

I Centri Regionali di LHC ed il TIER 1 unico

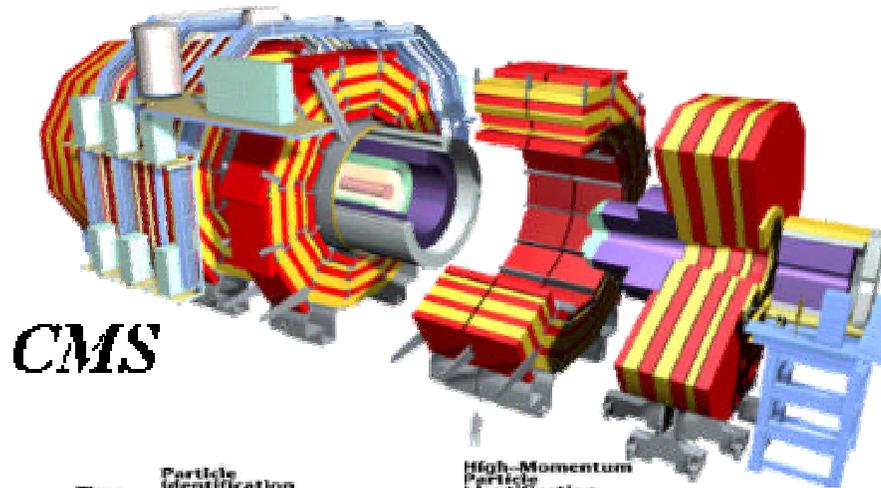
F. Ruggieri

IV Incontro di GARR-B

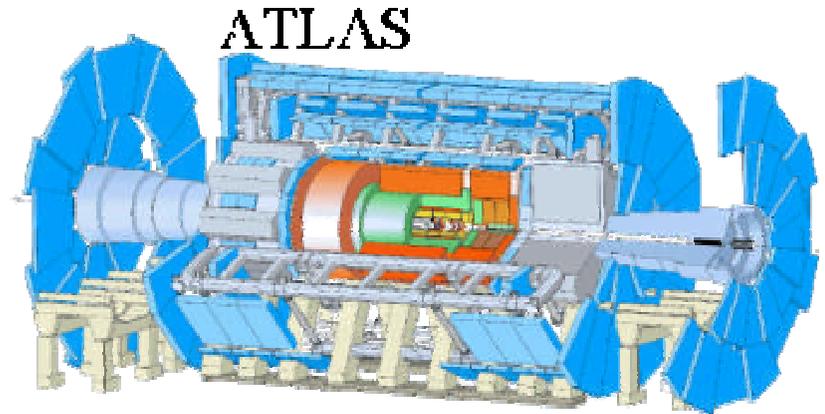
Bologna 24-26 Giugno 2002

Esperimenti a LHC

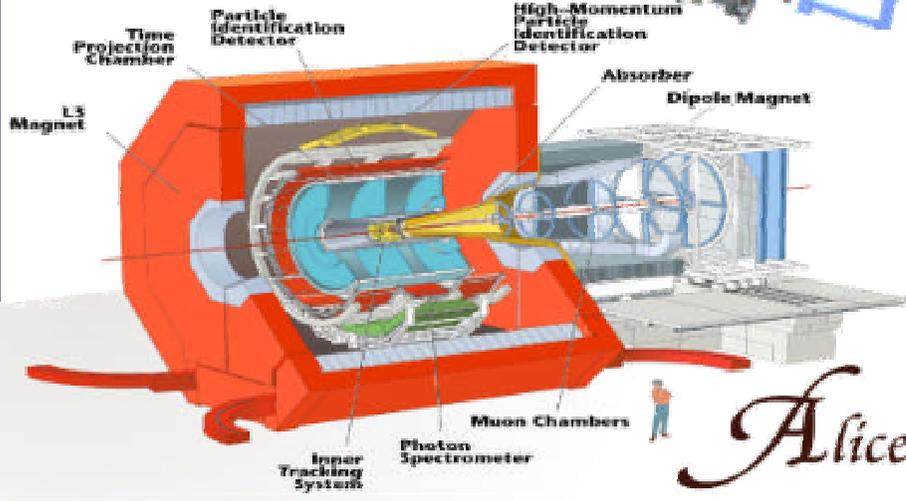
3.5 PetaBytes / year
 $\sim 4 \times 10^9$ events/year



