

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea Triennale in Scienze di Internet

IL PROBLEMA DEL RISPARMIO ENERGETICO NEI DATACENTER

Tesi di Laurea in Algoritmi e Strutture Dati

Relatore:
Chiar.mo Prof.
MORENO MARZOLLA

Presentata da:
ALESSANDRO PULICE

Sessione I
Anno Accademico 2010 - 2011

*Ai miei genitori
che mi hanno dato la possibilità
di raggiungere questo traguardo.*

Introduzione

Con la crescente diffusione del web e dei servizi informatici offerti via internet, è aumentato in questi anni l'utilizzo dei data center e conseguentemente, il consumo di energia elettrica degli stessi. Il problema ambientale che comporta l'alto fabbisogno energetico, porta gli operatori di data center ad utilizzare tecniche a basso consumo e sistemi efficienti. Organizzazioni ambientali hanno rilevato che nel 2011 i consumi derivanti dai data center raggiungeranno i 100 milioni di kWh, con un costo complessivo di 7,4 milioni di dollari nei soli Stati Uniti, con una proiezione simile anche a livello globale.

La seguente tesi intende valutare le tecniche in uso per diminuire il consumo energetico nei data center, e quali tecniche vengono maggiormente utilizzate per questo scopo. Innanzitutto si comincerà da una panoramica sui data center, per capire il loro funzionamento e per mostrare quali sono i componenti fondamentali che lo costituiscono; successivamente si mostrerà quali sono le parti che incidono maggiormente nei consumi, e come si devono effettuare le misurazioni per avere dei valori affidabili attraverso la rilevazione del PUE, unità di misura che valuta l'efficienza di un data center. Dal terzo capitolo si elencheranno le varie tecniche esistenti e in uso per risolvere il problema dell'efficienza energetica, mostrando alla fine una breve analisi sui metodi che hanno utilizzato le maggiori imprese del settore per risolvere il problema dei consumi nei loro data center.

Lo scopo di questo elaborato è quello di capire quali sono le tecniche e le strategie per poter ridurre i consumi e aumentare l'efficienza energetica dei data center.

Indice

Introduzione	iii
1 Struttura di un data center	3
1.1 Introduzione ai data center	3
1.2 I data center e il web	4
1.3 Parti del data center	5
1.3.1 Sistema di raffreddamento	6
1.3.2 Power delivery	7
1.3.3 IT equipment	8
1.4 Distribuzione dei costi	9
2 Considerazioni energetiche nei data center	11
2.1 Necessità di ridurre i consumi	12
2.2 Misurare i consumi	14
2.3 Punti di intervento	18
3 Tecniche di risparmio energetico	21
3.1 Ottimizzazione della struttura del data center	22
3.1.1 Right-sizing	22
3.1.2 Progettazione di un sistema efficiente	23
3.1.3 Scelta di componenti efficienti	24
3.2 Consumo proporzionale al lavoro	25
3.2.1 Transizione di stato	28
3.2.1.1 ACPI	30

3.2.2	Cambio dinamico di frequenza e voltaggio	32
3.2.3	Gestione energetica	33
3.3	Virtualizzazione e consolidamento	36
4	Casi di studio	43
4.1	Analisi e tecniche in uso	44
4.2	Risultati recenti	47
4.2.1	Open Compute	48
	Conclusioni	51
	Bibliografia	53

Elenco delle figure

1.1	Tipici componenti elettrici di un Data Center [10]	5
1.2	Tipico sistema di raffreddamento di un data center [13]	6
1.3	Flusso elettrico in un tipico data center [15]	10
2.1	Proiezioni future del consumo elettrico, tutti gli scenari[9]	13
2.2	Comparazione del consumi con la capacità disponibile [14]	15
3.1	Consumo energetico del server [7]	26
3.2	Contributo della CPU ai consumi del server [7]	28
3.3	Consumo energetico del server [7]	29
3.4	Pianificazione di cinque processi [6]	36
4.1	Classificazione dei data center analizzati [10]	45
4.2	PUE totale con utilizzo del sistema di raffreddamento libero [10]	47

Capitolo 1

Struttura di un data center

Un datacenter è un impianto utilizzato per ospitare sistemi informatici e i componenti associati necessari al suo funzionamento, come la rete ed i sistemi di storage. Comprendono un doppio sistema di alimentazione, una rete di comunicazione ridondante, un sistema di controllo ambientale (ad esempio, raffreddamento e climatizzazione) e sistemi di sicurezza.

1.1 Introduzione ai data center

I data center hanno origine nelle grandi sale computer, che contenevano i primi sistemi informatici. Data la loro complessità, richiedevano un ambiente particolare in cui operare. Dovendo stendere molti cavi e utilizzare strutture apposite per ospitare i vari componenti in modo sicuro: come telai standard per montare le attrezzature, pavimenti sopraelevati e supporti per cavi, fu necessario costruire stanze apposite per contenerli ed organizzarli. Inoltre, tutti i componenti richiedendo molta energia, nell'operare si surriscaldano facilmente, e dovevano essere continuamente raffreddati attraverso sistemi di raffreddamento, più o meno complessi, che richiedevano ulteriore spazio.

Inizialmente non c'era molta organizzazione e ogni sistema era differente dagli altri ma, dopo il 1980 con i computer che cominciarono ad essere più diffusi nel mondo e con la crescita di complessità delle operazioni, le imprese

diventarono consapevoli della necessità di organizzare e controllare le risorse IT. Con l'avvento del client-server computing e la disponibilità a basso costo delle attrezzature di rete è stato possibile utilizzare un disegno gerarchico e ordinato che permetteva di inserire i server in un locale specifico all'interno delle aziende. È a questo punto che nasce la definizione vera e propria di data center.

1.2 I data center e il web

Il boom dei data center è avvenuto con la bolla del dot-com, sempre più imprese per distribuire i loro servizi, ed avere una presenza sul web, avevano bisogno di connettività Internet a grande velocità con una continua disponibilità. Per molte piccole imprese non era possibile sostenere i costi d'installazione di queste attrezzature, quindi si cominciarono a costruire grandi impianti, chiamati *Internet Data Centers* (IDC), che permettevano a più aziende di utilizzare un'ampia gamma di soluzioni per la distribuzione di servizi e operazioni. Nel corso degli anni, sono state sviluppate nuove tecnologie e pratiche specifiche per la gestione di operazioni su larga scala, che successivamente, sono state largamente adottate a causa della loro praticità in quasi tutti i data center.

A partire dal 2007, il design di data center, la costruzione e il loro funzionamento, sono diventate una disciplina molto conosciuta, ed esistono documenti standard, creati da gruppi di professionisti e organizzazioni, che specificano i requisiti per la progettazione dei data center e le migliori pratiche da utilizzare. Si stanno sviluppando, inoltre, tecniche di progettazione e strumenti eco-sostenibili, per abbassare i consumi e l'inquinamento considerando che in genere i data center sono molto costosi sia da costruire, ma soprattutto da mantenere, ma soprattutto per il loro aumento di diffusione e uso negli ultimi anni.

Esigenze di business nel mercato di Internet, richiedono sempre di più l'aumento del numero di data center e di energia ad uso dei server. Più le

aziende aumentano la loro offerta di servizi digitali, più la domanda per la potenza di calcolo e di memoria aumenta. Alcuni esempi di servizi digitali in aumento, in questi anni, sono: on-line banking, e-commerce, monitoraggio online delle spedizioni postali e il download e streaming di musica e video.

1.3 Parti del data center

In questo paragrafo si descriveranno le parti in cui è composto un data center. In particolare, si parlerà dei componenti principali che permettono il funzionamento del data center e che richiedono un'elevato consumo di energia elettrica: i sistemi di raffreddamento, i sistemi di approvvigionamento energetico e le macchine informatiche; senza entrare nel dettaglio delle altre parti esistenti come la struttura fisica, gli uffici e i laboratori per i tecnici e lo staff.

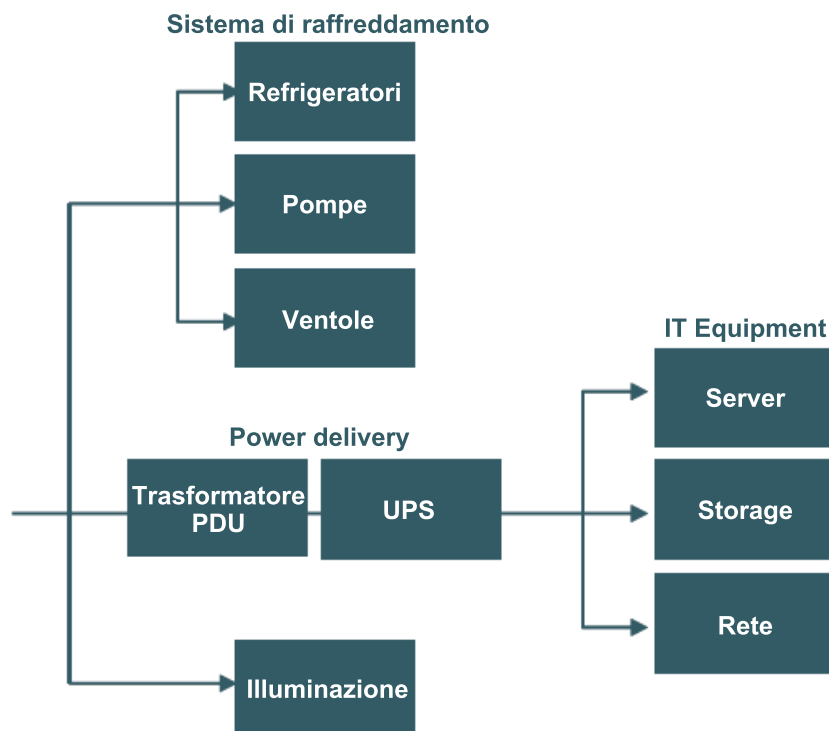


Figura 1.1: Tipici componenti elettrici di un Data Center [10]

1.3.1 Sistema di raffreddamento

L'infrastruttura di raffreddamento occupa una parte significativa di un data center. Il numero di server e macchine presenti all'interno, durante il loro funzionamento, producono un'elevato calore. Per evitare il surriscaldamento e il conseguente danneggiamento dei componenti IT, è necessario raffreddarli continuamente, e tenerli ad una temperatura costante. Per questo scopo si utilizza un sistema di raffreddamento che comprende refrigeratori, pompe e ventole.

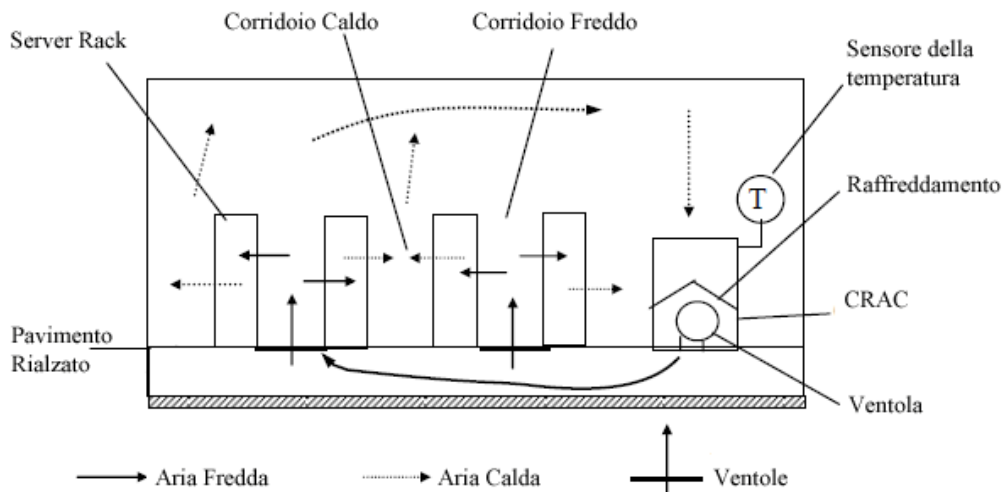


Figura 1.2: Tipico sistema di raffreddamento di un data center [13]

La Figura 1.2 mostra un tipico sistema di raffreddamento, con unità di condizionamento modulari e con un pavimento rialzato che permette la circolazione di aria fredda. La dimensione e la potenza del sistema è dimensionato in base al carico previsto dei server, durante la fase di design del data center.

