

RAPPORTI TECNICI INGV

Il Backup dei dati.
L'esempio dell'INGV - Osservatorio
Vesuviano



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

419

Direttore Responsabile

Valeria DE PAOLA

Editorial Board

Luigi CUCCI - Editor in Chief (luigi.cucci@ingv.it)
Raffaele AZZARO (raffaele.azzaro@ingv.it)
Christian BIGNAMI (christian.bignami@ingv.it)
Mario CASTELLANO (mario.castellano@ingv.it)
Viviana CASTELLI (viviana.castelli@ingv.it)
Rosa Anna CORSARO (rosanna.corsaro@ingv.it)
Domenico DI MAURO (domenico.dimauro@ingv.it)
Mauro DI VITO (mauro.divito@ingv.it)
Marcello LIOTTA (marcello.liotta@ingv.it)
Mario MATTIA (mario.mattia@ingv.it)
Milena MORETTI (milena.moretti@ingv.it)
Nicola PAGLIUCA (nicola.pagliuca@ingv.it)
Umberto SCIACCA (umberto.sciacca@ingv.it)
Alessandro SETTIMI (alessandro.settimi1@istruzione.it)
Andrea TERTULLIANI (andrea.tertulliani@ingv.it)

Segreteria di Redazione

Francesca DI STEFANO - Coordinatore
Rossella CELI
Barbara ANGIONI
Tel. +39 06 51860068
redazionecen@ingv.it

REGISTRAZIONE AL TRIBUNALE DI ROMA N.174 | 2014, 23 LUGLIO

© 2014 INGV Istituto Nazionale
di Geofisica e Vulcanologia
Rappresentante legale: Carlo DOGLIONI
Sede: Via di Vigna Murata, 605 | Roma



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA

RAPPORTI TECNICI INGV

Il Backup dei dati.
L'esempio dell'INGV - Osservatorio Vesuviano

Data Backup.
INGV - Vesuvius Observatory example

Giuseppe Borriello, Rosario Peluso, Giovanni Scarpato

Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Napoli, Osservatorio Vesuviano

Accettato 26 febbraio 2020 | Accepted 26 February 2020

Come citare | How to cite Borriello G., Peluso R., Scarpato G., (2020). Il Backup dei dati. L'esempio dell'INGV - Osservatorio Vesuviano.
Rapp. Tec. INGV, 419: 1-36.

In copertina Arte astratta | Cover Abstract-art

419

INDICE

| | |
|--|-----------|
| Introduzione | 7 |
| <i>Introduction</i> | 7 |
| 1. I sistemi di backup | 7 |
| 1.1 I dati da salvare e la loro dimensione | 8 |
| 1.2 I supporti di backup | 9 |
| 1.3 La configurazione hardware | 11 |
| 1.4 Il software | 12 |
| 1.5 I tipi di backup | 14 |
| 2. Il costo del backup | 16 |
| 2.1 Il costo del cloud | 19 |
| 3. Disponibilità ed affidabilità dei sistemi di storage | 19 |
| 3.1 Le configurazioni analizzate | 19 |
| 3.2 Il calcolo dell'affidabilità, disponibilità ed MTBF | 22 |
| 4. La soluzione ideale | 29 |
| 5. Upgrading del sistema | 30 |
| 6. Conclusioni | 30 |
| Sitografia | 31 |

Introduzione

Una delle principali esigenze di un'organizzazione è quella di impedire che i dati, creati e stoccati con dispendio di risorse, vadano perduti in seguito ad eventi che non esiteremmo a definire "catastrofici", per le enormi conseguenze che essi comportano.

La misura da adottare si definisce "backup dei dati". Il concetto è semplice: effettuare copie di sicurezza per impedire che i nostri dati vitali diventino uno spiacevole ricordo a causa di problemi hardware, software, incendi, problemi elettrici od altri eventi.

