



ISTITUTO NAZIONALE DI FISICA NUCLEARE

Progetto di Massima per un Centro Regionale di Calcolo per l'INFN

Autori

**Paolo Capiluppi¹, Domenico Galli¹, Alberto Masoni², Mauro Morandin³, Laura Perini⁴,
Fulvio Ricci⁵, Federico Ruggieri⁶**

¹Sezione INFN di Bologna

²Sezione INFN di Cagliari

³Sezione INFN di Padova

⁴Sezione INFN di Milano

⁵Sezione INFN di Roma

⁶INFN – CNAF, Bologna

Versione 1 del 10 Aprile 2001

Indice

1	Introduzione.....	3
2	Le necessità degli Esperimenti LHC.....	3
2.1	ALICE.....	5
2.1.1	CENTRO REGIONALE ITALIANO PER L' ESPERIMENTO ALICE	5
2.2	ATLAS.....	6
	CMS.....	9
	Tier1 di CMS in comune con altri Esperimenti INFN.....	9
	Premessa.....	9
	Tier1 CMS–INFN.....	9
	Funzioni.....	9
	Personale.....	11
	Ruolo verso i Tier2(3) CMS–INFN.....	11
	Ruolo verso la Collaborazione CMS.....	12
	Sviluppo del Tier1 CMS–INFN nel Modello di calcolo CMS Italia	12
4.1	LHCb.....	12
4.1.1	Il modello di calcolo di LHCb.....	13
4.1.2	Risorse di calcolo necessarie.....	13
4.1.3	Distribuzione del calcolo di LHCb–Italia tra i Tier–n.....	14
5	Le necessità di altri Esperimenti.....	15
5.1	BaBar.....	15
5.1.1	Il centro Tier A per BaBar.....	15
5.1.2	Il Tier A: requisiti.....	16
5.1.3	Tempi di realizzazione.....	18
	d)Il personale necessario.....	18
5.2	VIRGO.....	18
5.2.1	Il flusso dei dati acquisiti in Virgo.....	19
5.2.2	Attività di analisi off–line.....	19
5.2.3	Il Tier–1 di Virgo–Italia e le necessità complessive.....	21
5.2.4	Risorse Umane.....	22
6	Considerazioni Generali.....	22
6.1	Gestione del Centro.....	22
6.2	Spazi ed Infrastrutture al CNAF.....	23
6.3	Personale del CNAF.....	23
7	Fase I – Tier A per BaBar, Tier 1 per VIRGO e Sperimentazione del Tier 1 unico per LHC.	24
7.1	Risorse Hardware.....	24
7.2	Rete.....	25
7.3	Personale.....	26
7.4	Definizione del piano.....	27
7.5	Scala dei Tempi.....	28
8	Fase II di Realizzazione Tier 1 unico per LHC.....	29
9	Costi.....	30
10	Conclusioni.....	30
11	Referenze.....	31

1 Introduzione

Il progetto fornisce una descrizione di massima dell'implementazione (installazione e funzionamento) di un centro di calcolo regionale di tipo Tier 1^[1], Tier-A .

Già da tempo i futuri esperimenti, in particolare quelli a LHC, hanno formulato previsioni sull'entità del calcolo e dello storage richiesti negli anni a partire dal 2005 in poi ed hanno in corso grosse attività di simulazione^{MONARC Phase 2 final report: <http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf>} e sperimentazione di tecnologie^{DATAGrid Project: <http://www.cern.ch/Grid/>}. Nel corso del LHC Computing Review^[3], recentemente conclusasi al CERN, è stato posto in evidenza come il sistema di Calcolo per gli esperimenti ad LHC non possa, per motivi pratici, economici e politici, essere totalmente installato al CERN.

La grande comunità della Fisica delle Alte Energie, o High Energy Physics (HEP), partecipa a pochi grandi esperimenti internazionali con un insieme di diverse migliaia di ricercatori distribuiti in centinaia di Istituzioni ed Università in decine di Paesi ed in vari Continenti.

Non è ragionevole pensare che le risorse di calcolo, necessarie a questa grande comunità di utenti sofisticati, siano concentrate in un unico laboratorio. Già da tempo al CERN viene adottata una politica di suddivisione del calcolo che prevede che, per ogni esperimento, circa un terzo venga fatto presso il laboratorio e due terzi vengano forniti dalle sedi di origine dei ricercatori.

Le necessità che si prospettano nell'INFN per esperimenti futuri, come quelli ad LHC, prossimi, come VIRGO, o addirittura presenti, come BaBar, suggeriscono, come vedremo, una impressionante "escalation" delle necessità di archiviazione dati (dalle centinaia di TB a diversi PB) e, conseguentemente, delle potenze di calcolo necessarie all'analisi della enorme quantità di dati raccolta e da simulare.

Tali enormi quantità di dati, che costituiscono già una sfida per l'era LHC, comportano un problema di archiviazione e gestione che, con le attuali tecnologie e le attuali previsioni dei costi, rendono la problematica "critica" e hanno portato gli esperimenti a strutturare il loro modello di calcolo prevedendo grandi Centri Regionali. Questo progetto assume che in Italia la sede del Centro Regionale dove installare grandi risorse di archiviazione e processamento dati, sia comune per il Tier-1 degli esperimenti LHC e di Virgo e per il Tier-A di Babar.

Nello stesso tempo, queste necessità hanno spinto la comunità ad indirizzarsi verso la realizzazione di un software, di tipo GRID, che consenta di accedere a tutte queste risorse di dati e calcolo grandi e piccole in tutto il mondo in maniera facile, uniforme e trasparente all'utente.

Questo documento è il risultato delle valutazioni fatte dagli esperimenti, della loro analisi comparata finalizzata alla realizzazione di un servizio comune. E' stata effettuata un'analisi preliminare delle sinergie e dei vincoli che questo comporta tenendo presente l'evoluzione delle tecnologie sul calcolo distribuito (GRID) attualmente in atto.

2 Le necessità degli Esperimenti LHC

Tra le tante review ufficiali delle necessità degli esperimenti ad LHC la più recente è quella fornita dalla LHC Computing Review del CERN ^{Report of the Steering Group of the LHC Computing Review: CERN/LHCC/2001-004, CERN/RRB-D 2001-3, 22 February 2001.}

La review ha consentito di stabilire il corrente stato di sviluppo dei progetti di "Computing" degli esperimenti, del CERN (IT) e delle Istituzioni coinvolte in tutto il Mondo, mettendo in risalto le possibili prospettive comuni, gli investimenti ed i punti deboli che richiedono maggior attenzione ed immediate azioni.

I modelli dei quattro esperimenti si basano sulla comune esperienza del Modello sviluppato da MONARC^{MONARC Phase 2 final report:}

<http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf> opportunamente integrato dalle funzionalità offerte dai progetti di sviluppo di GRID. I Modelli si basano pertanto su una gerarchia di Centri Regionali (RC) con funzionalità diverse e interoperanti tra loro (Tier0, Tier1, Tier2, Tier3 e Tier4). I Tier0 dei quattro esperimenti saranno implementati in un unico sistema presso la IT del CERN, opportunamente partizionato e/o condiviso. Ogni esperimento

disporrà dei propri Tier1 (eventualmente condivisi con altri) e delle gerarchie inferiori di Tier (2, 3 e 4). Per tutti gli esperimenti, uno di questi Tier1 è previsto risiedere al CERN e pertanto verranno integrati nel sistema dei Tier0.

Le principali funzionalità del Tier0 sono lo storage dei Raw data, la ricostruzione degli eventi ed il riferimento per il mantenimento/supporto del software comune.

Ai Tier1 di ogni esperimento è demandata la funzionalità del re-processing, della selezione ricorsiva degli eventi, della produzione degli eventi simulati, di parte dell'attività di analisi e del supporto/sviluppo del software, anche specifico di esperimento). I Tier1 sono quindi nazionali o sopra-nazionali.

Ai Tier2 è demandata la funzionalità, insieme al Tier1 al quale fanno riferimento, della analisi e parzialmente della simulazione. Per questo motivo i Tier2 sono anch'essi la sede più opportuna per lo sviluppo del software specifico di esperimento. I Tier2 sono nazionali o infra-nazionali.

I Tier3 e Tier4 sono considerati risorse di utilizzo locale dei singoli gruppi di ricerca e degli individui, quindi, pur facendo parte delle funzionalità, sono per ora esclusi dalle valutazioni delle risorse del modello. I Tier3 e Tier4 sono pertanto presso le Istituzioni.

Va rilevato che il Modello è funzionale nella distribuzione dei compiti e dell'accesso ai dati soltanto (e tanto più) attraverso lo sviluppo di evoluti tools di Grid, possibilmente sviluppati in comune e con alta affidabilità.

La review del CERN ha stimato che circa 1/3 delle risorse per il Computing di LHC sia concentrato nel Tier0/1 comune al CERN e che quindi i rimanenti 2/3 siano dislocati per circa 1/3 presso i Tier1 e circa 1/3 presso i Tier2/3/4.

L'investimento totale per tutto il Computing di LHC è stato stimato (con l'incertezza della proiezione dei costi su tempi così lunghi per il mercato dei calcolatori) essere intorno ai 240 MCHF (senza considerare la maggior parte del costo del personale e le connessioni di rete tra i Centri Regionali di ogni livello, eccettuate le attuali connessioni col CERN).

Tale investimento, principalmente di infrastrutture di Computing, è previsto in un piano temporale compreso tra il 2005 ed il 2007, con un profilo di spesa pressochè piatto.

Va sottolineato che poichè i costi dei componenti di Calcolo sono previsti calare nel tempo, il profilo piatto comporta che durante ognuno dei tre anni si acquisiscono risorse superiori fino a raggiungere nel 2007 la piena funzionalità.

Dal 2008 in avanti è previsto di sostituire ogni anno circa un terzo delle capacità con investimento costante. Questo consente il necessario upgrade delle risorse rese obsolete e l'atteso necessario aumento delle risorse di calcolo per far fronte alla maggior mole di dati ed analisi.

Un'altro rilevante risultato della review riguarda la necessità di sperimentare la complessità dei modelli di Computing attraverso prototipi di dimensioni crescenti. Tali prototipi devono essere in grado di misurare la complessità e contemporaneamente permettere agli esperimenti di verificare il proprio software e le potenzialità di analisi fisica dei detectors attraverso quelli che vengono chiamati Data Challenges (tutti gli esperimenti, seppur con tempistiche e scopi lievemente diversi, li prevedono).

Le risorse così realizzate costituiscono una comune risorsa (opportunamente partizionata nel tempo ed eventualmente nelle dimensioni) per tutti gli esperimenti.

Alla fase di prototipizzazione devono partecipare il CERN, alcuni dei maggiori Tier1 e alcuni Tier2, per permettere di verificare la complessità e le funzionalità di tutte le componenti del modello.

A tal fine la review ha ritenuto che sia necessario realizzare al CERN un prototipo dell'ordine del 50% del sistema finale di uno degli esperimenti e che i Tier1 e Tier2 partecipanti debbano avere almeno le dimensioni del 10% di quelli finali per ogni esperimento che su di essi insisterà.

La possibilità di raggiungere, in tempo utile per le scelte sul sistema finale, una scala del sistema appropriata per una valutazione è stata ampiamente discussa all'interno dello LHC Computing Review e considerate un punto critico nello sviluppo del calcolo per LHC. L'esperienza passata, confermata dai Data Challenges già effettuati per LHC, ha infatti ampiamente dimostrato che estrapolazioni di un ordine di grandezza sulla scalabilità dei sistemi di calcolo non sono assolutamente affidabili.

La fase di prototipizzazione deve durare dal 2001 al 2003/4, utilizzando l'ultimo anno per definire i risultati, prima di procedere alla costruzione dei sistemi finali dal 2005 in poi. Le risorse dovrebbero seguire un piano di realizzazione nei tre anni che porti alla disponibilità del 10%, 30%, 100%.

Qui di seguito sono brevemente riassunte le necessità dei quattro esperimenti ad LHC per quel che riguarda la parte Italiana e con particolare riguardo alle risorse da installare nel Tier1.

2.1 ALICE

2.1.1 CENTRO REGIONALE ITALIANO PER L' ESPERIMENTO ALICE

Scopo del centro regionale italiano e' di fornire le risorse necessarie di calcolo per consentire una piena partecipazione della collaborazione italiana al programma di fisica di ALICE.

