

UNIVERSITÀ
DEGLI STUDI
DI PADOVA



Tesi di Laurea Triennale in Ingegneria Elettronica

Sostenibilità ambientale e consumo energetico nei data center

Relatore:

Prof. Antonio Rodà

Candidato:

Davide Antonelli

Anno Accademico 2023/2024

Sommario

L'elaborato illustra l'impatto inquinante di alcune tecnologie proprie della società dell'informazione e della comunicazione prestando particolare attenzione al ruolo svolto dalle infrastrutture che si occupano di elaborare e archiviare dati: i Data Center.

Si forniscono nozioni sul loro funzionamento, illustrando i componenti principali che li costituiscono e analizzando il loro impatto inquinante. Infine viene svolto un focus su diverse tecniche hardware e software ai fini di una maggiore efficienza energetica.

Tale elaborato pone come fine ultimo quello di illustrare ai più un problema ambientale che la transizione digitale sta creando e che risulta ancora sottovalutato (a fini economici) e non compreso degli stessi utilizzatori finali di tali servizi.

Indice

1	Tecnologie digitali e ambiente	1
1.1	Le tecnologie digitali	1
1.2	L'impatto ambientale	4
1.2.1	Tecnologie hardware	7
1.2.2	Reti di telecomunicazione	7
2	I Data Center	9
2.1	Diffusione dei data center	10
2.2	Struttura di un data center	11
2.2.1	IT Equipment	12
2.2.2	Power Delivery	12
2.2.3	Sistema di raffreddamento	13
2.3	Classificazione dei data center: Tier classification	15
3	Il consumo energetico	18
3.1	Il PUE	18
3.2	Consumi ed impatto ambientale	21
3.2.1	Focus sull'Unione Europea	23
4	Tecniche risparmio energetico	29
4.1	La progettazione	30
4.1.1	Dimensionamento	31
4.1.2	Data Center modulari	31
4.2	Tecniche per l'IT Equipment	34
4.2.1	Blade Server	35
4.2.2	Storage	35
4.2.3	Virtualizzazione	36
4.2.4	Consolidamento	37
4.2.5	L'Intelligenza Artificiale	38
5	Conclusioni	41

1 Tecnologie digitali e ambiente

Nel contesto dei movimenti e dell'etica ambientalista, il concetto di sostenibilità si riferisce a pratiche e comportamenti che durano nel tempo perché rispettano limiti ecologici tali da portare benefici alla generazione presente e a quelle future[3]. Le tecnologie digitali possono rappresentare un'opportunità ai fini dello sviluppo sostenibile: basta pensare all'uso di dati sul traffico urbano per ottimizzare le infrastrutture e ridurre l'inquinamento oppure ai modelli per stabilire in anticipo le quantità di prodotti di cui hanno bisogno ogni giorno negozi di alimentari e consumatori, così da ridurre al minimo gli sprechi. Allo stesso tempo, elaborare e immagazzinare grandi quantità di dati e smaltire i dispositivi digitali, sono tutte azioni che hanno un elevato impatto ambientale e sociale. Il tema della sostenibilità ambientale deve quindi essere riconosciuto come tema di fondamentale importanza. Un progresso tecnologico consapevole e sostenibile deve andare di pari passo con la salvaguardia del pianeta[9]. L'elaborato illustra l'impatto inquinante di alcune tecnologie proprie della società dell'informazione e della comunicazione prestando particolare attenzione al ruolo svolto dalle infrastrutture che si occupano di elaborare e archiviare ingenti quantità di informazioni: i Data Center.

1.1 Le tecnologie digitali

La diffusione delle tecnologie digitali in tutte le attività umane è un fenomeno che riflette al meglio i tempi in cui ci troviamo. I dispositivi e i servizi delle tecnologie dell'informazione e della comunicazione hanno assunto un ruolo centrale nelle nostre vite e hanno trasformato radicalmente il modo in cui lavoriamo, comunichiamo, viaggiamo e giochiamo negli ultimi decenni. Infatti, mentre la popolazione umana è solo raddoppiata negli ultimi 50 anni, il consumo globale di dispositivi elettronici e digitali è cresciuto di sei volte nello stesso arco di tempo. Tali tecnologie vengono definite digitali in quanto fondate sul sistema di cifratura binario, in inglese binary digit (o bit). Un bit può assumere valore 0 o 1. Il cardine delle tecnologie digitali è il computer anche definito calcolatore elettronico [8]. Hardware e software sono le due parti fondamentali di un computer. Il componente fisico è l'hardware, composto da una varietà di dispositivi elettronici, mentre il software è costituito da insiemi di istruzioni (programmi) che l'hardware esegue. I primi computer sono apparsi negli anni '50 ma la loro evoluzione è stata fin da subito un fenomeno senza precedenti per la sua rapidità. Ai giorni nostri siamo arrivati a mettere in una tessera di silicio di pochi millimetri di lato (i chip) quello che inizialmente occupava un'intera stanza. Con il diminuire delle dimensioni è diminuito anche il costo e di

conseguenza il consumo di energia. Il fenomeno è descritto dalla legge di Moore, enunciata nel 1965, affermando che ogni anno e mezzo si raddoppia la densità dei componenti elettronici realizzati su un chip e parallelamente se ne riduce il costo. Tale fatto sta alle fondamenta dell'enorme diffusione dei computer in ogni loro forma ai giorni nostri. L'altro lato dell'era digitale è Internet. Le origini risalgono agli anni Sessanta grazie al progetto ARPANET (Advanced Research Project Agency NETwork), una rete che collegava i centri di ricerca americani in grado di resistere agli attacchi nucleari. ARPANET era una rete con una struttura a maglie, in cui i messaggi tra due nodi potevano muoversi su percorsi diversi, passando attraverso altri nodi della rete. Per eseguire tale operazione, ogni nodo doveva essere in grado di reindirizzare i messaggi e quindi di elaborarli. La rete era perciò basata su tecniche digitali, tipiche del computer, mentre all'epoca le telecomunicazioni erano di tipo analogico, ovvero basate sulla modulazione di onde elettromagnetiche. Così il computer, nato come strumento di calcolo, si è evoluto anche in uno strumento di comunicazione. Si verifica quindi la convergenza e successiva integrazione tra due settori sviluppatisi indipendentemente l'uno dall'altro, l'informatica e le telecomunicazioni. Questa evoluzione da origine ad una nuova era tecnologica, quella delle ICT (Information and Communication Technologies). Nel 1986 tutti i centri di ricerca degli Stati Uniti erano connessi in rete. La diffusione della rete telematica era iniziata fino ad arrivare alle cronache dei giorni nostri, con Internet che collega centinaia di milioni di computer nel mondo e rappresenta uno strumento senza precedenti di diffusione e condivisione delle conoscenze e di stimolo alla crescita del sapere. L'interrelazione computer-Internet sta muovendo una vera e propria rivoluzione nel modo di comunicare, lavorare, studiare, divertirsi, con sviluppi e conseguenze ancora oggi imprevedibili. Si pensi alla diffusione di dispositivi mobili (tablet, smartphone, PDA, ecc.) in grado di accedere a Internet, comunicare, creare e trasmettere immagini oltre che fornire servizi altamente tecnologici come il GPS (Global Positioning System), da cui è ormai impossibile fare a meno. Alla stessa simbiosi corrisponde l'orientamento attuale verso il cosiddetto computer sulla nuvola (cloud computing), in cui archivi, programmi e risorse di elaborazione sono virtualizzati e disponibili per tutti con grandi vantaggi economici e gestionali. I precedenti esempi vengono comunemente raggruppati nella definizione generale di tecnologie digitali.

Tale concetto è immensamente ampio ed al suo interno si possono individuare 5 macrocategorie che ben caratterizzano le aree in cui queste innovazioni si stanno sviluppando [11]:

- *La tecnologia mobile* è un'unione fra hardware, software, networking e sistemi operativi, la quale consente a individui e organizzazioni di accedere a contenuti e servizi sempre e ovunque, al bisogno. Gli esempi più rappresentativi sono tablet e smartphone;
- *La tecnologia di comunicazione machine-to-machine (M2M)* consente lo scambio automatizzato di informazioni tra varie apparecchiature o centri di comando senza la necessità dell'intervento umano, fornisce agli utenti migliori prestazioni e minori costi attraverso l'utilizzo di informazioni disponibili in tempo reale. Aiuta anche a tracciare le risorse, controllarle e mantenerle in remoto;
- *Il cloud computing* consente alle applicazioni e ai contenuti basati sul Web di essere ospitati in un'unica posizione e resi disponibili a numerosi dispositivi contemporaneamente senza la necessità di scaricare o installare alcun software. I dispositivi dispongono di una connessione di rete, solitamente tramite Internet. Per alcuni modelli di business come piattaforme di condivisione la "nuvola" è un elemento chiave;
- *La tecnologia social* sono strumenti per la comunicazione e l'interazione sociale che aiutano gli utenti a connettersi e rimanere in contatto. Forum online, blog e software collaborativi sono tutti esempi di tecnologia social, oltre a piattaforme ben note come Facebook, Twitter, Whatsapp, LinkedIn e WeChat;
- *I big data analytics* sono una tecnologia che coinvolge 2 elementi; "big data" si riferisce a una raccolta di set di dati estremamente grandi e complessi che sono troppo complessi per essere gestiti o analizzati dalle applicazioni tradizionali, invece, il termine "big data analysis" si riferisce all'analisi di tali dati per produrre approfondimenti che facilitino decisioni ottimizzate.

Le tecnologie digitali stanno portando infine allo sviluppo e alla realizzazione della cosiddetta "quarta rivoluzione industriale" o 'Industria 4.0'. Il termine fa riferimento a un modello di produzione e gestione aziendale. Secondo una definizione che ne dà il Mise (Ministero delle Imprese e del Made in Italy), gli elementi che caratterizzano il fenomeno sono la connessione tra sistemi fisici e digitali, analisi complesse attraverso Big Data e adattamenti real-time. Semplificando: utilizzo di macchinari connessi alla rete, analisi delle informazioni ricavate da essa e la possibilità di una gestione più flessibile del ciclo produttivo. Le tecnologie abilitanti, menzionate dal Mise, vanno dalle stampanti 3D ai robot programmati per compiti specifici, passando per la gestione di dati in cloud e l'analisi dei dati per individuare punti deboli e punti di forza della produzione.

1.2 L'impatto ambientale

Le tecnologie digitali rappresentano senza ombra di dubbio un importante opportunità di sviluppo sostenibile. Di fatto un rapporto delle Nazioni Unite ha assegnato alla rivoluzione digitale e ai big data un ruolo cruciale nella realizzazione degli obiettivi per uno sviluppo sostenibile [7]. Il documento illustra come il crescente volume di dati, generati passivamente e raccolti automaticamente, consente ai responsabili politici di rilevare e sostenere il raggiungimento di alcuni obiettivi: tra cui l'“azione per il clima” è favorita dall'uso di dati satellitari atti a monitorare la deforestazione in corso ad esempio. Il documento delle Nazioni Unite trascura però un rischio cruciale per lo sviluppo sostenibile: il fatto che la rivoluzione digitale e i big data hanno un considerevole impatto ambientale. Il manifesto scritto nel 2013, dai partecipanti della prima conferenza ICT for Sustainability (ICT4S), afferma che nonostante le tecnologie informatiche abbiano la capacità di rendere i modelli di produzione e consumo più sostenibili, la storia della tecnologia ha mostrato che una maggiore efficienza energetica non contribuisce automaticamente allo sviluppo sostenibile [2]. L'economia circolare e la decarbonizzazione di tutti i settori possono essere supportate da soluzioni digitali, che contribuiranno a raggiungere gli obiettivi di sostenibilità delineati nel Green New Deal europeo, ma non è necessariamente scontato. In realtà, fino a questo punto, i cambiamenti digitali hanno sostenuto modelli di crescita ad alta intensità di risorse e di emissioni di gas serra che causano il riscaldamento globale. Il consumo di elettricità per computer, dispositivi elettronici e infrastrutture digitali è in costante aumento. Inoltre, l'elettricità che non proviene da risorse rinnovabili emette gas serra. Il carbon footprint (tradotto, “l'impronta del carbonio”), è un parametro che viene utilizzato per stimare tali emissioni di gas serra causate da un prodotto, da un servizio, da un'organizzazione, da un evento o da un individuo, espresse generalmente in tonnellate di CO₂ equivalente). Le emissioni delle tecnologie digitali utilizzate per la trasmissione, la ricezione e l'elaborazione dei dati sono aumentate dal 2% nel 2008 al 3,7% nel 2020 e sono proiettate all'8,5% nel 2025, pari alle emissioni di tutti i veicoli leggeri attualmente in circolazione [17].

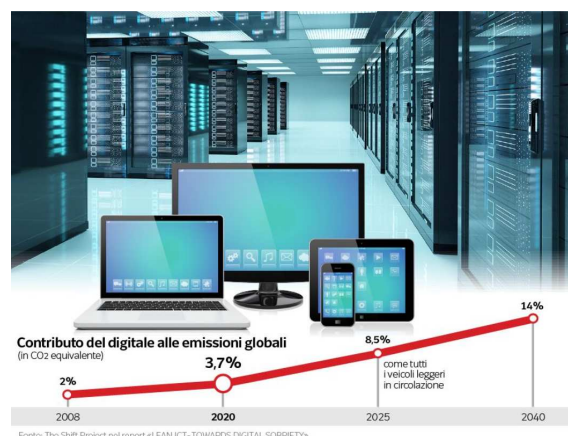


Figura 1: Contributo del digitale alle emissioni globali

Fonte: The Shift Project, lean-ict towards digital sobriety, (2019)

Territorial (MtCO₂)

Rank	Country	MtCO ₂
1	China	11472
2	United States of America	5007
3	India	2710
4	Russian Federation	1756
5	Japan	1067
6	Iran	749
7	Germany	675
8	Saudi Arabia	672
9	Indonesia	619
10	South Korea	616
11	Canada	546
12	Brazil	489
13	Turkey	446
14	South Africa	436
15	Mexico	407
16	Australia	391
17	United Kingdom	347
18	Italy	329
19	Poland	329
20	Vietnam	326

Figura 2: Classifica produzione MtCO₂ primi 20 Stati

Fonte: www.globalcarbonatlas.org, (2021)

Attualmente le tecnologie digitali sono infatti responsabili del 4% di tutte le emissioni di gas serra; perciò, confrontando le emissioni del digitale nel 2021 (ultimi dati disponibili su www.globalcarbonatlas.org) con le emissioni di tutti i Paesi si può vedere che se le infrastrutture digitali fossero uno Stato, sarebbe il quinto più grande consumatore di energia al mondo con 1485 MtCO₂. Secondo lo studio condotto da Assessing ICT global emissions footprint, l'impatto della tecnologia digitale sarà del 14% nel 2040 [1].

