
Il Progetto di Centro di Reprocessing di BaBar: Requisiti per il Monitoring e il Management



Moreno Marzolla

INFN Padova – BaBar Collaboration &
Dip. Informatica, Università di Venezia



Workshop su Management e Monitoring di Cluster Linux
Padova, 6/12/2001

Sommario

- Introduzione alla farm di Reprocessing dell'esperimento BaBar
- Requisiti per il management
- Requisiti per il monitoring
- Alcune considerazioni sul monitoring

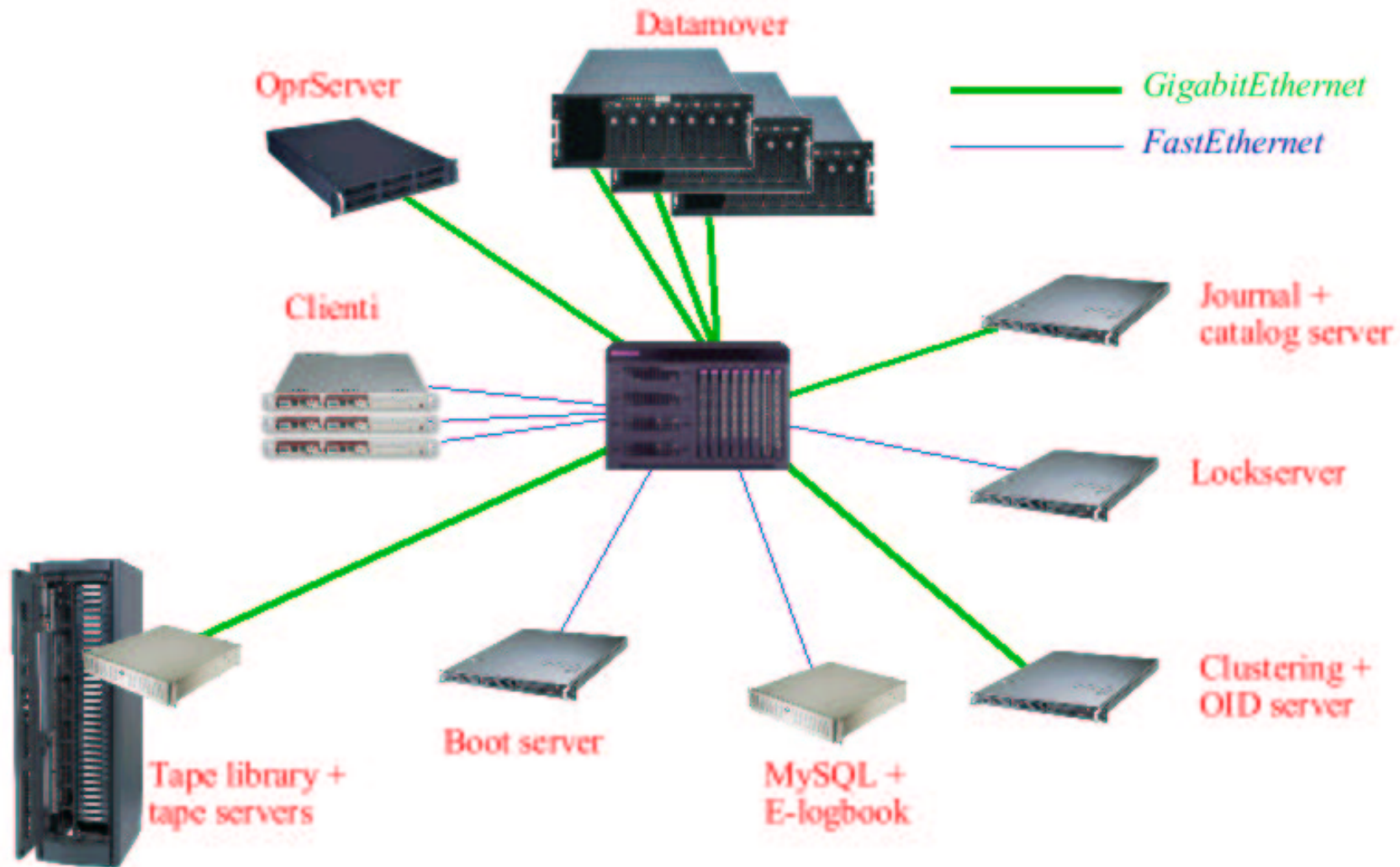
La farm di Reprocessing (1)