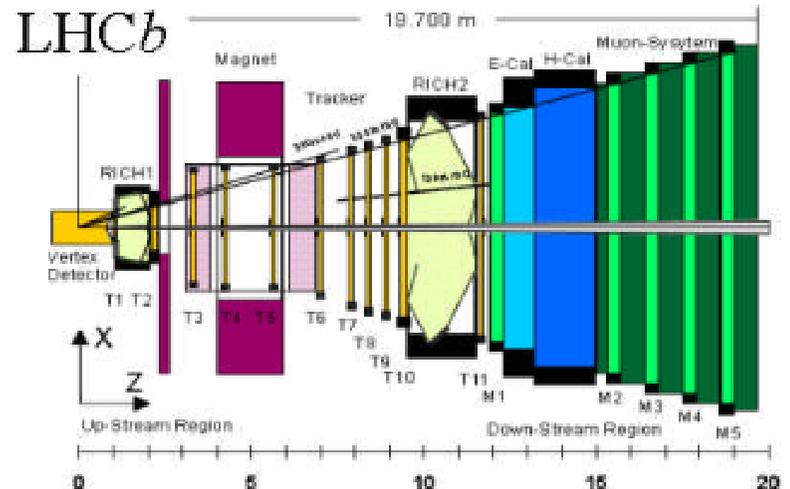
CMS



ATLAS



ALICE



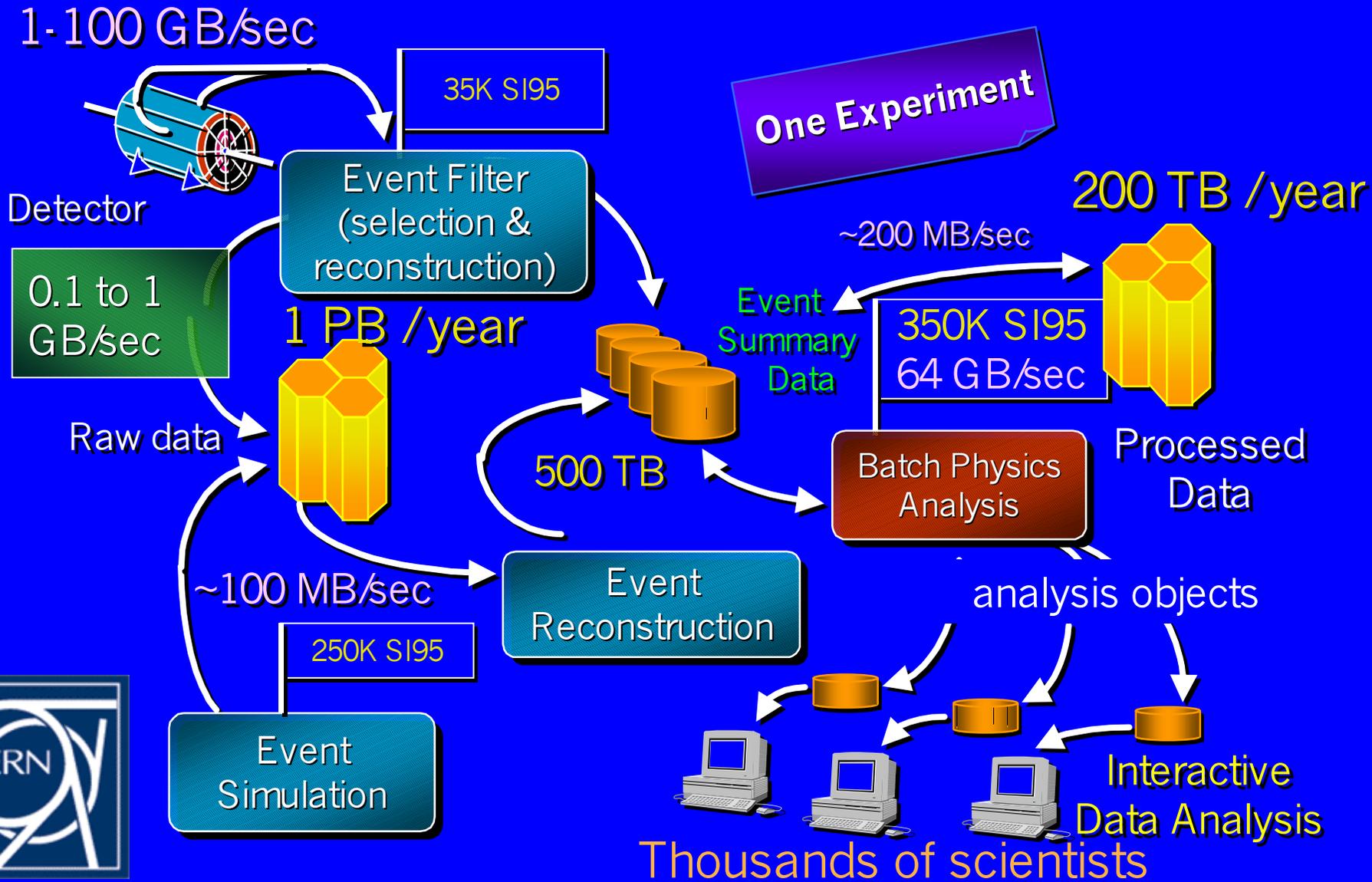
LHCb

19.700 m

0 5 10 15 20

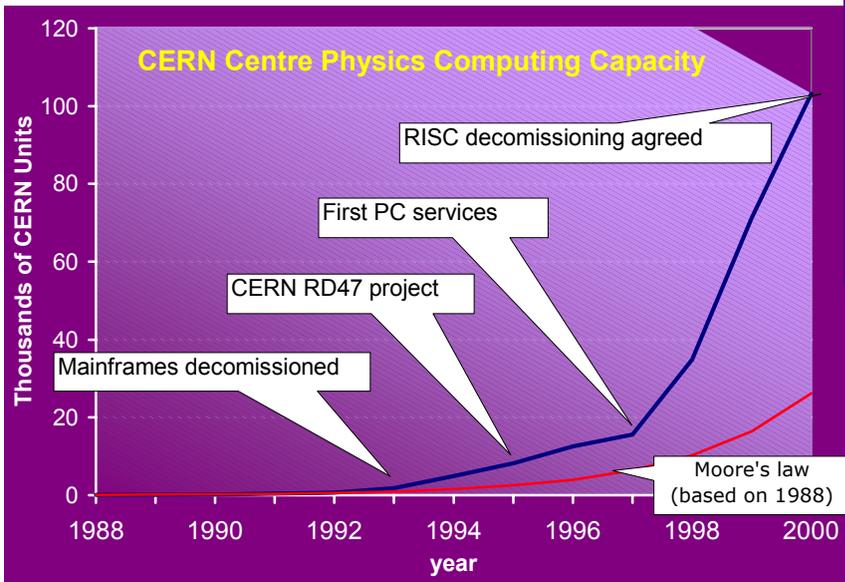
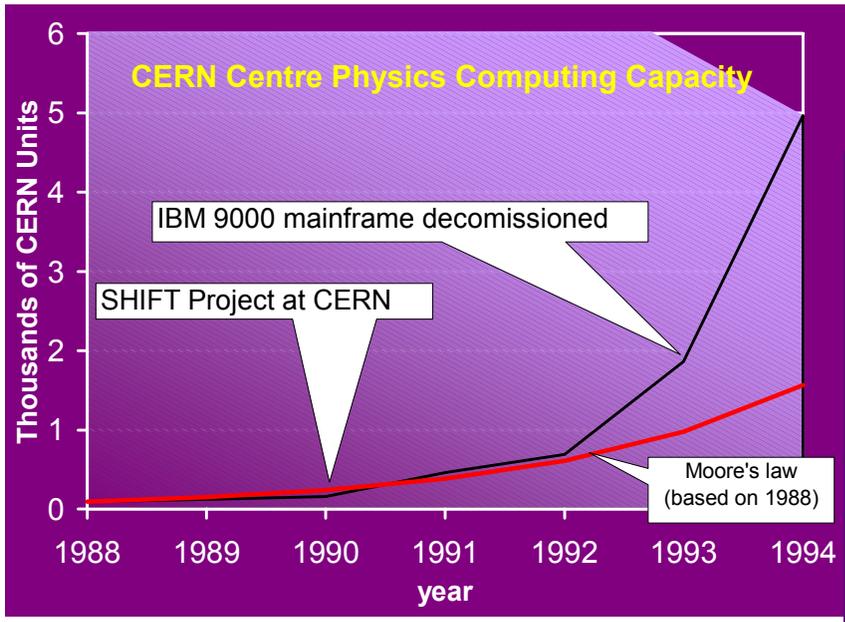


We have a very big job ahead of us... as in the past, together we will make it.