Se il concetto che sta alla base è semplice, molto più complessa è la strategia per mettere in atto il salvataggio dei dati.

Il backup periodico deve far parte di un'efficiente strategia di "disaster recovery", che permette di ripristinare i dati in tempi ragionevolmente brevi.

Diverse strategie e tecnologie sono disponibili per implementare un corretto piano di backup. Analizzeremo brevemente i punti di forza e le criticità dei principali sistemi disponibili, al fine di pianificare ed applicare una politica di backup e ripristino, per poter usufruire dei dati ed averli disponibili nel minor tempo di latenza possibile.

Configurazione ottimale, costi, supporti, affidabilità e disponibilità sono alcuni degli aspetti da valutare con attenzione per minimizzare il rischio di una perdita di dati.

Effettuare un'analisi dettagliata di tutti gli aspetti non è un compito semplice. Lungi dall'essere esaustivo, lo scopo del presente lavoro è di effettuare una breve panoramica dei principali aspetti del processo di backup e fornire un ordine di grandezza dei costi sia in termini economici che di risorse tecniche necessarie all'attuazione del piano di salvataggio dati.

Saranno presi in considerazione la configurazione hardware, l'affidabilità e disponibilità dei sistemi, il software, i tipi di backup e le dimensioni del backup.

Introduction

One of the main goals of an organization is to prevent data loss as a result of catastrophic events.

The measure to be adopted is called "data backup".

It's a simple concept: make backup copies to prevent data from becoming an unpleasant memory due to hardware, software, fire, electrical problems or other events.

If concept is simple, the strategy for saving data is much more complex. Periodic data backup must be part of an efficient "disaster recovery" strategy.

Different strategies and technologies are available to implement a correct backup plan. We will briefly analyze the strengths and weaknesses of the main systems available, in order to plan and apply a backup and recovery policy, taking advantage of the data and have it available in the shortest possible latency time. Optimal configuration, costs, backup mediums, reliability and availability are some of the aspects to be carefully evaluated to minimize the risk of data loss.

Carrying out a detailed analysis of all aspects is not a simple task. Far from being exhaustive, the aim of this work is to provide a brief overview of backup process main aspects and provide magnitude of costs in economic and technical resources necessary for the data backup plan implementation. Hardware configuration, reliability and availability of the systems, software, types and size of backup will be taken into consideration.

1. I sistemi di backup

Un'indagine di Frost & Sullivan del 2008 [<https://assets.ext.hpe.com/is/content/hpedam/documents/a00056000-6999/a00056041/a00056041enw.pdf>], afferma che le organizzazioni devono

affrontare notevoli difficoltà nel processo di backup, causate sia dal crescente flusso di dati, sia da altri aspetti, quali:

- Maggior numero di formati rispetto al passato.
- Sorgenti multiple.
- Punti accesso multipli.
- Lungo ciclo vitale: La necessità di accedere anche a dati più vecchi, prima relegati in backup offline.

Per definire una strategia di backup bisogna definire diversi aspetti, quali:

- Identificare i dati da salvare e la loro dimensione.
- Decidere dove verranno conservati i dati. (nastri, dischi).
- La configurazione hardware.
- Il software.
- Tipologia di backup.

1.1 I dati da salvare e la loro dimensione.

I dati possono, a grandi linee, essere suddivisi in dati statici e dinamici, in rapporto alla frequenza di modifica alla quale sono sottoposti. Le macchine virtuali sono un esempio di dati statici, mentre i dati sismici provenienti dall'acquisizione real-time sono dinamici.