Che impatto ha il Digitale sull'ambiente?

Emissioni di CO2 stimate in relazione all'uso dei dispositivi per persona e anno (in kg)

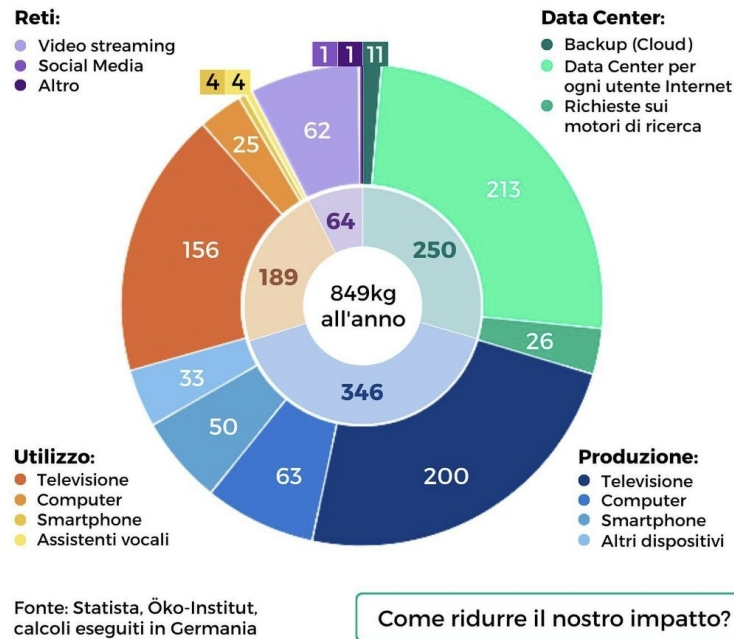


Figura 3: Divisione KgCO2 prodotta dai diversi settori dell'ICT
Fonte: Statista, Oeko-institut, Germania, (2022)

Analizzando nel dettaglio quali siano i maggiori fattori dell'industria ICT che generano tali emissioni risulta possibile identificare tre macrocategorie: tecnologie hardware che coinvolgono produzione e utilizzo di smartphone e desktops, reti di telecomunicazioni che comprendono lo streaming video e i social media ed infine i data center di cui verrà svolto uno studio più approfondito nel capitolo successivo. La precedente immagine vuole rendere un quadro generale sulle emissioni di CO2 stimate in relazione all'uso dei dispositivi per persona e anno (in kg).

1.2.1 **Tecnologie hardware**

Quando si parla di impatto ambientale con particolare riferimento alle tecnologie hardware si intende la quantità di CO₂ emessa da dispositivi quali smartphone, PC, tablet e tutti i dispositivi elettronici che vengono utilizzati nel quotidiano da ognuno di noi, durante il loro ciclo di vita, il quale si articola in tre fasi: produzione, utilizzo e smaltimento. Analizzando i devices che più usiamo nella nostra giornata, gli smartphone emerge che produrre, usare e smaltire gli smartphone ha un costo che si paga anche in termini di inquinamento ambientale. Il rapporto di Deloitte, che riporta la stima della quantità di CO₂ prodotta globalmente fino al 2022, fa luce sul fenomeno. La società di analisi prevede che entro la fine dell'anno vi saranno in circolazione 4,5 miliardi di smartphone in tutto il mondo, un numero di dispositivi che genereranno 146 milioni di tonnellate di CO₂ in totale. Esattamente l'83 per cento delle emissioni attribuibili agli smartphone sono prodotte nelle prime fasi del loro ciclo di vita, dall'estrazione delle materie prime necessarie per la produzione al trasporto. Le emissioni in fase di utilizzo, in particolare dopo il primo anno di acquisto, sono molto più contenute (11%) rispetto al totale. L'ultima quota di inquinamento prodotta dagli smartphone (5%) è quella riconducibile al recupero e ripristino di dispositivi giunti a fine vita [6].

1.2.2 **Reti di telecomunicazione**

A causa della pandemia, il confinamento domestico ha portato ad un aumento degli utenti e delle richieste di informazioni. La quarantena globale è stata indubbiamente positiva per l'ambiente, soprattutto considerando il forte calo delle emissioni legate ai trasporti. Tuttavia, è importante riconoscere gli effetti dell'utilizzo di piattaforme di streaming, chiamate in conferenza e archiviazione e trasmissione di dati su Internet. Un comune servizio di streaming richiede 7 GB per un'ora di streaming in alta qualità (Ultra HD o 4k) con un'impronta di carbonio di 441 g CO₂e/ora (mediana globale). Lo streaming video a questa qualità per quattro ore al giorno comporterebbe un'impronta di carbonio mensile di 53 kg CO₂e. Tuttavia, abbassando la qualità dei video da HD a standard, l'impronta mensile scenderebbe a 2,5 kg CO₂e. Se 70 milioni di abbonati allo streaming abbassassero la qualità video dei loro servizi, ci sarebbe una riduzione delle emissioni di CO₂e di 3,5 milioni di tonnellate, l'equivalente dell'eliminazione di 1,7 milioni di tonnellate di carbone negli Stati Uniti. Allo stesso modo, un servizio di videoconferenza standard ha un'impronta di carbonio di 157 gCO₂e/ora. Se si tenessero 15 riunioni di un'ora alla settimana, l'impronta di carbonio mensile sarebbe di 9,4 kg di CO₂e. Spegnendo semplicemente il video, tuttavia, ridurrebbe le emissioni mensili a 377 g CO₂e. Questo risparmierebbe le

emissioni prodotte da una ricarica di uno smartphone ogni notte per oltre 3 anni (1151 giorni). Se 1 milione di utenti di videoconferenze facesse questo cambiamento, ridurrebbe collettivamente le emissioni di 9023 t di CO₂e in un mese, l'equivalente delle emissioni di energia elettrica di una città di 36.000 persone per un mese con il carbone. Una breve e-mail inviata da telefono a telefono tramite Wi-Fi equivale a 0,3 grammi di CO₂, una breve e-mail inviata da laptop a laptop emette 17 g di CO₂ e una lunga e-mail con allegato inviata dal laptop potrebbe produrre 50 g di CO₂ [17]. Secondo Carbonfootprint, società di consulenza ambientale, una ricerca su Google può causare da 1 g a 10 g di emissioni di CO₂. Detta così non sembra molto ma teniamo presente che Google elabora circa 47.000 ricerche al secondo, 3.5 miliardi al giorno. Tenendo conto che gli apparati informatici contribuiscono a circa il 10% del consumo di energia elettrica globale, è stato calcolato che un utente medio emette, navigando nel web, circa 244 kg ogni anno.

2 I Data Center

La rete ed i sistemi informatici vengono percepiti come immateriali. La metafora del cloud, la nuvola eterea e senza limiti, dove dati, foto e video possono essere immagazzinati e resi accessibili da qualunque postazione o dispositivo collegato alla rete, è indicativa di questa errata prospettiva sul digitale. Nella pratica questi sistemi necessitano di vere e proprie infrastrutture industriali che occupano terreni, urbani ed extraurbani, consumano una quantità rilevante di energia, rilasciano gas serra nell'atmosfera, con conseguenze importanti sulla crisi climatica, ed infine terminano il loro ciclo di utilizzo in discariche, rilasciando sostanze tossiche. Gli edifici che ospitano tale infrastruttura sono denominati Data Center. Conosciuto anche con l'acronimo di CED (ossia Centro Elaborazione Dati) così come col termine server farm, un data center è il luogo fisico dotato di tutte le infrastrutture e gli impianti base adatti ad ospitare elaboratori elettronici che gestiscono le informazioni di una o più istituzioni; oltre ai componenti associati come sistemi di telecomunicazione e di memorizzazione. Generalmente includono un doppio sistema di alimentazione, una rete di comunicazione ridondante, un sistema di controllo ambientale (ad esempio, raffreddamento e climatizzazione) e importanti sistemi di sicurezza. Essi sono il sistema nervoso centrale del XXI secolo [10]. I CED consumano una quantità crescente di energia per svolgere le loro operazioni e raffreddare i server. Dovendo operare 24 ore su 24 e dovendo garantire ad ogni costo una certa continuità di servizio necessitano pure di generatori a diesel per mantenere i server operativi nel caso di interruzione della fornitura di energia elettrica, producendo così ulteriori emissioni di gas serra. Si stima che i CED hanno la più rapida crescita della carbon footprint di tutto il settore delle tecnologie dell'informazione e della comunicazione.

2.1 Diffusione dei data center

Le grandi sale computer, dove sono stati ospitati i primi sistemi informatici, sono i luoghi in cui sono nati i data center. I primi sistemi erano estremamente difficili da gestire e necessitavano di un settaggio specifico per funzionare. È stato necessario costruire appositi locali per alloggiarli e sistemarli perché era necessario far passare numerosi cavi e utilizzare strutture di alloggiamento, per i vari componenti, in modo sicuro, come: pavimenti sopraelevati, passerelle portacavi e telai standard per il montaggio delle apparecchiature. Tutti i componenti consumano molta energia e, durante l'uso, erano molto soggetti a surriscaldamenti e necessitavano di un raffreddamento costante tramite complessi sistemi di raffreddamento, che occupavano molto spazio. La sicurezza è stata una considerazione chiave perché queste tecnologie sono state sviluppate principalmente in un ambiente militare. Ogni sistema era inizialmente unico rispetto agli altri, ma dopo il 1980, con la diffusione dell'uso dei computer in tutto il mondo e la complessità dei processi aziendali, le aziende si sono rese conto della necessità di stabilire regole per la gestione delle risorse IT. È stato possibile utilizzare un design gerarchico e ordinato che consentisse di collocare i server in una stanza specifica all'interno delle aziende grazie anche all'avvento dell'elaborazione client-server e la disponibilità a basso costo delle apparecchiature di rete. Un'ampia diffusione dell'utilizzo dei data center si è verificata durante la bolla del dot-com, quando più aziende avevano bisogno di connettività Internet veloce e disponibilità di servizi 24 ore su 24 per distribuire i propri servizi e avere una presenza online. Per molte piccole imprese non era possibile sostenere i costi di installazione o aggiornamento dell'hardware di questa apparecchiatura. A tal fine sono state fondate e sviluppate rapidamente società per costruire questi grandi sistemi, noti come Internet Data Center (IDC), che consentivano a più aziende di utilizzare una varietà di soluzioni per la distribuzione di servizi e operazioni. La progettazione, la costruzione e il funzionamento dei data center si sono trasformati nel tempo in una disciplina riconosciuta. I documenti standard, prodotti da esperti e organizzazioni, delineano le specifiche per la progettazione del data center e le migliori pratiche da seguire [15]. Il numero di data center e la quantità di energia utilizzata dai server devono entrambi aumentare per soddisfare le esigenze aziendali nel mercato online. La domanda di potenza di calcolo e archiviazione aumenterà man mano che sempre più aziende amplieranno la loro selezione di servizi digitali. L'online banking, l'e-commerce, il monitoraggio delle spedizioni online, il cloud gaming e il download e lo streaming di musica e video sono alcuni esempi di servizi digitali che sono diventati molto popolari di recente.

2.2 Struttura di un data center

Nel seguente paragrafo verranno descritte le parti di cui è formato un data center. Verranno trattati i componenti principali che consentono il regolare funzionamento del data center e che causano un notevole consumo di energia elettrica: macchine informatiche, sistemi di rifornimento di energia e sistemi di raffreddamento; tralasciando la trattazione delle altre componenti presenti come la struttura, gli uffici e i laboratori dei tecnici e dello staff.

