Stage 2011 Informatica

Progettazione sotto il sistema operativo Unix di un centro di calcolo per un sistema dati che gestisce almeno 1800TByte

Stage 2011 Informatica

TUTOR : Fabio Fortugno

STUDENTI:

Sara Reda

Simone Italo Botticelli

Simone Agostinelli

FUNZIONE DI UN CENTRO DI CALCOLO

- **Acquisizione dati** dall' esperimento
- **Re-processing** dei dati e **analisi**
- **Storage** medio – lungo termine
- **Interfacciamento** verso l'utenza e offerta di kit di analisi.

Unix

Unix e` un **sistema operativo** inizialmente sviluppato da un gruppo di ricerca dei laboratori AT&T e Bell Laboratories, nel quale figurarono sulle prime anche Ken Thompson e Dennis Ritchie (gli sviluppatori del linguaggio di programmazione C).

CARATTERISTICHE:

- **MULTIUSER** : piu` utenti possono interagire contemporaneamente con il sistema
- **MULTITASKING** : gestisce l'esecuzione contemporanea di piu` processi a divisione di tempo.
- **GESTIONE DELLA MEMORIA VIRTUALE** : il sistema della memoria in Unix si basa su paginazione e segmentazione
- **PORTABILE E FLESSIBILE**
- **ARCHITETTURA CLIENT-SERVER**
- **ARCHITETTURA STRATIFICATA**
- **PROGRAMMAZIONE IN C**: Unix mantiene un legame stretto con il linguaggio di programmazione C infatti contiene un insieme ricco di strumenti per lo sviluppo di applicazioni in "C"

Unix: Versione Aix

AIX , acronimo di **A**dvanced **I**nteractive e**X**ecutive, è una serie di sistemi operativi proprietari Unix sviluppati e commercializzati da IBM per molte delle sue piattaforme.
La famiglia di sistemi operativi AIX ha debuttato nel 1986, è diventato il sistema operativo standard per la serie RS/6000 sul suo lancio nel 1990, ed è ancora attivamente sviluppato da IBM.

AIX ha la reputazione di essere un sistema operativo UNIX-Like, sebbene siano presenti numerose differenze sostanziali :

- Filesystem : Simulazione architettura I-NODE ;
- ODM : Database che gestisce i device in memoria invece che sul disco;
- WLM (Work Load Manager) che si occupa del bilanciamento dinamico dei carichi sul processore e in memoria .

COMANDI IN UNIX

I principali comandi della Shell (l'interfaccia tra l' utente con il sistema operativo) **Unix sono:**

