



# INFRASTRUTTURE TECNOLOGICHE GLI IMPATTI DELL'AI

---

***L'Intelligenza Artificiale per il Nucleare e il Nucleare per l'intelligenza artificiale***

*22 Gennaio 2026*

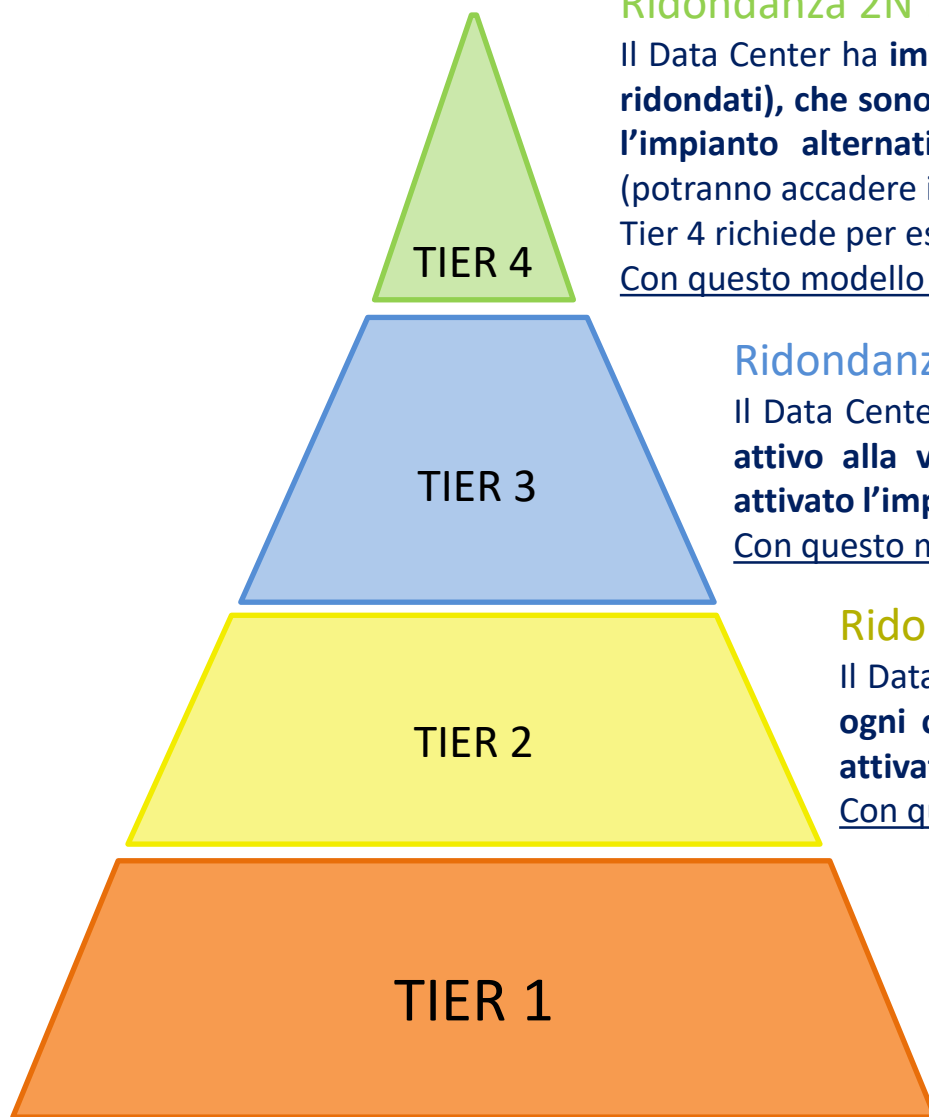
---

*Ing. Angelo Priolo*



# INDICE DEGLI ARGOMENTI

- La Ridondanza dei Data-Center
- Panoramica Impianto Elettrico
- Il Problema del Calore nei Data-Center
- Efficienza Energetica dei Data-Center
- Caratteristiche delle Sale Server
- Sistemi di Condizionamento Innovativi nell'era dell'AI
- Conclusioni dell'Impatto dell'AI nei Data-Center



### Ridondanza 2N : $2(N + 1)$ – Fault Tolerant

Il Data Center ha **impianti interamente duplicati, ognuno di capacità sufficiente (anche con componenti ridondati), che sono contemporaneamente attivi**, quindi in caso di guasto o manutenzione programmata **l'impianto alternativo consente il funzionamento del Data Center senza interruzioni di servizio** (potranno accadere interruzioni di servizio giusto per allarmi antincendio per esempio).

Tier 4 richiede per esempio apparati ICT con doppia alimentazione.

Con questo modello si prevedono disservizi dell'ordine di minuti per anno

### Ridondanza N+1 – 2N – Contemporaneamente manutenibile

Il Data Center ha **impianti duplicati ognuno di capacità sufficiente, ma un solo impianto attivo alla volta**, quindi in caso di guasto o manutenzione programmata **può essere attivato l'impianto alternativo**, riducendo ulteriormente il fuori servizio del Data Center.

Con questo modello si prevedono disservizi inferiori poche ore per anno

### Ridondanza N + 1 – Componenti ridondati

Il Data Center ha gli impianti con componenti ridondati (**almeno uno in più del necessario per ogni componente**), quindi in caso di guasto o manutenzione programmata **possono essere attivati i componenti ridondati riducendo il fuori servizio del Data Center**.

