



ISTITUTO NAZIONALE DI FISICA NUCLEARE

Sezione di Pisa

INFN/code-xx/xxx
dd/mm/yy



CCR-23/2008/P

IL “GRID DATA CENTER” DELL’INFN DI PISA

Silvia Arezzini¹, Tommaso Boccali¹, Federico Calzolari², Alberto Ciampa¹, Simone Marini¹,
Enrico Mazzoni¹, Subir Sarkar¹, Sonia Taneja¹, Giuseppe Terreni¹

¹⁾ *INFN-Sezione di Pisa, Largo B. Pontecorvo, 3, I-56127 Pisa, Italy*

²⁾ *Scuola Normale Superiore di Pisa, Piazza dei Cavalieri 7, I-56126 Pisa, Italy*

Abstract

La candidatura della Sezione INFN di Pisa come membro del consorzio italiano dei Tier 2 di CMS è stata proposta nel Giugno 2005. Partendo dall’inizio del *commissioning* avvenuto alla fine del 2006, si descrive l’attuale stato di ciò che, nel corso del progetto, è divenuto un “GRID Data Center”. Tale descrizione è organizzata tematicamente seguendo quelle che sono le diverse componenti del centro: infrastrutture, calcolo (*farm*), rete, storage e organizzazione GRID.

Si conclude descrivendo le attività progettuali in corso e la direzione che si intende seguire per lo sviluppo futuro.

1 IMPOSTAZIONE DEL PROGETTO

L'idea del progetto nasce circa due anni e mezzo fa, nel Giugno 2005, e viene presto caratterizzata da tre concetti cardine:

- il progetto, così come la struttura che ne sarebbe risultata, avrebbe dovuto avere un carattere di Sezione e non coinvolgere solo l'esperimento CMS;
- l'Ente invitava fortemente i siti candidati a trovare fonti alternative di finanziamento, necessarie al raggiungimento degli ambiziosi obiettivi dell'operazione;
- sarebbe stato premiato in particolar modo l'impegno da parte dei siti candidati ad attrezzarsi localmente per ciò che riguardava gli investimenti in infrastrutture, cronologicamente i primi da effettuare.

Lungo queste direttrici si è mosso il gruppo di persone coinvolte a vario titolo, con i seguenti risultati:

- a Pisa il progetto CMS T2 si è gradualmente trasformato, inglobandosi in un più comprensivo progetto per la realizzazione di un "GRID Data Center" rivolto all'utenza del calcolo tecnico e scientifico di tutta la comunità afferente alla sezione INFN di Pisa, e ancor più in generale a livello nazionale;
- attorno al progetto si sono riunite, oltre alla Sezione, il Dipartimento di Fisica e la Scuola Normale Superiore;
- è stato innanzitutto individuato un sotto-progetto per la parte infrastrutturale al fine di reperire localmente le risorse necessarie.

Come si vedrà in seguito la situazione delle componenti del GRID Data Center è attualmente in diverse fasi di sviluppo. In particolare:

- le infrastrutture di base (linee elettriche e produzione di acqua refrigerata) sono già realizzate tenendo conto anche dei potenziali sviluppi futuri;
- le infrastrutture di sala (distribuzione elettrica e condizionatori di sala) sono realizzate per gestire la situazione attuale ed i prossimi sviluppi, ma non per sfruttare le complete potenzialità a disposizione;
- i *server* sono dimensionati sulla potenza di calcolo finale richiesta dal T2 di CMS ma, all'occorrenza, vi è un ampio margine di crescita;
- la rete e lo *storage* sono in una condizione di "best effort" cioè si usa al meglio ciò che si ha, sapendo che è lontano da ciò che si renderà necessario in produzione. A tal fine un progetto è stato definito e verrà brevemente riportato;
- la "qualità del servizio" è ad un buon livello ma necessita ulteriori interventi per raggiungere quanto richiesto in produzione, anche su questo punto un progetto è in via di definizione e anch'esso verrà brevemente presentato.

2 DIMENSIONAMENTO E COMPITI DEL T2-PISA DI CMS

Il percorso di realizzazione e crescita del T2 di Pisa è segnato da due requisiti posti dall'esperimento CMS, relativi a due specifici momenti:

- fine 2008: potenza di calcolo di circa 1 MSI2k (Mega SpecInt2000), capacità di *storage on line* di circa 200 TB;
- 2010 (dimensioni finali): potenza di calcolo 2.2 MSI2k, capacità di *storage* di 600 TB.

Pisa ospita una delle più grandi comunità italiane di fisici afferenti all'esperimento CMS (circa 30 fra ricercatori e associati) ed ha attivamente partecipato alla costruzione del *silicon tracker* di CMS, realizzando presso i suoi laboratori l'integrazione della parte interna.

Una delle principali attività di calcolo é quindi legata a questo aspetto: già adesso i dati raccolti a Ginevra durante la fase di *test* con raggi cosmici vengono inviati a Pisa e sono qui analizzati con strumenti di analisi distribuita. Oltre a rappresentare un'importante attività per la comprensione del *detector*, questo modello permette anche di testare i *tool* per lo spostamento di dati reali e per l'analisi via GRID che CMS ha preparato, in tutto e per tutto identici a quelli che saranno utilizzati non appena i dati delle interazioni protone-protone di LHC cominceranno ad essere raccolti. Nel 2007, quando meno di un quarto del *tracker* era in presa dati, la quantità di dati raccolta é stata ancora modesta (dell'ordine di 20 TB); nel corso dell'anno corrente invece é prevista una presa dati con *tracker* completo e di maggiore durata, che dovrebbe raccogliere una quantità di dati intorno agli 80-100 TB. Anche per questi dati é previsto uno spostamento a Pisa e un'analisi sulla *farm* locale via GRID.

