati (backup/recovery), anche automatici nonché approntare adeguate misure e/o sistemi software di salvaguardia per la protezione dei dati personali (antivirus, firewall, IDS);
7 – impartire a tutti gli incaricati istruzioni organizzative e tecniche che prevedano le modalità di utilizzo dei sistemi di salvataggio dei dati con frequenza almeno settimanale;
8 – adottare procedure per la custodia delle copie di sicurezza dei dati e per il ripristino della disponibilità dei dati e dei sistemi;
9 - organizzare i flussi di rete, la gestione dei supporti di memorizzazione, la manutenzione hardware, la verifica di eventuali tentativi di accessi non autorizzati al sistema provenienti da soggetti terzi quali accesso abusivo al sistema informatico o telematico (articolo 615 *ter*), frode informatica (articolo 640 *ter*), danneggiamento di informazioni, dati e programmi informatici (articoli 635 *bis* e *ter*), danneggiamento di sistemi informatici e telematici (articoli 635 *quater* e *quinques*);
10 – predisporre, anche in contraddittorio con il Titolare dei trattamenti, un piano di controlli periodici, da eseguirsi con cadenza almeno semestrale, atti a verificare l'efficacia delle misure di sicurezza adottate in azienda/studio professionale;
11 – coadiuvare, se richiesto, il Titolare del trattamento nella predisposizione e/o aggiornamento e/o integrazione del Documento Programmatico sulla Sicurezza (D.P.S.) nonché alla stesura del documento denominato "Disciplinare in Materia di Utilizzo di Strumenti Informatici".

E' compito dell'Amministratore di Sistema monitorare costantemente lo stato di sicurezza di tutti i processi di elaborazione dati di cui sopra, mantenendo aggiornati tutti i supporti hardware e software e, se del caso, comunicando al Titolare tutte le attività da porre in essere al fine di garantire un adeguato livello di sicurezza in proporzione alla tipologia e quantità dei dati personali trattati.

L'operato dell'Amministratore di Sistema sarà oggetto, con cadenza annuale, ad una attività di verifica da parte del Titolare del trattamento, tesa a controllare la sua rispondenza alle misure organizzative, tecniche e di sicurezza rispetto ai trattamenti dei dati personali previsti dalle norme vigenti.

Pisa, lì

Il Titolare

L'Amministratore di Sistema

La preghiamo di restituirci copia della presente, firmata per accettazione.