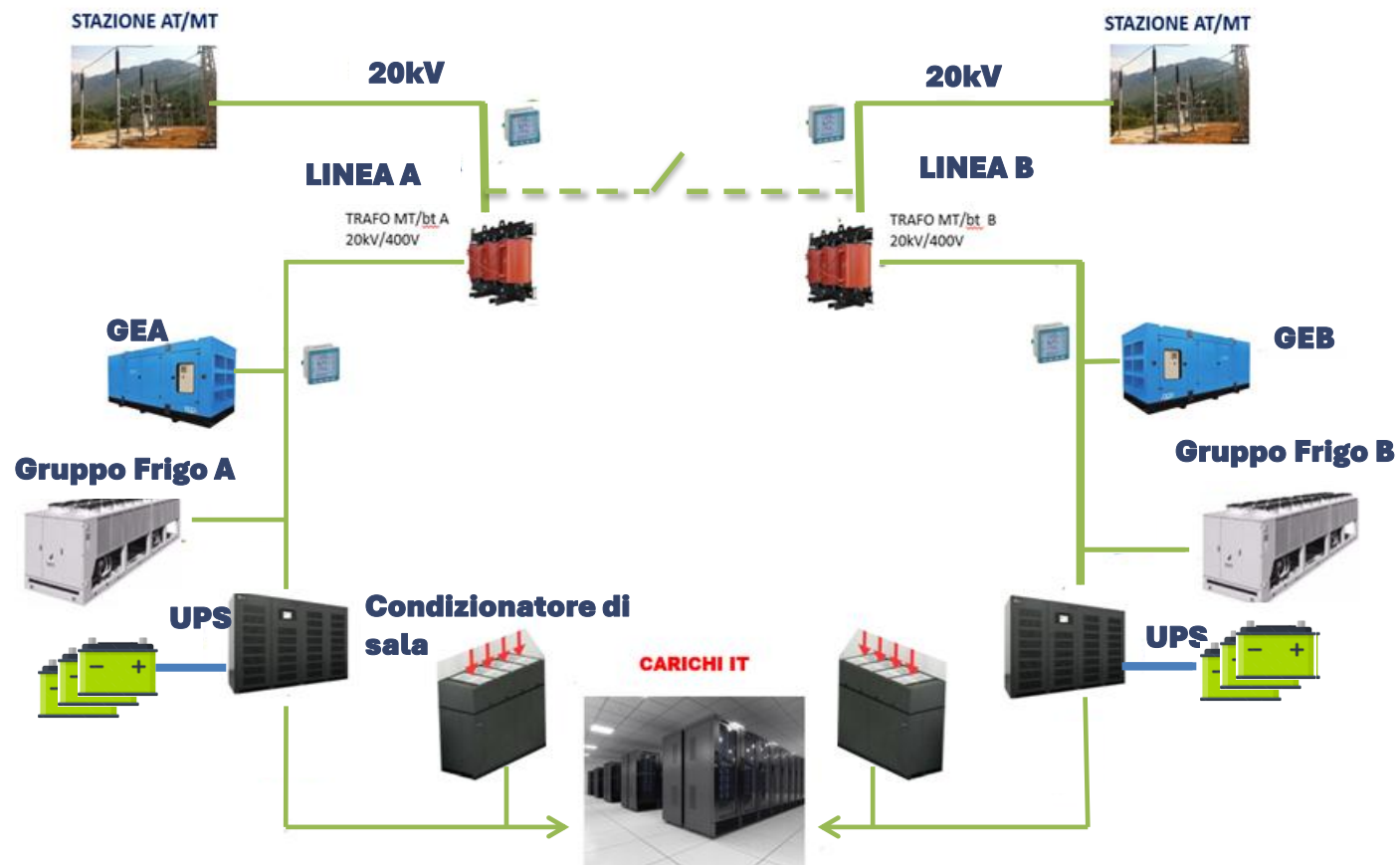
Con questo modello si prevedono disservizi inferiori ad un giorno per anno

### Ridondanza N – Nessun componente ridondato

Ogni componente o impianto del Data Center è singolo, quindi **ogni guasto crea un fuori servizio del Data Center** più o meno grave, così come le attività di manutenzione programmata, per il tempo necessario all'intervento.

Con questo modello si prevedono fuori servizio dell'ordine dei giorni per anno

## Panoramica Impianto Elettrico

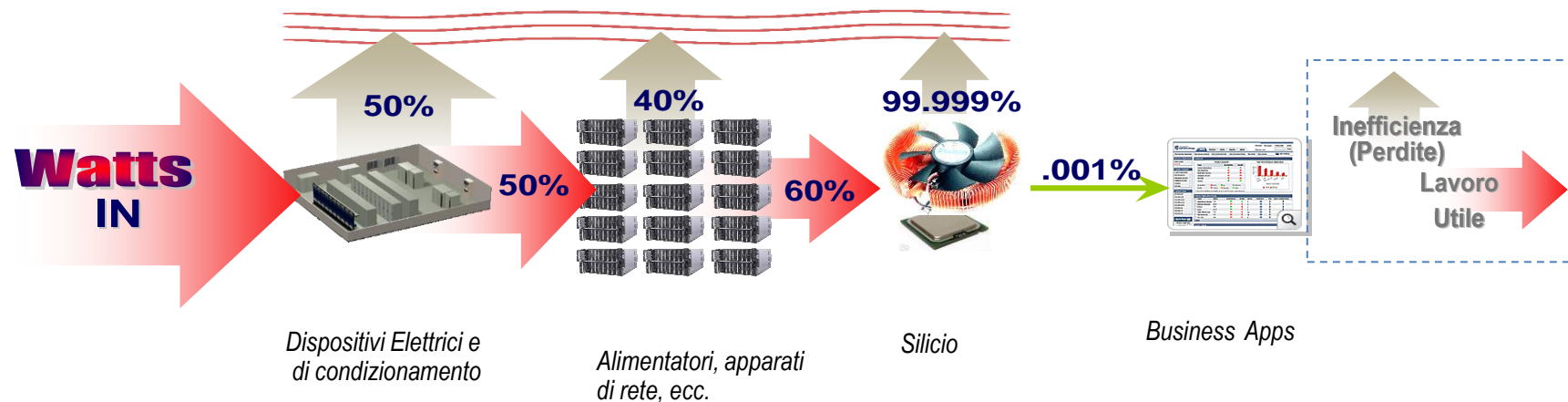


I GRANDI DATA CENTER SONO TIPICAMENTE DI CLASSE TIER III / IV QUINDI CON IMPIANTI DUPLICATI OGNUNO DI CAPACITA' SUFFICIENTE PER IL FUNZIONAMENTO DEL DATA CENTER

## Il Problema del Calore nei Data-Center

I Data Center sono dei divoratori di energia, che per la maggior parte viene utilizzata dagli impianti di supporto o dispersa sotto forma di calore.

Della parte di energia che arriva ai sistemi ICT il 99,999% è disperso in calore e solo lo 0,001% è utilizzato per la elaborazione delle informazioni.



Il calore generato deve essere estratto dai sistemi e dissipato, altrimenti si provocano malfunzionamenti e nei casi peggiori anche la distruzione dei circuiti.

Nel 2007 nasce il consorzio Green Grid che definisce il PUE (Power Usage Effectiveness) attualmente armonizzato nella norma ISO/IEC 30134-2.

Tale parametro consente di misurare l'efficienza energetica di un centro di calcolo ed è data dal rapporto fra tutta la potenza assorbita dal Data Center rispetto a quella assorbita dai soli apparati IT:

$$PUE = \frac{Potenza\ totale}{Potenza\ IT}$$

Minore è il valore del PUE ( Più vicino ad 1) e maggiore è l'efficienza del Data Center

Il PUE si migliora ottimizzando l'efficienza di tutti gli impianti di servizio del Data Center, ma principalmente degli UPS e del Condizionamento che sono impianti energivori e sempre attivi.

Un approccio ampiamente utilizzato è quello di aumentare la temperatura di esercizio dell'impianto di condizionamento dai 18 °C - 20 °C del passato a temperature più alte, per esempio 26 °C, che sono accettabili dagli attuali sistemi IT

Ad oggi sono state introdotte altre metriche:

WUE: Water Usage Effectiveness

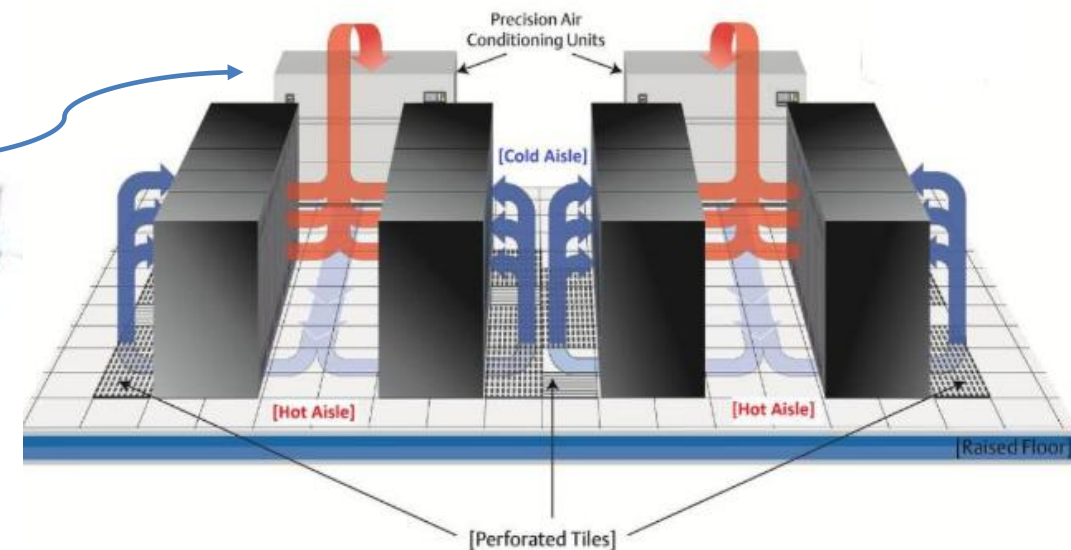
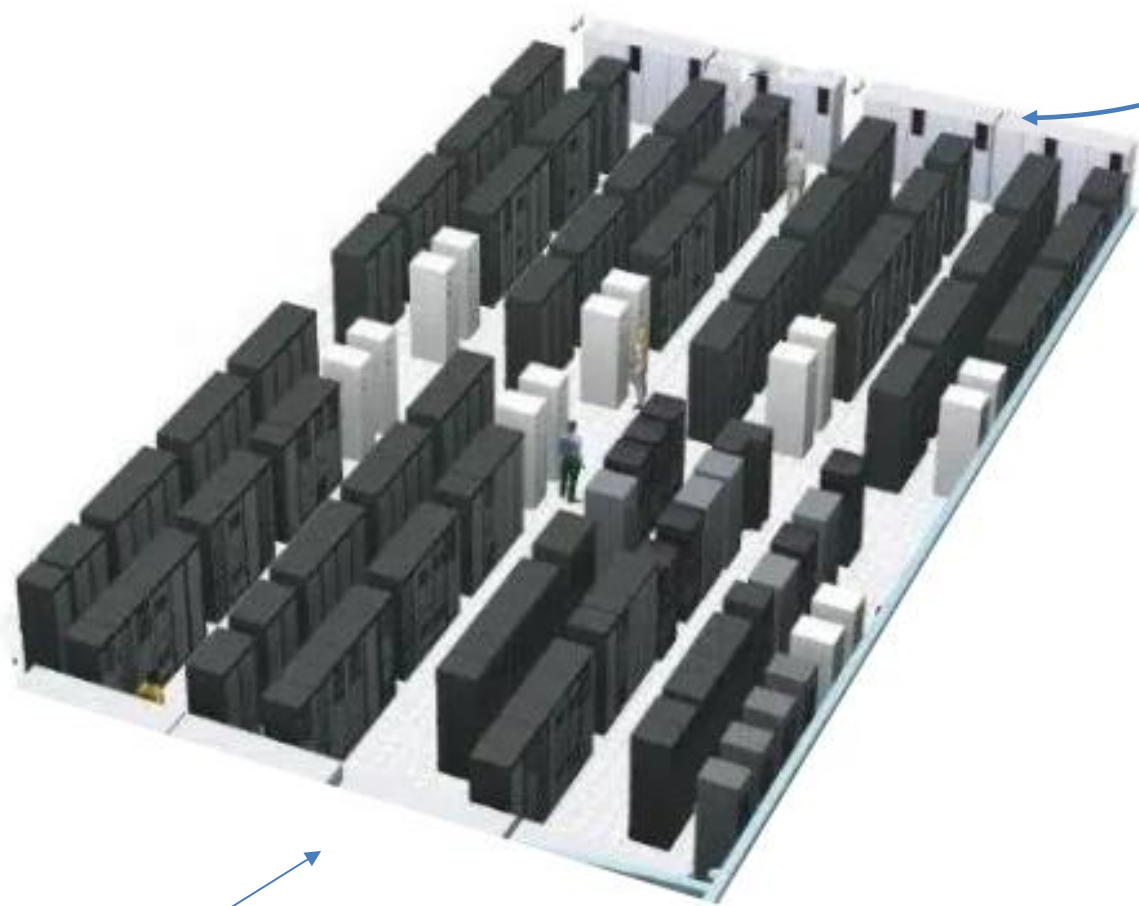
DCRE: Data Center Reusable Effectiveness che comprende tutte le risorse consumate dai Data Center

L'inefficienza di un DATACENTER è dovuta a impianti e macchine vetuste.  
L'abbattimento di forti inefficienze di un DC può essere attuato mediante la sostituzione di uno o di tutti e tre i seguenti elementi chiave che contribuiscono all'inefficienza del sito

- ✓ SISTEMI UPS
- ✓ SISTEMI DI CONDIZIONAMENTO
- ✓ EFFICIENTAMENTO DI ALTRI IMPIANTI A SERVIZIO DEL DATA CENTER  
(impianto di illuminazione, impianto idronico ecc.)

Dal punto di vista impiantistico possono essere presi dei provvedimenti che consentono un'ulteriore miglioramento dell'efficienza relativamente alla Power Quality

- ✓ Aumento del fattore di potenza  $\cos\varphi \geq 0.95$
- ✓ Riduzione del THDi%
- ✓ Riduzione delle perdite di potenza negli impianti
- ✓ Flicker Transitori Sovratensioni