Un'altra attività di *commissioning* riguarda il controllo e la verifica del modello di *computing* di CMS sotto carico. I *link* geografici che legano Pisa (Tier2 nel modello di calcolo) al nostro Tier1 di riferimento (CNAF) sono tenuti sotto pressione: procedure automatiche creano traffico fittizio in ingresso e uscita per 1-2 TB al giorno, e permettono di valutare la stabilità dei *link*. Dal lato del *farming*, analoghe procedure costantemente producono sulla *farm* pisana simulazioni Monte Carlo di canali di fisica e "background", usati per la preparazione delle analisi di fisica da CMS. Pisa, con la sua *farm* di Opteron, risulta uno dei 4 - 5 siti mondiali con più eventi prodotti, superando anche molti Tier1.

Dal luglio 2008, grazie all'impegno del GARR, il GRID Data Center può avvalersi di un *link* geografico da 1 Gb/s dedicato, costituendo anche un importante sito di *test* per l'architettura di rete geografica che sarà adottata a livello nazionale per i siti coinvolti in esperimenti LHC.

Durante l'anno 2008 il *focus* del gruppo di Pisa si sposterà gradualmente su attività legate all'analisi di fisica: dal lato *computing* questo implica un uso delle risorse da parte di un numero maggiore di utenti, spesso con poca esperienza di calcolo e con *use case* differenziati, oltre alla necessità di razionalizzare lo spostamento di dati in modo da garantire ad ogni utente i *samples* locali necessari. Già adesso Pisa, pur essendo in una situazione non ideale dal punto di vista dello *storage* presente, ospita numerosi *samples* di QCD, τ , top e Z' che i fisici pisani stanno utilizzando per preparare le stesse analisi che, a partire dal 2009, saranno effettuate su dati veri.

3 SALA CALCOLO E INFRASTRUTTURE

La sala di Pisa ha le seguenti caratteristiche:

- 95 m², spazio per 34-40 *rack* 19" (dipende dalla scelta dell'architettura di condizionamento);
- pavimento tecnico rialzato per la distribuzione dell'acqua refrigerata;
- distribuzione elettrica attraverso il contro-soffitto.

Lato "Molo Novo"

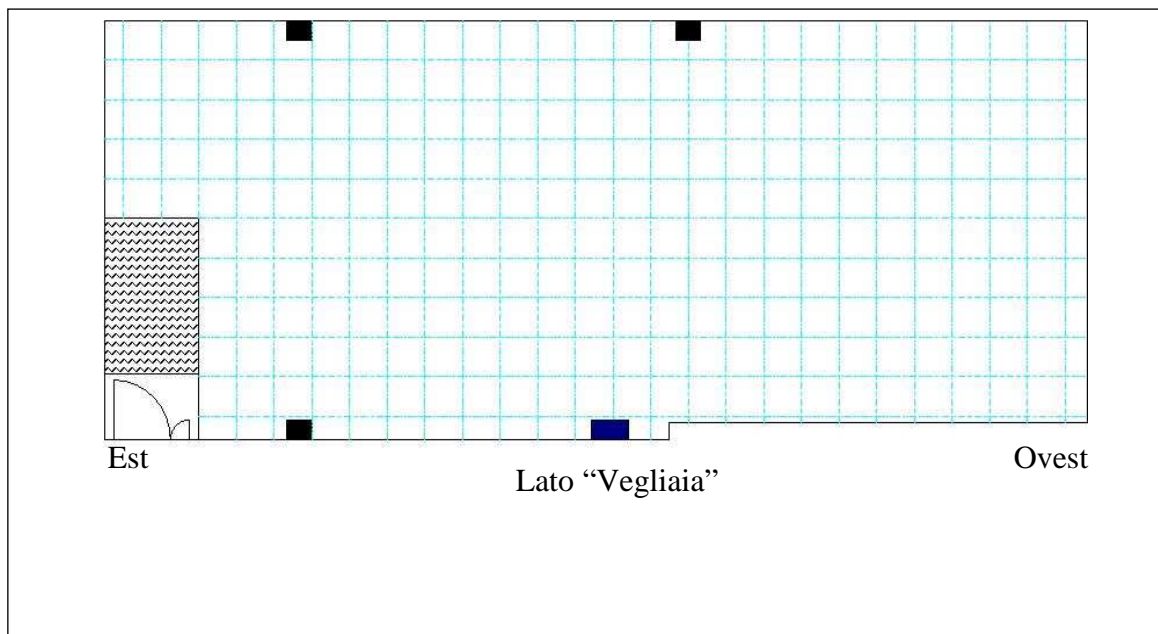


Fig. 1 - Pianta della Sala Calcolo.

3.1 Impianto elettrico

Sono installate le seguenti linee elettriche:

- 250 kW per servizi, *worker node* e *storage*:
 - Trifase 400V 200A (no UPS) per *worker node*
 - Trifase 400V 110A (UPS) per servizi e *storage*
 - Trifase 400V 50A (no UPS) per *worker node*
- 70 kW per gli impianti di condizionamento in sala: trifase 400V 100A
- 70 kW per gli impianti di condizionamento in copertura (*chiller*): trifase 400V 100A

3.2 Impianti di condizionamento

La sala è dotata di due impianti di condizionamento indipendenti, anche se è in via di realizzazione un sistema di *monitoring* e *management* che li unifichi:

- condizionamento ad espansione diretta: 3* 25 kW installati in sala;
- condizionamento a liquido refrigerato, composto da:

- 6*18 kW condizionatori APC InRow dedicati ad una “isola APC” in sala
- 2*80 kW *chiller* installati in copertura



Fig.2 - I due *chiller* in copertura e il condizionatore APC InRow.

Totale condizionamento installato: 183 kW.

Tenendo conto che l'impianto in copertura è predisposto per l'aggiunta di un terzo *chiller* da 80 kW, il totale del condizionamento potenziale raggiunge i 291 kW, in linea con la potenza elettrica disponibile.