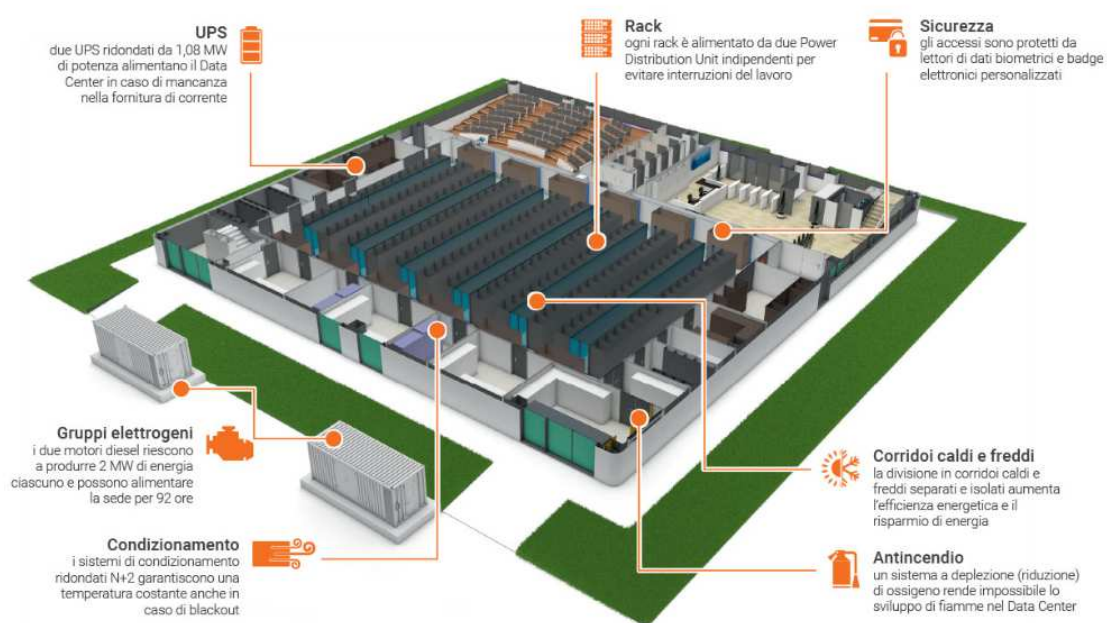


Figura 4: Layout data center di Pisa
Fonte: www.citconsult.it/datacenter

2.2.1 IT Equipment

Hardware, software, sistemi informativi e linguaggi di programmazione sono solo alcuni esempi delle varie tecnologie e strumenti che rientrano nell'ambito dell'IT (Information Technology). In altre parole, tutto ciò che può essere distribuito attraverso qualsiasi supporto multimediale e rende percepibili dati, informazioni o conoscenze in un formato visivo. È il componente principale di un data center, ne consente le operazioni essenziali e ospita le apparecchiature elettroniche necessarie per l'elaborazione dei dati (server), l'archiviazione dei dati (storage) e le comunicazioni (rete). L'insieme di questi componenti è definito Information Technology Equipment. Queste parti lavorano insieme per elaborare, archiviare e inviare informazioni digitali. Per evitare il surriscaldamento e gli incidenti che coinvolgono macchine e personale di supporto, hanno bisogno di un ambiente privato in cui siano posizionati e organizzati con cura. All'inizio della fase di progettazione del centro, è necessario determinare il numero di componenti delle apparecchiature informatiche. Al fine di accogliere futuri aggiornamenti o l'aggiunta di nuovi componenti evitando anche spese irrazionali o non necessarie, è fondamentale tenere conto della potenza richiesta all'inizio e prevedere la potenza futura. Questo è uno degli aspetti che incide maggiormente sui costi del data center, sia per l'acquisto che per la manutenzione. Hanno anche un costo in termini di consumo di energia elettrica, compresi i costi legati al sistema di raffreddamento; più grande è l'apparecchiatura IT, maggiore è la potenza necessaria per raffreddare correttamente i suoi componenti.

2.2.2 Power Delivery

È il sistema di alimentazione per il data center. In genere, l'infrastruttura è costituita da una connessione alla rete, alimentazione per il sistema di raffreddamento, batterie di backup e generatori. L'energia viene recuperata dalla rete elettrica esterna. Un generatore all'interno del data center interviene in caso di interruzione di corrente e fornisce al sistema l'energia necessaria. In questi apparecchi elettrici vengono utilizzate batterie tampone ricaricate, che hanno il vantaggio di mantenere una potenza costante anche in caso di interruzioni della rete elettrica e del generatore interno al fine di sopperire a fluttuazioni di rete o a brevi perdite di potenza. Non prima di essere adeguatamente regolata alla tensione operativa dei componenti IT, utilizzando l'unità PDU (unità di distribuzione dell'alimentazione), l'energia viene distribuita alle macchine del data center attraverso l'UPS (gruppo di continuità). Per aumentare l'affidabilità del sistema, è inoltre possibile utilizzare una doppia connessione a due diverse PDU. Pertanto, il livello di ridondanza nel sistema è un fattore cruciale nell'infrastruttura. Il data center deve avere sempre accesso

all'energia, quindi sono previste più fonti di energia e batterie di backup. Più il sistema è ridondante, meglio sarà in grado di resistere alle emergenze, ma allo stesso tempo i costi di installazione e manutenzione di questo sistema saranno maggiori.

2.2.3 Sistema di raffreddamento

L'infrastruttura di raffreddamento di un data center occupa molto spazio all'interno di esso. I numerosi computer e macchine che operano lì generano molto calore mentre sono in funzione. Devono essere continuamente raffreddati e mantenuti a una temperatura costante per evitare il surriscaldamento e il conseguente danneggiamento dei componenti. Refrigeratori, condizionatori, pompe e ventole sono utilizzate in un sistema di raffreddamento per questo scopo.

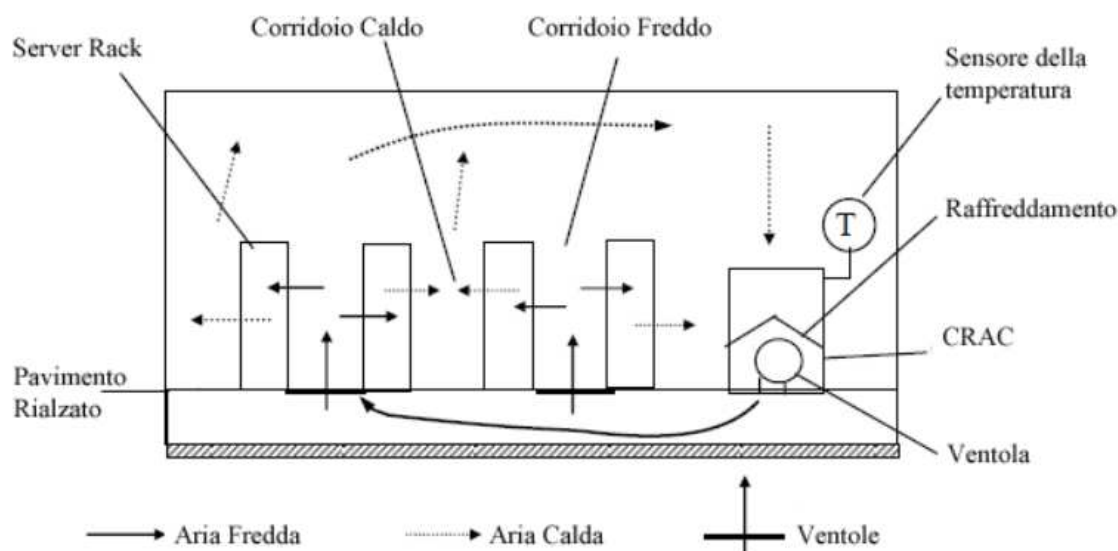


Figura 5: Sistema di raffreddamento

Fonte: Auslander D., *Viability of dynamic cooling control in a data center environment*, (2004)

Un tipico impianto di raffreddamento è rappresentato in Figura 5, composto da condizionatori modulari e pavimentazione sopraelevata che permette la circolazione dell'aria fredda. Durante la fase di progettazione del data center, le dimensioni e la potenza del sistema vengono dimensionate in base al carico previsto dei server. Il CRAC (computer room air conditioning) distribuisce l'aria in modo uniforme, controllando la temperatura con un sensore che misura tipicamente la temperatura interna a 20°C. Ventole e pompe diffondono l'aria fredda dal pavimento della stanza, raffreddando i server. L'aria calda che sale poi ritorna al condizionatore per essere nuovamente raffreddata e distribuita in tutto il data center. La combinazione di

questi elementi produce l'ambiente ideale e garantisce una maggiore longevità dei componenti. Al tempo stesso questo ecosistema ha un costo rilevante nell'economia del data center in termini di dispendio energetico.

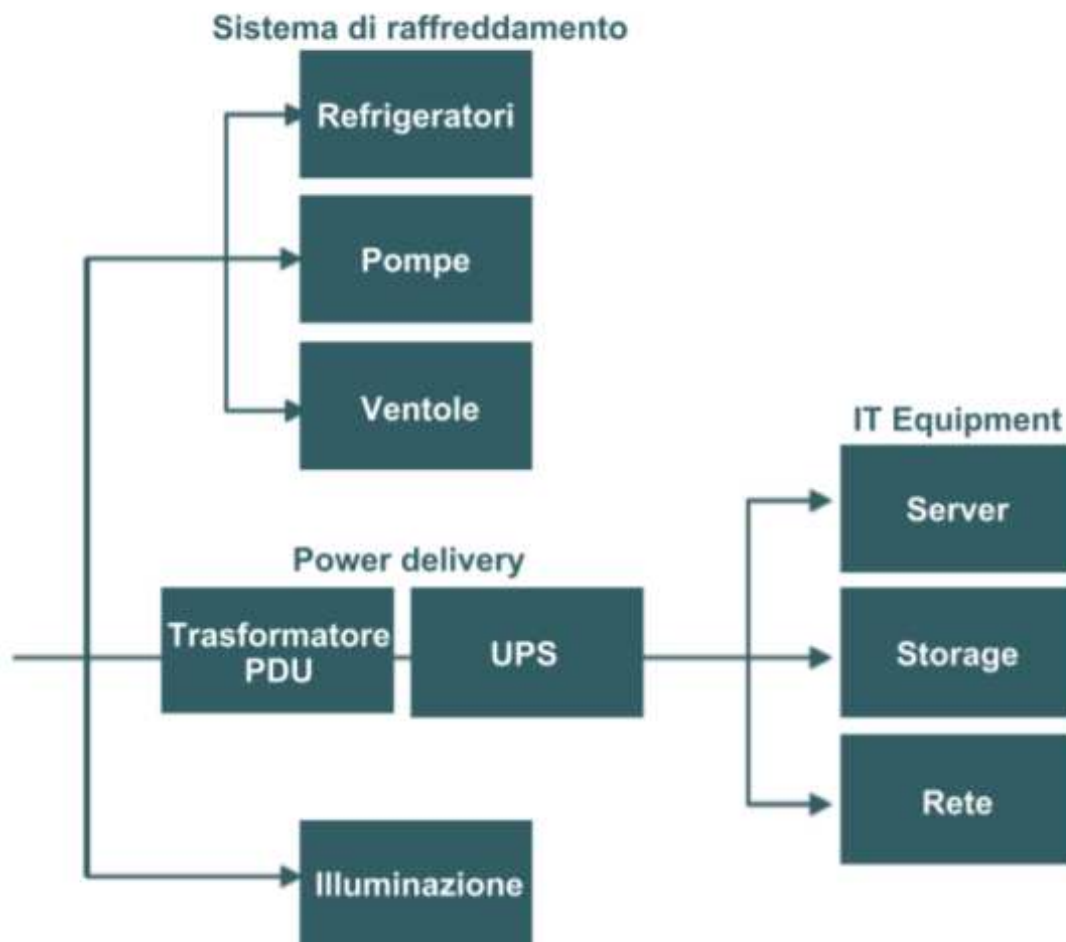


Figura 6: Componenti principali di un data center

Fonte: Ciaramella A. Roveda M., *Data Center: localizzazione, caratteristiche e prestazioni delle nuove fabbriche dati*, (2018)

2.3 Classificazione dei data center: Tier classification

La progettazione di un “Centro dati” è un procedimento estremamente articolato e complesso: connettività, ambienti ed edifici, operatività e continuità, sistemi di alimentazione e raffreddamento, sicurezza fisica e controllo degli accessi sono tutti aspetti critici che vanno presi in considerazione fin da subito. Per la definizione di standard a cui fare riferimento The Uptime Institute (la Global Data Center Authority) ha elaborato uno standard di classificazione a livelli, grazie a cui è possibile classificare i data center, come mezzo di valutazione efficace dell’infrastruttura di un data center, principalmente in termini di sicurezza, affidabilità e disponibilità dei sistemi. Nel 2005, sulla base di questo sistema di valutazione, Telecommunications Industry Association (TIA), associazione di categoria accreditata da ANSI (American National Standard Institute) ha elaborato la prima serie di standard, denominata ANSI/TIA-942, dell’infrastruttura per data center, che ne ha definito quattro livelli di classificazione con metodo confrontabile, quantificabile e approfondito [15]. La Tia-942 è stata aggiornata successivamente nel 2008 e nel 2010 e descrive i requisiti per l’infrastruttura dei Data Center (che oggi si basa su quattro livelli detti TIER) fornendo all’industria un metodo coerente e universale che consente di confrontare le strutture in base a dei criteri unici e specifici come le prestazioni dell’infrastruttura del sito (intesa come ambiente complessivo del Data Center, non solo come sistema IT) ed il tempo di attività (ovvero “livello di servizio garantito”). Le categorie fanno chiaro riferimento alle prestazioni di continuità operativa. Tier 1 è la categoria più basilare e corrisponde, ad una sala server, realizzabile, seguendo le linee guida di base per l’installazione di sistemi informatici. Il livello più elevato è un Data Center Tier 4, progettato per ospitare sistemi informatici mission-critical, dotato di sottosistemi ridondanti e zone di sicurezza suddivise e controllate con metodi biometrici di controllo degli accessi. Vediamo allora in dettaglio quali sono le differenziazioni tra i vari livelli di classificazione di un data center.

Tier Level	Requisiti
1	(Basic – 99,671%) <ul style="list-style-type: none">- Suscettibilità a interruzioni a causa di attività pianificate e non pianificate;- Mancanza di ridondanze e con singolo sistema di alimentazione e di raffreddamento;- Presenza o meno di UPS, generatori e pavimento flottante;- Fermo del Data Center: 28,8ore/anno;- Totale spegnimento durante le manutenzioni preventive.
2	(Redundant component – 99,741%) <ul style="list-style-type: none">- Meno suscettibilità a interruzioni a causa di attività pianificate e non pianificate;- Componenti ridondati e con singolo sistema di alimentazione e di raffreddamento;- Presenza di UPS, generatori e pavimento flottante;- Fermo del Data Center: 22ore/anno;- Totale spegnimento durante le manutenzioni su alimentazione e altre parti dell'infrastruttura.