Nella tabella 1 sono indicate le dimensioni di alcuni dei principali backup effettuati nella sezione di Napoli- Osservatorio Vesuviano aggiornate al 30/07/2019. La media è circa 3 Terabyte/anno.

| Dati al 30/07/2019 | Totale (GB) | Media Per Anno (GB) | Media Per Mese (GB) |
|--------------------|----------------|---------------------|---------------------|
| GEOCHIMICA | 1,19 | 0,22 | 0,018 |
| GEOELETTRICA | 30,9 | 7,1 | 0,58 |
| RADON | 1,95 | 0,37 | 0,03 |
| DEFORMAZIONI | 1994,91 | 222,41 | 18,54 |
| SISMICA Mobile | 5869,97 | 407,23 | 33,93 |
| SISMICA Permanente | 28144,98 | 2266,83 | 188,89 |
| TOTALI | 36043,9 | 2904,16 | 241,99 |

Tabella 1 Le dimensioni di alcuni dei principali backup presso OV.

Table 1 Size of some of the main backups at OV.

La tipologia di dati che vengono sottoposti a salvataggio è varia. Quelli di maggior dimensioni sono costituiti principalmente da file come i dati elencati nella tabella 1, altri sono backup dei numerosi database in uso.

Nella figura 1, a titolo di esempio, è mostrata l'interfaccia pubblica del database sismologico dei vulcani della Campania denominato "Plinio". Nella figura 2 sono elencate le dimensioni dei backup di alcuni dei database della sezione di Napoli, mentre in figura 3 le dimensioni di alcuni dei backup basati su file.

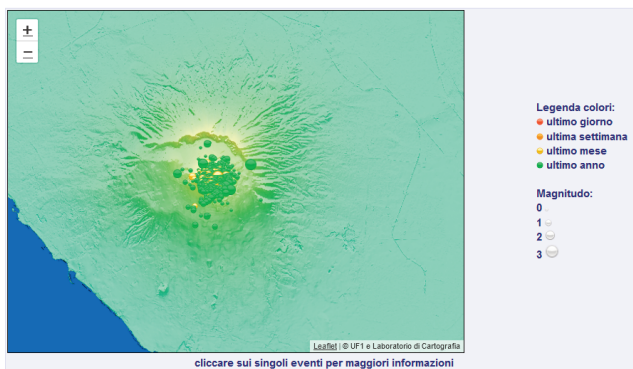


Figura 1 Il database “Plinio”.

Figure 1 “Plinio” database.

LISTA EVENTI (TOTALE 932)

[cliccare sui singoli eventi per maggiori informazioni](#)

| Tempo origine (UTC) | Magnitudo | Latitudine N | Longitudine E | Profondità (km) | Note |
|------------------------|-----------|--------------|---------------|-----------------|-------------|
| 2020/04/15 22:27:16.00 | 0.3 | 40.8217 | 14.4263 | 0.20 | Preliminare |
| 2020/04/15 22:24:5.00 | 0.6 | 40.8197 | 14.4272 | 0.17 | Preliminare |
| 2020/04/12 21:20:51.00 | 0.2 | 40.8220 | 14.4310 | 2.00 | Preliminare |
| 2020/04/12 03:11:21.00 | 0.4 | 40.8183 | 14.4313 | 0.30 | Preliminare |
| 2020/04/08 21:07:52.21 | 0.0 | 40.8190 | 14.4323 | 0.06 | Revisionato |
| 2020/04/08 16:46:58.07 | 0.8 | 40.8237 | 14.4263 | 0.15 | Revisionato |
| 2020/04/08 12:37:37.57 | 0.4 | 40.8208 | 14.4255 | 0.45 | Revisionato |
| 2020/04/07 19:38:54.38 | 0.0 | 40.8220 | 14.4257 | 0.10 | Revisionato |
| 2020/04/07 16:27:1.98 | 1.1 | 40.8223 | 14.4272 | 0.19 | Revisionato |

Biggest backup jobs

| Job name | Bytes | Files |
|------------------------------------|-----------|---------|
| sql_server_2000_server_5000_2_tape | 98.94 GB | 2 |
| server5000_sql_server_diff_full | 45.13 GB | 375 |
| server5000_sql_server | 44.54 GB | 352 |
| dms_geoves_server_5000_diff_full | 32.84 GB | 215.797 |
| dms_geoves_server_5000 | 32.76 GB | 215.236 |
| dms_geoves_2_tape | 32.39 GB | 212.304 |
| server5000_sql_server_diff_full | 969.07 MB | 37 |
| server5000_sql_server_diff_full | 846.43 MB | 33 |
| parco_auto_2_tape_full | 767.09 MB | 2 |
| parco_auto_full | 756.46 MB | 47.213 |

Figura 2 Dimensioni dei backup dei database.