Il flusso dell'aria viene distribuito uniformemente attraverso il *CRAC* (computer room air conditioning), che controlla la temperatura attraverso un sensore tipicamente settato a 20°C, che misura la temperatura interna. L'aria fredda viene distribuita attraverso il pavimento della stanza con l'utilizzo di ventole e pompe, raggiunge i server raffreddandoli e successivamente l'aria calda, che va in alto, ritorna nel condizionatore per essere nuovamente raffreddata e distribuita nel data center.

L'insieme di questi componenti crea un'ambiente ottimale e garantisce longevità ai vari componenti. Allo stesso tempo, però, questo ecosistema ha un costo consistente nell'economia del data center.

1.3.2 Power delivery

È il sistema di approvvigionamento energetico del data center. Tipicamente l'infrastruttura comprende il collegamento con la rete elettrica, generatori, batterie di backup e l'energia per il sistema di raffreddamento.

L'energia viene recuperata dalla rete elettrica esterna. In caso di mancanza di corrente interviene un generatore all'interno del data center che provvede a fornire l'energia necessaria al sistema. Attraverso questi apparati elettrici si ricaricano delle batterie di backup, che hanno l'utilità di mantenere una potenza costante anche in caso di interruzioni sia dalla rete elettrica che dal generatore interno e per compensare fluttuazioni della rete o temporanee perdite di potenza.

L'energia, viene distribuita alle macchine del datacenter attraverso l'*UPS* (gruppo statico di continuità), non prima di essere opportunamente adattata al voltaggio di funzionamento, dei componenti IT, utilizzando l'unità *PDU* (power distribution unit). Si può avere anche un doppio collegamento a due differenti *PDU* per migliorare l'affidabilità del sistema.

Il grado di ridondanza nel sistema, di conseguenza, è una considerazione molto importante nell'infrastruttura. Il data center non deve rimanere mai senza energia per questo sono previste varie fonti di approvvigionamento e batterie di riserva, più il sistema è ridondante maggiore sarà la resistenza alle

emergenze ma, allo stesso tempo, i costi d'installazione e di manutenzione di tale sistema saranno ancora più elevati.

1.3.3 IT equipment

IT (*Information Technology*) è l'area che comprende diverse tecnologie e apparecchiature tra cui ma non limitatamente a queste: hardware, software, sistemi informativi e linguaggi di programmazione. In sintesi, tutto quello che rende i dati, informazioni o conoscenze percepibili in un formato visivo, e che possono essere distribuite attraverso qualsiasi mezzo multimediale.

In un data center è la parte principale, che rende possibili le sue operazioni fondamentali, e contiene gli apparati elettronici che sono utilizzati per l'elaborazione dei dati (server), l'immagazzinamento dei dati (storage) e le comunicazioni (network). L'insieme di questi componenti è conosciuto come «information technology equipment». Collettivamente, questi componenti processano, memorizzano e trasmettono informazioni digitali. Richiedono un'ambiente esclusivo, dove vengono posizionati e organizzati in modo ottimale, per ridurre il surriscaldamento e per impedire incidenti alle macchine e allo staff di supporto.

Il numero di elementi dell'IT equipment deve essere deciso in fase di progettazione del centro. È molto importante considerare la potenza necessaria in fase di partenza e prevedere quella futura, per lasciare spazio a nuovi aggiornamenti o all'inserimento di nuovi componenti e, allo stesso tempo, per evitare una spesa eccessiva o inutile. Questa è una delle parti, che incide maggiormente nei costi del data center, sia per l'acquisto sia per la manutenzione; ma hanno anche un costo dal punto di vista dei consumi elettrici compresi quelli derivanti dal sistema di raffreddamento, più è grande la dimensione dell'IT equipment, maggiore sarà la potenza necessaria per raffreddare i componenti.

1.4 Distribuzione dei costi

I componenti descritti nel paragrafo precedente, richiedono molta energia elettrica e di conseguenza hanno un'elevata incidenza sui costi del datacenter.

Nonostante il fatto che i costi derivanti dall'energia elettrica, nella vita di un data center, possono superare i costi dei sistemi elettrici come l'*UPS* e dell'*IT equipment*, durante la fase di design del data center, tipicamente, l'utilizzo di energia elettrica non è un criterio di progettazione e non viene gestito efficacemente come un costo. Questo accade per vari motivi:

- non sono disponibili adeguate informazioni durante la pianificazione e per le decisioni sull'acquisto dei componenti elettrici;
- non sono ampiamente disponibili o conosciuti strumenti di modellazione dei costi e non vengono utilizzati durante la progettazione;
- la bolletta energetica arrivando dopo che i consumi sono avvenuti, si pensa che non siano collegabili a nessuna particolare decisione o pratica e quindi vengono visti come costi inevitabili;
- spesso i costi energetici non sono sotto la responsabilità o al budget del gruppo operativo del centro;
- la bolletta energetica del data center può essere inclusa in una bolletta elettrica generale e non è disponibile separatamente.

Senza un'adeguata progettazione si hanno dei costi e consumi che potrebbero essere evitati. Semplici decisioni no-cost inserite nel design di un nuovo data center possono risultare in un risparmio di 20-50% della bolletta energetica, e con uno sforzo sistematico si può arrivare fino ad un risparmio del 90% [15]. Nel corso degli anni varie tecniche e studi sono stati effettuati per ridurre i consumi nelle varie aree di un data center, prima di intervenire, bisogna capire in che modo l'energia è consumata. La Figura 1.3 mostra come è suddiviso, il flusso elettrico, tra i vari componenti di un tipico centro.

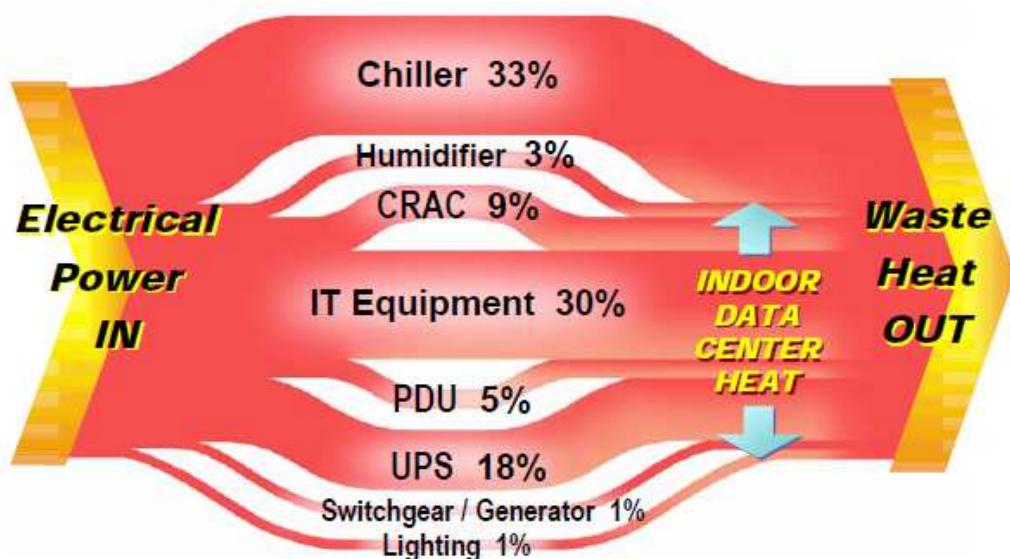


Figura 1.3: Flusso elettrico in un tipico data center [15]

Come si vede, circa la metà o meno dell'energia usata dal data center è consumata dalle apparecchiature dell'area IT. L'altra metà è utilizzata per la parte circostante di supporto, cioè alla *NCPI* (Network-Critical Physical Infrastructure) che comprende l'infrastruttura di alimentazione, di raffreddamento e illuminazione.

La figura mostra il flusso dell'energia elettrica, dall'ingresso ai consumi delle varie parti di un data center che opera a circa il 30% delle capacità, fino all'uscita di calore di scarto che viene immesso all'esterno nell'atmosfera.

Capitolo 2

Considerazioni energetiche nei data center

Come visto nel capitolo precedente, i costi dell'energia elettrica possono avere un'alta incidenza nel bilancio di un data center se non si applicano dei sistemi o strategie di risparmio energetico.

Considerando le numerose parti in cui è composto un data center, è importante capire come misurare i consumi e individuare i punti in cui ottimizzare l'impianto. Non è un compito facile ridurre i consumi perché, spesso, durante la costruzione del data center i vari manager che gestiscono la struttura non guardano ad un aspetto generale del problema, ma si concentrano sulle rispettive aree di competenza, comunicando con difficoltà tra loro: lo staff di supporto, lotta per limitare lo spazio occupato dalle attrezzature IT e per connetterle all'impianto per l'energia elettrica; nel frattempo i tecnici IT provano ad assicurare la potenza di calcolo, la banda di rete, e capacità di memorizzazione necessaria alle funzioni IT, predisponendo una sufficiente ridondanza per evitare interruzioni del sistema.

In questo capitolo si cercherà di capire perché è importante ridurre i consumi, e come individuare le parti in cui intervenire dopo aver utilizzato gli strumenti giusti per misurare i costi.

2.1 Necessità di ridurre i consumi

Se si considera che il costo tipico per kWh all'ora è di 0,12\$, il costo annuo per kW può raggiungere circa i 1.000\$. Nella vita tipica di un data center, che è di 10 anni, il costo raggiunge i 10.000\$ per kWh, questo solo per i componenti IT, come i server, le memorie e i processori. Queste attrezzature consumano circa la metà dell'energia utilizzata da un data center, quindi, l'intera struttura raggiungerà approssimativamente i 20.000\$ per kWh. Per esempio un data center con una potenza di 200 kW, in 10 anni avrà una spesa elettrica complessiva di 4.000.000\$ [15].

Con i consumi derivanti dalle infrastrutture IT sono strettamente collegati quelli dell'energia utilizzata per le attrezzature di raffreddamento, quindi sta diventando sempre più importante trovare e applicare delle soluzioni per ridurre i consumi, considerando inoltre, l'impatto ambientale che, con la diffusione negli ultimi anni dei data center e la crescente domanda di energia necessaria a queste strutture, causando l'emissione di gas serra, è diventato un problema di primaria importanza. Purtroppo molto spesso, questo aspetto, non viene considerato.

Per questi motivi gli operatori di data center stanno cercando di ottimizzare le loro con tecnologie verdi (green IT). I costi risparmiati e i benefici ambientali dovuti con l'introduzione di questi strumenti sono notevoli e per massimizzarli è importante una partnership tra i manager aziendali e quelli dell'area IT per puntare allo stesso obiettivo poiché, ottimizzare l'efficienza energetica dei data center richiede un'attenta pianificazione e la giusta disponibilità di componenti come alimentazione, raffreddamento e sistemi di rete, che incontrino le necessità correnti e di scala per futuri ampliamenti, riducendo al minimo i costi.