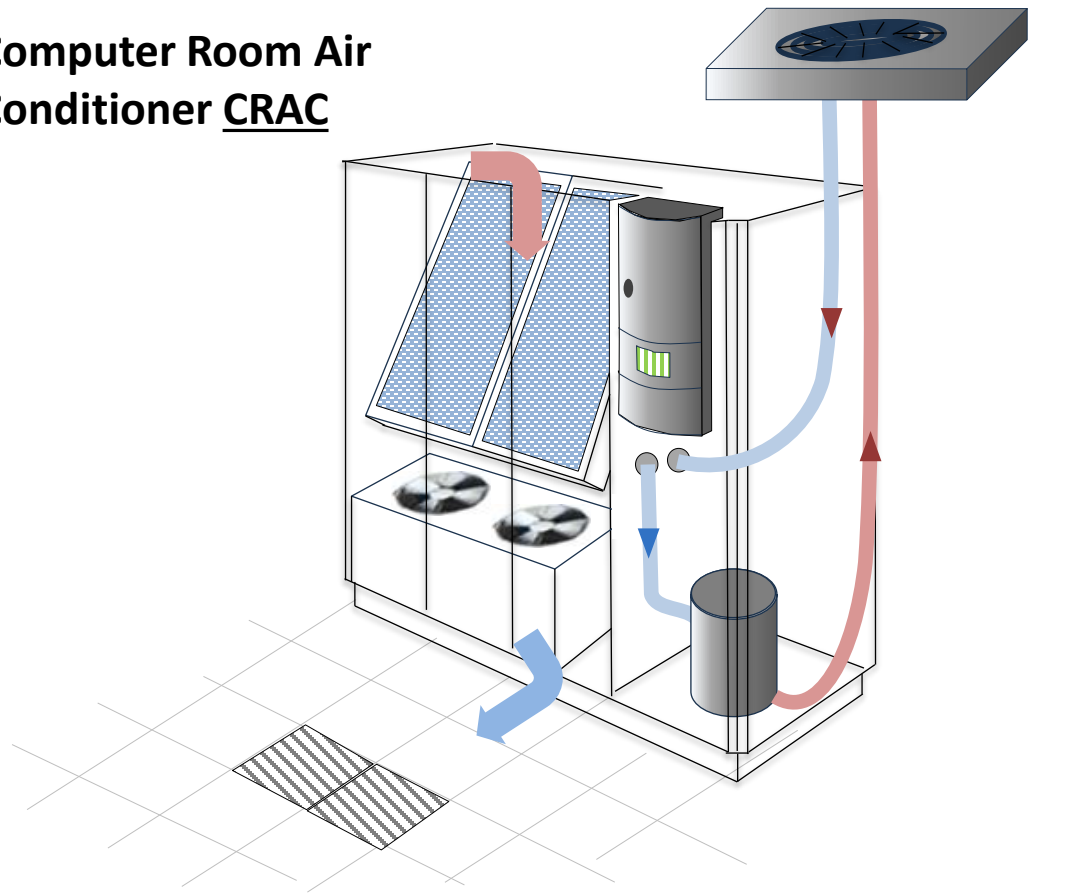


Schema di raffreddamento classico con condizionamento perimetrale e aria fredda convogliata al di sotto del pavimento flottante «Plenum» e rack disposti back-to-back in modo da creare corridoi caldi e freddi.

Pavimento Flottante



### Computer Room Air Conditioner CRAC



Condizionamento di precisione a Gas Refrigerante (ad espansione diretta DX) costituito da tante unità indipendenti, ognuna col proprio compressore ed unità esterna

Le unità sono disposte nel perimetro delle sale

L'aria calda viene aspirata dall'alto attraverso la griglia fredda e viene soffiata al di sotto del pavimento flottante (plenum)

Più adatto per data center medio piccoli.

### Impianto di condizionamento ad acqua refrigerata.

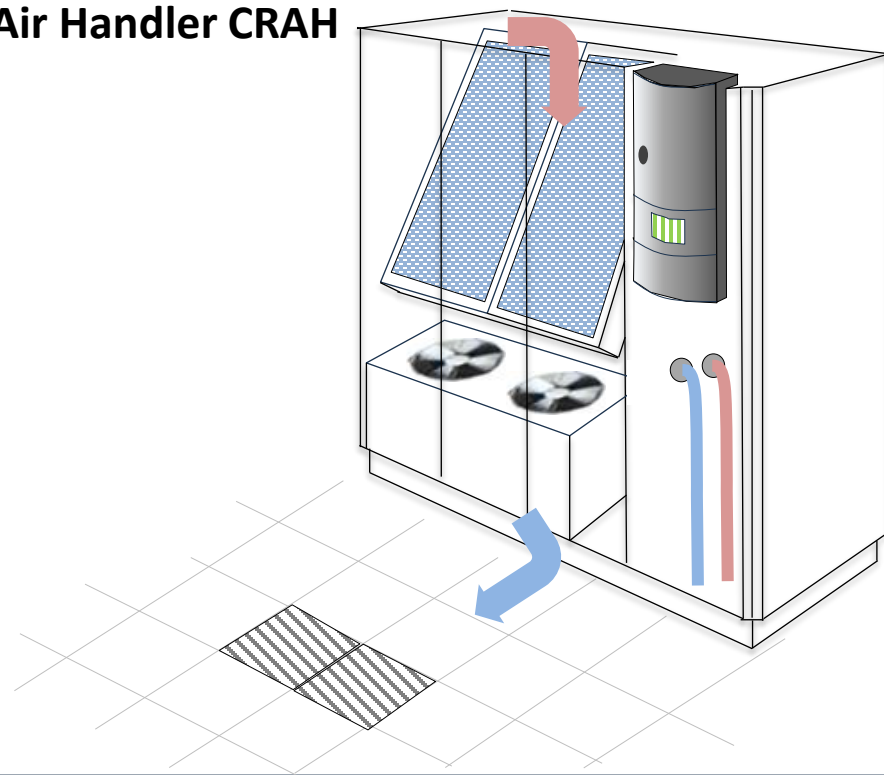
Si preferisce per grandi Data Center:

- sono più efficienti su larga scala
- L'acqua può essere prelevata da fonti di acqua fredda (acqua di falda)
- Minore impatto ambientale per la riduzione di gas refrigeranti
- Free Cooling che con un ulteriore impianto consente di raffreddare l'acqua se la temperatura esterna è bassa (es. inverno)

-----

Chiaramente è richiesto un impianto complesso, con tubazioni di mandata e ritorno per i CRAH, pompe per la circolazione del liquido, un grande refrigeratore (chiller) e unità di evaporazione o torri evaporative all'esterno.