La collaborazione prevede di raccogliere, ogni anno, dati Pb-Pb per 10^6 sec in 4 settimane dati p-p per 10^7 secondi nel resto dell' anno. Si prevede di raccogliere 5×10^7 eventi Pb-Pb e 10^9 eventi p-p per anno. Il fascio di Pb sarà disponibile un anno più tardi rispetto al fascio p-p, questo aspetto è stato tenuto in considerazione nel progetto di evoluzione del sistema di calcolo.

La potenza di calcolo complessiva, stimata per i Tier-0 e i Tier-1 di ALICE, e riportata nel documento del LHC Computing Review è stimata in **1800 KSI95** (cpu) e **1600 TB** (disco). Al momento il contributo dei Tier-2 non è stato ancora stimato. Gli elementi chiave di questa valutazione sono:

- il volume dei dati (2.7 PB dati RAW acquisiti per anno)
- la complessità del singolo evento (oltre 80000 tracce per un evento centrale Pb-Pb).

La potenza di CPU necessaria per la ricostruzione di un singolo evento e' di **100 kSI95/sec**, la sua dimensione media **25 MByte** (2.5 Mbyte dopo la ricostruzione).

La potenza di CPU per la simulazione di un evento Pb-Pb e' **2250 kSI95/sec**, la sua dimensione **600 MByte**.

La stima per il **contributo italiano** è di **450 KSI95** e **400 TB** di disco (Tier-1+Tier2). Questa stima è basata su tre aspetti.

- B) La valutazione dell'impegno globale del calcolo di ALICE (cfr. quanto sopra)
- C) Il ruolo dei gruppi italiani all'interno della collaborazione: 92 FTE (17% del totale) di cui 26 sul calcolo.
- D) I principali contributi previsti per il calcolo dell'esperimento: Tier-1 sono previsti in Francia, Germania, Italia e Gran Bretagna (la GB ha però un gruppo di solo 8 persone e non è chiaro quale sarà l'effettiva disponibilità del centro di RAL per ALICE)

La frazione del calcolo prevista nel Tier-1 è stimata in 150 KSI95 per la CPU e 150 TB di disco. Tale stima è naturalmente da considerarsi del tutto preliminare in quanto basata su un modello di calcolo che dovrà essere validato dall'attività di sperimentazione. Il modello può, comunque, facilmente evolvere verso una ripartizione di risorse Tier-1 - Tier-2 più o meno concentrata. Il mass storage (nastri o il sistema più appropriato nella scala dei tempi di LHC) sarà interamente concentrato nel Tier-1. Il volume totale è valutato come segue:

- § volume ESD per anno per pass di ricostruzione: 270 TB,
- § volume dati Raw per anno: 270 TB (10% del totale)

Il problema dell'export dei Raw data dal CERN e della ripartizione, fra i Tier-1, del carico di CPU relativo alla ricostruzione RAW->ESD e' attualmente in discussione all' interno di ALICE. La frazione di dati RAW esportata nei Tier-1 potrebbe essere più elevata.

Tenendo presente quanto sopra il modello attuale prevede una ripartizione di ruoli di questo tipo:

- a) Tape repository per raw data e ESD: Tier-1
- b) Manutenzione supporto middleware : Tier-1 / Tier-2
- c) Storage su disco ESD (e livelli successivi): Tier-1 / Tier-2

- d) Simulazione: Tier-2/ Tier-1
 e) Manutenzione e supporto codice di produzione e analisi: Tier-2
 f) Manutenzione e supporto database: Tier-1/Tier-2

Nell'attuale modello di calcolo di ALICE Italia, che, come detto precedentemente, sarà oggetto di sperimentazione e valutazione, la ripartizione del carico di lavoro fra i diversi siti è legata all'effettiva disponibilità di risorse e competenze nella collaborazione. In questo senso è funzionale il contributo e il ruolo di tutti i siti partecipanti. Al momento sono previsti Bari, Bologna, Catania e Torino come siti Tier-2, Cagliari, Padova, Roma, Salerno, Roma e Trieste come siti Tier-3. Nella fase prototipale è previsto il seguente schema di sviluppo per il sistema di calcolo di ALICE e per il centro Tier-1 in particolare, con l'obiettivo di raggiungere alla fine del 2003 una dimensione pari a circa il 10% di quella finale (i dati esprimono l' integrale)

Tabella 1 –Capacity Targets per ALICE

Capacity targets for ALICE				
	units	2001	2002	2003
CPU capacity	SI95	4,000	22,000	45,000
Disk capacity	TBytes	6	18	40
Tape storage – capacity	TBytes	12	36	80

Tabella 2 –Capacity Targets per ALICE al Tier1

Capacity targets for ALICE @ Tier-1				
	units	2001	2002	2003
CPU capacity	SI95	1000	8,000	15,000
Disk capacity	TBytes	3	9	20
Tape storage – capacity	TBytes	12	36	80

Per ciò che concerne il personale si valuta siano necessarie 4/5 persone, presenti al Tier-1 per il supporto specifico per ALICE.

2.2 ATLAS

In questa sezione si riporta la stima delle risorse di computing necessarie globalmente per ATLAS, come presentata ed accettata nel corso della CERN LHC computing Review, e si procede poi a stimare le risorse necessarie in Italia e loro distribuzione gerarchica e geografica.

Le risorse di calcolo stimate necessarie da ATLAS

La giustificazione di queste stime non è fornita in questo documento e nessuna stima si può considerare precisa a meglio del 10–20%.

L'esigenza da soddisfare, sulla cui base sono fatte le stime di ATLAS, è il trattamento completo (comprendente calibrazione, ricostruzione, analisi, MC etc.) dei dati raccolti in un anno a luminosità $10^{33} \text{ cm}^{-2} \text{ sec}^{-1}$. L'anno si considera costituito da 10^7 sec effettivi di presa dati.

Il numero di eventi da trattare per ATLAS risulta quindi:

- $2.7 \cdot 10^9$ Raw Data (di cui $1.55 \cdot 10^9$ non di fondo).
 La migliore stima attuale di ATLAS valuta infatti il rate di presa dati in 270 Hertz. Il target originale era 100 Hertz e sono in corsi studi per valutare se e come sia possibile avvicinarsi maggiormente a questi 100 Hertz.
- $1.2 \cdot 10^8$ Eventi MC completamente simulati.

Stime di ATLAS sui formati dei dati e sulla CPU per processare un singolo evento:

- Raw Data (anche generati con MC) 2 MB/evento.
La stima originale era 1 MB e si sta cercando di riavvicinarsi a questa prima stima.
- ESD, Event Summary Data, output della ricostruzione 500 KB/evento
- AOD, Analysis Object Data, formato "pubblico" di analisi 10 KB/evento
- DPD, Derived Physics Data, formato privato di analisi, tipo ntupla, 1 KB/evento
- CPU per ricostruire un evento 640 SpecInt95*sec
- CPU per simulare un evento 3000 SpecInt95*sec
- CPU per generare AOD di un evento 25 SpecInt95*sec
- CPU per generare DPD di un evento 5 SpecInt95*sec
- CPU per analizzare AOD o DPD di un evento 3 SpecInt95*sec

Stime sulle risorse di calcolo globalmente necessarie e sulla parte non basata al CERN

Nell'elenco che segue il limite inferiore dell'intervallo è dato dalla valutazione per 270 Hertz e 2 MB; quello inferiore corrisponde ai valori originali 100 Hertz e 1 MB.

- CPU per Tier-0 al CERN + insieme di tutti i Tier-1 e Tier-2: 1760–1944 K SI-95
- CPU nelle installazioni puramente locali (Tier-3,4): 250–350 K SI-95 (qui l'incertezza non è connessa a rate e size degli eventi). Esistono anche stime considerevolmente più elevate, che potrebbero fare salire perfino di un 25% la stima globale della CPU di ATLAS; in questo documento ci si attiene alla stima che sembra più probabile, ma bisogna tenere conto che ci potrebbero essere revisioni al rialzo.
- CPU al CERN 506 K SI-95.
Questa cifra suppone che tutta la prima produzione dei Raw Data acquisiti e un completo re-processing (della durata di 3 mesi) si svolgano al CERN.
- Disco in un Tier-1 (compresi eventuali Tier-2 satelliti): 260–365 TB.
Questo valore è dominato dal 35% degli ESD che si suppone di mantenere sempre su disco e che ammontano da soli a 175–272 TB.
- Disco globale per ATLAS (esclusi Tier-3,4 e calcolando CERN + 6 Tier-1 come sopra): 1.9–2.57 PB
- Nastri (archivio) in un Tier-1 (compresi eventuali Tier-2 satelliti): 1260–1839 TB
- Nastri global per ATLAS (esclusi Tier-3,4 e calcolando CERN + 6 Tier-1 come sopra): 11.5–21.5 PB

La valutazione per ATLAS-Italia

Si assume l'ipotesi 270 Hertz e 2 MB.

La consistenza della parte italiana di ATLAS si può stimare in 10–11% della collaborazione totale. Sembra perciò coerente richiedere di installare in Italia (in totale : Tier-1,2,3,4)

- CPU ~250 K SI-95
- Disco ~400 TB
- Nastro ~2PB

Le esigenze di personale "s/w experts", sia per il supporto al s/w specifico di esperimento, sia per il supporto ai fisici-utenti, relativo sempre al s/w specifico di ATLAS, è valutabile in ~6–8 FTE.

Questo personale dovrà essere attivamente coinvolto nella fase di sviluppo s/w dell'esperimento, nei prossimi anni, in modo da poter poi svolgere efficientemente il ruolo di supporto. Attualmente sono in servizio solo 2 esperti s/w di questo tipo in ATLAS, con posizioni temporanee. Da 4 a 6

ulteriori FTE devono quindi essere acquisiti, preferibilmente nei prossimi 2 anni, e le competenze così formate devono essere acquisite stabilmente nell'esperienza.

A regime 1-2 FTE dovranno essere collocati presso il Tier-1, ma la maggior parte dovrà mantenere uno stretto contatto con i fisici impegnati nell'esperienza. La struttura Tier-2 potrebbe fornire un buon bilanciamento fra la necessità di contatto con i fisici e l'esigenza di efficienza che porta ad un certo grado di centralizzazione in alcune delle attività di supporto s/w.

Per quanto riguarda le funzioni dei diversi Tiers, in linea con le posizioni di ATLAS internazionale, si assegna in esclusiva al Tier1 tutto ciò che ha Raw come input (reprocessing etc.) e tutto ciò che ha ESD come input. I Tier2 hanno storage completo degli AOD, e sono usati per l'analisi e la produzione MC. I Tier3 producono MC nei cicli spare e storano i dati che servono per le analisi "locali". La suddivisione delle funzioni Tier1/2 e Tier2/3 sarà determinata dopo i 3 anni di sperimentazione GRID. Per la fase di sperimentazione prevediamo 2-3 siti in sezioni ATLAS per studiare le funzionalità Tier2 (in rapporto a quelle Tier1 e Tier3). Il numero finale di Tier2 di ATLAS in Italia sarà probabilmente 0 o 1, ma l'opzione 0 è ancora aperta.

Per il modello di ATLAS un Tier-1 è necessario; dato che in ATLAS non c'è un gruppo di sezioni con risorse umane (di computing experts) sufficienti per supportare un Tier-1 distribuito, la scelta naturale (anche tecnologicamente meno demanding) è quella di un Tier-1 in un singolo sito. La soluzione multiesperimento presenta il vantaggio di una potenziale ottimizzazione delle risorse umane, e il potenziale svantaggio di una minore rispondenza alle necessità di ATLAS. La gestione del Tier-1 e i meccanismi per lo sharing delle risorse dovranno perciò essere studiati con cura.

Linee per la suddivisione delle risorse fra i vari Tiers a regime (nel 2007);

- CPU: ~50 K SI-95 nei Tier-3 (da 4 a 10 K SI-95 per ciascuno a seconda della dimensione) ; ~200 nell'insieme Tier-1,2 riservando comunque almeno ~140 K SI-95 al Tier-1
- Disco: almeno ~ 300 TB per il Tier-1 e almeno ~30 TB nell'insieme dei Tier-3.
- Nastri: almeno ~1.6 PB nel Tier-1. ATLAS prevede di usare i nastri come archivio esclusivamente i quindi **non ritiene necessario un "full MSS system" neppure nei Tier-1.**
Si studierà la possibilità di concentrare l'archivio su Tier-1,2 liberando così i Tier-3 da questa necessità.