Figure 2 Database backups size.

Biggest backup jobs

| Job name | Bytes | Files |
|--|----------|-----------|
| backup_pool_sismica_perm_camp_velo_bigmama2 | 18.42 TB | 3.445.013 |
| backup_pool_sismica_perm_camp_velo_bigmama2 | 18.10 TB | 3.401.011 |
| backup_pool_sismica_mobile_bigmama2 | 6.71 TB | 8.940.221 |
| backup_pool_sismica_mobile_bigmama2 | 6.29 TB | 8.558.707 |
| backup_pool_sismica_perm_camp_accel_bigmama2 | 3.90 TB | 3.022.963 |
| backup_pool_sismica_perm_stromboli_bigmama2 | 3.38 TB | 6.721.382 |
| backup_pool_sismica_perm_stromboli_bigmama2 | 3.37 TB | 6.693.933 |
| backup_pool_sismica_perm_camp_clienti_array_microfoni_bigmama2 | 3.18 TB | 3.537.504 |
| backup_pool_deformazioni_bigmama2 | 2.43 TB | 6.889.969 |
| backup_pool_deformazioni_bigmama2 | 2.35 TB | 6.107.475 |

Figura 3 Dimensioni dei backup file-based.

Figure 3 File-based backups size.

1.2 I supporti di backup

I principali supporti di storage sono:

- Raid di dischi.
- Nastri.

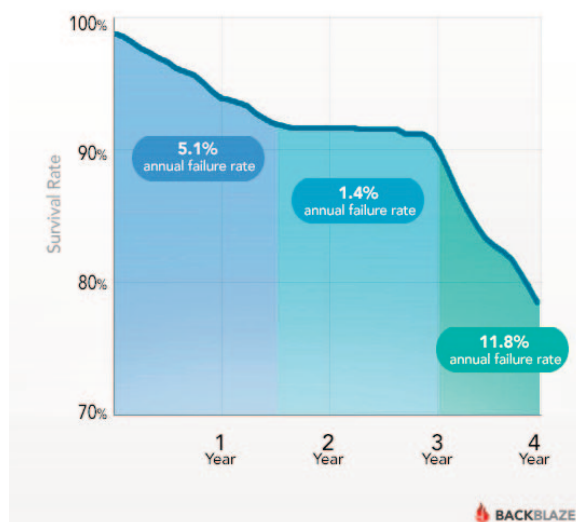
Un sistema di dischi Raid è una soluzione conveniente. Esistono in commercio librerie di dischi dal costo abbordabile che offrono al contempo una buona espandibilità ed una notevole velocità di lettura e scrittura, specialmente se supportati da interfacce appropriate.

I nastri, però, durano più a lungo. Se conservati nelle giuste condizioni, possono durare anche 30 anni ed hanno il vantaggio di essere facilmente trasportabili. [<https://www8.hp.com/it/it/campaigns/storage-media/lto-media.html>].

La vita media di un disco è di circa 5 anni (figura 4). [<https://www.backblaze.com/blog/hard-drive-stats-q2-2019/>].

Figura 4 Hard disk life-span
[\[https://www.prosofteng.com/blog/how-long-do-hard-drives-last/\]](https://www.prosofteng.com/blog/how-long-do-hard-drives-last/).

Figure 4 Hard disk life-span
[\[https://www.prosofteng.com/blog/how-long-do-hard-drives-last/\]](https://www.prosofteng.com/blog/how-long-do-hard-drives-last/).