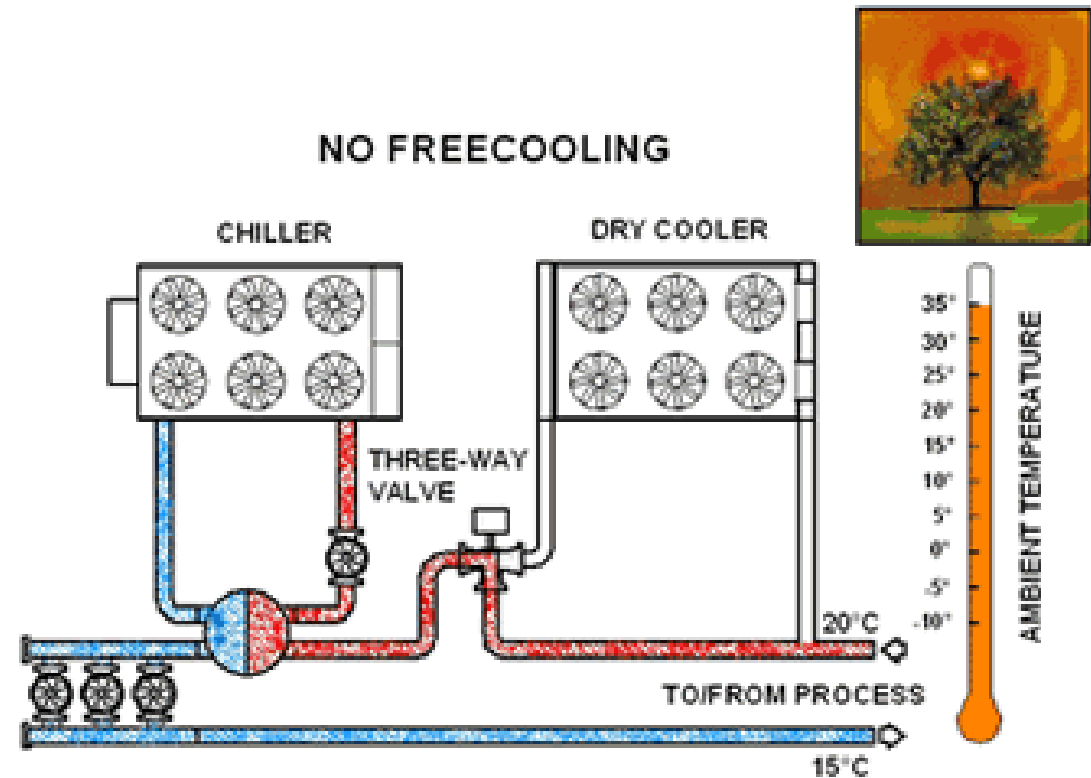
**Computer Room  
Air Handler CRAH**



# FUNZIONAMENTO DEL FREE COOLING

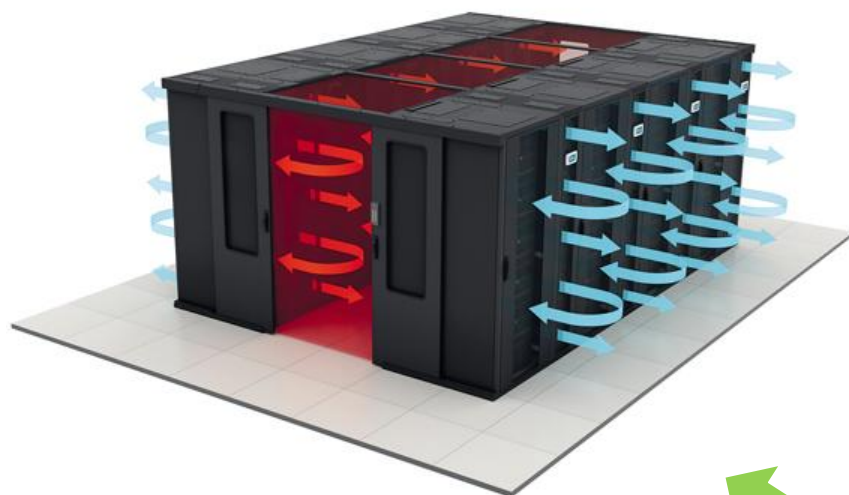
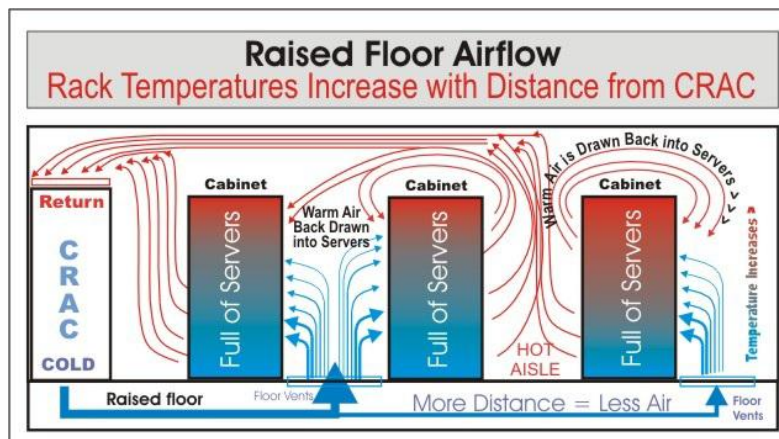


DRY COOLER





## Caratteristiche delle sale server



Il condizionamento perimetrale è poco efficiente:

- i flussi di aria fredda non sono uniformi
- notevoli dispersioni di aria in tutto l'ambiente della sala server
- il plenum di aria fredda viene disturbato da elementi presenti al di sotto del pavimento flottante (cavi, tubi, ecc)

Nel tempo sono state realizzate nuove soluzioni di condizionamento più efficienti di cui le ultime generazioni sono denominate ad «**isole chiuse**» nelle quali **il corridoio caldo o il corridoio freddo vengono chiusi** e separati fisicamente dal resto della sala.

Un esempio è condizionamento «in row» nel quale i CRAH sottili e alti quanto i rack e sono inseriti nella fila di rack a **tutta altezza**

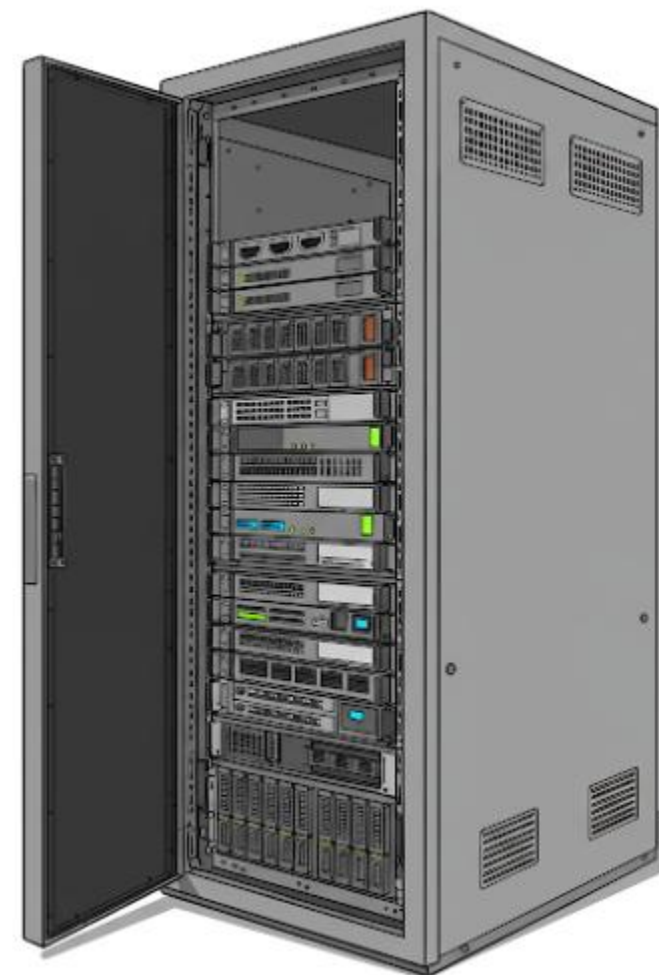
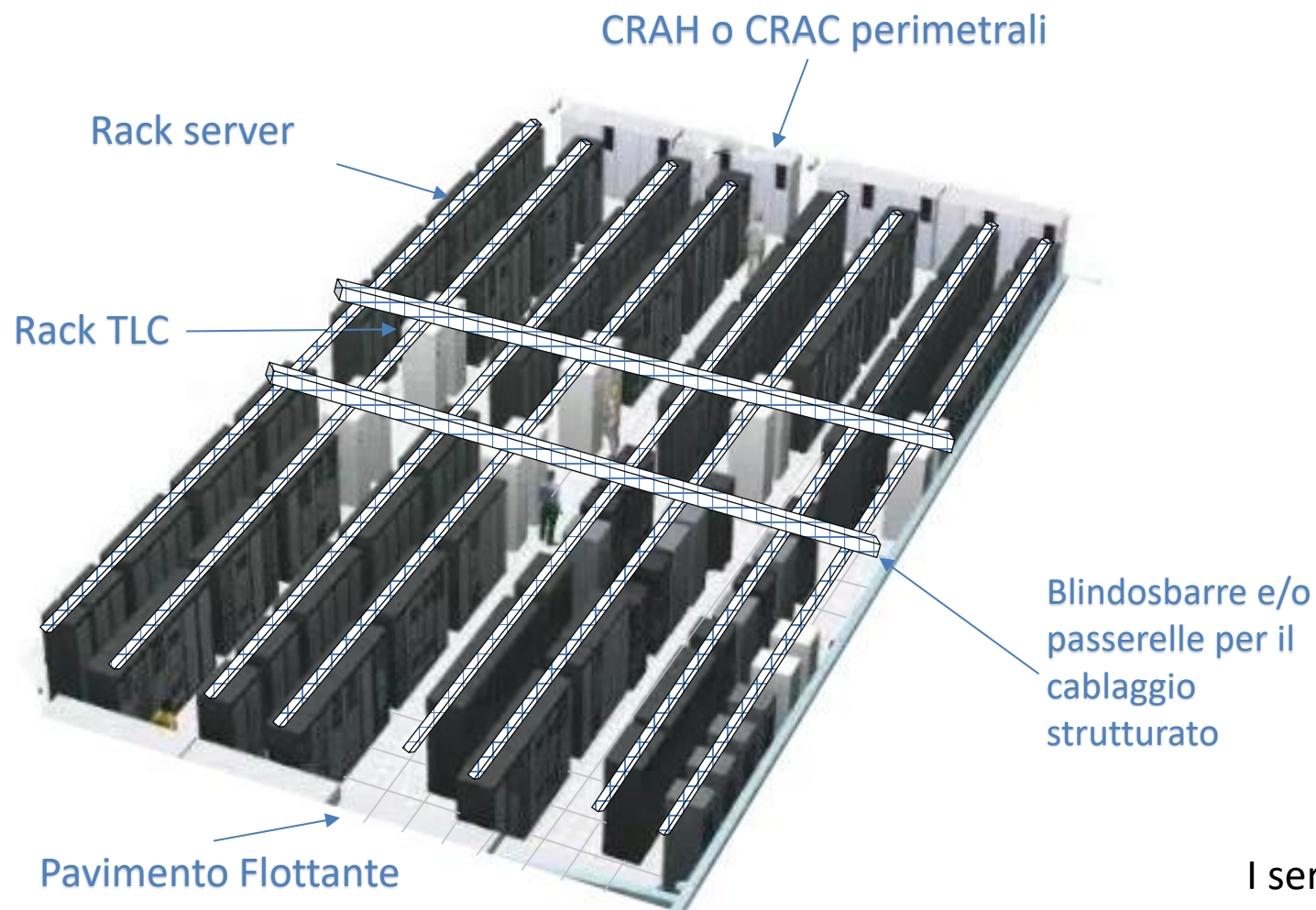
Esempio di condizionamento «in row» con compartimentazione dei corridoi





