



CERTIFICATI BIANCHI

Allegato 2.12 alla Guida Operativa

Guide Settoriali

CENTRI DI ELABORAZIONE DATI (CED)

INDICE

1	DESCRIZIONE DEI CED	3
1.1	Infrastruttura IT	5
1.1.1	<i>Server (Elaborazione)</i>	7
1.1.2	<i>Storage (Archiviazione)</i>	8
1.1.3	<i>Classificazione in base alla struttura dei dati</i>	8
1.1.4	<i>Rete (Comunicazione)</i>	10
1.2	Impianti elettrici e cablaggi	10
1.2.1	<i>Impianti Elettrici</i>	10
1.2.2	<i>Cablaggio</i>	11
2	DESCRIZIONE DELLE MIGLIORI TECNOLOGIE DISPONIBILI E DEGLI INTERVENTI DI EFFICIENZA ENERGETICA	13
2.1	Efficienza energetica: aspetti relativi al raffreddamento	13
2.1.1	<i>Ottimizzazione delle condizioni termoigrometriche</i>	14
2.1.2	<i>Efficienza nella distribuzione dell'aria</i>	14
2.1.3	<i>Free cooling</i>	16
2.1.4	<i>Altre tecnologie</i>	18
2.2	EFFICIENZA ENERGETICA: ASPETTI RELATIVI ALLA COMPONENTE INFORMATICA	20
2.2.1	<i>Consolidamento dei server tramite virtualizzazione e riduzione dell'hardware sottoutilizzato</i>	23
2.2.2	<i>Allocazione dinamica delle risorse</i>	23
2.2.3	<i>Scheduling energetico e ottimizzazione predittiva</i>	24
2.2.4	<i>Ottimizzazione dei flussi di rete e data locality</i>	24
2.2.5	<i>Ottimizzazione e virtualizzazione dello storage</i>	25
2.2.6	<i>Gestione ottimizzata dei consumi</i>	27
3	INDICI ENERGETICI	29
4	ALGORITMI PER IL CALCOLO DEL RISPARMIO ENERGETICO	30
4.1	Indicatore del servizio utile: il Workload funzionale erogato, WLE	30
4.2	Grandezze e indicatori utilizzati	31
4.3	Algoritmi per Data Center nel caso di nuova installazione	32
4.4	Algoritmi per Data Center nel caso di sostituzione e di efficientamento integrato	33
5	BIBLIOGRAFIA	35

1 DESCRIZIONE DEI CED

I Centri di Elaborazione Dati (CED), o Data Center, rappresentano luoghi fisici in cui ospitare l'infrastruttura tecnologica e gli impianti necessari al corretto funzionamento degli apparati (e.g., climatizzazione, alimentazione, antincendio, apparati di interconnessione) per la gestione, l'elaborazione e l'archiviazione delle informazioni digitali. Essi costituiscono il cuore operativo dei sistemi informativi di enti pubblici e privati, garantendo la continuità dei servizi digitali, la sicurezza dei dati e l'erogazione delle applicazioni critiche.

Un CED è composto da un insieme integrato di server, sistemi di storage, apparati di rete e componenti di sicurezza fisica e logica, alloggiati in ambienti controllati che assicurano condizioni ottimali di funzionamento. Le apparecchiature IT operano in rack o armadi dedicati, organizzati in file che favoriscono un'efficiente distribuzione del flusso d'aria e consentono di gestire in modo ottimale il raffreddamento e la manutenzione. Il corretto esercizio del centro dati richiede la presenza di impianti di raffreddamento ad alta efficienza – come sistemi di condizionamento di precisione, free cooling o soluzioni a contenimento dei corridoi caldi/freddi – indispensabili per mantenere la temperatura operativa entro limiti di sicurezza e prevenire surriscaldamenti che possano compromettere l'affidabilità dei sistemi. Inoltre, il CED richiede altresì la distribuzione elettrica ridondata, sistemi di continuità (UPS) e gruppi elettrogeni, che garantiscono stabilità e affidabilità anche in caso di interruzioni di rete o sovraccarichi.

A fine 2024 l'Italia dispone quindi di 513 MW "da computing", il 17% in più rispetto al 2023. Questa crescita è interessante anche in prospettiva, considerando che nel biennio 2023-2024 sono stati spesi circa 5 miliardi di euro per la costruzione, l'approntamento e il riempimento di server IT di nuovi spazi nei Data Center, il Politecnico di Milano stima che gli investimenti saliranno a 10,1 miliardi nel biennio 2025-2026 e che un numero crescente di nuove aperture di Data Center siano previste anche oltre il 2026¹.

In Italia, sono attualmente censiti 153 Data Center attivi, 50 a Milano, 16 a Roma, 10 a Torino, 6 a Padova e diverse città con 4 (Bari, Bologna, Genova). Più del 70% dei siti si trova nelle città del Nord (Emilia-Romagna compresa), il resto nei pressi delle principali città del Centro-Sud².

La capacità dei Data Center commerciali italiani raggiungerà i 1.522 MW, quasi sei volte quella del 2024, con un Cagr (Compound Annual Growth Rate) medio del 40%. Attualmente, sono in costruzione impianti per 343 MW, mentre 1.684 MW risultano già pianificati³.

Dal punto di vista energetico, il funzionamento di un CED comporta consumi elevati legati non solo all'alimentazione dei server, ma anche ai sistemi ausiliari di condizionamento e sicurezza. Un approccio integrato che consideri il CED come un ecosistema complesso – tecnologico, energetico e gestionale – consente di ridurre l'impatto ambientale, migliorare la sostenibilità e garantire al contempo elevati livelli di affidabilità operativa.

1 <https://www.impresacity.it/news/33716/data-center-litalia-cresce-con-qualche-difficolta.html>

2 https://www.key4biz.it/data-center-in-italia-ce-ne-sono-gia-153-ecco-la-lista-a-roma-e-milano/515539/?utm_source=chatgpt.com

3 https://www.corrierecomunicazioni.it/report/data-center-in-italia-e-boom-di-investimenti-e-nuovi-hub-regionali-600-di-capacita-entro-il-2031/?utm_source=chatgpt.com

Esistono diversi tipi di Data Center, che possono essere classificati nelle seguenti categorie:

- *Data Center Enterprise*: sono gestiti da aziende o istituzioni ad uso interno. Generalmente presentano dimensioni ridotte e un livello di efficienza inferiore rispetto ad altri tipi di Data Center;
- *Data Center di Colocation e Service Provider*: offrono spazi in affitto ai clienti per ospitare la propria infrastruttura di calcolo e storage (colocation) oppure forniscono sia lo spazio sia l'equipaggiamento informatico (service provider). Entrambe le tipologie possono ospitare centinaia o migliaia di clienti;
- *Data Center Hyperscale*: sono strutture di grandi dimensioni gestite dalle principali aziende tecnologiche. Utilizzano infrastrutture scalabili e ad alta efficienza per supportare servizi cloud, web hosting e, in misura crescente, applicazioni di intelligenza artificiale (AI).

Un'ulteriore classificazione dei Data Center si basa sul modello a 4 Tier⁴, che rappresentano i diversi livelli di affidabilità, ridondanza e resilienza dell'infrastruttura. A ciascun Tier corrisponde un livello tecnico e strutturale del Data Center, nonché la sua capacità di garantire la business continuity in presenza di circostanze avverse. In generale, maggiore è il Tier, più elevate sono le prestazioni che la struttura può offrire, in termini di sicurezza fisica, cybersecurity e sostenibilità ambientale/efficienza energetica.

Ogni livello è definito da una specifica normativa Tier, che regola sia la fase di progettazione sia quella di implementazione. L'Uptime Institute utilizza tali normative come riferimento per le proprie valutazioni di conformità.

- Tier 1: singolo percorso per l'alimentazione elettrica e il raffreddamento, con componenti ridondanti o di backup limitati. Uptime previsto pari a 99,671% mentre il Downtime annuo stimato è circa 28,8 ore.
- Tier 2: singolo percorso per alimentazione e raffreddamento, con presenza di alcuni componenti ridondanti e di backup. Uptime previsto pari a 99,741% mentre il Downtime annuo stimato è circa 22 ore.
- Tier 3: percorsi multipli per alimentazione e raffreddamento, con sistemi che consentono operazioni di manutenzione senza interrompere i servizi. Uptime previsto è pari a 99,982% mentre il Downtime annuo stimato è pari a circa 1,6 ore.
- Tier 4: infrastruttura completamente fault tolerant, con ridondanza totale di ogni componente. Uptime previsto: 99,995% mentre il Downtime annuo stimato è circa 26,3 minuti.

Le componenti di un Data Center possono essere così classificate [2]:

- *Componente IT*: costituita dagli elementi di storage, server e networking, che saranno approfonditi nella sezione 1.1;
- *Sistemi di raffreddamento e controllo ambientale*: comprendono le apparecchiature che regolano temperatura e umidità per mantenere l'equipaggiamento IT in condizioni operative ottimali. La quota dei sistemi di raffreddamento sul consumo totale dei Data Center varia da circa il 7% per i Data Center hyperscale efficienti a oltre il 30% per i Data Center enterprise meno efficienti;
- *Sistemi di alimentazione di emergenza (UPS) e generatori di backup*: garantiscono l'alimentazione del Data Center durante eventuali interruzioni di corrente. Sebbene siano raramente utilizzati, sono indispensabili per assicurare gli elevati livelli di affidabilità richiesti dai Data Center.

⁴ <https://tiaonline.org/products-and-services/tia942certification/ansi-tia-942-standard/>

- *Altre infrastrutture*: ad esempio gli impianti di illuminazione, ecc.

La quota di consumo elettrico attribuibile a queste diverse componenti varia notevolmente a seconda del tipo di Data Center, in funzione della natura e dell'efficienza delle apparecchiature installate. In particolare, dipendono:

- *dalla Capacità IT installata*, ovverosia dai server, dispositivi di storage e di rete operativi, ed è misurata in megawatt (MW). La capacità totale installata include sia la capacità IT sia la capacità elettrica delle apparecchiature ausiliarie. In molti casi, i Data Center non sono completamente riempiti di server. La capacità massima progettata indica la capacità massima di un Data Center qualora fosse completamente riempito di server; in molti casi, essa risulta inferiore alla capacità totale installata;
- *dal Tasso di utilizzo dell'equipaggiamento IT*: misura la quantità di risorse di calcolo effettivamente utilizzate in un determinato periodo. Ad esempio, i Data Center enterprise più piccoli e meno efficienti presentano tassi medi di utilizzo inferiori al 20%, mentre i Data Center hyperscale con carichi ottimizzati possono raggiungere tassi medi fino al 50%;
- *dalla Potenza a vuoto (idle power)*: rappresenta la quantità di elettricità consumata da un dispositivo per svolgere operazioni di base quando non elabora attivamente carichi di lavoro. È solitamente espressa come percentuale della potenza nominale massima. Una potenza a vuoto inferiore indica una maggiore efficienza.

1.1 Infrastruttura IT

La componente IT comprende l'insieme delle apparecchiature e dei sistemi dedicati all'elaborazione, all'archiviazione e alla trasmissione dei dati, come riportato in Figura 1 e 2. Tali dispositivi sono ospitati all'interno di armadi detti "rack", che accolgono server, sistemi di storage e apparati di rete, ottimizzando l'utilizzo dello spazio verticale, che viene misurato in "unit (u)". L'infrastruttura di cablaggio prevede l'utilizzo di percorsi e condotti dedicati, che sono tipicamente collocati sotto il pavimento flottante o al di sopra dei rack al fine di prevenire l'ingombro dei cavi, agevolare le operazioni di manutenzione e favorire un corretto flusso d'aria e conseguentemente un migliore raffreddamento delle apparecchiature.

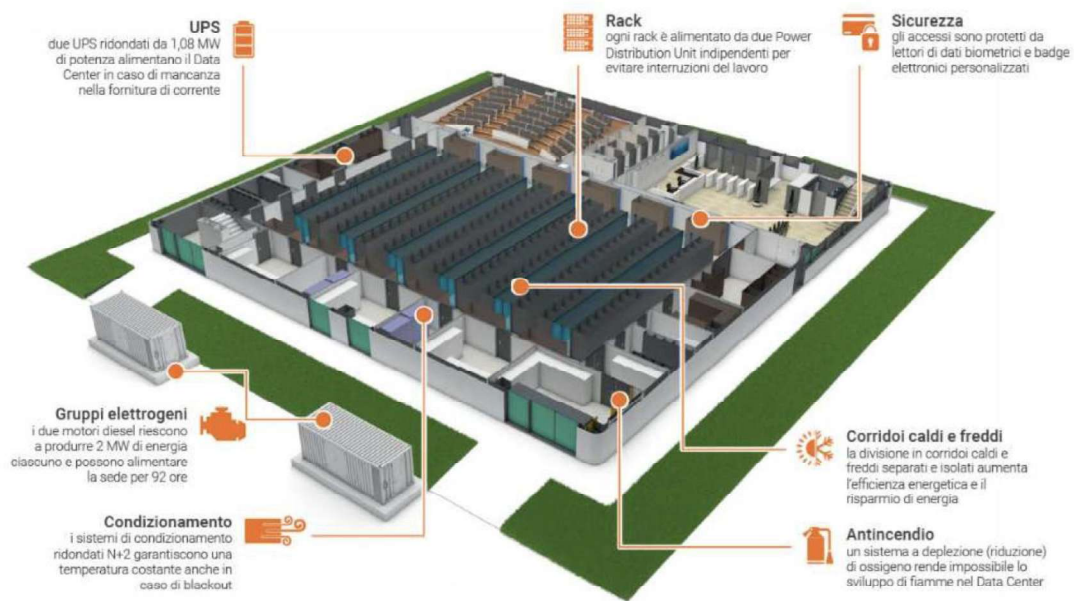


Figura 1 - Datacenter Serverfarm AltoAdige Hosting. Fonte: <https://www.citconsult.it/datacenter>

Gli apparati IT possono essere distinti nelle seguenti categorie principali:

- **Server**: macchine fisiche o virtuali, che processano e memorizzano dati, destinate all'esecuzione di applicazioni e processi di calcolo. Esse possono essere equipaggiate con Central Processing Units (CPUs) e/o acceleratori specializzati, quali ad esempio Graphic Processing Units (GPUs). In media, il loro contributo è pari a circa il 60% della domanda di energia nei moderni Data Center, sebbene tale percentuale dipenda fortemente dal tipo di Data Center considerato [2]. L'impiego di tecnologie di virtualizzazione e containerizzazione consente di ottimizzare l'uso delle risorse hardware, aumentare la flessibilità operativa e ridurre la frammentazione dei carichi di lavoro;
- **Storage**: sistemi dedicati alla memorizzazione dei dati, sia strutturati (ad esempio database) sia non strutturati (file, oggetti). Il loro contributo è pari a circa il 5% della domanda di energia [2].
- **Rete**: apparati di comunicazione, quali switch, router e firewall, che garantiscono la connettività interna ed esterna, assicurando prestazioni, sicurezza e affidabilità della rete. Il loro contributo è pari a circa il 5% della domanda di energia [2].