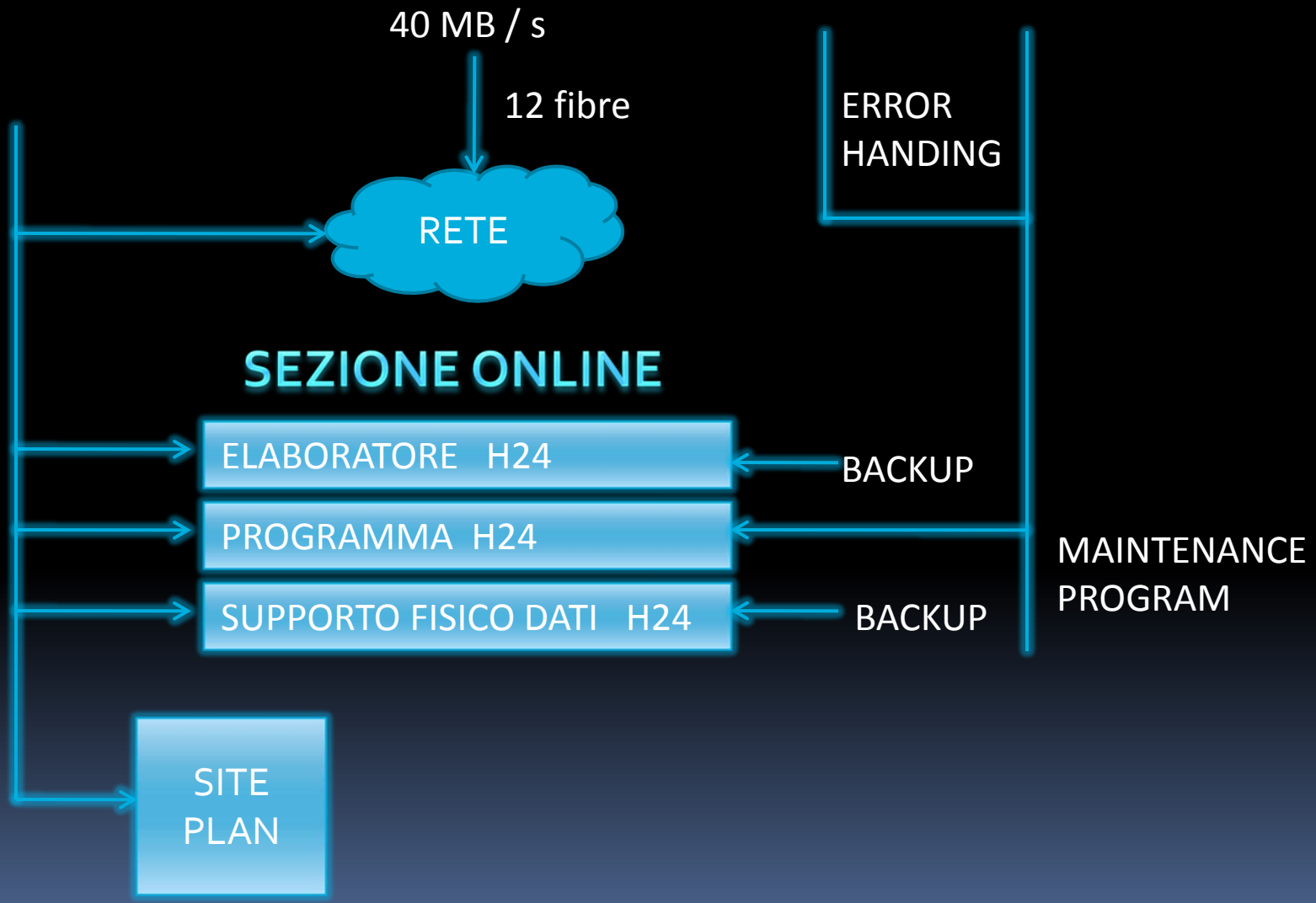
- ls = fornisce l'elenco dei file contenuti nella directory corrente;
- ls – a = permette di vedere i file nascosti e molti dettagli;
- ls –l = per informazioni sui diritti dei file e delle directory;
- cd <directory> = permette di muoversi nella directory indicata;
- cd .. = permette di spostarsi nella directory immediatamente precedente;
- pwd = mostra la directory corrente
- man <comando> = apre la pagina informazioni del comando digitato;
- more = divisione dell`output per pagine
- kill <PID processo>= permette di uccidere un processo;

I principali comandi per operare con i file e con le directory:

- mkdir <directory> = crea una directory con il nome indicato;
- rmdir <directory> = cancella la directory con il nome indicato;
- rm <file> = cancella il file con il nome indicato;
- cp <file sorgente> <destinazione> = copia e incolla il file nella destinazione indicata;
- mv <file sorgente> <destinazione> = taglia e copia il file nella destinazione indicata;
- chmod = gestisce i permessi in Unix
- telnet <nome server > oppure < indirizzo IP> = consente il collegamento con una macchina remota

PROBLEMI DELLA CONTINUITA' DELLA PRESA DATI

Questo settore si occupa dell' acquisizione dati ed e' chiamato **ONLINE**



Attenzione: Usando un solo elaboratore la continuita' e' a rischio

Usando piu' elaboratori si deve pensare un modo per far cooperare il programma di presa dati.