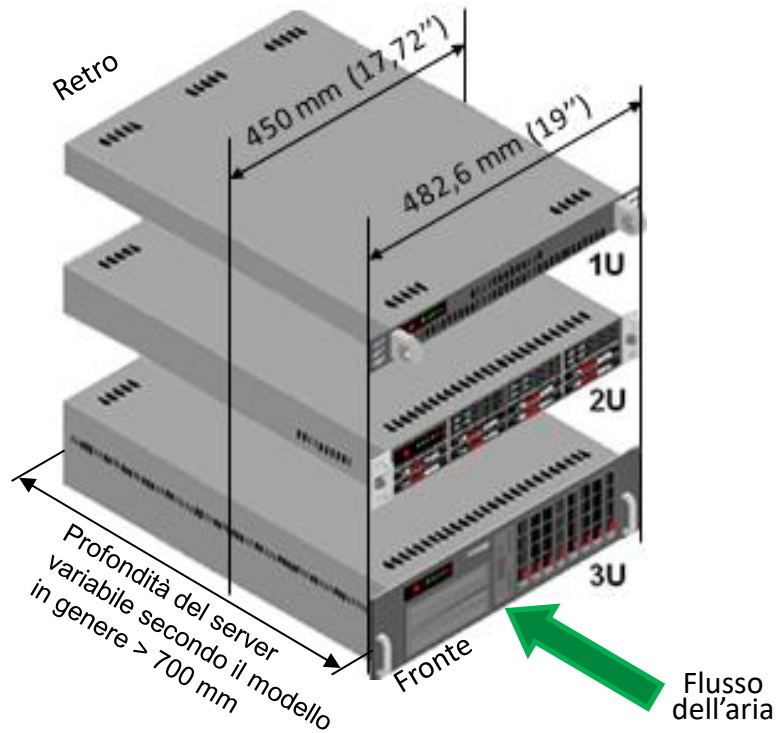
## Caratteristiche delle sale server





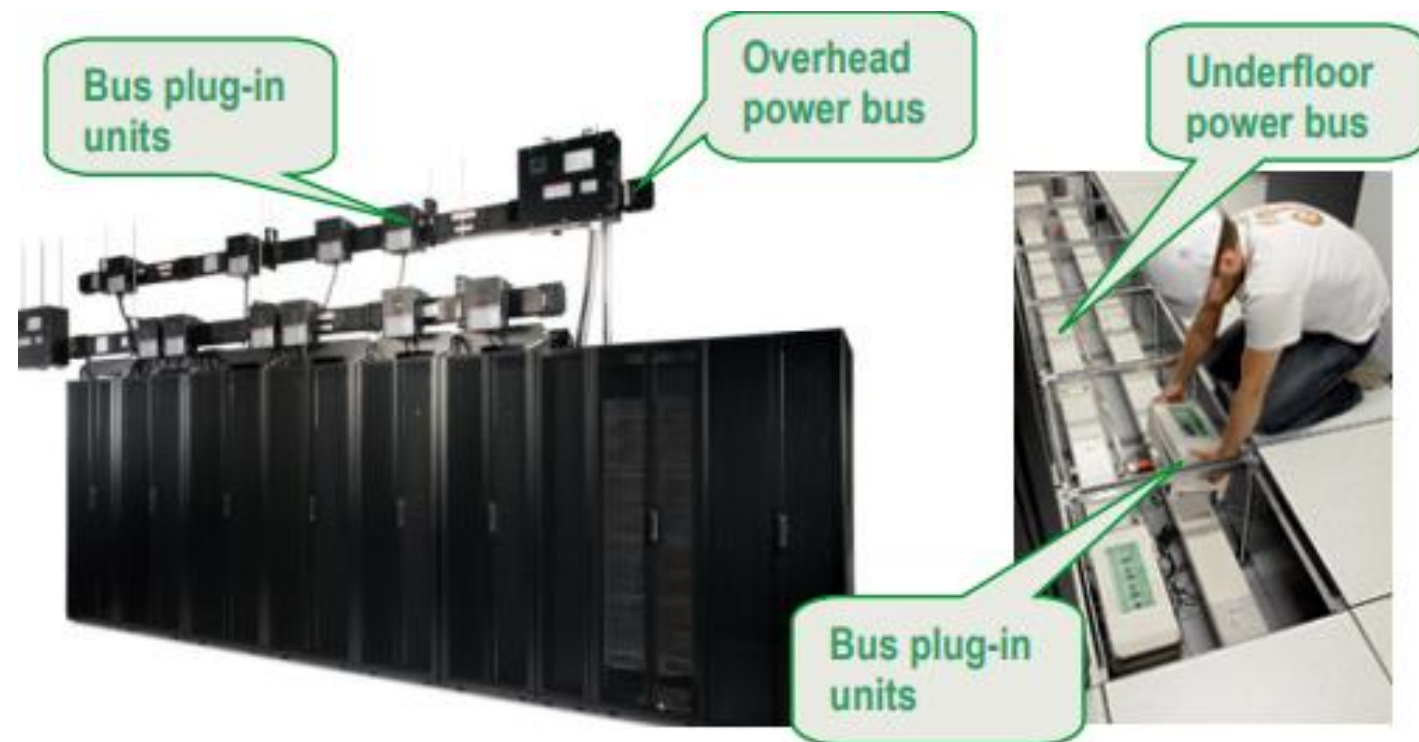
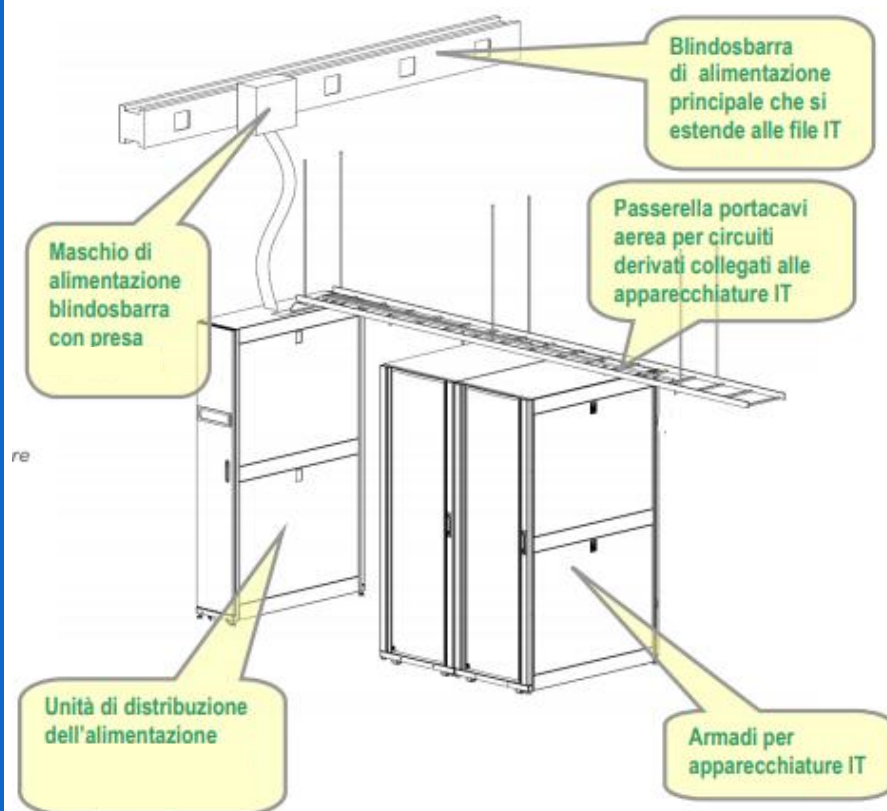
I server sono **montati su slitte** che consentono l'estrazione per le manutenzioni





