



Classificazione Decimale Dewey

— **006.3 (23.)** Intelligenza Artificiale e computazione naturale

Thema

— Soggetto: **U.** Informatica e Tecnologia dell'Informazione

— Qualificatori: **4CT.** Per l'istruzione superiore/terziaria/universitaria

GERARDO IOVANE
ROLANDO ROSTOLIS

DATA CENTER NELL'ERA DELL'INTELLIGENZA ARTIFICIALE

INFRASTRUTTURE DI CALCOLO, INGEGNERIA
ED ECONOMIA DEL DATO, SOSTENIBILITÀ
ENERGETICA ED AMBIENTALE E TENDENZE FUTURE





©

ISBN
979-12-218-2685-2

PRIMA EDIZIONE
ROMA 27 APRILE 2026

Indice dei contenuti

Prefazione.....	23
Introduzione	25
Struttura e articolazione del volume.....	25
Il framework AI Data Center Architecture Stack: presentazione formale	28
Come leggere questo libro: percorsi e livelli di lettura.....	30
Note su formalismo matematico, notazione e riferimenti bibliografici	31
Il contesto: la trasformazione strutturale dell'infrastruttura digitale globale.....	32
Riferimenti bibliografici della Prefazione e dell'Introduzione.....	35
CAPITOLO 1 - L'ECONOMIA DEL DATO: IL DATO COME RISORSA STRATEGICA, PIATTAFORME DIGITALI, ECOSISTEMI HYPERSCALE E DOMANDA COMPUTAZIONALE DELL'INTELLIGENZA ARTIFICIALE.....	37
1.1 Il dato come risorsa economica fondamentale: inquadramento teorico e misurazione.....	37
1.2 Piattaforme digitali e meccanismi di creazione del valore: effetti di rete, lock-in e strutture di mercato	40
1.3 Ecosistemi hyperscale: definizione, caratteristiche strutturali e soglie di scala	43
1.4 La domanda computazionale dell'intelligenza artificiale: struttura, traiettorie e implicazioni infrastrutturali.....	46
1.5 Traiettorie di investimento, dinamiche geopolitiche e prospettive strutturali per l'infrastruttura digitale	49
1.6 Riferimenti bibliografici del Capitolo 1	53
CAPITOLO 2 - EVOLUZIONE STORICA DELL'INFRASTRUTTURA COMPUTAZIONALE: DAI MAINFRAME ALLE AI FACTORY.....	57
2.1 L'era dei mainframe e del calcolo centralizzato (1950–1980): fondamenti architettonici e modello economico.....	57
2.2 La rivoluzione del personal computer e l'enterprise data center (1980–2000): decentralizzazione, client-server e l'esplosione di Internet	60
2.3 L'emergere del cloud computing e i data center hyperscale (2000–2015): virtualizzazione, utilità e scala planetaria	62
2.4 L'era del deep learning e la nascita dell'AI factory (2015–oggi): GPU computing, cluster di training e la nuova infrastruttura.....	66
2.5 Leggi di scala, traiettorie tecnologiche e transizioni di paradigma: un'analisi comparativa.....	69
2.6 Riferimenti bibliografici del Capitolo 2	73

CAPITOLO 3 - GEOGRAFIA GLOBALE DEI DATA CENTER: HUB COMPUTAZIONALI, CAVI SOTTOMARINI E CORRIDOI DI LATENZA	77
3.1 Fondamenti della geografia computazionale: perché la localizzazione conta nell'era del cloud	77
3.2 I principali hub computazionali globali: Northern Virginia, Silicon Valley, Francoforte, Singapore e Pechino	79
3.3 Infrastruttura dei cavi sottomarini: fisica della trasmissione, geopolitica e vulnerabilità sistemiche	83
3.4 Corridoi di latenza e architettura geografica delle reti: modelli quantitativi e ottimizzazione del placement	86
3.5 Mercati emergenti, reshaping geopolitico e traiettorie future della distribuzione geografica del compute	89
3.6 Riferimenti bibliografici del Capitolo 3.....	93
CAPITOLO 4 - ENTERPRISE DATA CENTER: PATTERN PROGETTUALI, CARICHI DI LAVORO TIPICI E CAPACITY PLANNING.....	97
4.1 Definizione e posizionamento dell'enterprise data center nell'ecosistema delle infrastrutture digitali.....	97
4.2 Pattern architeturali dei data center enterprise: topologie fisiche, layout e flussi d'aria.....	100
4.3 Carichi di lavoro tipici degli enterprise data center: caratterizzazione e modellazione	103
4.4 Metodologie di capacity planning: modelli quantitativi, rightsizing e gestione della crescita	106
4.5 Governance operativa, gestione del ciclo di vita e metriche di performance degli enterprise data center	109
4.6 Evoluzione dell'enterprise data center verso il modello ibrido: integrazione cloud, software-defined infrastructure e modernizzazione	112
4.7 Riferimenti bibliografici del Capitolo 4	115
CAPITOLO 5 - COLOCATION FACILITY: ARCHITETTURA MULTI-TENANT, INTERCONNECTION FABRIC E CARRIER NEUTRALITY	119
5.1 Il modello di business della colocation: struttura del mercato, tipologie di offerta e dinamiche competitive.....	119
5.2 Architettura fisica multi-tenant: design della segmentazione, sicurezza e isolamento dei tenant	122
5.3 Interconnection fabric: architettura tecnica, tipologie di cross-connect e mercati dello scambio del traffico.....	125
5.4 Carrier neutrality: definizione formale, modelli di peering e implicazioni per l'architettura di rete dei tenant.....	128
5.5 Modelli di cloud on-ramp, virtual interconnection e il ruolo delle colocation nell'ecosistema hybrid cloud	131
5.6 Riferimenti bibliografici del Capitolo 5	135