Le strategie verdi mostrano il meglio durante l'espansione e l'aggiornamento del data center, permettendo un aumento dell'efficienza energetica, di vecchie strutture, diminuendo il numero dei server in funzione, riducendo così i costi e aumentando la durata della vita del data center. Infatti, quando si raggiunge l'85-95% dello spazio per i vari componenti e per la capacità di

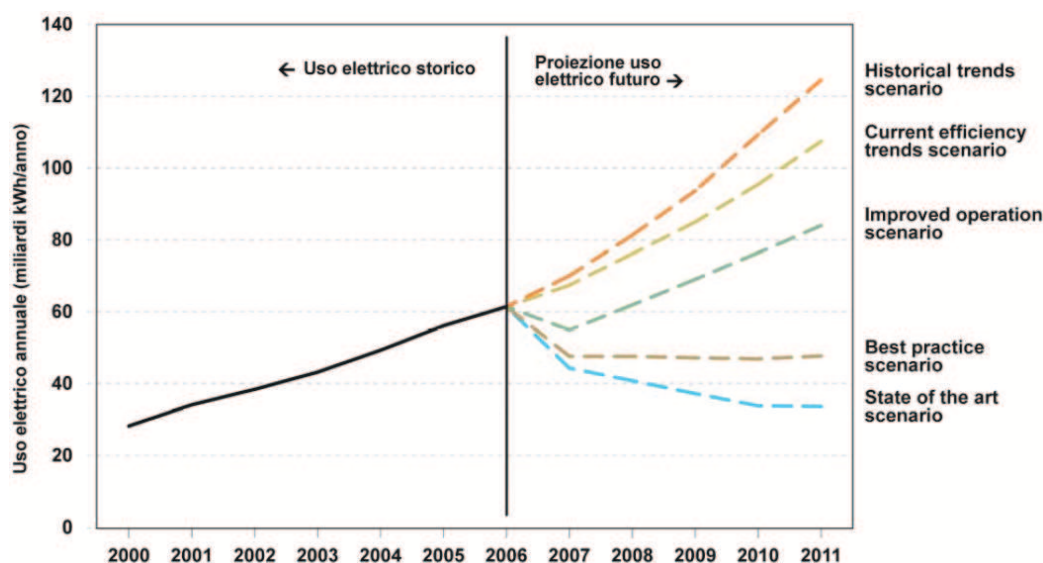


Figura 2.1: Proiezioni future del consumo elettrico, tutti gli scenari[9]

rete, gli operatori del data center dovrebbero prendere in seria considerazione di espandere la struttura, o di costruirne una nuova, è una decisione molto importante che può avere un grosso impatto sull'economia dell'azienda.

In base agli attuali trends, nel 2007, è stato stimato che il consumo di elettricità nazionale degli Stati Uniti derivante da server e data center raggiungerà, nel 2011, i 100 milioni di kWh, con una bolletta energetica corrispondente a 7,4 miliardi di dollari. La domanda di energia, durante i picchi di carico dei data center, nel 2007 era di 7 gigawatts, l'equivalente della produzione di 15 centrali elettriche, e si stimava che nei successivi 5 anni avrebbe raggiunto i 15 gigawatts, cioè ad altre 10 centrali elettriche[9]. Lo stesso incremento si può notare anche a livello globale, e lo si può vedere graficamente, nella Figura 2.1, che indica la tendenza futura divisa in vari scenari che indicano un possibile risparmio energetico dovuto al miglioramento tecnologico e all'ampio utilizzo di nuove strategie e pratiche energetiche.

C'è un significativo potenziale miglioramento dell'efficienza energetica nei data center, anche se alcuni miglioramenti dell'efficienza energetica sono pre-

visti se le attuali tendenze, ad attuare nuove tecniche, saranno confermate anche in futuro. Molte tecnologie sono già commercialmente disponibili o lo saranno presto, questo potrebbe portare ad un'ulteriore miglioramento nell'uso energetico nei microprocessori, server, dispositivi di storage, apparati di rete e in tutte le altre infrastrutture del sistema. Le migliori tecniche verranno spiegate nel capitolo successivo, ma è già possibile, con le tecnologie esistenti e utilizzando le nuove strategie di design, ridurre l'utilizzo di energia di un tipico server del 25 per cento o più. La definizione dei vari scenari sarà indicata nel capitolo 4, dove verranno mostrati alcuni casi di studio reali e i risultati che si stanno ottenendo.

2.2 Misurare i consumi

Per pianificare un sistema energetico efficiente, il primo passo è quello di misurare l'uso corrente di energia. Il sistema di alimentazione è la parte più critica dell'infrastruttura del data center. È difficile individuare in modo preciso i punti in cui l'energia viene utilizzata e persa nei punti deboli dell'impianto.

Per misurare il consumo d'energia bisogna calcolare la *PUE* (Power Usage Effectiveness) del data center, che indica quanto è efficiente la struttura dal punto di vista energetico. Questo parametro fa capire quanta potenza elettrica è dedicata all'alimentazione degli apparati IT rispetto ai servizi ausiliari come il condizionamento o l'UPS.

In pratica, è il rapporto tra la potenza utilizzata da tutti gli apparati IT (*Energia IT*) e quella totale assorbita dal data center (*Energia Totale*), il rapporto da calcolare è:

$$PUE = \frac{Energia\ Totale}{Energia\ IT}$$

Un valore di PUE pari a 1, indica che tutta l'energia assorbita dal data center è utilizzata dai componenti IT, e rappresenta teoricamente il valore

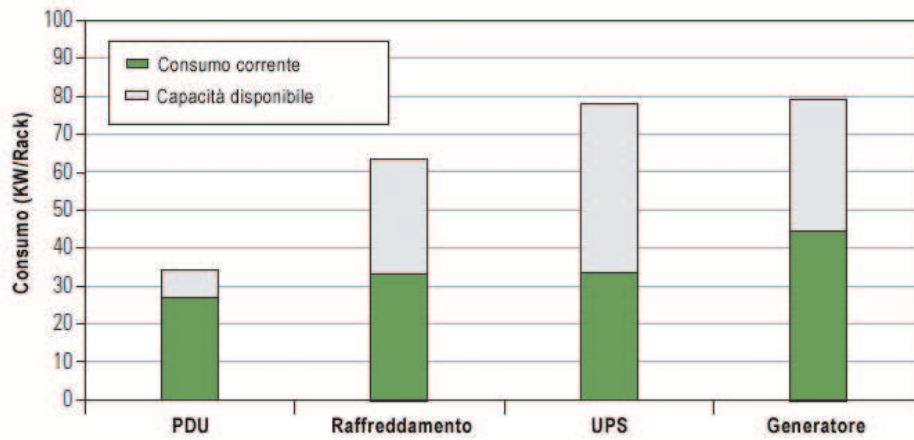


Figura 2.2: Comparazione del consumi con la capacità disponibile [14]

perfetto d'efficienza. Mentre, per esempio, un PUE uguale a 2 indica che per ogni watt consumato dalle attrezzature IT, un'altro watt è necessario per i componenti di supporto e per il resto della struttura.

Molte imprese non hanno sistemi di misura accurati che permettano agli operatori IT di verificare i risultati delle loro decisioni, perché solitamente sono i reparti amministrativi che vedono la bolletta energetica mensile che include l'energia complessiva utilizzata dal data center. Per riuscire ad avere un dato e monitorare il consumo d'energia si dovrebbe iniziare con la creazione di un «profilo di potenza» per ogni rack (armadio che contiene i server). Il profilo è utile per controllare il consumo a livello individuale, in questo modo è possibile controllare ogni server e assicurare che i consumi restino nei limiti previsti, identificando il peso di ogni elemento sui consumi elettrici e stabilendo una base di calcolo per futuri interventi o modifiche.

Il grafico, nella Figura 2.2, mostra un esempio dei dati rilevati con misurazioni individuali, riuscendo a rilevare il livello attuale di consumo rispetto al consumo massimo raggiungibile dai vari componenti.

Per fare delle misurazioni corrette, un gruppo di aziende americane, operanti nel campo energetico, in un'incontro del 13 Gennaio 2010, ha proposto

delle linee guida, indicando misure e convenzioni per misurare l'efficienza energetica nei data center [5].

Le raccomandazioni da seguire sono:

- Il PUE è la misura preferita per misurare l'efficienza energetica di un data center.
- Nel calcolare il PUE, il consumo energetico deve essere misurato almeno all'uscita dell'UPS. Con il progressivo aumento delle capacità di misurazione le imprese dovrebbero misurare direttamente dai server.
- In un data center dedicato, il totale dell'energia nell'equazione del PUE, deve comprendere tutte le fonti energetiche disponibili dal proprietario o dall'operatore del data center.
- Nei data center situati in edifici misti, vanno inclusi nel conteggio totale anche i consumi di raffreddamento, illuminazione, e delle infrastrutture di supporto alle operazioni del data center.

Queste linee guida hanno l'intento di aiutare gli operatori di data center ad avere una comune metrica per capire l'efficienza energetica, che possa generare un dialogo che possa favorire l'efficienza dei centri e ridurre il consumo energetico, rendendo possibile il confronto dei consumi tra vari data center individuando quello che utilizza le migliori tecniche di risparmio.

Il PUE, inoltre, è stato suddiviso in 4 categorie, da 0 a 3, in base al grado di precisione con cui vengono effettuate le misurazioni, dal più basso al più alto. Ogni categoria stabilisce come ricavare i valori da inserire nella formula del PUE.

- La **categoria 0** rappresenta il livello più basso di precisione nel determinare il PUE. Nel calcolo, l'Energia IT si ricava dalla lettura dei consumi annuali dell'energia direttamente all'uscita dell'UPS, mentre l'Energia Totale si rileva dal contatore generale, quindi dal consumo indicato dalla bolletta energetica. Basandosi solo su una lettura istantanea, si perde completamente l'impatto delle fluttuazioni energetiche

dipendenti dal carico di lavoro o dalle perdite del sistema, ma resta comunque una misura utile che può fornire informazioni sull'efficienza energetica del data center. Questa categoria non prevede altre fonti d'energia utilizzate dal data center, mentre nelle altre categorie l'Energia Totale si ricava sommando i consumi di altre fonti d'energia.

- La **categoria 1** del PUE rappresenta una misurazione più precisa. Rispetto alla categoria precedente l'Energia IT si ricava dalla somma delle letture, dei consumi mensili, dei contatori posizionati all'uscita dell'UPS e in altri punti di controllo posizionati nell'impianto elettrico. In questo caso l'efficienza indicata dal PUE è più accurata, perché si riescono a misurare anche le fluttuazioni dipendenti dal carico di lavoro dei server e dalle perdite del sistema.
- La **categoria 2** rispetto alla precedente aggiunge precisione, rimuovendo dai consumi IT l'impatto delle perdite associate ai trasformatori, misurando l'Energia IT subito dopo il trasformatore collegato alle apparecchiature IT.
- La **categoria 3** rileva la massima precisione sull'efficienza energetica del data center, riuscendo a riportare esclusivamente i consumi IT sul consumo generale. In questa categoria l'Energia IT si calcola dalla somma delle letture, dei consumi mensili, prese nei punti di collegamento tra l'impianto elettrico e le attrezzature IT, riuscendo, in questo modo, a rimuovere tutti i consumi derivanti dalle attrezzature non IT.

La categoria del PUE deve essere indicata dagli operatori di data center, nella pubblicazione dei rapporti sul consumo, facilitando la comprensione dei dati e il confronto trasparente con altri impianti. Nel calcolare il PUE, i data center che utilizzano diverse fonti d'energia per avere un'indicazione più precisa sull'efficienza del loro sistema, devono «pesare» i consumi di queste in base al tipo di fonte energetica utilizzata.