↑  
SMART  
PDU

- La potenza usata con le densità attuali rack è 7 - 10 kW per rack (@ 32 A monofase – 16 A trifase)
- Il flusso dell'aria di raffreddamento è sempre dal fronte al retro
- I cavi di interconnessione sono **sempre** sul retro del rack
- I server sono **montati su slitte** che consentono l'estrazione per le manutenzioni





Blindosbarre

Bus plug-in Unit

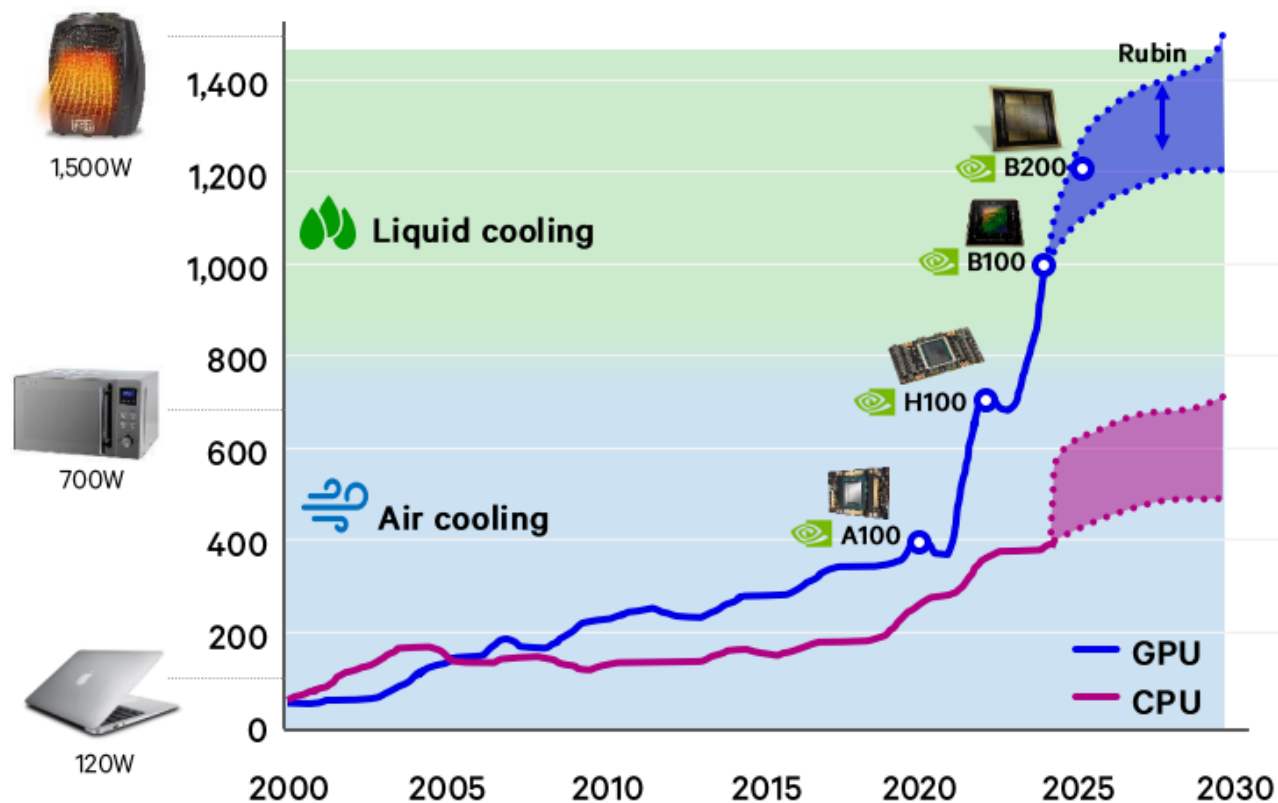


Unità CRAH in Row

Tubazioni di mandata e ritorno unità CRAH

## CPU and GPU power consumption forecast

Thermal Density Power - TDP (watts)



Above 700-800W TDP per chip, liquid cooling quickly becomes a necessity.