CAPITOLO 6 - HYPERSCALE DATA CENTER: ARCHITETTURE, INNOVAZIONI INGEGNERISTICHE E STRATEGIE OPERATIVE DEI PRINCIPALI OPERATORI GLOBALI.....	137
6.1 Definizione e caratteristiche strutturali del paradigma hyperscale: oltre la scala, verso la co-progettazione software-hardware	137
6.2 L'architettura dei data center Google: modularità, custom silicon e gestione dell'energia a 48V	140
6.3 Amazon Web Services: l'architettura a zone di disponibilità, il custom silicon Graviton/Trainium e i principi ingegneristici del cloud pubblico.....	143
6.4 Microsoft Azure e Meta: approcci differenziati alla scala hyperscale e all'ottimizzazione per AI	146
6.5 Il framework ADCAS applicato all'architettura hyperscale: validazione empirica e generalizzabilità del modello	149
6.6 Riferimenti bibliografici del Capitolo 6	153
CAPITOLO 7 - EDGE DATA CENTER: INFRASTRUTTURA A BASSA LATENZA PER IoT, 5G E SISTEMI AUTONOMI	157
7.1 Fondamenti concettuali dell'edge computing: dalla gerarchia centralizzata alla distribuzione capillare del compute	157
7.2 Architettura fisica degli edge data center: vincoli di deployment, fattori di forma e infrastruttura tecnica.....	160
7.3 Edge computing e reti 5G: architettura MEC, network slicing e requisiti di latenza per applicazioni critiche	163
7.4 Edge computing per l'Internet of Things industriale: architetture, protocolli e casi d'uso manifatturieri	167
7.5 Edge data center per i sistemi di trasporto autonomo e la smart city: requisiti, architetture e prospettive di deployment	170
7.6 Riferimenti bibliografici del Capitolo 7	174
CAPITOLO 8 - DATA CENTER PER INTELLIGENZA ARTIFICIALE E HIGH PERFORMANCE COMPUTING: CLUSTER GPU, AI FACTORY E INFRASTRUTTURA DI TRAINING E INFERENZA.....	177
8.1 La convergenza tra HPC e AI: traiettorie storiche, similitudini architetture e divergenze operative	177
8.2 Architettura dei cluster GPU per il training di modelli AI: topologie di rete, parallelismo e scalabilità	180
8.3 AI Factory: il paradigma della produzione industriale di intelligenza artificiale e i requisiti infrastrutturali	183
8.4 Infrastruttura di inferenza AI: ottimizzazione per throughput e latenza, batching e quantizzazione	186
8.5 Storage per AI: filesystem paralleli, object storage e pipeline dati per il training e l'inferenza	190
8.6 Riferimenti bibliografici del Capitolo 8	193

CAPITOLO 9 - SELEZIONE DEL SITO PER DATA CENTER: DISPONIBILITÀ DI ENERGIA, CLIMA, RISCHIO SISMICO, CONNETTIVITÀ IN FIBRA E RISCHIO GEOPOLITICO	197
9.1 Il processo decisionale di site selection: framework multicriteria e struttura degli investimenti.....	197
9.2 Disponibilità e qualità dell'energia elettrica: analisi della rete, tariffe e fonti rinnovabili	199
9.3 Fattori climatici e rischi naturali: temperatura, umidità, rischio idrogeologico e rischio sismico	203
9.4 Connettività in fibra ottica: valutazione dell'infrastruttura esistente, dark fiber e latenza verso i mercati	206
9.5 Rischio geopolitico, sovranità digitale e framework normativi: la dimensione istituzionale della site selection	208
9.6 Riferimenti bibliografici del Capitolo 9	213
CAPITOLO 10 - INGEGNERIA STRUTTURALE DEI DATA CENTER: DISTRIBUZIONE DEI CARICHI, PROGETTAZIONE ANTISISMICA E COSTRUZIONE MODULARE	217
10.1 Fondamenti dell'ingegneria strutturale applicata ai data center: carichi, normative di riferimento e requisiti funzionali specifici	217
10.2 Distribuzione dei carichi nel white space: solai ad alta densità, pavimento sopraelevato e percorsi tecnici.....	220
10.3 Progettazione antisismica dei data center: forze sui componenti non strutturali, ancoraggi e isolamento sismico alla base	223
10.4 Costruzione modulare prefabbricata: principi ingegneristici, tipologie di moduli e ottimizzazione del processo costruttivo.....	227
10.5 BIM, gemello digitale strutturale e monitoraggio della salute strutturale nel ciclo di vita del data center.....	230
10.6 Riferimenti bibliografici del Capitolo 10	235
CAPITOLO 11 - LAYOUT DESIGN DEL DATA CENTER: WHITE SPACE, CONTENIMENTO DEI CORRIDOI CALDO/FREDDO E SALE MODULARI.....	237
11.1 Principi fondamentali del layout design: dalla pianificazione funzionale all'ottimizzazione termoidraulica	237
11.2 Architettura hot aisle / cold aisle: fondamenti, varianti di contenimento e analisi delle prestazioni termoidrauliche	240
11.3 Layout delle sale modulari: progettazione per la scalabilità, ridondanza e flessibilità operativa	244
11.4 Layout per data center ad alta densità e raffreddamento a liquido: gestione dello spazio, distribuzione dei fluidi e integrazione con l'IT.....	247
11.5 Ottimizzazione del layout attraverso modelli di simulazione, gestione dinamica dello spazio e tendenze evolutive.....	250
11.6 Riferimenti bibliografici del Capitolo 11	255

CAPITOLO 12 - TECNICHE COSTRUTTIVE PER DATA CENTER: MODULI PREFABBRICATI, SISTEMI CONTAINERIZZATI E INDUSTRIALIZZAZIONE DEL PROCESSO EDILIZIO.....	259
12.1 Evoluzione del processo costruttivo dei data center: dalla cantieristica tradizionale all'industrializzazione produttiva	259
12.2 Sistemi containerizzati: progettazione strutturale, integrazione impiantistica e prestazioni operative	262
12.3 Moduli prefabbricati di grande scala: Light Steel Frame, sistemi sandwich e ingegneria delle connessioni inter-modulo	266
12.4 Commissioning e qualità nella costruzione modulare: metodologie di test, certificazione e integrazione con la produzione in serie.....	269
12.5 Analisi del ciclo di vita, sostenibilità e prospettive di innovazione nella costruzione modulare dei data center	273
12.6 Riferimenti bibliografici del Capitolo 12	277
CAPITOLO 13 - MODELLAZIONE DELLA DOMANDA DI POTENZA NEI DATA CENTER: SCALING IN MW, DENSITÀ DI POTENZA E CONSUMO DEI CLUSTER AI.....	281
13.1 Fondamenti della modellazione della domanda di potenza: gerarchia dei carichi, metodologie di stima e framework analitici	281
13.2 Densità di potenza per rack: evoluzione storica, drivers tecnologici e impatto sull'infrastruttura	284
13.3 Modelli di consumo energetico dei server: dynamic power scaling, ACPI e caratterizzazione della curva di carico	287
13.4 Modellazione del consumo dei cluster AI: training, inferenza e proiezioni di scaling per large language model	290
13.5 Proiezioni della domanda globale di potenza dei data center: scenari al 2030, implicazioni per la rete elettrica e politiche di efficienza.....	293
13.6 Riferimenti bibliografici del Capitolo 13	298
CAPITOLO 14 - ARCHITETTURA DI DISTRIBUZIONE ELETTRICA DEI DATA CENTER: SOTTOSTAZIONI, TRASFORMATORI, QUADRI DI MANOVRA E RIDONDANZA	301
14.1 Fondamenti dell'architettura elettrica dei data center: gerarchia della distribuzione, livelli di tensione e criteri di ridondanza	301
14.2 Trasformatori di potenza per data center: tipologie, perdite, dimensionamento e gestione termica.....	304
14.3 Quadri di manovra in media e bassa tensione: tecnologie, selettività e coordinamento delle protezioni	308
14.4 Topologie di ridondanza avanzata: sistemi 2N, 2(N+1) e distributed redundancy per data center hyperscale	312
14.5 Power Distribution Unit, cablaggio di rack e gestione della qualità dell'energia: armoniche, fattore di potenza e misura intelligente	315
14.6 Riferimenti bibliografici del Capitolo 14	319