- Deve riprocessare ogni anno la statistica raccolta dall'esperimento BaBar negli anni precedenti
 - Tempo di riprocessamento \approx 6 mesi
- Sarà composta da circa 150 nodi client. In ogni nodo client:
 - 2 CPU Pentium, almeno 1GHz
 - Dischi locali
 - 1 GB Ram
 - 1 Ethernet 100Mbit/s
- Il numero di nodi aumenterà ogni anno.

La farm di Reprocessing (2)

- In più vi sono un certo numero di server, con caratteristiche leggermente diverse
 - Più disco locale
 - Scheda di rete Gigabit Ethernet
- Altre componenti:
 - Libreria robotizzata di nastri
 - 20 TB di spazio disco
 - Switch di rete ad alte prestazioni
 - Gruppi di continuità
 - Sistema di condizionamento ambientale

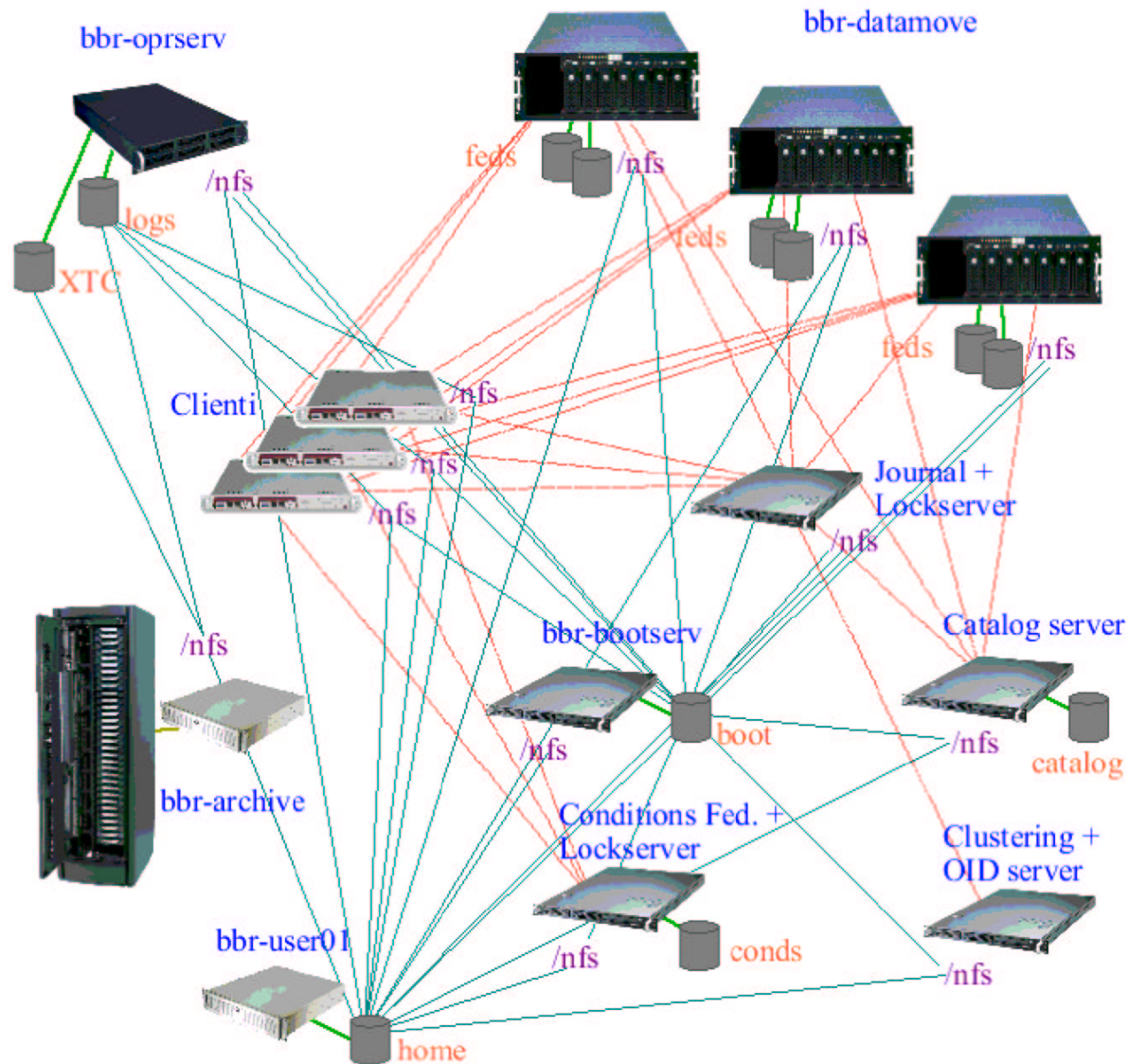
Schema della Farm



Requirement per il Management

- Installazione e configurazione di pacchetti software su tutti i nodi della farm in modo semplice.
