Highlights dalla conferenza Hepix Fall 2008

Giovanni Organtini
Sapienza Università di Roma &
INFN-Sez. di Roma

Outline

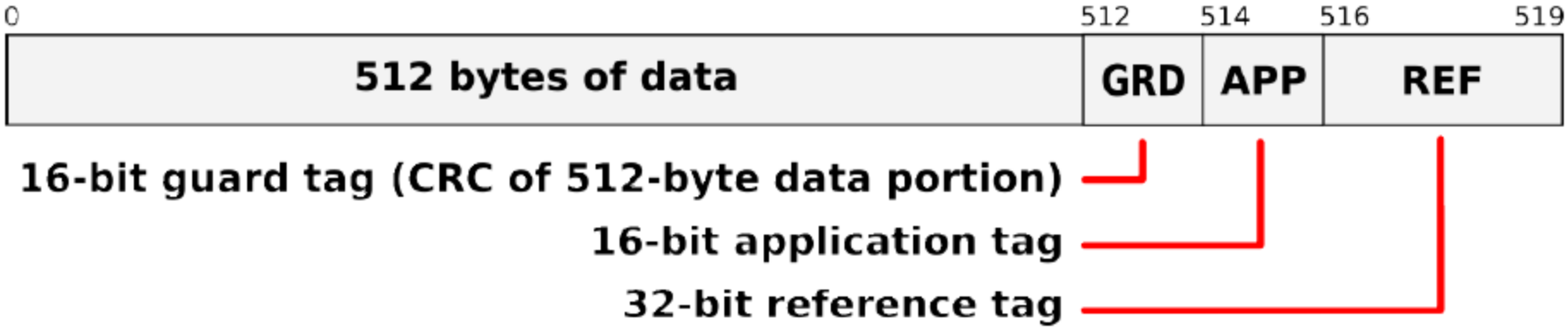
- Avviso: selezione topics personale
- Tecnologie emergenti
- Site Reports
- Il Tier-2 di Roma

Teconologie emergenti

- Cloud Computing/SaaS (Software as a Service)
  - *IT-enabled capabilities delivered as a service to external customers using Internet technology*
  - Esempi: salesforce.com, google app engine, amazon webservices
- Outsourcing di servizi
  - Ostacoli economici
  - Ostacoli culturali

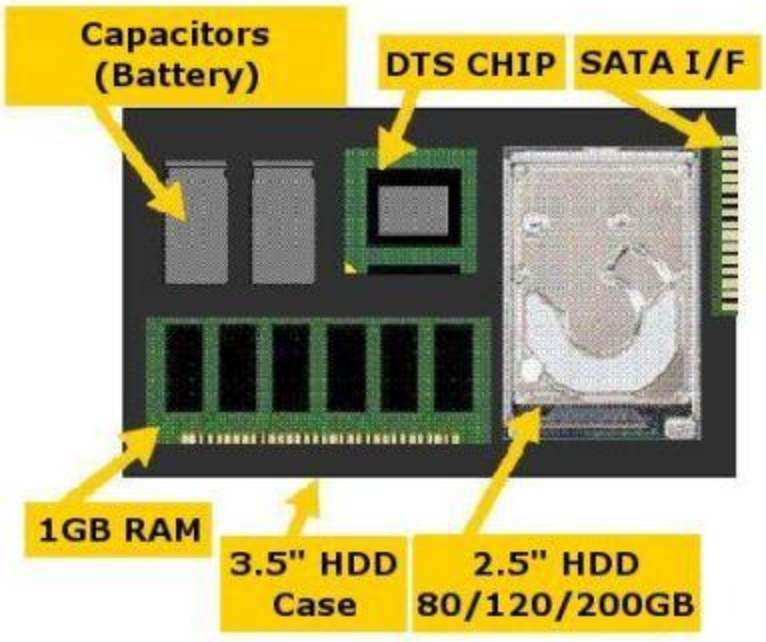
Tecnologie emergenti

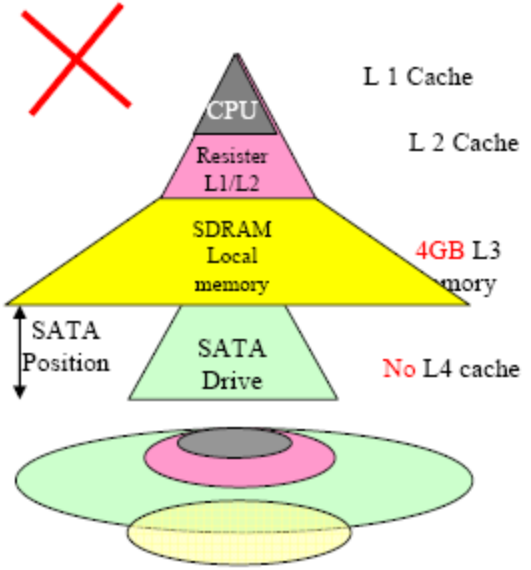
- End-to-end data protection
  - Silent data corruption
  - Sistemi in uso verificano solo una parte
  - DIF: aggiunta di 512 B/blocco su cui scrivere dati d'integrità che dipendono dall'applicazione
  - Hardware compatibili/applicazioni trasparenti
  - 1.5% di overhead



Tecnologie emergenti

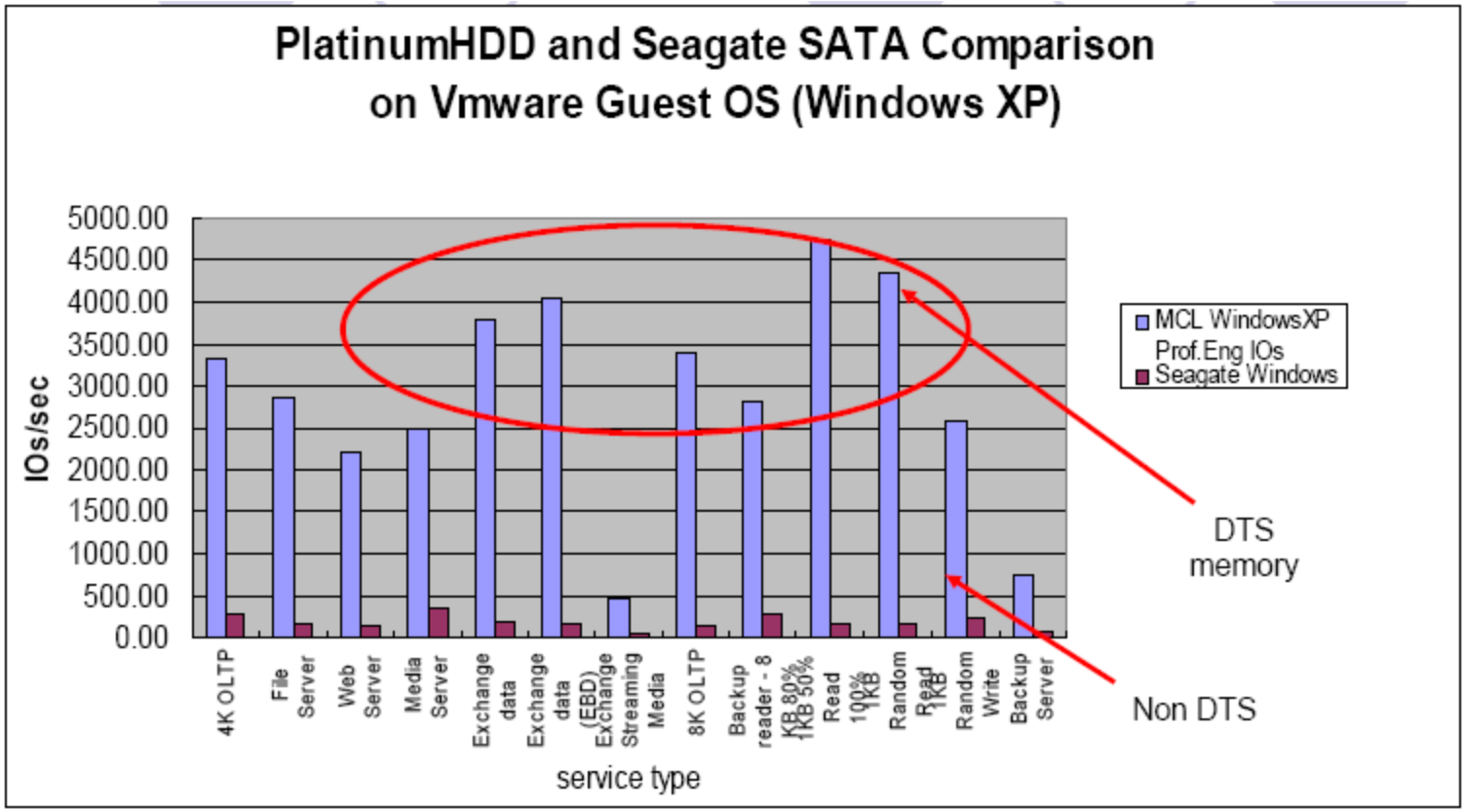
- DTS
  - Intelligenza a bordo per algoritmo di caching
  - Efficienza energetica
  - Reliability