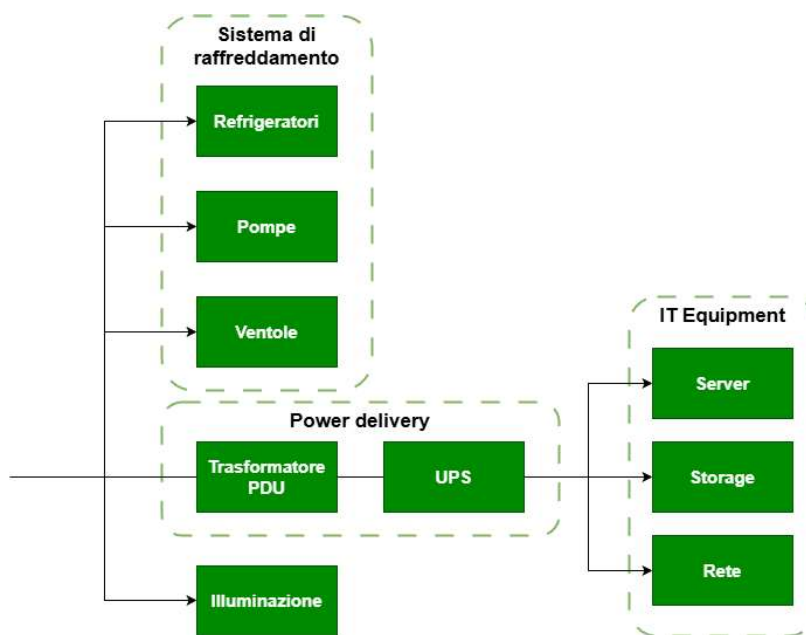


Figura 2 - Componenti principali di un Data Center [3]

1.1.1 Server (Elaborazione)

I Server sono le macchine fisiche o logiche dedicate all'esecuzione di applicazioni e processi di calcolo, fungendo da potenza di elaborazione (*compute*) essenziale del Data Center.

Essi possono essere classificati in due macrocategorie: server fisici o virtuali. I server fisici, anche chiamati bare metal, rappresentano l'hardware fisico dedicato su cui un sistema operativo è installato direttamente senza alcuno strato di virtualizzazione intermedio. Tali server possono essere singole unità dotati di alimentazione e raffreddamento propri da inserire in rack oppure moduli ultracompatti (blade) che condividono chassis, alimentazione e I/O per massimizzare la densità di calcolo.

Invece, i server virtuali (Virtual Machines (VM)) sono ambienti operativi completi creati e gestiti da un software chiamato Hypervisor (installato sul server *Bare Metal*), come VMware ESXi o KVM. Ogni VM emula le componenti hardware di un server fisico (CPU virtuale, RAM, disco virtuale) e ospita il proprio sistema operativo e le proprie applicazioni. Il loro vantaggio è la flessibilità ed ottimizzazione (provisioning) nell'uso di risorse hardware del server fisico sottostante.

La Virtualizzazione è il processo di creazione di una rappresentazione basata sul software, anziché fisica, di una risorsa informatica. Questa può applicarsi a diversi livelli dell'infrastruttura IT:

- *Virtualizzazione di Storage (Software-Defined Storage - SDS)* astrae e aggrega la capacità di storage fisica da dispositivi diversi (es. array SAN multipli o dischi locali) e la presenta come un pool di risorse unificato. Permette la gestione centralizzata, l'allocazione dinamica e il provisioning dello storage indipendentemente dall'hardware sottostante. Ha come obiettivo quello di aumentare la flessibilità, la scalabilità e la resilienza dei dati.

- *Virtualizzazione di Rete (Software-Defined Networking - SDN)* separa il piano di controllo (la logica che decide come instradare il traffico) dal piano dati (che si occupa del forwarding effettivo dei pacchetti) degli apparati di rete (switch e router). Ciò consente di configurare, gestire e segmentare la rete tramite software (ad esempio, creando reti virtuali, o VLAN, o utilizzando Network Function Virtualization - NFV per virtualizzare firewall e bilanciatori di carico). Ha come obiettivo quello di semplificare la gestione della rete, velocizzare il deployment e migliorare la sicurezza creando segmenti logici isolati.

La maggior parte dei Data Center moderni opera in modalità virtualizzata al fine di ottimizzare l'utilizzo delle risorse, permettendo al server *Bare Metal* di eseguire decine o centinaia di VM contemporaneamente, creando ambienti applicativi isolati e portabili.

1.1.2 Storage (Archiviazione)

I sistemi di storage costituiscono uno dei pilastri fondamentali di un Data Center, in quanto responsabili della memorizzazione persistente, affidabile e accessibile dei dati gestiti dalle applicazioni e dai servizi IT. Essi si differenziano in base alla struttura dei dati, al livello di prestazioni richiesto, al tipo di supporto fisico e al modello di connessione utilizzato.

1.1.3 Classificazione in base alla struttura dei dati

È possibile classificare lo storage in storage strutturato e non strutturato in base alla struttura dei dati. I primi riguardano i dati strutturati e indicizzati, come tabelle relazionali, anagrafiche, registri contabili e flussi di transazioni, che richiedono elevata velocità di I/O, bassa latenza e disponibilità costante. Per la loro realizzazione si progettano tipicamente Storage Area Network (SAN) ad alte prestazioni che utilizzano protocolli specializzati come Fibre Channel (FC) o iSCSI e si basano su connessioni dedicate punto-punto o fabric con switch FC e Host Bus Adapter (HBA) installati nei server. In queste architetture, lo storage viene esposto come blocchi logici (LUN) direttamente gestiti dai sistemi operativi o dall'hypervisor garantendo prestazioni paragonabili ai dischi locali ma con maggiore affidabilità e scalabilità.

Invece, lo storage non strutturato comprende la gestione di contenuti eterogenei, come ad esempio documenti, immagini, video, log, file multimediali e dati generati da sensori IoT. Questi dati non richiedono strutture rigide né transazioni complesse e vengono tipicamente gestiti tramite Network Attached Storage (NAS) o object storage. I sistemi NAS, basati su protocolli standard come Network File System (NFS) o SMB/CIFS permettono la condivisione dei file su rete ethernet e offrono un buon compromesso tra capacità, costi e semplicità di gestione. Le piattaforme di object storage adottano un approccio scalabile e distribuito, ideale per l'archiviazione massiva e per applicazioni cloud native.

1.1.3.1 Classificazione in base alla tipologia di supporto fisico

L'efficienza, la capacità e i consumi energetici di un sistema di archiviazione dipendono in larga parte dal tipo di supporto fisico utilizzato. Essi possono essere classificati in Hard Disk Drive (HDD) e Solid State Drive (SSD). Gli HDD sono basati su piatti magnetici rotanti e testine di lettura/scrittura, offrono grandi capacità a basso costo per terabyte sebbene presentano tempi di accesso più elevati (in media 5–10 ms) e un consumo energetico costante, anche in idle, dovuto alla rotazione continua dei dischi. Essi sono ideali per archiviazione massiva, backup, o repository di dati freddi, cioè dati che non vengono acceduti spesso dove la priorità è la capacità, non la velocità. Gli hard disk si distinguono anche in base alla velocità di rotazione dei piatti,

misurata in giri al minuto (RPM), la quale influisce direttamente sulle prestazioni, sulla latenza e sul consumo energetico. In particolare:

- 5400 RPM: hard disk lenti ma economici, con basso consumo energetico. Sono utilizzati principalmente per archiviazione di massa e backup, dove la velocità non è un fattore critico;
- 7200 RPM: hard disk che rappresentano un buon equilibrio tra prestazioni e costo. Sono comuni nei server general-purpose, nei desktop e nei NAS di fascia media;
- 10.000 RPM: hard disk che offrono alte prestazioni, tempi di accesso più ridotti e una maggiore velocità di trasferimento dati. Utilizzati in server ad alte prestazioni e applicazioni dove la rapidità di accesso ai dati è essenziale;
- 15.000 RPM: hard disk più veloci tra gli HDD tradizionali, oggi meno diffusi perché sostituiti dagli SSD. Erano tipicamente impiegati in Storage Area Network (SAN) e in sistemi enterprise con carichi di lavoro intensivi.

Gli SSD utilizzano memoria flash NAND, garantendo latenze inferiori al millisecondo e throughput elevatissimo. Sebbene siano molto efficienti dal punto di vista energetico nei carichi intensivi, essendo il consumo per operazione I/O molto inferiore rispetto agli HDD, hanno una capacità inferiore rispetto agli HDD. Inoltre, gli SSD basati su interfacce NVMe collegati tramite PCIe, rappresentano oggi lo standard per gli ambienti virtualizzati e database ad alte prestazioni sebbene i loro costi per unità capacitiva rimangano molto elevati rispetto alla controparte, e la loro durata dipende dal numero di cicli di scrittura (wear-leveling).

Nei Data Center moderni, la soluzione ottimale è spesso un approccio ibrido (tiered storage), che combina HDD per i dati statici e SSD per i carichi di lavoro ad alte prestazioni. I software di gestione automatica del tiering spostano dinamicamente i dati tra i diversi livelli in base alla frequenza di accesso, ottimizzando contemporaneamente prestazioni e consumi energetici.

Con l'evoluzione verso architetture virtualizzate e cloud, lo storage fisico è sempre più un ambiente eterogeneo che viene gestito in modo centralizzato tramite opportuni virtualizzatori di storage, che consentono di aggregare dischi fisici e risorse distribuite in un unico pool logico di archiviazione, scalabile e resiliente. Essi interagiscono con i server fisici e le SAN mediante HBA o interfacce NVMe over Fabrics (NVMe-oF), che estendono le prestazioni NVMe su rete Ethernet o InfiniBand o switch SAN o fabric Ethernet dedicati con ridondanza e bassa latenza. I virtualizzatori di storage permettono la virtualizzazione dei volumi e delle LUN, trattandoli come risorse software non legate a un singolo apparato fisico. Essi gestiscono automaticamente la replica e la deduplicazione dei dati, ottimizzano il caching tra SSD e HDD e distribuiscono dinamicamente i carichi di I/O, evitando colli di bottiglia e riducendo i consumi concentrati su un singolo nodo.

I sistemi Data Center Infrastructure Management (DCIM) possono monitorare in tempo reale: l'assorbimento energetico dei controller storage e degli shelf di dischi, la temperatura dei rack dedicati allo storage e il carico sulle PDU e la correlazione tra IOPS effettivi e potenza assorbita.

Tali dati vengono poi utilizzati per ottimizzare la distribuzione dei workload, spostando dinamicamente i carichi intensivi verso sistemi più efficienti o meno saturi, e per adattare il raffreddamento in base alla reale dissipazione termica degli apparati.

1.1.4 Rete (Comunicazione)

La componente di rete rappresenta il tessuto connettivo critico che abilita lo scambio di dati tra server e risorse di archiviazione all'interno del Data Center. L'organizzazione della componente di rete è cruciale per la scalabilità e la performance del Data Center stesso. L'architettura prevalente nei Data Center è costituita da un approccio *spine-leaf* nel quale vi sono switch Leaf, che, collocati in cima ad ogni rack, connettono i server e le unità di storage presenti in un rack tramite connessioni in fibra ottica o rame, e switch Spine, apparati centrali ad alta velocità che connettono tutti gli apparati leaf. Ogni switch Leaf è connesso a ogni switch Spine in modo da assicurare che il traffico tra due server in rack diversi debba attraversare un massimo di due salti. Infine, router e firewall possono essere utilizzati per gestire il traffico tra reti eterogenee e implementano le politiche di sicurezza perimetrale.