CAPITOLO 15 - TECNOLOGIE UPS PER DATA CENTER: BATTERIE AL LITIO, SISTEMI A VOLANO, ARCHITETTURE DI RIDONDANZA E GESTIONE DEL CICLO DI VITA	323
15.1 Fondamenti degli Uninterruptible Power Supply: topologie circuitali, parametri prestazionali e normativa di riferimento	323
15.2 Batterie al litio per UPS: chimica, caratterizzazione elettrochimica, gestione BMS e confronto con le tecnologie piombo-acido	326
15.3 Sistemi UPS a volano di inerzia (flywheel): principi fisici, tecnologie e confronto con l'accumulo elettrochimico	330
15.4 Dimensionamento del sistema UPS: capacità di potenza, capacità energetica, perdite e ottimizzazione del ciclo di vita	333
15.5 Architetture di ridondanza N+1 e 2N: analisi della disponibilità, transient fault management e integrazione con i sistemi di gestione	336
15.6 Riferimenti bibliografici del Capitolo 15	341
CAPITOLO 16 - GENERAZIONE DI BACKUP PER DATA CENTER: GRUPPI ELETTRICI DIESEL, TURBINE A GAS, SISTEMI A IDROGENO E TRANSIZIONE VERSO LA GENERAZIONE SOSTENIBILE	345
16.1 Ruolo e architettura della generazione di backup nei data center: requisiti di continuità, tempi di intervento e integrazione con l'UPS	345
16.2 Gruppi elettrogeni diesel: termodinamica del ciclo Diesel, efficienza, caratterizzazione delle emissioni e gestione del combustibile	348
16.3 Turbine a gas per la generazione di backup e CHP: ciclo Brayton, efficienza e applicazioni nei data center di grande scala	352
16.4 Sistemi a idrogeno per la generazione di backup: celle a combustibile PEM, elettrolizzatori e prospettive dell'hydrogen economy per i data center	355
16.5 Parallelismo dei generatori, sistemi di controllo della generazione e integrazione con le energie rinnovabili	358
16.6 Riferimenti bibliografici del Capitolo 16	363
CAPITOLO 17 - FONDAMENTI DI GESTIONE TERMICA NEI DATA CENTER: TRASMISSIONE DEL CALORE, BILANCI ENERGETICI E METRICHE DI EFFICIENZA TERMICA.....	367
17.1 La termodinamica applicata ai data center: leggi fondamentali, flussi di calore e il vincolo fisico del raffreddamento	367
17.2 La catena termica dal transistor all'ambiente: resistenza termica, junction temperature e limiti operativi dei componenti IT	371
17.3 Meccanismi di scambio termico nei white space: fluidodinamica computazionale, modelli di miscelazione e indici di performance	374
17.4 Metriche globali di efficienza termica: PUE termico, WUE, ERE e indicatori emergenti per i data center AI	377
17.5 Standard ASHRAE per l'ambiente termico IT: classi operative, limiti di temperatura e umidità e implicazioni per il progetto.....	381
17.6 Riferimenti bibliografici del Capitolo 17	385

CAPITOLO 18 - SISTEMI DI RAFFREDDAMENTO AD ARIA PER DATA CENTER: CRAC, CRAH, CIRCUITI AD ACQUA REFRIGERATA E ARCHITETTURE DI CONTENIMENTO TERMICO	389
18.1 Principi del raffreddamento ad aria nei data center: ciclo psicrometrico, portate d'aria e bilancio termico del white space	389
18.2 Unità CRAC e CRAH: architetture costruttive, cicli frigoriferi, coefficienti di prestazione e modalità di controllo.....	392
18.3 Chiller centrifughi e a vite: termodinamica del ciclo reale, diagramma di Mollier del refrigerante e ottimizzazione operativa	396
18.4 Circuiti idronici dell'acqua refrigerata: topologie, pompaggio variabile, bilanciamento e ridondanza	399
18.5 Sistemi di contenimento termico, in-row cooling e rear-door heat exchanger: efficienza e applicazioni nelle alte densità	402
18.6 Riferimenti bibliografici del Capitolo 18	407
CAPITOLO 19 - RAFFREDDAMENTO A LIQUIDO PER DATA CENTER: DIRECT LIQUID COOLING, IMMERSION COOLING A SINGOLA E BIFASE E INTEGRAZIONE CON L'INFRASTRUTTURA TERMICA	411
19.1 Motivazioni fisiche e architetture del raffreddamento a liquido: limiti del raffreddamento ad aria e transizione tecnologica verso le alte densità	411
19.2 Direct Liquid Cooling monofase (warm water cooling): cold plate, circuiti idronici CDU e caratterizzazione termofluidodinamica.....	414
19.3 Direct Liquid Cooling bifase: ebollizione sul cold plate, correlazioni di Rohsenow, crisi del flusso termico e fluidi di lavoro	418
19.4 Immersion cooling monofase: vasche dielettriche, fluidi fluorurati e idrocarburici, scambiatori primario-secondario e gestione dei server	421
19.5 Immersion cooling bifase: meccanismo di ebollizione in pool, geometrie di raffreddamento, condensatori e prospettive per i data center AI	425
19.6 Riferimenti bibliografici del Capitolo 19	428
CAPITOLO 20 - FREECOOLING E ARCHITETTURE DI RAFFREDDAMENTO OTTIMIZZATE PER IL CLIMA: ECONOMIZZATORI ARIA E ACQUA, RAFFREDDAMENTO ADIABATICO E PROGETTAZIONE CLIMATICA DEI DATA CENTER	431
20.1 Principi del freecooling: definizione, ore di economizzazione annua e analisi climatica con il metodo dei gradi-ora.....	431
20.2 Economizzatori ad aria (airside economizers): freecooling diretto, filtrazione, controllo dell'umidità e rischi di contaminazione	434
20.3 Economizzatori ad acqua (waterside economizers): freecooling indiretto, dry cooler, cooling tower e integrazione con il chiller.....	438
20.4 Raffreddamento adiabatico e pre-raffreddamento evaporativo: sistemi diretti, indiretti e a due stadi per climi caldi e secchi	441
20.5 Progettazione climatica dei data center: zonizzazione termica globale, architetture ottimali per macroclima e strategie ibride di raffreddamento	445
20.6 Riferimenti bibliografici del Capitolo 20	449