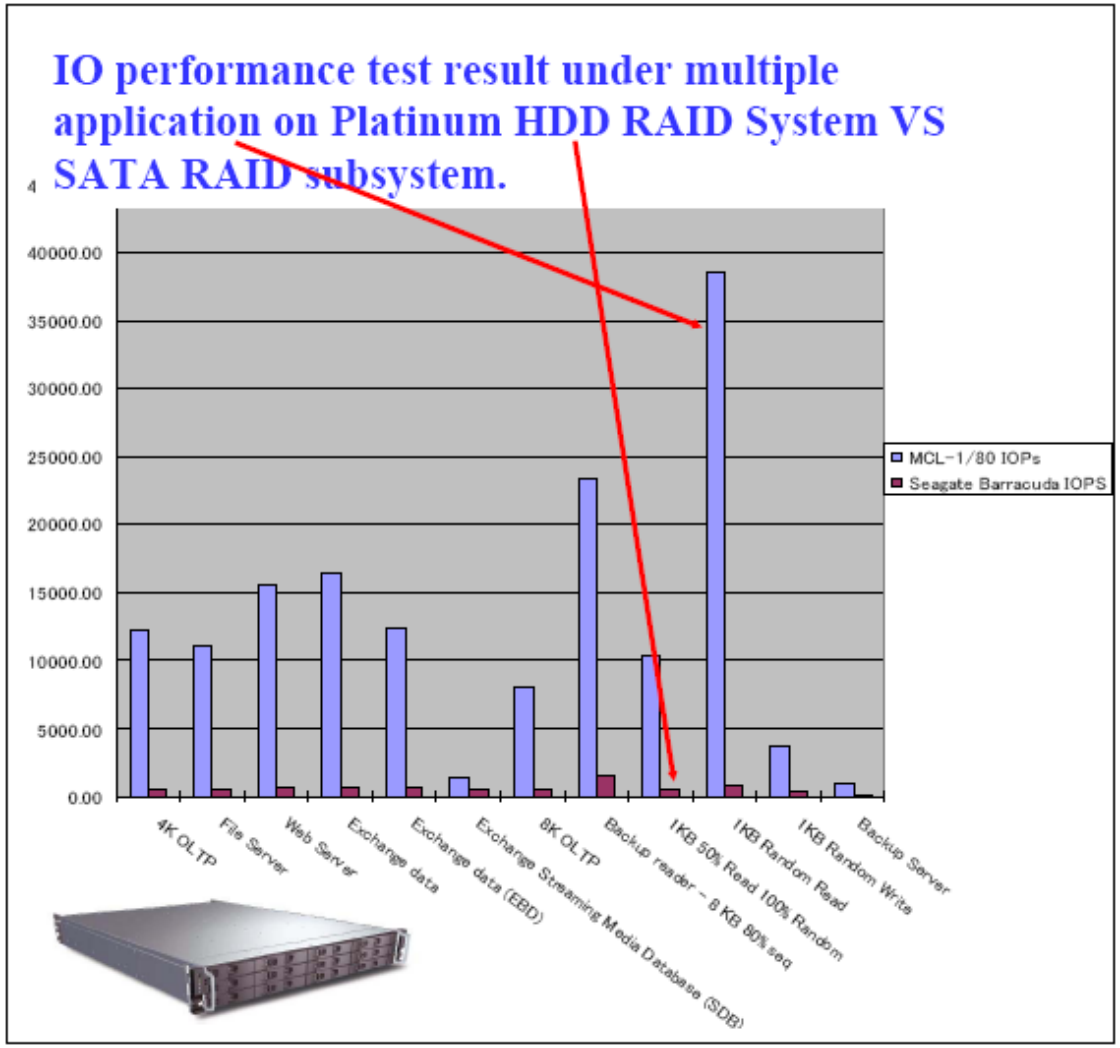


Traditional Computer memory hierarchy.

Tecnologie emergenti



Tecnologie emergenti



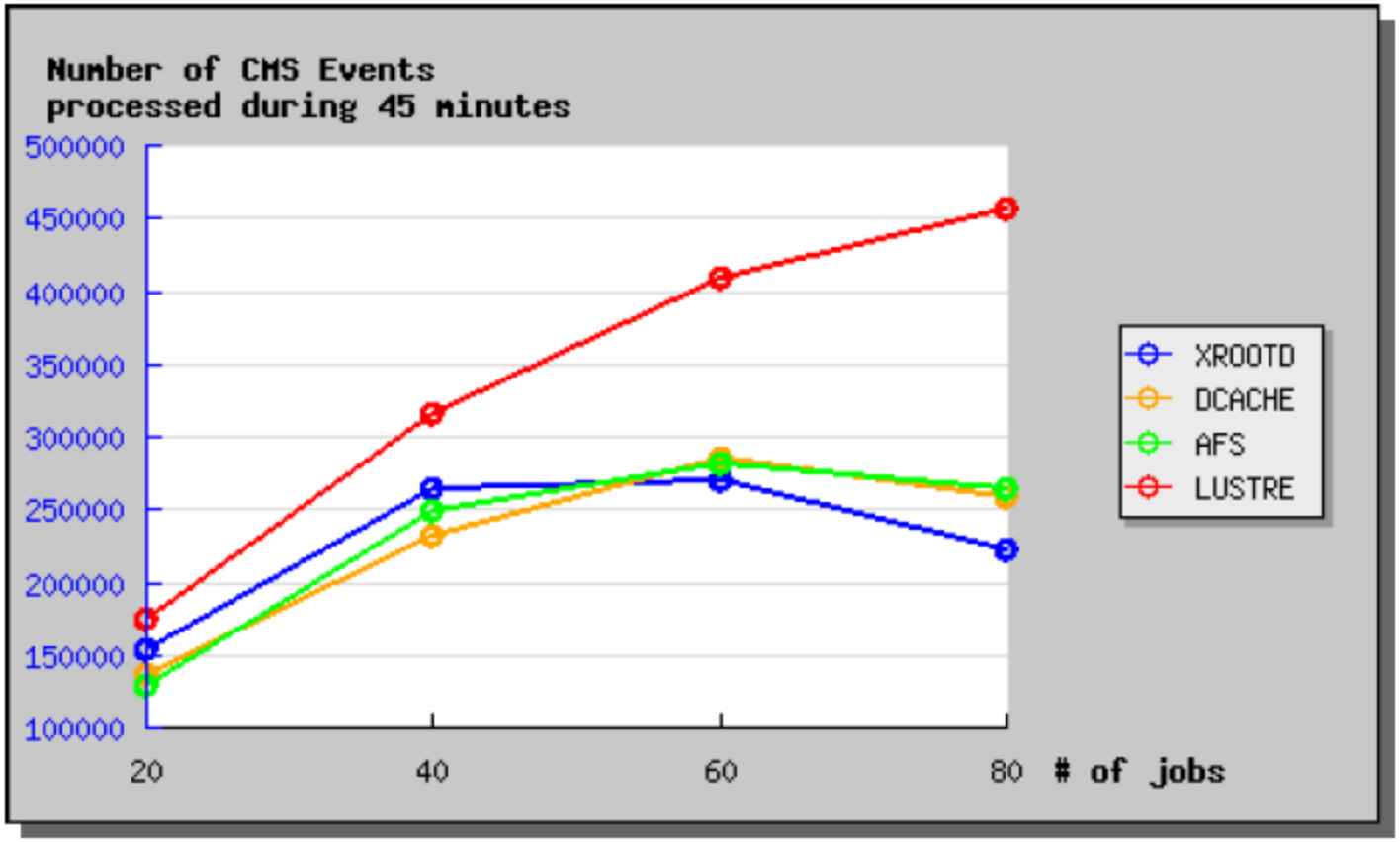
Tecnologie emergenti

- Lustre
  - Filesystem distribuito Sun Open Source
  - Prestazioni eccezionali (miglioramento di fattori 2 su molti job concorrenti)
  - Interessante il talk di Roth (problemi di Lustre)