La soluzione a questo problema passa attraverso l'architettura UNIX

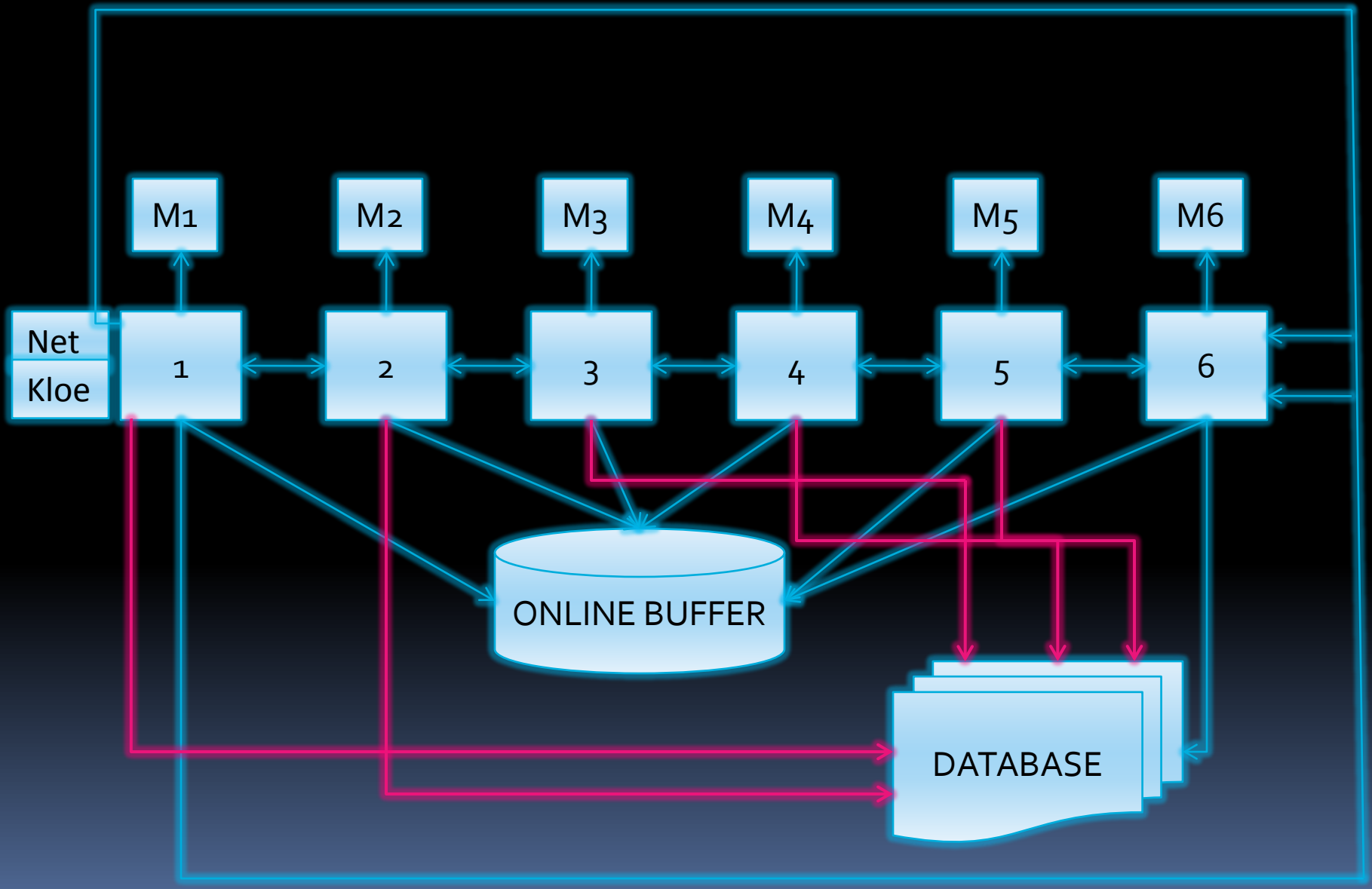
Sfruttando :

- Intersocket communication
- Shared memory
- Semaphores

E' stato realizzato un **circular buffer** che mette in **comunicazione** tutti gli elaboratori che acquisiscono i dati.