Per ottimizzare le risorse e migliorare la sicurezza, i moderni sistemi di rete integrano concetti di virtualizzazione a livello logico, chiamate Virtual Local Area Network (VLAN). La VLAN è una tecnica di segmentazione logica che permette di suddividere un singolo switch fisico o una rete in più reti virtuali isolate. Un'altra tecnica di virtualizzazione di rete è la Software-Defined Networking (SDN) che consente di astrarre l'infrastruttura di rete dall'hardware fisico. Tale tecnica semplifica la gestione della rete che è effettuata tramite software senza la necessità di riconfigurare fisicamente gli apparati. Questo abilita la creazione di overlay network flessibili e la gestione dinamica dei flussi di dati, essenziale per gli ambienti multi-tenant e cloud. Infine, una parte essenziale della connettività è dedicata al Management Out-of-Band (OOB), una rete fisica separata dalla rete principale che ha principalmente il compito di monitorare la salute dei sistemi e gestire remotamente l'alimentazione.

1.2 Impianti elettrici e cablaggi

Le componenti IT non possono operare senza una solida infrastruttura di supporto, che include la gestione dell'energia e delle interconnessioni fisiche.

1.2.1 Impianti Elettrici

L'infrastruttura elettrica di un Data Center costituisce la spina dorsale dell'intero sistema, garantendo la continuità di alimentazione e la sicurezza operativa dei carichi critici. Essa è progettata secondo criteri di ridondanza, affidabilità e scalabilità, in conformità agli standard internazionali (ad esempio TIA-942, EN 50600, Uptime Institute Tier Classification System). L'alimentazione dell'infrastruttura IT è il requisito più critico per la continuità operativa. Se la potenza apparente massima del Data Center è superiore a 10–20 MVA, il Data Center è collegato alla rete esercita in alta tensione (AT), mediante una stazione AT/MT dedicata e installata in loco. Se la potenza è inferiore a tale soglia, il collegamento avviene in media tensione (MT) e il Data Center sarà dotato di una cabina di trasformazione MT/BT. L'alimentazione dei server e delle apparecchiature IT avviene in bassa tensione (BT).

L'energia elettrica è distribuita attraverso quadri elettrici e Power Distribution Units (PDU) installate all'interno di ogni armadio rack. I quadri generali di bassa tensione distribuiscono l'energia verso i vari sottosistemi del Data Center e sono dotati di interruttori automatici, sezionatori e sistemi di protezione

contro sovratensioni e cortocircuiti. Tutti gli apparati elettrici e IT sono previsti di sistema di messa a terra unico. Inoltre, sono previsti scaricatori di sovratensione (SPD) in ogni punto strategico della rete. Nei Data Center moderni, i quadri sono predisposti per la misurazione intelligente dei carichi (smart metering) e l'integrazione con sistemi di monitoraggio energetico (DCIM).

Al fine di garantire continuità operativa, i Data Center sono progettati secondo uno schema di alimentazione ridondante (la configurazione N+1 prevede la capacità necessaria più un modulo di riserva; la configurazione 2N implica la duplicazione della capacità necessaria) per garantire che un guasto a un componente elettrico non causi interruzioni. Qualora entrambe le linee di alimentazione siano interrotte, vi sono due tipologie di apparati che supportano la continuità operativa:

- *Gruppi di continuità (UPS, Uninterruptible Power Supply)* che sono tipicamente di tipo statico e che sono costituiti da una batteria e da un insieme di dispositivi elettronici di conversione destinati a trasformare la tensione continua (CC) in tensione alternata (CA), idonea all'alimentazione delle apparecchiature informatiche collegate, e viceversa. Nell'UPS, il carico riceve normalmente l'alimentazione dalla rete elettrica tramite un doppio stadio di conversione: l'inverter è alimentato dal raddrizzatore e, in caso di interruzione della rete, la batteria, precedentemente caricata attraverso il raddrizzatore, fornisce energia ai carichi mediante l'inverter. Gli UPS includono inoltre un dispositivo di commutazione che provvede a trasferire automaticamente alla batteria la funzione dell'alimentazione in caso di indisponibilità della rete. Relativamente ai moduli di accumulo è bene evidenziare che le batterie al litio stanno progressivamente soppiantando le batterie al piombo grazie alla loro maggiore durata (fino a 10-15 anni), tempi di ricarica più rapida e minore impatto ambientale. Oltre a sopperire il carico in caso di mancata alimentazione, gli UPS possono avere anche il compito di garantire un'alimentazione di qualità ai carichi, in conformità alla norma EN 50600-2-2.;
- *Generatori di riserva* (gruppi elettrogeni), con motore primo a diesel o a gas che garantiscono la continuità operativa in caso di interruzioni prolungate, alimentando l'intera struttura per giorni se necessario. La commutazione del carico avviene mediante un sistema automatico di trasferimento (Automatic Transfer Switch - ATS) che trasferisce l'alimentazione del carico al gruppo elettrogeno in pochi secondi.

1.2.2 Cablaggio

Il cablaggio costituisce l'arteria con cui i dati sono scambiati all'interno del Data Center. Una pianificazione accurata delle dorsali, in linea con gli standard internazionali (come, ad esempio, TIA-942 o ISO/IEC 24764), consente di ridurre le complessità fisiche e le perdite associate a collegamenti ridondanti o troppo lunghi. Il cablaggio all'interno del Data Center segue una struttura gerarchica e ordinata:

- *Top-of-Rack (ToR)*: ogni rack ospita uno switch che collega i server contenuti nello stesso armadio, riducendo la quantità di cavi orizzontali;
- *End-of-Row (EoR) o Middle-of-Row (MoR)*: i server di più rack convergono verso uno switch centrale di corridoio;
- *Backbone*: connessioni in fibra ottica che collegano i vari moduli del Data Center (core network, sale server, sale UPS).

Per la posa, i cavi vengono disposti in canaline sospese o pavimenti flottanti separando i cavi elettrici da quelli dati per ridurre e facilitare la manutenzione. Ogni tratta è etichettata, testata e documentata. Tipicamente vi sono due tipi di connessioni: *inter-rack* ed *intra-rack*. Le connessioni *inter-rack* sono tipicamente realizzate mediante cavi in rame che sono tipicamente economici ma limitati in distanza e banda, o mediante fibra ottica usata per connessioni ad alta velocità, che può essere multimodale o monomodale. L'uso di fibra ottica per le connessioni di backbone ad alta capacità (40/100 GbE e superiori) è preferibile rispetto al rame, non solo per le migliori prestazioni e minori perdite, ma anche per la ridotta sensibilità alle interferenze elettromagnetiche e la minore necessità di schermature, con un impatto positivo sui consumi complessivi. Invece, le connessioni *intra-rack* sono tipicamente realizzate tramite InfiniBand, un'interconnessione ad altissima velocità progettata specificamente per ambienti di High-Performance Computing (HPC), intelligenza artificiale e sistemi di calcolo distribuito. A differenza delle tradizionali reti ethernet, InfiniBand è concepita per garantire latenze estremamente basse (nell'ordine di pochi microsecondi) e larghezze di banda molto elevate, che oggi possono superare i 400 Gbps per collegamento. Essa utilizza un'architettura a commutazione (switched fabric), in cui ogni dispositivo comunica con gli altri attraverso switch InfiniBand.

2 DESCRIZIONE DELLE MIGLIORI TECNOLOGIE DISPONIBILI E DEGLI INTERVENTI DI EFFICIENZA ENERGETICA

Con riferimento ai progetti di efficienza energetica presentati nell'ambito del meccanismo dei Certificati Bianchi, nonché alla letteratura di settore e alle soluzioni tecnologiche ad oggi installabili, si rappresenta di seguito una descrizione degli interventi di efficienza energetica relativi ai Centri di Elaborazione Dati rientranti nella tipologia di cui all'Allegato 2, Tabella 1 del D.M. 21 luglio 2025.

Tipologie di intervento del settore dei CED <i>Allegato 2, Tabella 1 del D.M.21 luglio 2025</i>	Vita utile		
	Nuova installazione	Sostituzione	Efficientamento integrato
CED	7	7	5

Tabella 1 - tipologie di intervento del settore dei CED dell'allegato 2, Tabella 1 del D.M. 21 luglio 2025

2.1 Efficienza energetica: aspetti relativi al raffreddamento

L'efficienza energetica nel raffreddamento si ottiene anzitutto applicando le best practice di progettazione e gestione termica. Queste misure permettono di ridurre le perdite (ad esempio, aria fredda bypassata o aria calda ricircolata) e di massimizzare l'efficacia del sistema di condizionamento. Le principali tipologie di sistemi di raffreddamento dei Data Center sono sintetizzate in Figura 3.

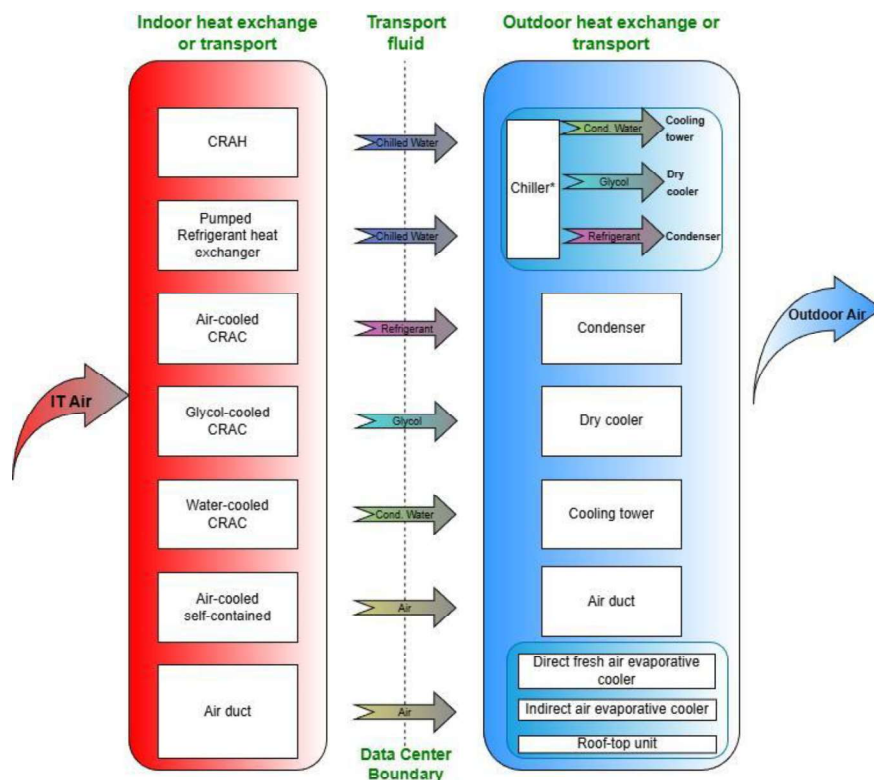


Figura 3 - Quadro sinottico delle principali tipologie di sistemi di climatizzazione dei Data Center [4].

Legenda: CRAC = Computer Room Air Conditioning, CRAH = Computer Room Air Handler. Il CRAC è un'unità di condizionamento dell'aria a espansione diretta, comprensiva di compressore frigorifero, ed è usato generalmente per impianti di piccole o medie dimensioni. Il CRAH è, in sostanza, un'unità di trattamento dell'aria, alimentata mediante un fluido freddo prodotto esternamente; ideale per Data Center di grandi

dimensioni, può integrarsi con soluzioni per il free-cooling e con sistemi di ricircolo dell'acqua, utili a ridurre il PUE.

2.1.1 Ottimizzazione delle condizioni termoigrometriche

Il mantenimento di temperature di refrigerazione eccessivamente basse determina un'inefficienza energetica sistematica. Storicamente, i Data Center tradizionali venivano gestiti adottando soglie termiche fortemente conservative, spesso inferiori ai 20 °C per i flussi d'aria in ingresso ai server. Al contrario, l'evoluzione degli standard settoriali favorisce oggi l'adozione di regimi termici più elevati. Le recenti linee guida ASHRAE stabiliscono, infatti, un intervallo raccomandato compreso tra i 18 °C e i 27 °C per l'aria in ingresso alle apparecchiature IT [5]. Tale soglia operativa può estendersi fino a un massimo di 45 °C qualora si impieghino sistemi hardware certificati in classe A4, come illustrato in Figura 2. L'orientamento verso i limiti superiori di questi intervalli genera vantaggi economici e ambientali immediati: l'incremento anche di un solo grado centigrado nella temperatura di mandata consente di ridurre il consumo energetico dei chiller in una misura stimata tra il 4% e il 5%, ottimizzando contestualmente la durata delle finestre di free-cooling [6].

La regolazione dei parametri igrometrici costituisce un'ulteriore e rilevante leva di efficientamento energetico. L'approccio gestionale tradizionale imponeva un dispendio energetico considerevole per stabilizzare l'umidità relativa entro un intervallo tipicamente compreso tra il 40% e il 55%. L'orientamento progettuale contemporaneo subordina tuttavia il fattore igrometrico alla componente termica; le direttive ASHRAE hanno infatti esteso i limiti tollerati di umidità relativa fino al 60%. Questa evoluzione normativa consente di escludere integralmente i sistemi di regolazione attiva, focalizzando l'infrastruttura sul solo controllo della temperatura. La disattivazione dei moduli di umidificazione a resistenze e dei cicli di deumidificazione forzata — ove le condizioni ambientali lo permettano — determina una contrazione sensibile dei consumi. Nei contesti climatizzati, il mantenimento di una soglia di tolleranza estesa (ad esempio 20–80%), o preferibilmente il monitoraggio del punto di rugiada (dew point), risulta sufficiente a inibire fenomeni di scarica elettrostatica o di condensazione. Di conseguenza, la parzializzazione o la totale sospensione del controllo igrometrico attivo si configura attualmente come una best practice settoriale consolidata.