4 FARM

I *worker node* sono gestiti da GRID come un insieme unico, ma per motivi storici ed organizzativi sono divisi in quattro *farm*:

- Scirocco (Opteron 280):
 - 300 processori Opteron *dual core* 2.4 GHz
 - 1 GB di *ram* per *core*
 - Uniwide 1U 2P
- Libeccio (Opteron 2218)
 - 300 processori Opteron *dual core* 2.6GHz
 - 1 GB di *ram* per *core*
 - Rete veloce Infiniband 10 Gb/s
 - IBM x3455 1U 2P
- Grecale (Opteron 8220)
 - 10 macchine SUN X4600 4U 8P
 - 80 processori Opteron *dual core* 2.8 GHz
 - 2 GB per *core*
- Teo (Xeon 5400)
 - 17 macchine Dell PowerEdge 1950, 1U 2P
 - 34 processori *quad core*
 - 1 GB per *core*

Attualmente sono installati quasi 1500 *core* per una potenza di calcolo di oltre 2 MSI2k.

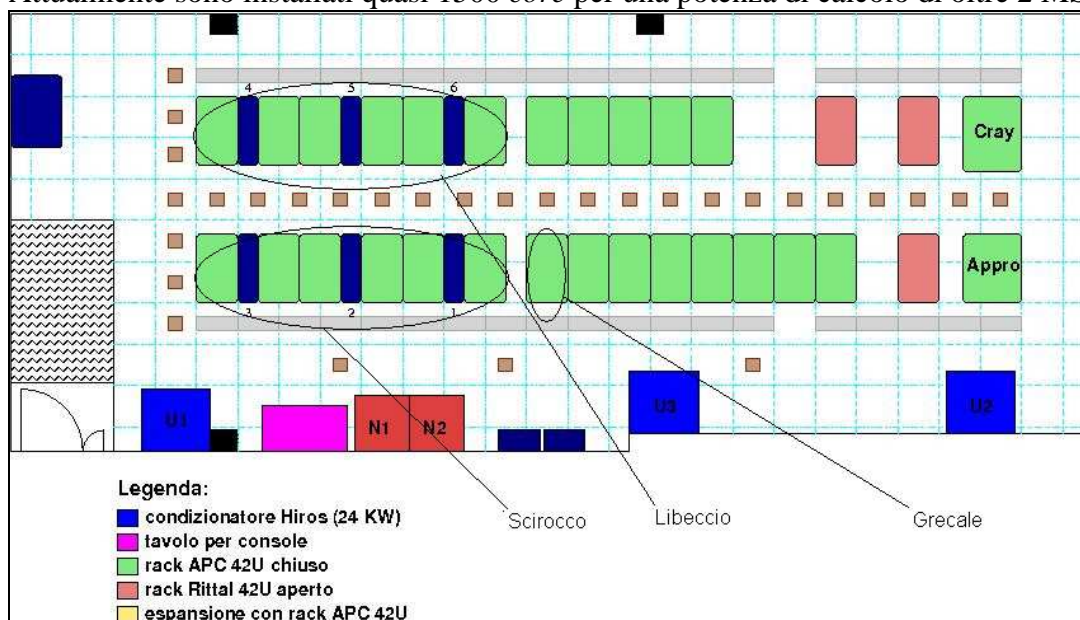


Fig.3 - Pianta della attuale Sala Calcolo.





Fig.4 – Immagini della sala calcolo.

Le *farm* Libeccio e Grecale sono in una situazione particolare, definita “in condominio”, per quanto riguarda il loro utilizzo in GRID. L’*hardware*, di proprietà di AMD, viene utilizzato per attività di *testing* e *benchmarking* da loro, con un impegno medio che non supera il 20% del tempo. Per il resto del tempo, oltre l’80%, i sistemi sono a nostra disposizione per l’uso in GRID. Per poter gestire in modo efficiente questa situazione è stato sviluppato internamente un meccanismo di *switching* automatico denominato “bottono rosso”.

4.1 Il “Bottone Rosso”

Il sistema è basato sull’ambiente PXE (ambiente di esecuzione *pre-boot*) e permette la rapida riconfigurazione dei sistemi (tutti o in parte) dalla modalità “AMD” a quella in uso per “INFN-Grid” e viceversa. Più precisamente nel caso vi sia l’esigenza di un passaggio veloce da “INFN-Grid” ad “AMD” i *job* in corso vengono re-indirizzati su altri sistemi in modo automatico, salvaguardando la loro integrità.

Per l’uso “AMD” viene installato di norma SUSE Linux SLES10 e i sistemi non vengono messi in GRID. Per l’uso “INFN-Grid” il sistema operativo è SLC4. Per rendere agevole questo cambio di sistema operativo si è così proceduto:

- ogni *server* è stato dotato di un secondo disco: sul disco “1” è installata la configurazione “AMD”, sul disco “2” è installata la configurazione “INFN-Grid”;

- nessuna delle due configurazioni è a conoscenza dell'esistenza dell'altra;
- per cambiare configurazione è sufficiente riavviare la macchina dal disco contenente la configurazione desiderata;
- per poter rendere automatico lo *switch* ciascuno dei sistemi si riferisce ad una sotto-rete diversa (la configurazione “AMD” e quella “INFN-Grid” generano un *pool* di macchine con indirizzi su reti diverse).

In questo modo con un solo comando, impartito via DHCP/PXE, si può trasmettere la configurazione richiesta a tutta la *farm/cluster*, ad una singola macchina o a gruppi di macchine procedendo nel modo seguente:

- viene inviato alle macchine coinvolte nello *switch* un *boot loader* via PXE in grado di selezionare il *boot loader* del disco rigido contenente la configurazione desiderata;
- prima che un *server* sia passato da una configurazione all'altra entra in gioco un meccanismo che interrompe e “riaccoda” eventuali *job* in esecuzione sulle macchine interessate dallo *switch*;
- il sistema è dotato di tutti i *tool* (*shell script* realizzati in loco) per il controllo delle configurazioni in produzione e lo *switch* delle stesse.