La valutazione per ATLAS in Italia nella fase dei prototipi

Nella fase conclusiva della sperimentazione (seconda metà del 2003) si valuta che le risorse necessarie in Italia saranno approssimativamente il 10% di quelle a regime:

- CPU: ~20-25 K SI-95 (corrispondenti a ~100 biprocessori): ~50-60% nel Tier-1, 4-6 K SI-95 nei 2-3 Tier-2 (con 2-2.5 ciascuno, equivalenti a 10-12 biprocessori). In ciascuno dei 5-6 Tier-3 andrebbero 0.6-1. K SI-95 (da 3-5 biprocessori) a seconda dell'utenza locale.
- Disco: ~30 TB di cui ~15 TB al Tier-1, ~1 TB per ogni Tier-3, ~3-4 TB per ogni Tier-2.
- Nastro ~ 60 TB, distribuiti come il disco.
- Personale di supporto al s/w specifico di ATLAS presso il Tier-1: 1FTE

- **CMS**

- **Tier1 di CMS in comune con altri Esperimenti INFN**

- **Premessa**

- Il Modello di calcolo di CMS è un modello *integrato di Funzionalità* dei Tier1, Tier2 e Tier3
- Un Tier1 unico per tutto l'INFN comporta una maggior importanza di funzioni nei Tier2 (e 3).
- Essenziale il ruolo dei Tier2 ancora di più nella fase in cui il Tier1 unico non è ancora "utilizzabile".
- Le Sedi candidate per i Tier2 sono: LNL+Pd, Ba, Bo, Pi, Rm1
- Le Sedi candidate per i Tier3 sono: Ct, Fi, Pg, Pv, To
- Le funzionalità del modello riguardano due principali attività: "scheduled" e "cahotic". Le prime sono coordiante a livello della Collaborazione (o dei Gruppi di Analisi), le seconde riguardano la analisi individuale (o di gruppi di ricercatori)

- **Tier1 CMS-INFN**

- **Funzioni**

§ Il Tier1 di CMS Italia è la Base per le attività "schedulate" dalla Collaborazione CMS

⇒ Supporto per lo storage dei RAW data

⇒ Supporto per la produzione degli ESD da re-processing coordinato dalla Collaborazione e/o dai Physics Groups

⇒ Controllo dei Database distribuiti in tutta la Collaborazione

⇒ Supporto per le produzioni di grandi Datasamples (anche AOD e/o simulazioni)

⇒ Parziale supporto sistemistico per gli Utenti CMS

⇒ Coordinazione "tecnica" verso gli altri Centri di CMS (ovviamente insieme ai responsabili di CMS Italia)

⇒ Supporto e sviluppo di "alcuni" tools di Middleware (GRID) di interesse comune anche di CMS

⇒ Etc.

§ Il Tier1 di CMS Italia contribuisce alle analisi di larghe dimensioni e che richiedono il contributo delle risorse (storage e CPU) colà locate

Dimensioni

§ Le tabelle che seguono danno la dimensione "media" di un Tier1 di CMS a regime (2007) per quanto riguarda le necessità di storage e di potenza di CPU (senza efficienza di uso!, ovvero sono le risorse che devono essere disponibili per gli utilizzatori)

Tabella 3 – Necessità Medie di Storage per un Tier 1 di CMS

STORAGE	# of events	Ev-size (MBytes)	Active Tape/Archive tape/Disk(Tbytes)
SIM.Out	1.E+08	2.	0/200/0
SIM.Rec.	1.E+08	0.4	40/0/30
Raw-sample	5.E+07	1.	50/0/0
Calibration			0/10/10

ESD	1.E+09	0.5	500/0/0
Re-proc.ESD	0.2E+09	0.5	0/100/40
Re-vised ESD	0.2E+09	0.5	0/100/40
General AOD	1.E+09	0.01	0/10/10
Revised AOD	2.E+08	0.01	0/2/2
TAG	1.E+09	0.001	0/1/1
Local AOD,TAG, DPD	2.E+08		0/10/10
User data			0/0/50
Cache Disk for active Tapes			0/0/120
Total			590/433/313

Tabella 4 – Necessità Medie di CPU per un Tier 1 di CMS

Data Processing	# events	CPU per event kSI95/ev.s	CPU total kSI95
Simulation	0.25E+08	5.	5
Rec-Simulation	0.25E+08	3.	3
Re-Processing	0.1E+09	3.	50
Selection	1.E+07–1.E+08	Up to 0.025	15
Analysis	1.E+07	0.010	80
TOTAL			153

- Il Tier1 italiano per CMS è pensato con un supporto "minimo" di storage su nastro, in quanto CMS Italia è confidente che sarà possibile fare a meno di "active tapes". La tecnologia ha oggi questa tendenza e, se mantenuta, renderà possibile l'uso di soli dischi e unità di backup su storage permanente di nuova concezione. Questo permette forti risparmi sul personale esperto di "mass storage" e sui costosi sistemi di automated tapes. Inoltre l'efficienza nell'accesso ai dati verrebbe fortemente aumentata.
- Le dimensioni a regime (2007) per il Tier1 di CMS Italia in comune con altri esperimenti risulta quindi, per quanto detto nella premessa, di circa (incluso l'efficienza d'uso):

Tabella 5 – Risorse necessarie nel Tier 1 Italiano di CMS

Tier1 di CMS in comune	Disk storage	CPU power	Archive tapes
	400 TB	100 kSI95	500 TB

- Le funzionalità e le risorse del Tier1 italiano di CMS sono comunque assicurate dalla sua integrazione con i Tier2 di CMS Italia, in modo da rendere il Tier1 di CMS Italia conforme alla media dei Tier1 di CMS. Infatti i Tier2 nel modello di CMS Italia contribuiscono per le seguenti risorse:

Tabella 6 – Risorse necessarie per tutti i Tier 2 Italiani di CMS

Tier2 di CMS integrati	Disk storage	CPU power	Archive tapes
	800 TB	240 kSi95	250 TB

- Alle risorse dei Tier2 vanno aggiunte quelle dei Tier3 di CMS Italia. Questo porta ad uno sharing delle risorse che è approssimativamente:
50% dello storage al Tier1 e 50% ai Tier2/3
35% della CPU al Tier1 e 65% ai Tier2/3
Tali frazioni sono giustificate dalla necessità di storage per le attività "collaboration coordinated" (svolta fortemente nel Tier1) e dalla maggior incidenza della richiesta di CPU nelle attività di analisi (dominata dai Tier2/3).
- La realizzazione del Tier1 dovrà seguire le analoghe implementazioni degli altri Tier1 di CMS (FNAL, RAL, CCIN2P3–Lyon, Moskow) e del Tier0/1 di CMS al CERN. Quest'ultimo è oggi previsto avere le seguenti frazioni percentuali realizzate negli anni: 2005 = 30%, 2006 = 60%, 2007 = 100%.
- La realizzazione dei Tier2 risulta parte integrante del modello e deve quindi seguire lo stesso profilo, in modo da assicurare le necessarie funzionalità integrate e realizzare un Tier1 "standard" nei confronti della Collaborazione (ed i necessari Tier2 "standard").

Personale

- Non va sottovalutato che il personale da reperire deve avere differenti specificità professionali: dal sistemista all'operatore, dall'architetto all'ingegnere del sw, dal fisico del computing al fisico delle procedure di analisi, dall'esperto di hardware all'esperto di software tools, etc.
Ovviamente senza considerare il personale di supporto: amministrazione, segreterie, guardianerie, etc.
Il "sistema" è 24x365 e così deve essere.
- Il personale coinvolto ha due caratterizzazioni: gestione del sistema e mantenimento/sviluppo delle specifiche necessità di CMS
- Il personale di gestione del sistema, per la parte che riguarda CMS, è stimato in circa 10 unità (non FTE). Le funzioni sono il controllo del sistema, la sua implementazione sistemistica, la gestione del network (locale e remoto), la gestione del(i) database(s), l'upgrade del sistema, il recovery, il supporto sistemistico, la coordinazione tecnica, il supporto dei Tools (di Grid e non), etc.
- Occorre anche che ci sia personale "CMS specific" che si occupa del centro Tier1. Tale personale è valutato dell'ordine di 4–5 FTE (non persone), di cui circa la metà presso il Tier1 e circa la metà dislocate presso i Tier2.
La funzione di questo personale è di fornire il supporto e lo sviluppo del software (in senso lato, cioè includendo le specifiche degli OS, compilatori, librerie, ambiente, grid tools specifici, etc) specifico di CMS. Il supporto agli utenti, ma solo per ciò che riguarda il sw di CMS deve essere garantito da questi FTE, essendo il resto del supporto agli utenti garantito dal personale del punto precedente.
- A questo personale va aggiunto il personale (comune?) che si preoccupa dello sviluppo e dell'upgrade tecnologico. La consistenza di questa forza lavoro può rendere il Tier1 competitivo e fare la differenza tra un progetto che ha successo e un progetto che fallisce in pochi anni.
- **Ruolo verso i Tier2(3) CMS–INFN**
- Il coordinamento delle attività che richiedono l'intervento dei Tier2 nei jobs "scheduled" deve essere garantito per CMS dal Tier1 verso i Tier2 e Tier3.
- La disponibilità delle risorse deve essere garantita ai Tier2/3 per quelle attività di analisi che richiedono anche la frazione di risorse dislocate presso il Tier1 di CMS (seppur in

comune), naturalmente in modo coordinato.

Qualora la diponibilità dei dati e delle risorse per queste attività non risultasse disponibile, occorre che il Tier1 se ne faccia carico sia come responsabilità che come operatività.

- Occorre pertanto personale (un responsabile?) con poteri decisonali presso il Tier1 per le succitate attività e che possa disporre operativamente delle risorse.
- I Tier2/3 di CMS Italia devono quindi vedere il Tier1 come una naturale e "trasparente" estensione delle loro capacità di analisi.

- **Ruolo verso la Collaborazione CMS**

- Il Tier1 di CMS Italia è parte integrante del "Computing" di CMS e pertanto rappresenta l'INFN nell'Esperimento. Responsabilità e disponibilità devono essere garantite come (e più) di un qualunque detector.

L'accessibilità da parte di utenti, gruppi di analisi, gruppi di produzione, re-processing dei dati, anche da parte di altri Tier1 od eventualmente altri Tier2 deve essere garantita.

- La rappresentatività del Tier1 CMS Italia nei comitati di CMS deve essere garantita da personale di CMS con potere effettivo sulle risorse e sulla gestione del Tier1 unico.

- **Sviluppo del Tier1 CMS-INFN nel Modello di calcolo CMS Italia**

Risorse necessarie negli anni di prototipizzazione

- Le risorse dei Tier-n italiani nel corso degli anni di sperimentazione sono valutabili sia con il processo Top-down (Sperimentazione) che Bottom-up (Produzione di Fisica con la simulazione). (I valori espressi sono per l'anno indicato, ovvero rappresentano l'incremento).
- La programmazione delle risorse dei Tier-n di CMS Italia risulta essere per il processo Top-down.

Tabella 7 – Target 2003 Profile 2001–3 [20%, 30%, 50%] TOP-DOWN

	CPU (kSI95)			Disk (TB)			Tape (TB)		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
Tier1 unico	2.0	3.0	5.0	8	12	20	15	22	38
5 Tier2	4.8	7.2	12.0	16	24	40	0	0	0
+ 5 Tier3	0.4	0.6	1.0	1	2	3	0	0	0
Total	7	11	18	25	38	63	15	22	38

- La programmazione delle risorse dei Tier-n di CMS Italia risulta essere per il processo Bottom-up.

Tabella 8 – Target 2003 Profile 2001–3 BOTTOM-UP (raddoppio risorse per anno)

	CPU (kSI95)			Disk (TB)			Tape (TB)		
	2001	2002	2003	2001	2002	2003	2001	2002	2003
Tier1 unico	1.8	3.7	5.5	6.7	13.3	20.0	17.0	33.0	50.0
5 Tier2	2.7	5.3	8.0	11.7	23.3	35.0	3.3	6.7	10.0
+ 5 Tier3	0.7	1.3	2.0	1.7	3.3	5.0	0.0	0.0	0.0
Total	5.2	10.3	15.5	20.1	39.9	60.0	20.3	39.7	60.0

4.1 LHCb

4.1.1 Il modello di calcolo di LHCb

Per molti aspetti il modello di calcolo di LHCb è assai simile al modello di calcolo degli altri esperimenti di LHC.

La principale differenza consiste nel fatto che nel modello di calcolo di LHCb **non si identificano esplicitamente gruppi di analisi e la prima fase dell'analisi** (ricostruzione degli stati finali e classificazione degli eventi) **viene eseguita in comune** per tutte le analisi che seguono, nella fase di produzione (analisi di produzione).