In questo modo da 2 a N elaboratori possono cooperare nella presa dati.

CIRCULAR BUFFER

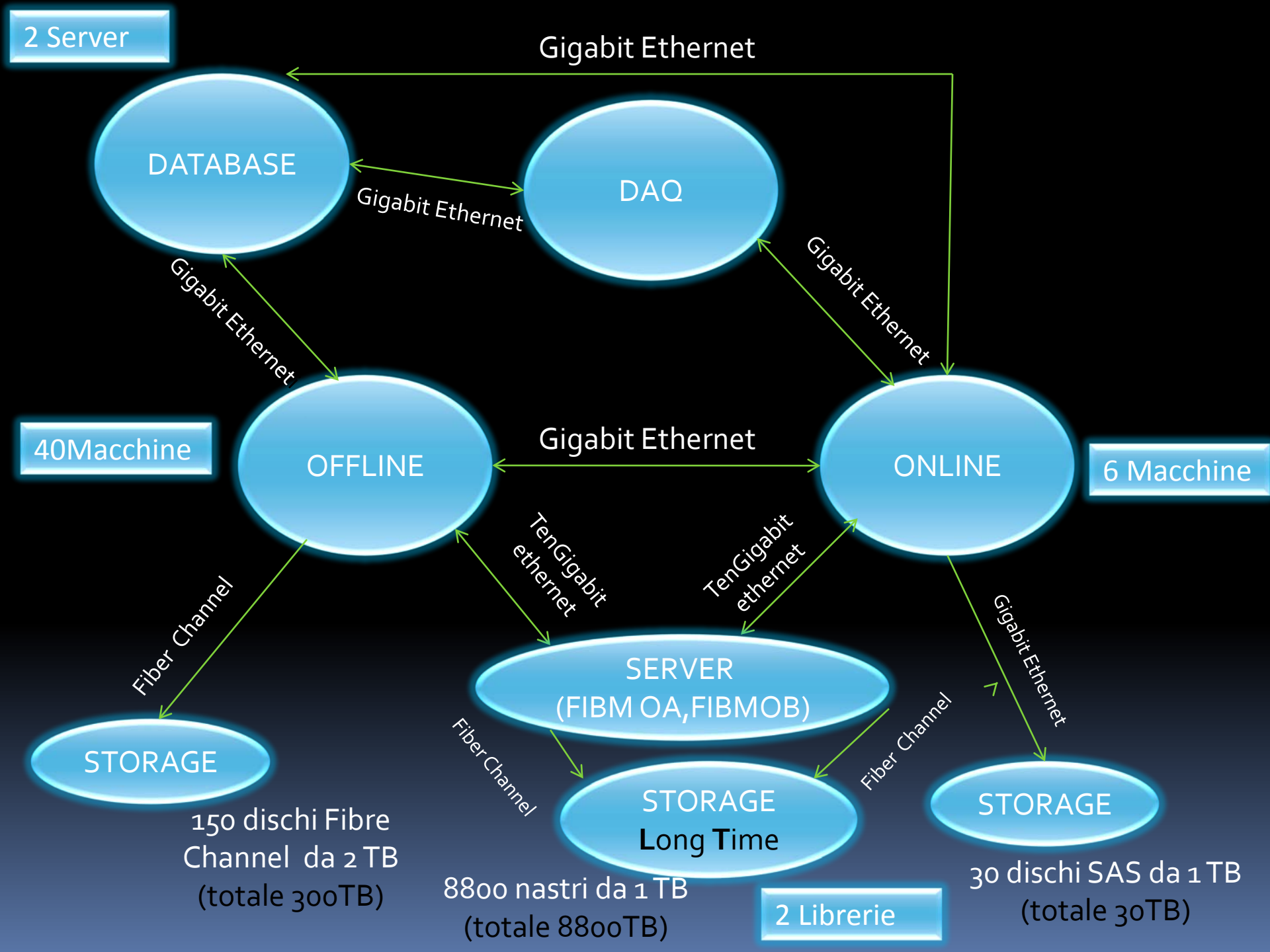