5 INFRASTRUTTURA DI RETE ATTUALE

L'infrastruttura di rete attuale si basa su *switch* da rack con *uplink* ad uno *switch* concentratore, tutti i *link* sono da 1Gb/s. Per il collegamento remoto si usa un *link* da 1Gb/s utilizzato da tutta la Sezione e condiviso con altre strutture dell'Università di Pisa; dal luglio scorso è attivo un secondo *link* da 1 Gb/s dedicato al collegamento con GARR del GRID Data Center.

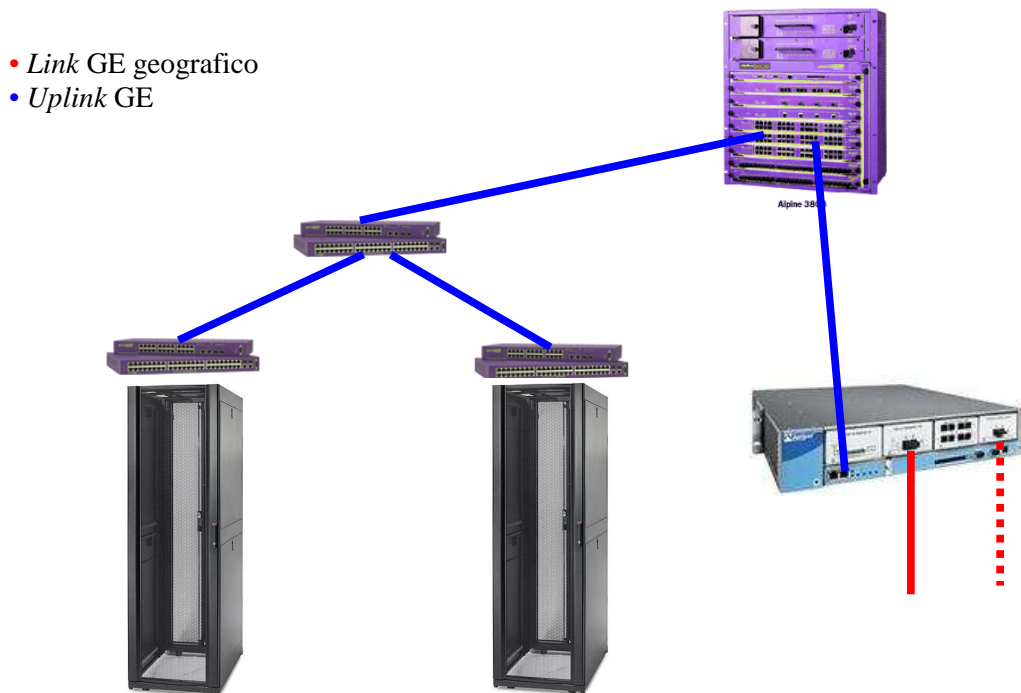


Fig.5 - Attuale configurazione di rete.

6 STORAGE: SITUAZIONE ATTUALE

La situazione attuale dello *storage*, come quella della rete esposta precedentemente, è da considerarsi provvisoria e su questo punto è in fase di attuazione un progetto già definito.

Sono utilizzati da GRID diversi sistemi con architettura sia SAN (*Storage Area Network*) sia NAS (*Network Attached Storage*), ma un orientamento per la soluzione da adottare in produzione è verso sistemi SAN, o meglio FAS (*Fabric Attached Storage*).

Come SRM (sistema di *Storage Resource Manager*) dopo una prima esperienza con DPM, ora utilizziamo dCache.

- Infrastruttura SAN:
 - 2 *switch* FC (CISCO MDS9020) per permettere la ridondanza dei collegamenti
 - Sistemi di *storage*:
 - RIO288 6x4=24TB
 - ER2510 7x4=28TB
 - Storm T2000 52TB
- 10 *server* di disco collegati;
- Sistemi NAS:
 - PANASAS: 5 TB
 - Thumper (Sun 4500): 24 TB
 - EMC2 NS20: massimo 20 TB, in via di installazione

7 ORGANIZZAZIONE GRID E TOOL COLLEGATI

L'organizzazione GRID comprende:

- 3 CE (*Computing Element*, gestione delle code):
 - per motivi storici, fino ad ottobre 2007, il primo annunciava le code SLC3 e l'altro SLC4;
 - ora, finita la transizione a SLC4 (fine ottobre), abbiamo ancora 2 CE (entrambi su Blade IBM) funzionanti in parallelo (ridondanza).
- 2 SE (*Storage Element*, gestione dello *storage*)
 - 1 di tipo classico esistente per motivi storici
 - 1 di tipo dCache formato da 1 "head node" e 4 *disk server*.
- HLR:
 - sistema di *accounting* attualmente ospitato su una macchina virtuale di tipo VMWare;
- vari servizi specifici di CMS (Phedex, squid, pa, user interface).

Due aspetti interessanti sui quali sono stati effettuati sviluppi in loco sono la gestione delle installazioni e delle code.

7.1 Gestione delle installazioni

Il sistema nasce nel 2001 perché a fronte di una infrastruttura UNIX/Linux definita (autorizzazione AFS, *syslog* centralizzato, politiche di sicurezza) mancavano riproducibilità e standardizzazione dell'installazione. Il punto nodale era garantire flessibilità per compensare differenze *hardware* e personalizzazioni specifiche dei gruppi.

All'inizio il sistema si basava su *floppy bootnet* + KS + *script* sviluppato appositamente eseguito al primo *boot*. Dal 2004 c'è stata una profonda rivoluzione dovuta a:

- diffusione capillare di PXE sulle NIC (*Network Interface Card*, schede di rete);
- consolidamento dell'infrastruttura di rete;
- aumento delle "distribuzioni" da gestire.