Questo significa che, qualora un evento soddisfi più di un criterio di *tagging*, su tale evento deve essere eseguito più di un algoritmo di analisi, ragion per cui questa fase dell'analisi è notevolmente impegnativa in termini di potenza di calcolo richiesta. Il vantaggio consiste nella possibilità di eseguire la prima fase dell'analisi in appendice al processo di produzione (al CERN per quanto riguarda i dati reali e nei centri Tier-1 per quanto riguarda i dati Monte Carlo) e di distribuire i dati a un livello di raffinazione più avanzato (gli AOD in luogo degli ESD), con notevole risparmio di memoria di massa e di carico della WAN. Si prevede di archiviare gli ESD soltanto nei centri in cui sono prodotti.

L'analisi di LHCb viene eseguita quasi esclusivamente sugli AOD. L'accesso a RAW ed ESD si prevede più intenso all'avvio dell'esperimento (10% degli eventi nei primi 2 anni), mentre a regime si prevede che l'accesso agli ESD sia limitato all'1% degli eventi e quello ai RAW allo 0.001%.

La fase di analisi richiede la **collaborazione di Tier-1 e Tier-3**. Il **Tier-1 seleziona** gli eventi sulla base di un determinato criterio e li invia al **Tier-3** il quale li **processa** e produce i DPD. Esiste tuttavia un canale di reazione particolarmente popolato ($B^0 \rightarrow D^* \pi$, con un ordine di grandezza in più di eventi rispetto al canale tipico) per il quale si ritiene più conveniente produrre i DPD direttamente nel Tier-1.

In sintesi, nel **Tier-1** si prevede di eseguire la produzione di eventi Monte Carlo, l'analisi di produzione dei medesimi eventi e la selezione degli eventi per l'analisi utente, mentre nei **Tier-3** si prevede di eseguire l'elaborazione degli eventi per l'analisi utente e l'analisi interattiva dei DPD.

Per quanto concerne la memoria di massa, nei **Tier-1** devono essere archiviati i RAW e gli ESD prodotti nel centro stesso, insieme all'intera produzione degli AOD. Nei **Tier-3** devono invece essere archiviati gli AOD selezionati per l'analisi utente e i DPD utilizzati per l'analisi interattiva.

4.1.2 Risorse di calcolo necessarie

Gran parte della potenza di CPU dei centri regionali (circa l'80%) è richiesta per la generazione di eventi Monte Carlo. Le stime che seguono sono preliminari e verranno di volta in volta corrette con l'avanzamento della comprensione delle esigenze di calcolo.

L'esperimento LHCb utilizza un trigger fortemente selettivo (ricerca del vertice secondario al I livello ed efficiente particle identification basata sui RICH al III livello) per la selezione degli eventi di decadimento del mesone B. Se si suppone, come è ragionevole, di simulare un numero di eventi di segnale pari a 10 volte gli eventi reali, occorre produrre circa 10^7 eventi Monte Carlo per anno. Tale produzione richiede all'incirca 110 kSi95.

Per quanto riguarda il fondo $b\bar{b}$, in un anno si producono circa 10^{12} eventi. Se si supponesse di simulare un numero di eventi pari al numero di eventi reali, la potenza di calcolo richiesta sarebbe circa 3×10^7 Si95. Ovviamente occorre trovare il modo di ottimizzare la produzione degli eventi di fondo. È in fase di studio la possibilità di archiviare e riutilizzare i dati prodotti a livello del generatore; di ottimizzare la fase di generazione senza bias fisici; di concentrarsi sui tipi di fondo particolarmente pericolosi per uno specifico canale di analisi. L'obiettivo è quello di ridurre la richiesta di più di un ordine di grandezza. Si può supporre perciò di richiedere circa 440 kSi95. La potenza totale di CPU richiesta per la produzione Monte Carlo è perciò 550 kSi95, da dividersi per 5 centri regionali Tier-1.

Tabella 9 – Tabella dei Requisiti del Tier-1 di LHCb

Fase dell'elaborazione	Dati in uscita	Potenza di CPU [SI95]	Spazio su Disco [TB]	Spazio su nastro attivo [TB]	Spazio su nastro di backup [TB]
Dati reali	AOD + TAG	?	40	80	0
Simulazione / Ricostruzione	RAW + ESD	110000	23	70	40
Analisi di produzione MC	AOD + TAG	8000	18	35	0
Calibrazione	?	?	10	0	10
Cache su disco per lo staging	?	?	15	0	0
Analisi utente	DPD	23000	5	0	5
Totale	?	141000	111	185	55

Si prevede di ripetere l'analisi di produzione degli eventi Monte Carlo (generazione degli AOD) una volta al mese; la CPU richiesta è stimata in circa 40 kSi95 da dividersi per 5 centri regionali Tier-1.

La potenza di CPU richiesta per l'analisi utente (produzione dei DPD) è stimata in circa 160 kSi95, supponendo che circa 140 fisici siano attivi nell'analisi (20 per ogni centro regionale e 40 al CERN), che essi eseguano in media 2 job al giorno e che il processo si concluda in 4 ore. La ripartizione di tale potenza di calcolo sarà: 46 kSi95 al CERN, 23 kSi95 per ogni centro Tier-1. La tabella 1 riporta le risorse di calcolo e di memoria di massa previste per un centro regionale Tier-1 di LHCb.

4.1.3 Distribuzione del calcolo di LHCb-Italia tra i Tier-n

Poiché nell'esperimento LHCb la distribuzione dei dati avviene a un livello di raffinazione più avanzato di quello previsto negli altri esperimenti LHC (AOD invece che ESD), la funzione dei centri Tier-2 in LHCb risulta pure diversa.

In LHCb si prevede di costruire centri Tier-2 negli stati privi di Tier-1 o laddove pre-esistano consistenti concentrazioni di calcolo in siti diversi dal Tier-1 che giustifichino il riutilizzo di risorse di calcolo e umane (come nel caso degli oltre 300 PC di MAP a Liverpool).

Poiché in Italia si prevede di costruire un Tier-1 e tuttora non esistono consistenti concentrazioni di risorse di calcolo, le ragioni del Tier-2 per LHCb vengono meno.

Si prevede invece la costruzione di Tier-3 (uno per sede) con potenza di calcolo sufficiente a consentire l'esecuzione dell'analisi. La distribuzione del calcolo in Italia tra i diversi strati è mostrata nella tabella 2.

Tabella 10 – Tabella dell'Organizzazione dei Tier-n di LHCb in Italia.

centro	Numer o	locazione	risorse
Tier-1	1	CNAF	141 kSI95, 111 TB disco, 185 TB nastro attivo, 55 TB nastro di backup.
Tier-2	?	?	?

Tier-3	9	Bologna, Cagliari, Ferrara, Firenze, Frascati, Genova, Milano, Roma1, Roma2	Possono variare dipendentemente dalla dimensione del gruppo e dall'attività di analisi. Valori medi normalizzati a 10 fisici: 5 kSI95, 10 TB disco.
--------	---	---	---

Il personale che si stima necessario al funzionamento dei centri Tier-n di LHCb in Italia è riportato nella tabella 3.

Tabella 11 – Personale necessario al calcolo dell'esperimento LHCb in Italia.

	Tier-1 [FTE]	Tier-3 [FTE]
Supporto per funzioni di ricerca e sviluppo e software tools generali	4	?
Supporto per il software specifico dell'esperimento LHCb	2	0.5
Amministrazione dei sistemi	2	0.5

5 Le necessità di altri Esperimenti

Le necessità di calcolo associate con le attività dell'INFN non si esauriscono con LHC. In questo documento sono presi in considerazione altri due esperimenti che hanno anch'essi un impatto rilevante: BaBar e Virgo. Entrambi hanno tempi di realizzazione più stretti degli esperimenti ad LHC, ma richiedono risorse confrontabili a quelle dei test di LHC in un contesto dove, a differenza di questi ultimi, il servizio fornito diventerebbe immediatamente requisito fondamentale per il successo degli esperimenti stessi. In questo senso essi possono fungere da vero banco di prova per le funzionalità che il Centro dovrà fornire per LHC a regime.

Nel seguito discuteremo in modo più dettagliato le necessità di BaBar e Virgo.

5.1 BaBar

5.1.1 Il centro Tier A per BaBar

Per potenziare le risorse di calcolo necessarie alla prevista crescita della luminosità che PEP-II dovrebbe alimentare nei prossimi anni (v. seguente tabella), la collaborazione BaBar ha deciso di adottare un nuovo modello di calcolo che è stato approvato alla fine del 2000.

Tabella 12 – Luminosità attesa negli anni 2001-2005

Anno	Luminosità di picco ($10^{33} \text{cm}^{-2}\text{s}^{-1}$)	Luminosità integrata per anno (fb^{-1})
2000	2	25
2001	5	40
2002	8	80
2003	10	115
2004	13	135
2005	24	225

Esso stabilisce che, accanto ai centri regionali (Tier B, del tipo di quelli già funzionanti presso il Caspur e RAL) e ai centri locali presso le istituzioni partecipanti (Tier C), sia possibile costituire dei centri Tier A in grado di condividere con SLAC l'onere di mantenere una parte dell'event

store, fornire potenza di calcolo per l'analisi e contribuire alla produzione e al reprocessing dei dati.

Gli eventuali nuovi Tier A, oltre a venire considerati come contributi in-kind da parte dei paesi ospitanti, porterebbero ad un miglioramento delle potenzialità a disposizione della collaborazione e condurrebbero ad una diminuzione complessiva del common fund gravante sulle agenzie finanziatrici.

5.1.2 Il Tier A: requisiti

I requisiti che un centro Tier A dovrà soddisfare sono:

- A) Garantire accesso interattivo e batch a tutti i collaboratori (> 500 persone), 24x7.
- B) Utilizzare piattaforme standard di BaBar.
- C) Provvedere risorse di calcolo pari ad almeno il 30% di quelle globali minime richieste, in ipotesi di volume totale dei dati dell'esperimento (incluse le simulazioni) residente a SLAC. Parte di tali dati saranno mantenuti in unica copia presso ciascuno dei Tier A dell'esperimento.
- D) Mantenere un ambiente di produzione completo.
- E) Fornire accesso al mass storage tramite un sistema di staging locale, attraverso l'interfaccia ufficiale di BaBar.
- F) Garantire una banda passante sufficiente al normale aggiornamento via rete dell'event store al rate prodotto dall'esperimento.

Finora l'unico centro Tier A approvato, al di fuori di SLAC, è quello di Lione che dovrebbe entrare in funzione a ottobre 2001. Altre espressioni di interesse sono state formulate dai gruppi inglese e tedesco.

Le dimensioni del Tier A che conseguono dai requisiti precedenti sono stimate nella seguente tabella dove è indicato l'andamento temporale in funzione della luminosità prevista.

Tabella 13 – Tier A totale (senza contingency)

			200 1	200 2	200 3	200 4	200 5
Disco	Incremento	TB	20	17	12	14	23
	<i>Integrale</i>		20	37	49	62	85
CPU	Incremento	KSI2000	90	165	221	264	481
	<i>Integrale</i>		90	255	476	739	1220
Nastro	Incremento	PB	0.19	0.2	0.26	0.3	0.5
	<i>Integrale</i>		0.19	0.4	0.65	0.95	1.45
Rete	<i>Bandwidth per Micro</i>	MBytes/s	1.0	1.1	1.1	1.1	1.8
	<i>Bandwidth per Tutto</i>		11.2	19.5	24.4	28.6	47.7

La tabella mostra le risorse di calcolo fondamentali necessarie al primo ordine. Vi sono poi da considerare risorse accessorie che vanno incluse nelle stime dei costi.

Per esempio i dischi utilizzati necessitano di servers che aumentano il costo globale di circa il 25%. Inoltre BaBar avrà bisogno, almeno fino a quando non vi siano alternative disponibili sul mercato, di server del tipo di quelli utilizzati a SLAC (Sun), sia per motivi di prestazioni, che per necessità di utilizzare software (p.es.: AMS server) non immediatamente disponibile per Linux.

Vi è poi la necessità di avere a disposizione un sistema di mass storage per la gestione dello staging e della migrazione automatica da disco a nastro e viceversa. Sia a SLAC che a Lione viene utilizzato HPSS (High Performance Storage System), ma esso probabilmente risulterebbe troppo

costoso ed oneroso da mantenere presso il Centro Regionale, se venisse utilizzato solo da BaBar. Alternative esistono ma vanno attentamente valutate: questa è la parte dell'infrastruttura di calcolo di BaBar, che per il notevole investimento finanziario e di personale rappresenta l'elemento critico della tabella di marcia per la allestimento del Tier A.

