ISTITUTO NAZIONALE DI FISICA NUCLEARE

INFN/CCR-10/04
7 Settembre 2010

CCR-37/2010/P

PROPOSTA INFN PER LA RETE DEI TIER2 DI LHC IN GARR-X

A. Brunengo[1], A. De Salvo[2], D. Di Bari[3], G. Donvito[3], R. Gomezel[4], P. Lo Re[5], G. Maron[6], E. Mazzoni[7], A. Spanu[2], S. Zani.[8]

[1)]INFN-Sezione di Genova, Via Dodecaneso 33 - 16146 Genova
[2)]INFN-Sezione di Roma, P.le Aldo Moro 2 - 00185 Roma
[3)]INFN-Sezione di Bari, Via E. Orabona 4 - 70126 Bari
[4)]INFN-Sezione di Trieste, Via A. Valerio 2 - 34127 Trieste
[5)]INFN-Sezione di Napoli, Complesso Univ. di Monte Sant'Angelo, Via Cintia - 80126 Napoli
[6)]INFN-Laboratori Nazionali di Legnaro, Viale dell'Università 2 - 35020 Legnaro (Padova)
[7)]INFN-Sezione di Pisa, Edificio C, Polo Fibonacci, Largo Bruno Pontecorvo 3 - 56127 Pisa
[8)]INFN-CNAF, Viale Berti Pichat 6/2 - 40127 Bologna

Abstract

La presente proposta è frutto di incontri con i rappresentanti di tutti i Tier di LHC italiani e quindi rappresenta una visione condivisa su come l'INFN intende connettere i propri centri di calcolo di secondo livello al Tier1 nazionale e agli analoghi centri di altre nazioni, attraverso la costituenda rete GARR-X.

A questi incontri di indirizzo sono seguite numerose discussioni tecniche promosse dal gruppo NetArch della Commissione Calcolo e Reti di INFN con lo scopo di delineare i requisiti principali dei collegamenti dei Tier2 alla rete ottica di GARR-X.

L'interazione con i colleghi del GARR che si occupano della nuova rete è stata frequente e il documento "Progetto di Rete GARR-X: La rete dei Tier2 dell'INFN in GARR-X" di M. Marletta, M. Carboni e C. Battista del GARR, che sostanzialmente propone gli scenari tecnici possibili per la rete dei Tier2, rappresenta il riferimento tecnico principale della presente proposta.

*Published by **SIS–Pubblicazioni***
Laboratori Nazionali di Frascati

1 REQUISITI DI BANDA E DI CONNETTIVITA'

La tabella n. 1 mostra una versione aggiornata delle previsioni di banda di trasmissione dei centri Tier1 (CNAF) e Tier 2 italiani[1]. La tabella riporta sia il rate di trasferimento massimo continuativo previsto che quello stimato per esigenze straordinarie e temporanee. I valori riportati derivano dal documento della CCR/14/07/p "Evoluzione delle esigenze di rete geografica dell'INFN negli anni 2008-2011 e prospettive offerte dal progetto GARR-X" aggiornato con le stime più recenti fornite dagli esperimenti.

Banda Garantita di accesso (Gbit/s) **Range di utilizzo tra caso massimo continuativo e caso straordinario**								
	2010		2011		2012		2013	
	Continuativo	Straordinario	Continuativo	Straordinario	Continuativo	Straordinario	Continuativo	Straordinario
CNAF	10,2	11,9	15,5	17,3	20,7	23,2	22,2	24,6
Catania	1,5	1,9	1,8	2,6	5,3	6,6	5,6	6,7
LNL	1,9	2,3	2,3	3,6	2,8	4,5	4,2	6,8
Milano	0,7	1,4	1,0	2,3	1,4	3,7	1,4	3,6
Napoli	0,7	1,4	1,0	2,3	1,3	3,6	1,4	3,6
Pisa	0,8	1,6	1,3	2,7	1,5	3,3	2,5	5,1
Roma1	1,4	2,2	2,3	3,7	2,8	5,1	3,7	6,3
Torino	0,8	1,2	1,0	1,9	1,4	2,9	1,7	2,9
Bari	0,9	1,3	1,4	2,3	1,8	3,2	2,3	3,5
Frascati	0,6	1,2	0,7	1,4	0,9	2,3	1,0	2,2

Tabella n. 1

Dalla tabella si vede che, anche con le stime più aggiornate, un singolo collegamento a 10 Gbit/s per Tier2 è sufficiente mentre per il CNAF sarà necessario passare dal 2011 ad un link a 40-100 Gbit/s.