-
- | | |
|----------|--|
| 3 | (Concurrently maintainable – 99,982%) <ul style="list-style-type: none">- Possibilità di effettuare manutenzioni pianificate senza interruzione, ma suscettibilità a interruzioni a causa di attività non pianificate;- Componenti ridondati e collegamenti multipli per alimentazione e raffreddamento;- Presenza di UPS, generatori e pavimento flottante;- Fermo del Data Center: 1,6ore/anno,- Non necessario lo spegnimento totale durante le manutenzioni, prevista deviazione su altri collegamenti per alimentazione ed infrastruttura. |
| <hr/> | |
| 4 | (Fault tolerant– 99,995%) <ul style="list-style-type: none">- Possibilità di effettuare manutenzioni pianificate senza interruzione, ma suscettibilità a interruzioni a causa di attività non pianificate;- Componenti ridondati e collegamenti multipli per alimentazione e raffreddamento;- Presenza di UPS, generatori e pavimento flottante;- Fermo del datacenter:0,4ore/anno;- Non necessario lo spegnimento totale durante le manutenzioni, prevista deviazione su altri collegamenti per alimentazione ed infrastruttura. |
-

Figura 7: Classificazione Tier dei Data Center

Fonte: Ciaramella A. Roveda M., Data Center: localizzazione, caratteristiche e prestazioni delle nuove fabbriche dati, (2018)

3 Il consumo energetico

I data center sono sistemi in funzione 24 ore su 24 per garantire la costante continuità di servizio. Tuttavia, due fenomeni rappresentano la maggior parte delle criticità dei data center. In primo luogo, la massiccia esplosione del numero di grandi data center, ognuno dei quali ha migliaia (se non milioni) di server a causa della continua crescita della domanda di calcolo, elaborazione, archiviazione dei dati e servizi cloud da parte di una varietà di organizzazioni come Facebook e Google, e da operatori di telecomunicazioni, come Telecom. In secondo luogo, sono state sviluppate infrastrutture informatiche su larga scala per supportare una varietà di applicazioni, da quelle che vengono eseguite per un breve periodo di tempo a quelle che vengono eseguite continuamente su piattaforme con hardware condiviso, che hanno un impatto ambientale significativo in termini di consumo energetico.

3.1 Il PUE

È importante illustrare alcuni indicatori utilizzati nella valutazione dell'efficienza energetica dei data center e nel confronto tra essi prima di iniziare l'analisi vera e propria. Si è deciso, a livello internazionale, di introdurre un nuovo parametro, il PUE, nell'ambito dei data center per evitare confusione con il tradizionale concetto di efficienza. Un'organizzazione di professionisti IT focalizzata sul miglioramento dell'efficienza energetica dei data center, Green Grid, ha introdotto la metrica PUE (Power Usage Effectiveness). Questa metrica consente di valutare le prestazioni di vari centri di elaborazione confrontandoli tra loro. Il PUE viene definito come il rapporto tra il totale dell'energia assorbita dall'intero data center e l'energia che effettivamente raggiunge le apparecchiature IT.

$$PUE = \frac{ENERGIA\ TOTALE}{ENERGIA\ IT} \quad (1)$$

Risulta essere l'inverso di un'efficienza, per cui è sempre maggiore di 1 e si riduce all'aumentare dell'efficienza energetica. Un valore PUE pari a 1 rappresenterebbe l'efficienza ottimale del data center. Idealmente, vorremmo che tutta l'energia che entra nel data center fosse utilizzata per alimentare il carico IT (server, storage e rete). Realisticamente, tuttavia, parte di questa potenza deve essere deviata per supportare il raffreddamento, illuminazione e altre infrastrutture di supporto e parte dell'energia residua viene consumata a causa di perdite nel sistema di alimentazione. L'energia rimanente va quindi a servire l'IT Equipment.

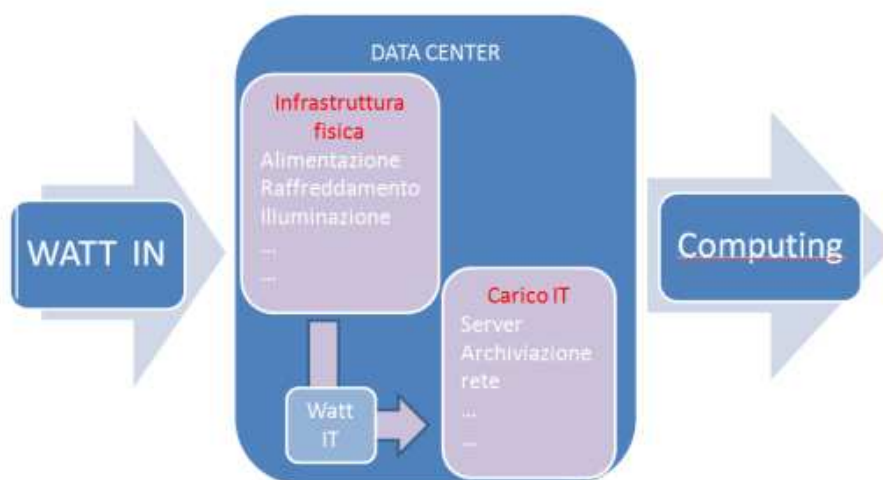


Figura 8: Distribuzione di potenza

Fonte: ENEA, Definizione di algoritmi e indicatori per l'efficientamento dei centri di elaborazione dati (CED), www.enea.it, (2011)

Un valore PUE di 2,0 è abbastanza tipico per un data center. Ciò significa che per ogni watt necessario per alimentare un server, in realtà consumiamo 2 watt di potenza. È importante ricordare che stiamo pagando per l'energia che entra nel data center, quindi ogni watt rappresenta un costo aggiuntivo. Ridurre questo sovraccarico ridurrà i nostri costi operativi complessivi per il data center.

Volendo migliorare l'efficienza energetica del data center, ci sono due aree in cui è possibile influenzare il cambiamento. Riuscendo a ridurre la potenza che va all'infrastruttura di supporto o riducendo le perdite nel sistema di alimentazione, una maggiore quantità di energia che entra nel data center arriverà al carico IT. Ciò migliorerà la nostra efficienza energetica e ridurrà il nostro PUE. Per eseguire misurazioni corrette, un collettivo di aziende americane, operatrici nel campo energetico, hanno proposto linee guida, indicando misure e convenzioni, per ricavare l'efficienza energetica nei data center. Le linee guida da seguire sono:

- Il PUE è il parametro preferito per valutare l'efficienza energetica di un data center.
- Il consumo di energia deve essere misurato almeno all'uscita dell'UPS per determinare il PUE. Le aziende dovrebbero misurare direttamente dai server man mano che le capacità di misurazione si espandono.
- In un data center dedicato, tutte le fonti energetiche disponibili per il proprietario o per l'operatore del data center devono essere incluse nell'energia totale nell'equazione PUE.
- Negli edifici a uso misto con data center, deve essere preso in considerazione anche il consumo totale di raffreddamento, illuminazione e infrastrutture necessarie per le operazioni del data center.

Queste raccomandazioni hanno lo scopo di aiutare gli operatori dei data center ad avere un metro comune per comprendere l'efficienza energetica, che può innescare una conversazione volta a promuovere l'efficienza e a ridurre il consumo di energia, rendendo possibile confrontare il consumo tra diversi data center determinando quale utilizza il migliori tecniche di risparmio. Inoltre, il PUE è stato suddiviso in 4 categorie, che vanno da 0 a 3, a seconda della precisione con cui vengono effettuate le misurazioni, dal più basso al più alto. Per ogni categoria è stabilito il metodo per ottenere i valori da inserire nella formula PUE.

- Nel calcolo del PUE, la categoria 0 ha il livello di precisione più basso. L'Energia Totale viene calcolata utilizzando la lettura del contatore generale, che corrisponde al consumo riportato sulla bolletta energetica, mentre l'Energia IT viene calcolata utilizzando la lettura dell'uscita dell'UPS del consumo energetico annuo. Anche se si perdono completamente gli effetti delle fluttuazioni di potenza dipendenti dal carico di lavoro o delle perdite di sistema basate solo su una lettura istantanea, è comunque un indicatore utile dell'efficienza energetica del data center. Mentre nelle altre categorie l'Energia Totale è calcolata sommando il consumo di altre fonti energetiche, questa categoria non include altre fonti energetiche utilizzate dal data center.
- La misurazione della categoria 1 del PUE è più precisa. A differenza della categoria precedente, l'energia IT viene calcolata utilizzando il totale delle letture, del consumo mensile e dei contatori installati presso la presa dell'UPS e altri punti di controllo in tutto il sistema elettrico. Poiché è possibile misurare anche le fluttuazioni in funzione del carico di lavoro dei server e delle perdite del sistema, l'efficienza indicata dal PUE in questo caso è più accurata.
- La categoria 2 rispetto alla categoria 1 aggiunge precisione, escludendo dai consumi IT l'impatto delle perdite dovute ai trasformatori, misurando l'Energia IT subito dopo il trasformatore collegato alle apparecchiature IT.

- La categoria 3 riesce a mettere in relazione i consumi IT esclusivamente con i consumi generali, rilevando l'efficienza energetica del data center con il massimo grado di accuratezza. L'Energia IT di questa categoria è calcolata sommando i consumi mensili rilevati nei punti di connessione del sistema elettrico con le apparecchiature IT, eliminando tutti i consumi originati dalle apparecchiature non IT[21].

Gli operatori dei data center sono tenuti a includere la categoria PUE quando pubblicano i report sui consumi per rendere i dati più facili da comprendere e confrontare in modo trasparente con altre installazioni. I data center che utilizzano varie fonti di energia devono soppesare il loro consumo durante il calcolo del PUE per ottenere un quadro più accurato dell'efficienza del loro sistema.

3.2 Consumi ed impatto ambientale

Possiamo inviare messaggi, condividere foto, scaricare musica e guardare video istantaneamente grazie a Internet, ma i nostri comportamenti online hanno un impatto ambientale sorprendente. Come risultato dell'energia utilizzata per far funzionare i dispositivi e alimentare le reti wireless a cui accediamo, viene rilasciata una piccola quantità di anidride carbonica per ciascuna di queste attività online. Anche se solo una piccola quantità di energia viene utilizzata per ogni ricerca su Internet o e-mail, Internet è ora utilizzata da 4,1 miliardi di persone, pari al 53,6% della popolazione mondiale. Con ogni attività online, l'energia viene sprecata e vengono rilasciati gas serra associati. Secondo alcune stime, l'impronta di carbonio della nostra tecnologia, di Internet e dei sistemi che la supportano, contribuisce a circa il 3,7% delle emissioni mondiali di gas serra. Secondo Mike Hazas, ricercatore presso la Lancaster University, è paragonabile alla quantità prodotta dall'industria aerea globale. E si prevede che entro il 2025 queste emissioni raddoppieranno. Se dovessimo dividere in modo piuttosto approssimativo gli 1,7 miliardi di tonnellate di emissioni di gas serra che si stima vengano prodotte nella produzione e nel funzionamento delle tecnologie digitali tra tutti gli utenti di Internet in tutto il mondo, significa che ognuno di noi è responsabile della produzione di 414 kg di anidride carbonica all'anno. Meno ovvi, ma forse ancora più energivori, sono i data center necessari per supportare Internet e archiviare i contenuti a cui accediamo tramite esso. Essendo infrastrutture informatiche globalmente diffuse, con un ruolo rilevante all'interno della industria IT, esse hanno un importante impatto ambientale in termini di consumo di energia. Sono infrastrutture che funzionano 24 ore su 24, 7 giorni su 7, tutto l'anno, e sono molto intensivi dal punto di vista energetico con densità di potenza tipiche di 538-2153 W/m² che a volte possono arrivare fino a 10 KW/m². Alcuni dei più grandi data center del mondo possono contenere decine di migliaia

di dispositivi IT e richiedono più di 100 megawatt (MW) di capacità di potenza, sufficienti per alimentare circa 80.000 famiglie americane. Secondo l'Agenzia Internazionale dell'Energia il consumo di elettricità dei data center nel mondo raggiunge all'incirca 220-320 TWh/anno, pari a circa l'1% del consumo totale : in paragone, nel 2020 il fabbisogno annuale di energia elettrica in Italia è stato pari a 301,2 TWh. Prima del conflitto Russo-Ucraino che ha sconvolto il mercato dell'energia e le economie mondiali, il mercato dei Data Center si stava già muovendo verso un percorso di efficientamento energetico, motivato da una crescente spinta verso una linea di sviluppo economico sostenibile. Infatti, nonostante l'impatto notevole del loro consumo energetico globale, tra il 2010 e il 2018, il consumo energetico dei data center è aumentato solo del 6%, a fronte di un aumento delle istanze di calcolo del 550% [14].

Il consumo energetico dei data center può essere suddiviso in due: l'utilizzo di energia ad opera delle apparecchiature IT (server, networks ecc) e l'utilizzo di energia ad opera dell'infrastruttura dell'impianto, quali il sistema di raffreddamento e condizionamento dell'energia. Sia l'efficienza dell'infrastruttura che il design del data center hanno un impatto sulla quantità di energia utilizzata da queste due parti.

La figura sottostante mostra le percentuali di consumo energetico per ciascun componente del data center. In particolare, l'infrastruttura utilizzata per il raffreddamento rappresenta il 50% di tutta l'elettricità consumata, con server e dispositivi di archiviazione che arrivano al 26%.