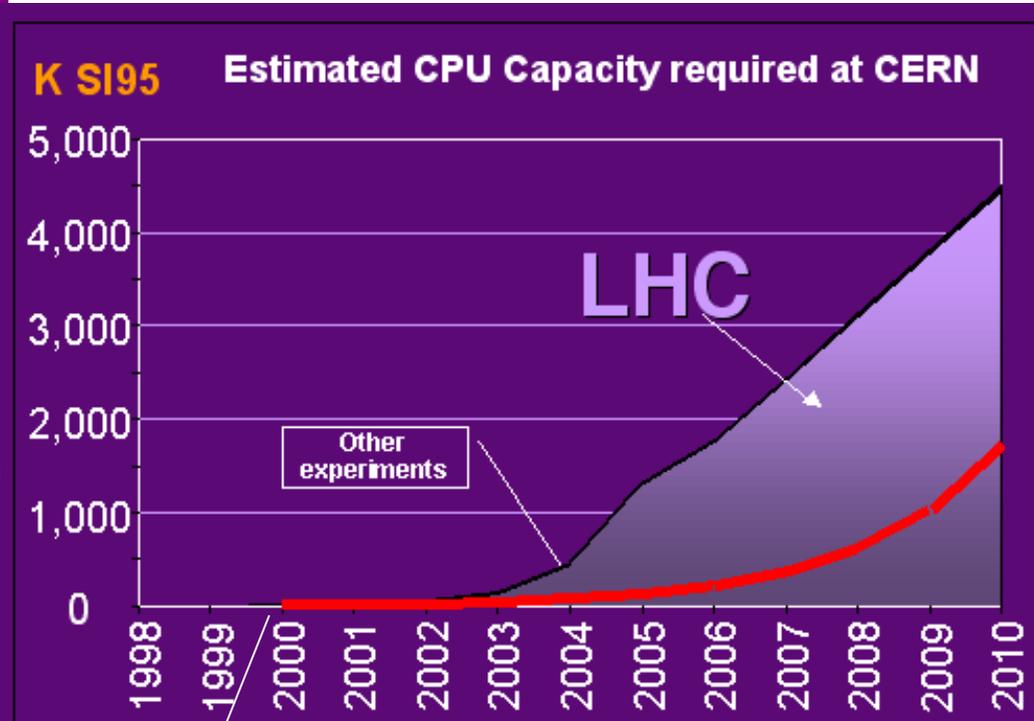


Evolution of Computing Capacity at CERN

LEP



LHC



Jan 2000:
3.5K SI95

Moore's law – some measure of the capacity technology advances provide for a constant number of processors or investment (based on 2000)

