

1 Storage architectures, technologies and systems

1.1 Evoluzione

1.1.1 Fattori legati al Business

I dati sono importanti per le aziende moderne e bisogna proteggerli per non perderli e garantire sempre il loro accesso. Bisogna gestire le situazioni critiche mettendo i dati in un posto sicuro, garantendo la continuità al business. I costi per la manutenzione degli storage sono maggiori degli acquisti e c'è bisogno di semplificare l'architettura.

Le infrastrutture devono essere: **efficienti, affidabili, scalabili, gestibile, basate su standard e ragionevolmente economico.**

1.1.2 Fattori tecnici

Le ultime tecnologie sviluppate hanno permesso di sviluppare nuove applicazioni (internet, multimedia, BI, ...). Di conseguenza c'è stata una maggiore richiesta di: **capacità d'archiviazione, performance, affidabilità.**

1.1.3 Insidie degli approci tradizionali

Alla fine del '97 la velocità del network supera quello dello storage. Gli approci tradizionali hanno diversi vantaggi quali: **"attaccati direttamente al server, bus parallelo SCSI usato come interconnessione, accessi diretti al server, se il server è inattivo l'archiviazione non è accessibile, numero di dischi per controller è limitato in un server, ...** Questo approccio non è una buona architettura per la scalabilità!

1.1.4 Cambio di paradigma

Per risolvere il problema delle reti veloci che mettono sotto pressione i server, che generano un overload delle richieste, si è dovuto cambiare dallo **storage collegato direttamente a quello condiviso**, con il seguente schema:

- L'archiviazione può essere accessibile da più server
- Il carico di lavoro può essere distribuito su più server
- Gli storage possono essere spostati spostati da un server all'altro
- Gestione centralizzata dello storage aziendale

1.1.5 Crescita dei dati

Con la nascita degli IoT la quantità di dati sono esplosi, proprio perché lavorano 24/24, rispetto all'interazione dell'uomo con il device (videocamere, sensori, ecc...). Questa crescita ha aumentato drasticamente i costi del IT con: più capacità di storage (e costi), grandi backup, ecc...

ILM (Information Lifecycle Management) permette le strategie per l'amministrazione di sistemi di archiviazione su dispositivi informatici.

1.1.6 Aspetti legali

Regole nazionali e internazionali impongono alle aziende delle regole per l'archiviazione, la conservazione e la sicurezza dei dati. La conformità è quindi un problema che influisce sul ICT e ILM è un modo per supportarla.

1.2 Storage

Uno storage system è una tecnologia designata e usata per la conservazione dei dati digitali. Ci sono 4 gerarchie di storage systems:

- **Primario** - Accesso diretto dalla CPU (RAM, ROM, cache)

- **Secondario** - Non diretto dalla CPU (HDD, SSD)
- **Terziario** (nearline) - Uso di robot per caricare/scaricare in un drive
- **Off-line** - Supporti d'archiviazione mantenuti offline e richiede l'intervento umano per caricare/scaricare in un drive

Parametri per la differenziazione dei livelli di gerarchia: performance, mutability, latency, addressability, capacity, security, energy use, cost per capacity.

1.2.1 Tipi di storage

- **Blocchi** - Su dischi e memorie (tracce e settori)
- **File** - Cartella in ordine logico fisso (file path, file name, Date)
- **Oggetti** - Container di dimensione flessibile (UID, Data e metadati)

1.2.2 Storage a blocchi

I dati sono salvati in blocchi di dimensione fissa, senza nessun metadato di alto livello. Accessibile da OS con un "mount" e tramite un file system si decide come i blocchi sono accessibili, combinati e modificabili. L'applicazione scrive sui data block, ideale per l'archiviazione primaria ad alte prestazioni. Casi d'uso: strutture DB, volumi virtuali, Workloads, High change content, Random R/W e Bursty I/O.

Accesso ai dati: Applicazione scrive blocchi → Iniettor SW/HW → Controller → dati scritti sul device come data block.

1.2.3 File storage

I file sono accessibili in modo casuale, chiamati con operazioni IO attraverso un manipolatore, vengono organizzati in directory

(file strutturati) e possono avere diversi livelli (anche combinati): **file system logici, virtuali, fisici e accesso di rete.** Hanno supporto POSIX e possono essere trattati attraverso dati non strutturati (stream o oggetti), dati strutturati e blocchi.

1.2.4 Storage ad oggetti

Contiene tipicamente le repository, URL, metadati, ObjectID e versioni. I casi d'uso sono: **per i cloud pubblici/privati/ibridi, archiviazione, archiviazione senza server, IoT, machine learning, ecc....** Non richiede un "mount" per accedere ai dati e sono accessibili da qualsiasi endpoint.

1.3 Storage Area Network (SAN)

1.4 iSCSI

1.5 Inband

2 High-end systems and RAS (Reliability, Availability e Serviceability)

3 Clustering for high performance computing and high-availability

4 Virtualization

5 Cloud