SISTEMA OFFLINE

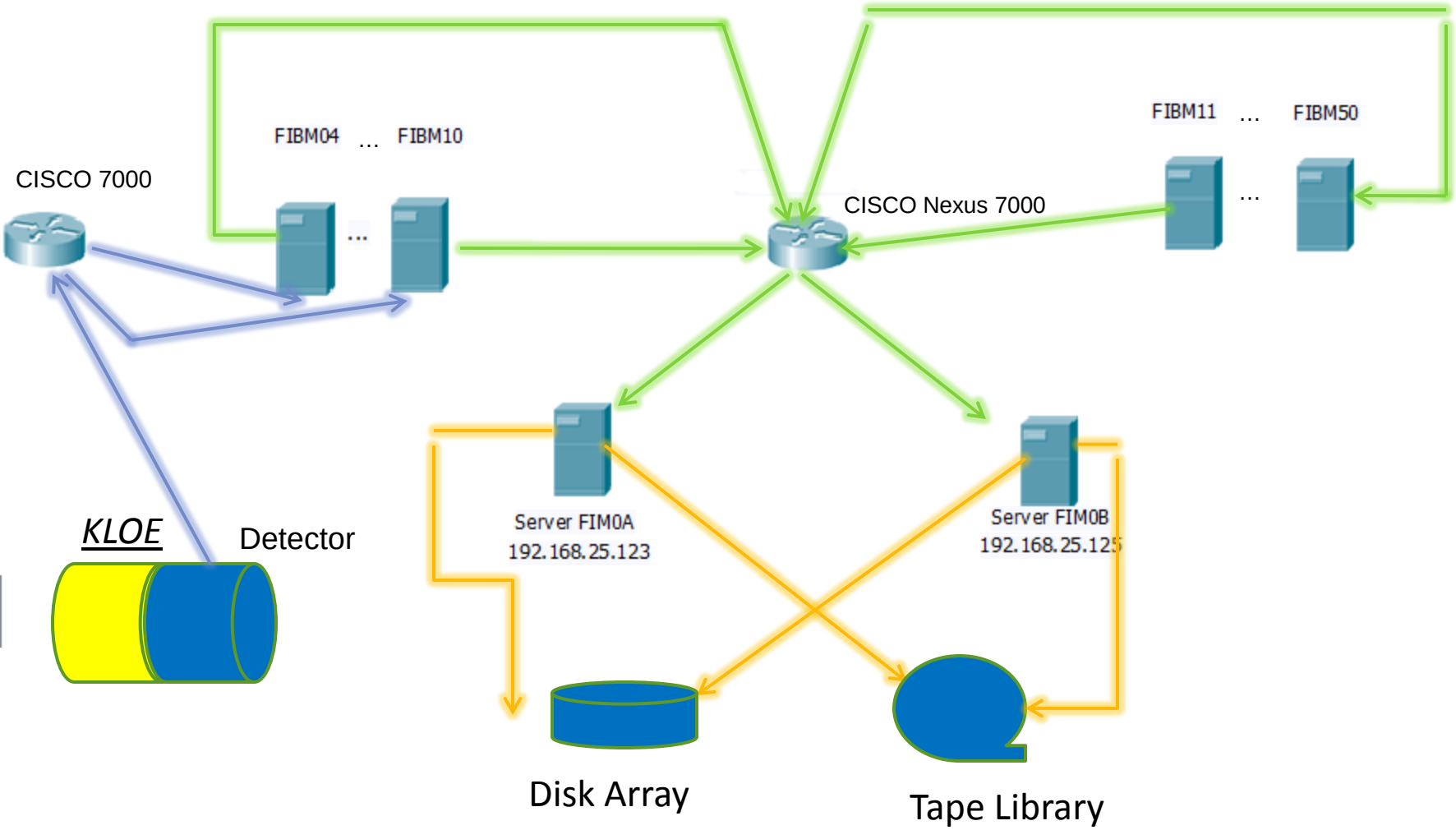
Ogni volta che un segmento dati e' completato viene passato in carico all' **Offline**