La formula si modifica in questo modo:

$$PUE = \frac{\textit{Elettricit\`a consumata} + \textit{Energia alternativa}}{\textit{Energia IT}}$$

L'*Elettricit\`a consumata* sono i kWh consumati dall'energia della rete elettrica (con peso 1 che viene omesso), mentre l'*Energia alternativa* viene calcolata nel seguente modo: per ogni tipo di energia utilizzata si moltiplicano i kWh consumati per il fattore di peso per quel tipo di energia. I pesi da moltiplicare si ricavano dai rapporti forniti dall'EPA. Nel caso fossero presenti diversi tipi d'energia, si sommeranno anche quei consumi moltiplicati per i relativi fattori di peso.¹

2.3 Punti di intervento

Dopo aver effettuato le misurazioni \`e importante fare una previsione di crescita delle apparecchiature IT o l'attivazione di nuovi progetti e ogni altro significativo cambiamento che potrebbe avere un impatto nelle operazioni del data center. Con una migliore pianificazione si possono attuare le migliori soluzioni energetiche, evitando sorprese e impatti negativi sull'organizzazione.

In base alle valutazioni fatte sui livelli di consumo e le previsioni future, \`e possibile esaminare l'infrastruttura per effettuare le modifiche necessarie e mettere in pratica le tecniche per migliorare l'efficienza energetica del data center. Attuando le «best practices» \`e possibile ridurre i consumi energetici del 10-50%[11]. Le aree del data center in cui si pu\`o intervenire sono diverse, tra cui:

- il sistema di condizionamento, migliorando la gestione del flusso d'aria ed eliminando gli ostacoli che non permettono un'ottimale circolazione dell'aria;

¹I fattori energetici dell'EPA sono disponibili all'indirizzo:

http://www.energystar.gov/ia/business/evaluate_performance/site_source.pdf

- la gestione energetica, utilizzando sistemi a risparmio energetico, e sostituendo progressivamente, negli anni, i componenti più vecchi dell'impianto;
- nelle attrezzature IT (server, processori, memorie, network), razionalizzando, virtualizzando e consolidando, per diminuirne il numero complessivo ed avere un consumo inferiore immediato.

Capitolo 3

Tecniche di risparmio energetico

Nel capitolo precedente, abbiamo visto i metodi per misurare i consumi di un data center e come individuare le parti da ottimizzare per ottenere un risparmio energetico. Sono diverse le tecniche e le strategie che negli anni sono state ricercate e sperimentate per questo scopo, e in questo capitolo verranno mostrate alcune delle tecniche più utilizzate e che hanno un impatto migliore nell'aumentare l'efficienza energetica dei data center.

Sono diversi i punti in cui ottimizzare i data center, si può iniziare da una progettazione accurata dell'intero sistema e della struttura per eliminare le perdite fisse di energia; oppure, considerando che i server sono spesso sotto utilizzati rispetto al loro massimo livello di funzionamento, è possibile implementare meccanismi o strategie per aumentare l'efficienza delle apparecchiature IT, riducendo il numero di elementi in funzione ed utilizzandoli al massimo delle capacità.

3.1 Ottimizzazione della struttura del data center

La costruzione di un data center richiede un enorme investimento in termini finanziari per l'acquisto e la gestione delle numerose apparecchiature che risiedono al suo interno. Ogni data center è diverso e unico in base alle specifiche operazioni che deve eseguire e richiede impianti elettrici e meccanismi studiati appositamente per il singolo data center. È importante, quindi, non solo dimensionare correttamente il data center selezionando le apparecchiature adatte all'impianto, ma è necessario anche stimare correttamente la potenza dei componenti IT e delle strutture di supporto. Queste decisioni hanno un grosso impatto sull'efficienza energetica del data center e per evitare perdite nell'impianto è importante progettare il sistema in base al carico di lavoro (right sizing) necessario, e scegliere apparecchiature efficienti per avere consumi ridotti di energia, anche per i componenti non IT.

3.1.1 Right-sizing

Tra le tecniche disponibili il corretto dimensionamento del sistema in base al carico di lavoro è quella che ha il maggiore impatto sull'efficienza finale del data center. Spesso, durante la costruzione, non viene considerato il problema che all'interno del sistema esistono delle perdite fisse d'energia derivanti per esempio dall'impianto di raffreddamento che per mantenere l'ambiente alla temperatura ottimale resta in funzione sia con i server a pieno carico sia nei momenti di inattività. Le perdite infrastrutturali sono proporzionali alla potenza complessiva del sistema e possono occupare un'alta percentuale sul consumo complessivo della rete di supporto (distribuzione elettrica e raffreddamento) alle attrezzature IT. Negli impianti piccoli, che hanno bassi carichi di lavoro è facile che queste perdite superino i consumi IT; oppure, nel caso contrario, in sistemi con alti carichi di lavoro e con un sistema sovradimensionato, le perdite avranno un peso maggiore sulla bolletta energetica. Se si dimensiona correttamente un data center, è possibile risparmiare quasi il

50% dei consumi [15], con una limitata spesa. Quindi è 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.

3.1.2 Progettazione di un sistema efficiente

Uno degli errori più comuni che si ha in fase di progettazione del data center per aumentarne l'efficienza, è quello di pensare ai consumi dei singoli componenti, senza guardare ad una progettazione più accurata dell'intero impianto. Confrontando due diversi data center con differenti progettazioni ma con le stesse apparecchiature e potenza installata, si potranno notare due livelli di consumi energetici diversi a favore del data center che ha una progettazione più attenta dell'intero sistema.

Per questo motivo il design dell'intero sistema è importante, almeno quanto la scelta della potenza e dei sistemi di raffreddamento nel determinare l'efficienza di un data center. Gli errori più comuni di progettazione che riducono l'efficienza, possono essere:

- UPS e trasformatori che lavorano al di sotto delle loro capacità, provocando perdite energetiche;
- posizionamento sbagliato dei condizionatori che riscaldano nella stessa stanza dove altri raffreddano;
- condizionatori che vengono spinti al massimo per inviare l'aria fredda a lunga distanza, invece di utilizzare apposite ventole;
- condizionatori che lavorano a temperature eccessivamente basse rispetto a quella necessaria;
- valvole inserite all'interno del flusso d'aria fredda per controllare il livello di raffreddamento, invece di regolare opportunamente le pompe.

Come si può notare, la maggior parte dei problemi sono relativi alla progettazione del sistema di raffreddamento, che possono portare ad un consumo pari al doppio del necessario alla rete di supporto alle apparecchiature IT; consumi che sarebbero evitabili con una piccola o quasi nulla spesa attraverso l'applicazione di semplici regole di progettazione e organizzazione.

Le strategie trovate per risolvere questi problemi sono principalmente due. La prima è quella di progettare tutti i componenti del sistema espressamente per il data center in cui saranno installati, effettuando dei test mirati ad evitare i problemi elencati precedentemente, posizionando in modo opportuno tutti gli elementi. La seconda è costruire il sistema utilizzando moduli pre-costituiti che vengono già testati dai costruttori per evitare quei tipi di problemi, come i singoli componenti che formano l'impianto di condizionamento e di distribuzione energetica.

La seconda alternativa per evidenti motivi pratici e di spesa è ormai diventato lo standard per la costruzione dei data center, anche per l'alta variabilità del primo approccio.

3.1.3 Scelta di componenti efficienti

Anche se ha un impatto minore, rispetto a quanto non facciano le attrezzature IT, il corretto dimensionamento e la progettazione del sistema, anche la selezione di dispositivi efficienti è un'elemento importante nella progettazione di un data center a basso consumo.

Tra due apparecchiature dello stesso tipo da utilizzare a supporto dell'IT equipment, ci possono essere sostanziali differenze di consumo. In un rapporto del 2005 da parte della U.S. Electric Power Research Institute, è dimostrato che diversi UPS che funzionano al 30% della loro capacità, hanno delle perdite energetiche che vanno dal 4% al 22%, che è una variazione del 500%[15]. Spesso queste differenze non sono accertabili dal manuale tecnico dell'apparecchio e quindi è difficile effettuare una scelta da parte del progettista del data center. Per effettuare una giusta scelta è importante sottoporre i singoli componenti a test specifici di efficienza e compararne i

risultati ed eventualmente, se si hanno le capacità, effettuare delle modifiche sui componenti per abbassare i loro consumi.

3.2 Consumo proporzionale al lavoro

La parte che incide maggiormente nei consumi di un data center, sono sicuramente i dispositivi IT. Questi consumi sono causati principalmente dalle operazioni svolte dei server e dai componenti per fornire le varie funzioni offerte del data center.

Nonostante gli elevati consumi, però, i server sono utilizzati per la maggior parte del tempo al di sotto del 50% della loro capacità massima e restano poche volte completamente inattivi o utilizzati al massimo livello. Proprio nella fascia in cui i server sono più utilizzati, questi mostrano una scarsa efficienza energetica consumando di più di quanto necessario. Questo accade per il semplice fatto che, anche il server più efficiente, nel momento in cui è quasi inattivo e praticamente non è impegnato a svolgere operazioni consuma circa la metà dell'energia; mentre quelli progettati con ancora meno attenzione raggiungono livelli ancora più elevati di consumo. L'energia impiegata dai server cambia al variare del livello di utilizzo dello stesso e delle operazioni che deve svolgere. La Figura 3.1 mostra il consumo d'energia di un tipico server e l'efficienza energetica ottenuta in base al grado di utilizzo. Per misurare l'efficienza energetica del server, si divide la percentuale d'utilizzo con il corrispondente valore dell'energia consumata. Nel grafico, si può notare come il picco dell'efficienza si raggiunge quando il server lavora al massimo della sua capacità, proprio perché vengono sfruttate al massimo le sue possibilità, e cala velocemente con la diminuzione del grado di utilizzo. Proprio nella fascia di maggiore utilizzo del server, si vede come l'efficienza scenda a meno della metà rispetto al picco massimo raggiungibile.

Il motivo per cui i server non sono utilizzati al 100%, è per il fatto di mantenere un certo margine di capacità e riuscire ad accettare tutte le richi-

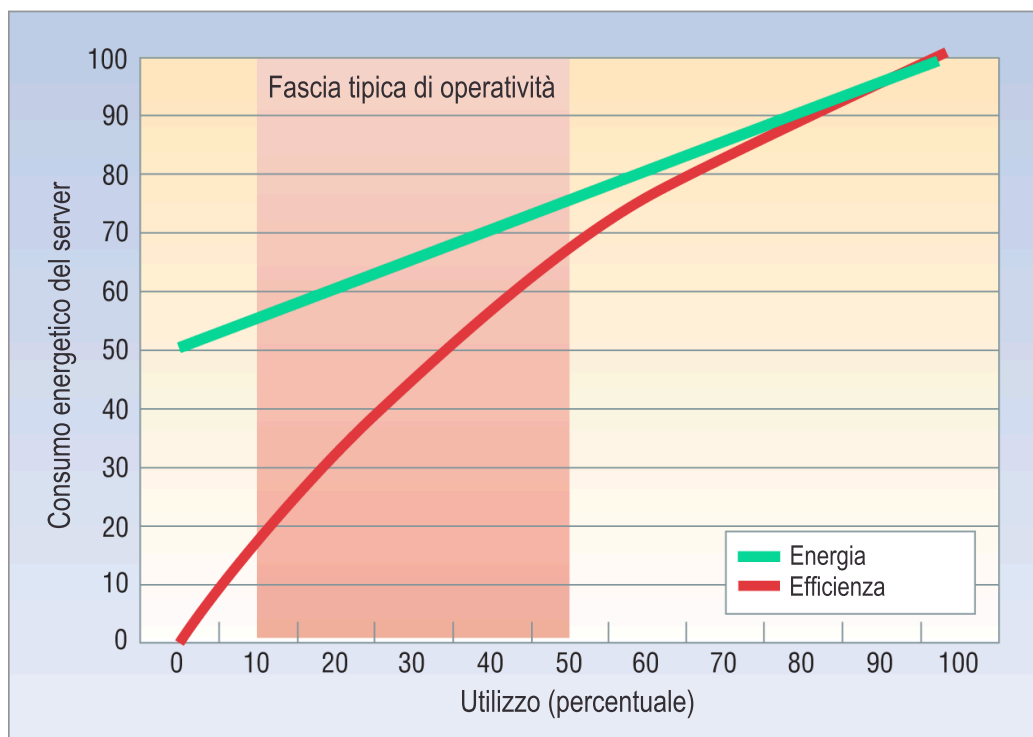


Figura 3.1: Consumo energetico del server [7]

este 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 per l'esecuzione di funzioni più complesse, senza questa flessibilità si rischierebbero frequenti interruzioni di servizio. Non è possibile neanche 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.