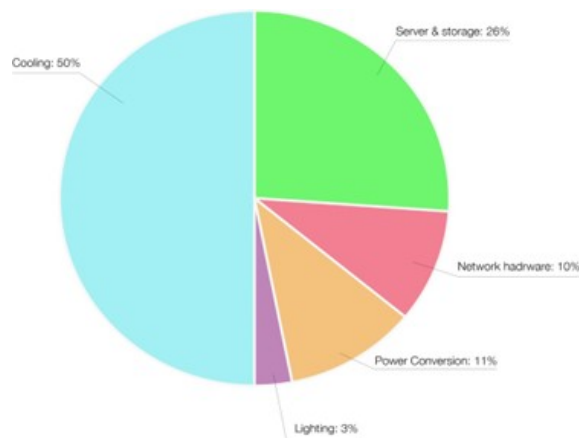


Figura 9: Ripartizione consumo di energia tipico di un data center
Fonte: The Shift Project, lean-ict towards digital sobriety, (2019)

3.2.1 Focus sull'Unione Europea

Il consumo energetico dei data center nell'UE, in una delle ultime stime pubblicate in un rapporto dalla Commissione europea [19], è aumentato da 53,9 TWh/a a 76,8 TWh/a tra il 2010 e il 2018, ovvero il consumo dei data center è cresciuto di quasi il 42%. Ciò significa che nel 2018 i data center rappresentavano il 2,8% di tutta la domanda di elettricità nell'UE. Per quanto riguarda l'impatto ambientale, il rapporto ha notato che, nonostante l'assenza di dati precisi, i data center potrebbero emettere tra lo 0,4% e lo 0,6% del totale dei gas serra generati nell'Ue. La digitalizzazione in corso e soprattutto la crescente disponibilità di servizi cloud portano a una crescita significativa delle capacità dei data center. Questa crescita è così forte che ha più che compensato i significativi guadagni di efficienza realizzati a tutti i livelli (hardware, software, infrastruttura del data center) e il consumo di energia totale dei data center in Europa è aumentato. La questione del futuro sviluppo del consumo energetico dei data center ha attirato una crescente attenzione del pubblico, soprattutto dal 2018. Oltre a riviste specializzate e media, l'argomento è ora sempre più ripreso dalle principali testate come Forbes, Guardian e Frankfurter Allgemeine. Un articolo sulla rivista scientifica Nature dal titolo "Come impedire ai data center di assorbire l'elettricità mondiale" è stato sicuramente un fattore scatenante per l'aumento dell'interesse dei media. L'articolo faceva riferimento a uno studio del 2015 di Huawei dei ricercatori Andrae e Edler in cui hanno previsto che il consumo di energia dei data center aumenterebbero drasticamente entro il 2030. Nello scenario previsto, l'energia consumata dai data center aumenterebbe da circa 200 terawattora (TWh) per anno nel 2010 a quasi 3.000 TWh/a nel 2030. In tal caso, i data center rappresenterebbero circa l'8% del consumo globale di elettricità entro il 2030. Anche altri studi hanno previsto un forte aumento del consumo energetico dei data center in tutto il mondo. Il think tank no-profit francese "Shift Project" prevedeva un consumo di energia dei data center nel mondo da 323 TWh/a nel 2013 a 1.918 TWh/a nel 2025. Per il 2018 lo studio ha calcolato un consumo energetico di circa 700 TWh/a.

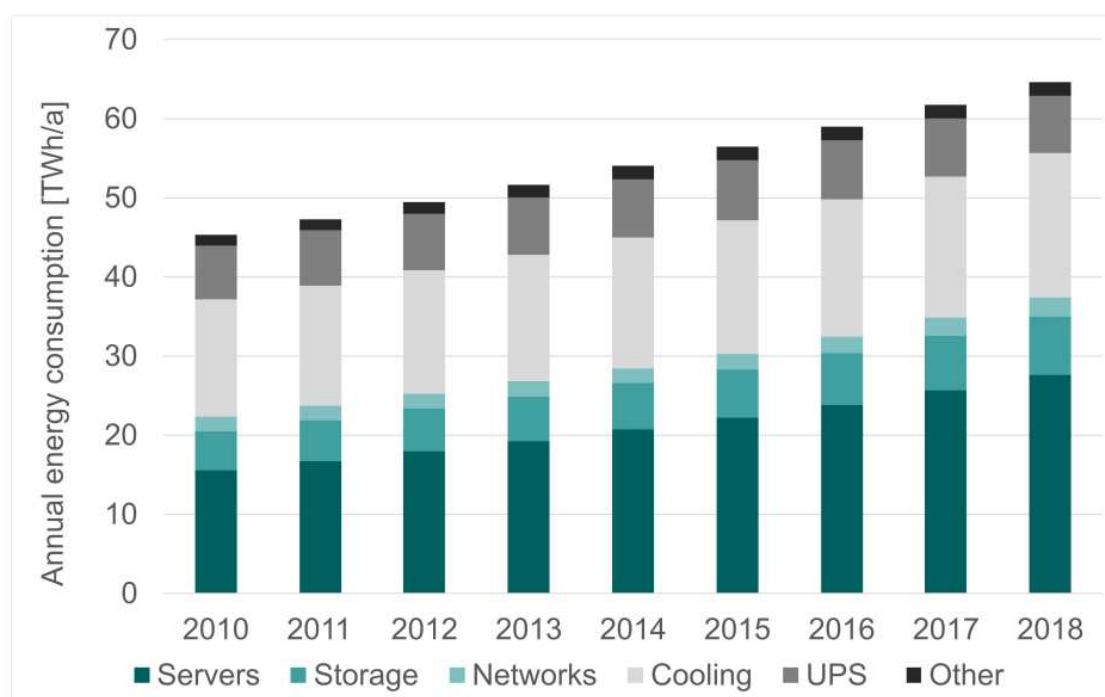


Figura 10: Evoluzione della domanda di energia dei data center nell'UE a partire dal 2010 al 2018.

Fonte: Montevecchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)

La Figura 10 distingue nel dettaglio il consumo energetico dei componenti IT (server, storage, rete) e l'infrastruttura (raffreddamento, UPS, altro). Il consumo di energia delle componenti informatiche sono aumentate del 65% passando da 26,5 TWh/a a 43,8 TWh/a nel periodo in esame. Al contrario, il consumo energetico delle infrastrutture è aumentato solo del 20%, da 27,3 a 33,0 TWh/a. L'efficienza delle infrastrutture è quindi migliorata. Questo sviluppo si riflette in un miglioramento del valore PUE medio nell'UE da 2,03 nel 2010 a 1,75 nel 2018.

A livello regionale, la maggior parte della capacità del data center si trova in Europa settentrionale e occidentale. Queste regioni erano responsabili del 82% del consumo energetico dei data center nel 2018. Entro l'anno 2025 questa percentuale salirà all'87%. Soprattutto per l'energia consumata dai data center nel Nord Europa in forte aumento del 48%, da 26,3 a 38,9 TWh/a previsto per il periodo 2018-2025. Il mercato dei data center in Europa è particolarmente concentrato in Germania, Regno Unito, Francia e Paesi Bassi. Insieme, i data center in questi paesi, rappresentavano il 56% del consumo energetico dei data center nell'UE nel 2018.

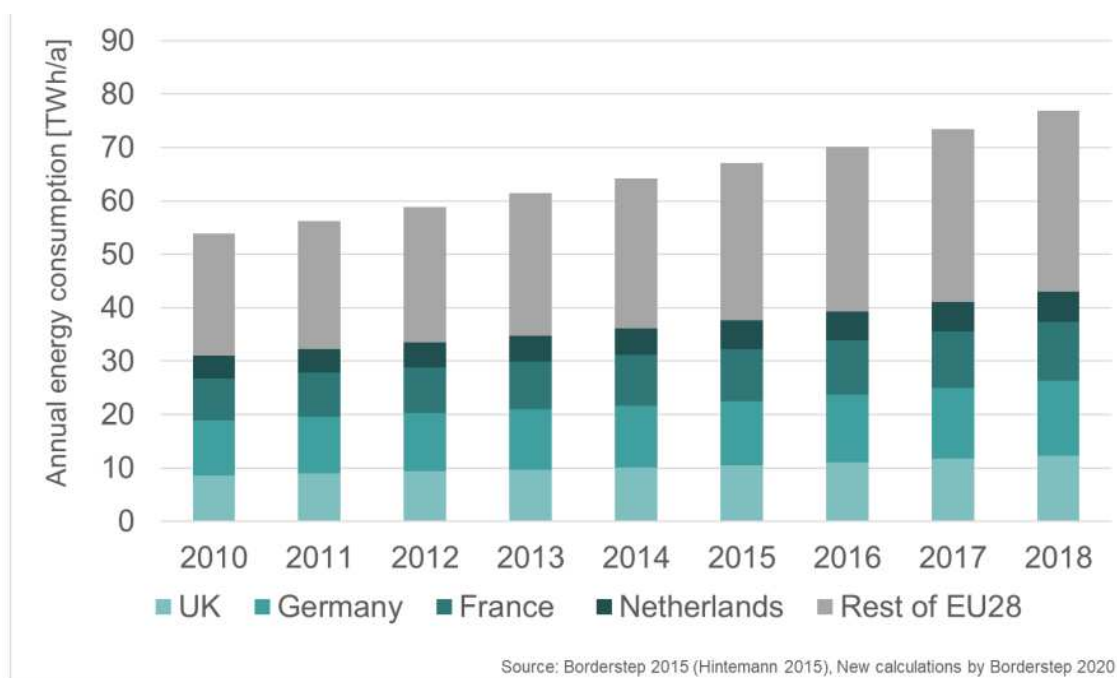


Figura 11: Distribuzione della domanda di energia del data center tra questi quattro paesi e il resto dell'UE28.

Fonte: Montevicchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)

La gamma di possibili sviluppi futuri del consumo energetico di data center in Europa è ampia. Si può presumere che la digitalizzazione dell'economia e la società nel suo insieme porteranno ad un ulteriore aumento dei consumi da parte dei data center. Per illustrare i possibili sviluppi, quattro scenari sono stati creati; Nel caso di tendenza, i data center in Europa consumeranno 98,5 TWh/a nel 2030. Tuttavia, sfruttando tutte le potenzialità tecniche, sarà persino possibile ridurre il loro consumo di energia al livello del 2010. In seguito, una breve descrizione per ogni scenario. Scenario di tendenza: lo scenario di tendenza descrive lo sviluppo delle vendite di hardware e gli aumenti di efficienza dello stesso e nell'infrastruttura del data center. Si presume che nel 2025 il 20% della capacità totale disponibile nei data center verrà installata nei data center edge. In questo scenario, il consumo energetico dei data center aumenterà a 98,5 TWh/a entro il 2030.

Espansione dei data center Edge: per questo scenario, si presume che la tendenza sia verso l'edge; l'informatica sarà molto più pronunciata rispetto allo scenario di tendenza ed in aggiunta verranno costruiti un considerevole numero di data center edge. Per l'anno 2030, si presume che i data center edge rappresentano il 40% della capacità totale dei server. Poiché è probabile un aumento dell'edge computing, accompagnato da un aumento significativo della quantità totale di dati da trattare, il consumo di energia calcolato in questo scenario è superiore rispetto allo scenario di tendenza. Un tale sviluppo non è improbabile. Con l'espansione delle reti mobili 5G e gli sviluppi hightech come l'industria 4.0 e la guida autonoma, la domanda di piccoli centri di edge computing decentralizzati potrebbero vedere un enorme aumento. In questo caso, il consumo energetico di tutti i data center nell'UE potrebbe salire a circa 120 TWh/a. Se i data center edge sono costruiti in modo efficiente vengono utilizzati principalmente per sostituire l'elaborazione dei dati su larga scala dei data center cloud; un forte aumento dei data center edge potrebbe essere possibile anche senza un aumento del consumo energetico dei data center in generale

Efficienza: nello scenario dell'efficienza si assume un espansione dei data center leggermente inferiore. Il tasso di crescita delle apparecchiature hardware nei data center dovrebbe essere 10% in meno rispetto allo scenario di tendenza. Il potenziale tecnologico per aumentare l'efficienza delle infrastrutture del data center, dell'hardware IT e soprattutto la gestione delle architetture sia stato in gran parte esaurito. Vengono utilizzati data center edge principalmente per migliorare l'efficienza dell'intera infrastruttura ICT. Una volta che tutto il potenziale di efficienza sarà stato sfruttato, sarà possibile che il consumo di energia dei data center in Europa diminuisca tornando al livello del 2010 (52 TWh/a). Tuttavia, tale sviluppo è improbabile che si verifichi a meno che non sia massicciamente promosso dal governo attraverso

incentivi e regolamenti.

Caso peggiore: nello scenario peggiore si presume che la crescita attuale accelererà ancora di più e che le capacità dei data center nell'UE saranno ampliate più che nello scenario di tendenza. Si presume che il tasso di crescita delle apparecchiature hardware nei data center sia pari a 10% in più rispetto allo scenario di tendenza. Si presume inoltre che ci possa essere una minore efficienza. Un tale sviluppo non è improbabile, poiché ulteriori aumenti di efficienza potrebbero essere più difficili a causa dei limiti alla miniaturizzazione della tecnologia dei semiconduttori che è comunemente usata oggi ed è una tendenza che è stata indicata come la fine della legge di Moore. Nello scenario peggiore, è possibile un aumento del consumo energetico del data center fino a 160 TWh/a nel 2030.

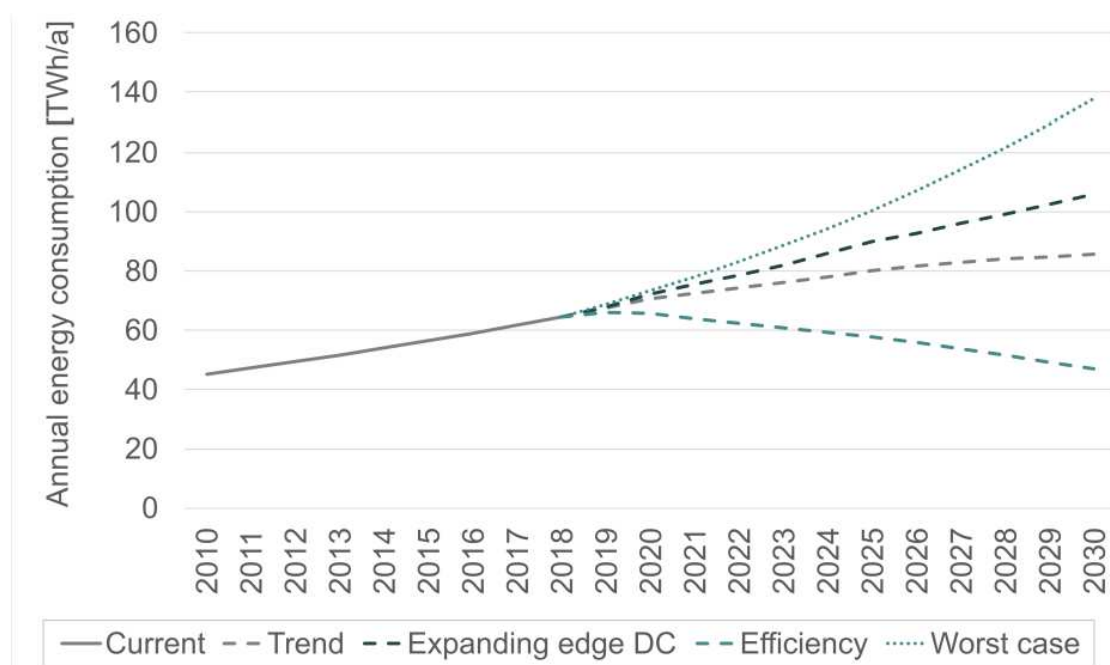


Figura 12: Consumo energetico del data center come percentuale della domanda dell'energia elettrica finale nell'UE28 per gli anni dal 2010 al 2030.