- Facilitare l'aggiunta di nuove macchine alla farm
- Bilanciare due proposte estreme:
 - Tutto centralizzato
 - Ovvero tutto il software installato in una directory condivisa via NFS da tutte le macchine della farm
 - NFS collo di bottiglia? Carico eccessivo della rete?
 - Tutto distribuito
 - Ogni macchina ha una copia locale di tutto il software
 - Come aggiornare tutte le macchine se, ad es., cambia la versione del software di ricostruzione?

Schema dei mount NFS



Requirements per il Monitoring (1)

- I parametri da monitorare appartengono alle categorie seguenti:
 - Stato dell'Hardware
 - Crash generico delle macchine
 - Crash dei dischi
 - Altri guasti hardware diagnosticabili
 - Stato della libreria dei nastri
 - Temperatura delle CPU
 - Spazio libero sui dischi

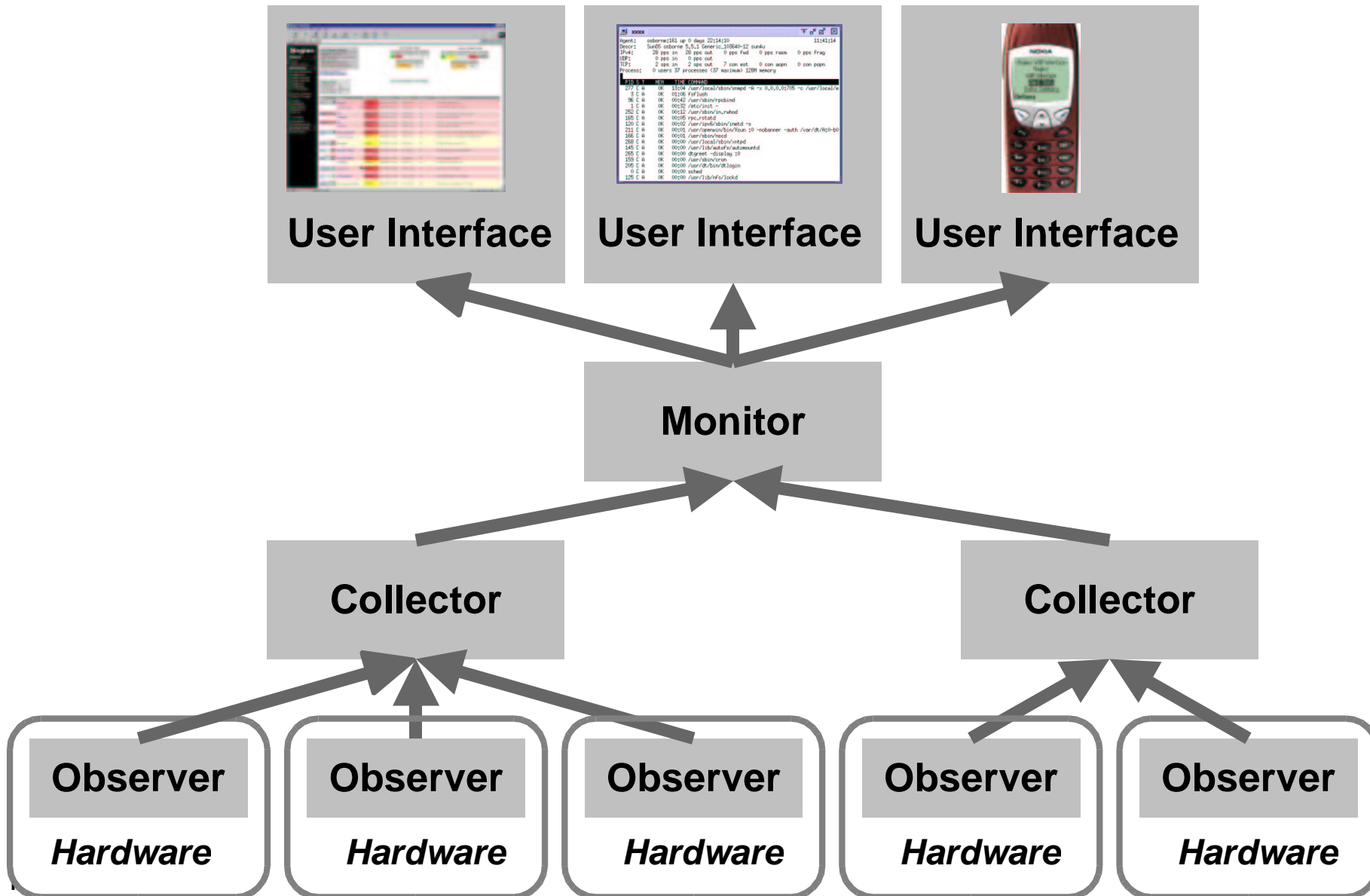
Requirements per il Monitoring (2)