Le stime della precedente tabella sono il risultato di una proiezione dettagliata, basata sulle risorse di calcolo resesi necessarie durante il primo anno e mezzo di funzionamento dell'esperimento. Lo spazio dello storage è stimato assumendo una riduzione progressiva delle dimensioni degli eventi e della frazione tenuta su disco, come riportato nella seguente tabella. La potenza di calcolo comprende sia le necessità dell'analisi, che della produzione (MC, reprocessing, ecc.).

Tabella 14 – Stime delle dimensioni degli eventi nei vari Formati

Formato	Utilizzo	Grandezza Eventi Adronici		Frazione su Disco	
		2001 KB	2005 KB	2001 %	2005 %
Tag	Selezione veloce degli eventi	0.7	0.2	100	100
Micro (AOD)	Selezione degli eventi più raffinata, Analisi Fisica	8	2	100	50
Mini (ESD)	Analisi di Fisica dettagliata e studi sul rivelatore, ricalibrizioni	20	20	30	10
Reco	Studi dettagliati sul rivelatore, analisi particolarmente dettagliata	230	100	5	2
Raw	Dati Raw, possono essere usati per ricostruzione.	140	<100	5	2

Sul versante del software, l'elenco seguente riporta la lista di ciò per cui deve essere attualmente garantita installazione e manutenzione:

- a) AFS Andrew File System (remote access to SLAC files)
- b) bison, (GNU), project parser generator, yacc replacement.
- c) CERNLIB, CERN's program library.
- d) CVS, (GNU), Concurrent Version System.
- e) flex, (GNU flex), lexical analyzer generator .
- f) gcc, (GNU cc), C, C++, and Objective-C compiler.
- g) gdiff, (GNU diff), display differences between two files.
- h) gfind, (GNU find), search for files that matches regular expression.
- i) gmake, (GNU make), remake portion of software that need to be re-compiled.
- j) gtar, (GNU tar), read/write multiple files from/to single file.
- k) Objectivity, Objectivity/DB federated databases
- l) patch, (GNU patch), apply a diff file to an original.
- m) Perl, shell script language.
- n) RCS, (GNU), Revision Control System, manages multiple revisions.
- o) STL, Standard Template Library.
- p) Tcl/Tk, a shell script language. Tcl and Tk together provide Graphical User Interface.
- q) Tools.h++ (Rogue Wave), the C++ foundation class library from RogueWave.
- r) X11, the X11 graphic interface.
- s) Motif, the X11 window manager.
- t) Java
- u) database relazionale

In gran parte si tratta di strumenti che non richiedono un particolare impegno, eccezion fatta per Objectivity, il database ad oggetti utilizzato per la gestione dell'event store e degli altri principali database (condition, configuration, ecc.) sui quali è basata la produzione e l'analisi dei dati.

Le prestazioni nell' accedere e nel memorizzare gli eventi dipendono criticamente dall' ottimizzazione dell' utilizzo di Objectivity e presentano ostacoli non banali di scalabilità. Inoltre, la manutenzione ordinaria si è rivelata molto impegnativa per la complessità del sistema (v., per esempio, la gestione dei lock sul database).

Di conseguenza è stata necessaria a SLAC e lo è anche in un Tier A la presenza di personale tecnologo specifico.

Per le normali funzionalità di bookkeeping, occorre poi avere a disposizione un DBMS di tipo relazionale che richiede anch'esso di essere seguito da personale specializzato.

5.1.3 Tempi di realizzazione.

La creazione di un Tier A INFN di tali dimensioni non si può realizzare in tempi brevissimi, trattandosi di costruire un centro completamente ex-novo. D'altro canto, sembra ragionevole programmare l'inizio di operazione di un tale centro, al più tardi, per l'inizio del 2003, in quanto altrimenti:

- b) la sua realizzazione entrerebbe in conflitto con quella dei Tier 1 per gli esperimenti LHC
- c) il contributo al calcolo di BaBar si concretizzerebbe troppo tardi.

d) Il personale necessario

L' esigenze di personale per il Tier A di BaBar sono simili a quelle degli altri esperimenti LHC, con delle ulteriori necessità specifiche che sono così riassumibili:

- § 2 FTE per il mass storage
- § 2 FTE per gestione Objectivity
- § 1 FTE per import/export dei dati
- § personale per supporto di una larga base di utenti residenti parte in Europa e parte in America, per i quali occorre provvedere un servizio on-call 7x24 in lingua inglese.
- § 1 persona che si occupi dei problemi di sicurezza e di affidabilità nell' accesso al centro.
- § 2 FTE per il software di esperimento.

Attualmente il gruppo BaBar ha formato un nucleo di tecnologi del software che hanno acquisito familiarità con l'ambiente di calcolo dell' esperimento. Questo pool, in larga misura creato ex-novo, comprende:

- § un tecnico ed un tecnologo INFN:
- § un tecnico assunto con l'art. 15
- § tre borsisti tecnologi INFN (ai quali speriamo se ne aggiunga presto almeno un quarto)

Esso è stato essenziale per:

- § creare e mantenere la farm per analisi e produzione simulazione di Roma, presso il CASPUR,
- § fornire la quota italiana di Computing Professionals a SLAC,
- § collaborare nello sviluppo dei tools di import/export di dati e MC.

e certamente le competenze acquisite potrebbero essere riutilizzate nel Tier A, specialmente per il supporto al software specifico di esperimento. Tuttavia, perché ciò si realizzi, compatibilmente on gli impegni già in essere, è necessario che il gruppo venga potenziato rapidamente con nuove posizioni a più lungo termine.

5.2 VIRGO

L'interferometro centrale di VIRGO (CITF) produrrà dati nel 2001. Questi devono essere analizzati

a fini diagnostici della macchina ed allo scopo di testare e raffinare gli algoritmi di analisi.

Dati dell'interferometro completo sono previsti per il 2003. Entro questa data, le procedure d'analisi devono essere mature ed un forte nucleo delle risorse di calcolo necessarie deve essere disponibile.

L'ipotesi secondo cui è stato sviluppato il Computing Model di VIRGO (VCM) è basata sull'assunzione dell'esistenza di un Tier di base con almeno 300 Gflop disponibili per la ricerca di segnali generati nella fase terminale della coalescenza di sistemi binari (chirp), 1 Tflop per la ricerca di segnali continui emessi da stelle di neutroni rotanti nella nostra galassia ed un sistema di archiviazione dati (includente un robot per il retrieval process). A questo sito si aggiungono i Tier di livello inferiore che fungono da polo di sviluppo e complemento delle due specifiche attività quali la ricerca dei segnali continui ed il riprocessamento dati relativo alla ricerca dei segnali tipo "chirp".

Il VCM prevede la duplicazione dei dati originali prodotti dall'interferometro (Raw Data) così da costituire due archivi in siti differenti. Allo stato attuale in Francia la discussione è concentrata sull'uso delle risorse disponibili a Lione, in particolare per la creazione di un archivio dei dati e per un'attività di off line computing comprensiva di re-processing su Raw Data e di test per lo sviluppo di tecniche di analisi per la rivelazione di impulsi brevi da collassi stellari.

5.2.1 Il flusso dei dati acquisiti in Virgo

Il sistema d'acquisizione dati di Virgo genera il flusso di dati che viene sintetizzato nella tabella seguente.

Tabella 15 – Flusso dei Raw e Processed Data e dei Selected Data per differenti scale dei tempi

	Raw data	Processed data	Selected data
Flow/s	4.0 MB	.3 MB	.34 MB (1%) .7 MB (10%)
Flow/day	350 GB	26 GB	30 GB (1%) 60 GB (10%)
Flow/year	126 TB	9.5 TB	11 TB (1%) 22 TB (10%)
Tapes/year (50 GB DLT)	2500	190	220 (1%) 440 (10%)

Il sistema d'acquisizione è già attivo a Cascina, tuttavia questi numeri sono ancora il frutto di una valutazione che include un fattore di compressione stimato pari a 2. I dati raccolti in "framè" costituiscono l'insieme dei Raw Data. Data la natura ed il livello di frontiera del rivelatore, questi flussi sono destinati a variare per eccesso o per difetto in funzione dell'abilità degli sperimentatori nel comprendere e modellare tutte le sorgenti di rumore proprie dell'antenna. Il numero dei nastri magnetici che deve essere utilizzato in un anno è stato valutato assumendo nastri del tipo DLT di 50 GB di capacità. Il valore in parentesi corrisponde alla frazione di dati selezionati dai trigger on-line. I "processed" e i "selected" data sono immagazzinati su dischi del cluster di computer on line e sono accessibili tramite un sistema di distribuzione dati.

5.2.2 Attività di analisi off-line

Riportiamo qui di seguito una lista necessariamente parziale, relativa alle attività di calcolo off-line previste per Virgo .

Simulazione dati di Virgo. Questa attività è basata sul processamento di dati simulati in ambiente SIESTA. Il flusso dati, così come il formato, è analogo a quello dei dati reali. La simulazione è fondamentale per definire l'efficienza dei vari passi di analisi previsti lungo la catena d'elaborazione on-line.

Caratterizzazione delle sorgenti di rumore dell'interferometro. L'analisi è strettamente connessa con lo sviluppo della simulazione ed è prevista già per l'interferometro centrale. Essa implica un pesante lavoro di riprocessamento dei Raw Data ed ha come obiettivi primari la caratterizzazione delle sorgenti di rumore dell'apparato, lo sbiancamento dei dati, la verifica dell'efficienza di trigger e la ricostruzione degli eventi.

Ricerca off-line di segnali da binarie coalescenti. Le necessità di calcolo si riferiscono essenzialmente alla ricerca di possibili candidati per sistemi binari in coalescenza; il metodo di base è quello del filtro adattato con template ad un livello prefissato dell'approssimazione post Newtoniana parametrizzata (PPN). Questo richiede un ampio spazio disco per lo storage dei template e notevoli risorse di elaborazione. Nella tabella 2 sono riportati i risultati delle necessità di calcolo previste per questa attività (P. Canitrot, L. Milano, A. Vicerè: *Computational costs for coalescent binaries detection in Virgo using matched filters*; Virgo Internal Note, VIR-NOT-PIS-1390-148-1, May, 5 2000). Vengono prese in considerazione tre differenti masse minime del sistema emittente ed una frequenza di inizio inseguimento del segnale di 20 Hz. La potenza di calcolo necessaria è espressa in 10^9 operazioni floating point al secondo (Gflops). Il numero di nodi riportato in tabella è quello di un ipotetico cluster di Pentium III ed è stato valutato assumendo un'efficienza di 1.2 Gflops per nodo. A titolo di paragone riportiamo in tabella anche una previsione fatta dal gruppo GEO.

Tabella 16 – Stima della potenza computazionale necessaria per la ricerca di segnali da stelle binarie nella fase finale della loro coalescenza.

Frequenza di inizio inseguimento del segnale: 20 Hz								
m / M_{\oplus}	Gflops Virgo	Gflops GEO	Nodes # Virgo	Nodes # GEO	GBytes Virgo	GBytes GEO	SNR recovery	event Recovery
.25	1500	1800	1250	1500	6400	48000	96.8 %	90.7 %
.5	300	360	250	300	4400	32000	96.8%	90.7 %
1	56	67	47	56	280	1980	96.8 %	90.7 %

Ricerca di segnali continui da stelle di neutroni asimmetriche in rapida rotazione. La ricerca è basata sullo sviluppo di tecniche di analisi "sub ottimali" che consentono di affrontare la ricerca in tutto cielo di sorgenti di frequenza di rotazione sconosciuta, limitando la potenza di calcolo necessaria nel range di qualche Tflops, potenza ragguardevole ma non irraggiungibile. Va sottolineato che l'appetibilità della full sky search è legata alla previsione di 10^9 stelle di neutroni galattiche previste nei modelli correnti, contro le 10^3 attualmente note. L'algoritmo di ricerca sviluppato è basato sulla trasformata di Hough. Si mappa il cielo assegnando la probabilità di falso allarme e si ritorna con passi successivi ed iterativi sulle regioni in cui la probabilità di avere una sorgente è più significativa; il costo computazionale è relativamente basso, poiché si può far uso di FFT su brevi sequenza di dati. Riducendo la probabilità di "falso allarme" si opera una selezione nello spazio dei parametri e, dopo un' opportuna operazione di "zooming", si procede a processare sequenze di dati selezionati con FFT complesse a più alta risoluzione (S. Frasca: *Hierarchical search for periodic sources*; Proceedings of the IX Marcel Grossmann Meeting, Rome 2000). A titolo di esempio si riporta il costo computazionale per un tipico schema di calcolo con strategia gerarchica basato su una opportuna separazione delle bande di frequenza nelle quali operare la ricerca.