# The Future of NVIDIA AI Infrastructure

Accelerated computing innovations enabling efficient deployment of trillion-parameter-scale AI

**GB200 NVL72**


36 GRACE CPUs  
72 BLACKWELL GPUs  
Fully Connected  
NVLink Switch Rack  
Rack-level liquid cooling integration


720 PFlops FP8 training  
1440 PFlops FP4 inference

**GB200 NVL72**

**GRACE  
BLACKWELL**

**Real-Time Trillion-Parameter Models LLM & MoE**

  
Graph Neural Networks

  
Massive Scale Model Training & Inference

**405B - 1T+**

**Rack Power: ~132 kW**

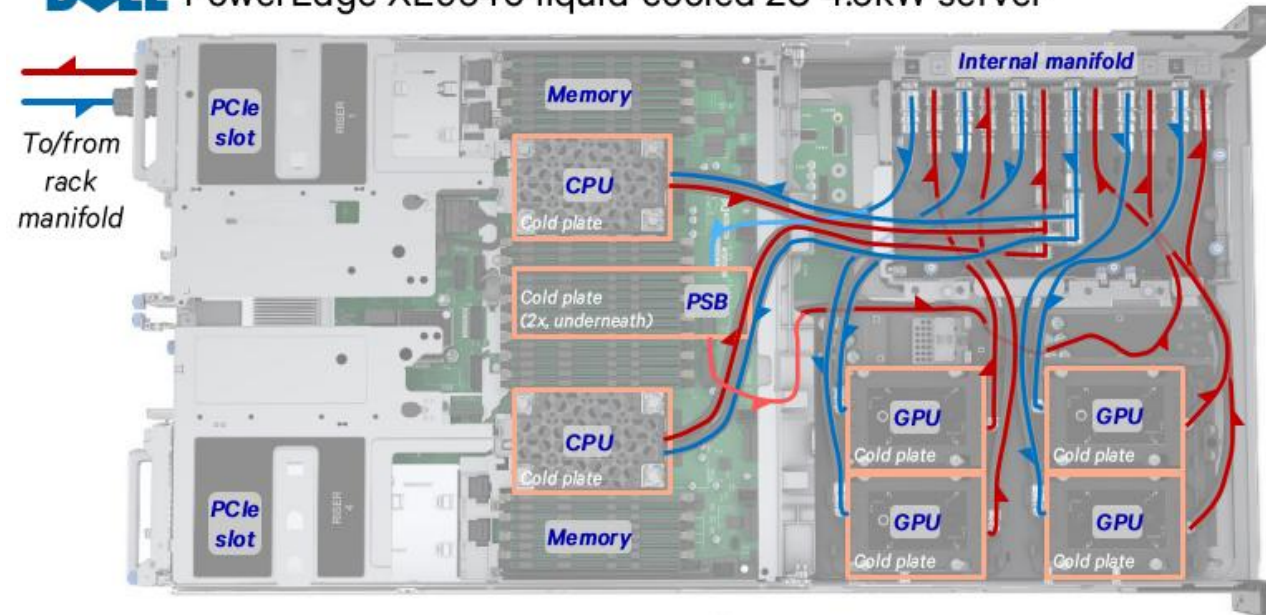
**NVLink Domain 72**

Max GPUs per NVLink Domain  
Max Performance & Capability



### Example:

#### DELL PowerEdge XE9640 liquid-cooled 2U 4.6kW server



#### Liquid-cooled components

- ✓ GPUs
- ✓ CPUs
- ✓ PSB



#### Air-cooled components

- ✓ PSUs
- ✓ Memory
- ✓ Storage
- ✓ PCIe cards



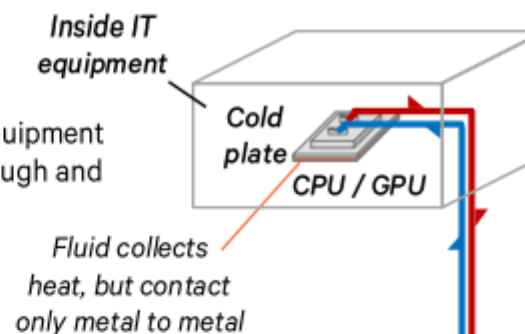
In general, cold plates capture 70-90% of server heat depending on server and cold plate design.





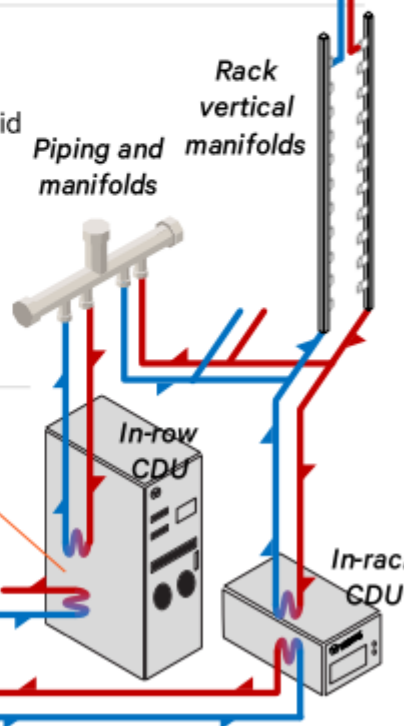
## Cold plate

- ✓ Highly conductive metal in contact with IT equipment pierced by micro-channels for fluid to go through and collect heat
- ✓ Considerable variety in designs
- ✓ Critical with little room for redundancy



## Secondary fluid network

- ✓ Piping, hoses and manifolds carrying the secondary fluid from CDU to cold plate
- ✓ Closed loop with minimal fluid load
- ✓ Equipped with quick disconnects for easier service
- ✓ Requirement to be able to handle system pressure and crucial to avoid leaks and contamination



## Coolant distribution unit

- ✓ Heat coming from cold plates transferred to primary fluid loop
- ✓ Crucial role in controlling flow rate, system pressure and filtration
- ✓ Redundancy ensured with multiple pumps and connection to UPS



I rack per AI dedicati a grandi modelli LLM richiedono una potenza 40 – 80 kW destinata ad aumentare.

Le alte potenze in gioco richiedono profonde modifiche alle infrastrutture tecnologiche:

- Impianto di raffreddamento:
  - disponibilità di raffreddamento a liquido «direct to chip» per le GPU e CPU o a tendere per immersione
  - incremento del condizionamento convenzionale per i flussi di aria fredda che devono raffreddare gli altri componenti dei server (memoria, chip interni di supporto, alimentatori, ecc.)
- Impianto elettrico
  - Cabine di trasformazione, UPS, G.E.
  - Quadri elettrici
  - Cablaggi adeguati alle nuove correnti
  - PDU nei rack adeguate alle nuove richieste di prese elettriche



Manifold per la distribuzione del liquido di raffreddamento ai server



<https://www.youtube.com/watch?v=Jf8EPSBZU7Y&t=746s>



### Data Center per implementazione massiva della IA

- Necessità di potenza elettrica adeguata – allo stato attuale anche 80 kW per rack, a tendere anche 130 -150 kW per rack
- Impianto elettrico adeguato per veicolare ai rack server le potenze richieste
- Gruppi elettrogeni e sistemi di continuità assoluta (UPS) di capacità adeguata
- Impianto di Raffreddamento più complesso, a liquido Direct to CHIP e raffreddamento classico ad aria per il resto dei componenti dei server
- Infrastruttura di rete adeguata a livello di cabling (esclusivamente in fibra ottica) e apparati di rete e con architetture progettate per offrire altissima larghezza di banda, latenza ultra-bassa e funzionamento senza perdite su larga scala (RDMA - Remote Direct Memory Access o Infiniband allo stato attuale)
- Infrastruttura di storage ad alte prestazioni con storage a stato solido per supportare le richieste dei Cluster di GPU





# GRAZIE PER L'ATTENZIONE