Tecnologie emergenti

INFN Istituto Nazionale di Fisica Nucleare

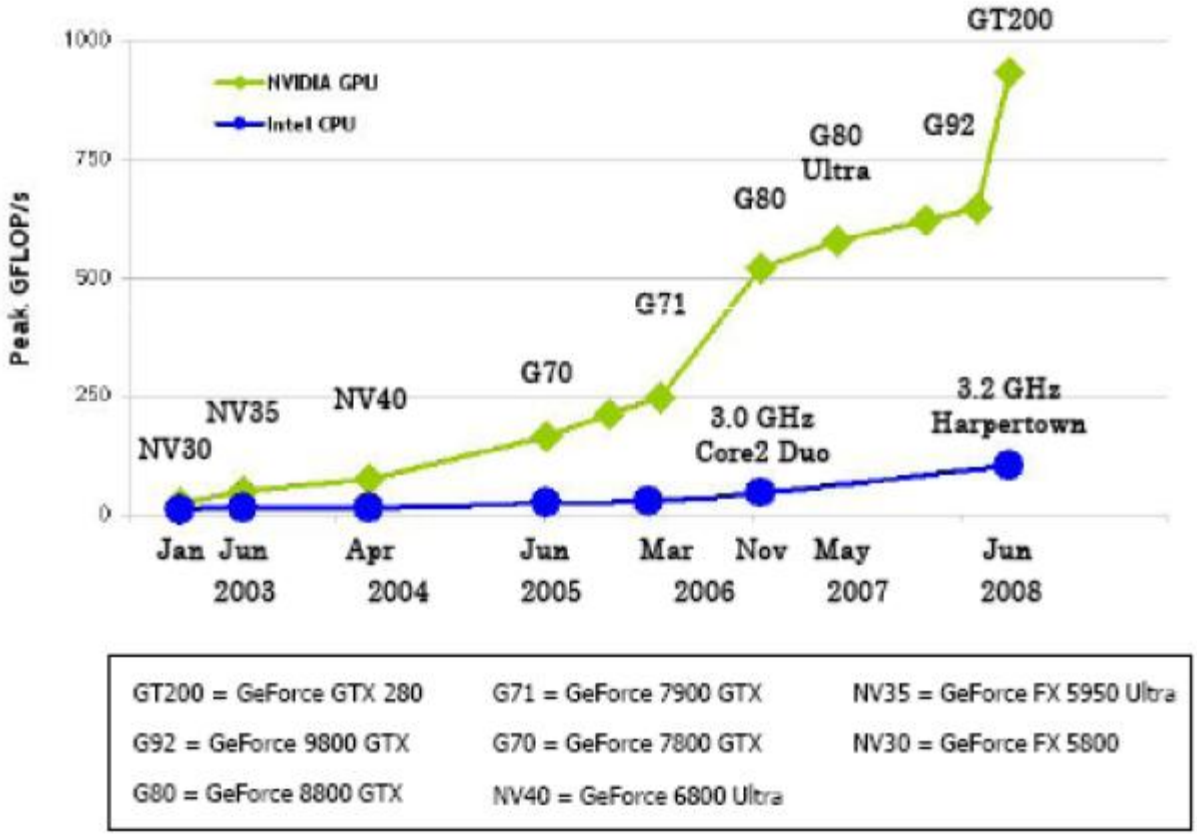
First results (16/10/2008)



Tecnologie emergenti

- Monitoring
  - Monitoring del monitoring
  - CluMan
- Workflow
- CVS
  - Obsoleto: dismesso a fine 2009
  - si va verso SVN: a partire da gennaio
- Parallel Computing
  - Uso dei processori delle graphic card (GPU Computing): nVidia e CUDA

Tecnologie emergenti



Simple "C" Extension to Express Parallelism

```
void saxpy_serial(int n, float a, float *x, float *y)
{
    for (int i = 0; i < n; ++i)
        y[i] = a*x[i] + y[i];
}
// Invoke serial SAXPY kernel
saxpy_serial(n, 2.0, x, y);
```

Standard C Code

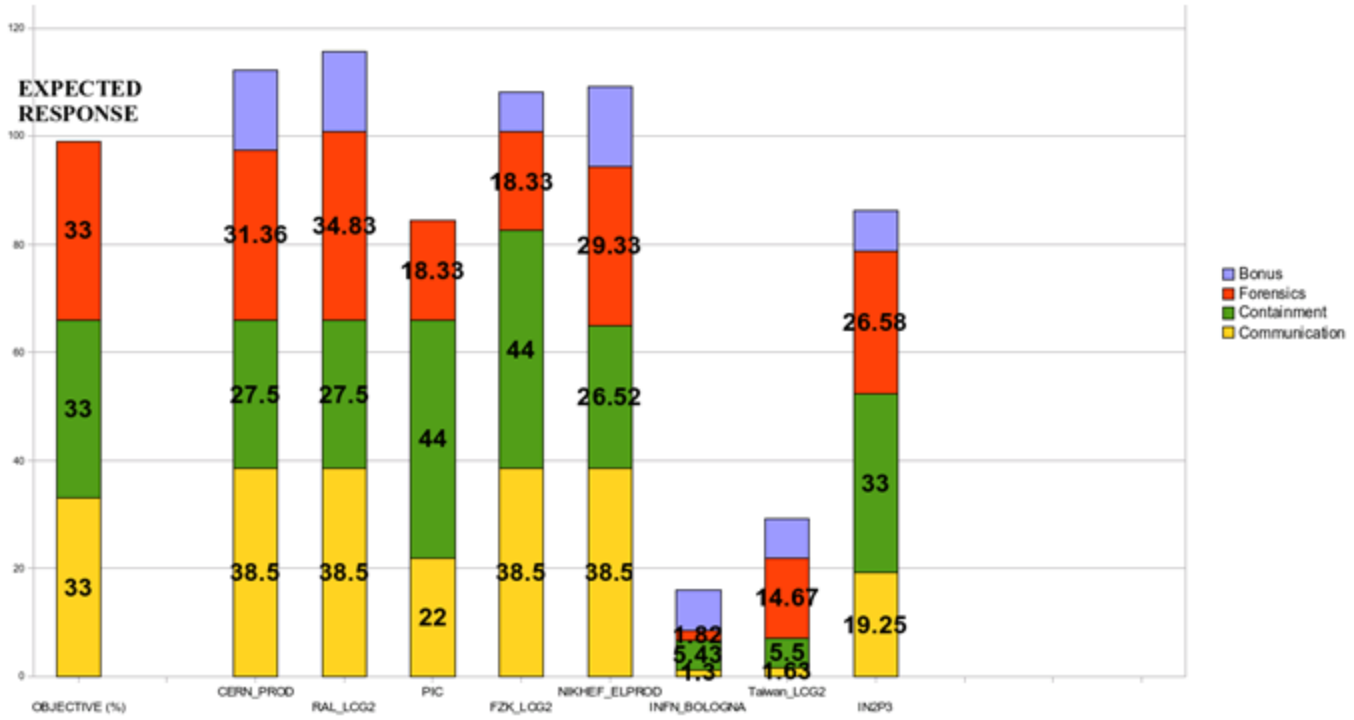
```
__global__ void saxpy_parallel(int n, float a, float *x, float
*y)
{
    int i = blockIdx.x*blockDim.x + threadIdx.x;
    if (i < n)  y[i] = a*x[i] + y[i];
}
// Invoke parallel SAXPY kernel with 256 threads/block
int nblocks = (n + 255) / 256;
saxpy_parallel<<<nblocks, 256>>>(n, 2.0, x, y);
```