Per risolvere questo problema, è importante che i progettisti di componenti e sistemi mirino a costruire server che consumino energia in modo proporzionale al livello di utilizzo. Si dovrebbe puntare a costruire server in grado di non consumare energia nei momenti di inattività e che aumentino progressivamente i consumi all'aumentare delle operazioni da svolgere. Per ottenere questo i sistemi dovrebbero implementare meccanismi dinamici di gestione energetica che consentano il cambio della potenza nei componenti fondamentali, come il processore, ma anche nelle memorie, nei dischi e nei componenti di rete. Nei processori, tecniche di questo tipo, sono già state implementate negli ultimi anni, infatti i consumi dei processori non occupano più una parte consistente dei consumi di un intero sistema, come mostra la Figura 3.2, che mostra il progressivo calo che hanno avuto dal 2005 al 2007 e la differenza di consumi in fase di attività e inattività, dimostrando un comportamento proporzionale. Comportamento che altri componenti non mostrano, senza sfruttare completamente i vantaggi di queste strategie.

Le tecniche utilizzate che permettono ai processori di avere questa efficienza e che altri componenti non implementano in pieno sono: il «power down mechanism» (transizione di stato) e il «dynamic speed scaling» (cambio dinamico di frequenza e voltaggio). Utilizzando questi sistemi dinamici di gestione energetica in maniera più completa, è possibile tagliare di circa la metà il consumo energetico del data center [7]. Nel cambio dinamico

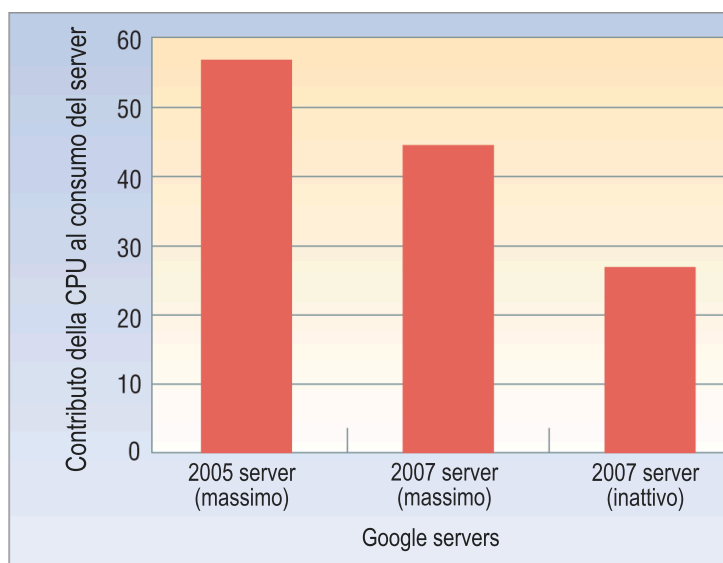


Figura 3.2: Contributo della CPU ai consumi del server [7]

di frequenza e voltaggio, il sistema continua ad essere attivo e funzionante; mentre la transizione di stato prevede la disattivazione graduale di alcune parti o dell'intero sistema, in base ai componenti necessari alle operazioni.

La Figura 3.3 mostra la curva di un server efficiente che utilizza sistemi di consumo proporzionale al lavoro. Si può notare come l'efficienza di questo sistema aumenti rispetto all'esempio precedente.

3.2.1 Transizione di stato

La tecnica della transizione di stato, per ottenere un risparmio energetico, prevede la sospensione e quindi la disattivazione di alcune parti o dell'intero sistema nel momento in cui i vari componenti o il sistema non vengono utilizzati per svolgere funzioni. Solo nel momento di necessità queste parti vengono riattivate.

Per effettuare questa operazione, sono stati previsti degli stati energetici che possono essere semplicemente «attivo» o «inattivo» oppure comprendere

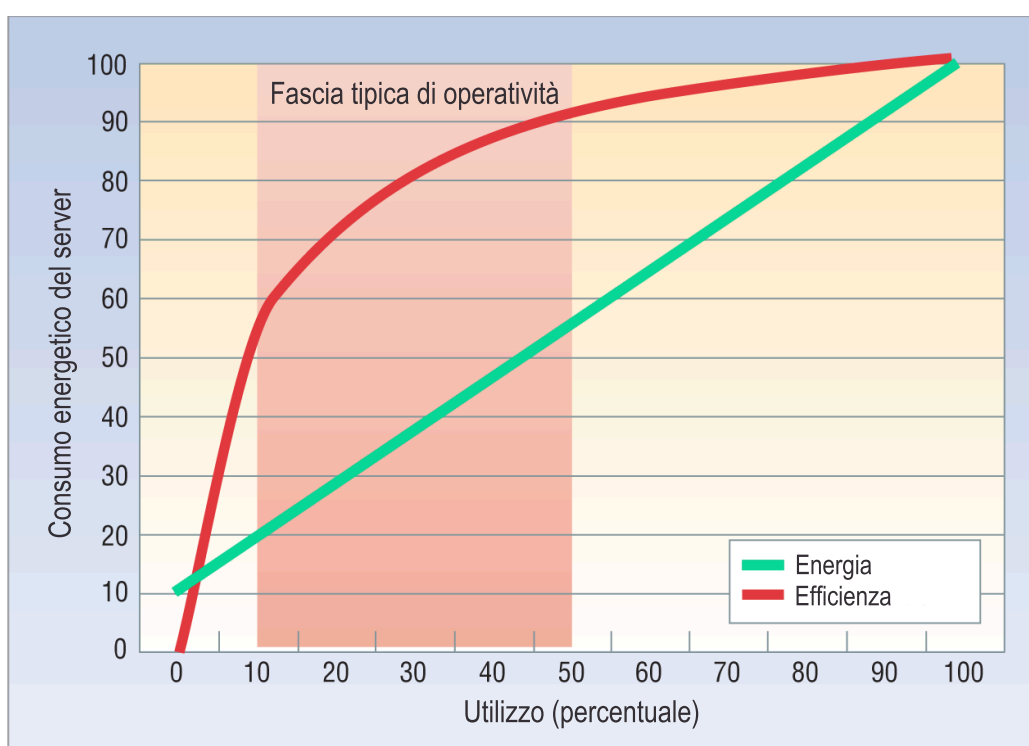


Figura 3.3: Consumo energetico del server [7]

diversi gradi intermedi che permettono una riduzione graduale delle funzioni di un componente fino alla completa disattivazione. I primi componenti che hanno implementato gli stati sono i processori, per ridurre i loro consumi visto l'alta incidenza sul consumo totale del sistema. Successivamente però, sono stati implementati anche negli altri componenti; si pensi ad esempio ai monitor dei computer che dopo un periodo di inattività, si spengono, oppure ai laptop che passano in stand-by se rimangono inattivi per un certo periodo. Gli unici effetti negativi di questo sistema si hanno nel momento in cui si passa dallo stato inattivo a quello attivo, nell'energia che si consuma nel tempo necessario per riattivare il sistema o il componente. Alcune tecniche, anche a livello software, cercano di ridurre i tempi di transizione, prevedendo l'uso del sistema e la finestra temporale di inattività, prima di scegliere in quale stato portare il sistema; infatti, in questo tipo di meccanismo è da specificare il periodo di tempo di inattività del sistema, o di un componente, prima di essere disattivato.

3.2.1.1 ACPI

Uno standard, che comprende più standard, molto diffuso per la gestione dell'energia è l'ACPI (Advanced Configuration and Power Interface). Questa specifica è stata rilasciata per la prima volta nel 1996 ed è stata sviluppata con la cooperazione tra diverse aziende del settore: HP, Intel, Microsoft, Phoenix e Toshiba. Nello standard sono stati definiti diversi strumenti e proprietà ma, l'aspetto principale è la definizione dei controlli necessari all'attuazione del controllo elettrico del sistema.

Nell'ACPI vengono definiti diversi gruppi di stati energetici in cui si possono trovare i vari componenti. Gli stati sono suddivisi per categorie: sette per l'intero sistema (S0-S6), quattro per il processore (C0-C3) e altri quattro per le periferiche (D0-D3).

Lo stato indicato con il numero zero indica lo stato di normale attività mentre con numero crescente viene indicato lo stato di progressiva inattività dei componenti con un consumo minore di energia.

Gli stati dell'intero sistema (S0-S6) sono stati inclusi in gruppi globali (G0-G3):

- G0 (S0) *Esecuzione*: è lo stato di normale attività del sistema. Il sistema operativo e le applicazioni sono in funzione e il processore esegue le istruzioni. Da questo stato è possibile, modificare lo stato del processore e delle periferiche (C0-C3, D0-D3).
- G1 *Idle*: raggruppa quattro stati del sistema (S1-S4):
 - S1 *Power On Standby*: è la modalità di riposo che consuma più energia. Restano attivi sia il processore sia la RAM ma la CPU non esegue istruzioni; inoltre, vengono disattivate le periferiche che non sono necessarie. Le macchine più recenti non prevedono più questo stato ed è utilizzato solo da quelle più vecchie al posto della S3.
 - S2: questo stato non è molto implementato, rispetto allo stato precedente, richiede meno energia in quanto viene disattivato anche il processore.
 - S3 *Sleep*: questo stato è chiamato anche *Standby*. L'unico componente che resta alimentato è la RAM che mantiene in memoria tutti i dati in esecuzione, permettendo all'utente di riprendere il lavoro precedente all'entrata in questo stato.
 - S4 *Ibernazione*: in questo stato tutto il contenuto delle ultime applicazioni in esecuzione precedentemente è spostato nel disco fisso permettendo la disattivazione anche della RAM. Rispetto agli stati precedenti, per transitare a questo stato è richiesto più tempo, necessario alla copia dei dati nel disco fisso.
- G2 (S5) *Soft Off* e G3 (S6) *Mechanical Off*: sono due stati molto simili, in entrambi, per riattivare il sistema è necessario un riavvio completo. Nel primo, tutti i componenti sono disattivati, tranne alcuni

per permettere il risveglio attraverso qualche input che può essere da tastiera, mouse, rete, ecc. Nel secondo, il sistema è completamente spento ed è lo stato in cui vanno i sistemi nel caso di blackout.

Gli stati relativi al processore che supporta le specifiche ACPI, sono:

- C0: che è lo stato normale di funzionamento, il processore è attivo ed esegue tutte le istruzioni.
- C1 *halt*: il processore non esegue le istruzioni, ma può essere riportato allo stato precedente in tempi rapidi.
- C2 *stop clock*: il processore continua a mantenere il contenuto dei registri per riprendere i calcoli, ma non esegue operazioni.
- C3: il processore è disattivato, non mantiene i dati e richiede più tempo per essere riattivato.

Gli stati delle periferiche sono compresi tra D0 completamente attivo, a D3 totalmente spento, gli stati intermedi sono diversi per ogni periferica.