In questa nuova versione, che è utilizzata anche per le installazioni dei sistemi in GRID, vengono coinvolti i seguenti *server*:

- *Server* DHCP:
 - fornisce la configurazione di rete alla macchina;
 - passa il file da caricare (filename "pxelinux.0");
 - passa l'informazione sul nodo ("next-server swsrv.pi.infn.it").
- *Server* TFTP:
 - fornisce al PXE della NIC il *boot loader* e la sua configurazione grazie alle informazioni ricevute via DHCP.
- *Repository* delle "distribuzioni":
 - macchina con copia delle "distribuzioni" accessibili tramite protocollo anonimo (NFS, FTP, HTTP).

- *Server* DNS:
 - fornisce al sistema le informazioni riguardanti la macchina (tipo di installazione, utente/gruppo “proprietario”): queste informazioni in precedenza erano immesse a mano dall'operatore.

Il sistema attuale ha le seguenti caratteristiche:

- diverse reti IP, ciascuna dedicata a tipologie di macchine diverse:
 - *server* centrali e *farm*
 - *desktop*
 - portatili
 - ospiti
 - sistemi DAQ
- per ciascuna rete c'è un file di configurazione del *boot loader* diverso: diverse condizioni per l'installazione;
- *default = boot* dal disco;
- per tutti sono disponibili *utility* generiche come *memtest* e *rescue system*;
- *file* specifici per nodi particolari, ad esempio macchine Grid;
- possibilità di portare (*soft link*) i *file* di configurazioni fuori dal TFTP *server* (AFS) per dare autonomia di installazione ai gruppi.

7.2 LSF: gestione code

Il *tool* principale di gestione delle code è LSF. Sono utilizzati anche altri *tool* per gruppi che ne comprovino la necessità. Caratteristiche della gestione delle code:

- un nodo *master* di LSF visto come “Servizio di Sezione” (LSF viene usato infatti anche per *cluster* e *farm* non GRID);
- ad oggi:
 - una coda per ciascuna VO supportata;
 - peso relativo delle code determinato in base al *fairshare*, controllabile mediante “*lsfmon*” (vedi seguito);
 - coda locale (principalmente per *test* ed utenti “senza VO”).
- nonostante ne esistesse già uno in ambiente INFN, internamente e in collaborazione con CMS-Pisa è stato sviluppato un *tool* di *monitoring* dell'attività GRID gestita mediante LSF (<http://farmsmon.pi.infn.it/lsfmon>):
 - il *tool* è trasferibile agli altri siti interessati.

8 “LAVORI IN CORSO” E FUTURO

8.1 Infrastrutture e progetto “S+R”

Come accennato in precedenza il dimensionamento della corrente elettrica in sala è in linea con le potenzialità dei sistemi di condizionamento. Lo schema di distribuzione in sala è stato oggetto di una profonda ristrutturazione completata nel giugno scorso. Il prossimo passo riguarderà la messa in esercizio dei voltmetri ed degli amperometri delle varie linee, in

grado di essere letti via *ethernet* dal sistema di monitoraggio (progetto “S+R”).

Di seguito è riportato lo schema della distribuzione elettrica nella sua configurazione finale. Questa attività si è recentemente conclusa con il raggiungimento dell’obiettivo di fondo: la disponibilità di alimentazione sia UPS sia non UPS facilmente raggiungibile in qualsiasi *rack* presente.

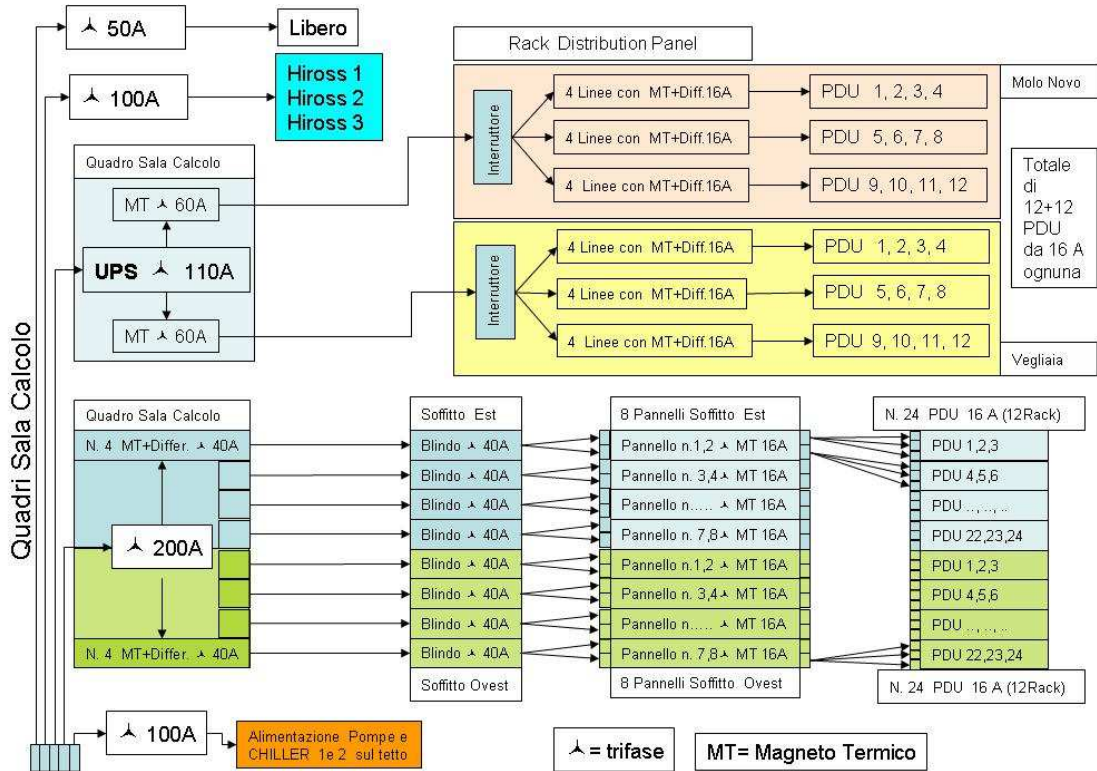


Fig. 6 - Schema della distribuzione elettrica.