CAPITOLO 21 - ARCHITETTURA DEI SERVER NEI DATA CENTER AI: CPU, GPU, ACCELERATORI HARDWARE E SISTEMA DI MEMORIA — MODELLI DI PRESTAZIONE E CRITERI DI SELEZIONE	453
21.1 Architettura dei processori general-purpose: pipeline, gerarchia di cache, NUMA e limiti del parallelismo di istruzione per i workload AI	453
21.2 Architettura GPU per il calcolo AI: CUDA core, Tensor Core, gerarchia di memoria on-chip e modello di esecuzione SIMT	456
21.3 Acceleratori AI dedicati: TPU, Gaudi, Trainium e architetture systolic array — confronto con le GPU NVIDIA	460
21.4 Sistemi di memoria per il calcolo AI: HBM, LPDDR, memoria di capacità e gerarchia di memoria nei server multi-acceleratore	463
21.5 Server AI di riferimento: NVIDIA DGX H100, architettura NVLink, interconnessione InfiniBand e criteri di selezione del server per i cluster AI	466
21.6 Riferimenti bibliografici del Capitolo 21	470
CAPITOLO 22 - SISTEMI DI STORAGE NEI DATA CENTER AI: SAN, NAS, OBJECT STORAGE, FILE SYSTEM DISTRIBUITI E DIMENSIONAMENTO DELLE PRESTAZIONI	473
22.1 Tassonomia dei sistemi di storage: block, file e object storage — modelli di accesso, stack protocollare e trade-off fondamentali	473
22.2 Storage Area Network (SAN): architetture Fibre Channel e NVMe-oF, zoning, multipathing e dimensionamento delle prestazioni	475
22.3 Network Attached Storage (NAS): protocolli NFS e SMB, architetture scale-out e file system paralleli per i cluster AI	478
22.4 Object storage distribuito: architettura S3-compatible, consistency model, erasure coding e dimensionamento della capacità per i dataset AI	482
22.5 Prestazioni di storage per workload AI: metriche IOPS, latenza e throughput, modello coda M/M/1 e ottimizzazione del data loading pipeline	485
22.6 Riferimenti bibliografici del Capitolo 22	489
CAPITOLO 23 - NETWORKING NEI DATA CENTER AI: TOPOLOGIA SPINE-LEAF, INTERCONNESSIONI OTTICHE, RETI AD ALTA LARGHEZZA DI BANDA E PROTOCOLLI RDMA PER IL CALCOLO DISTRIBUITO	493
23.1 Evoluzione delle topologie di rete nei data center: da three-tier a spine-leaf — analisi della latenza, del diametro del grafo e della bisection bandwidth	493
23.2 Interconnessioni ottiche nei data center: fisica della trasmissione in fibra, moduli ottici (QSFP-DD, OSFP, CPO) e tecnologia WDM per i cluster AI	496
23.3 InfiniBand HDR e NDR: architettura, routing adattivo e prestazioni per cluster AI — confronto con Ethernet RoCE	499
23.4 Protocolli RDMA e comunicazioni collettive: NCCL, AllReduce, operazioni MPI e ottimizzazione del communication overlap nei training loop	502
23.5 Gestione della congestione e Quality of Service nelle reti AI: ECN, PFC, DCQCN e dimensionamento dei buffer degli switch	505
23.6 Riferimenti bibliografici del Capitolo 23	509

CAPITOLO 24 - VIRTUALIZZAZIONE E HYPERVISOR NEI DATA CENTER AI: FONDAMENTI ARCHITETTURALI, VIRTUALIZZAZIONE HARDWARE-ASSISTED, GPU VIRTUALIZATION E MODELLI DI CONSOLIDAMENTO	513
24.1 Fondamenti della virtualizzazione: il teorema di Popek-Goldberg, ring di protezione x86 e virtualizzazione hardware-assisted con Intel VT-x e AMD-V ...	513
24.2 Architetture hypervisor Type-1 bare-metal e Type-2 hosted: KVM, Xen, VMware ESXi, Hyper-V e confronto di overhead, scalabilità e sicurezza	516
24.3 IOMMU, SR-IOV e passthrough PCIe: virtualizzazione dell'I/O ad alte prestazioni per GPU, NIC ad alta velocità e SSD NVMe nei server AI	519
24.4 Virtualizzazione GPU: NVIDIA vGPU, MIG Multi-Instance GPU e SR-IOV GPU — modelli di partizione e prestazioni per i workload AI di inferenza e training	523
24.5 Consolidamento dei server e modelli di workload: overcommitment di CPU e memoria, live migration e schedulazione delle risorse nelle piattaforme AI.....	526
24.6 Riferimenti bibliografici del Capitolo 24	530
CAPITOLO 25 - CONTAINER E ORCHESTRAZIONE: KUBERNETES, CALCOLO DISTRIBUITO E PIATTAFORME AI-NATIVE PER I DATA CENTER DI NUOVA GENERAZIONE	533
25.1 Fondamenti dei container Linux: namespace, cgroup, overlay filesystem e confronto con la virtualizzazione a VM per i workload AI	533
25.2 Architettura di Kubernetes: API server, etcd, scheduler, kubelet, CNI e CRI — progettazione dei cluster AI con node pool GPU e topology-aware scheduling .	536
25.3 Gestione delle GPU in Kubernetes: NVIDIA GPU Operator, device plugin, topology-aware scheduling con NUMA e MIG-aware resource allocation	539
25.4 Framework di calcolo distribuito AI su Kubernetes: Kubeflow, PyTorch Operator, Ray e gestione del ciclo di vita dei job di training e inferenza.....	542
25.5 Scalabilità e osservabilità dei cluster Kubernetes AI: autoscaling orizzontale e verticale, Prometheus, resource quotas e cost governance nei data center multi-tenant	546
25.6 Riferimenti bibliografici del Capitolo 25	550
CAPITOLO 26 - ARCHITETTURA CLOUD: MODELLI IAAS, PAAS E SAAS, CLOUD-NATIVE DESIGN, MULTI-CLOUD E CLOUD PRIVATO PER I DATA CENTER AI	553
26.1 Modelli di servizio cloud IaaS, PaaS e SaaS: definizioni NIST, pila tecnologica, responsabilità condivisa e deployment model per i workload AI.....	553
26.2 Cloud-native design: microservizi, API RESTful, service mesh, serverless computing e il pattern di architettura degli AI inference endpoint	556
26.3 Architetture multi-cloud e cloud ibrido: progettazione della connettività, latenza e coerenza dei dati tra cloud pubblici e data center privati.....	559
26.4 Storage cloud per l'AI: object storage S3-compatible, data lakehouse, delta lake e accesso ad alta larghezza di banda per le pipeline di training dei grandi modelli	563