Nella tabella 2 sono sintetizzati i principali vantaggi e svantaggi di hard disk e nastri.

| | Vantaggi | Svantaggi |
|-------------|--|---|
| Hard Drives | Basso Costo. Buona espandibilità. Velocità di lettura e scrittura. | Rumorosi. Produzione di Calore. Consumo energia. Fragilità. |
| Nastri | Basso costo. Basso consumo di energia. Alta Capacità. Basso ingombro. Facilità di gestione. Storage off-site. Facilità di rimozione e trasporto. Scalabilità. | Accurata manutenzione. Necessità di rotazione periodica. Ambiente di stoccaggio con precise caratteristiche. Unità di lettura/scrittura costose. |

Tabella 2 Vantaggi e svantaggi di Hard Drives e nastri.
Table 2 Hard Drives and tapes pros and cons.

| Environmental Factor | Operating | Operational Storage ¹ | Archival Storage ² | Shipping |
|-----------------------------------|--|----------------------------------|-------------------------------|------------------------------|
| Temperature | 10 to 45°C (50 to 113°F) | 16 to 32°C (61 to 90°F) | 16 to 25°C (61 to 77°F) | -23 to 49°C (-9 to 120°F) |
| Relative humidity (noncondensing) | 10 to 80% | 20 to 80% | 20 to 50% | 5 to 80% |
| Maximum wet bulb temperature | 26°C (79°F) | 26°C (79°F) | 26°C (79°F) | 26°C (79°F) |
| Magnetic field | Stray magnetic field at any point on tape not to exceed 50 oersteds (4000 Ampere/meter). | | | |

¹ Operational storage equals less than 6 months
² Archival storage equals greater than 6 months.

Tabella 3 Condizioni ambientali nastri LTO.
Table 3 LTO tapes operational environment.

Dischi e nastri necessitano di un idoneo spazio fisico di stoccaggio che rispetti i giusti parametri ambientali ed essere lontano da campi magnetici dannosi.

In tabella 3 sono riassunte le condizioni ambientali relative ai nastri di tipo LTO Ultrium, tratte dal sito IBM [https://www.ibm.com/support/knowledgecenter/sv/STCMML8/com.ibm.storage.ts3500.doc/ipg_3584_meesi.html].

Attualmente i nastri utilizzati presso la nostra sezione sono conservati in una stanza climatizzata atta allo scopo. In figura 5 sono rappresentati gli standard attualmente disponibili per i nastri LTO Ultrium.

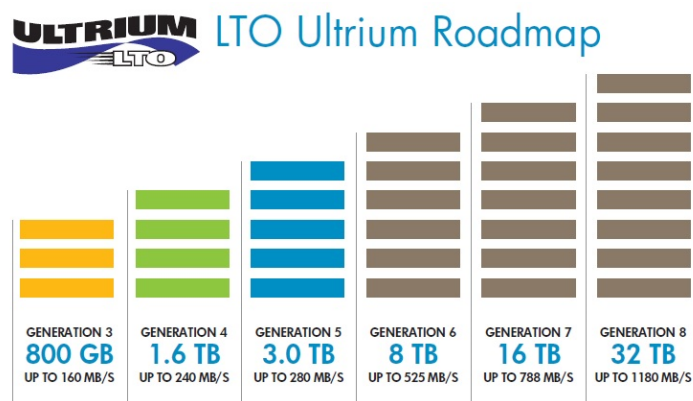


Figura 5 Standard LTO. [https://www8.hp.com/uk/en/pdf/hp-ltoultrium-datasheet_tcm_183_1268652.pdf].

Figure 5 LTO standard. [https://www8.hp.com/uk/en/pdf/hp-ltoultrium-datasheet_tcm_183_1268652.pdf].