HTC piuttosto che HPC

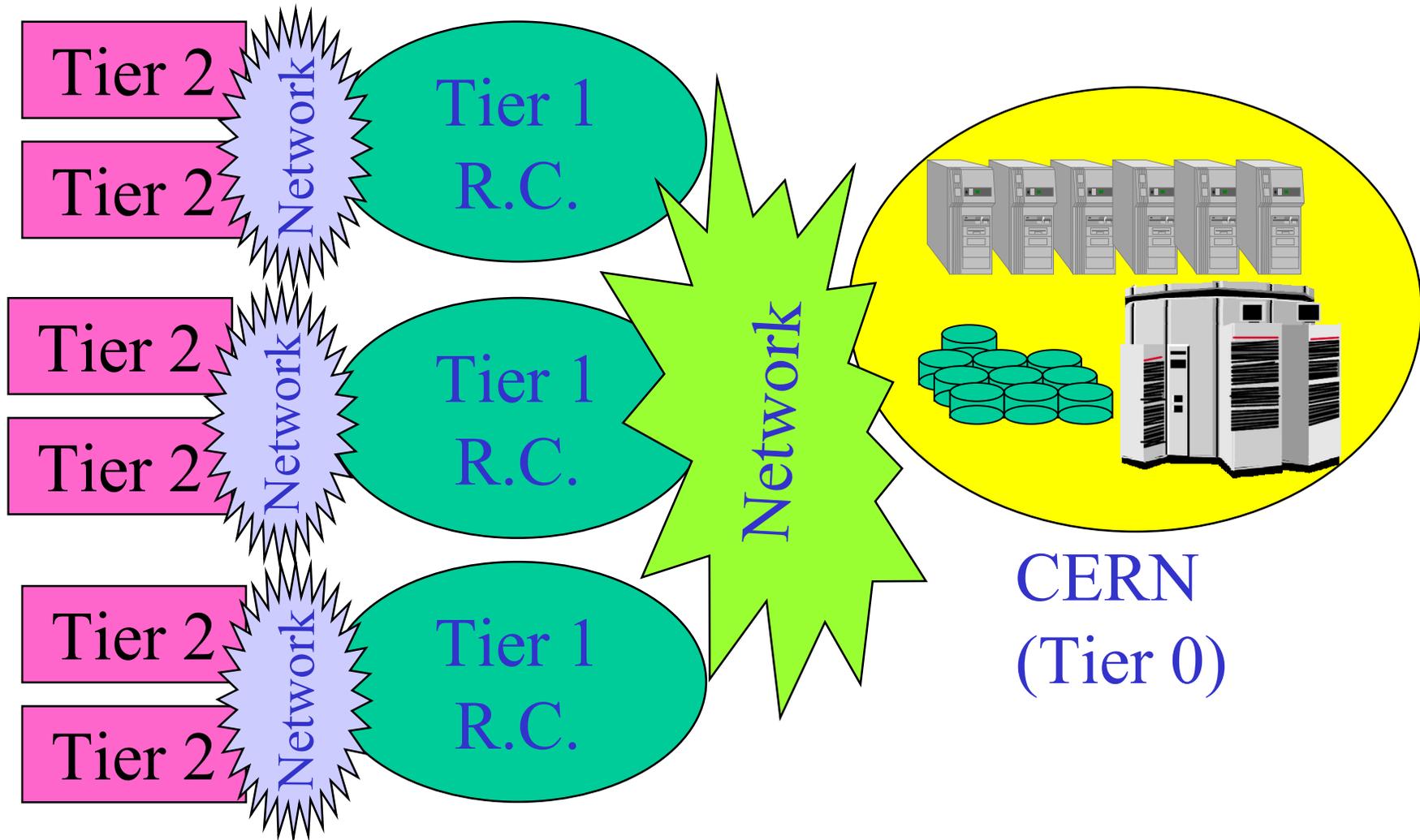
- High Throughput Computing
 - Un numero molto elevato di problemi indipendenti (Eventi)
 - Calcolo in parallelo e non calcolo parallelo.
 - Prestazioni di insieme piuttosto che alte prestazioni sul singolo programma.
 - Ridondanza piuttosto che totale affidabilità dei singoli componenti.
- Questo permette di rivolgersi al mercato consumer per i componenti meno critici.
- Servono però dei sistemi di management altamente scalabili e, possibilmente, altrettanto economici.
- Pur essendo tipica della HEP, non è unica come problematica di calcolo (Biologia, Astronomia, ecc.)

CERN's Network in the World

267 institutes in Europe, 4603 users
208 institutes elsewhere, 1632 users
some points = several institutes



Architettura Gerarchica



Le Capacità Richieste

- CERN (Somma di tutti gli esperimenti):
 - Mass Storage: 10 Peta Bytes (10^{15} B)/anno
 - disk: 2 PB
 - CPU: 20 MSPECint2000
- Per ogni Tier 1 Multi-esperimento :
 - Mass Storage: 3 PB/anno
 - disk: 1.5 PB
 - CPU: 10 MSPECint2000
- Networking Tier 0 --> Tier 1: 2 Gbps