L' **Offline** ha le seguenti caratteristiche:

- Lavora in modo logicamente separato dall' Online
- Ha il compito di mettere in **sicurezza** i dati prima di ogni ulteriore operazione di ricostruzione o di analisi dei dati.
- Ricostruisce automaticamente gli eventi di fisica contenuti nei dati. (datarec)
- Costruisce dei file utili alla statistica chiamati DST .
- Fornisce a chi fa analisi attraverso un sistema di code batch e attraverso un sistema HSM per il recupero dei dati dallo storage una serie di tools che permettono di sfuttare a pieno le risorse del CED e di selezionare sotto insiemi di dati


2 Server
Gigabit Ethernet
DATABASE
DAQ
Gigabit Ethernet
Gigabit Ethernet
Gigabit Ethernet
40Macchine
OFFLINE
Gigabit Ethernet
ONLINE
6 Macchine
TenGigabit ethernet
TenGigabit ethernet
Fiber Channel
Gigabit Ethernet
SERVER
(FIBM OA,FIBMOB)
Fiber Channel
Fiber Channel
STORAGE
STORAGE
Long Time
STORAGE
150 dischi Fibre Channel da 2 TB (totale 300TB)
8800 nastri da 1 TB (totale 8800TB)
2 Librerie
30 dischi SAS da 1 TB (totale 30TB)

Le reti del centro di calcolo



1 Rete Fiber channel

Rete 1 ethernet

Rete 2 ethernet

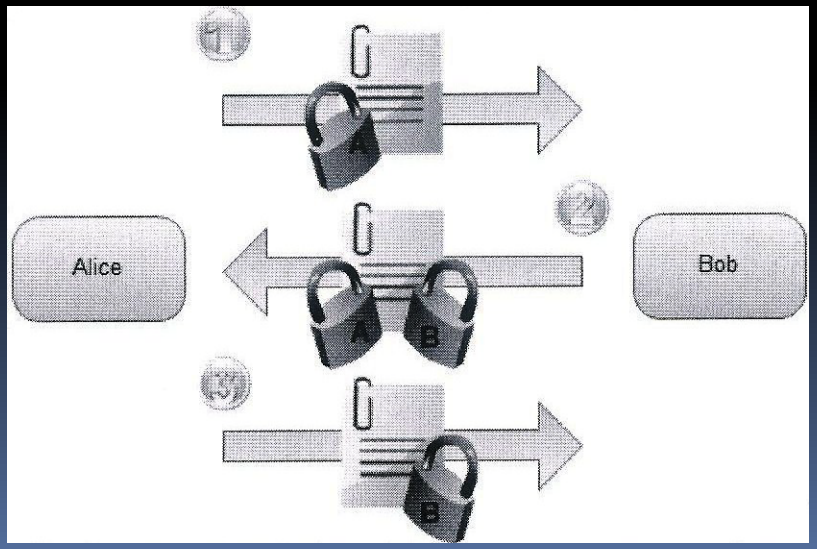
Crittografia asimmetrica

Per salvaguardare la sicurezza dei dati mentre viaggiano nella rete e quindi affinchè essi vengano preservati dall'esterno, viene utilizzata la tecnica della crittografia asimmetrica.

Il problema dello scambio della chiave

Come inviare un messaggio criptato a qualcuno, senza che la chiave di decriptazione viaggi, o abbia viaggiato in precedenza sulla rete?