CUDA C Code

Tecnologie emergenti

- ARC CE
  - Un CE con Advanced Resource Connector
  - Prodotto NorduGRID
  - Rende inutile il middleware grid su WN
- GRID Security
  - Nuove policy per testare la responsivity dei siti
  - Si immettono job *maligni* e si misura la risposta

Tecnologie emergenti



Site Reports

- INFN Tier-1
  - 2.4 PB disco (GPFS)
  - 700 server (3400 cores)
  - T2 di Roma citato come *good practice*
  - Mail: gestione locale
  - WG su IPv6
  - Esaurimento delle licenze LSF
  - Problemi finanziari (upgrade banda e altro)

Site Reports

- CERN
  - Incidenti di security in discesa
  - Fabric su SLC4 (prossimamente su SLC5)
  - No supporto per Mathematica e ingegneria elettronica
  - Il filesystem non è pronto per LHC
  - Numerosi problemi *spiccioli*, ma importanti per il manpower

Site Reports

Fabric Infrastructure and Operation (3)

CERN IT Department

- Procurements
  - Going ahead (more or less) as foreseen
  - Considering power consumption in tenders is bearing fruits
  - Selected 'fun' points:
    - Major supplier went out of business in spring
      - Orders for millions of CHF not delivered
      - 1700 systems in CC under warranty no longer maintained
    - Late deliveries
    - Rails that don't fit (mismatch with evaluation sample)
    - Spurious error messages due to defective firmware
    - Defective memory modules
    - Non-homogeneous components across batch
    - Disk drives dropped out of RAID under heavy load
    - Insufficient cooling (fans different from evaluation sample)
    - IPMI monitoring of PSUs not functional
    - Systematic mainboard failures after 1…1.5 years
    - Windows Blue Screen of Death when re-establishing power on redundant PSU
    - Surprising (for us!) route for IPMI traffic into blades
    - …

CERN Site Report for HEPiX Fall 2008 – Helge Meinhard at cern.ch - 8

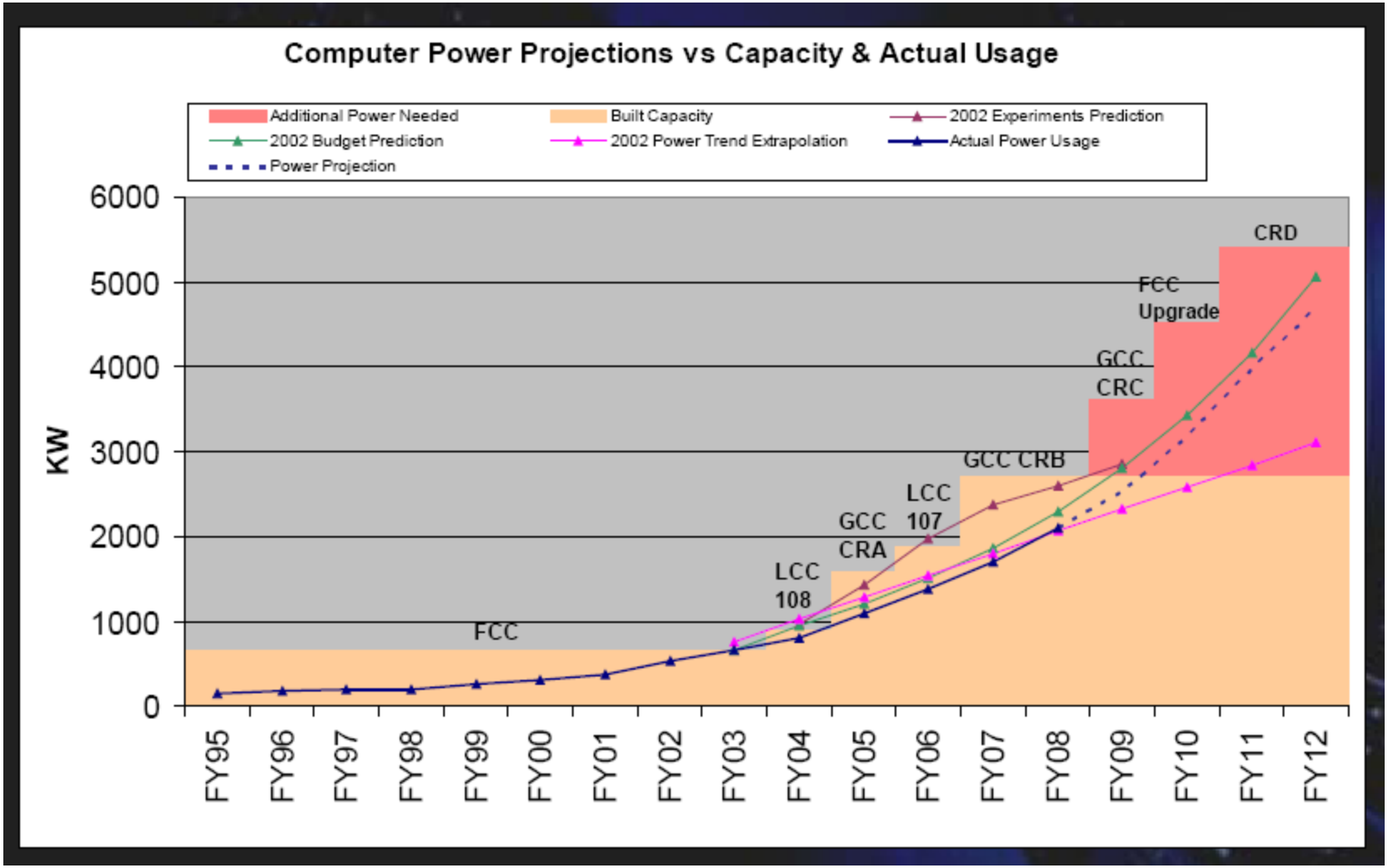
Site Reports

- CERN
  - 2.5 MW di potenza disponibile nel nuovo Centro (ma già si sa che non bastano)
  - Decommissioning di box vecchie (nel 2008 saldo positivo)
  - Si va verso il raffreddamento a acqua con 60 rack da 10 kW: un po' di confusione
  - Long term: costruzione di nuovo building
    - Problemi
    - Costi
    - Scelta entro fine 2008
    - Backup solution: affitto o sinergia con T1

Site Reports

- Fermilab
  - Piattaforme molto eterogenee (Linux/Windows)
  - 16 persone full time per il T2
    - 8 sale
    - 30000 m^2
    - 1 PB di disco

Site Reports



Site Reports

- IN2P3
  - Personale: 56 unità per production e 17 per amministrazione e management (no personale di esperimento)
  - 40 % non permanente
  - Sistema di code batch home made
  - 7200 core + 50 macchine interattive + 200 macchine per storage
  - Consumo: 1.5 MW (air cooling)
  - Centro non presidiato 24/7

Site Reports

- SLAC [National Accelerator Laboratory]
  - Non siamo soli: anche loro spendono soldi ed energie per il *marketing* ☺
  - Hanno cambiato nome pure a GLAST (FGST: Fermi Gamma ray Space Telescope)

Site Reports

- RAL
  - Nuovo building in fase di ultimazione
  - Piccoli errori di progettazione (cooling)
  - Limite agli acquisti imposto dal limite rack 15 kW
  - AntiSPAM in outsourcing
  - Problemi di performance dell'accesso ai dischi su grid
  - CASTOR si usa con molta pena

Site Reports

- DESY
  - Forzano uso di XP nel caso Microsoft
  - Usano le temperature dei PC per mappare le temperature del centro, ma è complicatissimo nonostante i soli 5 tipi diversi di hardware
  - In luglio incidente di sicurezza (chiavi rubate)