- Stato dei processi
 - Utilizzazione delle CPU nei client e nei server
 - Utilizzazione della rete
 - I/O su disco, specialmente nei server
 - Crash dei processi server o client
- Parametri ambientali
 - Temperatura nella stanza che ospita la Farm
 - Stati dei condizionatori
 - Stato dell'UPS

Altri requirements

- Il sistema di monitoring dovrebbe essere:
 - *Scalabile*
 - Efficiente (richiedere la minor quantità di risorse possibili)
 - Flessibile e facilmente adattabile a nuove esigenze
 - *Facile da configurare*
 - Possibilmente gestibile anche da riga di comando
- E' importante che possa correlare eventi diversi per evidenziare eventuali problemi

II monitoring in generale



Alcune Considerazioni

- Al momento ci siamo concentrati prevalentemente sul monitoring piuttosto che sul management
 - La ricostruzione avviene a SLAC su macchine Sun Solaris, e siamo particolarmente interessati a misurare le prestazioni dell'architettura Linux/Intel in questo contesto.
- Abbiamo preso in esame alcuni prodotti *Open Source*
 - MRTG, Remstats, NetSaint, Snips...
 - La situazione si presenta alquanto caotica...

SNMP

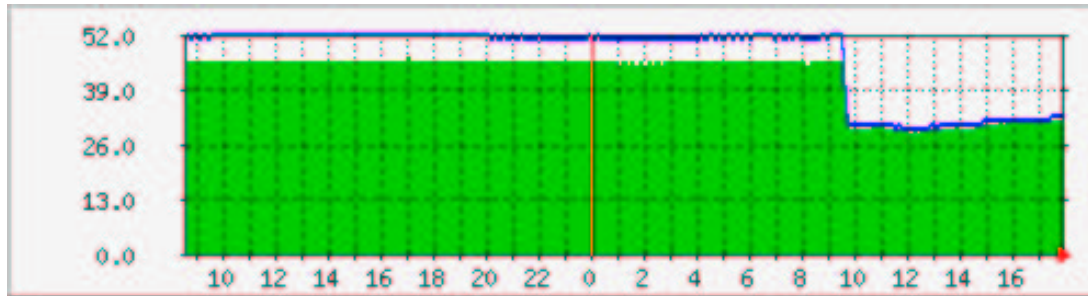
- In base ad alcuni esperimenti preliminari, sembra ragionevole basare il futuro sistema di monitoring su **SNMP** (*Simple Network Management Protocol*)
 - Si tratta di un protocollo ormai maturo
 - È supportato da una vasta gamma di dispositivi
 - È ragionevolmente flessibile
 - Esiste una ottima implementazione *open source* per Unix/Linux (<http://ucd-snmp.ucdavis.edu/>)