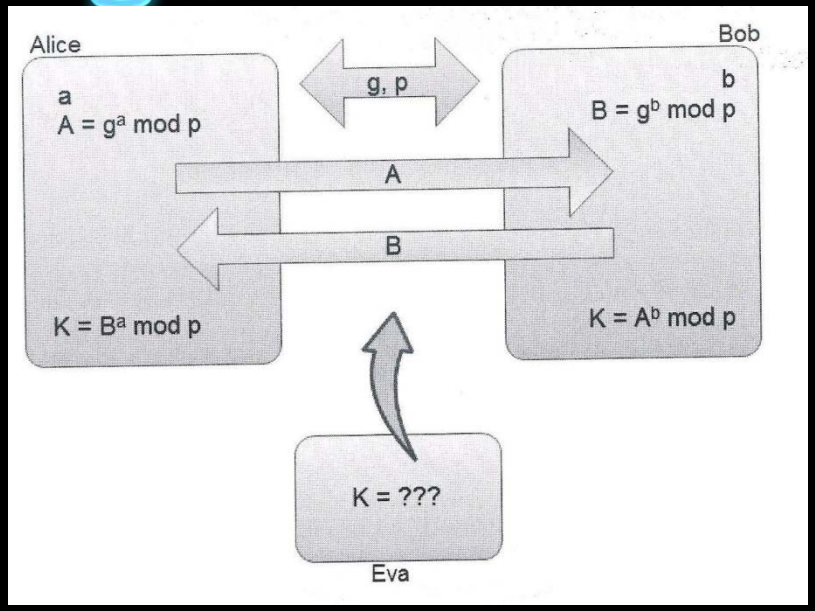
Uno scenario interessante:



1. Alice manda a Bob una scatola chiusa con un suo lucchetto
2. Bob aggiunge il proprio lucchetto e rispedisce la scatola ad Alice
3. Alice rimuove il proprio lucchetto e rispedisce la scatola a Bob

•Il contenuto della scatola è stato inviato senza alcuno scambio di chiavi!

Algoritmo di Diffie-Hellman



- **Usando l'algoritmo sviluppato da Diffie-Hellman nel 1976 è possibile condividere un numero segreto(la chiave) senza trasmetterlo**

- **Non è necessario un canale di comunicazione sicuro.**

L'unico modo che una terza persona (Eva) ha di trovare la chiave è quello di calcolare un logaritmo discreto ⟶ COMPUTAZIONALMENTE DIFFICILE! Tuttavia non sono noti algoritmi che risolvano questo problema in maniera efficente (tempo di calcolo esponenziale)

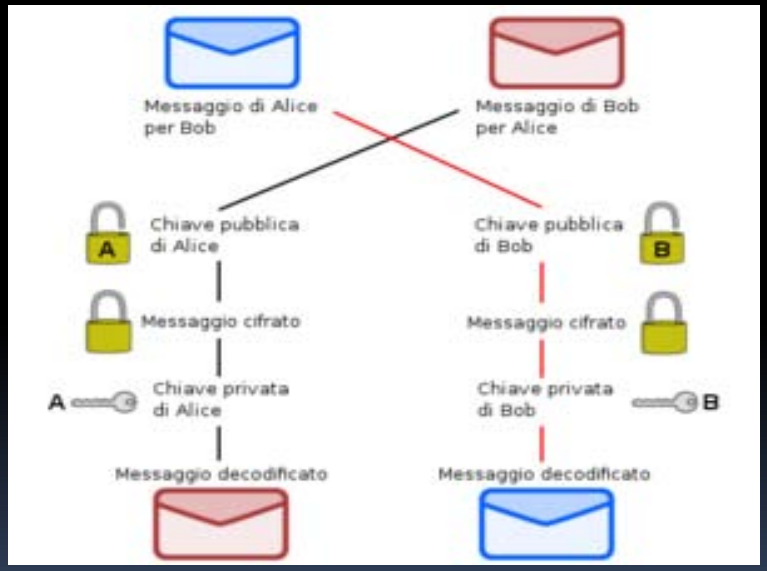
•Per una chiave a 54 bit bastano poche ore di calcolo tramite un attacco a forza bruta per trovare la chiave. Mentre da 128 bit in su anche anni.

- **In seguito alle scoperte di Diffie e Hellman la ricerca in campo crittografo divenne molto attiva**

- **Si giunse così all'invenzione della crittografia asimmetrica (detta anche a "chiave pubblica") , forse l'unica vera rivoluzione nella storia della crittografia.**

- La Chiave Pubblica, deve essere distribuita, serve a *cifrare* un documento destinato alla persona che possiede la relativa chiave privata.