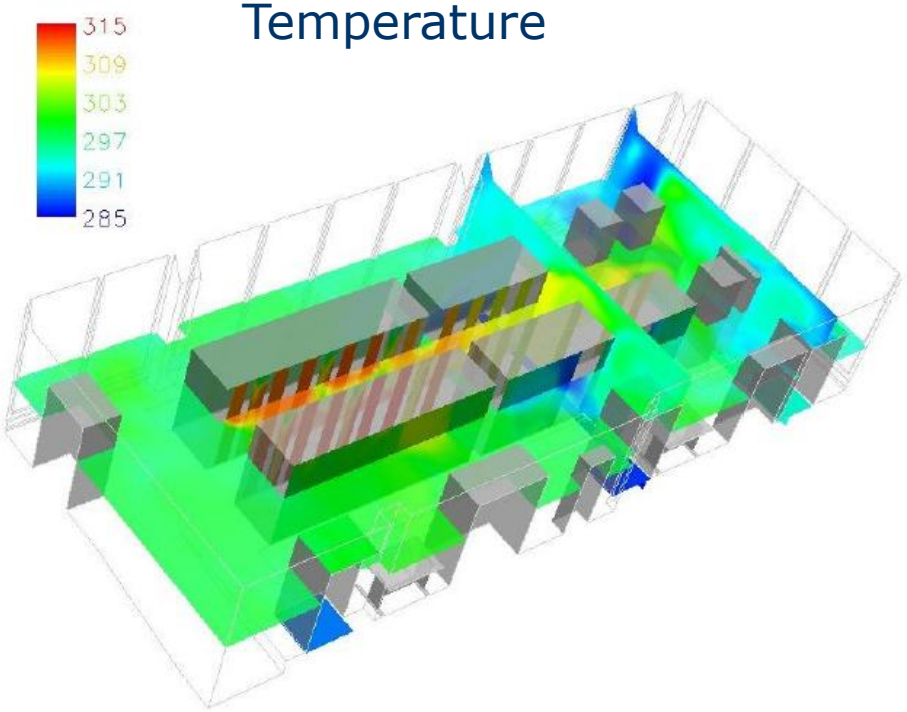
Site Reports

- GSI
  - Ottima soddisfazione degli utenti con Lustre (25 Gbps)
  - Racks a acqua @ 24 °C: soddisfattissimi

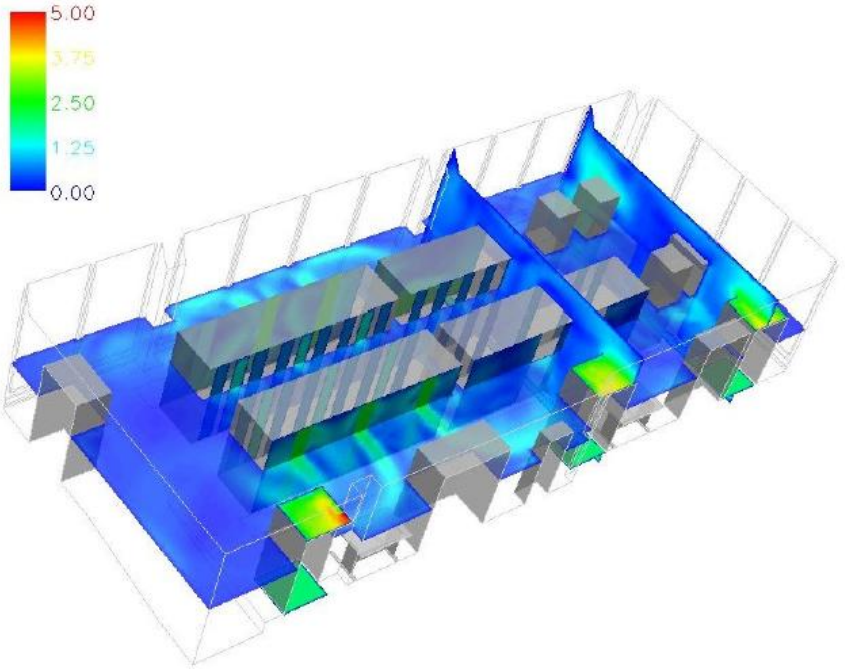
Site Reports

- Pisa
  - Simulazione fluidodinamica della sala calcolo

Temperature



Velocità aria



Site Reports

- Miscellanea
  - Sullo storage non capisce niente nessuno
  - Infiniti test con infiniti parametri: o tutti uguali oppure inconsistenti

The INFN/Roma Tier-2 Green Computing Centre

Giovanni Organtini
Sapienza Università di Roma
& INFN-Sez. di Roma

General Layout

Cooling: requirements

- Server heat production:
  - 1U multi-core server average: (220 ± 60) W
  - single core average: (38.5 ± 7.8) W
- 1 rack ~ 10 kW → 1 row ~ 70 kW
  - Disk Server size: 4U $P_D = P_S/4$
  - $P = N/2(P_S + P_S/4) = 5/8 \times (NP_S) \sim 0.6 \times 70 = 42$ kW
- Further heat sources
  - UPS, common services, insulation