Esperienza con SNMP

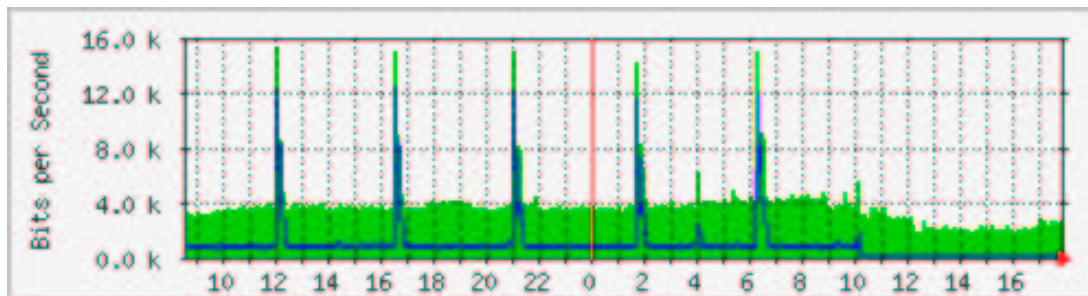
- UCD–SNMP (default su Linux) non supporta nativamente i sensori interni delle macchine, né la misura dell'utilizzazione della CPU
- Siamo passati ad una versione preliminare di NET–SNMP (<http://net-snmp.sourceforge.net/>)
 - Funziona in maniera soddisfacente
 - L'agente SNMP è stato modificato da noi per misurare l'I/O su disco, cosa che non è supportata su Linux al momento

Alcune considerazioni

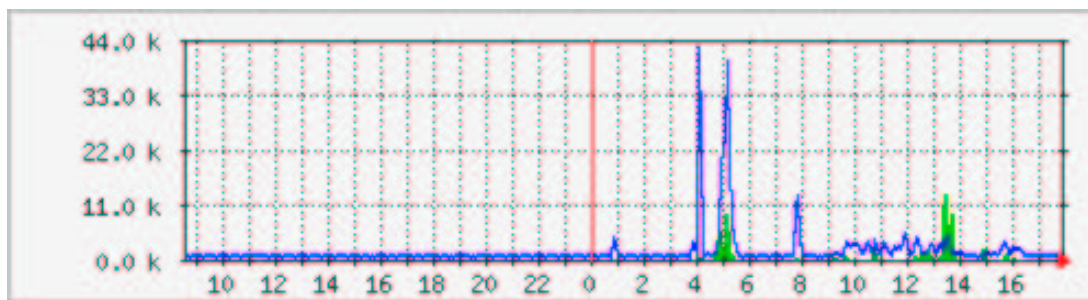
- L'attuale monitoring è basato su MRTG
(<http://people.ee.ethz.ch/~oetiker/webtools/mrtg/>)



Temperatura
CPU



Utilizzazione
Rete



I/O Disco

Problemi generali

- **Uso eccessivo di CPU sul monitor**
 - Molti programmi di monitor sono scritti in Perl
 - Quasi nessuno fa interrogazioni SNMP asincrone
- **Impossibilità di monitorare con granularità sufficientemente piccola (MRTG)**
 - Per diagnosticare certi problemi è utile una granularità di circa 10 secondi.
- **Problemi di scalabilità**
 - Monitoring distribuito?
- **Difficoltà di configurazione**
 - Dozzine di file di configurazione diversi
 - Sintassi spesso *barocca*

Soluzioni