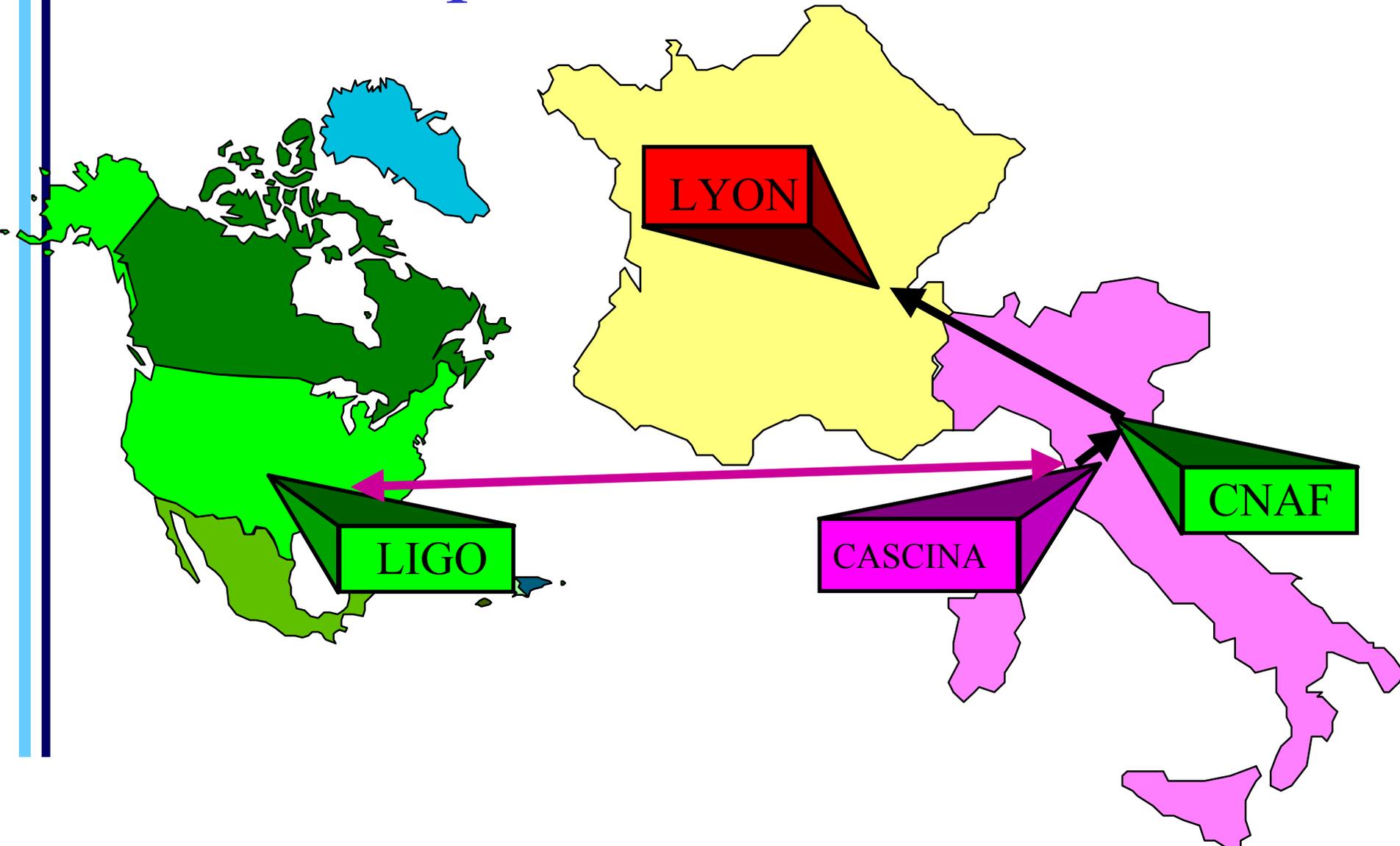
LHC Computing Grid

- Progetto CERN approvato a settembre.
- Si indirizza al calcolo per LHC e per la prototipizzazione dei sistemi Hw e SW necessari.
- Cerca di coordinare lo sforzo con i centri regionali (tier1 e tier2 almeno).
- Strettamente legato agli esperimenti, cerca di coordinarne le attività comuni.

La GRID dei TIER per LHC



Esperimento VIRGO



Progetto TIER1

- Obiettivi:
 - Costruzione e sperimentazione di un prototipo di Tier1 INFN unico per i 4 esperimenti LHC.
 - Ottimizzazione delle risorse.
 - Formazione di personale.
 - Rappresentatività verso gli altri Tier1 e verso il CERN.
 - Coordinamento con i Tier2 e fornitura di servizi ed attrezzature non replicabili o costose.

I Servizi Tecnici Necessari

- Potenza di calcolo (FARM di CPU)
- Accesso ai Dati in Linea (Dischi)
- Archiviazione Dati (Mass Storage/Nastri)
- Rete di Accesso ad Alta Banda e Qualità di Servizio
- Gestione Sistemi Operativi e File System
- Gestione DataBase
- Gestione Librerie Software Generali
- Supporto all'Utenza e servizi accessori
- Coordinamento con gli altri Centri Regionali (Tier0, Tier1, Tier2, ecc.).

Programma delle Installazioni

| Anno | Virgo (Tier 1) | | | LHC (Tier 1) | | | Totali Centro Regionale | | |
|-------------|------------------|--------------|---------------|------------------|--------------|---------------|-------------------------|--------------|---------------|
| | CPU (kSI2000) | DISK (TB) | TAPES (PB) | CPU (kSI2000) | DISK (TB) | TAPES (PB) | CPU (kSI2000) | DISK (TB) | TAPES (PB) |
| 2001 | 10 | 1 | 0.001 | 63 | 11 | 0.03 | 73 | 12 | 0.03 |
| 2002 | 40 | 7 | 0.04 | 183 | 25 | 0.07 | 223 | 32 | 0.11 |
| 2003 | 300 | 100 | 0.16 | 328 | 44 | 0.10 | 628 | 144 | 0.26 |
| 2004 | 1000 | 160 | 0.2 | 800 | 100 | 0.30 | 1,800 | 260 | 0.50 |
| 2005 | 500 | 100 | 0.2 | 1,600 | 200 | 0.60 | 2,100 | 300 | 0.80 |
| 2006 | 500 | 100 | 0.2 | 3,200 | 400 | 2.00 | 3,700 | 500 | 2.20 |
| 2007 | 500 | 100 | 0.2 | 3,200 | 400 | 3.00 | 3,700 | 500 | 3.20 |