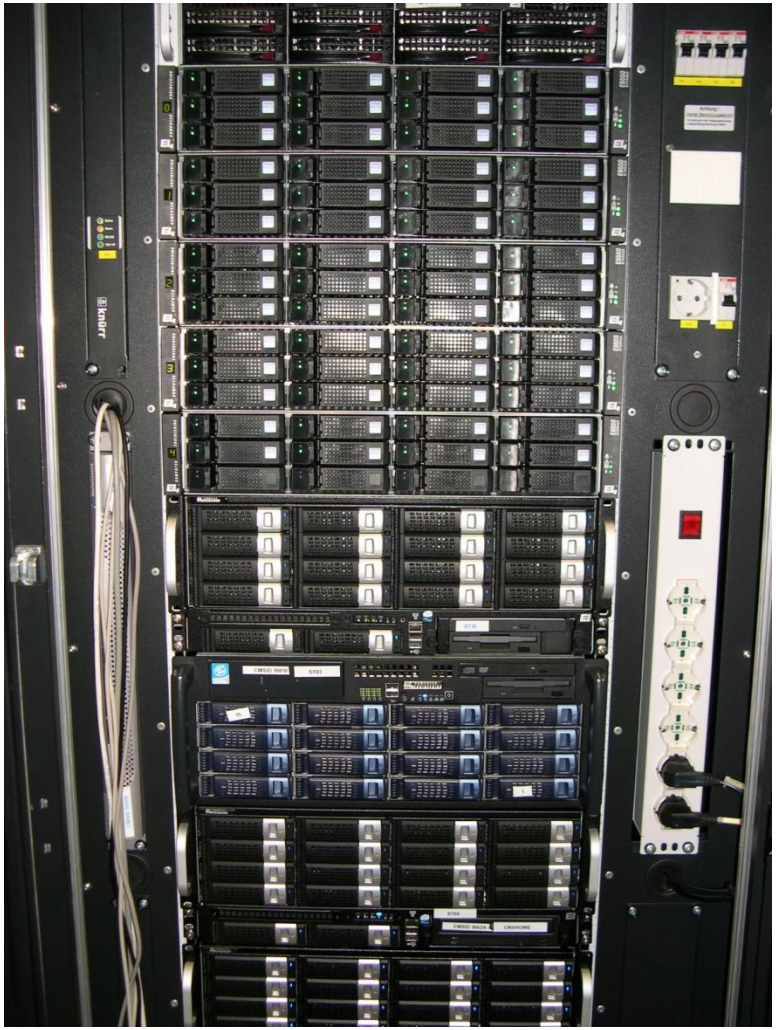
Cooling: racks

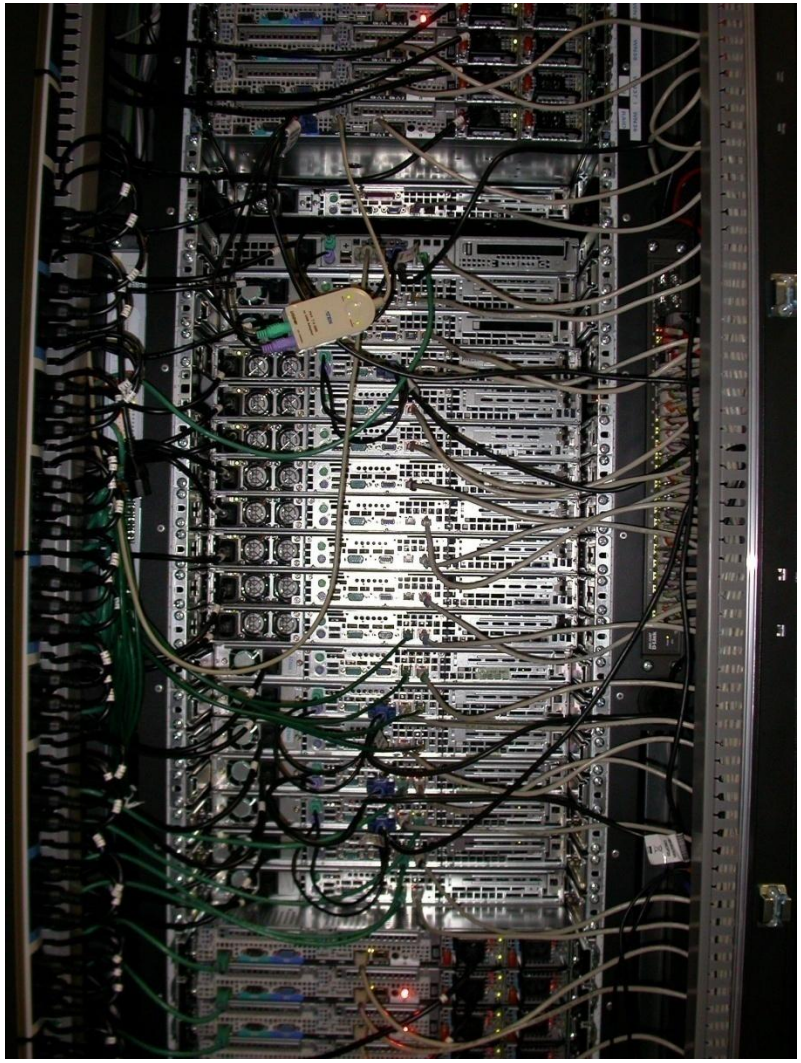
- Knürr CoolTherm 42U 17 kW
- Remote Monitoring System (RMS)
- KVM + Switches in service slots
- Home-made structured cabling

Cooling: racks

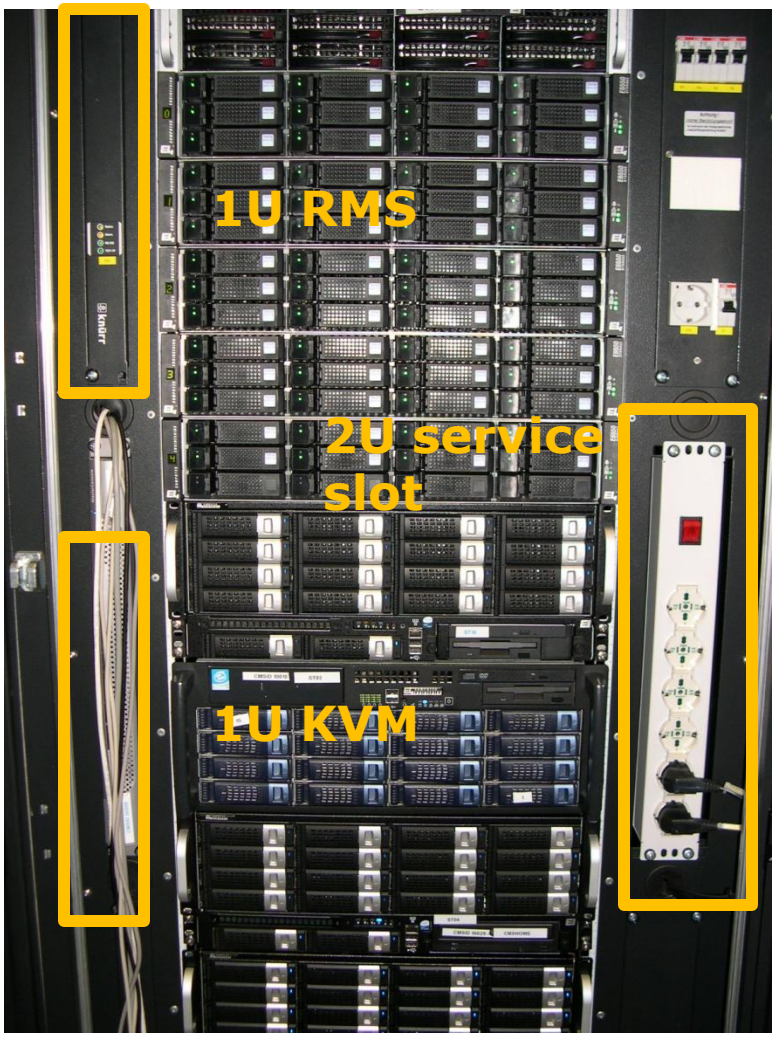
- Knürr CoolTherm 42U 17 kW
- Remote Monitoring System (RMS)
- KVM + Switches in service slots
- Home-made structured cabling