Tabella 17 – Parametri Computazionali e Potenza di Calcolo per la ricerca di segnali continui

	Banda 1	Banda 2	Banda 3	Banda 4
Frequenza Max della banda (Hz)	2500	625	156.25	39.0625
Banda Doppler (Hz)	0.25	0.0625	0.0156	0.0039
Numero di Picchi nella banda Doppler	1643684	410921	102730	25683
Risoluzione angolare nel cielo (rad)	1.9073E-03	3.8147E-03	7.6294E-03	1.5259E-02
Numero of pixels nel cielo	3.4542E+06	8.6355E+05	2.1589E+05	5.3972E+04
Numero di frequenze indipendenti	3.1457E+06	1.5729E+06	7.8643E+05	3.9322E+05
Parametri di Spin down (solo l'ordine 1)	839	419	210	105
Numero Totale di parametri	2.8976E+09	3.6220E+08	4.5275E+07	5.6594E+06
Numero di operazioni per ogni picco	6.5884E+03	3.2942E+03	1.6471E+03	8.2355E+02
Numero totale di operazioni	2.8577E+19	8.9302E+17	2.7907E+16	8.7209E+14

Potenza di calcolo per passo (GFlops)	1.8109E+03	5.6592E+01	1.7685E+00	5.5265E-02
Potenza di calcolo complessiva (GFlops)	3.6219E+03	1.1318E+02	3.5370E+00	1.1053E-01

Il valore richiesto è di circa 3 Tflops, valore che può essere abbassato (ma purtroppo non abbattuto) con un'ottimizzazione dell'algoritmo usato (trasformata di Hough). L'ipotesi secondo cui è stato sviluppato il Computing Model di VIRGO è partita sempre dall'assunzione dell'esistenza di un Tier principale con 1 Tflop per questo tipo di ricerca, a cui si aggiungono i Tier di livello inferiore.

–*Ricerca relativa ai transienti generati da collassi stellari.* Nell'attuale fase di sviluppo la componente francese della collaborazione sta lavorando alla definizione degli algoritmi di ricerca per questa importante categoria di segnali.

– *Analisi dati di una rete di antenne.* Lo scambio dati con altri rivelatori è fondamentale per questo tipo di esperimenti; in particolare ciò è vero per i transienti brevi ed i segnali tipo chirp. Virgo scambierà dati con i due interferometri americani di LIGO e con altri rivelatori distribuiti nel mondo. Al fine di definire i protocolli di scambio sono previsti dei test già nel 2001. L'analisi in rete per la ricerca dei chirp concorrerà certamente ad aumentare le necessità di calcolo e di velocità di trasmissione dati dell'esperimento.

5.2.3 Il Tier-1 di Virgo-Italia e le necessità complessive

Come è stato riportato sopra il laboratorio VIRGO di Cascina è il luogo di produzione dei dati ed il Tier 1 italiano deve svolgere funzione di archiviazione dei "Raw Data" ed avere sufficiente potenza di calcolo per coprire le necessità che qui riassumiamo in due tabelle, ipotizzando anche la distribuzione sui siti di livello inferiore. Vista la situazione di fluidità dello stato del calcolo di Virgo, ma anche le necessità ormai stringenti per centrare in tempo le milestone che Virgo si è assegnato, riteniamo utile riportare qui di seguito un quadro complessivo includente la somma delle risorse del sistema di Tiers e quindi il quadro relativo al Tier 1.

Quadro complessivo delle risorse di calcolo off-line di Virgo – Italia

<u>—units</u>	<u>end 2001</u>	<u>end 2003</u>
—CPU capacity*	80 kSI2000 (350 Gflops)	800 kSI2000 (3.5Tflops)
—disk capacity	TBytes 10	100
—disk I/O rate	GBytes/sec 5	5

* [comprensivo delle risorse di tutti i laboratori (Cascina, Firenze/Urbino, Napoli, Perugia, Pisa, Roma1)]

Quindi, ipotizzando lo scenario in cui la richiesta in Commissione II sia approvata (la richiesta prevedeva strutture di calcolo per Napoli e a Roma per un totale di 150 Gflops), le necessità di Virgo per il Tier 1 sono qui sintetizzate

Quadro di sviluppo del Tier 1 di Virgo Italia

<u>—units</u>	<u>end 2001</u>	<u>end 2003</u>
—CPU capacity*	46 kSI2000 (200 Gflops)	300 kSI2000 (1.3Tflops)
—disk capacity	TBytes 8	100
—disk I/O rate	GBytes/sec 5	5

5.2.4 Risorse Umane

Una valutazione di ordine zero della necessità di risorse umane per lo start-up di questa struttura di Tier 1 è di: 4 unità uomo di operatore e 2 tecnologi di livello più alto.

6 Considerazioni Generali

Nel prosieguo di questo documento si cercherà di delineare un progetto di massima per la creazione di un Centro Regionale che serva agli esperimenti ad LHC, BaBar e VIRGO.

L'ipotesi iniziale che verrà presa in considerazione è quella di realizzare un unico centro in Bologna presso il Centro Nazionale CNAF dell'INFN. Questo Centro Sperimentale sarà in coordinamento con le sedi che sperimentano i prototipi dei Tier 2 all'interno dell'INFN e con il Tier 0 del CERN ed i Tier 1 delle altre nazioni che per il momento sono individuati a Fermilab (USA/CMS), Brookhaven (USA/ATLAS), Lione (FR), Rutherford (UK). Per quel che riguarda BaBar e VIRGO il coordinamento sarà con il centro di IN2P3 a Lione per entrambi, e con SLAC e Cascina rispettivamente, nonché con altri centri Tier A di BaBar o di secondo livello per entrambi. La realizzazione di tale centro ed il suo progetto dettagliato sono soggetti ad alcune condizioni di cui bisogna tener conto:

- § è necessaria, almeno per quel che riguarda gli esperimenti ad LHC e VIRGO, una fase di sperimentazione per valutare le soluzioni tecniche ed organizzative proposte e verificarne la validità insieme con gli utilizzatori (esperimenti);
- § la realizzazione finale per il Tier 1 di LHC che, secondo quanto proposto dalla LHC Computing Review, potrebbe essere diviso negli anni dal 2005 al 2007 secondo una proporzione 30%, 30%, 40%, necessita di una revisione del progetto nell'intorno dell'anno critico 2004 quando, finita la fase sperimentale, dovrà essere aggiornata, sia la natura tecnico-economica del progetto, sia quella organizzativa e realizzativa per tener conto dell'evoluzione tecnologica intervenuta e dei risultati della sperimentazione effettuata.
- § per BaBar è critico realizzare un "servizio in produzione" entro i tempi delineati in questo documento, pena la decadenza dell'utilità stessa del Tier A.

Il Progetto è pertanto diviso in due fasi parzialmente sovrapposte:

- 1) Fase I (2001–2004) in cui:
 - verranno sperimentate le tecnologie, l'organizzazione, verrà formato il personale necessario e
 - sarà avviata, previa individuazione degli spazi e delle risorse necessarie, una prima fase di servizio per BaBar e VIRGO a partire dal 2003.
- 2) Fase II (2004–2007) in cui sarà rivisto e definito il piano di realizzazione del Centro per LHC e sarà effettuata la sua installazione finale per gli esperimenti ad LHC.

La prima fase includerà la utilizzazione di un "Servizio" da parte di altri esperimenti non-LHC che potranno fornire un banco di prova per l'organizzazione e la sperimentazione dei servizi del Centro.

6.1 Gestione del Centro

Uno dei temi più importanti riguarda l'organizzazione della gestione del Centro, in maniera tale da garantire agli esperimenti i servizi richiesti. Le risorse di calcolo necessarie, sono parte integrante dei piani di calcolo degli esperimenti che saranno soggetti all'approvazione delle competenti Commissioni Scientifiche Nazionali.

Il Centro dovrà fornire adeguati strumenti amministrativi ed organizzativi per far sì che gli esperimenti siano certi delle risorse a loro allocate e ne possano controllare l'utilizzo.

L'ipotesi più semplice è che vengano formati due comitati di indirizzo dell'attività del Centro: uno con caratteristiche politico-amministrative ed uno con caratteristiche tecnico-organizzative.

- § Comitato di Direzione Scientifica del Centro Regionale: è composto dal direttore del CNAF e da un rappresentante per ogni esperimento che abbia presso il Centro un suo Tier 1 o Tier A (ad esempio nella prima fase: ALICE, ATLAS, BaBar, CMS, LHCb, VIRGO). Questo comitato avrebbe le funzioni di "consiglio di amministrazione" del Centro con potere

obbligatorio sulla allocazione delle risorse e sull'indirizzamento della spesa relativamente alle attività del Centro stesso, in armonia con gli obiettivi e le priorità dell' INFN.

§ Comitato Tecnico di Gestione del Centro Regionale: dovrebbe essere composto dal Responsabile Tecnico del Centro Regionale e da un rappresentante presso il Centro per ogni esperimento che ivi abbia un suo Tier 1 o Tier A. Il Comitato Tecnico avrebbe la funzione di proporre le soluzioni tecniche da utilizzare all'interno del Centro e validare quelle già in uso, di facilitare la gestione quotidiana, di preparare un rapporto trimestrale di utilizzo delle risorse e di elaborare, in collaborazione con il Comitato di Direzione Scientifica un piano annuale di allocazione delle risorse. I rapporti e i piani dovranno essere approvati dal Comitato di Direzione Scientifica.

§ Gli esperimenti decideranno, autonomamente, chi designare a rappresentare l'attività di Tier1, a loro relativa, nei loro eventuali comitati internazionali di coordinamento.

Per garantire una migliore integrazione delle problematiche degli esperimenti nell'attività del Centro, è opportuno che personale degli esperimenti partecipi alle attività del Centro stesso in loco con adeguati meccanismi di trasferimento e congedo o con posizioni temporanee, secondo la necessità.

Tutte queste soluzioni saranno avviate, in fase sperimentale, già dal primo anno della Fase I e definite con precisione nel corso della stessa.

6.2 Spazi ed Infrastrutture al CNAF

Gli attuali spazi ed infrastrutture disponibili al CNAF non sono sufficienti, come vedremo, per realizzare la Fase I nella sua completa attuazione (2004). Le attuali infrastrutture sono, al più, utilizzabili per il primo anno (2001) in cui solo una modesta parte delle installazioni verrà provata. Uno dei fattori critici del piano proposto è quindi il reperimento degli spazi adeguati, sia per la sala macchine, sia per il personale, anche proveniente dall'esterno del CNAF, che dovrà essere coinvolto nell'attività del Centro. È in fase avanzata una richiesta di spazi presso il Rettore dell'Università di Bologna che si è riservato di dare una risposta in tempi brevi.

Questa parte critica dovrà essere necessariamente superata e risolta entro il primo anno (2001) del piano pena una revisione totale della tempistica e degli obiettivi previsti.

A tal fine è prevista una prima importante milestone nel 2001 su questo tema.

6.3 Personale del CNAF

Altro fattore critico per la riuscita del piano è il reperimento di personale adeguato, per numero e preparazione, all'attività prevista. L'attuale pianta organica del CNAF prevede 17 posti in organico la cui distribuzione per tipologia è descritta nella tabella seguente. E' anche da precisare che detto personale, non essendo coinvolto al momento in alcun esperimento, non ha la preparazione e l'esperienza necessaria per svolgere, nell'immediato, compiti di supporto per software specifico di esperimento.

Tabella 18 – Organico Attuale del CNAF.

N.	Tipo	Situazione
1	Tecnologi Liv. I	
3	Tecnologi Liv. II	1 in Congedo
5	Tecnologi Liv. III	1 Bandito ed 1 da Bandire
4	CTER	1 da Bandire
3	Coll. Amministrazione	2 da Bandire
1	Funz. Amministrazione	Da ricoprire
17	Totale Posti in Organico	

Una parte del personale è dedicata ad attività di Sperimentazione e Ricerca, sia nel campo delle Reti, che, in generale, in Informatica applicata alle attività dell'INFN. Tale personale, attualmente

impegnato nei progetti di GRID, non potrà essere coinvolto, in maniera apprezzabile, nella prima fase del presente progetto. Anche dopo la Fase II, non è assolutamente pensabile di far coincidere tutta l'attività svolta al CNAF con quella del Centro Regionale e risulta difficile valutare quanta parte di sperimentazione sarà comunque fatta nell'interesse del CR stesso e quanta attività del CR sarà, a sua volta, utile per la sperimentazione.