3.2.2 Cambio dinamico di frequenza e voltaggio

Oltre alla transizione di stato, una seconda strategia per ottenere un consumo energetico proporzionale al carico di lavoro, consiste nell'adattare le performance del processore in base ai calcoli da eseguire. Questo è possibile cambiando in modo dinamico la frequenza (Dynamic Frequency Scaling), il voltaggio (Dynamic Voltage Scaling), oppure intervenendo su entrambi con la tecnica combinata Dynamic Voltage and Frequency Scaling (DVFS).

Il cambio dinamico di voltaggio è una tecnica di gestione energetica dove si interviene sul voltaggio dei componenti che in base alle circostanze viene diminuito o aumentato. Facendo funzionare i componenti ad un basso voltaggio, è possibile ridurre i consumi di energia ma, allo stesso tempo, si riducono anche le prestazioni del sistema. Invece, con il cambio dinamico di frequenza si interviene sulla frequenza di operatività del processore. Questi sistemi

sono ampiamente utilizzati, per esempio, negli apparecchi mobili o nei computer laptop che, funzionando a batteria hanno energia limitata e devono continuamente controllare i livelli di consumo.

Gli attuali processori possono operare a diverse velocità permettendo l'utilizzo di queste tecniche, altri componenti, invece, non offrono le medesime opportunità. Oltre alla riduzione dei consumi energetici e la possibilità di ottenere un consumo più proporzionale, con queste tecniche è possibile controllare la temperatura del processore, facendo calare anche i consumi necessari per il raffreddamento dello stesso.

Con l'utilizzo delle tecniche del cambio dinamico di frequenza e di voltaggio è possibile ottenere diversi vantaggi, ma vanno utilizzate con attenzione, perché aumentano l'usura dei componenti, e inoltre, con la diminuzione delle performance si provoca l'aumento del tempo di esecuzione delle operazioni fornite con il conseguente calo della qualità di servizio desiderata.

3.2.3 Gestione energetica

Nella gestione energetica è importante tenere conto dei tempi di risposta del servizio offerto dal data center che molto spesso viene garantito dagli operatori ai clienti. Durante l'esecuzione delle varie applicazioni, se alcune di queste devono essere svolte in tempi fissati, non è possibile disattivare o ridurre i consumi di alcuni componenti per offrire la massima capacità per l'esecuzione di questi processi, ma probabilmente nello stesso tempo è possibile disattivare altri server o componenti non fondamentali per la stessa quindi, è necessario un meccanismo di controllo e degli algoritmi studiati appositamente per questo.

L'obbiettivo è quello di progettare algoritmi energetico-efficienti che riescano a gestire il sistema, garantendo la massima efficienza e che permettano il consumo minimo di energia senza causare interruzioni o cali di qualità del servizio. Negli anni sono stati progettati diversi algoritmi che cercano di determinare il giusto periodo di tempo per transare il sistema da uno stato all'altro e che scelgono anche in quale stato deve essere portato il sistema,

per offrire il giusto compromesso tra consumi e servizio. Per trovare una soluzione, negli algoritmi studiati si assume che i consumi del sistema in stato operativo siano fissi e ci si concentra nel minimizzare i consumi durante i periodi di inattività.

Per risolvere questo problema, l'algoritmo deve trovare una soluzione, senza avere sufficienti informazioni e in tempo reale. I problemi di questo tipo sono definiti *online problem*. Durante l'esecuzione del sistema, l'algoritmo non conosce la durata del periodo di inattività, ma rispetto a questi difetti, per riuscire a trovare una soluzione ed avere una garanzia dell'efficacia della stessa, si ricorre ad un'analisi competitiva, comparando l'algoritmo *online* con un algoritmo *offline* ottimale, che risolve lo stesso problema con tutte le informazioni necessarie, calcolando un piano completo di transizioni tra gli stati per il sistema.

Nei sistemi a due stati un algoritmo offline ottimale, sapendo in anticipo la fine del periodo di inattività, confronta il consumo d'energia durante questo periodo con il costo che si crea nel riattivare il sistema, durante la transizione dallo stato a basso consumo a quello attivo. Se il costo del periodo di inattività è inferiore al costo dovuto alla transizione, si resta nello stato attivo, perché non c'è convenienza nel disattivare il sistema; altrimenti, si effettua la transizione e si passa allo stato a basso consumo.

Senza conoscere la durata del periodo di inattività, gli algoritmi online studiati cercano di seguire il comportamento dell'algoritmo ottimale, provando a determinare la durata del periodo di inattività prima di effettuare la transizione. Il tempo di inattività può essere determinato con una misura fissa oppure in base ad un calcolo delle probabilità. Nel caso della misura fissa, non si nota una grande efficienza e si rischia di dover riattivare il sistema subito dopo essere passati allo stato inattivo, pagando il costo di transizione; con un calcolo probabilistico, invece, si aumenta l'efficienza dell'algoritmo riuscendo a modellare anche specifiche situazioni con frequenti periodi molto lunghi o corti di inattività [6].

Nei sistemi con più stati, il confronto viene fatto tra i costi di transizione

da tutti gli stati di inattività presenti allo stato attivo e i consumi energetici durante il periodo di inattività in cui il sistema resta attivo. Un algoritmo che conosce in anticipo la durata del periodo, sceglierà di mettere il sistema nello stato che minimizza i costi di transizione. Mentre gli algoritmi online, anche in questo caso sono basati sull'algoritmo offline, cercando di stabilire la finestra di tempo in maniera prestabilita o tramite un calcolo delle probabilità e scegliendo di conseguenza lo stato che minimizza i costi di transizione. [6].

Per gestire il cambio dinamico di potenza e voltaggio, invece, gli algoritmi devono cambiare dinamicamente la velocità del processore per far eseguire le operazioni con il minimo consumo ma mantenendo sempre la qualità di servizio desiderata. È un problema di pianificazione (scheduling). Al momento dell'arrivo di un nuovo processo, un pianificatore (scheduler), deve decidere a quale velocità eseguirlo e con quale priorità rispetto agli altri. Di conseguenza, diversi ricercatori sono impegnati nel progettare e studiare algoritmi di scheduling efficienti, da usare per risolvere questo problema. Uno di questi si basa sui tempi massimi in cui deve essere eseguito un lavoro e la sua priorità, definendo in questo modo il livello di velocità del processore per quel lavoro e calcolando la densità dei vari processi in base al loro arrivo per l'esecuzione. Un processo con bassa priorità può essere eseguito in più tempo dal processore, e può essere sospeso, per essere ripreso, successivamente all'esecuzione di un processo con priorità più alta e che deve essere svolta in breve tempo, aumentando la velocità del processore e accorpendo ulteriori processi che arrivano con alta priorità.

Un esempio di scheduling di cinque processi si può vedere nella Figura 3.4, i processi sono rappresentati dai vari rettangoli. L'altezza dei rettangoli corrisponde alla velocità del processore per eseguire quel processo. Il tempo è indicato dall'asse orizzontale. Il processo A non è prioritario ed è possibile eseguirlo a in più tempo a bassa velocità. Il processo B arriva al tempo 3, deve essere eseguito entro il tempo 8 e quindi viene processato ad una velocità più alta, accorpendo il processo C al suo arrivo nel tempo 5. La stessa cosa accade ai processi D ed E, che vengono accorpati al momento dell'esecuzione.

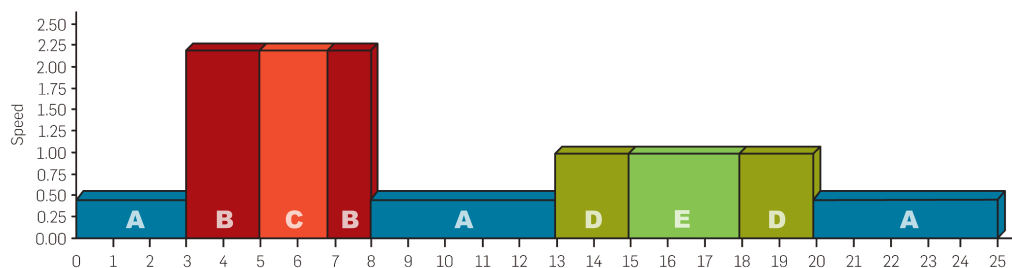


Figura 3.4: Pianificazione di cinque processi [6]

3.3 Virtualizzazione e consolidamento

Tra i metodi più utilizzati per ridurre i consumi delle apparecchiature IT del data center, c'è sicuramente la virtualizzazione, che consente la riduzione del numero dei server in esecuzione, aumentandone il livello di utilizzo e l'efficienza, anche per il motivo, come visto in precedenza, che i server in un tipico data center sono sotto utilizzati.

Possono essere virtualizzati diversi componenti hardware e software come: memorie, sistemi operativi o risorse di rete, attraverso la creazione di una macchina virtuale che si comporta a tutti gli effetti come un vero computer con un proprio sistema operativo, realizzando a tutti gli effetti una versione virtuale di una qualche risorsa del sistema, che normalmente è fisica. Il vantaggio di questa tecnica è che le macchine virtuali sono contenitori isolati rispetto all'hardware sottostate e questa proprietà permette l'esecuzione contemporanea di diverse macchine virtuali, con differenti applicazioni eseguite separatamente in ogni contenitore.

Con la virtualizzazione è possibile ridurre il numero dei server in un data center, condividendo le risorse di più utenti in un unico server, aumentando in questo modo la loro percentuale di utilizzo, ed eliminando i server non necessari permettendo un risparmio energetico immediato. Grazie a queste

possibilità, negli ultimi anni, sono aumentate la diffusione e gli interessi su questa tecnologia.

I tipi di virtualizzazione esistenti sono due: *hypervisor-based* come ad esempio: VMware, Microsoft Virtual Server e Xen, che permette la virtualizzazione di un intero sistema; e *operating system level virtualization* come OpenVZ, LinuxVServer e Solaris Zones, che permette l'esecuzione di diversi sistemi operativi su un'unica macchina reale. L'hypervisor (Virtual Machine Monitor) è un software che crea la base per ospitare le diverse macchine virtuali che sono eseguite simultaneamente. Il Virtual Machine Monitor opera in maniera trasparente, in modo che con la propria attività non pesi in modo eccessivo sulle prestazioni del sistema svolgendo un'attività di controllo sulle macchine virtuali in esecuzione e permettendo lo sfruttamento delle risorse in maniera condivisa. L'hypervisor controlla l'allocazione delle risorse necessarie di volta in volta, stabilendo in maniera dinamica lo spazio e le risorse necessarie, all'esecuzione delle applicazioni di ogni macchina virtuale, in maniera efficiente.

Altri vantaggi, che hanno fatto diventare la virtualizzazione una diffusa tecnologia nei data center, è la riduzione dei tempi di installazione e di manutenzione dei server, rendendo più semplice anche la loro gestione, che può essere eseguita da remoto dagli operatori dell'impianto.

Per utilizzare la virtualizzazione, nei data center già esistenti che non utilizzano tale tecnologia, è necessario tenere conto di diversi punti e considerare i rischi associati alla virtualizzazione, per i maggiori carichi che dovranno supportare i server. È importante controllare se l'infrastruttura possiede i requisiti minimi necessari a supportare l'aumento della potenza e se si hanno le necessarie capacità di raffreddamento, considerando inoltre, la possibilità di non tenuta di singoli server nel sopportare l'eccessivo carico.