26.5 Sicurezza e conformità nell'architettura cloud AI: modello zero-trust, cifratura end-to-end, compliance GDPR e gestione delle identità IAM per i data center multi-tenant.....	566
26.6 Riferimenti bibliografici del Capitolo 26	570
CAPITOLO 27 - CLUSTER GPU E SUPERCOMPUTER AI: ARCHITETTURA DEI SISTEMI DI CALCOLO AD ALTISSIMA PRESTAZIONE, SCALING LAWS, PARALLELISMO MULTI-DIMENSIONALE E INFRASTRUTTURA DEI GRANDI AI FACTORY	573
27.1 Architettura dei cluster GPU: topologia hardware, fabric di interconnessione NVLink-NVSwitch, gerarchie di comunicazione e il concetto di AI Factory.....	573
27.2 Scaling laws per i grandi modelli AI: leggi di potenza di Kaplan-Hoffmann, compute-optimal training e implicazioni per il dimensionamento dei cluster GPU	576
27.3 Parallelismo multi-dimensionale per il training distribuito di LLM: data parallel, tensor parallel, pipeline parallel e sequence parallel — analisi del communication overhead	579
27.4 Efficienza del training distribuito: MFU, hardware utilization, gradient checkpointing, mixed precision e ottimizzazione dei kernel CUDA per i cluster GPU.....	582
27.5 Case study dei supercomputer AI: Frontier, Eagle, Selene, Meta RSC, xAI Colossus — architettura, record di prestazione e pianificazione dei cluster di nuova generazione.....	585
27.6 Riferimenti bibliografici del Capitolo 27	589
CAPITOLO 28 - AI NETWORKING: INFINIBAND, NVLINK, ETHERNET AD ALTE PRESTAZIONI E FABRIC DI INTERCONNESSIONE PER I CLUSTER GPU DI NUOVA GENERAZIONE	593
28.1 Fondamenti del networking AI: requisiti di larghezza di banda, latenza e semantica di comunicazione per il training distribuito di grandi modelli linguistici	593
28.2 InfiniBand HDR e NDR: architettura del protocollo, topologie fat-tree e dragonfly+, adaptive routing e prestazioni nei cluster AI di grande scala	596
28.3 NVLink e NVSwitch: evoluzione generazionale, topologia all-to-all intra-nodo, NVLink Network per cluster multi-nodo e confronto con InfiniBand per workload AI	599
28.4 RoCE v2 ed Ethernet ad alte prestazioni per l'AI: DCQCN, PFC, ECN e confronto con InfiniBand per i data center AI di scala hyperscale	602
28.5 Tendenze evolutive del networking AI: Silicon Photonics, Co-Packaged Optics, 800G/1.6T Ethernet, liquid-cooled network fabrics e architetture di rete per cluster da milione di GPU.....	606
28.6 Riferimenti bibliografici del Capitolo 28	610
CAPITOLO 29 - DATA PIPELINE E STORAGE PER IL MACHINE LEARNING: ARCHITETTURA DEI SISTEMI DI INGESTIONE, PREPROCESSAMENTO, VERSIONING E ACCESSO AD ALTA VELOCITÀ PER I WORKLOAD AI NEI DATA CENTER	613

29.1 Architettura end-to-end delle data pipeline per l'AI: dal dato grezzo al tensore di training — ingestione, validazione, feature engineering e data lineage	613
29.2 File system distribuiti e storage parallelo per l'AI: GPFS, Lustre, WekaFS e ottimizzazione del throughput di I/O per il training su cluster GPU di grande scala	616
29.3 Sistemi di caching e data staging per l'AI: Alluxio, CacheLib e architetture di storage gerarchico per ridurre la latenza di accesso ai dataset di training	620
29.4 Streaming data pipeline per il training online e continuo: Apache Kafka, Flink, architetture Lambda e Kappa e gestione del concept drift nei modelli AI di produzione	623
29.5 Formati di dataset per l'AI: WebDataset, TFRecord, MosaicML StreamingDataset, serializzazione ad alta velocità e gestione di dataset da petabyte su object storage distribuito.....	626
29.6 Riferimenti bibliografici del Capitolo 29	630
CAPITOLO 30 - SICUREZZA FISICA DEI DATA CENTER: DIFESA PERIMETRALE, SISTEMI BIOMETRICI, SORVEGLIANZA INTELLIGENTE E MODELLI DI CONTROLLO DEGLI ACCESSI PER I DATA CENTER AI DI NUOVA GENERAZIONE.....	633
30.1 Fondamenti della sicurezza fisica dei data center: modello a strati concentrici, standard ANSI/TIA-942 e classificazione dei livelli di sicurezza in funzione del tier di disponibilità.....	633
30.2 Sistemi di controllo degli accessi fisici: tecnologie RFID, smart card, PIN pad, lettori biometrici e il protocollo PACS — architettura dei sistemi IAM fisici per data center multi-tenant	636
30.3 Sistemi biometrici per i data center AI: riconoscimento delle impronte digitali, dell'iride, del volto e del palmo — metriche FAR/FRR/EER e protezione dei template biometrici	640
30.4 Videosorveglianza intelligente per i data center AI: TVCC IP ad alta risoluzione, video analytics con deep learning, riconoscimento comportamentale e integrazione con il PACS	643
30.5 Minacce fisiche avanzate e contromisure: insider threat, social engineering fisico, protezione delle infrastrutture critiche e gestione degli incidenti di sicurezza fisica nei data center AI.....	647
30.6 Riferimenti bibliografici del Capitolo 30	651
CAPITOLO 31 - CYBERSECURITY DEI DATA CENTER AI: ARCHITETTURE ZERO-TRUST, OPERAZIONI SOC, THREAT INTELLIGENCE E DIFESA DELLE INFRASTRUTTURE DI CALCOLO DI NUOVA GENERAZIONE.....	655
31.1 Panorama delle minacce cyber ai data center AI: superfici di attacco, vettori di compromissione e impatto economico degli incidenti di sicurezza sul ciclo di vita dei modelli	655
31.2 Architettura Zero-Trust per i data center AI: principi NIST SP 800-207, micro-segmentazione, policy engine e implementazione con Istio, OPA e SPIFFE/SPIRE	659

31.3 Security Operations Center per i data center AI: SIEM, SOAR, threat hunting e metriche MTTD/MTTR per la risposta agli incidenti nei cluster GPU	662
31.4 Hardening dell'infrastruttura AI: sicurezza dei container, CIS Benchmarks per Kubernetes, gestione delle vulnerabilità CVE e Supply Chain Security con SBOM e Sigstore	665
31.5 Threat Intelligence e hunting nei data center AI: TTP degli APT contro i laboratori AI, condivisione STIX/TAXII e rilevamento proattivo delle minacce persistenti avanzate	668
31.6 Riferimenti bibliografici del Capitolo 31	673
CAPITOLO 32 - DISASTER RECOVERY E BUSINESS CONTINUITY PER I DATA CENTER AI: METRICHE RTO/RPO, ARCHITETTURE MULTI-SITO, STRATEGIE DI BACKUP DEI MODELLI E PIANIFICAZIONE DELLA CONTINUITÀ OPERATIVA NEI CLUSTER GPU DI GRANDE SCALA.....	677
32.1 Fondamenti della continuità operativa per i data center AI: framework ISO 22301, analisi di impatto sul business e quantificazione del rischio di interruzione nei cluster GPU	677
32.2 Architetture DR multi-sito per i cluster GPU AI: active-active, active-passive, modalità warm standby e cold standby — trade-off RTO/costo nelle topologie geograficamente distribuite	681
32.3 Backup e protezione dei dati AI: strategie 3-2-1-1-0, backup dei pesi del modello, versionamento dei dataset e immutable storage per la protezione dal ransomware.....	684
32.4 Resilienza applicativa dei cluster GPU: checkpoint distribuiti, failure recovery nel training, meccanismi di retry e architetture fault-tolerant per il training di LLM a scala di migliaia di GPU	688
32.5 Test e validazione dei piani DR: tabletop exercises, DR drills, chaos engineering e metriche di maturità della continuità operativa per i data center AI	691
32.6 Riferimenti bibliografici del Capitolo 32	695
CAPITOLO 33 - IMPATTO AMBIENTALE DEI DATA CENTER AI: EMISSIONI DI CARBONIO, CONSUMO IDRICO, CICLO DI VITA DELLE APPARECCHIATURE, METRICHE DI SOSTENIBILITÀ E QUADRO NORMATIVO INTERNAZIONALE..	699
33.1 Il metabolismo energetico e carbonico dei data center AI: dalla potenza elettrica alle emissioni di gas serra — Carbon Usage Effectiveness, Scope 1-2-3 e intensità carbonica del mix elettrico	699
33.2 Il consumo idrico dei data center AI: Water Usage Effectiveness, evaporazione nelle torri di raffreddamento, stress idrico locale e strategie di riduzione del consumo d'acqua.....	703
33.3 Ciclo di vita delle apparecchiature IT ed embodied carbon: produzione, trasporto, uso e smaltimento delle GPU AI — Life Cycle Assessment e strategie di economia circolare per i data center	707
33.4 Metriche di sostenibilità dei data center AI: PUE, CUE, WUE, REF, ERE e gli standard ISO/IEC 30134 — reporting ambientale e confronto con i framework GRI, TCFD e EU Taxonomy	710