1.3 La configurazione hardware

Scegliere la configurazione adatta alle proprie esigenze di backup non è una questione di poco conto. Analizzati costi, affidabilità, ridondanza, prestazioni, all'INGV-OV è stata implementata un'infrastruttura hardware costituita da due sistemi gemelli implementati su server Dell Poweredge R815 [<https://www.dell.com/it-it/work/shop/povw/poweredge-r815>], a cui sono collegati Array di storage iSCSI SAN PowerVault MD3200i [<https://www.dell.com/it-it/work/shop/productdetailstxn/powervault-md3200i>], con capacità di 36 TB ciascuno e MD1200 [<https://www.dell.com/en-us/work/shop/povw/powervault-md1200>] che montano array di dischi di tipo SAS con capacità di 48 TB ciascuno, in grado di garantire ottime performance.

Gli array di dischi sono organizzati secondo il livello di RAID 5. È stato scelto tale livello di ridondanza in quanto ampiamente collaudato ed affidabile. Inoltre presenta un elevato tasso di *fault tolerance* ed una buona velocità di lettura/scrittura.

I server sono affiancati da due tape library Dell PowerVault TL4000 collegate tramite interfaccia SAS [<https://www.dell.com/it-it/work/shop/cty/pdp/spd/powervault-tl4000?~ck=anav>] con 48 slots che utilizzano nastri ultrium6 da 2.5 TB ciascuno, che possono diventare 6.25 in caso di compressione dati.

Un altro sistema è costituito da un server HP Proliant DL360G6 al quale è collegata, tramite interfaccia iscsi, una libreria a nastri robotica HP msl2024 con 24 slots compatibile con nastri di tipo Ultrium3 da 400 GB (800 GB in caso di compressione dei dati) ciascuno ed è destinata ai backup di minori dimensioni.

I server sono affiancati, per backup non strategici, da server con prestazioni di livello inferiore, collegati ad unità di backup su dvd blu ray che hanno una capacità di 25 GB/disco.



Figura 6 Le Tape Library Dell PowerVault TL4000 e HP MSL2024.

Figure 6 Dell PowerVault TL4000 and HP MSL2024 Tape Libraries.

1.4 Il Software

Il software di backup è un'applicazione, o un insieme di applicazioni, destinata alla gestione dei processi di backup, ripristino, schedulazione, catalogazione dei supporti di archiviazione, gestione attraverso la rete delle macchine oggetto di backup, nonché tenere traccia dell'ubicazione dei file e della loro data di stoccaggio.

La distribuzione del software può avvenire secondo due modelli principali:

- SoP (Software on-premises);
- SAAS (software as a service o in Cloud computing).

Il **SoP** è il metodo tradizionale, che consiste nell'installazione del software direttamente in una macchina locale.

Presenta diversi vantaggi, quali:

- Controllo diretto sui sistemi;
- Gestione interna dei dati.

Lo svantaggio è rappresentato dal costo d'acquisto del software, anche se, come vedremo, esistono soluzioni open source. In tale infrastruttura i programmi si possono installare solo su macchine che soddisfino i requisiti di compatibilità hardware e software. **SAAS** è il metodo che si è andato affermando a partire dal 2005. Ne sono esempi Google Doc o Microsoft Office 365. Il software è distribuito da un provider di servizi e può essere utilizzato da qualsiasi client attraverso un web browser indipendentemente dalla sua configurazione.

L'infrastruttura hardware e software risiede esclusivamente presso il provider che la gestisce e ne assicura la sicurezza. L'organizzazione stipula con il provider un contratto di fornitura, solitamente a canone mensile. Non sono richieste risorse per la gestione dei sistemi, che sono facilmente scalabili ed offrono la possibilità di lavorare in mobilità da qualunque postazione che abbia accesso alla rete internet.