Fonte: Montevocchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)

Viene evidenziato che il consumo energetico dei data center nell'UE è in costante aumento da anni. Sebbene il cloud computing stesso sia in molti casi un modo efficiente di fornire risorse IT, la forte crescita dei data center cloud è una delle principali ragioni di questo aumento complessivo. L'utilizzo conveniente e flessibile porta ad un aumento significativo della domanda di servizi cloud e quindi a un aumento del consumo energetico dei data center. Infine, il notevole uso di energia elettrica dei data center suscita anche preoccupazioni per le loro emissioni di anidride carbonica. Purtroppo, non è ancora possibile stimare con precisione le emissioni totali di CO₂, a causa della mancanza di dati sulla localizzazione della maggior parte dei data center globali e sull'intensità delle emissioni (misurata in grammi di CO₂ per chilowattora) delle loro effettive fonti di elettricità. Solo poche aziende, tra cui Google, Apple, Switch e Facebook, riportano pubblicamente tali dati, indicando una tendenza crescente tra alcuni dei più grandi operatori di data center del mondo verso l'approvvigionamento di energia rinnovabile.

4 Tecniche risparmio energetico

Nel capitolo precedente, è stato esposto come misurare i consumi di un data center e come col passare del tempo, a causa del traffico dati sempre più intenso e di un utilizzo sempre più massivo di servizi di rete, tali consumi incrementino anno dopo anno. I consumi di un datacenter si possono scomporre in tre aree principali; in ordine crescente alimentazione, raffreddamento ed IT equipment. Il lavoro utile è rappresentato dall'elaborazione dati nei dispositivi IT, il cui consumo però è circa la metà della totale energia richiesta, che deve coprire anche le esigenze di raffreddamento e accessorie (antincendio, antintrusione, etc). La maggior parte dell'energia utilizzata dalle macchine elettriche e dalle apparecchiature elettroniche viene convertita in calore da disperdere; di conseguenza, migliorandone l'efficienza, oltre ad abbassarne i consumi diretti, si ridurrà anche la necessità di impianti di raffreddamento (circa un terzo del totale).

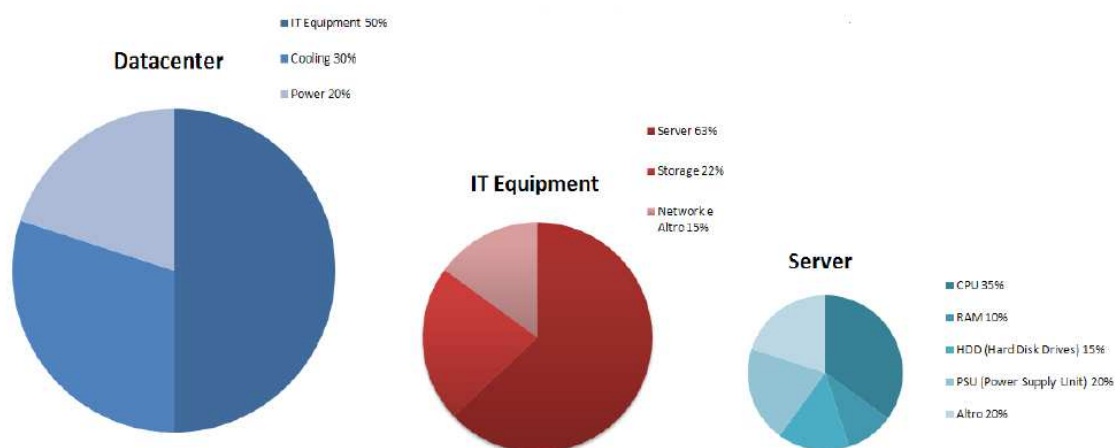


Figura 13: Percentuali tipiche di consumo in un data center dei vari componenti
Fonte: ENEA, Definizione di algoritmi e indicatori per l'efficientamento dei centri di elaborazione dati (CED), www.enea.it, (2011)

Esistono molti modi per ottimizzare i data center. Si può iniziare con una progettazione precisa dell'intero sistema e della struttura per eliminare le perdite energetiche fisse, oppure, dato che i server spesso non sono utilizzati al massimo della loro capacità operativa, si possono mettere in atto meccanismi o strategie per aumentare l'efficienza delle apparecchiature informatiche riducendo il numero di elementi operativi e sfruttandoli al massimo delle loro potenzialità. Risulta inoltre proficuo, in termini di risparmio energetico, adottare particolari accorgimenti in merito alla progettazione e all'aggiornamento degli impianti ausiliari; in primis su tutti, l'impianto di alimentazione elettrica e l'impianto di raffreddamento in quanto contribuiscono

con rilevanza ai consumi dell'intero data center, inoltre se adeguatamente progettati permettono di rallentare drasticamente il deterioramento dell'IT Equipment. Sono diverse le tecniche e le strategie che negli anni sono state elaborate e sperimentate allo scopo di ridurre più o meno efficacemente i consumi dei data center e in questo capitolo ne verranno illustrate alcune tra le più utilizzate legate all'aspetto elettronico e quindi all'IT Equipment, aventi risultati migliori nell'aumentare l'efficienza energetica dei data center.

4.1 La progettazione

La progettazione di un data center efficiente dal punto di vista energetico è complessa e costosa, ma se eseguita correttamente offre notevoli vantaggi in termini di gestione. Un centro informatico è costituito da una varietà di apparecchiature, tutte scelte in base alla struttura complessiva piuttosto che individualmente. Sia i data center dedicati che i centri di calcolo possono essere integrati in altre strutture (questo è lo scenario più tipico). È fondamentale valutare accuratamente i flussi interni di energia termica ed elettrica in entrambe le situazioni. I tre fattori principali da tenere in considerazione nella progettazione di un datacenter sono:

l'affidabilità, caratteristica che viene garantita dalla ridondanza delle apparecchiature;

la scalabilità, aspetto assicurato dall'uso di componenti modulari, in grado di permettere l'adattamento a situazioni che possono mutare nel tempo e di evitare inutili sovradimensionamenti;

la scelta di macchine e strumentazioni ad alta efficienza [5].

Tali caratteristiche devono essere prese in considerazione nella selezione di tutti i componenti, che si tratti di IT Equipment, di componenti della catena di alimentazione, o di parti dell'impianto di condizionamento. Queste scelte hanno un impatto significativo sull'efficienza energetica del data center, quindi è fondamentale dimensionare il sistema in modo appropriato e scegliere apparecchiature ad alta efficienza energetica per avere un consumo energetico inferiore, anche per i componenti non IT, al fine di ridurre al minimo le perdite nell'impianto.

4.1.1 Dimensionamento

La tecnica che incide maggiormente sull'efficienza finale del data center è quella di dimensionare correttamente il sistema in base al carico di lavoro. Spesso, in fase di realizzazione di un impianto, non si tiene conto che all'interno dell'impianto ci sono dispersioni fisse di energia, come quelle causate dall'impianto di raffreddamento, che deve essere sempre in funzione per mantenere la temperatura ideale dell'ambiente, anche quando non ci sono server in uso. Le perdite infrastrutturali sono inversamente correlate alla potenza del sistema e possono rappresentare una parte considerevole del consumo complessivo della rete utilizzata per raffreddare e distribuire l'elettricità alle apparecchiature IT. È semplice che queste perdite superino il consumo IT in piccoli impianti con carichi di lavoro leggeri; d'altra parte, in impianti con carichi di lavoro elevati e impianti sovradimensionati, le perdite aumenteranno sensibilmente la bolletta energetica. Un data center adeguatamente dimensionato può ridurre i consumi di quasi il 50% con costi minimi [16]. Ragion per cui risulta di fondamentale importanza stimare attentamente la dimensione della struttura e il numero di componenti necessari al suo interno, calcolando la giusta potenza che il data center dovrà avere in base alle funzioni da svolgere, individuando il numero di computer e server necessari.

4.1.2 Data Center modulari

Pensare al consumo dei singoli componenti senza considerare una progettazione più accurata dell'intero sistema è uno degli errori più frequenti commessi durante la fase di progettazione del data center. Confrontando due diversi data center con design diversi ma con le stesse apparecchiature e potenza installata sarà possibile osservare due differenti livelli di consumo energetico a favore del data center con una progettazione più attenta dell'intero sistema. Ci sono stati principalmente due approcci utilizzati al fine di ottenere una progettazione efficiente.

Il primo passo è progettare ogni componente del sistema specificamente per il data center in cui verrà installato, conducendo test mirati per prevenire i problemi sopra menzionati e posizionando ogni componente nel posto giusto. Il secondo è quello di realizzare l'impianto utilizzando moduli precostruiti che sono già stati sottoposti a collaudo da parte dei costruttori al fine di prevenire problemi di questa natura, anche con le singole parti dell'impianto di climatizzazione e di distribuzione dell'energia. A causa dell'ovvio costo e dei vantaggi pratici la seconda opzione, è ora diventata la norma per la costruzione di data center.

Le soluzioni modulari prendono le migliori idee per design, affidabilità ed efficienza e le impacchettano in un modulo prefabbricato, ripetibile e ottimizzato dal punto di

vista operativo. La struttura di un data center modulare è costruita con una serie di moduli di costruzione aventi diversi tipi di funzionalità col fine di formare un'intera struttura data center avente un design architettonico pre-approvato e standardizzato. Tutte le istanze di un particolare tipo di modulo avranno all'incirca la stessa planimetria e lo stesso progetto architettonico. Una serie iniziale di moduli può essere costruita su un lotto di terreno, ed in seguito, man mano che aumentano le esigenze di spazio e di capacità della struttura del datacenter modulare, è possibile aggiungere rapidamente ulteriori moduli di diversi tipi al set iniziale di moduli di costruzione [13]. Esistono in genere due diverse tipologie di forma per i data center modulari. Il tipo più diffuso, noto anche come data center containerizzato o data center edge, comporta l'imballaggio dell'hardware del data center (server, storage e apparecchiature di rete) in un normale container di spedizione, che viene quindi spedito nella posizione desiderata. La maggior parte dei data center containerizzati dispone di sistemi di raffreddamento indipendenti. Un altro tipo di data center modulare incorpora l'hardware del data center in un edificio costituito da parti prefabbricate che possono essere rapidamente assemblate in loco ed espanse al variare dei requisiti di capacità. Ad esempio, la versione HP di questo tipo di data center modulare è costituita da componenti in lamiera che collegano quattro sale del data center ad un edificio operativo centrale. Grazie alla rapida evoluzione delle tecnologie ci sono molti motivi e vantaggi per cui conviene scegliere un approccio modulare nella realizzazione di data center.

- Velocità di distribuzione: I tempi dall'ordine all'implementazione per le soluzioni modulari sono incredibilmente brevi. Si tratta di un prodotto pronto all'uso che può essere ordinato, personalizzato e consegnato alla sede del data center in pochi mesi(o meno). E' una soluzione personalizzabile in grado di essere rapidamente implementata all'interno di un ambiente.

- Scalabilità: È semplice ridimensionare rapidamente l'infrastruttura per soddisfare la domanda quando si dispone di un design ripetibile e standardizzato. L'infrastruttura di supporto presso il sito del data center e il terreno disponibile sono gli unici vincoli di scalabilità per un data center modulare. La flessibilità che offre, avendo moduli che possono essere rapidamente sostituiti quando diventano obsoleti o quando è richiesta una tecnologia più recente, è un altro vantaggio della scalabilità. Ciò significa che le aziende possono pianificare con anticipo di alcuni mesi i cambiamenti tecnologici. Di conseguenza, non ci vuole molto tempo per pianificare una soluzione di data center cloud.

-Agilità: Non basta la semplice scalabilità per creare rapidamente un ambiente di data center. Essere flessibili con le piattaforme di data center implica essere in gra-

do di reagire rapidamente alle mutevoli esigenze aziendali. I data center modulari sono creati appositamente per supportare l'agilità aziendale e dell'infrastruttura, sia che ciò significhi aggiungere un nuovo servizio o ridurre i tempi di inattività. Altre organizzazioni utilizzano data center modulari per le loro operazioni di ripristino di emergenza altamente efficaci, mentre alcune organizzazioni costruiscono i propri ambienti modulari per scopi di pianificazione della capacità.

- *Mobilità e posizionamento:* Un data center modulare può essere installato ovunque desideri l'utente finale. In quanto metodo di trasporto internazionale che ha ricevuto l'approvazione ISO, un container può vantare la massima mobilità. Una soluzione modulare può essere trasportata in pezzi e assemblata rapidamente sul posto, rendendola mobile. Quando si cercano soluzioni di ripristino di emergenza modulari, la mobilità è una qualità desiderabile perché può essere rapidamente implementata nel sito di ripristino e messa in funzione. La sfida per i fornitori di data center sarà rimanere il più agili possibile mentre cercano di accettare nuove offerte. Per aiutare a soddisfare le crescenti esigenze di capacità, ciò potrebbe benissimo comportare la costruzione di data center più modulari.

- *Densità e PUE:* Lo spazio viene utilizzato in modo molto efficace in una soluzione modulare e ha densità fino a 20 kilowatt per armadio. Poiché il modulo è pre-progettato e standardizzato, il PUE può essere calcolato al momento della messa in servizio ed è compreso tra 1,1 e 1,4.

- *Efficienza:* Il fatto che i moduli siano prodotti ingegnerizzati significa che i sottosistemi interni sono strettamente integrati, il che si traduce in guadagni di efficienza in termini di alimentazione e raffreddamento nel modulo. Esistono opportunità per utilizzare piattaforme di gestione dell'energia all'interno dei moduli, con tutti i sottosistemi progettati nel loro insieme.