2.1.2 Efficienza nella distribuzione dell'aria

Esistono tre approcci di base per distribuire l'aria in un Data Center [8]:

- *sistema allagato*: gli unici vincoli al flusso d'aria di mandata e di ripresa sono le pareti, il soffitto e il pavimento della sala. Ciò comporta un forte mescolamento tra aria calda ed aria fredda;
- *sistema mirato*: un meccanismo (per esempio una canalizzazione, una piastrella forata o un'unità di raffreddamento inserita tra le file IT) convoglia il flusso d'aria di mandata e ripresa entro 3 metri dalle prese d'aria e dagli sfiati delle apparecchiature IT;
- *sistema contenuto*: il flusso d'aria di mandata e ripresa delle apparecchiature IT è completamente confinato, eliminando ogni mescolamento tra le correnti d'aria calda e quelle fredde.

Ciascuno dei tre approcci può essere adottato sia sul percorso di mandata sia su quello di ripresa. Questo genera nove possibili combinazioni, o tipi di distribuzione dell'aria, tutti effettivamente impiegati nella pratica; talvolta, approcci diversi coesistono nello stesso Data Center.

Le modalità di implementazione della distribuzione dell'aria e le pratiche di gestione del flusso influenzano fortemente la densità di potenza raggiungibile in ciascuno dei nove tipi. Ad esempio, le unità di raffreddamento possono essere posizionate lungo il perimetro oppure all'interno delle file di rack o all'esterno.

Allineare i rack in file alternate fronte-retro, secondo lo schema hot aisle/cold aisle consente di ottenere un raffreddamento efficiente. In questo layout, i lati frontali dei server, detti prese d'aria, si fronteggiano formando corridoi freddi, mentre sul retro l'uscita di aria calda forma corridoi caldi. Così, l'aria fredda immessa viene aspirata solo dalle prese frontali, e l'aria calda espulsa dai retro si accumula nel corridoio caldo.

Questa semplice disposizione può aumentare significativamente la capacità di raffreddamento disponibile senza modificare gli impianti, semplicemente evitando che l'aria calda di scarico torni a mescolarsi con l'aria fredda di mandata. In termini di efficienza, il corretto layout a corridoi caldo/freddo può offrire risparmi energetici nell'ordine del 10–35% nel carico di raffreddamento rispetto a un ambiente non organizzato, grazie alla riduzione dei ricircoli e dei bypass.

Le best practice di settore raccomandano inoltre moduli prefabbricati di contenimento e sensori onboard per ottimizzare la gestione termica [6].

Per massimizzare il beneficio, è fondamentale implementare il contenimento dei corridoi aggiungendo barriere fisiche per separare completamente l'aria fredda da quella calda, come ad esempio: pannelli rigidi, porte, tende in PVC trasparente, ecc.. L'obiettivo in entrambi i casi è impedire ogni miscelazione tra flussi caldi e freddi. Con un buon contenimento, l'aria calda di scarico viene confinata e restituita ai climatizzatori a temperatura molto più elevata, anche intorno ai 30 °C, incrementando l'efficienza degli scambi termici e l'eventuale uso del free cooling, consentendo inoltre ai chiller di operare con COP più alti.

Inoltre, i CRAC/CRAH possono ridurre la velocità delle ventole e aumentare la temperatura di mandata quando non c'è rischio che l'aria calda contamini il corridoio freddo.

Anche il controllo della velocità dei ventilatori in base alle esigenze delle apparecchiature IT è fondamentale per ottenere risparmi. Sono ormai ampiamente disponibili variatori di velocità per i ventilatori di mandata delle unità di condizionamento ad espansione diretta (DX CRAC) delle sale server.

Ulteriori esempi di gestione efficiente che rientrano in questa categoria sono descritti di seguito: [6].

- rimozione di ostacoli e gestione dei cablaggi: sotto i pavimenti sopraelevati (plenum di mandata) spesso passano cavi che, se non ordinati, possono ostruire il flusso d'aria dai fori del pavimento. È raccomandabile l'implementazione di un piano di "cable management": mantenere libera un'altezza minima di plenum (ad es. ≥ 60 cm) e instradare i cavi preferibilmente sopra i rack o in canaline laterali, evitando matasse disordinate sotto il pavimento. Inoltre, è bene adottare un programma di "cable mining" per rimuovere periodicamente cavi non più in uso che ingombrano inutilmente. Anche le aperture per il passaggio dei cavi nel pavimento o nel soffitto vanno sigillate accuratamente con apposite guarnizioni "brush grommets", per impedire fughe di aria attraverso quei punti. Una gestione diligente dei cavi e delle vie d'aria assicura che l'aria fredda raggiunga i server con la portata prevista, evitando al contempo ricircoli di aria calda.
- bilanciamento delle portate e contenimento dei bypass: come già accennato, in molti Data Center tradizionali, una parte dell'aria fredda immessa bypassa i carichi IT, ad esempio fuoriuscendo da griglie

in zone vuote, o filtrando attraverso fessure e una parte dell'aria calda ricircola verso le prese prima di essere raffreddata. Questi fenomeni comportano inefficienze: l'aria fredda bypassata rappresenta uno spreco di energia, mentre l'aria calda ricircolata alza la temperatura di ingresso ai rack, creando hotspot. Idealmente, è necessario bilanciare la portata d'aria: un eccesso di aria in sovrappressione genera bypass inutili, mentre una portata insufficiente provoca ricircoli, poiché i server aspirano aria calda dai lati o dal retro. Si possono utilizzare piastrelle forate nei corridoi freddi e in quantità adeguata al carico, installare paratie sotto il pavimento per convogliare l'aria nelle zone desiderate, e regolare la velocità dei ventilatori CRAC/CRAH, se possibile in automatico, con VFD, per fornire l'aria necessaria, senza eccessi. L'adozione di sensori termici presso i rack, soprattutto in alto, dove l'aria è più calda, aiuta a guidare queste regolazioni: l'obiettivo è mantenere tutte le prese entro i limiti di temperatura voluti, con il minimo flusso d'aria sufficiente. Se persistono hotspot in alto, occorrerà una portata maggiore o una migliore distribuzione, se ci sono zone molto fredde in basso, potrebbe esserci eccesso di portata o perdite da occludere. Come detto, "blanking panel" nei buchi dei rack e coperture sui lati aperti di armadi non standardizzati, o dispositivi montati in modo non front-to-back, sono esempi di misure per eliminare ricircoli interni al rack.

- distribuzione dell'aria ottimizzata: nei sistemi tradizionali a pavimento sopraelevato, l'aria fredda immessa dalle griglie tende a uscire con maggior pressione nei primi fori vicino ai CRAC, lasciando gli ultimi rack potenzialmente sotto-riforniti. Per mitigare questo effetto, si possono usare ventilatori booster in-row o sistemi di air supply overhead mandati dall'alto, che distribuiscono più uniformemente l'aria per induzione. In alcuni casi, condotte dedicate o sistemi "close-coupled cooling", ad es. unità di raffreddamento in fila tra i rack, possono portare l'aria fredda più vicino ai carichi, aumentando l'efficienza. L'importante è assicurare un approvvigionamento d'aria uniforme e sufficiente a ogni presa dei server, evitando al contempo inefficienze: un ambiente ben gestito permette di mantenere temperature uniformi lungo tutto il rack, minimizzando il gradiente dal basso verso l'alto. In ambienti molto regolari, si può mirare ad avere differenze di temperatura verticali contenute (sotto i 5° C) tra base e top rack; se i corridoi sono aperti, in alto, verso fine fila, si avrà il punto più caldo, ma con un buon contenimento questo gradiente può essere pressoché annullato.

2.1.3 Free cooling

Sfruttare le condizioni ambientali favorevoli per raffreddare i Data Center – il cosiddetto free cooling – è una delle strategie più efficaci in termini di risparmio energetico, permettendo di ridurre al minimo l'uso dei chiller quando l'aria esterna o l'acqua di torre è sufficientemente fredda da assicurare un adeguato raffreddamento [1, 6, 7, 4, 9, 10].

Si possono distinguere due approcci, descritti di seguito.

- *economizzatore ad aria (air-side economizer)*: dell'aria esterna fresca viene immessa direttamente nella sala dati, previa filtrazione, quando la sua temperatura lo consente, espellendo all'esterno l'aria calda di estrazione. I climatizzatori si spengono e grandi serrande convogliano l'aria esterna attraverso i filtri verso il plenum o la sala, oppure un sistema a ventilazione controllata mantiene la sala pressurizzata con aria fresca. Questo metodo è molto efficace in climi temperati o freddi: in questi casi, l'air economizer può coprire il carico nella maggior parte delle ore dell'anno, dato che, come visto, l'IT può tollerare fino a circa 27 °C in ingresso, o anche più alte. Di notte e in inverno quasi tutti i Data Center potrebbero essere raffreddati con sola aria esterna, purché l'umidità e la qualità dell'aria siano sotto controllo. Una

preoccupazione comune è infatti la contaminazione: polveri, salsedine o gas inquinanti nell'aria esterna potrebbero danneggiare i circuiti. In realtà, con filtri adeguati il rischio è molto basso. È opportuno implementare controlli per chiudere l'ingresso di aria esterna se l'aria esterna è troppo umida o troppo secca, usando sistemi di “lockout” basati sulla temperatura di rugiada. In ambienti industriali estremi, quali acciaierie, impianti chimici, ecc., il rischio di contaminanti va valutato caso per caso, ma per la gran parte dei siti un economizzatore è fortemente raccomandato, e va pianificato preferibilmente in fase di progettazione, poiché l'intervento richiede grossi condotti e aperture; inserirlo in un Data Center esistente può essere quindi complesso. Quando applicabile, questo sistema garantisce risparmi enormi: alcune strutture in climi freddi funzionano senza chiller per tutto l'anno, usando solo economizzatori ad aria e/o ad acqua.

- *economizzatore ad acqua (water-side economizer)*: utilizza l'aria esterna indirettamente, tramite le torri evaporative o i dry coolers. In pratica, quando fa abbastanza freddo, l'acqua delle torri di raffreddamento può raffreddare l'acqua del circuito dei chiller attraverso uno scambiatore a piastre, consentendo di climatizzare la sala senza accendere i compressori frigoriferi. Spesso, l'economizzatore ad acqua è implementato inserendo uno scambiatore a valle delle pompe, prima dei chiller: in tal modo, quando la torre può soddisfare il setpoint, ad es.: acqua a 15 °C, i chiller vengono bypassati. Questa soluzione è utile se l'uso di aria esterna diretta è sconsigliato per problemi di polveri, sicurezza o layout dell'edificio. Di norma, è efficace in climi con temperatura di bulbo umido inferiore a circa 12–13 °C per un buon numero di ore all'anno. Molti Data Center in climi freddi utilizzano solo water-side economizer, senza chiller tradizionali. Una variante sono i *sistemi adiabatici/evaporativi diretti*: l'aria esterna viene raffreddata facendola passare su pacchi bagnati evaporando acqua e poi introdotta; è una soluzione tipica di grossi Cloud Data Center, in climi secchi, nei quali è possibile usare il fenomeno spontaneo dell'evaporazione dell'acqua in seno a una corrente d'aria non satura per raffreddare quest'ultima e poi immetterla nella struttura. Si distinguono due tipologie:
 - sistemi diretti (DEC), nei quali l'acqua da far evaporare viene direttamente spruzzata in seno alla corrente d'aria da trattare, raffreddandola e al contempo umidificandola, prima della sua immissione nel Data Center;
 - sistemi indiretti (IEC), nei quali invece l'acqua evapora in un flusso di aria secondaria che successivamente, attraverso un normale scambiatore di calore, raffredda quella primaria, che sarà poi immessa nel Data Center.

Il sistema diretto è più semplice ed economico rispetto a quello indiretto, ma comporta l'umidificazione dell'aria immessa nel Data Center ed è adatto a climi particolarmente secchi, mentre il secondo può essere utilizzato efficacemente in una più ampia varietà di condizioni climatiche, ma, essendo più costoso, è generalmente usato in DC di grandi dimensioni.

Sono ovviamente possibili, e largamente utilizzate, anche soluzioni ibride, più costose ma anche più flessibili ed efficienti, nonché in grado di adattarsi alle specifiche esigenze di qualsiasi Data Center, e nelle quali viene generalmente realizzato prima un EC diretto e poi quello indiretto.

Implementare il free cooling richiede di progettare correttamente i setpoint: più alta è la temperatura di mandata e di ritorno, maggiore è il numero di ore in cui si può usare questo regime di funzionamento. Ad esempio, se si può accettare acqua refrigerata a 20 °C invece che 7 °C, le torri evaporative potranno fornire

raffreddamento per gran parte dell'anno in molti climi. Per questo motivo, pratiche come l'innalzamento dei setpoint vanno di pari passo col free cooling. In sostanza, corridoi separati, combinati con setpoint elevati, garantiscono il massimo vantaggio.

Infine, esistono soluzioni complementari come il raffreddamento di emergenza senza compressori: con corridoi caldi chiusi, se la temperatura esterna è sotto il limite di sicurezza, ad es. < 35 °C, un Data Center può operare per un certo tempo in sola ventilazione forzata free cooling anche in caso di guasto ai refrigeratori, aumentando la resilienza.

Si possono inoltre installare sistemi di accumulo di acqua refrigerata o ghiaccio che immagazzinano "freddo" nelle ore notturne fredde per usarlo nei picchi diurni, riducendo i carichi di punta e proteggendo dai guasti. Questa è una tecnica avanzata, utilizzata talvolta in grandi impianti, con obiettivi di peak shaving del carico elettrico.

In conclusione, l'utilizzo del free cooling, dove possibile, va attualmente considerato come una best practice primaria.