Rete

- Accessi attuali tra 34Mbps e 2.5 Gbps
- Nel corso dei prossimi anni un Tier1 dovrebbe avere un accesso di almeno 1-2 Gbps
- Attualmente il CNAF ha un accesso a 100 Mbps e si sta migrando verso 1 Gbps.
- La rete interna è già basata su core switches con interfacce Gbit: 8600 Cabletron (rete CNAF); Extreme 7i (Tier1); Cisco 6509 (GRID e Test)..

FARMING

- Racks da circa 40 macchine biprocessori collegati in dual fast ethernet ed uplink Gigabit Ethernet.
- Boot ed installazione automatica del sistema e del sw (LCFG).
- Gestione remota delle console (KVM Analogici e Digitali).
- Monitoring via SNMP (temperature, carico CPU, ecc.).



Architettura dei sistemi (2)

- DISCO
 - Pool di circa 2-10 TB per staging (CASTOR, o altro SW).
 - 2-4 TB spazio utente (AFS/NFS) per esperimento.
 - Moduli di 1-2 TB spazio disco su richiesta.
 - Architettura attualmente preferita: SCSI e/o Fibre Channel (nativo o su SCSI) con accesso tramite Gigabit Ethernet su Servers ridondati (Fail-over + Accesso distribuito).
 - Sperimentazioni in corso su: FC-IDE, HA, iSCSI, uso di Volumi superiori a 2 TB, Lock Managers, File Systems Distribuiti (GFS, GPFS, ecc.)

Architettura dei sistemi (3)