- *Messa in produzione:* In quanto soluzione ingegnerizzata e standardizzata, il modulo del data center può essere commissionato dove è costruito e richiede meno passaggi da eseguire una volta collocato nel sito del data center

- *Standardizzazione:* La soluzione modulare è un metodo standardizzato per costruire un data center, proprio come fece Henry Ford quando costruiva automobili, ed è vista come parte dell'industrializzazione dei data center. I moduli del data center fabbricati vengono costruiti in un sito diverso dal sito del data center utilizzando un set di modelli di componenti. Le procedure operative standard possono essere utilizzate su tutta la linea grazie all'infrastruttura standardizzata dei moduli. Le procedure operative sono identiche perché il modulo è prefabbricato e possono essere impacchettate con la soluzione modulare per fornire una documentazione standardizzata per i sottosistemi all'interno del modulo.

-Ridimensionamento: Il design modulare consente infine un approccio di distribuzione ottimizzato per soddisfare le esigenze IT. Questa capacità di dimensionare correttamente l'infrastruttura in base alla crescita delle esigenze IT consente l'allineamento dell'azienda con le strategie IT e del data center. Il modulo o il contenitore può anche fornire capacità quando necessario rapidamente per progetti o adeguamenti temporanei della capacità. Questo è importante perché le risorse sono costose. I data center modulari possono aiutare a dimensionare correttamente le soluzioni in modo che le risorse siano utilizzate in modo ottimale. Il provisioning eccessivo o insufficiente delle risorse del data center può essere estremamente costoso e difficile da correggere [4].

Approcci di data center modulari si sono evoluti in parallelo con il cambiamento nell' IT Equipment.

4.2 Tecniche per l'IT Equipment

Le apparecchiature IT sono l'utilizzatore finale che svolge il lavoro nel data center. La loro funzionalità viene garantita dalla presenza di un adeguato sistema di alimentazione e condizionamento. Mentre il primo deve garantire continuità e qualità elettriche, il secondo deve dissipare il calore prodotto dagli altri componenti, evitando surriscaldamenti che porterebbero al danneggiamento e al conseguente fermo dei sistemi, con i relativi danni per il blocco delle attività. La scelta di server efficienti permette di ottenere vantaggi indiretti quali una minore produzione di calore che dovrà essere smaltito e una minore necessità di potenza elettrica di cui garantire qualità e continuità.

Nonostante l'elevato consumo, i server raramente sono completamente inattivi o funzionano alla massima capacità. Al contrario, vengono generalmente utilizzati a meno del 50% della loro capacità. La quantità di energia utilizzata dai server varia a seconda del loro livello di utilizzo e delle attività che devono completare. I server non sono utilizzati al 100%, per mantenere un certo margine di capacità e riuscire ad accettare tutte le richieste con un tempo di risposta accettabile, questo a causa delle fluttuazioni del traffico che influenzano le richieste di servizio, considerando inoltre le eventuali manutenzioni o l'esecuzione di funzioni più complesse. Senza questa flessibilità si rischierebbero frequenti interruzioni di servizio. Non è nemmeno possibile lasciare i sistemi completamente inattivi anche in caso di poche operazioni da svolgere, perché in molti casi i vari servizi, specialmente quelli su larga scala, richiedono l'utilizzo di diversi server con il peso del lavoro suddiviso tra di loro, con anche i dati necessari alle varie funzioni, distribuiti tra i server impedendo così la disattivazione di anche una piccola parte dei server inutilizzati in quel momento.

Al fine di risolvere questo problema, è importante che i progettisti di componenti e sistemi costruiscano server che consumino energia in modo proporzionale al livello di utilizzo. Si dovrebbe puntare a costruire server capaci di non consumare energia nei momenti di inattività e che aumentino progressivamente i consumi all'aumentare delle operazioni da svolgere. In seguito vengono indicate pratiche e tecnologie informatiche efficienti, da tenere in considerazione nella progettazione di un nuovo data center o nella ristrutturazione di uno esistente, per quanto riguarda l'IT Equipment.

4.2.1 Blade Server

È un tipo specifico di server con molti componenti dei sistemi informatici (CPU, storage, ecc.) che condividono i dispositivi ausiliari (alimentatori e ventole). Sia i nuovi rack che la configurazione tradizionale possono essere sostituiti con questo tipo di server. Ad esempio, 14 blade occupano lo spazio di 7 server tradizionali. Di conseguenza, questa soluzione consente la sostituzione di alcune unità di tipo classico con unità di tipo blade, che possono essere considerate componenti ad alta densità con elevata potenza computazionale. Tuttavia, concentrare più potenza di calcolo in un'area più piccola comporterà un aumento significativo del calore per unità di superficie, fino a 28 kW per rack, per il quale è necessario installare sistemi di raffreddamento adeguati. A parità di potenza di calcolo, questo tipo di server promette una riduzione dei costi complessivi del 10% rispetto alla configurazione convenzionale [5].

4.2.2 Storage

I sistemi di storage a stato solido, che utilizzano SSD (Solid State Drive) piuttosto che i più tradizionali HDD (Hard Disk Drive), possono offrire vantaggi sia dal punto di vista dei consumi che della velocità di selezione dei dati. In realtà, gli SSD consumano meno energia e generano meno calore. Le operazioni di input e output al secondo (IOPS, operazioni di input/output al secondo) vengono utilizzate per misurare la velocità con cui un sistema di archiviazione elabora i dati di selezione. Con gli HDD, sono possibili 300 IOPS a velocità di rotazione di 15.000 rpm, mentre con gli SSD, il massimo IOPS per unità è 30.000. Sono necessarie 100 unità HDD per immettere/esportare i dati alla stessa velocità di un SSD. Sun/Oracle stima che il rapporto IOPS/Watt per gli SSD sia di circa 10.000, rispetto a 20 per gli HDD. (Figura 2) Il consumo degli HDD è circa 500 volte maggiore a parità di IOPS. (10.000/20). La diffusione degli SSD è frenata dai costi ancora molto elevati rispetto ai tradizionali HDD, differenza che dovrebbe assottigliarsi in futuro [5].

4.2.3 Virtualizzazione

La virtualizzazione del server è il processo che consente di suddividere un server fisico in più server virtuali univoci e isolati per mezzo di un'applicazione software. Tramite questo processo è quindi possibile installare sistemi operativi su hardware virtuale; l'insieme delle componenti hardware virtuali (Disco fisso, RAM, CPU, Scheda di rete) prende il nome di macchina virtuale e su di esse può essere installato il software come, appunto, i sistemi operativi e relative applicazioni. Quando si confronta un server fisico con un server virtuale, il server virtuale si rivela un modo efficace per risparmiare sui costi dell'hardware fisico. Inoltre, utilizza una quantità relativamente inferiore di energia per ulteriori risparmi sui costi e vantaggi ambientali. La virtualizzazione dei server semplifica la riallocazione delle risorse e l'adattamento ai carichi di lavoro dinamici. Le risorse fisiche vengono ripartite dagli hypervisor in modo che gli ambienti virtuali possano utilizzarle. Esistono tre modi per creare server virtuali: virtualizzazione completa, paravirtualizzazione e virtualizzazione a livello di sistema operativo. Condividono tutti alcuni tratti comuni. Il server fisico è chiamato host. I server virtuali sono chiamati guest. I server virtuali si comportano come macchine fisiche. Ogni sistema utilizza un approccio diverso per allocare le risorse del server fisico alle esigenze del server virtuale.

La virtualizzazione completa utilizza un tipo speciale di software chiamato hypervisor. L'hypervisor interagisce direttamente con la CPU e lo spazio su disco del server fisico funge da piattaforma per i sistemi operativi dei server virtuali. L'hypervisor mantiene ogni server virtuale completamente indipendente e all'oscuro degli altri server virtuali in esecuzione sulla macchina fisica. Ogni server guest viene eseguito sul proprio sistema operativo. L'hypervisor monitora le risorse del server fisico. Mentre i server virtuali eseguono le applicazioni, l'hypervisor inoltra le risorse dalla macchina fisica al server virtuale appropriato. Gli hypervisor hanno le proprie esigenze di elaborazione, il che significa che il server fisico deve riservare alcune risorse e potenza di elaborazione per eseguire l'applicazione hypervisor. Ciò può influire sulle prestazioni complessive del server e rallentare le applicazioni.

L'approccio alla paravirtualizzazione è leggermente diverso. A differenza della tecnica di virtualizzazione completa, i server guest in un sistema di paravirtualizzazione sono consapevoli l'uno dell'altro. Un hypervisor di paravirtualizzazione non ha bisogno di tanta potenza di elaborazione per gestire i sistemi operativi guest, perché ogni sistema operativo è già a conoscenza delle richieste che gli altri sistemi operativi impongono al server fisico. L'intero sistema funziona insieme come un'unità coesa. Un approccio di virtualizzazione a livello di sistema operativo non utilizza affatto un hypervisor. Invece, la capacità di virtualizzazione fa parte del sistema operativo

host, che esegue tutte le funzioni di un hypervisor completamente virtualizzato. La più grande limitazione di questo approccio è che tutti i server guest devono eseguire lo stesso sistema operativo. Ogni server virtuale rimane indipendente da tutti gli altri, ma non è possibile combinare sistemi operativi tra loro. Poiché tutti i sistemi operativi guest devono essere uguali, si parla di ambiente omogeneo [12].

La virtualizzazione dei server aiuta a ridurre i costi IT riducendo la necessità di server fisici e hardware associato, migliora l'utilizzo dei server e consente un uso più efficiente delle risorse, fornisce maggiore flessibilità e scalabilità alle organizzazioni in quanto possono facilmente aggiungere o rimuovere server virtuali secondo necessità. Quando si confronta un server fisico con un server virtuale, il server virtuale si rivela un modo efficace per risparmiare sui costi dell'hardware fisico. Inoltre, utilizza una quantità relativamente inferiore di energia per ulteriori risparmi sui costi e vantaggi ambientali. La conversione di un server fisico in più server virtuali consente alle organizzazioni di utilizzare la potenza di elaborazione e le risorse in modo più efficiente eseguendo più sistemi operativi e applicazioni su un server partizionato. La maggior parte delle organizzazioni utilizza i server virtuali per limitare le spese relative all'hardware dei server e ridurre i costi energetici e di alimentazione. L'evoluzione delle tecnologie software e hardware la hanno portata ad essere uno strumento indispensabile.

4.2.4 Consolidamento

Il consolidamento del data center si riferisce a metodi e tecniche che rendono le architetture IT più efficaci, ciò può comportare la combinazione fisica di diversi data center o semplicemente la gestione di un singolo data center di dimensioni considerevoli in modo più efficiente utilizzando meno risorse. Esistono molti modi in cui le aziende possono consolidare i propri data center e rendere i loro sistemi IT più efficienti. Alcune strategie comuni includono la virtualizzazione di server o di dispositivi di archiviazione, in cui i sistemi di rete fisica vengono sostituiti con sistemi di accesso logico o reti virtuali che utilizzano software per ricoprire il ruolo di dispositivi hardware tradizionali come server e singole macchine. Consolidando più data center fisici in un unico data center virtualizzato, le organizzazioni sono in grado di ridurre l'ingombro dell'infrastruttura IT, con conseguenti risparmi sui costi. Inoltre, le organizzazioni possono beneficiare di una maggiore efficienza, scalabilità e sicurezza. Possono essere utilizzati anche tipi alternativi di server, come i server blade. I nuovi sistemi di cloud hosting possono sostituire i sistemi hardware e software interni e i servizi di fornitori terzi possono aiutare a fornire l'automazione dei processi aziendali o altri miglioramenti che possono ridurre le risorse necessarie per eseguire

un data center aziendale. Le organizzazioni devono considerare una serie di best practices quando implementano il consolidamento dei data center. Queste includono la valutazione della compatibilità delle risorse IT prima della virtualizzazione, la pianificazione dell'implementazione del data center virtualizzato e la garanzia che l'implementazione soddisfi i requisiti IT. Inoltre, le organizzazioni dovrebbero rivedere regolarmente il proprio data center virtualizzato per identificare eventuali aree di miglioramento. Il consolidamento dei data center può anche offrire una serie di vantaggi ambientali. Consolidando più data center fisici in un unico data center virtualizzato, le organizzazioni sono in grado di ridurre il consumo energetico, con conseguente riduzione delle emissioni di carbonio. Inoltre, le organizzazioni possono essere in grado di ridurre la quantità di spazio fisico richiesto per la loro infrastruttura IT, con conseguente riduzione degli sprechi.