E' comunque evidente che il CNAF, per continuare ad essere efficace come CR, dovrà mantenere un'adeguata, sia per quantità che per qualità, attività di ricerca e sperimentazione nel campo informatico. L'esperienza e le competenze del personale coinvolto nel servizio di sperimentazione e ricerca, sarà di fondamentale importanza per l'evoluzione del CR, perché essa verrà riversata principalmente nell'attività del centro stesso e quindi esso ne sarà il principale beneficiario.

L'esempio attuale è quello dell'attività di GRID che si propone di sviluppare software e tools che saranno di primaria importanza per il funzionamento del CR come Tier1 in collaborazione con gli altri Centri Regionali (Tier 0, Tier1 e Tier2) sparsi per il mondo. GRID risulta quindi funzionale all'attività del CR e la completa garantendo competenze e risorse non facilmente reperibili in Italia e all'estero.

Ciò premesso, una ragionevole stima del personale che potrà, essere coinvolto praticamente full-time sull'attività del CR è la seguente:

Tabella 19 – Organico del CNAF che potrebbe essere coinvolto nel Centro Regionale.

N.	Tipo
1	Tecnologi Liv. II
2	Tecnologi Liv. III
1	CTER
1	Coll. Amministrazione
5	Totale Posti in Organico

Le altre risorse di personale dovranno quindi necessariamente venire con posizioni non permanenti ed, in particolare, le soluzioni prevedibili e che saranno sperimentate durante la Fase I sono:

- § Contratti temporanei (art. 23, art. 15 CCNL ed art. 2222 C.C.)
- § Outsourcing, per i profili professionali più semplici, da ditte in grado di fornire personale formato.
- § Lavoro interinale, per alcune tipologie di personale che non fossero reperibili con i meccanismi precedenti.
- § Contratti di Formazione: Borse di Studio, Assegni di Ricerca Tecnologica per giovani laureati da formare.

Tutti i suddetti meccanismi saranno esaminati approfonditamente e, ad ogni reclutamento, sarà proposto quello più adeguato.

La definizione con maggiore dettaglio delle necessità di personale e del suo reclutamento nei vari profili è oggetto di una milestone nel corso del 2001.

7 Fase I – Tier A per BaBar, Tier 1 per VIRGO e Sperimentazione del Tier 1 unico per LHC.

7.1 Risorse Hardware

La quantificazione indicativa delle risorse hardware richieste per esperimento, nel Centro Regionale in ogni anno per il raggiungimento degli scopi previsti è presentata nella tabella seguente. Il piano dettagliato, per esperimento e per anno, delle risorse installate ed il relativo piano temporale degli acquisti, di questa prima fase, è oggetto di uno specifico deliverable (documento D1.3).

Tabella 20 – Previsione indicativa delle risorse installate per la Fase I*

Esperimento	Risorsa	Anno		
		2001	2002	2003
BaBar	CPU (kSI2000)	10	30 (20)	460 (430)
	DISK (TB)	0.5	10 (9.5)	50 (40)
	TAPES (TB)	1	20 (19)	580 (560)
VIRGO	CPU (kSI2000)	10	50 (40)	350 (300)
	DISK (TB)	1	8 (7)	108 (100)
	TAPES (TB)	1	40 (39)	200 (160)
ALICE	CPU (kSI2000)	20	60 (40)	150 (90)
	DISK (TB)	2	6 (4)	14 (8)
	TAPES (TB)	1	10 (9)	20 (10)
ATLAS	CPU (kSI2000)	5	25 (20)	125 (100)
	DISK (TB)	0.5	5 (4.5)	15 (10)
	TAPES (TB)	1	15 (14)	30 (15)
CMS	CPU (kSI2000)	20	50 (30)	100 (50)
	DISK (TB)	7	20 (13)	40 (20)
	TAPES (TB)	17	50 (33)	100 (50)
LHCb	CPU (kSI2000)	18	71 (53)	159 (88)
	DISK (TB)	2	6 (4)	12 (6)
	TAPES (TB)	13	26 (13)	39 (13)

La quantificazione di ATLAS nel 2001 sconta già l'impossibilità pratica di ottenere nel 2001 un servizio consistente dal CNAF; per ATLAS è però importante che le risorse richieste nel 2001 siano disponibili a partire almeno da inizio luglio e quelle 2002 a febbraio.

7.2 Rete

Per la fase di sperimentazione si utilizzeranno sulla rete locale delle tecnologie di tipo Gigabit Ethernet e sue evoluzioni. Per la rete geografica sarà effettuata la sperimentazione sul Pilot di GARR-G e sulla infrastruttura GARR-B per quelle sedi che non sono coinvolte nel Pilot GARR-G.

VIRGO richiede l'immediata messa in funzione di una connessione di rete dedicata con Cascina (connessione punto-punto a 50 Mbit/s) e l'alta efficienza di connessione con le Sezioni coinvolte nel progetto a partire dal 2002.

La rete sarà anche di fondamentale importanza per il servizio di Tier A di BaBar a partire dalla seconda metà del 2002 in cui partirà una fase intensa di commissioning con inizio di trasferimento di dati da SLAC. Per tali trasferimenti è prevedibile l'uso di una banda dedicata equivalente almeno a quella prevista per VIRGO (50 Mbit/s) di tipo altamente affidabile e ridondata su circuiti separati. Il Tier 1 unico per LHC richiederà, a regime (2005) bande dell'ordine di alcuni Gigabit/s, mentre per la fase iniziale di sperimentazione dovrebbe essere sufficiente una banda garantita fra 155 e 622 Mbit/s nel periodo 2001-2004.

In totale queste esigenze portano a ritenere indispensabile, per le sole necessità del Centro, un attacco a 155 Mbit/s con garanzia di continuità di banda, senza strozzature, almeno verso Cascina, Lione, CERN e SLAC già a partire dal 2001 e, successivamente, un raddoppio almeno ogni anno della banda d'accesso.

* I numeri fra parentesi sono le risorse da installare nell'anno in questione

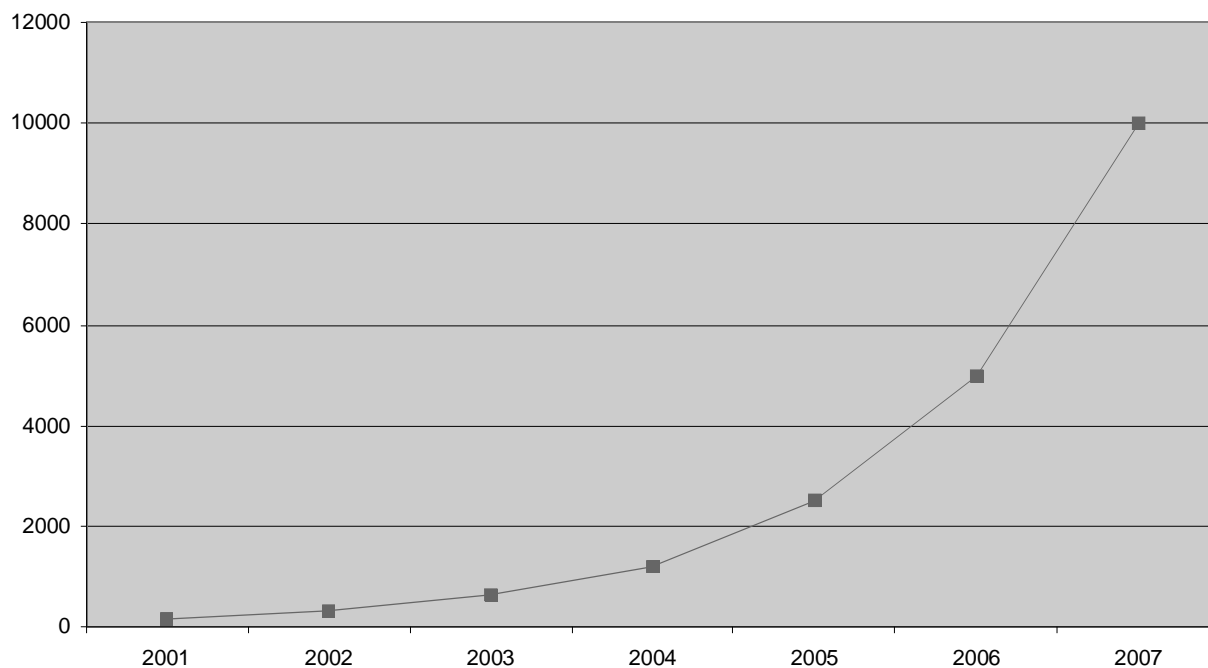


Figura 1 – Velocità Minima di Accesso al Centro negli anni

7.3 Personale

La parte più delicata del progetto riguarda la necessità di personale per avviare, gestire e mantenere costantemente aggiornato il centro. Le esigenze degli esperimenti, qui di seguito riassunte non sono molto diverse od incompatibili, ma richiedono una frazione di personale a loro dedicata.

Le persone che saranno "dedicate" agli esperimenti, dovranno essere Tecnologi o Ricercatori con profilo orientato al calcolo ed, in prima istanza, verranno ricercati fra il personale che già partecipa agli esperimenti e, solo nel caso che non sia possibile reperirne, sarà reclutato con contratti a termine opportuni e formato dall'esperimento.

BaBar prevede una necessità di almeno 2 persone con competenze specifiche sul software dell'esperimento ed un numero di tecnici addetti ai servizi di base fra 5 e 10 variabile in funzione del tipo di supporto e livello di servizio si vorrebbe garantire. Il supporto a SLAC prevede che, in caso di problemi relativi ai servizi di calcolo, gli utenti abbiano la possibilità di chiamare un numero telefonico e lasciare un messaggio e poi, generalmente entro un quarto d'ora, vengano richiamati da una persona il turno nei giorni lavorativi e on-call di notte, nei sabati e nei festivi.

VIRGO ritiene necessari circa 4 operatori e 2 Tecnologi di alto profilo.

LHC: per gli esperimenti ad LHC viene nel seguito fatta la distinzione fra personale con competenze specifiche sul s/w del singolo esperimento e quello con altri profili.

Per gli altri profili la valutazione dettagliata sarà fatta nel documento previsto come deliverable D1.2. Una prima valutazione per la Fase I è comunque fornita nella Tabella 22, dove il personale si intende già operativo e non sono quindi compresi gli eventuali periodi di formazione ove richiesti. Nella tabella seguente vengono invece sintetizzate le necessità espresse, a regime, dagli esperimenti.

Tabella 21 – Personale richiesto dagli esperimenti presso il Centro Regionale

Esperimento	N	Descrizione
BaBar	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
	8	Tecnici di supporto dei servizi di base

VIRGO	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
	4	Tecnici Operatori/Supporto Servizi di base
ATLAS	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
ALICE	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
CMS	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
LHCb	2	Ricercatori/Tecnologi esperti sul software dell'esperimento
LHC Totali	15	Tecnici di supporto dei servizi di base
Totali	12	Ricercatori/Tecnologi esperti sul software dell'esperimento
	27	Tecnici di supporto dei servizi di base

Tabella 22 – Personale necessario per i primi tre anni

<i>Descrizione Compiti</i>	Unità	2001	2002	2003	
Responsabile del Centro e Vice	Person e	1	2	2	
Collaboratori di Amministrazione		2	3	3	
Supporto Software Esperimenti ¹		6	10	14	
Gestione Farm e planning	FTE	1	4	3	
DataBase e Data Management		1	3	3	
Active Mass Storage		0	2	3	
Import/Export Dati		1	2	2	
Sistemisti ²		2	4	4	
Reti e Sicurezze		1	2	2	
Servizi Generali e Infrastrutture		1	1	1	
Operatori (+Backup e LAN) ³		0	3	7	
<i>Totali Persone</i>			9	15	19
<i>Totali FTE</i>			7	21	25

7.4 Definizione del piano

Il piano di Fase I può essere quindi sintetizzato nella realizzazione nel periodo 2001–2004 del prototipo di Centro Regionale Tier 1 unico per gli esperimenti LHC, secondo le necessità elencate in precedenza.

La realizzazione nel Centro di un Tier1 per Virgo ha alte probabilità di poter essere integrato nella realizzazione precedente se le infrastrutture verranno terminate nella prima parte del 2002.