2.1.4 Altre tecnologie

Oltre alle soluzioni "tradizionali" precedentemente descritte, negli ultimi anni si sono affermate tecnologie di raffreddamento innovative, anche per far fronte all'aumento delle densità di calcolo e agli obiettivi via via più stringenti di sostenibilità. Queste tecnologie includono metodi di raffreddamento a liquido, sistemi ibridi aria/liquido, il recupero del calore dissipato e l'uso di controlli intelligenti per ottimizzare dinamicamente la climatizzazione del Data Center [6].

2.1.4.1 Raffreddamento a liquido

Le tecnologie di raffreddamento a liquido possono essere suddivise in due macrocategorie: raffreddamento a liquido diretto e raffreddamento a immersione.

Raffreddamento a liquido diretto (direct-liquid cooling)

Si distinguono due tipologie:

- a) Scambiatore di calore aria-liquido localizzato (rear-door heat exchangers): prevede l'uso di uno scambiatore di calore aria-liquido (coil) posizionato sul retro di un rack di server che cattura il calore generato dagli stessi, garantendo così un miglioramento dell'efficienza energetica grazie al trasferimento del calore a un liquido più vicino alla sorgente (anziché ai sistemi di trattamento aria della sala computer o alle unità di trattamento aria). Le varianti includono: scambiatore di calore posteriore a porta (rear-door heat exchanger, anche detto cooling door), armadio chiuso (enclosed cabinet), raffreddatore in-row (in-row cooler) e raffreddamento dall'alto (overhead cooling), tutti basati su uno scambiatore aria-liquido accoppiato in prossimità della sorgente IT. Le opzioni per le porte raffreddanti comprendono sistemi attivi (active system), dove ventole aggiuntive aspirano l'aria attraverso le serpentine, oppure sistemi passivi (passive system), dove sono le ventole dei server a spingere l'aria attraverso le serpentine.
- b) Piastra fredda (cold plate): i dissipatori standard a lamelle montati sui chip vengono sostituiti da piastre fredde attraversate da canali in cui scorre un liquido per rimuovere il calore (la soluzione

generalmente prevede l'impiego di un'unità di distribuzione del refrigerante – Coolant Distribution Unit, CDU). Esiste un'ampia varietà di soluzioni tecnologiche, e questo approccio può essere applicato anche alla memoria e ad altri componenti che generano calore.

I vantaggi energetici del liquid cooling sono duplici: l'acqua liquida ha una capacità di trasporto del calore circa 3500 volte superiore a quella dell'aria, e il raffreddamento mirato consente di utilizzare acqua refrigerata a temperature più elevate (20-30°C). Ciò si traduce in un COP notevolmente migliorato per i chiller e apre la strada all'uso del "free cooling a secco" per gran parte dell'anno, senza consumo di acqua.

Alcuni studi indicano che il PUE può essere ridotto del 10-20% passando da sistemi ad aria a raffreddamento a liquido con setpoint elevati [6].

La densità di potenza per rack è aumentata del 38% tra il 2022 e il 2024, con picchi che raggiungono 80-120 kW e più nei cluster di AI. Questo incremento ha reso il raffreddamento a liquido non più una soluzione di nicchia, ma una necessità tecnica per la gestione di carichi di lavoro moderni come quelli dell'AI.

Le linee guida ASHRAE prevedono che sopra una certa potenza per socket (ad es., 400 W per CPU/GPU), il passaggio al liquido diventi la scelta preferenziale [5]. Grandi operatori del settore, stanno adottando il direct-to-chip per i loro cluster di AI, e il mercato del liquid cooling è previsto in forte crescita nei prossimi anni [11, 12].

Raffreddamento a immersione

In questo caso, l'elettronica è immersa in un fluido dielettrico (non conduttivo); gli approcci includono:

- a) raffreddamento a immersione monofase (single-phase immersion cooling), in cui il carico termico viene trasferito a un olio/fluido dielettrico, che viene pompato attorno ai componenti elettronici tramite una CDU (Coolant Distribution Unit, ovvero unità di distribuzione del fluido refrigerante);
- b) raffreddamento a immersione bifase (two-phase immersion cooling), in cui un fluido dielettrico appositamente progettato, con un punto di ebollizione inferiore alla temperatura massima di esercizio dei componenti IT, rimuove il calore passando dallo stato liquido a quello di aeriforme; il vapore trasferisce poi il calore a uno scambiatore di calore vapore-liquido (vapor-to-liquid heat exchanger), che lo condensa nuovamente.

I benefici del raffreddamento a immersione sono potenzialmente notevoli:

- efficienza molto elevata: si elimina il consumo energetico delle ventole dei server; i valori del PUE (Power Usage Effectiveness, rapporto tra l'energia elettrica complessivamente consumata e quella usata per le sole apparecchiature IT), per i Data Center a immersione, possono scendere fino a 1.05, avvicinandosi quindi al limite ideale di 1;
- densità elevata: si possono raggiungere densità computazionali estreme, fino a 50-100 kW per rack, in spazi ridotti;
- resilienza: i fluidi hanno un'alta capacità termica, permettendo al sistema di assorbire calore per diversi minuti in caso di interruzione del raffreddamento attivo;
- sostenibilità: l'uso di acqua si riduce quasi a zero, e si può abbattere l'impronta ambientale complessiva del Data Center fino al 60%.

Sebbene l'adozione sia ancora limitata da sfide come i costi iniziali e la gestione operativa, l'immersione è vista come la "nuova frontiera" per applicazioni HPC e AI.

2.1.4.2 Recupero del calore dissipato (Heat Reuse)

Una delle tendenze più importanti in ottica di sostenibilità è il recupero del calore prodotto dai Data Center, che viene trasformato da scarto in una risorsa. Le applicazioni del calore di scarto includono:

- teleriscaldamento urbano: fornire calore a reti che servono abitazioni e uffici, un'applicazione sempre più comune in Europa settentrionale;
- riscaldamento di edifici adiacenti;
- impiego in processi industriali e agricoli (ad esempio, per il riscaldamento di serre).

Le metriche ERE (Energy Reuse Effectiveness) e ERF (Energy Reuse Factor), che verranno meglio definite in seguito, sono state introdotte proprio per misurare l'efficacia del recupero.

L'adozione di queste pratiche, quando possibile, può considerarsi un punto di svolta, trasformando il calore da un reflujo da smaltire in un'opportunità di creare valore. Questo cambiamento è stato riconosciuto in Europa attraverso recenti normative [15, 16]: la Direttiva sull'Efficienza Energetica (EED) dell'UE del 2023 richiede ai Data Center con potenza IT superiore a 500 kW di rendicontare dati operativi, inclusi il recupero di calore. La progettazione dei nuovi Data Center dovrebbe quindi essere basata anche su questo nuovo paradigma, che vede queste strutture non solo come forti consumatori di energia, ma anche come potenziali fornitori di calore utile.

2.1.4.3 Controllo intelligente dei setpoint e AI per HVAC

L'ultima frontiera dell'innovazione nel raffreddamento riguarda l'uso di controlli avanzati e intelligenza artificiale per ottimizzare la gestione termica in tempo reale [1, 13]. A differenza dei controlli tradizionali a logica fissa, l'AI e il machine learning (ML) possono agire in modo dinamico, usando dati provenienti da migliaia di sensori per prevedere le esigenze di raffreddamento e regolare i setpoint dei chiller, le velocità delle ventole e delle pompe, per minimizzare il consumo totale.

L'AI rappresenta un "livello successivo" di efficienza, che amplifica i benefici delle buone pratiche fisiche. Tuttavia, la sua adozione è prevalentemente limitata ai grandi Data Center hyperscale e cloud, che dispongono dell'abbondanza di dati e delle risorse computazionali necessarie per implementare tali sistemi.

Per i Data Center più piccoli, si possono ottenere benefici già mediante controlli automatizzati più semplici o soluzioni DCIM (Data Center Infrastructure Management), con moduli di analisi meno complessi.

2.2 EFFICIENZA ENERGETICA: ASPETTI RELATIVI ALLA COMPONENTE INFORMATICA

Per ottenere una gestione realmente efficiente e sostenibile del Data Center, è indispensabile disporre di un sistema di monitoraggio integrato capace di unire in un'unica piattaforma le metriche provenienti sia dall'infrastruttura elettrica che dalle componenti informatiche. L'efficienza energetica di un data center non

dipende unicamente dall'infrastruttura elettrica e di raffreddamento, ma anche dall'ottimizzazione dell'intero ecosistema IT. Le strategie di intervento sulla componente informatica mirano a ridurre i consumi energetici, migliorare l'utilizzo delle risorse e garantire la flessibilità operativa del sistema. Tali opportunità – alcune specifiche per l'intelligenza artificiale (AI) e altre applicabili più in generale ai Data Center e al calcolo ad alte prestazioni – possono essere generalmente classificate in tre categorie [2]: hardware, software e approcci trasversali.

Hardware

- Processori a basso consumo energetico: processori progettati per minimizzare il consumo di energia, ad esempio CPU basate su architettura ARM o Intel Atom;
- Acceleratori per AI: hardware specializzato in grado di eseguire compiti di AI in modo rapido ed efficiente, ad esempio GPU Nvidia, TPU Google, o dispositivi come NPU e Apple Neural Engine;
- Processori ibridi ottimizzati per attività specifiche: processori che combinano unità di elaborazione specializzate ("chiplet") per determinate funzioni all'interno di un unico package, al fine di massimizzare prestazioni ed efficienza energetica (es. AMD Epyc CPUs);
- Circuiti fotonici integrati: tecnologie che utilizzano la luce (fotoni) invece dell'elettricità (elettroni) per l'elaborazione delle informazioni, riducendo la dispersione energetica e consentendo una gestione dei dati più rapida ed efficiente;
- Memorie e sistemi di storage ad alta efficienza energetica: utilizzo di tecnologie di memoria e archiviazione che riducono il consumo di energia, come memorie DDR5 a basso consumo e unità SSD NVMe;
- Prossimità della memoria (memory proximity): posizionamento dei dati più vicino al processore per ridurre le distanze di trasferimento e il consumo energetico, ad esempio tramite memorie ad alta larghezza di banda (HBM) integrate con GPU;
- Tecnologie di raffreddamento innovative: metodi avanzati per rimuovere il calore dai Data Center in modo più efficiente, riducendo l'energia necessaria per il raffreddamento, come i sistemi di raffreddamento a liquido (direct-to-chip, immersion cooling);

Software

- Algoritmi ad alta efficienza energetica: sviluppo di algoritmi di intelligenza artificiale che richiedono un minor consumo energetico;
- Modelli specifici per attività (task-specific models): modelli di AI più piccoli e specializzati, progettati per compiti specifici invece che per funzioni generiche di ampia portata;
- Ottimizzazione di modelli e codice: miglioramento delle architetture dei modelli esistenti, del codice e del software al fine di ridurre l'utilizzo di risorse computazionali e il consumo di energia;

Approcci trasversali

- Co-progettazione hardware/software (co-design): progettazione congiunta di hardware e software per sfruttare le sinergie e massimizzare efficienza energetica e prestazioni;
- Edge computing: esecuzione dei processi di inferenza AI più vicino agli utenti finali (dispositivi o server edge), riducendo la trasmissione dei dati e distribuendo il carico energetico su più dispositivi efficienti e distribuiti;

- Virtualizzazione: esecuzione di più macchine virtuali su un singolo server fisico per aumentare il tasso di utilizzo e ridurre il numero di server fisici necessari;
- Gestione intelligente dell'energia: sistemi intelligenti di gestione della potenza e dei carichi di lavoro, ad esempio assegnando le attività a hardware più efficienti dal punto di vista energetico, oppure utilizzando l'AI per monitorare e ottimizzare il raffreddamento e la distribuzione delle risorse computazionali a livello di Data Center;
- Calcolo quantistico (quantum computing): paradigma di calcolo basato sulla meccanica quantistica, in grado di eseguire operazioni estremamente complesse rispetto al calcolo classico,
- Calcolo neuromorfico (neuromorphic computing): approccio che imita l'architettura neuronale del cervello per elaborare dati e calcoli in modo più efficiente rispetto ai sistemi classici.

La Tabella 2 riassume i livelli attuali di adozione di tali tecnologie, il probabile grado di diffusione entro il 2030 e la scala potenziale di risparmio energetico associata [2]. Tali stime non tengono conto degli effetti di rimbalzo (rebound effects) che potrebbero controbilanciare i miglioramenti ottenuti in termini di efficienza energetica.

Tecnologia/Approccio	Adozione attuale	Adozione prevista nel 2030	Potenziale di risparmio energetico
<i>Hardware</i>			
Processori a bassa potenza	●●	●●●	●●●●
Acceleratori AI	●●●	●●●●	●●
Processori ibridi ottimizzati per compiti	●●	●●●	●●
Circuiti integrati fotonici	●	●●	●●●
Memoria e archiviazione ad alta efficienza energetica	●●●	●●●●	●●
Prossimità della memoria	●●	●●●	●●
Tecnologie innovative di raffreddamento	●●	●●●●	●●
<i>Software</i>			
Algoritmi ad alta efficienza energetica	●●	●●●●	●●●●
Modelli specifici per compito	●●	●●●●	●●●●
Ottimizzazione di modelli e codice	●●	●●●	●●●
<i>Trasversale</i>			
Codesign software/hardware	●●	●●●	●●
Edge computing	●●	●●●	●●●
Virtualizzazione	●●●●	●●●●	●●
Gestione intelligente dell'energia	●●●	●●●●	●●
Calcolo quantistico	●	●	●●●
Calcolo neuromorfico	●	●●	●●●●

Tabella 2 - Attuali e potenziali risparmi energetici nei Data Center tratta da [2].