4.2.5 L'Intelligenza Artificiale

Nonostante negli ultimi anni sia stato fatto molto per ottimizzare l'infrastruttura (consolidando e virtualizzando server, storage e reti), nonché per gestire in modo più intelligente l'alimentazione e il raffreddamento (attraverso sistemi energetici e di raffreddamento), sfruttare le peculiarità tecnologiche di alcune soluzioni basate sull'intelligenza artificiale potrebbe portare ad ulteriori progressi. L'intelligenza artificiale può essere sfruttata in ambito IT e nella gestione efficiente dei data center. Nel 2018 Google annunciò di aver elaborato un sistema di gestione di data center utilizzando l'intelligenza artificiale con il fine ultimo di ridurre i costi di gestione ed aumentare il risparmio energetico. Il Massachusetts Institute of Technology (MIT) di Boston ha recentemente compiuto un ulteriore passo avanti in questa direzione creando un nuovo sistema che apprende automaticamente come programmare le operazioni di elaborazione dei dati su migliaia di server al fine di massimizzare l'utilizzo delle risorse (comprese quelle necessarie per l'energia e raffreddamento), portando ad una diminuzione dei costi. L'allocazione dinamica delle risorse infrastrutturali per supportare l'elaborazione dei dati nei data center è stata tradizionalmente lasciata agli algoritmi. Tuttavia, fino a poco tempo fa, questi algoritmi erano sviluppati da esseri umani che preferivano politiche e linee guida aziendali che, nella maggior parte dei casi, tenevano poco conto dell'efficienza energetica (se non come controllo sui costi complessivi), privilegiando prestazioni e affidabilità a supporto di processi e servizi. I ricercatori del MIT di Boston hanno creato nuove tecniche di machine learning basate sull'apprendimento per rinforzo incoraggiando il sistema ad apprendere da solo sulla base di regole, prove ed errori in modo da personalizzare le decisioni relative alla pianificazione del carico di lavoro, anche molto complesse,

allocando carichi di lavoro specifici a specifici cluster di server al fine di ottenere l'ottimizzazione del carico e una migliore efficienza energetica. Nell'uso effettivo, il sistema dei ricercatori testa una varietà di possibili schemi di distribuzione per i carichi di lavoro in entrata tra i server prima di stabilire una combinazione di utilizzo delle risorse di calcolo e velocità di elaborazione che sia efficiente ed efficace. In tutto ciò è sufficiente una semplice direttiva che dica al sistema di "minimizzare i tempi di completamento dei lavori"; non è richiesto alcun coinvolgimento umano. Gli scienziati che hanno preso parte al progetto affermano che il sistema sviluppato dal MIT porta a termine lavori dal 20% al 30% più velocemente rispetto ai migliori algoritmi di programmazione scritti dagli esseri umani. Il sistema ha finora capito come comprimere in modo efficiente i carichi di lavoro per ridurre gli sprechi (questo è in realtà da dove viene l'efficienza energetica). I risultati indicano tuttavia uno sbocco aggiuntivo, che potrebbe consentire al data center di gestire lo stesso carico di lavoro più rapidamente e con meno risorse.[18].

Oltre che per ottimizzare i carichi di lavoro l'intelligenza artificiale può essere utilizzata nei data center anche per ottimizzarne il funzionamento per evitare guasti. Questo è uno degli obiettivi del gruppo di ricerca ICT di ENEA che sta testando tecniche innovative sul proprio supercalcolatore, partendo dalla raccolta dei dati sui consumi energetici, termici e computazionali del data center HPC di ENEA per individuare, attraverso strumenti avanzati di intelligenza artificiale, eventuali sovracarichi di lavoro, sprechi, malfunzionamenti o 'falle' nel sistema di raffreddamento. Un surriscaldamento frequente e, di conseguenza, un sistema di raffreddamento poco efficace, potrebbero causare, ad esempio, il degrado dell'hardware, una ridotta affidabilità dei server nonché un dispendioso consumo energetico. Grazie a questa metodologia innovativa è possibile estrarre informazioni utili dai dati reali di consumo dell'infrastruttura di calcolo e elaborare strategie di gestione che non sarebbero ottenibili con le tecniche classiche di data analysis. Nello specifico, con la metodologia ENEA vengono raccolti e messi in relazione i dati sugli effettivi carichi di lavoro del data center e sul consumo energetico a diversi livelli (nodo, server, rack e stanza). Questa mole di informazioni viene poi elaborata attraverso tecniche di apprendimento supervisionato che permettono di definire modelli previsionali in grado di predire il comportamento dei server in base alla temperatura e al carico computazionale e di mappare eventuali fenomeni di surriscaldamento (hotspot o punti caldi). In questo modo, risulta possibile effettuare interventi mirati ed efficaci per garantire il corretto funzionamento delle singole componenti dell'infrastruttura di calcolo, a partire, ad esempio, dall'ottimizzazione del flusso d'aria di raffreddamento, oppure, individuare esattamente gli hotspot termici in modo da consentire al cluster

di lavorare a temperature simili su tutti i nodi di calcolo, evitando rischi legati al surriscaldamento e consentendo di modulare le unità di raffreddamento su potenze più basse. Risulta possibile definire modelli che mappano il comportamento degli utenti in termini di utilizzo e consumo delle risorse e poter predire, sulla base di dati storici, i carichi di lavoro e, quindi, il consumo futuro di processore, memoria e rete. La comunità dei data center potrebbe sfruttare questo approccio per migliorare le condizioni termiche in cui operano le infrastrutture informatiche, con azioni mirate su quei server che più frequentemente vanno incontro a un surriscaldamento, visto che i consumi energetici di un data center dipendono fortemente dalla temperatura del luogo dove si trova l'infrastruttura informatica, così come dalle prestazioni dei sistemi IT e di raffreddamento [20].

5 Conclusioni

In seguito ad una breve introduzione sulle tecnologie digitali che pervadono la nostra quotidianità e al loro impatto rilevante sull' ambiente in termini di consumi energetici e produzione di CO₂, sono stati presi in esame i centri di elaborazione dati ovvero i Data Center. Ne è stata analizzata la struttura, prendendo in esame i suoi componenti principali e ausiliari, la diffusione in termini di evoluzione temporale nel loro utilizzo e le criticità rispetto il consumo di energia di cui le apparecchiature IT sono le maggiori responsabili. Sono state illustrate metodologie per analizzare i consumi e l' impatto ambientale di un data center, definendo la metrica PUE, ed in seguito sono stati discussi i potenziali scenari all' interno dell' Unione europea che potrebbero presentarsi in futuro in base alla loro evoluzione tecnologica e alla loro riduzione dei consumi. In ultima è stato mostrato come attraverso studi ingegneristici, di sviluppo di nuove tecniche e tecnologie, test pratici su supercalcolatori e analisi di dati, è possibile migliorare l' efficienza energetica dei data center riducendo sempre di più il loro impatto sull' ambiente. A tal fine risulta possibile ottenere buoni risultati attraverso una scrupolosa progettazione del sistema, accompagnata dall' utilizzo di componenti efficienti e tecnologie di efficientamento energetico quali virtualizzazione e consolidamento, e in data center più moderni grazie alla fondamentale, ma ancora in fase di sperimentazione, collaborazione con l' intelligenza artificiale. Nei prossimi anni l' industria dei data center si adopererà per accelerare ulteriormente lo sviluppo di queste tecnologie e l' adozione di pratiche sostenibili in quanto limitare il proprio impatto sul cambiamento climatico a livello mondiale rappresenta tuttora una enorme sfida.

Elenco delle figure

1	Contributo del digitale alle emissioni globali Fonte: The Shift Project, lean-ict towards digital sobriety, (2019) . . .	5
2	Classifica produzione MtCO ₂ primi 20 Stati Fonte: www.globalcarbonatlas.org, (2021)	5
3	Divisione KgCO ₂ prodotta dai diversi settori dell'ICT Fonte: Statista, Oko-institut, Germania, (2022)	6
4	Layout data center di Pisa	11
5	Sistema di raffreddamento Fonte: Auslander D., Viability of dynamic cooling control in a data center environment, (2004)	13
6	Componenti principali di un data center Fonte: Ciaramella A. Roveda M., Data Center: localizzazione, caratteristiche e prestazioni delle nuove fabbriche dati, (2018)	14
7	Classificazione Tier dei Data Center Fonte: Ciaramella A. Roveda M., Data Center: localizzazione, caratteristiche e prestazioni delle nuove fabbriche dati, (2018)	17
8	Distribuzione di potenza Fonte: ENEA, Definizione di algoritmi e indicatori per l'efficientamento dei centri di elaborazione dati (CED), www.enea.it, (2011) . .	19
9	Ripartizione consumo di energia tipico di un data center Fonte: The Shift Project, lean-ict towards digital sobriety, (2019) . .	22
10	Evoluzione della domanda di energia dei data center nell'UE a partire dal 2010 al 2018. Fonte: Montevecchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)	24
11	Distribuzione della domanda di energia del data center tra questi quattro paesi e il resto dell'UE28. Fonte: Montevecchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)	25
12	Consumo energetico del data center come percentuale della domanda dell'energia elettrica finale nell'UE28 per gli anni dal 2010 al 2030. Fonte: Montevecchi F. Stickler T. Hintemann R. Hinterholzer S, Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market, (2020)	27

- 13 Percentuali tipiche di consumo in un data center dei vari componenti
Fonte: ENEA, Definizione di algoritmi e indicatori per l'efficiamento dei centri di elaborazione dati (CED), www.enea.it, (2011) . . . 29

Riferimenti bibliografici

- [1] BELKHIR L. ELMELIGI A. «Assessing ICT global emissions footprint: Trends to 2040 recommendations». In: *Journal of Cleaner Production* 177 (2018), pp. 448–463. URL: <https://www.sciencedirect.com/science/article/pii/S095965261733233X>.
- [2] HILTY L.M. et al. *ICT4S 2013: Proceeding of the first International Conference on Information and Communication Technologies for Sustainability*. Zurich, Switzerland, 14-16 February 2013. URL: <https://doi.org/10.3929/ETHZ-A-007337628>.
- [3] R. ATTFIELD. «Sustainability». In: *International Encyclopedia of Ethics, Wiley* (2013). URL: <https://doi.org/10.1002/9781444367072.wbiee033>.
- [4] KLEYMAN B. «Why Consider a Modular Data Center? A DCK Guide». In: *Data Center Knowledge* (2013). URL: <https://www.datacenterknowledge.com/archives/2013/04/09/why-consider-a-modular-data-center>.
- [5] BRAMUCCI M. DI SANTO D. FORNI D. *Linee guida per la progettazione di datacenter ad alta efficienza*. 2010. URL: <https://www.fire-italia.org/prova/wp-content/uploads/2014/03/Linee-guida-per-la-progettazione-di-datacenter-ad-alta-efficienza.pdf>.
- [6] DELOITTE. «Digital (Green) Evolution L’impatto ambientale degli smartphone e i comportamenti dei consumatori.» In: *Focus Italia-Digital Consumer Trends Survey 2021* (2022). URL: https://www2.deloitte.com/content/dam/Deloitte/it/Documents/technology-media-telecommunications/Digital_green_evolution_Deloitte.pdf.
- [7] IEAG – UNITED NATIONS INDEPENDENT EXPERT ADVISORY GROUP ON DATA REVOLUTION FOR SUSTAINABLE DEVELOPMENT. *A World That Counts: Mobilising the Data Revolution for Sustainable Development*. 2014. URL: <https://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>.
- [8] FILIPPAZZ F. OCCHINI G. SALA F. «Tecnologie digitali». In: *Il Contributo italiano alla storia del Pensiero - Tecnica* (2013). URL: [https://www.treccani.it/enciclopedia/tecnologie-digitali_\(Il-Contributo-italiano-alla-storia-del-Pensiero:-Tecnica\)/](https://www.treccani.it/enciclopedia/tecnologie-digitali_(Il-Contributo-italiano-alla-storia-del-Pensiero:-Tecnica)/).

- [9] FOSSA F. LUCIVERO F. SCHIAFFONATI V. TAMBURRINI G. «Sostenibilità ambientale della società dell'informazione». In: *Automi e persone, Introduzione all'etica dell'intelligenza artificiale e della robotica*, Carrocci editore (2021), pp. 153–171.
- [10] WHITEHEAD B. ANDREWS D. SHAH A. MAIDMENT G. «Assessing the Environmental Impact of Data Centers Part 1: Background, Energy Use and Metrics». In: *Building and Environment* 82 (2014), pp. 151–159. URL: <https://www.sciencedirect.com/science/article/abs/pii/S036013231400273X>.
- [11] LACY P. LAMONICA B. RUTQVIST J. *Circular Economy: Dallo spreco al valore*. Prima Edizione. Milano:EGEA, 2016.
- [12] STRICKLAND J. *How Server Virtualization Works*. 2008. URL: <https://computer.howstuffworks.com/server-virtualization.htm>.
- [13] CHRISTOPHER J. CROSBY JR. «Truly modular building datacenter facility». In: *United States Patent Application Publication* (2014). URL: <https://patentimages.storage.googleapis.com/1b/fc/8f/8fdb53ec78986f/US20140352231A1.pdf>.
- [14] OBRINGER R. RACHUNOK B. MAIA-SILVA D. ARBABZADEH M. NATEGHI R. MADANI K. «The overlooked environmental footprint of increasing Internet use». In: *Resources, Conservation and Recycling* 167 (2021). URL: <https://www.sciencedirect.com/science/article/pii/S0921344920307072>.
- [15] CIARAMELLA A. ROVEDA M. *Data Center: localizzazione, caratteristiche e prestazioni delle nuove fabbriche dati*. Milano: FrancoAngeli, 2018.
- [16] RASMUSSEN N. «Implementing energy efficient data centers.» In: *APC White Paper* (2006). URL: http://book.itep.ru/depository/green_computing/NRAN-6LXSHX_R0_EN.pdf.
- [17] THE SHIFT PROJECT. *Lean ict- towards digital sobriety. report of the working group directed by hugues ferreboeuf for the think tank the shift project*. 2019. URL: https://theshiftproject.org/wp-content/uploads/2019/03/Lean-ICT-Report-The-Shift-Project_2019.pdf.
- [18] MATHESON R. «Artificial intelligence could help data centers run far more efficiently». In: *MIT News on campus and around the world* (2019). URL: <https://news.mit.edu/2019/decima-data-processing-0821>.

- [19] MONTEVECCHI F. STICKLER T. HINTEMANN R. HINTERHOLZER S. *Energy-efficient Cloud Computing Technologies and Policies for an Eco-friendly Cloud Market*. 2020, pp. 52–64. URL: <https://d110erj175o600.cloudfront.net/wp-content/uploads/2020/11/FINALSTUDYEnglishKK-03-20-210-EN-N13072020pdf.pdf>.
- [20] ENEA-Agenzia nazionale per le nuove tecnologie l'energia e lo sviluppo economico sostenibile. *ENEA testa intelligenza artificiale per ottimizzare funzionamento dei data center*. 2023. URL: <https://it.overleaf.com/project/641445ceb0569c573cca22b0>.
- [21] ENERGY STAR. «Recommendations for Measuring and Reporting Overall Data Center Efficiency». In: *Version 1 – Measuring PUE at Dedicated Data Centers* (2010). URL: https://www.energystar.gov/ia/partners/prod_development/downloads/Data_Center_Metrics_Task_Force_Recommendations.pdf.