Il progetto “S+R” riguarda gli obiettivi di *Security* e *Robustness* per le infrastrutture nell’ambito della qualità di servizio necessaria in fase di produzione. È da considerarsi propedeutico e sottostante al sistema di “GRID *monitoring* e *management*”, che già può contare per quel che riguarda il *monitoring* su tool quali GridIce, Ganglia e il già citato “LSF monitor” mentre per la parte di *management* si sta valutando “LeMon”.

Lo scopo del progetto “S+R” è la realizzazione di un sistema di controllo e gestione degli impianti infrastrutturali asserviti al GRID Data Center.

Gli impianti in oggetto sono quello della distribuzione elettrica e quello di condizionamento.

Il sistema sarà composto da 3 elementi integrati tra loro:

- monitoraggio: sia locale sia remoto;
- generazione e gestione di allarmi: sia in modo automatico sia semi-automatico, in locale e in remoto;
- analisi comparative di serie storiche di parametri caratteristici degli impianti in

oggetto.

Il sistema da realizzare si baserà sull'integrazione di analoghi sistemi già funzionanti, che coprono sezioni parziali degli impianti, forniti insieme alle apparecchiature e sullo sviluppo di una piattaforma *software* comune. Questa attività di sviluppo ed integrazione è già in corso. Tra gli elementi già parzialmente sviluppati c'è un "sinottico" della sala che, una volta completato, permetterà di controllare in tempo reale:

- parametri di funzionamento dei sistemi di condizionamento (condizionatori e *chiller*);
- stato delle temperature, sia in sala sia su singoli *rack* (temperatura dell'aria in ingresso ai *server*);
- assorbimento di corrente ai quadri;
- stato di alimentazione e consumo delle singole PDU (*Power Distribution Unit*, le "ciabatte" che distribuiscono l'alimentazione all'interno dei *rack*);
- stato degli UPS.

Il "Sinottico" è stato impostato in modo totalmente parametrico, così da poter risultare facilmente "portabile" e quindi applicabile anche ad altre sale calcolo (anche non a Pisa).

<http://farmsmon.pi.infn.it/sinottico/>

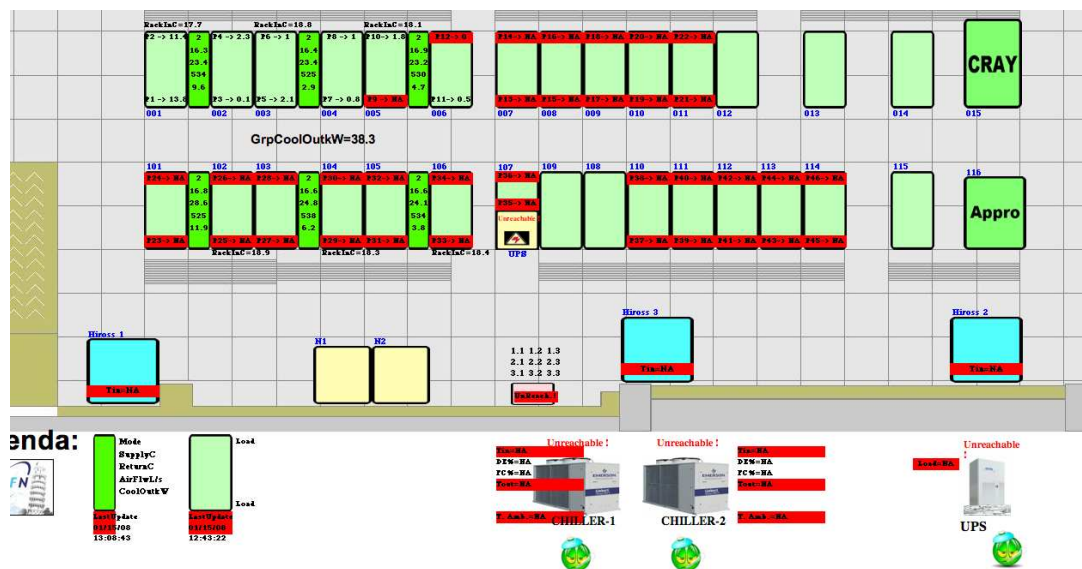


Fig. 7 - Esempio della schermata attuale del "sinottico".

Del progetto "S+R" fa parte anche il sistema, già in funzione, di video-sorveglianza dedicato alla sala e basato su telecamere ad infrarossi per un monitoraggio anche in notturna.



Fig. 8 – Particolare del sistema di video-sorveglianza.

8.2 Progetto “N+S”

Come si è già avuto modo di dire, la situazione della rete LAN e dello *storage* è carente rispetto alle necessità della fase di produzione. Per questo motivo nel corso della seconda metà del 2007 è stato elaborato un progetto specifico, denominato “N+S” (*Network + Storage*).

Due linee guida (“drivers”) sono state identificate per la definizione dell’architettura più adatta (riportiamo i termini inglesi originali che compaiono nel progetto):

- *manageability & robustness*: qui la necessità è di limitare al massimo l’intervento umano, necessità imposta dalle esigue risorse che sono e saranno disponibili;
- *self-balancing*: come confermato dall’esperimento, non è possibile elaborare un modello dell’uso del Data Center in termini di distribuzione dei *job* sulle *farm* e dell’allocazione dei *file* sullo *storage*. Quindi non abbiamo la possibilità di studiare ed implementare una efficiente strategia di ottimizzazione per il bilanciamento del traffico di I/O tra i *worker node* (wn) e lo *storage*.