33.5 Quadro normativo e target ambientali internazionali per i data center AI: EU Energy Efficiency Directive, Corporate Sustainability Reporting Directive, Pact for Skills e impegni net-zero degli hyperscaler	714
33.6 Riferimenti bibliografici del Capitolo 33	718
CAPITOLO 34 - INTEGRAZIONE DELLE ENERGIE RINNOVABILI NEI DATA CENTER AI: SOLARE, EOLICO, IDROELETTRICO, SISTEMI DI ACCUMULO ENERGETICO, POWER PURCHASE AGREEMENT E CARBON-FREE ENERGY PER I CLUSTER GPU DI GRANDE SCALA	721
34.1 Fondamenti dell'approvvigionamento energetico rinnovabile per i data center AI: profilo di carico, variabilità delle fonti intermittenti e dimensionamento ottimale degli impianti	721
34.2 Power Purchase Agreement e mercati dell'energia rinnovabile: struttura contrattuale dei PPA, certificati GO/REC, addizionalità e strategie di hedging del rischio energetico per i data center AI	725
34.3 Sistemi di accumulo energetico per i data center AI: batterie LFP e NMC, pompaggio idroelettrico, idrogeno verde e dimensionamento ottimale degli storage per la continuità rinnovabile	729
34.4 Carbon-Free Energy e 24/7 matching: l'approccio Google CFE, la piattaforma EnergyTag e l'ottimizzazione dei workload AI in funzione dell'intensità carbonica oraria della rete	732
34.5 Casi studio di integrazione rinnovabile nei data center AI: Google Hamina, Meta Odense, Microsoft Cheyenne e il progetto nucleare per i mega-cluster del 2025-2030	736
34.6 Riferimenti bibliografici del Capitolo 34	740
CAPITOLO 35 - NET-ZERO E CARBON-AWARE COMPUTING PER I DATA CENTER AI: STRATEGIE DI DECARBONIZZAZIONE, ARCHITETTURE DI CARBON ACCOUNTING, ALGORITMI DI SCHEDULING CARBON-AWARE E PERCORSI VERSO LA NEUTRALITÀ CLIMATICA	743
35.1 Definizioni operative di net-zero, carbon neutrality e carbon negative per i data center AI: il framework SBTi, gli standard ISO 14064 e il ruolo dei carbon removal rispetto ai carbon offset	743
35.2 Architetture di carbon accounting in tempo reale per i data center AI: pipeline di misurazione delle emissioni, Carbon Intensity API, attributional vs consequential LCA e il modello ADCAS per la sostenibilità	747
35.3 Algoritmi di carbon-aware scheduling per i cluster GPU AI: scheduling temporale, geografico e ibrido — formulazione come problema di ottimizzazione, euristiche e soluzioni di reinforcement learning	751
35.4 Efficienza computazionale come leva di decarbonizzazione: FLOP per joule, scaling laws, quantizzazione e pruning per la riduzione del carbon budget del training AI	755
35.5 Percorsi settoriali verso il net-zero dei data center AI entro il 2030-2050: scenari di transizione, ruolo del carbon removal permanente, iniziative di settore (RE100, EV100, EP100) e indicatori di progresso	758
35.6 Riferimenti bibliografici del Capitolo 35	762

CAPITOLO 36 - STRUTTURA DEI COSTI DEI DATA CENTER AI: CAPEX, OPEX, COSTO TOTALE DI PROPRIETÀ, MODELLI DI PRICING DEL CLOUD E OTTIMIZZAZIONE ECONOMICA DELLE INFRASTRUTTURE DI CALCOLO AD ALTA DENSITÀ	765
36.1 Tassonomia dei costi di un data center AI: anatomia del CAPEX, driver dell'investimento iniziale e modelli di ammortamento per infrastrutture ad alta intensità di capitale.....	765
36.2 Costi operativi dei data center AI: decomposizione dell'OPEX, modelli di pricing dell'energia elettrica, costo del personale specializzato e gestione dei contratti di manutenzione	769
36.3 Total Cost of Ownership dei cluster GPU AI: modello TCO analitico, costo per FLOP, costo per token generato e benchmarking con i prezzi del cloud pubblico	772
36.4 Modelli di pricing dei servizi cloud GPU e strategie di ottimizzazione della spesa: on-demand, reserved instances, spot instances, committed-use contracts e architetture multi-cloud	776
36.5 Analisi economica delle decisioni di investimento nei data center AI: NPV, IRR, payback period, analisi di sensitiva e impatto dei sussidi pubblici e degli incentivi fiscali	779
36.6 Riferimenti bibliografici del Capitolo 36	784
CAPITOLO 37 - MERCATI GLOBALI DEI DATA CENTER AI E TREND DI INVESTIMENTO: GEOGRAFIA DEL CAPITALE, DINAMICHE COMPETITIVE, CONCENTRAZIONE DEL MERCATO E PROIEZIONI AL 2030.....	787
37.1 Dimensioni e struttura del mercato globale dei data center AI: stock di capacità installata, flussi di investimento annuo, segmentazione per operatore e analisi della concentrazione di mercato	787
37.2 Hub geografici dei data center AI: la triade USA-Europa-Asia, le localizzazioni emergenti e la geopolitica del calcolo come determinante degli investimenti transfrontalieri.....	790
37.3 Struttura degli investimenti privati e istituzionali nei data center AI: private equity, REIT, infrastrutture da debito e dinamiche di valorizzazione degli asset	794
37.4 Modelli economici dei mercati dell'energia per i data center AI: meccanismi di determinazione del prezzo dell'elettricità, congestion pricing, PPA e il problema dell'addizionalità	796
37.5 Proiezioni di mercato al 2030: scenari di crescita della domanda AI, bottleneck della supply chain dei semiconduttori, evoluzione dei prezzi GPU e impatti macroeconomici del settore.....	799
37.6 Riferimenti bibliografici del Capitolo 37	804
CAPITOLO 38 - REGOLAZIONE E SOVRANITÀ DIGITALE NEI DATA CENTER AI: QUADRO NORMATIVO GLOBALE, LOCALIZZAZIONE DEI DATI, AI ACT EUROPEO, STANDARD TECNICI DI SICUREZZA E ARCHITETTURE DI CONFORMITÀ.....	807
38.1 Il quadro normativo globale per i data center AI: GDPR, Cloud Act, trasferimenti internazionali di dati e la tensione tra sovranità digitale nazionale e architetture cloud transfrontaliere.....	807