Un modello a livelli proposto da vari ricercatori [18, 17], definisce cinque livelli che comprendono diverse azioni da eseguire dagli operatori di un data center per implementare e gestire la virtualizzazione nei loro centri in maniera sicura, ottenendo gli obiettivi desiderati. I passaggi suggeriti sono:

- Inventario delle risorse
- Scelta del tipo di virtualizzazione
- Potenziare l'hardware
- Controllo dell'infrastruttura
- Gestione della virtualizzazione

Il processo di inventario consiste nel controllo di tutte le risorse hardware e software presenti all'interno del data center, inclusi i server e le risorse collegate per il loro funzionamento. Nell'inventario vanno contati sia i server attivi sia quelli in stato inattivo e che non vengono utilizzati. Dopo l'analisi e la classificazione dei server e delle altre risorse, il secondo passaggio è quello di scegliere il tipo di virtualizzazione, tra i vari tipi esistenti, che si adatti e soddisfi i dati ricavati nel passaggio precedente. Il livello più difficile e importante è quello dell'*hardware maximization*. Siccome i server dovranno eseguire diverse macchine virtuali deve essere considerato il problema dell'hardware perché quello disponibile potrebbe non essere sufficiente per eseguire i maggiori carichi di lavoro. In questo caso, un cambiamento nell'hardware è obbligatorio, e andranno installati nuovi componenti che possano sostenere i nuovi carichi e che forniscano le migliori performance. Nel quarto passaggio, si deve verificare l'intero impianto e la struttura fisica del data center, controllando sia i componenti di distribuzione elettrica sia gli apparecchi di raffreddamento. Nell'ultimo passaggio, invece, sono incluse le decisioni finali che valutano la fattibilità economica della virtualizzazione e che vede coinvolti i manager del data center. Per prendere questa decisioni, si guarda a fattori come costi, ricavi, sicurezza, ma soprattutto se si riesce ad offrire il livello di servizio desiderato agli utenti finali.

La tecnologia della virtualizzazione promette di cambiare il modo in cui vengono offerti i servizi di un data center, facilitando la gestione dell'impianto, migliorando l'utilizzo delle risorse disponibili e l'efficienza delle

stesse, riducendo anche i costi complessivi avendo una buona flessibilità nella scelta di software e hardware.

Insieme alla virtualizzazione, con l'utilizzo delle virtual machine, è possibile attuare anche una politica di consolidamento, che permette di consolidare i carichi di lavoro dai server sotto utilizzati su poche macchine, o addirittura su una singola macchina. Ottenendo un beneficio, nella riduzione dell'hardware in funzione.

Una delle maggiori cause di inefficienza nei data center è l'energia che viene consumata dai server nei periodi di inattività. Come visto nei paragrafi precedenti, anche con bassi carichi di lavoro, l'energia consumata dai server e dai suoi componenti è di circa il 50% rispetto al massimo consumo. I server sono sotto utilizzati, perché le diverse applicazioni in esecuzione nel data center vengono divise tra i diversi server, per evitare i sovraccarichi e per riuscire a soddisfare tutta la domanda degli utenti finali, per mantenere un livello stabile e bilanciato dei carichi di lavoro in tutti i server.

Utilizzando il consolidamento, è possibile, senza abbassare la qualità del servizio, consolidare i carichi di lavoro delle varie macchine virtuali in un numero ridotto di server, aumentando la loro percentuale di utilizzato e l'efficienza del data center, potendo mettere i server rimasti vuoti e non utilizzati in stato di basso consumo.

Un consolidamento efficace non è semplice da ottenere: non basta concentrare tutte le applicazioni nel minor numero di server, perché si porterebbe l'utilizzo di ogni risorsa nei server al 100% e tale approccio non è consigliabile perché si avrebbe un aumento dell'energia utilizzata per ogni singolo servizio fornito, e si rischierebbe il danneggiamento dei componenti hardware; ma soprattutto un consolidamento efficiente, deve ottimizzare l'utilizzo di energia fornendo la qualità di servizio voluta. Per ottenere queste caratteristiche, un metodo di consolidamento deve considerare diversi aspetti.

Nell'effettuare il consolidamento dei dati, si deve decidere attentamente quali macchine virtuali consolidare nello stesso server, considerando ad esempio le applicazioni a più livelli (multi-tiered), che hanno i vari livelli in ese-

cuzione su diversi server e in macchine virtuali separate nello stesso momento, con un'incidenza diversa sul carico di lavoro complessivo, queste, potrebbero essere consolidate sullo stesso server, per migliorare i tempi di risposta; l'allocazione dei vari livelli in un numero ridotto di server deve permettere che le applicazioni continuino a operare al livello di performance desiderato, senza incorrere in interruzioni di servizio. Il secondo aspetto è considerare che esiste un punto ottimale di performance e di energia, questo perché il consolidamento può portare ad un degrado della performance che può causare l'aumento del tempo di esecuzione dei processi, perdendo così il risparmio energetico recuperato dalla riduzione dei periodi inattivi. Quindi determinare il punto ottimale delle performance in base al carico di lavoro totale è fondamentale per ottenere un consolidamento ad alta efficienza energetica, di conseguenza un algoritmo di consolidamento, ha come obiettivo quello di minimizzare il consumo di energia tenendo conto del vincolo di prestazioni da soddisfare. Con gli sviluppi delle tecnologie di virtualizzazione, è stato possibile semplificare il trasferimento di processi tra i server, riuscendo a spostare intere macchine virtuali mentre sono in esecuzione.

Sono diversi gli algoritmi proposti per effettuare un consolidamento efficace, il primo [16] si basa sul calcolo euristico che stima la distanza dal livello ottimale di capacità da raggiungere. Questo algoritmo nasce dall'idea di riuscire ad ottimizzare lo spazio e di utilizzare nel modo migliore le capacità disponibili. All'arrivo di ogni richiesta, si calcola il punto ottimale da raggiungere. Il server in cui allocare la macchina virtuale viene individuato in base alla distribuzione dei carichi di lavoro attuali presenti nei server; si calcola la nuova distanza dal punto ottimale, in tutti i server in base al peso dell'applicazione da sistemare, e verrà scelto il server in cui questa distanza è la minore. In questo modo è possibile utilizzare nel modo migliore le capacità disponibili. Nel caso in cui una richiesta non possa essere allocata, si dovrà mettere in funzione un nuovo server, ricalcolando, a sua volta, il livello ottimale dei server riallocando i processi. Questo sistema, ha bisogno di un decisore centrale che decide l'allocazione ad ogni ingresso, ogni volta che ar-

riva un nuovo processo, ma un sistema ottimale dovrebbe riuscire a fare una scelta anche con richieste simultanee. Esistono, infatti, anche algoritmi più efficienti che mostrano una maggiore elasticità.

I risultati che si possono ottenere con il consolidamento cambiano anche tra le diverse tecnologie esistenti di virtualizzazione. In base alle singole esigenze di un data center e alle applicazioni in esecuzione, alcune si comportano in modo migliore. In uno studio [12], sono state testate due diverse tecnologie di virtualizzazione: Xen e OpenVZ, con diverse configurazioni e consolidando applicazioni di diverso tipo con livelli multipli. Lo studio dimostra che i tempi di risposta in Xen aumentano del 400% rispetto al 100% di OpenVZ, per colpa dell'eccessiva virtualizzazione del primo. Questo dimostra che per un data center che vuole ridurre i costi attraverso il consolidamento deve effettuare una scelta corretta del tipo di virtualizzazione da usare in base alle proprie esigenze e funzioni da svolgere.

Questa tecnologia implica un investimento iniziale dovuto a tutti i passaggi da effettuare per adeguare il data center alla virtualizzazione, ma in tempi medi o brevi si riesce ad ottenere un ottimo risparmio nei consumi, infatti è la tecnologia che negli ultimi anni ha avuto il maggiore successo e per il momento non si vedono altre tecnologie così efficaci per la riduzione dei costi nei data center.

Capitolo 4

Casi di studio

Come visto nei capitoli precedenti, per risolvere il problema del risparmio energetico, l'obiettivo è quello di aumentare l'efficienza del data center attraverso vari sistemi, e tutte le aziende che operano in questo settore sono impegnate per risolvere questo problema e trovare nuove soluzioni che le aiutino a ridurre i consumi dei loro impianti. Gli operatori dei data center possono aumentare l'efficienza attraverso il consolidamento, la virtualizzazione e l'utilizzo di risorse efficienti aumentando la capacità di ogni singola unità IT. Per migliorare l'efficienza si comincia dalla progettazione o dalla modifica, se già esiste, dell'intera infrastruttura, concentrando le attenzioni dall'illuminazione, all'impianto elettrico ed ai sistemi di raffreddamento, senza dimenticare anche gli ambienti adibiti ad uffici e laboratori per i tecnici che gestiscono il centro.

In questo capitolo verranno mostrati i progressi che hanno ottenuto i maggiori operatori del settore nel migliorare l'efficienza energetica dei loro data center e quali sono le tecniche più utilizzate per questo obiettivo, e si mostreranno alcune tecniche e iniziative in fase di studio direttamente dalle aziende.

4.1 Analisi e tecniche in uso

In base ai dati pubblicati dalle diverse aziende e da organizzazioni ambientali è possibile confrontare le tecniche che sono utilizzate dai diversi operatori di data center, per vedere come hanno affrontato il problema del consumo energetico, classificando le differenze e le tecniche che hanno un'efficacia migliore.

Per un'analisi più corretta, si sono individuate delle categorie (dette scenari), in cui si possono trovare i data center in base ai progressi ottenuti nel risparmio energetico. Questa classificazione è stata definita dall'agenzia americana EPA, all'interno del programma Energy Star [8]. Ogni categoria indica una maturità tecnologica delle apparecchiature IT e dell'infrastruttura del data center; tra le cinque categorie due, *storico* e *attuale*, non verranno elencate perché sono calcolati su misurazioni passate e sugli attuali trend di efficienza energetica; mentre gli altri scenari che prevedono miglioramenti dei data center, sono:

- **Improved operation:** per essere in questa categoria si richiedono miglioramenti di natura organizzativa, che richiedono un piccolo investimento per gli operatori di data center. Rappresenta la categoria più bassa per ottenere un livello minimo di efficienza e il valore del PUE non deve superare 1,7;
- **Best practice:** in questa categoria si richiede un minimo utilizzo delle migliori tecniche a basso consumo esistenti e che sono utilizzate al giorno d'oggi. I data center in questa categoria devono attuare un moderato consolidamento dei server e mostrano una migliore efficienza energetica e gestione del sistema di raffreddamento, il PUE massimo per questi data center è di 1,3;
- **State of the art:** nella categoria “stato dell'arte” il PUE non deve superare il valore di 1,2. I data center in questa categoria vengano portati al massimo dell'efficienza possibile, adottando un consolidamento

più aggressivo, utilizzando le migliori pratiche di gestione energetica disponibili al momento.

Per posizionare i data center in una di queste categorie, nelle analisi fatte nel 2007 dell'EPA [10], gli impianti vengono suddivisi nelle loro parti principali, e analizzate singolarmente, registrando i consumi e le tecnologie impiegate. Dai dati in possesso si può notare che la maggior parte delle aziende rilevate, si trovano nello scenario *Best Practice* e *State of the Art*, per la parti relative alla distribuzione elettrica e per gli impianti di raffreddamento. Mentre per quanto riguarda la parte delle attrezzature IT, si nota un maggiore equilibrio, con solo un data center che riesce ad ottenere la posizione massima nello scenario *State of the Art*.

Initiative	Improved Operations	Best Practice	State of the Art
Data Center Site Infrastructure			
Data Center Cooling			
Data Center Airflow Management		XX	
Free Cooling in Large Scale Data Centers	X		XX
Data Center Cooling Optimization		X	
High Efficiency Chilled Water Systems		X	
Modular Cooling Systems		XX	XXX
Wireless Sensor Network Adaptive Cooling			X
Data Center Power Distribution			
High Efficiency Power Transformation		XX	
High Voltage AC Power			X
High Efficiency Stand-by Power Systems		XX	
IT Infrastructure			
Consolidation and Optimization			
IT Computing Resource Optimization		X	X
IT Consolidation and Virtualization	X	X	
Server Power Characterization & Modeling	X		

Figura 4.1: Classificazione dei data center analizzati [10]

Nel reparto IT la strategia che viene più utilizzata per aumentarne l'efficienza, è il consolidamento. Per i motivi visti in precedenza, attraverso questa tecnica si riesce ad utilizzare un numero ridotto di server in maniera più efficiente e performante, riducendo sia il costo dell'acquisto di server sia

i consumi derivanti dal sistema di condizionamento. La *Sun Microsystems*, ad esempio, attraverso il consolidamento è riuscita a dimezzare il numero di server installati e a ridurre di un terzo i dispositivi di storage, risparmiando circa il 60% sui consumi energetici ed evitando le spese per la costruzione di nuovi impianti. Risultati simili gli ha avuti anche *Symantec* riuscendo a diminuire i consumi ad un terzo rispetto all'anno 2006, senza dover sostituire il 12% dei server nei suoi impianti[10].