Nota: Un numero maggiore di punti indica una scala più alta

Di seguito sono riportati i principali interventi attuabili e le relative modalità di integrazione con l'infrastruttura elettrica. Le strategie di intervento sulla componente informatica mirano a:

- Ridurre i consumi energetici diretti dei sistemi di calcolo, storage e rete;
- Migliorare l'utilizzo delle risorse e ridurre l'hardware sottoutilizzato;
- Aumentare la flessibilità operativa e la capacità di adattamento ai carichi dinamici;
- Ottimizzare l'integrazione con i sistemi elettrici e di controllo (DCIM/BMS);
- Supportare la sostenibilità ambientale attraverso metriche e processi di gestione efficienti.

Di seguito sono riportati i principali interventi attuabili sulla componente IT e le relative modalità di integrazione con l'infrastruttura elettrica e di monitoraggio.

2.2.1 Consolidamento dei server tramite virtualizzazione e riduzione dell'hardware sottoutilizzato

Uno dei principali approcci per incrementare l'efficienza energetica e ottimizzare l'utilizzo delle risorse in un Data Center consiste nel consolidamento dei server fisici mediante tecniche di virtualizzazione. Nelle architetture tradizionali, i server dedicati a singole applicazioni o servizi risultano spesso sottoutilizzati, con tassi medi di utilizzo della CPU inferiori al 20–30%. Questo comporta un notevole spesa energetica, in quanto ogni unità fisica continua a utilizzare energia anche in condizioni di carico minimo.

Attraverso l'adozione di piattaforme di virtualizzazione è possibile eseguire più VM o container su un singolo host fisico, massimizzando il tasso di utilizzo delle risorse computazionali (CPU, memoria, storage e rete). La virtualizzazione consente inoltre una gestione centralizzata e flessibile dell'ambiente IT, permettendo di ridistribuire i carichi di lavoro in base alle esigenze operative e di ottimizzare dinamicamente le prestazioni e i consumi. Inoltre, altre funzionalità dell'hypervisor permettono di spostare una VM tra host (live migration) in funzione del carico e spegnere eventuali server inattivi. Quindi, l'utilizzo della virtualizzazione abilita una diminuzione della potenza assorbita nei quadri di distribuzione elettrica e minore dispersione termica, che si traduce in un carico inferiore sui sistemi HVAC, permettendo inoltre l'incremento dell'efficienza complessiva tramite una ottimizzazione delle risorse di calcolo. La piena valorizzazione di questa strategia si ottiene tramite l'integrazione con i sistemi DCIM e con i moduli di monitoraggio elettrico intelligente che permettono di correlare il consumo energetico di ciascun host fisico con il carico IT effettivo gestito e di identificare server sottoutilizzati o in idle per pianificare ulteriori consolidamenti o lo spegnimento controllato degli stessi.

In Data Center di nuova generazione, il consolidamento dei server è quindi non solo una misura di razionalizzazione infrastrutturale, ma anche un elemento chiave per la transizione verso modelli sostenibili e ad alta efficienza, in cui i domini elettrico e informatico operano in modo integrato e sinergico.

2.2.2 Allocazione dinamica delle risorse

L'Allocazione dinamica delle risorse rappresenta una delle strategie più efficaci per l'ottimizzazione energetica dei dati moderni. L'obiettivo di tale tecnica è di adattare dinamicamente la capacità computazionale alle reali esigenze operative, modulando l'impiego delle risorse in funzione del carico di lavoro nell'intervallo di tempo, consentendo di garantire un equilibrio ottimale tra prestazioni, disponibilità e consumo energetico. L'implementazione di tali tecniche si basa su tecnologie di orchestrazione e

automazione avanzate che monitorano costantemente i parametri di utilizzo (CPU, memoria, I/O) e sono in grado di migrare automaticamente i carichi di lavoro virtualizzati (macchine virtuali o container) tra i nodi del cluster, in modo da garantire una distribuzione ottimale delle risorse. Quando il carico complessivo del sistema diminuisce, la piattaforma può consolidare più workload su un numero ridotto di host attivi oppure porre i server non necessari in stato di sospensione, ibernazione o spegnimento controllato. Al contrario, durante i picchi di domanda, i nodi vengono attivati automaticamente o scalati orizzontalmente per mantenere il livello di servizio desiderato (Quality of Service – QoS) senza interruzioni.

Questa strategia consente una riduzione significativa del consumo medio di potenza, poiché il sistema IT non opera più a capacità massima in modo costante, ma in funzione del fabbisogno reale. Essa trasforma il modello di erogazione delle risorse da statico a intelligente e adattivo, consentendo una perfetta integrazione tra infrastruttura elettrica e logica IT, in un’ottica di efficienza operativa, resilienza e sostenibilità ambientale.

2.2.3 Scheduling energetico e ottimizzazione predittiva

Oltre alla gestione dinamica dei carichi, un ulteriore passo verso l’efficienza complessiva del Data Center è rappresentato dall’impiego di algoritmi di scheduling energetico e ottimizzazione predittiva. Tali algoritmi permettono di prevedere e gestire in modo intelligente i periodi di inattività dei server, adeguando in modo proattivo la disponibilità delle risorse alle effettive necessità operative. L’obiettivo principale è ridurre al minimo i consumi energetici derivanti dal funzionamento in stato di idle, ossia quando i server sono accesi ma non impiegano in modo significativo la propria capacità computazionale.

Gli algoritmi di scheduling energetico vengono tipicamente integrati nei sistemi di orchestrazione dei cluster virtualizzati o nei framework di gestione distribuita. Dal punto di vista energetico, l’adozione di tali algoritmi consente una riduzione significativa del consumo di potenza a riposo (idle power).

Gli algoritmi di scheduling contribuiscono non solo ai benefici energetici, ma anche ad estendere la vita utile delle apparecchiature, riducendo i cicli termici e meccanici sui componenti hardware oltre ad ottimizzare la capacità installata, evitando la sovrappopolazione di server attivi in condizioni di basso carico. Le soluzioni più avanzate implementano logiche di scheduling predittivo e autonomo, in cui l’intelligenza artificiale analizza grandi volumi di dati storici (telemetrie, log, sensori ambientali) per anticipare i picchi e i cali di domanda. In questo contesto, l’infrastruttura IT diventa un sistema auto-adattativo, capace sia di prevedere variazioni di carico in base a eventi pianificati (ad es. backup notturni, aggiornamenti, carichi batch) sia di calibrare l’erogazione di potenza in modo proporzionale al fabbisogno reale. L’impiego di algoritmi di scheduling energetico per pianificare accensione, ibernazione o spegnimento dei server in base a pattern di utilizzo e previsioni di carico. Inoltre, l’integrazione di logiche di machine learning per prevedere variazioni di domanda e adeguare preventivamente la capacità IT. Misurazione continua del consumo a riposo (idle power) e riduzione del fabbisogno di energia non produttiva.

2.2.4 Ottimizzazione dei flussi di rete e data locality

Un ruolo determinante nel consumo complessivo di un Data Center è svolto dalla componente di rete, che comprende switch, router, apparati di storage, connessioni backbone e infrastrutture di interconnessione intra e inter-rack. L’obiettivo dell’ottimizzazione è ridurre il consumo di banda, minimizzando il numero di operazioni di switching e routing necessarie per la trasmissione dei dati. Ciò consente di diminuire non solo i

consumi diretti degli apparati di rete, ma anche l'energia dissipata in calore e quindi il carico termico da smaltire nei sistemi di raffreddamento.

Differenti strategie possono essere adottate tra cui la segmentazione logica del traffico tramite VLAN che consente di isolare i flussi di dati in base a criteri funzionali o applicativi evitando instradamenti superflui e congestioni. Le reti SDN permettono, inoltre, di gestire dinamicamente i percorsi dei pacchetti, scegliendo le rotte più efficienti in termini energetici, ad esempio evitando l'attivazione di link o dispositivi non necessari. Il controller di una rete SDN può monitorare il traffico in tempo reale e adattare la topologia logica della rete alle condizioni operative, riducendo l'uso di risorse nei periodi di basso carico.

Un'altra strategia fondamentale consiste nel posizionare i processi di elaborazione il più vicino possibile ai dati che devono essere trattati (data locality), riducendo così la quantità di traffico che attraversa la rete interna. Questo approccio, strettamente legato alle architetture edge computing e software-defined storage, consente di minimizzare la latenza e le operazioni di trasmissione, riducendo sensibilmente l'energia richiesta dai sistemi di networking e di storage distribuito.

Altre tecniche come EtherChannel, Link Aggregation Control Protocol (LACP) e protocolli di power-aware routing consentono di aggregare più interfacce di rete per aumentarne la capacità complessiva, mantenendo attive solo le connessioni effettivamente necessarie in base al traffico.

Nei periodi di bassa attività, i canali ridondanti possono essere disattivati automaticamente o posti in modalità a basso consumo, riducendo così la potenza assorbita dagli switch e dalle Network Interface Cards (NIC). Allo stesso modo, algoritmi di routing adattivo selezionano i percorsi che minimizzano il numero di hop o di dispositivi intermedi attivi, bilanciando efficienza e prestazioni.

Le tecniche di Network Function Virtualization (NFV) consentono la disaggregazione delle funzioni di rete dall'hardware sottostante, permettendo la loro esecuzione su server general-purpose virtualizzati. Questa flessibilità apre la strada a strategie di gestione dinamica delle risorse, orientate all'efficienza energetica. L'integrazione di modelli di machine learning consente di prevedere i pattern di traffico e di ottimizzare l'attivazione, la migrazione e lo spegnimento delle istanze virtuali, riducendo il consumo energetico complessivo senza compromettere la qualità del servizio (QoS).

2.2.5 Ottimizzazione e virtualizzazione dello storage

L'ottimizzazione dei sistemi di archiviazione rappresenta un elemento chiave per migliorare l'efficienza complessiva delle infrastrutture IT e ridurre il consumo energetico. Il consolidamento dello storage attraverso architetture centralizzate, come le Storage Area Network (SAN), e soluzioni Software Defined Storage (SDS) consente di gestire in modo unificato risorse eterogenee, ottimizzando la capacità e le prestazioni in base ai carichi di lavoro.

Un approccio ibrido, che combina dischi rigidi tradizionali per la conservazione di dati statici o a bassa priorità con unità a stato solido per applicazioni ad alte prestazioni, permette di bilanciare costo, velocità di accesso e consumo energetico.

Il tiering dello storage è una tecnica di gestione automatica e gerarchica dei dati che consente di allocare le informazioni su diversi livelli (tier) di archiviazione, ciascuno caratterizzato da prestazioni, capacità e costi differenti. In un sistema di storage tiered, i dati vengono classificati in base alla loro frequenza di accesso e importanza operativa: i dati “caldi” (hot data) sono utilizzati frequentemente e vengono collocati su supporti ad alte prestazioni e bassa latenza, come SSD o NVMe; i dati “tiepidi” (warm data) sono soggetti ad accessi periodici e vengono conservati su supporti intermedi; i dati “freddi” (cold data) sono raramente consultati e vengono spostati su HDD o su sistemi di archiviazione a basso costo e consumo energetico, come storage object o cloud archive.

Il processo di tiering può essere gestito manualmente dall'amministratore di sistema o, più comunemente, in modo automatico e dinamico da software di gestione dello storage (Storage Management Software o Software Defined Storage), che monitorano i pattern di accesso ai dati e ne ottimizzano la collocazione in tempo reale. La gestione automatica del tiering dei dati consente di spostare dinamicamente le informazioni tra i diversi livelli di storage in funzione della loro frequenza di utilizzo (hot, warm e cold data), riducendo gli sprechi energetici e ottimizzando le prestazioni complessive. Dal punto di vista energetico, il tiering consente di ridurre i consumi complessivi, poiché mantiene attivi i dispositivi ad alte prestazioni solo per i dati effettivamente necessari, lasciando in stato di basso consumo o inattività i livelli di storage meno utilizzati. Ciò si traduce in una maggiore efficienza energetica, un minor calore generato e una riduzione dei costi di raffreddamento e manutenzione.

Altre due tecniche sono spesso adottate nell'ambito dell'efficientamento energetico dello storage: la compressione dei dati e le tecniche di deduplicazione. La compressione dei dati è un processo che riduce la dimensione complessiva dei file mediante algoritmi che eliminano la ridondanza interna dei dati, rappresentandoli in modo più efficiente senza perdita di informazioni (compressione lossless) o con perdita controllata (compressione lossy, meno comune nello storage enterprise). Questo consente di ottimizzare la capacità disponibile e di migliorare le prestazioni di trasferimento dei dati, in quanto vengono letti e scritti meno byte. La deduplicazione dei dati è una tecnica di ottimizzazione dello storage che elimina le copie ridondanti degli stessi dati, conservando una sola istanza univoca e sostituendo le duplicazioni con riferimenti o puntatori. In questo modo si riduce notevolmente lo spazio fisico necessario per l'archiviazione, soprattutto in ambienti dove più sistemi o utenti memorizzano file simili o identici (ad esempio backup o immagini di macchine virtuali).

Le tecniche di compressione e deduplicazione dei dati contribuiscono in modo significativo all'efficienza complessiva dei sistemi di archiviazione, riducendo la capacità fisica necessaria e il numero di dischi attivi. Questo si traduce in un sensibile abbattimento dei consumi elettrici e dei costi di raffreddamento, migliorando al contempo la sostenibilità ambientale dell'infrastruttura. Inoltre, tali tecniche permettono di ottimizzare i flussi di I/O e di aumentare la densità dei dati gestiti per unità di storage, consentendo una gestione più scalabile e performante. In ambienti virtualizzati e distribuiti, l'applicazione combinata di compressione e deduplicazione riduce anche il traffico di rete durante le operazioni di replica o backup, contribuendo a una maggiore efficienza energetica e operativa dell'intero ecosistema di storage.