38.2 L'AI Act europeo e le sue implicazioni architeturali per i data center AI: classificazione dei sistemi ad alto rischio, requisiti tecnici di conformità, governance e sandbox regolatori	811
38.3 Standard tecnici internazionali per la sicurezza e la conformità dei data center AI: ISO/IEC 27001, SOC 2, PCI-DSS, FedRAMP e i framework di sicurezza NIST CSF e CIS Controls.....	814
38.4 Sovranità digitale e requisiti di localizzazione dei dati: data residency, regimi di data sovereignty in Cina, Russia, India, UE e Medio Oriente e impatti sull'architettura dei cluster GPU AI	818
38.5 Architetture di conformità per i data center AI: crittografia omomorfa, secure enclaves, federated learning e trusted execution environments come strumenti tecnici di privacy by design	822
38.6 Riferimenti bibliografici del Capitolo 38	826
CAPITOLO 39 - METODOLOGIA DI PROGETTAZIONE END-TO-END DEI DATA CENTER AI: DAL BRIEF PROGRAMMATICO AL PROGETTO ESECUTIVO, INTEGRAZIONE MULTI-DISCIPLINARE E OTTIMIZZAZIONE ITERATIVA DELL'ARCHITETTURA	829
39.1 Il ciclo di vita del progetto di un data center AI: fasi progettuali, gating reviews, responsabilità disciplinari e modelli di consegna del progetto dalla commessa al collaudo.....	829
39.2 Il Basis of Design come strumento di integrazione disciplinare: modellazione del profilo di carico IT, sizing dei sistemi di alimentazione e raffreddamento e analisi delle ridondanze.....	833
39.3 Progettazione integrata dei sistemi IT e delle infrastrutture fisiche: co-design di rack GPU, sistemi DLC, distribuzione elettrica e architettura di rete come processo iterativo multi-obiettivo	836
39.4 Modellazione energetica e simulazione del comportamento termico: CFD, strumenti di simulazione del PUE e ottimizzazione del layout del data hall per la riduzione del thermal bypass	840
39.5 Value engineering, risk management del progetto e gestione delle modifiche: strumenti quantitativi per il controllo di costi, tempi e qualità durante le fasi di progettazione e costruzione	843
39.6 Riferimenti bibliografici del Capitolo 39	847
CAPITOLO 40 - GESTIONE DELLA COSTRUZIONE E COMMISSIONING DEI DATA CENTER AI: PIANIFICAZIONE COSTRUTTIVA, GESTIONE DELLA SUPPLY CHAIN, INTEGRATED SYSTEM TESTING E PROCEDURE DI ACCETTAZIONE FUNZIONALE	851
40.1 Pianificazione costruttiva dei data center AI: sequenze di lavoro, logistica del cantiere, coordinamento delle specializzazioni e gestione dei materiali critici con lunghi lead time	851
40.2 Il processo di commissioning dei data center AI: definizione, livelli di commissioning, protocolli di test pre-funzionali e funzionali e il ruolo del commissioning agent indipendente.....	855

40.3 Integrated System Testing e Acceptance Testing: test di resilienza, failover testing, Integrated System Testing (IST) a pieno carico e Operational Readiness Testing (ORT).....	859
40.4 Gestione della qualità in costruzione: piano qualità, ispezioni, test non distruttivi e documentazione as-built per i data center AI con raffreddamento a liquido ad alta densità	862
40.5 Acceptance Testing e transizione alle operazioni: criteri di accettazione contrattuale, gestione delle garanzie, periodo di presa in carico e strategia di ramp-up del carico IT	865
40.6 Riferimenti bibliografici del Capitolo 40	870
CAPITOLO 41 - OPERAZIONI E GESTIONE DEL CICLO DI VITA DEI DATA CENTER AI: DCIM, MANUTENZIONE PREDITTIVA, SITE RELIABILITY ENGINEERING, KPI OPERATIVI E STRATEGIE DI HARDWARE REFRESH	873
41.1 Architettura operativa dei data center AI: struttura organizzativa del team Operations, modelli di turnazione, escalation matrix e integrazione tra IT Operations e Facilities Management.....	873
41.2 Data Center Infrastructure Management (DCIM): architettura dei sistemi di monitoraggio, integrazione con BMS e ITSM, analytics predittivi e digital twin operativo	877
41.3 Manutenzione predittiva e preventiva dei sistemi critici: strategie di manutenzione per UPS, chiller, sistemi DLC e hardware GPU, modelli di affidabilità e ottimizzazione degli intervalli di manutenzione	881
41.4 Site Reliability Engineering per i data center AI: SLO, SLA, error budget, automazione del runbook e chaos engineering applicata all'infrastruttura fisica	885
41.5 KPI operativi dei data center AI: metriche di efficienza, benchmark di settore, sustainability reporting e strategie di ottimizzazione continua del PUE e del CUE nell'operazione quotidiana	889
41.6 Riferimenti bibliografici del Capitolo 41	893
CAPITOLO 42 - DATA CENTER AI-NATIVI: ARCHITETTURE HARDWARE-SOFTWARE CO-DESIGN PER IL TRAINING E L'INFERENZA DI MODELLI FONDAZIONALI, FABRIC DI INTERCONNESSIONE ULTRA-DENSA E ORCHESTRAZIONE AUTONOMA DELL'INFRASTRUTTURA	897
42.1 Definizione e tassonomia dei data center AI-nativi: differenze architetturali rispetto ai data center general-purpose, requisiti di co-design hardware-software e caratteristiche distintive del workload AI	897
42.2 Architetture di interconnessione ultra-densa per il training distribuito: NVLink Switch, InfiniBand XDR, topologie dragonfly+ e rail-optimized e le limitazioni fisiche della comunicazione inter-GPU	901
42.3 Architettura di memoria per i modelli fondazionali: HBM, NVMe disaggregata, disaggregazione della memoria e tecnologie emergenti di memoria computazionale per eliminare il memory wall nei LLM.....	905
42.4 Orchestrazione autonoma dell'infrastruttura AI: scheduling disaggregato delle risorse, preemption del training, gestione dei checkpoint e ottimizzazione dinamica dell'energia nei data center AI-nativi	908