Dai risultati disponibili, si ricava, come la virtualizzazione non sia ancora sfruttata adeguatamente e che venga utilizzata solo da una piccola parte dei data center analizzati, con l'attuazione di un consolidamento più aggressivo, invece, sarebbe possibile raggiungere il livello di "stato dell'arte" o addirittura superarlo.

Dalle specifiche dell'EPA, per avere un data center efficiente, non si dovrebbe superare il valore 1,20 del PUE, in realtà i trend del 2007 mostrano che i data center ottengono in media un PUE di 1,90 con quasi la metà dell'energia assorbita, consumata da apparecchiature non IT.

Esistono comunque delle eccezioni, che mostrano un'elevata efficienza strutturale. Il data center più efficiente è di *NetApp* che già nel 2007 poteva raggiungere un PUE di 1,10. Questo valore è raggiunto grazie ad una serie di fattori, l'illuminazione occupa una piccola percentuale dei consumi del data center (0,01 del PUE) ed è in funzione un UPS efficiente di ultima generazione, inoltre, il data center è progettato in modo ottimale e dimensionato correttamente. Il data center di *Yahoo! Wenatchee*, invece, è riuscito ad ottenere buone performance, dal punto di vista energetico, eliminando dal trasformatore i voltaggi che non servivano per fornire direttamente quello corretto alle attrezzature IT non subendo, in questo modo, le perdite energetiche che si hanno durante la trasformazione nei voltaggi intermedi, raggiungendo lo 0,015 su un PUE totale, non certo ottimale, di 1,74. Per il sistema di raffreddamento, invece, il data center che riesce ad ottenere migliori risultati è sempre quello di *NetApp* che implementa un sistema di raffreddamento cosiddetto "libero" (free cooling), che, essendo costruito in

una zona secca e fredda, sfrutta la temperatura dell'ambiente esterno per raffreddare l'aria calda interna, senza l'utilizzo di pompe o ventole con un conseguente consumo minimo di energia e pesa sul pue totale solo..

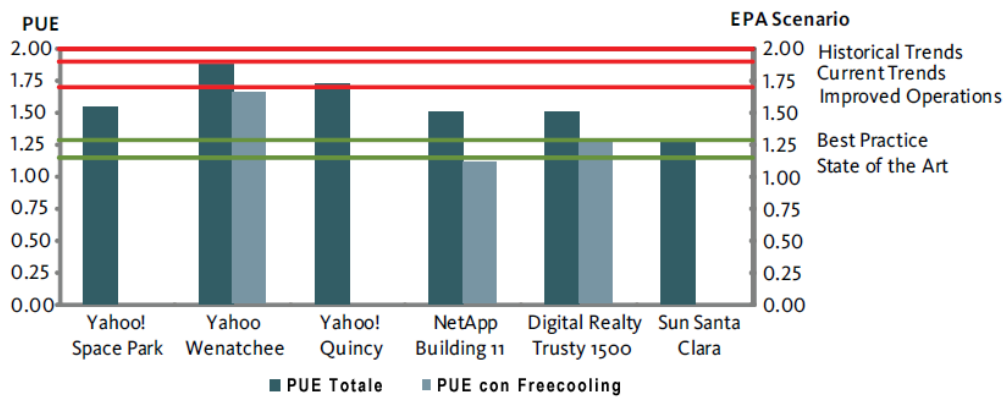


Figura 4.2: PUE totale con utilizzo del sistema di raffreddamento libero [10]

4.2 Risultati recenti

Da questi dati si può ricavare che le migliori tecniche in uso sono il consolidamento attraverso la virtualizzazione, e la scelta di componenti efficienti, ma senza trascurare la scelta della posizione del data center e il giusto dimensionamento.

Anche analisi più recenti confermano queste strategie e la media degli ultimi anni del PUE è di 1,5. I data center di *Google* [4] attualmente riescono a raggiungere un valore medio di 1,2 con il record del più recente a 1,1. Google, ha puntato molto sulla progettazione dei data center e dei server installati, implementando diverse tecniche per riuscire ad ottenere questi risultati. Hanno modificato i server per minimizzare i consumi, eliminando tutti i componenti non fondamentali, come i chip grafici; oppure attuando modifiche per aumentare l'efficienza, creando supporti specifici che permettono di montare un numero minore di ventole per raffreddare al meglio i

componenti, senza sprechi di energia. Google stima di risparmiare per ogni server 500kWh all'anno rispetto ad un tipico server, e insieme ai maggiori produttori di hardware stanno incentivando la costruzione e l'utilizzo di componenti energetico efficienti. Per quanto riguarda il raffreddamento in alcuni data center, Google, per risparmiare dai consumi di ventole e pompe ha combinato, in aggiunta ai classici sistemi, un complesso impianto di raffreddamento "libero" che riesce a raffreddare l'acqua necessaria, attraverso torri di raffreddamento, sfruttando l'evaporazione della stessa che si riscalda con il funzionamento del data center; oppure, come nel centro di Hamina in Florida, utilizzando l'acqua prelevata dall'ambiente esterno (da un fiume) per rinfrescare l'aria all'interno.

Tra le altre aziende importanti, abbiamo: *Yahoo!* che dimostra l'importanza della scelta di un ambiente favorevole in cui costruire un data center; trovandosi vicino alle Cascate del Niagara, da cui ricava il 90% dell'energia utilizzata [1], riesce ad ottenere un PUE di 1,08 [3], quasi alla soglia della perfezione.

Chi dimostra più efficienza tra le aziende analizzate, è *Facebook* con il data center di Prineville riuscendo a raggiungere un PUE di 1,07 che utilizza le tecnologie del progetto studiato da loro stessi: Open Compute [2].

4.2.1 Open Compute

Questo progetto è stato creato con l'intenzione di definire delle specifiche per la costruzione di data center efficienti che siano poco costosi e sostenibili. La creazione di questo progetto ha lo scopo di rendere disponibili a tutti le innovazioni trovate nei data center, in modo che tutte le industrie possano implementarle e aumentare l'efficienza dei loro impianti.

Durante la costruzione data center di Prineville da parte di Facebook, si sono utilizzati server di nuova concezione, che rispetto a quelli tradizionali vede rimosse tutte le parti non necessarie, come le rifiniture esterne e utilizzando meccanismi che non prevedono viti. In questo modo si utilizza il 20% di materiali in meno, risparmiando sulla costruzione stessa del server.

Nello stesso data center vengono montate solo le attrezzature fondamentali con l'utilizzo di componenti elettrici e di raffreddamento studiati apposta per l'impianto. Con questi accorgimenti, questo data center riesce ad ottenere un'efficienza superiore del 38% rispetto ai data center comuni che svolgono le stesse operazioni, risparmiando il 24% dei costi complessivi.

Attraverso la diffusione di questa tecnologia, con il supporto di aziende di hardware, sarà possibile implementare le stesse tecniche anche in altri data center per ottenere gli stessi benefici.

Conclusioni

Dopo aver analizzato la struttura dei data center e la loro diffusione tramite i servizi web negli ultimi anni, in questo elaborato, si è discusso del problema energetico che comporta l'utilizzo di queste strutture, mostrando il livello di incidenza delle varie parti dei data center sui consumi. Si è rilevato come la parte con i consumi maggiori sia quella relativa ai server e alle apparecchiature IT, discutendo della riduzione di questi consumi senza subire una perdita nella qualità del servizio fornita.

Come abbiamo visto, è possibile ottenere discreti risultati attraverso una corretta progettazione del sistema, insieme all'utilizzo di componenti efficienti e dalla costruzione di impianti elettrici e di raffreddamento studiati appositamente per ogni data center. I maggiori risultati si sono ottenuti nella riduzione del numero dei server attraverso tecnologie come il consolidamento e la virtualizzazione, riuscendo ad aumentare in maniera consistente l'efficienza dei data center, dimostrando come alcune aziende abbiano ridotto i consumi di oltre il 50% attraverso l'uso di queste tecnologie.

Abbiamo visto, però come non tutti i data center mostrino livelli di efficienza sufficienti, e come alcune delle più grandi imprese del settore si siano impegnate nella diffusione dell'utilizzo di tecniche energetico efficienti, per fare in modo che anche le imprese minori possano ottenere gli stessi risultati e a ridurre ulteriormente i consumi degli impianti. Con la collaborazione tra le più grandi imprese e i maggiori produttori di hardware, sarà possibile aumentare l'efficienza energetica dei data center e dei componenti hardware portando notevoli vantaggi, sia a livello economico sia a livello ambientale.

Bibliografia

- [1] <http://green.yahoo.com/blog/ecogeek/1125/yahoo-data-center-will-be-powered-by-niagara-falls.html>.
- [2] <http://opencompute.org/>.
- [3] <http://pressroom.yahoo.net/pr/ycorp/508872.aspx>.
- [4] <http://www.google.com/intl/en/corporate/datacenter/>.
- [5] *Recommendations for Measuring and Reporting Overall Data Center Efficiency*, July 2010.
- [6] Susanne Albers. Energy-efficient algorithms. *Commun. ACM*, 53:86–96, May 2010.
- [7] L.A. Barroso and U. Holzle. The case for energy-proportional computing. *Computer*, 40(12):33–37, dec. 2007.
- [8] ENERGY STAR Program Environmental Protection Agency. Report to congress on server and data energy efficiency. Technical report, 2007.
- [9] EPA. Epa report to congress on server and data center energy efficiency. Technical report, U.S. Environmental Protection Agency, 2007.
- [10] Silicon Valley Leadership Group. Data center energy forecast. Technical report, 2008.
- [11] Vales K.A. McNamara W., Seimetz G. Best practices for creating the green data center. *The Outsourcing Institute*, 2008.

-
- [12] Pradeep Padala, Xiaoyun Zhu, Zhikui Wang, Sharad Singhal, Kang G. Shin, Pradeep Padala, Xiaoyun Zhu, Zhikui Wang, Sharad Singhal, and Kang G. Shin. Performance evaluation of virtualization technologies for server consolidation. Technical report, 2007.
 - [13] Shah A. Patel, C. D. Cost model for planning, development and operation of a data center. Technical report, 2005.
 - [14] Webb T. Rad P., Thoene M. Best practices for increasing data center energy efficiency. *Dell Power Solutions*, 2008.
 - [15] N. Rasmussen. Implementing energy efficient data centers. *APC White Paper*, 2006.
 - [16] Shekhar Srikantaiah, Aman Kansal, and Feng Zhao. Energy aware consolidation for cloud computing. In *Proceedings of the 2008 conference on Power aware computing and systems*, HotPower'08, pages 10–10, Berkeley, CA, USA, 2008. USENIX Association.
 - [17] Mueen Uddin and Azizah Abdul Rahman. Server consolidation: An approach to make data centers energy efficient and green. *CoRR*, abs/1010.5037, 2010.
 - [18] Mueen Uddin and Azizah Abdul Rahman. Virtualization implement model for cost effective and efficient data centers. *IJACSA*, 2, 2011.