La virtualizzazione dello storage introduce un livello di astrazione che consente la gestione distribuita, dinamica e intelligente delle risorse di archiviazione tra nodi differenti. Questo approccio elimina la rigidità

dei sistemi di storage tradizionali, permettendo una allocazione flessibile della capacità e delle prestazioni in base alle reali esigenze operative. Dal punto di vista del risparmio energetico, tali soluzioni offrono un vantaggio significativo poiché consentono di ottimizzare il carico di lavoro tra i nodi, riducendo l'attività dei dispositivi meno utilizzati o attivandoli solo quando necessario. Grazie a meccanismi di bilanciamento automatico del carico (load balancing) e migrazione dinamica dei dati, il sistema può distribuire in modo efficiente le operazioni di I/O, evitando sovraccarichi localizzati e migliorando il rapporto tra prestazioni e consumo energetico.

2.2.6 Gestione ottimizzata dei consumi

L'uso dell'AI per il controllo dei sistemi di raffreddamento amplifica i benefici delle buone pratiche fisiche, tuttavia, la sua adozione risulta più vantaggiosa per i grandi Data Center hyperscale e cloud, che dispongono dell'abbondanza di dati e delle risorse computazionali necessarie per implementare tali sistemi. Per i Data Center più piccoli, si possono ottenere benefici già mediante controlli automatizzati più semplici o soluzioni Data Center Infrastructure Management (DCIM), con moduli di analisi meno complessi per l'ottimizzazione della distribuzione energetica attraverso il monitoraggio e la gestione delle risorse [19].

Nei Data Center di nuova generazione composti da rack ad alta densità dove i carichi termici sono più concentrati, l'integrazione di tecniche di raffreddamento avanzate come il liquid cooling con sistemi di controllo dinamico basati su AI rappresenta una delle soluzioni più promettenti per migliorare l'efficienza energetica e la resilienza operativa. Questi sistemi sfruttano reti di sensori, moduli predittivi e algoritmi di controllo adattivo per analizzare in tempo reale le condizioni termiche e regolare dinamicamente il flusso di liquido, la velocità delle pompe e la distribuzione del carico termico [17]. Questi sistemi operano su tre layer funzionali: l'acquisizione dei dati dei sensori, il controllo e l'ottimizzazione AI-based e l'attuazione fisica sui dispositivi di raffreddamento. In tale architettura, i modelli di Machine Learning (ML) consentono di prevedere variazioni di temperatura, ottimizzare i parametri operativi e migliorare il coefficiente di efficienza energetica (PUE), massimizzando il recupero del calore e la stabilità termica del sistema riducendo significativamente il consumo complessivo di energia destinato al raffreddamento.

Questi algoritmi utilizzano deep neural networks addestrate sui dati storici provenienti da migliaia di sensori (e.g., temperatura, potenza assorbita, velocità delle pompe e parametri di setpoint) per prevedere l'andamento del PUE e ottimizzare le impostazioni operative in tempo reale. Il sistema di controllo risultante è in grado di adattarsi dinamicamente alle condizioni interne ed esterne, riducendo l'energia richiesta per il raffreddamento e migliorando la stabilità termica complessiva.

Parallelamente, la letteratura più recente evidenzia come l'integrazione dei Digital Twin con algoritmi di ML consenta di replicare digitalmente la struttura fisica e il comportamento termico dei sistemi di raffreddamento dei Data Center. Questa rappresentazione virtuale permette di prevedere la formazione di hotspot e di regolare in tempo reale i parametri termici, adattando le strategie di controllo alle condizioni operative effettive. Tali approcci favoriscono un'ottimizzazione energetica dinamica e una gestione predittiva dell'infrastruttura, migliorando l'efficienza complessiva e riducendo i rischi di sovraccarico termico [21].

Infine, diversi studi recenti hanno esplorato l'applicazione del Reinforcement Learning (RL) al controllo dei sistemi di raffreddamento, mostrando notevoli guadagni in efficienza energetica e maggiore adattabilità rispetto ai controllori tradizionali [22, 23]. Gli agenti RL apprendono policy ottimali interagendo con

l'ambiente fisico o con simulazioni termiche, regolando parametri come la velocità dei ventilatori, il flusso di liquido o la distribuzione del carico termico in funzione delle variazioni ambientali. Questi modelli abilitano un controllo autonomo e proattivo, riducendo il PUE e migliorando la stabilità operativa. Tuttavia, richiedono dataset estesi, infrastrutture di calcolo per l'addestramento e un'integrazione stretta con i sistemi fisici di controllo, condizioni più facilmente soddisfatte da operatori hyperscale e cloud dotati delle risorse necessarie per supportare tali architetture intelligenti.

3 INDICI ENERGETICI

Il raffreddamento rappresenta una quota significativa del consumo energetico totale. Migliorare l'efficienza energetica dei sistemi di raffreddamento è dunque decisivo per ridurre la potenza assorbita dalle apparecchiature IT, senza comprometterne affidabilità e disponibilità. Un concetto chiave è separare l'efficienza dell'infrastruttura di raffreddamento da quella dell'IT: metriche dedicate permettono di quantificare le prestazioni energetiche sia globali che specifiche del raffreddamento. Data la complessità del tema, nei prossimi paragrafi vengono elencati gli indicatori suddividendoli in tre ambiti:

- Indicatori infrastrutturali: che rappresentano sostanzialmente tutto quanto non è direttamente dedicato all'IT, e più dettagliatamente descritti in Appendice;
- Indicatori ICT: che rappresentano le caratteristiche elaborative del CED;
- Indicatori di obiettivo: relativi a caratteristiche infrastrutturali, IT e di tipo energetico, che devono essere perseguiti in un adeguato arco temporale (massimo 3 anni).

Power Usage Effectiveness (PUE)

Il PUE (Power Usage Effectiveness) è il rapporto tra l'energia totale consumata dal Data Center IT (compresa di impianti di supporto) e l'energia consumata solo dalle apparecchiature IT. Misura l'efficienza complessiva dell'infrastruttura: un PUE pari a 1 indica che tutta l'energia alimenta l'IT: è ovviamente una situazione ideale; comunemente, il PUE si aggira intorno a 1.5, ovvero c'è un 50% di "overhead" (consumo non associato direttamente all'alimentazione delle apparecchiature IT). Recentemente, i Data Center di grandi dimensioni e tecnologicamente avanzati hanno raggiunto PUE vicini a 1.1. Il PUE, introdotto originariamente dall'organizzazione internazionale The Green Grid, è ora standard ISO/IEC 30134-2, ed è tra i più utilizzati. Tuttavia, quest'indice non misura l'efficienza dei server stessi, ma solo quella degli impianti di supporto (raffreddamento, UPS, etc.) [6, 7].

Inverso del PUE

Il reciproco del PUE esprime in percentuale l'efficienza complessiva delle infrastrutture ed è definito come il rapporto tra consumo energetico delle sole apparecchiature IT (server, storage, networking) e consumo energetico complessivo del Data Center. Quando tale valore raggiunge il 100% si ottiene la massima efficienza infrastrutturale sebbene valori tipici siano compresi tra il 50% e il 67%.

IT Equipment Average Efficiency Index (η_{IT})

η_{IT} (IT Equipment Average Efficiency Index) è definito come il rapporto tra il workload normalizzato WL_N , che rappresenta la capacità media effettivamente utilizzata delle risorse di elaborazione ($WL_{processing}$), storage ($WL_{storage}$) e rete ($WL_{networking}$) in un determinato periodo di monitoraggio Q , normalizzata rispetto a fattori rappresentativi della dimensione, capacità e complessità del Data Center, e il consumo di energia elettrica delle sole apparecchiature IT (server, storage e networking) nel medesimo periodo, espresso in kWh e calcolato su base oraria (E_{IT}).

Altri indicatori

Altri indicatori, quali il CUE (Carbon Usage Effectiveness) [5,20], WUE (Water Usage Effectiveness) [9], RCI (Rack Cooling Index) [19] e ERF (Energy Reuse Factor) [19]) sono reperibili in letteratura.

4 ALGORITMI PER IL CALCOLO DEL RISPARMIO ENERGETICO

Si prendono in considerazione interventi di efficientamento energetico relativi sia all'infrastruttura (cooling, impiantistica elettrica) che alle apparecchiature IT, che consentano una riduzione del rapporto tra consumo di energia e servizio erogato dal DC, sia per i casi di sostituzione e di efficientamento integrato di un Data Center ma anche per i casi di nuova installazione.

In questa guida non vengono considerati interventi che producono effetti esterni al perimetro del DC, quali, ad esempio, quelli basati sul recupero dell'energia termica, asportata per il raffreddamento del DC, per l'alimentazione di una rete di teleriscaldamento o di altre utenze esterne al DC. Interventi di questa tipologia possono essere valutati sulla base di programmi di misura e di algoritmi di calcolo del risparmio energetico già descritti in altre Guide Settoriali.

4.1 Indicatore del servizio utile: il Workload funzionale erogato, WLE

Per valutare in modo efficace l'efficienza energetica di un DC, l'analisi non può limitarsi alla sola valutazione delle metriche puramente energetiche. L'efficienza energetica è, per definizione, il rapporto tra i servizi erogati e l'energia consumata (EE= Servizi Erogati/ Energia Consumata).

Le metriche più note, come il PUE (Power Usage Effectiveness), benché fondamentali per monitorare l'efficienza infrastrutturale (overhead), non riescono a cogliere l'efficienza con cui l'hardware IT svolge il lavoro effettivo né il grado di utilizzo degli asset IT installati.

Dati i limiti pratici nel definire e misurare universalmente i servizi complessi forniti all'utente finale (ad esempio, l'elaborazione di intelligenza artificiale o lo streaming video), l'approccio più robusto e pragmatico consiste nell'adottare metriche basate sulle tre funzioni intrinseche dei DC:

- il *processing* (elaborazione dati);
- lo *storage* (archiviazione);
- il *networking* (trasmissione dati).

La metrica funzionale deve quantificare il lavoro erogato dal DC in un periodo di tempo, in relazione all'energia consumata. Poiché le unità di misura delle tre funzioni IT (elaborazione, archiviazione, trasmissione dati) non sono direttamente confrontabili, è necessario introdurre fattori di conversione, per creare un Workload erogato (WL_E) uniforme, che possa rappresentare un indicatore oggettivo e misurabile del carico di lavoro (Workload) erogato dal DC.

Una possibile formula aggregata del Workload erogato, WL_E , basata sul documento “Energy Efficiency Metrics for Data Centers” elaborato nel 2022 dall’Electronic Devices & Networks Annex – EDNA, dell’IEA, è la seguente:

$$WL_E = N_1 \cdot WL_{processing} + N_2 \cdot WL_{storage} + N_3 \cdot WL_{networking}$$

dove:

- $WL_{processing}$, $WL_{storage}$ e $WL_{networking}$ rappresentano il servizio effettivamente svolto dal DC, per ciascuna delle tre principali funzioni, nel periodo di monitoraggio, ovvero la capacità media utilizzata di server, storage e networking, rispettivamente;

- N_1 , N_2 e N_3 sono dei Fattori di Conversione, i cui valori rappresentano il consumo energetico convenzionale per unità di servizio erogato, ovvero la potenza elettrica media assorbita per unità di capacità di processing, storage o networking, come meglio specificato nella seguente tabella;
- WL_E si intende espresso in kWh, nell'intervallo temporale di monitoraggio.

I Fattori di Conversione, in linea di principio, andrebbero definiti, sulla base di dati riferiti alle Best Available Technologies (BAT): tuttavia, al momento, l'assenza di un quadro armonizzato ufficiale e standardizzato per test e reportistica, e quella di dati sufficientemente ampi, rappresentativi e consolidati, non consentono di definire in modo univoco questi coefficienti.

Pertanto, tali Fattori di Conversione sono propri di ogni Data Center e possono essere definiti sulla base delle potenze nominali, riportate nelle rispettive schede tecniche, ovvero sia come:

- N_1 = potenza elettrica nominale del server / capacità nominale del server;
- N_2 = potenza elettrica nominale dello storage / capacità nominale dello storage;
- N_3 = potenza elettrica nominale della rete / capacità nominale della rete.

I fattori di Conversione per il calcolo del Workload Erogato sono riepilogati nella tabella seguente:

Funzione IT	Unità di Misura per WL_i	Fattore di Conversione, N_i
Processing (Server)	GTOPS	Potenza massima assorbita/Potenza Computazionale (W/GTOPS)
Storage (Archiviazione)	Gbyte	Potenza massima assorbita/Capacità di storage (W/Gbyte)
Networking (Dati In/Out)	Gbps	Potenza massima assorbita/Capacità di trasmissione dati (W/Gbps)

4.2 Grandezze e indicatori utilizzati

Si elencano di seguito le principali grandezze utilizzate per il calcolo dei risparmi energetici addizionali:

- Power Usage Effectiveness, $PUE = \frac{E_{TOT}}{E_{IT}}$
- IT Equipment Average Efficiency Index, $\eta_{IT} = \frac{WL_E}{E_{IT}}$
- $E_{non-IT} = E_{TOT} - E_{IT}$

dove:

- E_{TOT} è il consumo complessivo di energia elettrica del Data Center, espresso in [kWh];
- E_{IT} è il consumo di energia elettrica relativo alle sole apparecchiature IT del Data Center, espresso in [kWh];
- E_{non-IT} è l'energia elettrica consumata dall'infrastruttura, al netto di quella assorbita dalle apparecchiature IT, espresso in [kWh].