Attualmente nel nostro sistema, si utilizza la tipologia **SoP**. In particolare il software Bacula [<http://blog.bacula.org/>], di cui esiste sia una versione a pagamento che una versione opensource. Esso è costituito da un set di programmi che permettono di gestire le procedure di backup in una architettura Enterprise di tipo client/server attraverso la LAN. È compatibile con macchine con sistemi operativi differenti (Linux, Windows, Mac), e può utilizzare supporti fisici diversi, inclusi i nastri. Di facile scalabilità, Bacula consente di gestire reti di computer anche di grandi dimensioni. L'installazione e la configurazione non sono propriamente intuitivi, ma una volta familiarizzato con la sua struttura, risulta abbastanza agevole gestire la propria

strategia di backup ed essere operativi in tempi ragionevolmente brevi. Per le note di installazione si rimanda alla documentazione online di Bacula, reperibile all'indirizzo: [<http://blog.bacula.org/documentation/documentation/>].

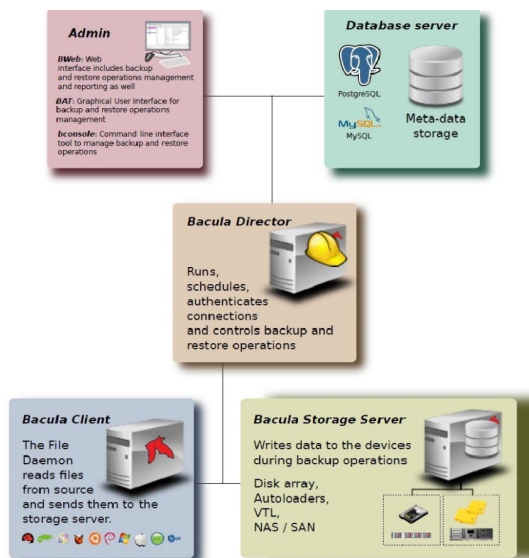


Figura 7 I componenti di Bacula [Kern Sibbald, “Bacula Main Reference”. <http://www.bacula.org/9.2.x-manuals/en/main/main.pdf>.

Figure 7 Bacula components [Kern Sibbald, “Bacula Main Reference”. <http://www.bacula.org/9.2.x-manuals/en/main/main.pdf>.

Nella figura 7 è schematizzata la struttura di Bacula. Il componente principale di Bacula è il “Director”, un servizio (o demone) da cui è possibile dirigere e controllare tutte le operazioni di backup, ripristino e configurazione.

Bacula Console service è lo strumento che permette di comunicare con il “Director”. È un programma a linea di comando, ma ne esistono anche versioni grafiche ed interfacce web come webmin (figura 8) [<http://www.webmin.com/>], che permette di modificare i file di configurazione e di amministrare quasi ogni aspetto di Bacula.

Bacula File service o Bacula client è installato sulle macchine di cui si deve eseguire il backup. Ne esistono diverse versioni in dipendenza del sistema operativo “ospite”.

Bacula Storage services permette la comunicazione con i dispositivi su cui avverrà lo storage finale dei dati. Coordina le attività di lettura e scrittura sui supporti fisici, siano essi DVD, dischi o nastri.

Bacula conserva le informazioni relative ai backup in un database, che può essere Postgres [www.postgresql.org] o Mysql [<https://www.mysql.com/it/>]. Della gestione di tale database si occupano i Catalog services.

Infine il servizio Bacula Monitor controlla lo stato di funzionamento di tutti i servizi.

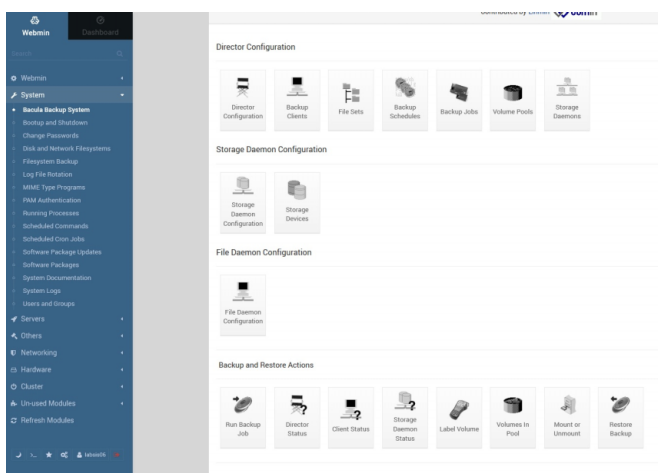


Figura 8 Bacula l'interfaccia webmin.