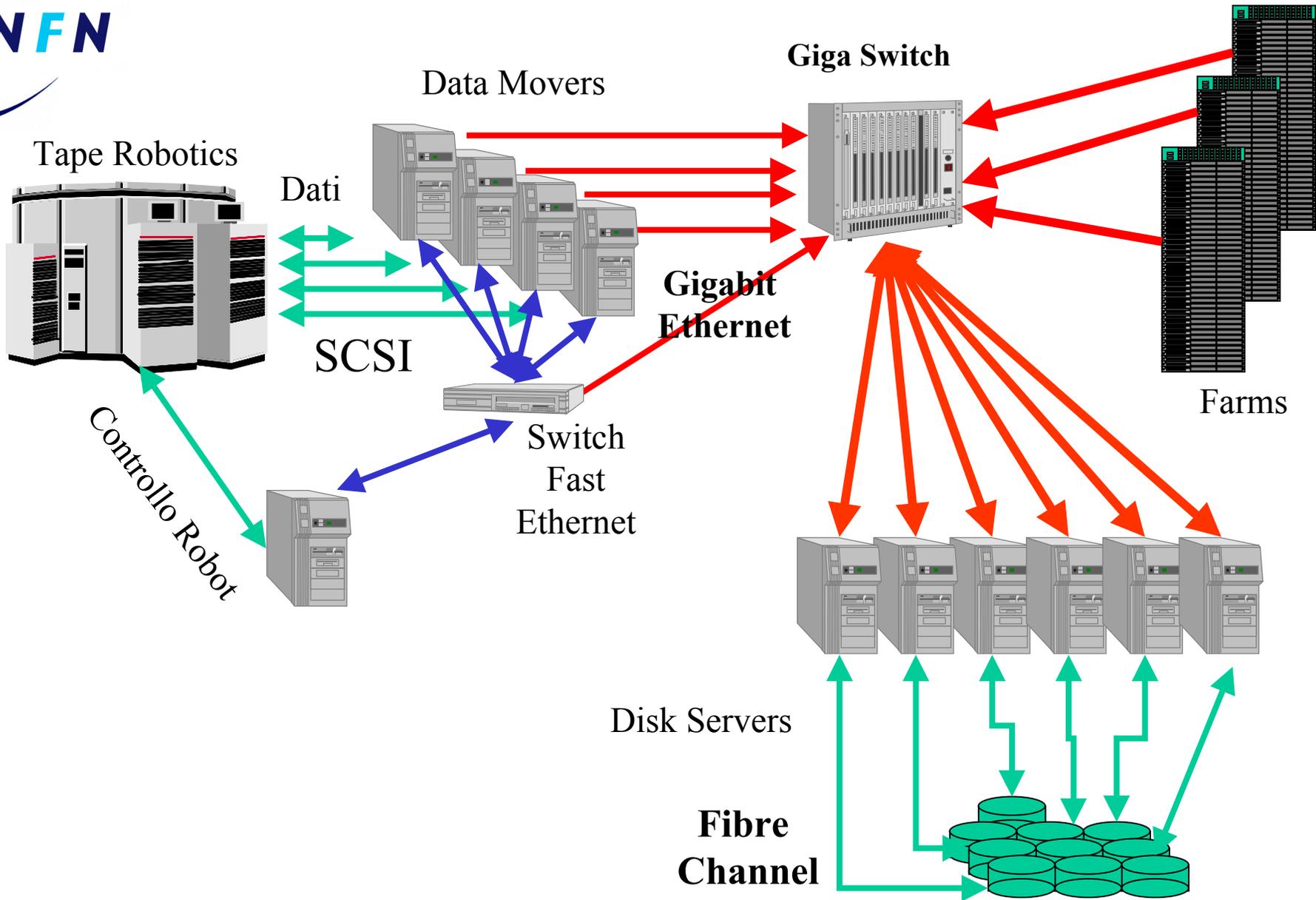
- NASTRI

- Legato Networker per Backup (spazio utente).
- CASTOR (o altro) per Staging dati da nastri LTO.
- Virtualizzazione della libreria L180 con ACSLS.
- Possibile uso di DLT per import/export se richiesto.
- Accesso tramite servers su Gigabit Ethernet con ridondanza (Fail-over + Distributed Access)
- Nuova possibile libreria da 5000-6000 cassette entro fine 2002 – inizio 2003.

Architettura dei sistemi (4)

- ACCESSO

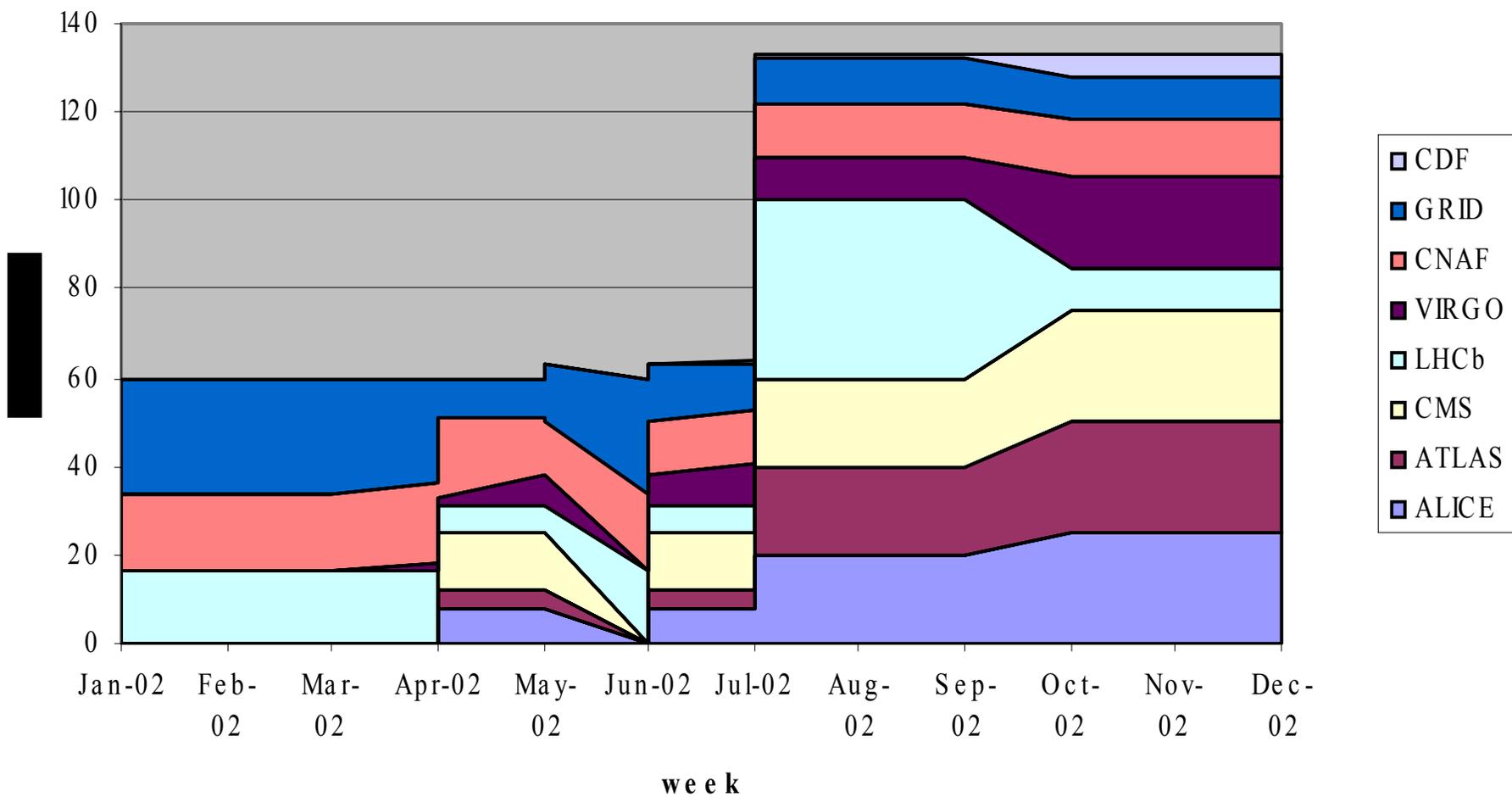
- GRID per le attività di utente sia schedulate che non schedulate (produzione e analisi).
- Portale web personalizzato per monitoraggio ed amministrazione delle risorse assegnate all'esperimento.
- Uso di certificati per l'autorizzazione.
- Autenticazione Kerberos 5 (in corso).
- Rete IP con indirizzi "visibili" dall'esterno e limitazioni sulle porte.
- Rete interna multi-gigabit con accesso a 1-10 Gbps.