4.3 Algoritmi per Data Center nel caso di nuova installazione

Per un nuovo Data Center, è necessaria la misura delle seguenti grandezze con una frequenza di campionamento oraria:

- $E_{TOT,h} = E_{IT,h} + E_{non-IT,h}$, è il consumo orario complessivo di energia elettrica del Data Center [kWh];
- $E_{IT,h}$ è il consumo orario di energia elettrica relativo alle sole apparecchiature IT del Data Center [kWh];
- $WL_{processing,h}$, è il servizio orario erogato dal Data Center per la funzionalità di server [GTOPS];
- $WL_{storage,h}$, è il servizio orario erogato dal Data Center per la funzionalità di storage [Gbyte];
- $WL_{networking,h}$, è il servizio orario erogato dal Data Center per la funzionalità di networking [Gbps];

Il soggetto proponente può proporre una frequenza di campionamento diversa purché dimostri che le fluttuazioni del consumo specifico in funzione delle variabili operative non intervengano a livello a orario, fermo restando che qualsiasi ipotesi dovrà essere cautelativa per l'Amministrazione. La medesima frequenza dovrà essere uguale o superiore a quella adottata per la determinazione del consumo di baseline.

Il risparmio energetico nel periodo di monitoraggio dovrà essere calcolato con una frequenza di campionamento oraria, in generale, come somma di due contributi, il primo riferito a interventi sull'infrastruttura IT, il secondo a interventi di altra natura (cooling, impianti elettrici, etc.):

$$RISP = RISP_{IT} + RISP_{non-IT}$$

dove:

$$RISP_{IT} = \sum_{h=1}^{h_{PdM}} \left(\frac{1}{\eta_{IT,RIF,h}} - \frac{1}{\eta_{IT,POST,h}} \right) \cdot WL_{E,h} \cdot f_e$$

$$RISP_{non-IT} = \sum_{h=1}^{h_{PdM}} (PUE_{RIF,h} - PUE_{POST,h}) \cdot E_{IT,h} \cdot f_e$$

dove:

- $WL_{E,h} = N_1 \cdot WL_{processing,h} + N_2 \cdot WL_{storage,h} + N_3 \cdot WL_{networking,h}$
- $E_{non-IT,h} = E_{TOT,h} - E_{IT,h}$
- Power Usage Effectiveness, $PUE_{POST,h} = \frac{E_{TOT,h}}{E_{IT,h}}$
- IT Equipment Average Efficiency Index, $\eta_{IT,POST,h} = \frac{WL_{E,h}}{E_{IT,h}}$
- f_e è il fattore di conversione da energia elettrica, in kWh, a energia primaria, in tep pari a $0,187 \cdot 10^{-3}$ tep/kWh;
- h_{PdM} ore del periodo di monitoraggio.

I valori di riferimento del PUE_{RIF} e di $\eta_{IT,RIF}$ saranno individuati dal Soggetto Proponente sulla base di analisi specifiche di settore rappresentative dello standard di mercato, anche in funzione delle condizioni di reale esercizio del CED, al fine di normalizzarli alle effettive condizioni di utilizzo post intervento (ad es.

$T_{ext,POST,h}; WL_{E,POST,h}$). Per quanto concerne il PUE, possibili dati di riferimento sono riportati nel documento “Assessment of the energy performance and sustainability of data centres in EU” (European Commission, 2025).

Nei casi di nuova installazione, il riconoscimento dei risparmi energetici aggiuntivi è subordinato all'applicazione congiunta degli algoritmi relativi sia alla componente IT sia alla componente non IT. Pertanto, in tali fattispecie non è ammessa l'applicazione parziale dell'algoritmo di calcolo limitata al solo termine $RISP_{IT}$ o al solo termine $RISP_{non-IT}$.

Resta ferma la possibilità di proporre un'applicazione parziale dell'algoritmo relativa alla sola componente non IT, $RISP_{non-IT}$, solo qualora venga dimostrato che i sistemi IT installati, rispettino gli standard internazionali di efficienza (ad es. conformità a Energy Star; etc.), tramite l'ausilio di documentazione tecnica (ad es. schede tecniche; etc.).

4.4 Algoritmi per Data Center nel caso di sostituzione e di efficientamento integrato

Nel caso di sostituzione o efficientamento energetico integrato di un Data Center esistente è necessario fornire i dati misurati nel periodo di monitoraggio di 12 mesi ante intervento delle seguenti grandezze, con una frequenza di campionamento oraria:

- $E_{TOT,h} = E_{IT,h} + E_{non-IT,h}$, è il consumo orario complessivo di energia elettrica del Data Center [kWh];
- $E_{IT,h}$ è il consumo orario di energia elettrica relativo alle sole apparecchiature IT del Data Center [kWh];
- $WL_{processing,h}$ è il servizio orario erogato dal Data Center per la funzionalità di server [GTOPS];
- $WL_{storage,h}$ è il servizio orario erogato dal Data Center per la funzionalità di storage [Gbyte];
- $WL_{networking,h}$ è il servizio orario erogato dal Data Center per la funzionalità di networking [Gbps];

Il soggetto proponente può proporre una frequenza di campionamento diversa purché dimostri che le fluttuazioni del consumo specifico in funzione delle variabili operative non intervengano a livello a orario, fermo restando che qualsiasi ipotesi dovrà essere cautelativa per l'Amministrazione. La medesima frequenza dovrà essere uguale o superiore a quella adottata per la determinazione del consumo di baseline.

In questa tipologia di intervento devono essere considerate quali variabili operative:

- la temperatura media oraria esterna, $T_{ext,h}$
- il Workload utile erogato orario, $WL_{E,POST,h}$

al fine di normalizzare il valore del $PUE_{baseline,h}(T_{ext,h}; WL_{E,POST,h})$ e del $\eta_{IT,baseline,h}(WL_{E,POST,h})$ alle condizioni di esercizio post-intervento.

Il risparmio energetico nel periodo di monitoraggio dovrà essere calcolato con una frequenza di campionamento oraria, in generale, come somma di due contributi, il primo riferito a interventi sull'infrastruttura IT, il secondo a interventi di altra natura (cooling, impianti elettrici, etc.):

$$RISP = RISP_{IT} + RISP_{non-IT}$$

dove:

$$RISP_{IT} = \sum_{h=1}^{h_{pdm}} \left(\frac{1}{\eta_{IT,baseline,h}(WL_{E,POST,h})} - \frac{1}{\eta_{IT,POST,h}} \right) \cdot WL_{E,POST,h} \cdot f_e$$

$$RISP_{non-IT} = \sum_{h=1}^{h_{pdm}} (PUE_{baseline,h}(T_{ext,POST,h}; WL_{E,POST,h}) - PUE_{POST,h}) \cdot E_{IT,h} \cdot f_e$$

con:

- $WL_{E,h} = N_1 \cdot WL_{processing,h} + N_2 \cdot WL_{storage,h} + N_3 \cdot WL_{networking,h}$
- $E_{non-IT,h} = E_{TOT,h} - E_{IT,h}$
- Power Usage Effectiveness, $PUE_{POST,h} = \frac{E_{TOT,h}}{E_{IT,h}}$
- IT Equipment Average Efficiency Index, $\eta_{IT,POST,h} = \frac{WL_{E,h}}{E_{IT,h}}$
- f_e è il fattore di conversione da energia elettrica, in kWh, a energia primaria, in tep pari a $0,187 \cdot 10^{-3}$ tep/kWh;
- h_{pdm} ore del periodo di monitoraggio.

Mediante regressione multivariata dei dati misurati nel periodo di monitoraggio ante intervento per il Data Center ante intervento, è possibile normalizzare alle condizioni di funzionamento post intervento i valori di baseline $\eta_{IT,baseline,h}(WL_{E,POST,h})$ e $PUE_{baseline,h}(T_{ext,POST,h}; WL_{E,POST,h})$. In particolare:

- Sulla base dei dati misurati ante intervento è possibile costruire una regressione lineare che correli il $\eta_{IT,ANTE}$ con il Workload utile erogato misurato nel periodo di monitoraggio ante intervento $WL_{E,ANTE}$. Una volta costruita la curva di regressione sarà possibile normalizzare il $\eta_{IT,ANTE}$ al variare del Workload utile erogato misurato nelle condizioni post intervento e calcolare il valore orario di $\eta_{IT,baseline,h}(WL_{E,POST,h})$;
- Sulla base dei dati misurati ante intervento è possibile costruire una regressione lineare che correli il PUE ante intervento con la temperatura esterna ($T_{ext,ante}$) e il Workload utile erogato ($WL_{E,ANTE}$) misurati nel periodo di monitoraggio ante intervento . Una volta costruita la curva di regressione sarà possibile normalizzare il PUE ex-ante al variare della temperatura esterna e del Workload utile erogato ,misurati nelle condizioni post intervento, e calcolare il valore orario di $PUE_{baseline,h}(T_{ext,POST,h}; WL_{E,POST,h})$.

Nei casi in cui gli interventi di efficienza realizzati riguardino solo interventi sull'infrastruttura IT, ovvero solo interventi non riconducibili all'IT, è facoltà del Soggetto Proponente proporre un'applicazione parziale dell'algoritmo, ovvero richiedere che il calcolo sia effettuato limitatamente al termine $RISP_{IT}$ oppure, rispettivamente, a $RISP_{non-IT}$.

Nei casi in cui le misure ante intervento siano incomplete o non disponibili, resta ferma la possibilità di determinare il consumo di baseline secondo quanto previsto dal punto 1.3 dell'allegato 1 al D.M. 21 luglio 2025.

5 BIBLIOGRAFIA

- [1] ISO/IEC 30134. Data Centers — key performance indicators (pue, wue, cue, etc.). White Paper, 2019.
- [2] IEA (2025), Energy and AI, IEA, Paris <https://www.iea.org/reports/energy-and-ai>, License: CC BY 4.0.
- [3] Ciaramella, Andrea, and Matteo Roveda. Data Centers: localizzazione, caratteristiche e prestazioni delle nuove fabbriche di dati. Franco Angeli, 2018.
- [4] Schneider Electric. The different technologies for cooling Data Centers. White Paper, 59, 2017.
- [5] ASHRAE. Thermal Guidelines for Data Processing Environments. ASHRAE, Peachtree Corners, GA, 5th edition, 2021.
- [6] O. Van Geet and D. Sickinger. Best practices guide for energy-efficient Data Center design. Technical report, US Department Of Energy (DOE), National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2024.
- [7] The Green Grid. PUE: A comprehensive examination of the metric. White Paper, 2019.
- [8] Schneider Electric. The different technologies for cooling Data Centers. White Paper, 55, 2019.
- [9] The Green Grid. Cue and wue metrics report. White Paper, 49, 2020.
- [10] J. Chu and X. Huang. Research status and development trends of evaporative cooling air-conditioning technology in Data Centers. Energy and Built Environment, 4(1):86–110, 2023, doi: <https://doi.org/10.1016/j.enbenv.2021.08.004>.
- [11] Microsoft. Sustainable datacenter design and operations, 2021.
- [12] Google. Sustainable datacenter design and operations, 2020.
- [13] Intel. Direct liquid cooling for high performance computing, 2021.
- [14] ASHRAE. TC9.9. Liquid cooling guidelines for datacom equipment centers, 2021.
- [15] Unione Europea. Directive (eu) 2023/1791 (energy efficiency directive, eed), 2023.
- [16] European Commission Joint Research Centre (JRC). Eu code of conduct for data centres (energy efficiency), 2023.
- [17] D. Al Kez, A. M. Foley, F. W. B. H. Wong, A. Dolfi, and G. Srinivasan. Ai-driven cooling technologies for high-performance data centres: state-of-the-art review and future directions. Sustainable Energy Technologies and Assessments, 82:104511, 2025, doi: <https://doi.org/10.1016/j.seta.2025.104511>.
- [18] Charles E. Leiserson et al. ,There’s plenty of room at the Top: What will drive computer performance after Moore’s law?.Science368,eaam9744(2020).DOI:10.1126/science.aam9744.
- [19] A. Chinnici, E. Ahmadzada, A.-L. Kor, D. De Chiara, A. Domínguez-Díaz, L. de Marcos Ortega, and M. Chinnici. Towards sustainability and energy efficiency using data analytics for hpc Data Centers. Electronics, 13(17):3542, 2024, doi: <https://doi.org/10.3390/electronics13173542>.

- [20] R. Evans and J. Gao. Deepmind ai reduces google Data Center cooling bill by 40%. <https://deepmind.google/discover/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40/>, July 2016.
- [21] J. Athavale, C. Bash, W. Brewer, M. Maiterth, D. Milojevic, H. Petty, and S. Sarkar. Digital twins for Data Centers. *Computer*, 57(10):151–158, 2024.
- [22] H. Kahil, S. Sharma, P. V. alisuo, and M. Elmusrati. Reinforcement learning for Data Center energy efficiency optimization: A systematic literature review and research roadmap. *Applied Energy*, 389:125734, 2025, doi: <https://doi.org/10.1016/j.apenergy.2025.125734>.
- [23] R. Wang, Z. Cao, X. Zhou, Y. Wen, and R. Tan. Phyllis: Physics-informed lifelong reinforcement learning for Data Center cooling control. In *Proceedings of the 14th ACM International Conference on Future Energy Systems*, pages 114–126, 2023, <https://doi.org/10.1145/3575813.3595189>.