Un punto sul quale è stata fatta una profonda riflessione è lo *switch* tra i *disc server* e lo *storage* (“Big FC or IB Switch”) di seguito indicato come lo “Storage Switch”. Sono stati analizzati due diversi scenari:

- Lo *storage switch* è incluso all'interno dell'architettura dello *storage* (esempi di tale soluzione sono: EMC² Symmetrix, Data Direct in configurazione HPC e, in parte, NetApp FAS6000).
- Lo *storage switch* è un oggetto fisico esterno che deve essere opportunamente dimensionato e configurato. In questo caso l'architettura dello *storage* può essere una normale SAN, non richiedendo al singolo sistema la possibilità di “scalare” fino all'ordine di IPB. Si possono usare, in questa architettura, più sistemi gemelli che vengono gestiti come un sistema logicamente unico a livello di *disc server*. Il numero di tali sistemi sarà determinato non dallo spazio disco ma dall'aggregato di I/O richiesto.

La prima soluzione è più costosa della seconda. Il maggiore costo è dovuto alle prestazioni generali di I/O che il sistema di *storage* garantisce in sé. Nel secondo caso viene garantito solo un I/O aggregato teorico, il compito di ottenerlo in produzione è responsabilità nostra ed è basato sulle possibilità di bilanciamento di traffico sui vari canali di I/O che collegano ai *server* le diverse porzioni del sistema di *storage* (i sistemi gemelli di cui sopra).

Per motivi principalmente economici i responsabili del calcolo dell'esperimento CMS hanno optato per la seconda soluzione, in via di implementazione mediante una serie di unità di *storage* montate in batteria (senza una architettura SAN unica).

8.3 Simulazione termofluidodinamica

Un importante impegno è stato e sarà messo nelle attività di modellizzazione e simulazione termofluidodinamica della sala, attività che si sta rivelando in forte espansione a livello internazionale per le possibilità di ottimizzazione che offre in fase di progettazione di Data Center (presentazione RAL Hepix-fall 2007).

Nel corso del 2007 è stato effettuato uno studio di analisi termica, condotto in collaborazione con il *Servizio Alte Tecnologie della Sezione INFN di Pisa*, con lo specifico scopo di confrontare le prestazioni del condizionamento di sala mediante espansione diretta e il condizionamento “a isola” con liquido refrigerato (soluzione APC InRow). Di seguito sono riportate una piantina con i punti di misura per una delle campagne condotte ed una termografia di esempio di una porzione frontale di un *rack* di *server*.

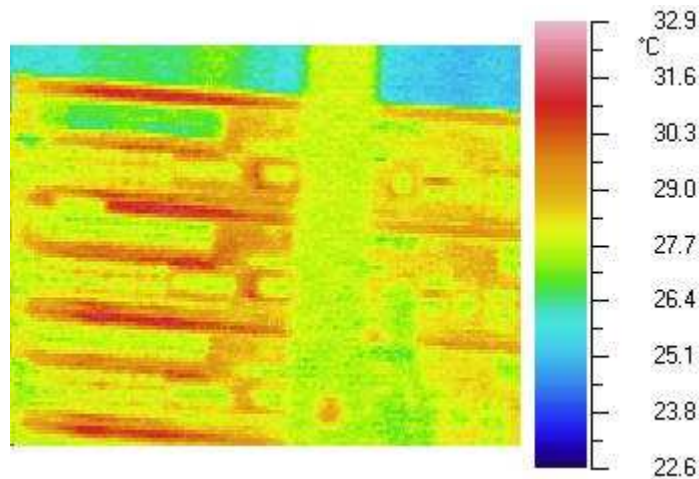
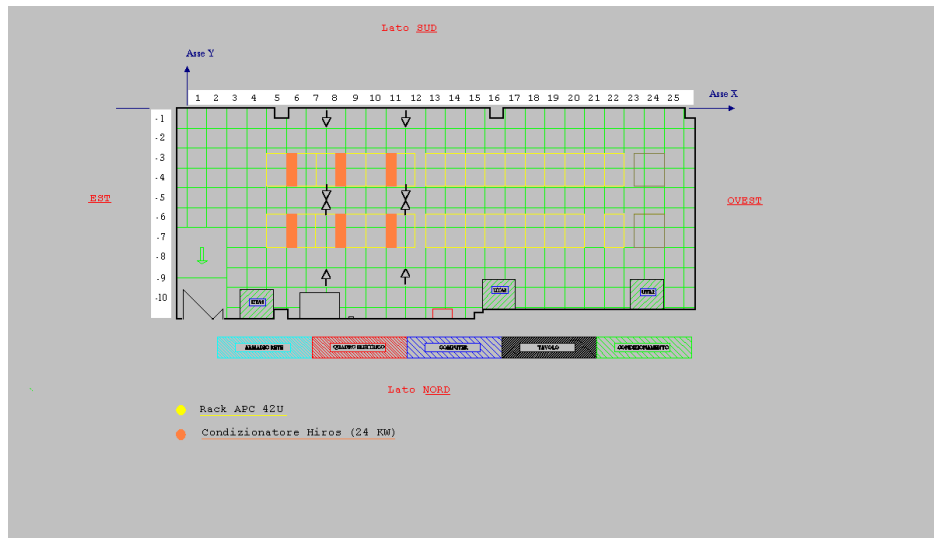


Fig. 9 – Studio di analisi termica della sala calcolo.

È in corso un'altra attività, volta questa alla simulazione termofluidodinamica della sala (usando il codice Fluent), condotta in collaborazione con il *Dipartimento di Ingegneria Aerospaziale dell'Università di Pisa*. In questo caso l'interesse è rivolto alla possibilità di fare simulazioni finalizzate all'ottimizzazione dell'espansione delle macchine presenti in sala. In particolare l'architettura APC InRow prevede la possibilità di "confinamento" parziale o totale dei flussi di aria fredda e/o calda utilizzando dei "tappi" da mettere sul fronte e/o sul retro dei *rack* per canalizzare il flusso d'aria verso e/o da i condizionatori.