Configurazione attuale

| <i>Tipo</i> | <i>Descrizione</i> | <i>Capacità</i> |
|---------------|------------------------------|----------------------------------|
| Farm | 14 biproc. 800Mhz | 10K SPECint2000 |
| | 55 biproc. 1000 MHz | 55K SPECint2000 |
| | 70 biproc. 1400Mhz | 100K SPECint2000 |
| Dischi | NAS Procom | 17 TB Raw RAID5 |
| | Server IBM | 396 GB Raw RAID5 |
| | Dischi SCSI | 2 TB Raw RAID5 |
| | Dischi FC | 8 TB Raw RAID5 |
| Nastri | Robot L180 StorageTek | 3.6 TB / 8 TB (compressi) |
| | Drives LTO | 15 TB / 30 TB (Compressi) |

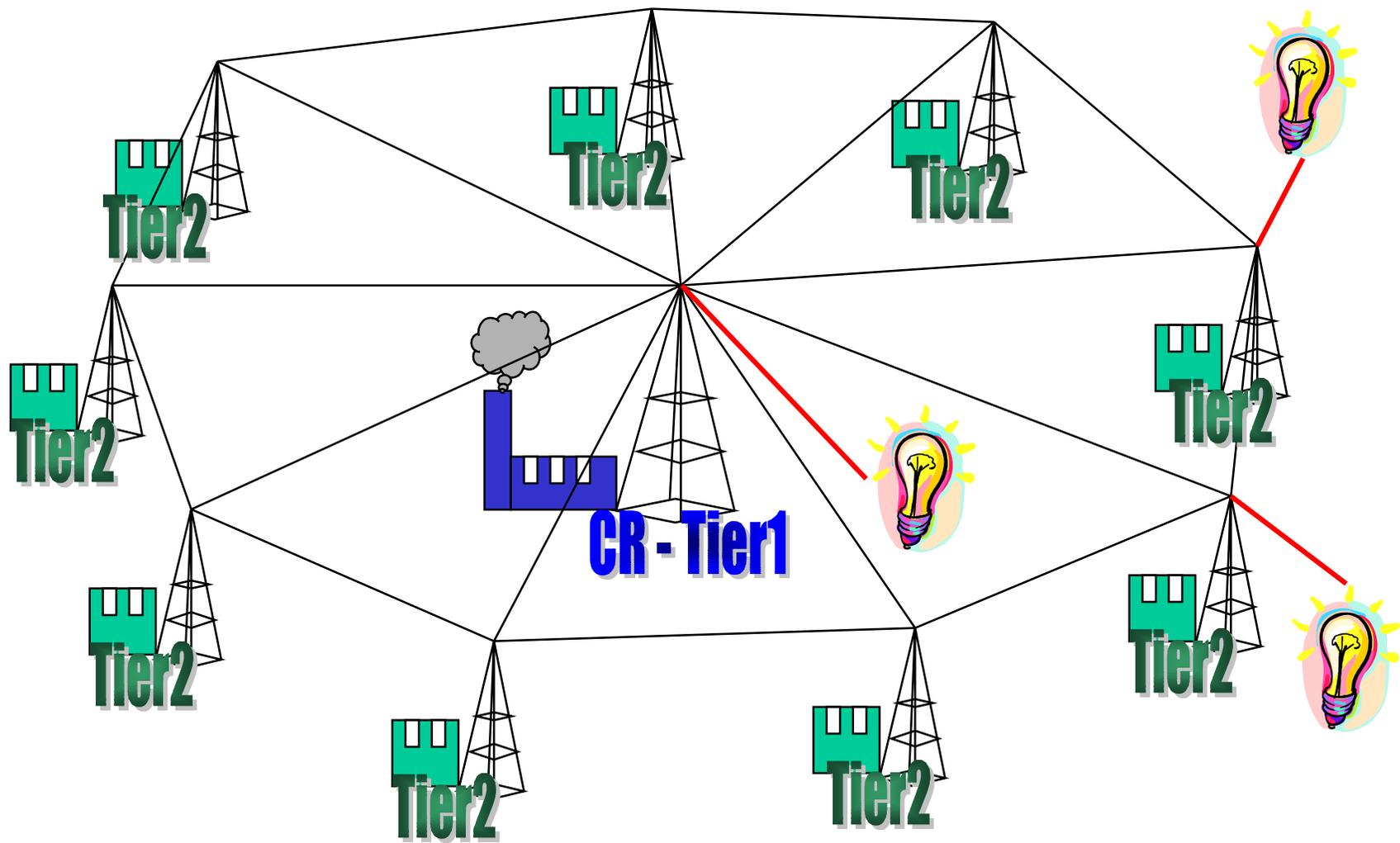
Esempio: Allocazione CPU 2002



Interazioni Tier1-LCG-GRID

- Il Tier1 è uno dei Centri Regionali per LHC e, come tale partecipa all'attività di LCG.
- L'attività di GRID è di fondamentale importanza per la realizzazione del sistema a Tiers previsto. L'architettura generale, definita dal progetto MONARC ha ora un'architettura pratica di riferimento.
- In prospettiva, il Tier1 dovrebbe anche assumere un ruolo di coordinamento e supporto generale.

Una “Centrale” di Servizi GRID



Scala dei tempi

- Giugno 2003: Completa operatività della Sala Macchine.
- Fine 2004: Fine della fase sperimentale.
- Inizio 2005: Revisione della sperimentazione ed aggiornamento tecnico-economico-logistico-organizzativo del piano.
- Fine 2005: Inizio della realizzazione del Centro per la fase di Produzione.