Per quanto riguarda la topologia logica delle connessioni, riassumiamo qui di seguito le esigenze dei centri Tier2.

Con l'eccezione dell'esperimento Atlas che ha un modello sostanzialmente gerarchico tra i Tier2 e il Tier1 nazionale, gli altri esperimenti presentano modelli che prevedono movimento dati non solo tra i Tier2 e il Tier1 nazionale, ma anche con i Tier1 dell'esperimento presenti negli altri Paesi. Il modello gerarchico di Atlas è però invalidato dalle procedure di calibrazione dell'apparato che coinvolgono il trasferimento diretto di dati dal T0 ai Tier2 di Roma e Napoli. Inoltre recentemente sia Alice che CMS hanno esteso il proprio modello prevedendo movimentazione di dati tra gruppi di Tier2 (solitamente legati ai gruppi di fisica), completando in questo modo la magliatura logica tra i propri centri di analisi del secondo livello.

[1] Sono inclusi nella lista, oltre ai centri Tier2 approvati dall'INFN, anche quelli di Bari e Frascati che contribuiscono a fornire risorse di calcolo aggiuntive agli esperimenti LHC.

Questa evoluzione dei modelli è riflessa nelle previsioni di banda verso i siti internazionali che, come mostra la tabella n. 2, sono dello stesso ordine di grandezza del traffico tra i siti INFN e il centro nazionale Tier1 del CNAF.

Banda Garantita di accesso (Gbit/s) ai siti internazionali Caso Straordinario					
Tipo Collegamento	Siti collegati	**2010**	**2011**	**2012**	**2013**
Tra siti INFN e Siti Internazionali	CNAF-Siti Internazionali	8,4	11,5	17,1	16,7
	Altri Siti INFN - Siti Internazionali	5,5	7,7	8,1	12,1
Tra siti INFN e CNAF	CNAF- Siti INFN	6,3	10,1	13,2	16,0

Tabella n. 2

Va qui sottolineato che i valori mostrati nelle due tabelle sono delle previsioni fornite dagli esperimenti che rappresenteranno gli user principali di questa nuova infrastruttura. Va quindi colto di questi valori l'ordine di grandezza e la linea di tendenza più che il loro valore assoluto che necessariamente è affetto da errori di stima e soggetto a variazioni man mano che il modello di calcolo si preciserà con la movimentazione di dati reali tra i vari Tier della struttura di calcolo di LHC.

Una topologia che vede il Tier1 nazionale al centro della stella formata da tutti i Tier2 italiani, sembra, da queste previsioni e dai requisiti degli esperimenti, limitata e non più attuale. Essa inoltre rappresenta un inutile "single point of failure" per tutti i Tier2 che in caso di problema grave al Tier1 nazionale non potrebbero in alcun modo fare analisi su dati di altri Tier1.

1.1 Gli scenari possibili

L'evoluzione recente dei modelli di calcolo e di distribuzione dei dati degli esperimenti porta quindi a scartare una soluzione di rete ottica a stella di tipo L2 che, ad un primo approccio, sembrava naturale. L'interconnessione dei siti Tier2 potrebbe infatti essere supportata da GARR-X tramite la fornitura diretta di circuiti lambda ai Tier2 (vedasi il par. 4 "*Interconnessione via lambda*" del citato documento GARR), configurando in questo modo una rete privata a stella tra i Tier2 italiani e il proprio Tier1 di riferimento. Questa soluzione che mappa direttamente il modello logico gerarchico in quello fisico della struttura ottica, ha l'innegabile vantaggio delle prestazioni e della bassa latenza, ma da una parte presenta problemi di affidabilità e dall'altra non risponde in modo adeguato ai requisiti citati qui sopra. Serve un approccio più sofisticato.

Crediamo infatti che un accesso di tipo IP in tecnologia ethernet tra i Tier2 e i Tier1 rappresenti in questa fase non solo la soluzione più flessibile, ma anche globalmente la più efficiente permettendo una connettività totale tra i nodi della rete internazionale formata dai

centri di analisi di secondo livello e centri Tier1.