42.5 Architetture emergenti per i data center AI-nativi: chip specializzati per AI (TPU, Gaudi, Trainium), fotonica integrata per le interconnessioni e il futuro dei data center liquid-immersed a densità estrema	911
42.6 Riferimenti bibliografici del Capitolo 42	915
CAPITOLO 43 - INFRASTRUTTURE PER IL QUANTUM COMPUTING NEI DATA CENTER: TECNOLOGIE DI QUBIT, REQUISITI CRIOGENICI, INTEGRAZIONE CON L'INFRASTRUTTURA CLASSICA E ROADMAP VERSO IL QUANTUM ADVANTAGE.....	919
43.1 Fondamenti del quantum computing per l'infrastruttura: qubit, sovrapposizione, entanglement, decoerenza e le metriche di qualità dei processori quantistici rilevanti per le decisioni infrastrutturali	919
43.2 Sistemi criogenici per il quantum computing: dilution refrigerator, catena di raffreddamento a stadi multipli, consumo energetico, footprint e scalabilità verso centinaia di qubit fisici	923
43.3 Integrazione dei sistemi quantum con l'infrastruttura classica del data center: elettronica di controllo, schermatura EMI, comunicazione ibrida quantum-classical e latenza del feedback classico	927
43.4 Quantum networking nei data center: protocolli di distribuzione dell'entanglement, quantum repeater, interfacce memoria-fotone e architetture di quantum internet per l'interconnessione di processori quantistici distribuiti.....	931
43.5 Roadmap verso il quantum advantage nei data center: traiettoria tecnologica, applicazioni ibride quantum-classical, modelli economici e ruolo dei data center quantistici nelle architetture HPC ibride del futuro	934
43.6 Riferimenti bibliografici del Capitolo 43	938
CAPITOLO 44 - DATA CENTER SPAZIALI E SUBACQUEI: ARCHITETTURE DI CALCOLO IN AMBIENTI ESTREMI, VINCOLI TERMODINAMICI, MODELLI DI CONNETTIVITÀ E PROSPETTIVE DI DEPLOYMENT PER L'EDGE AI DISTRIBUITA.....	941
44.1 Data center subacquei: il progetto Natick di Microsoft, termodinamica del raffreddamento oceanico, pressione idrostatica e analisi dell'affidabilità in ambienti iperbarici-corrosivi	941
44.2 Scalabilità, manutenzione remota e modelli operativi dei data center subacquei: ciclo di vita, sistema k-out-of-N, recupero e sostenibilità ambientale nel lifecycle assessment	945
44.3 Data center spaziali su orbita terrestre bassa: rationale ingegneristico, vincoli termodinamici del vuoto, generazione energetica solare e comunicazioni laser per il calcolo AI in orbita	949
44.4 Processori radiation-hardened, on-board AI processing e costellazioni di nanosatelliti come fabric di calcolo distribuito: CCR, guasti da radiazione e architetture TMR.....	953
44.5 Data center lunari e interplanetari: latenza a velocità della luce, energia nucleare spaziale, Delay-Tolerant Federated Learning e il calcolo AI per l'esplorazione autonoma del sistema solare	955
44.6 Riferimenti bibliografici del Capitolo 44	961

CONCLUSIONI - DATA CENTER NELL'ERA DELL'INTELLIGENZA ARTIFICIALE: SINTESI DEI PARADIGMI FONDATIVI, CONVERGENZE INTERDISCIPLINARI E PROSPETTIVE DI RICERCA.....	965
C.1 Il data center come infrastruttura critica dell'economia digitale: riepilogo dei paradigmi fondativi e delle trasformazioni strutturali documentate nelle tredici parti del volume	965
C.2 Il framework ADCAS come struttura teorica unificante: formalizzazione dei vincoli inter-layer e applicazioni alla progettazione olistica dei data center AI-nativi	968
C.3 Sostenibilità, sovranità digitale e geopolitica delle infrastrutture: le tensioni strutturali del prossimo decennio e le direzioni di ricerca aperte	971
C.4 Convergenza delle architetture future e sintesi dei contributi originali del volume: quantum-classical hybrid, edge planetario e il data center come sistema cibernetico autonomo	974
C.5 Prospettive di ricerca, direzioni per i dottorandi e il contributo atteso del settore dei data center alla civiltà digitale del XXI secolo	976
C.6 Riferimenti bibliografici delle Conclusioni	980
APPENDICE - GLOBAL DATA CENTER ATLAS: ANALISI TECNICA COMPARATIVA DEI PRINCIPALI DATA CENTER NEI PAESI DEL G20 E NEI PRINCIPALI HUB DIGITALI MONDIALI	983
A.1 Metodologia dell'Atlas, metriche di classificazione e framework comparativo	983
A.2 Nord America: Northern Virginia, Pacific Northwest, Texas e Canada come pilastri dell'infrastruttura cloud globale	985
A.3 Europa: Francoforte, Paesi Nordici, Dublino, Amsterdam e i nuovi hub della sovranità digitale europea	988
A.4 Asia-Pacifico: Singapore, Cina, Giappone, Corea del Sud e Australia tra crescita esplosiva e sfide climatiche	991
A.5 Mercati emergenti: Brasile, India, Emirati Arabi Uniti e la traiettoria di crescita del Global South	994
A.6 Riferimenti bibliografici dell'Appendice	998

Prefazione

Questo libro nasce da una doppia necessità: quella scientifica e formativa da un lato e realizzativa dall'altro. La necessità scientifica è quella di disporre di un testo che affronti con rigore analitico, profondità ingegneristica e ampiezza sistematica l'infrastruttura tecnologica che oggi sorregge l'intera economia digitale globale — il data center moderno, nella sua accezione più evoluta e più esigente, quella imposta dai carichi di lavoro dell'intelligenza artificiale. La necessità formativa è quella di offrire a studenti di dottorato, ricercatori e professionisti del settore uno strumento che sappia muoversi con uguale competenza tra la formalizzazione matematica delle relazioni inter-layer, la specificazione ingegneristica delle architetture di raffreddamento, l'analisi economica dei modelli di finanziamento e l'inquadramento geopolitico della sovranità digitale, in modo da avere professionisti pronti alla realizzazione.

Chiunque abbia lavorato con dati, modelli di apprendimento automatico, sistemi distribuiti o reti su larga scala conosce la vertigine che si prova quando si comprende davvero — non soltanto intuitivamente, ma in modo quantitativo e strutturale — quante risorse fisiche, quanta energia e quanta complessità ingegneristica si celano dietro un singolo ciclo di training di un Large Language Model o dietro il flusso ininterrotto di richieste di inferenza che supporta un servizio cloud di scala planetaria. Un modello con settecento miliardi di parametri addestrato su decine di migliaia di GPU per mesi consuma energie dell'ordine dei gigawattora; un hyperscale data center in piena operatività può assorbire centinaia di megawatt dalla rete elettrica; la latenza di un singolo packet tra server appartenenti a rack differenti, se moltiplicata per i milioni di scambi che avvengono ogni secondo in un cluster di GPU, diventa il collo di bottiglia che determina l'efficienza dell'intero sistema di training. Queste realtà non sono marginali: sono centrali, strutturali e meritano un trattamento all'altezza della loro complessità.

Il lettore ideale di questo libro è plurale. Può essere il dottorando in informatica, ingegneria informatica, ingegneria elettrica o ingegneria civile che necessita di una visione sistemica dell'infrastruttura fisica che ospita i propri esperimenti computazionali. Può essere il ricercatore in sistemi distribuiti, in cloud computing o in intelligenza artificiale che desidera comprendere le relazioni formali tra i vincoli di risorse ai livelli inferiori dello stack tecnologico e le prestazioni osservabili ai livelli