- La Chiave Privata, personale e segreta, e' utilizzata per *decodificare* un documento cifrato con la chiave pubblica.



Questo garantisce la riservatezza del messaggio ma non l'identità del mittente.

Soluzione ⟶ La firma digitale garantita da un ente e che permette

- Autenticazione
- Integrità
- Non ripudiabilità

Che aspetto avrebbe il nostro centro di calcolo…

IBM Power7 Systems

HDD SAS 2TB

SWITCH CISCO NEXUS 7000series

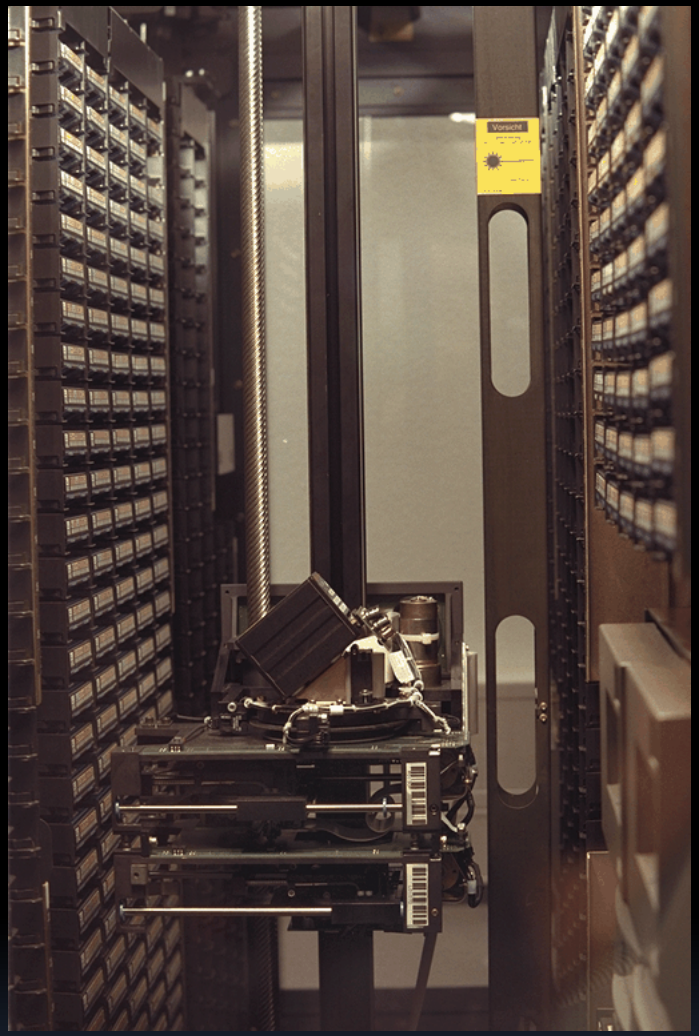
SWITCH CISCO fiber channel Nexus 4400

Disk Array

Perché i tape?

- Meno costosi
- Più capienti
- Non hanno bisogno di alimentazione perchè vengono prelevati solo al momento di lettura/scrittura
- No condizionamento perchè non surriscaldano
- Meno probabilità che si verifichi la rottura

VS



Tape Library IBM 3494

ELENCO COMPONENTI

ONLINE

- 6 elaboratori IBM Pseries POWER 7 , 8/16 CORE
- Switch CISCO serie 7000
- Disco SAS 30 TB

OFFLINE

- almeno 24 elaboratori IBM Pseries POWER 7 , 32/64 CORE
- Switch CISCO Nexus
- 2 Server per lo storage IBM Pseries POWER 7 , 32/64 CORE
- 2 Server per il database IBM Pseries POWER 7 , 8/16 CORE
- Dischi Fiber Channel 300 TB
- 2 Librerie IBM 3494 con almeno 3300 cartridges ciascuna da 1 TB
- 2 Switch fiber channel CISCO 4400

COSTO STIMATO: 2'500'000 euro

CONCLUSIONE

Per passare alla fase esecutiva del progetto aspettiamo un contratto dalla direzione dell’ Ente ;-)

THE END