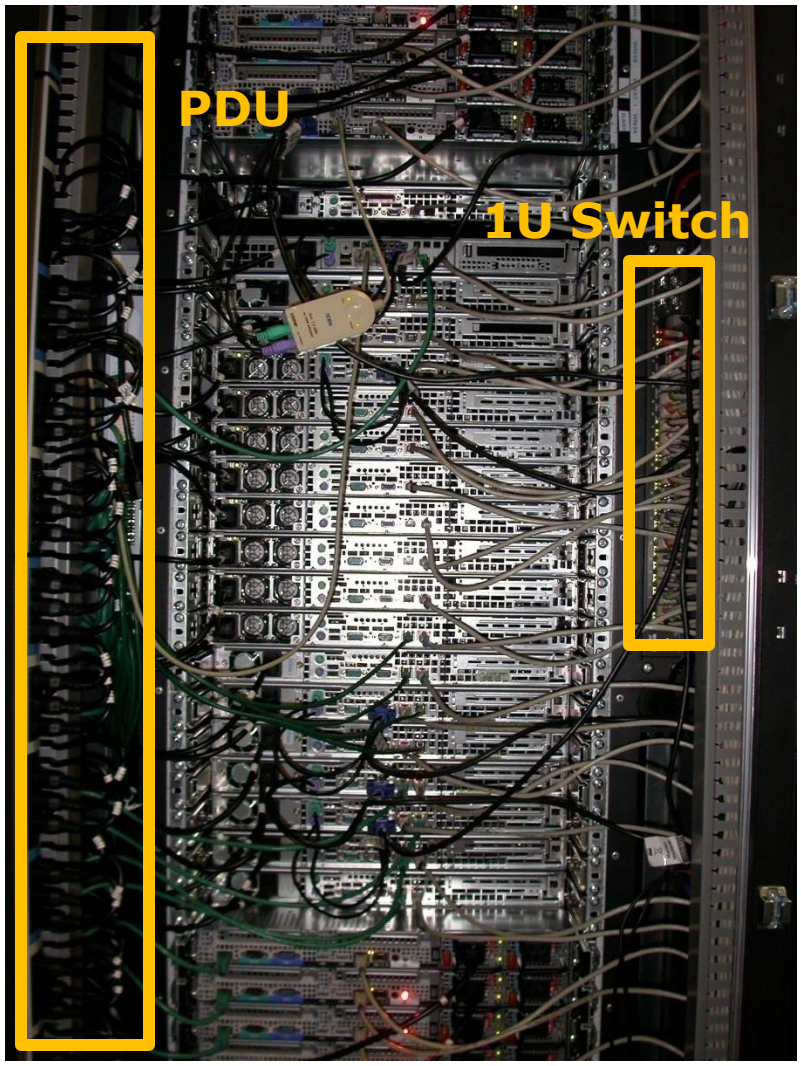
Rack layout



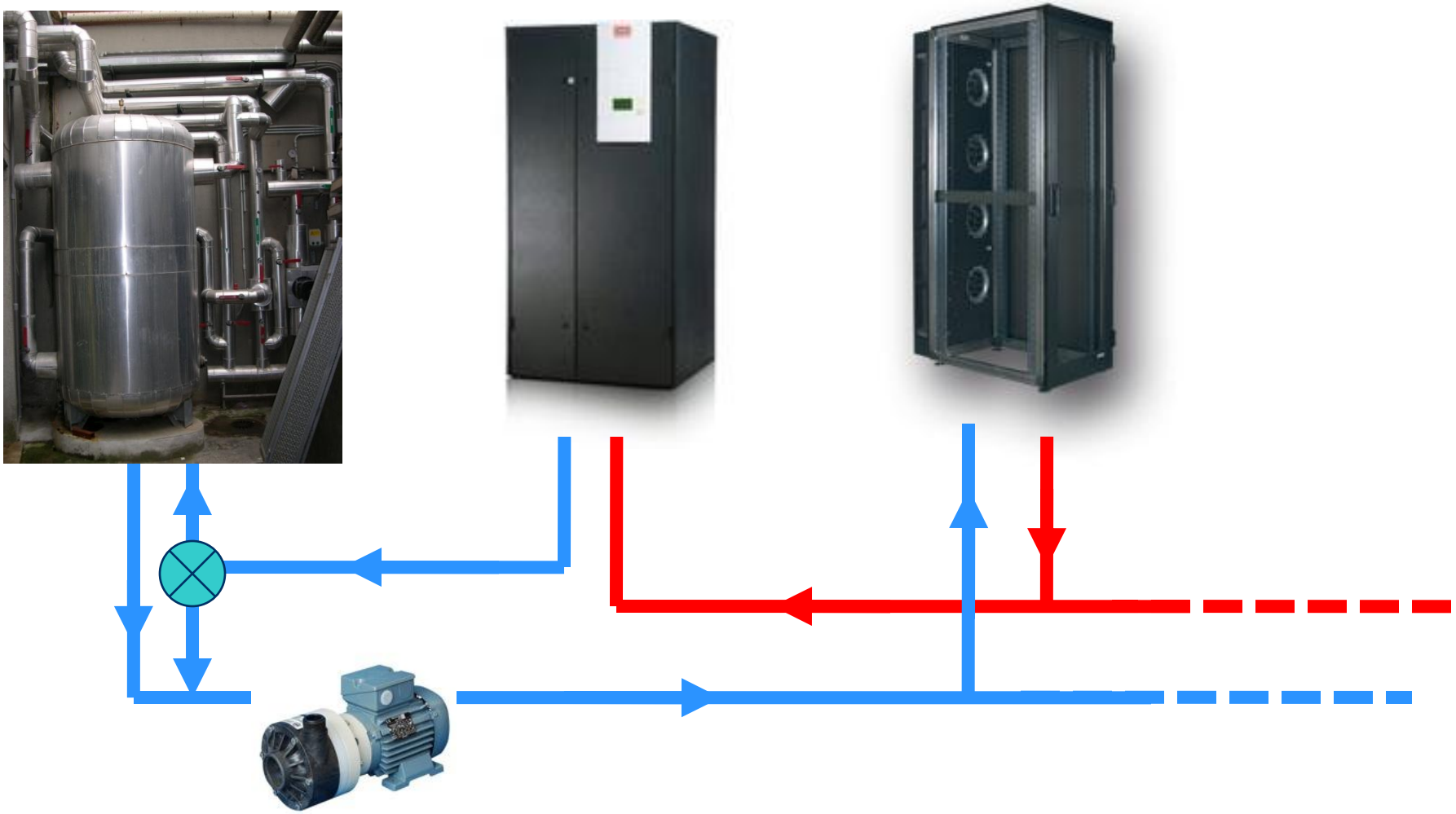


Rack layout





Cooling: circuit



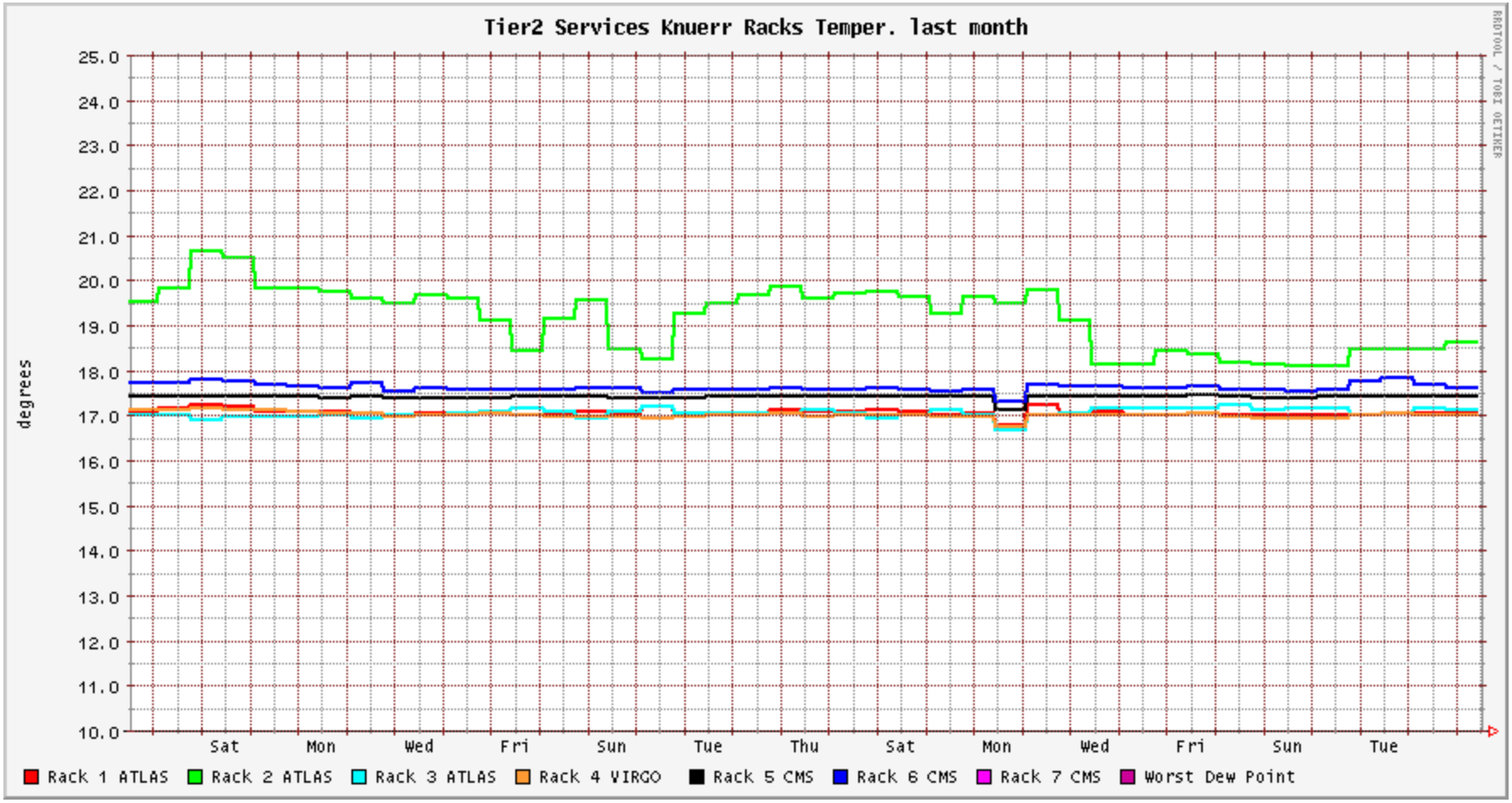
Electrical services

protected

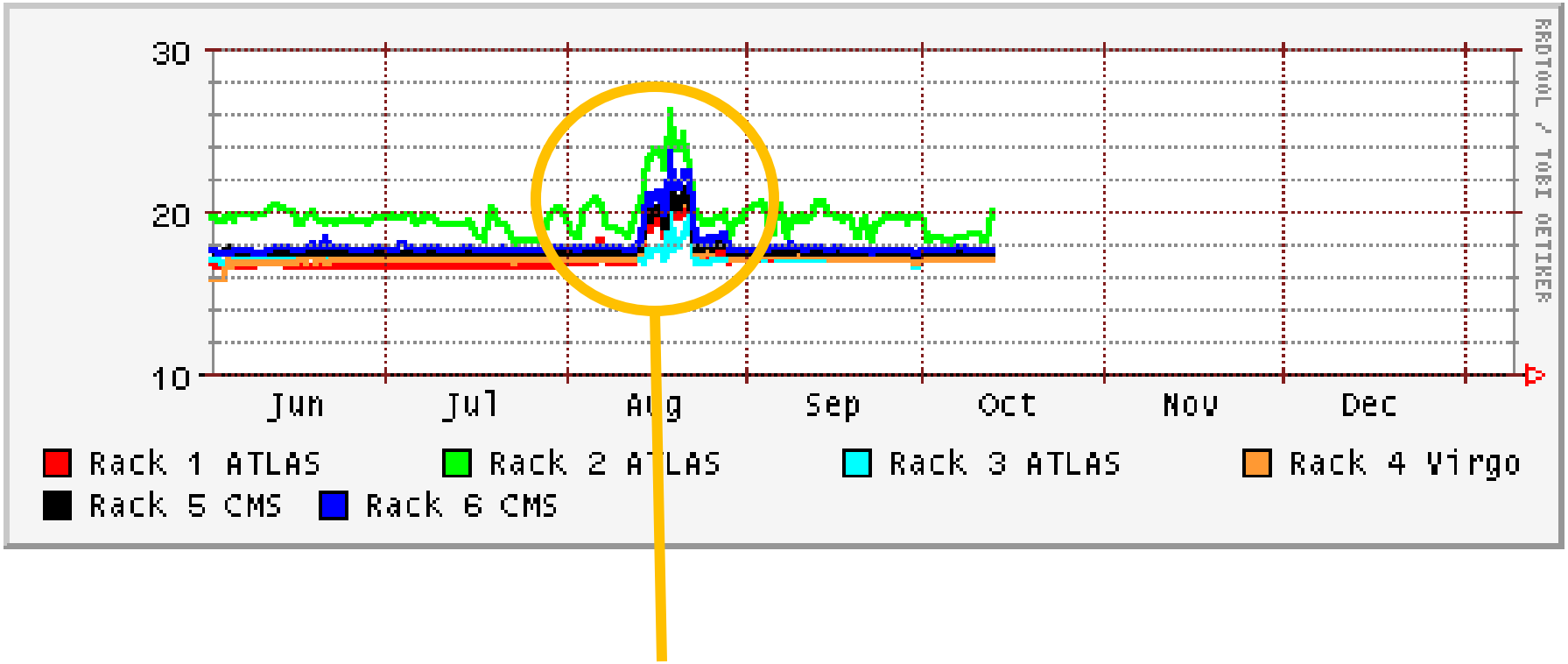
Electrical frame

unprotected

Temperature stability

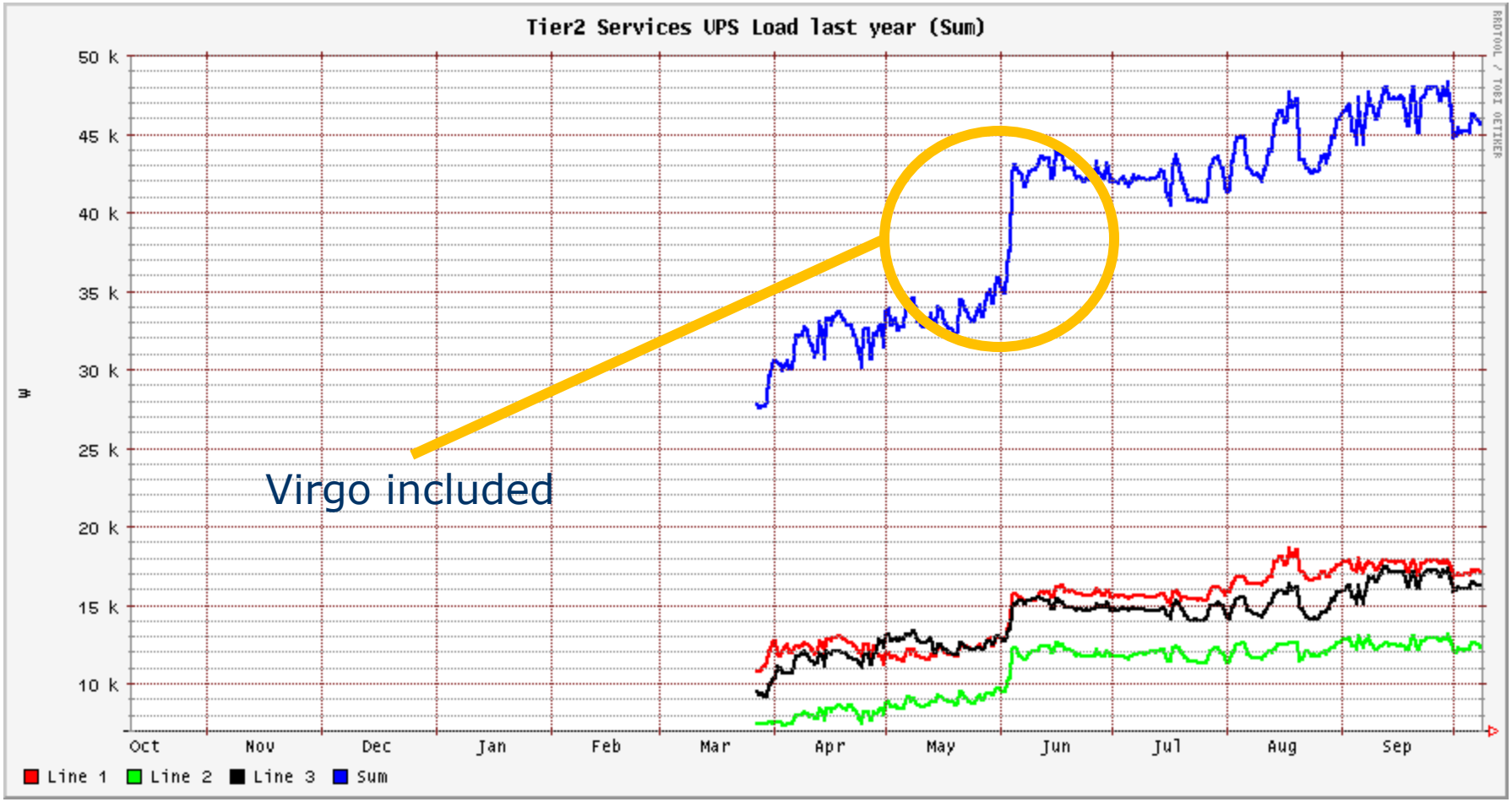


Temperature stability



Environmental conditioning failure +
Chiller 2 failure

Power consumption



Monitoring: active systems

- Custom monitoring daemon
  - Fully configurable via text configuration file
    - Periodic checks/Intercept traps
    - Customizable e-mail recipients for alarms
    - Customizable specific actions
- SMS gateway
  - Fully configurable via text configuration file
    - Standard action per user (phone call)
    - Respond to SMS executing pre-defined scripts specified in the message body
    - Software triggerable with custom message

Total power consumption

- Current power consumption for UPS-served systems: 45 kW
- Current total power consumption: 100 kW
- Chillers power consumption: 44 kW (160 kW refrigerating power)
  - Redundant system
  - Scroll type
  - Flat power consumption
- Condensers, UPS loss, etc.: 11 kW

Estimation at full load

- Power required for servers: 90 kW
- Power required for chillers: 44 kW
- Power required for services: 15 kW

- Total power requirement: 149 kW

- Power ratio = $(W_{servers}/W_{tot}) = 0.6$

Traditional cooling

- 80 W/m^2×100 m^2+90 kW+15 kW = 113 kW

- $W_{chillers}$ = 2×W/COP = 56 ÷ 113 kW
  - To be compared with 44 kW
- Large inefficiences
  - Turbulent motion
  - Larger volume and dispersion
- Power ratio tends to be lower
  - Power consumption even worse

$$COP = \frac{\Delta Q}{E_{spent}}$$