La realizzazione nel Centro di un Tier A per BaBar è invece molto condizionata sia da elementi di complessità non presenti negli esperimenti LHC in fase I (mass-storage, impiego di Objectivity su larga scala, apertura del centro a tutta la collaborazione, ecc.) sia dalla criticità dei tempi richiesti che impongono la partenza immediata dell'attività di progetto. Le difficoltà sono esacerbate dalla scarsità di personale, con il profilo e l'esperienza adeguate, disponibile sia all'interno della collaborazione italiana che al CNAF, e dai tempi necessari a reperire e formare nuovo personale.

Al livello di dettaglio a cui è arrivata la presente analisi non si possono trarre conclusioni definitive, ma è probabile che l'impegno richiesto per BaBar, prefigurandosi come il più oneroso per il centro nei primi due anni, possa non risultare compatibile con la totalità dei servizi verosimilmente richiesti dagli altri esperimenti.

Con queste premesse, il piano di realizzazione della Fase I diventa il seguente.

¹ Questo è il personale Tecnologo di cui alla tabella precedente

² Personale in Outsourcing

³

Nella Tabella seguente è elencato il materiale che deve essere installato in ogni anno, a partire dall'anno in corso, (2001) fino al 2003. L'impegno finanziario potrà, in parte, riguardare l'anno finanziario precedente.

Tabella 23 – Materiale da installarsi per anno nella Fase I

Ann o	CPU (kSI2000)	DISK (TB)	TAPES (PB)
2001	83	13	0.03
2002	200	42	0.13
2003	1050	184	0.9

Le risorse previste per l'acquisto ogni anno sono un ragionevole compromesso tra le esigenze espresse dagli esperimenti e tengono conto di diversi fattori che possono permettere una implementazione inferiore alle richieste. Alcuni fattori sono: l'economia di scala delle risorse, la capacità (da sperimentare) di utilizzo congiunto da parte degli esperimenti con un aumento dell'efficienza (partizione e condivisione dinamica delle risorse), la possibilità di programmare le attività cercando di evitare la sovrapposizione temporale (in particolare per Virgo ed LHC), etc. Le necessità di spazi riguardano le seguenti quantità di materiale corrispondente alla tabella risorse precedente.

Tabella 24 – Spazi necessari nella Fase I

	Racks CPU	Racks Disco	Robot Nastri	Racks Rete
Anno				
2001	2	3	1 x 6000 slots	1
2002	7	6	1 x 6000 slots	0
2003	9	9	1 x 6000 slots	0
	18	18	3 x 20 mq	1

Per tali attrezzature basta un'area fra 140 e 200 mq di spazi attrezzati. A questi spazi vanno sommati quelli per il gruppo elettrogeno e l'UPS.

7.5 Scala dei Tempi

La scala dei tempi è desumibile dalle esigenze degli esperimenti prospettate nella prima parte del documento. Allo stesso tempo, dati i vincoli dettati dalla tempistica molto stretta (in particolare per Babar e Virgo), le milestones sono necessariamente commisurate ad una ragionevole realizzabilità degli obiettivi.

I deliverables riguardano, sia documenti che, passo passo, definiranno nel dettaglio la parte realizzativa del progetto, sia installazioni e collaudi, sia l'avvio di servizi per gli esperimenti.

Tabella 25 – Milestones della Fase I

M	Data	Descrizione Milestone
1.1	15/05/2001	Definizione dei Gruppi di Progetto Tecnico Esecutivo e Gestione
1.2	01/06/2001	Decisione sulla realizzazione del Tier A Italiano di BaBar Definizione modello di Calcolo di VIRGO e del Tier 1 Italiano Individuazione degli spazi necessari e delle risorse di personale
1.3	01/11/2001	Inizio Realizzazione dei lavori di infrastruttura di base.
1.4	01/01/2002	Inizio preparazione del Tier 1 INFN come parte del Progetto Prototipo di calcolo distribuito per LHC
1.5	01/03/2002	Inizio installazione dell'hardware della Fase I secondo la tabella prevista.
1.6	01/06/2002	Inizio commissioning dell'attrezzatura.

1.7	01/01/2003	Fase operativa prototipizzazione Tier 1 LHC
1.8	01/02/2003	Partenza del servizio per BaBar
1.9	01/06/2003	Partenza del servizio per VIRGO
1.1	01/06/2004	Revisione Tecnico–Economico–Organizzativa e fine della Fase I
0		

Tabella 26 – Deliverables della Fase I

D	Data	Tipo	Descrizione
1.1	30/06/2001	Documento	Piano della realizzazione delle infrastrutture di base
1.2	30/06/2001	Documento	Piano di acquisizione e formazione del personale
1.3	30/06/2001	Documento	Piano dettagliato degli acquisti di materiale
1.4	01/07/2001	Hardware	20% di hardware 2001 installato e disponibile per gli esperimenti LHC.
1.5	01/10/2001	Hardware	Hardware 2001 installato e disponibile per gli esperimenti LHC.
1.6	01/10/2001	Capitolato	Capitolato delle infrastrutture hardware
1.7	01/01/2002	Prototipo	Tier 1 unico INFN prototipale per LHC
1.8	01/06/2002	Hardware	Hardware della Fase I /2002 Installato.
1.9	01/12/2002	Hardware	Hardware per 2003 installato
1.10	01/01/2003	Servizio	Servizio per Data Challenge esperimenti LHC
1.11	01/02/2003	Servizio	Servizio per BaBar
1.12	01/06/2003	Servizio	Servizio per VIRGO
1.13	01/03/2004	Documento	Primo rendiconto sull'attività svolta
1.14	01/06/2004	Documento	Revisione Tecnico–Economico–Organizzativa del Progetto

8 Fase II di Realizzazione Tier 1 unico per LHC

Il piano dettagliato per la seconda fase del progetto dovrà necessariamente essere completato dopo il superamento, con successo, della fase I. Pertanto la definizione di questa fase è del tutto schematica e non dettagliata.

In particolare nel 2004 sarà effettuata una revisione che riguarderà principalmente tre argomenti:

1. **Tecnica:** Le attuali previsioni sono basate sulla prevedibile evoluzione della tecnologia attuale; tali previsioni non possono, però, essere considerate sufficientemente affidabili su periodi di tempo che eccedano un paio d'anni. Si rende, pertanto, necessaria una revisione delle tecnologie da adottare per la seconda fase al termine della prima fase.
2. **Economica:** Sarà necessario valutare l'impatto economico della revisione tecnologica citata al punto precedente. L'attuale stima dei costi, basata sulla legge di Moore^[5] è soggetta ad incertezze anche piuttosto grandi (40%).
3. **Organizzativa:** La prima fase servirà anche a provare l'organizzazione del Centro Regionale, le procedure necessarie e la funzionalità ed efficacia degli organi di controllo e decisionali proposti.

La descrizione sintetica delle milestones e dei deliverables per la Fase II è elencata nelle tabelle seguenti.

Tabella 27 – Milestones della Fase II

M	Data	Descrizione Milestone
2.1	01/06/2004	Definizione Tecnico–Economico–Organizzativa della Fase II
2.2	01/11/2004	Decisione sulla realizzazione del Tier 1 unico Italiano per LHC

2.3	01/12/2004	Inizio eventuali modifiche alla infrastruttura di base.
2.4	01/01/2005	Inizio installazione dell'hardware della Fase II secondo la tabella prevista.
2.5	01/06/2005	Fine commissioning della prima tranche dell'attrezzatura.
2.6	01/02/2006	Installazione della seconda parte dell'attrezzatura.
2.7	01/06/2006	Inizio commissioning della seconda tranche dell'attrezzatura.
2.8	01/03/2007	Installazione della terza tranche dell'attrezzatura
2.9	01/06/2007	Commissioning del sistema finale
2.1 0	31/12/2007	Fine della Fase II

Tabella 28 – Deliverables della Fase II

D	Data	Tipo	Descrizione
2.1	30/06/2004	Documento	Piano della realizzazione delle Fase II
2.2	15/11/2004	Documento	Piano dettagliato degli acquisti di materiale
2.3	01/03/2005	Hardware	30% Hardware della Fase II Installato.
2.4	01/03/2006	Hardware	60% Hardware della Fase II Installato.
2.5	01/03/2007	Hardware	100% Hardware della Fase II Installato.
2.6	01/06/2007	Sistema	Sistema Tier 1 Unico completo in produzione
2.7	31/12/2007	Documento	Rendiconto sull'attività svolta
2.8	31/12/2007	Documento	Piano di mantenimento ed aggiornamento del Centro

9 Costi

La tabella seguente riassume la distribuzione dei costi previsti, per tipo e per anno, ed è puramente indicativa. Una valutazione dettagliata dei costi per la Fase I sarà oggetto dei documenti D1.1, D1.2 e D1.3 previsti come deliverables. Per la Fase II la valutazione dettagliata dei costi verrà effettuata nel corso del 2004 e sarà argomento dei documenti D2.1 e D2.2 previsti come deliverables.

Tabella 29 – Previsione indicativa dei costi del Progetto

Item	2001	2002	2003	2004	2005	2006	2007	Total
FARM	395,000	1,338,000	1,801,000	2,500,000	1,872,000	1,886,000	1,182,000	10,974,000
DISCHI	741,000	1,622,000	2,474,000	1,990,000	1,137,000	927,000	464,000	9,355,000
ROBOT	250,000	900,000	925,000	225,000	925,000	550,000	950,000	4,725,000
ALTRO	2,111,000	1,811,000	1,558,000	1,605,300	1,659,600	1,848,500	1,848,500	12,441,900
Totali	3,497,000	5,671,000	6,758,000	6,320,300	5,593,600	5,211,500	4,444,500	37,495,900

La voce "ALTRO" contiene sia costi indicativi di materiale inventariabile, di consumo, di rete e di contratti per servizi in outsourcing non contemplati nelle altre voci. Sono esclusi i costi di personale dipendente e di accesso alla rete GARR.

L'ammontare indicativo, necessario per la Fase I (2001–2003), è quindi di circa 16 M, pari a circa 31 Miliardi di Lire. La previsione, molto preliminare, per la Fase II (2004–2007) è, invece, di circa 22 M, pari a circa 42 Miliardi di Lire.

10 Conclusioni

Abbiamo analizzato l'ipotesi di costituzione di un unico Centro Regionale di Calcolo di primo livello per gli esperimenti Babar, Virgo ed a LHC.

Alla luce delle necessità e delle scale dei tempi degli esperimenti considerati concludiamo che la costituzione del centro è fattibile purché siano verificate le condizioni critiche di:

- reperimento degli spazi necessari;
- coinvolgimento del personale specializzato
- disponibilità per il progetto del personale esperto già in servizio;
- possibilità di acquisire, come previsto, con contratti a tempo, il restante personale necessario;

entro le date indicate nel documento, perché il progetto non venga vanificato o comunque subisca un radicale ridimensionamento degli obiettivi.

Va anche sottolineato che onde avere una efficace partecipazione italiana al Data Challenge per LHC è altrettanto importante il rispetto de tempi ipotizzati nel documento.

La costituzione del Centro consentirà di sfruttare al meglio le potenziali sinergie consentendo un efficiente sfruttamento delle risorse di calcolo che andranno installate nel sito.

La sede ipotizzata per la costituzione del Tier unico di primo livello è il CNAF di Bologna dove risulta possibile coinvolgere alcune unità di personale pur nel rispetto degli impegni scientifici già assunti da questa sede.

Sono state evidenziate tutte le fonti principali di spesa con una valutazione di massima destinata ad essere maggiormente dettagliata nel futuro. L'entità dell'investimento richiesto è importante, ma esso è commisurato alla complessità dei problemi di calcolo e alle responsabilità che la comunità dei ricercatori italiani si assume nel contesto di questi esperimenti a collaborazione internazionale.

Concludiamo sottolineando di nuovo l'urgenza della decisione sulla realizzazione di un tal progetto, dettata dall'esigua contingenza temporale che deriva dagli impegni internazionali degli esperimenti.

11 Referenze

[1]MONARC Phase 2 final report:

<http://monarc.web.cern.ch/MONARC/docs/phase2report/Phase2Report.pdf>

[2]DATAGrid Project: <http://www.cern.ch/Grid/>

[3]Report of the Steering Group of the LHC Computing Review: CERN/LHCC/2001-004, CERN/RRB-D 2001-3, 22 February 2001.

[4]<http://www.specbench.org/>

[5]Gordon E. Moore Intel co-founder: more information can be found in the following URL:

<http://www.sciam.com/interview/moore/092297moore2.html>

[6]GEANT: <http://www.dante.net/geant.html>