A questa rete logica deve però corrispondere nella parte italiana una rete fisica molto efficiente in grado di sfruttare tutte le potenzialità della nuova rete ottica di GARR-X. E' auspicabile che vi siano delle lambda dedicate al traffico LHC tra i POP relativi ai Tier2 ed il POP di Bologna, ma è comunque fondamentale che, se necessario, il Garr sia in grado di attivarle rapidamente.

Va notato come questo approccio comunque fornisca ai Tier2 nazionali una infrastruttura ottica verso il Tier1 del CNAF permettendo in futuro una facile estensibilità in banda (lambda dedicata, più lambda se necessario), ma anche l'utilizzo di eventuali tecnologie di connessione di livello 2 a bassissima latenza che si potranno affermare nei prossimi anni.

1.2 La proposta INFN

La nostra propensione all'accesso IP in tecnologia ethernet trova conferma di realizzabilità tecnica nel documento del GARR già citato in precedenza (si veda il par. 3 "*Interconnesione via Switching Ethernet"* del documento). In particolare la possibilità di configurare sul link a 10 Gbps del Tier2 delle VLAN ethernet soddisfa i requisiti menzionati più sopra in quanto sarà possibile configurarle in modo di avere i due accessi fondamentali più il terzo di backup che servono al centro Tier2:

- VLAN per collegamento dei Tier2 al Tier1 del CNAF;
- VLAN per accesso IP general purpose primario in modo da garantire l'accessibilità da e verso i Tier1 e Tier2 internazionali;
- VLAN per accesso IP general purpose di backup.

Per quanto riguarda i requisiti della rete ottica che collega un Tier2 al Tier1, essi sono già stati menzionati nel paragrafo precedente e si riducono alla necessità di avere una lambda garantita a 10 Gbps tra il pop di riferimento del Tier2 e il Tier1. In caso di necessità deve essere possibile applicare un meccanismo di "shaping" del traffico in modo da evitare che il traffico su una delle VLAN saturi il link fisico a scapito delle altre.

Il traffico "general porpouse" della sezione/laboratorio che ospita il Tier2 deve poter circolare su link separato da quello a 10 Gbps del Tier2. La sezione/laboratorio deve quindi essere connessa al pop-x con un ulteriore link fisico da almeno 1 Gbps. Il routing deve poi essere fatto in modo tale da garantire che il traffico verso i Tier1 e Tier2 internazionali passi per il link a 10 Gbps senza interferire con quello a 1 Gbps.

La banda complessiva prevista per il traffico internazionale è riportata in tabella n. 2. Sommando le esigenze del CNAF e di tutti gli altri siti INFN, essa varia da 15 Gbps per il 2010 al doppio per il 2013.

Qualora la sezione/laboratorio fosse connessa al pop-x attraverso un percorso metropolitano in "dark fiber", potrebbero essere previsti due link a 10 Gbps, con percorsi fisici separati, in modo da garantire un "path" di backup in caso di guasti. Attualmente i Tier2 che abbisognano di questa configurazione sono Legnaro e Pisa.

La proposta di INFN per il collegamento dei propri Tier2 al Tier1 del CNAF va quindi in questa direzione che, tra l'altro, permette ai centri di secondo livello di dotarsi di apparati di routing relativamente poco costosi in quanto non abbisognano dei protocolli più sofisticati come ad esempio il MPLS.

1.3 Connettività del Tier1

Allo stato attuale il Tier1 è collegato con 2 interfacce 10 Gb/s al POP del GARR (ospitato all'interno dei locali del Tier1 stesso) utilizzate principalmente per raggiungere il CERN, i Tier1 di LHC ed i Tier2.

Entro la fine del 2010 il CNAF dovrebbe avere almeno 3 connessioni 10Gb/s: una dedicata al traffico T0-Tier1 (già in opera), una per il traffico Tier1-Tier1 (parzialmente in opera) ed almeno una per il traffico verso i Tier2 che al momento transita sull'accesso general purpose del CNAF.

Nello specifico, per quanto riguarda la connettività del Tier1 verso i Tier2 italiani, si richiede vi sia una quantità di connessioni opportuna ad aggregare il traffico destinato ai Tier2 (il dimensionamento preciso avverrà in base all'utilizzo reale di tali link fino ad un massimo teorico di 10 Gb/s per ogni Tier2).

Si prevede inoltre di iniziare entro la fine del 2010 o l'inizio del 2011 la sperimentazione su scala geografica di interfacce a 100 Gb/s con un primo link di questo tipo fra CNAF e CERN.