Sono già stati prodotti il modello CAD della sala con le espansioni previste e con le varie configurazioni di "confinamento" d'aria (nella figura seguente è mostrato come esempio il CAD della sala con entrambi i flussi caldo e freddo confinati) e i modelli matematici (semplificati e linearizzati) del comportamento dei diversi tipi di condizionatore.

Questa attività ha prodotto una prima serie di simulazioni (mediante una tesi specialistica in Ingegneria Aerospaziale discussa lo scorso luglio) sulle quali si stanno studiando i necessari affinamenti.

Fig. 10 – Simulazione termofluidodinamica della sala calcolo: modello CAD.

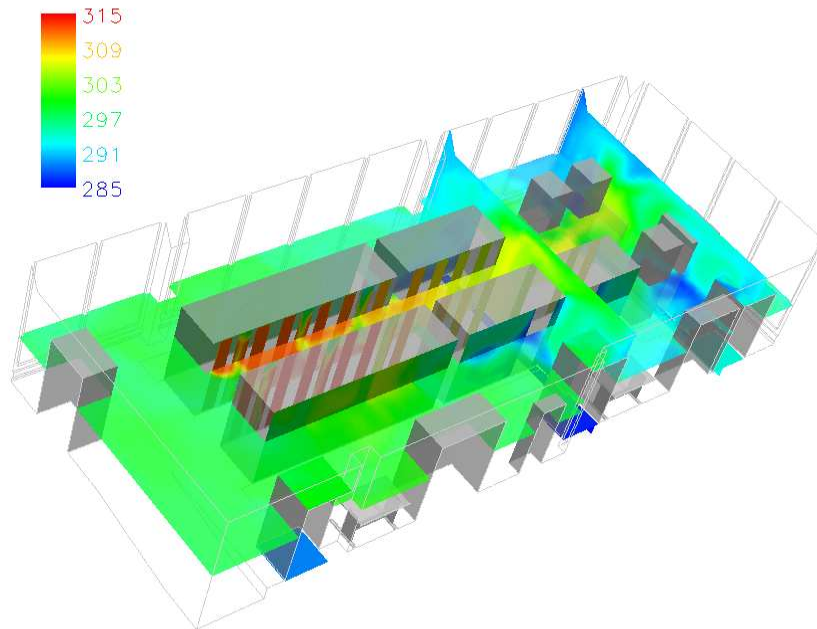


Fig. 11 - Simulazione termofluidodinamica della sala calcolo: esempio di campo delle temperature.

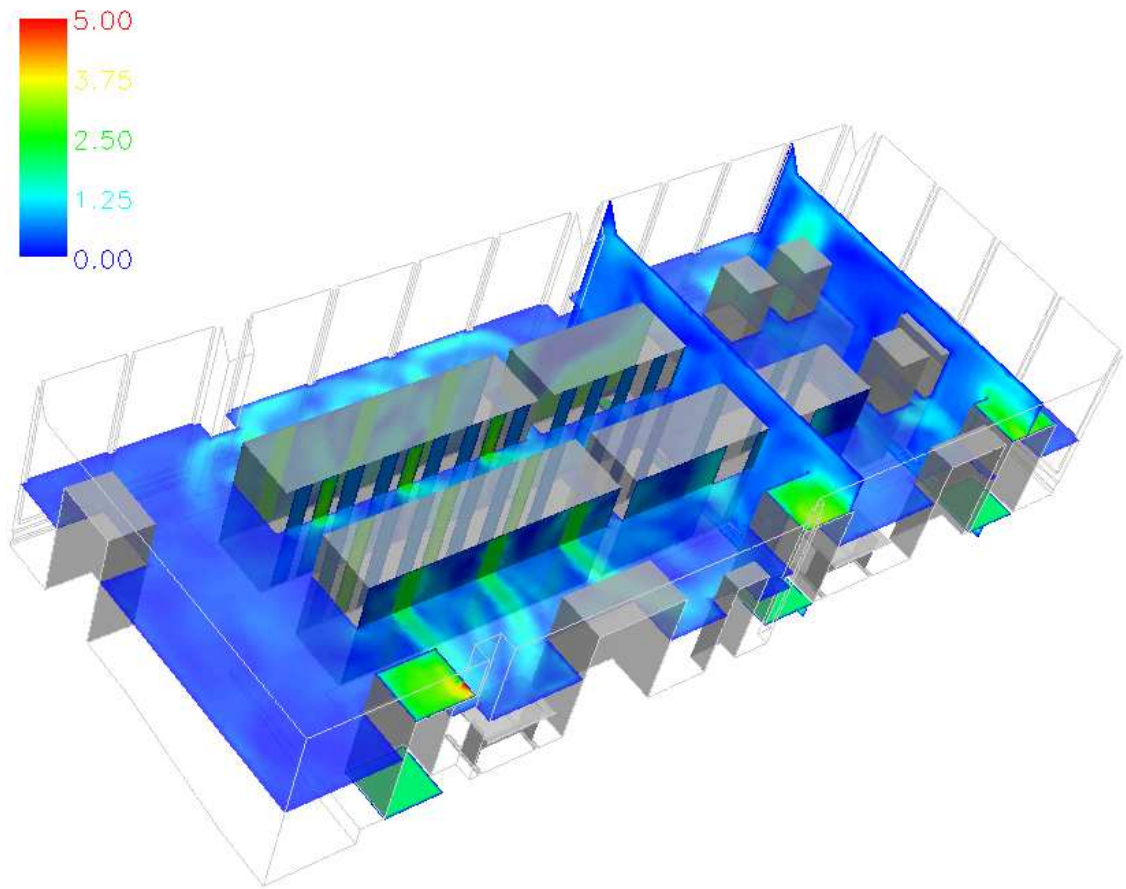


Fig. 12 - Simulazione termofluidodinamica della sala calcolo: esempio delle velocità dei flussi.

Chi fosse interessato ad ulteriori dettagli sulle attività esposte può rivolgersi al Servizio Calcolo e Reti della Sezione di Pisa.