Figure 8 Bacula webmin interface.

1.5 Tipi di backup

Le più comuni tipologie di backup sono:

- **Full backup.** Semplicemente si effettua una copia completa di tutti i dati.
- **Backup differenziale.** Si effettua una copia di tutti i dati modificati dall'ultimo backup completo, includendo le modifiche precedentemente salvate.
- **Backup incrementale.** Si effettua una copia dei dati modificati dal backup più recente, sia esso completo od incrementale, senza includere tutte le precedenti modifiche.

Inoltre si possono distinguere:

- **Cold backup.** Si effettua quando i dati non sono utilizzati da alcun utente.
- **Hot backup.** Si effettua anche quando i dati sono accessibili agli utenti.

In figura 9 è schematizzato il backup differenziale. È evidente che la dimensione dei backup differenziali cresce in maniera esponenziale, dal momento che ogni backup successivo include anche tutti i precedenti. Per il ripristino di dati salvati secondo tale schema bastano l'ultimo backup completo e l'ultimo differenziale.

In figura 10 è schematizzato il backup incrementale. I backup successivi sono di dimensioni minori, infatti includono solo i file modificati dall'ultimo backup incrementale.

Per il ripristino di dati salvati secondo tale schema, però, occorrono l'ultimo backup completo e tutti gli incrementali a partire da quest'ultimo. Figura 10 Il backup incrementale.

Figura 9 Il backup differenziale.

Figure 9 Differential backup.

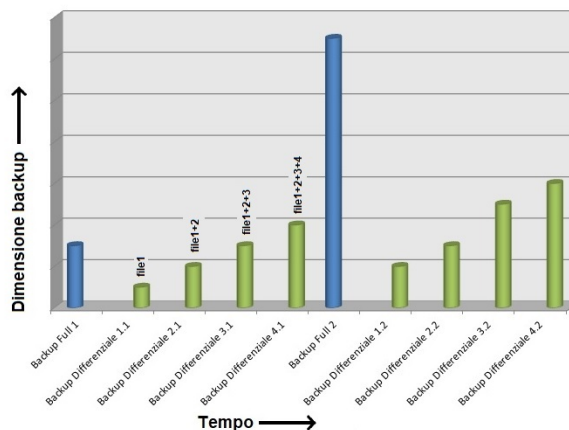
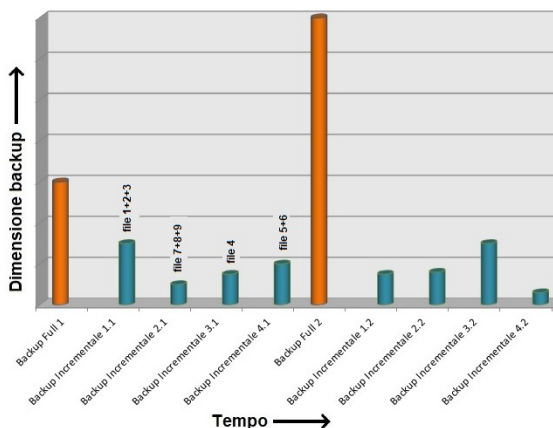


Figura 10 Il backup incrementale.

Figure 10 Incremental backup.



Quando si progetta un piano di backup su nastri, è utile prevedere uno schema di rotazione degli stessi. Una delle strategie più semplici da utilizzare è la cosiddetta GFS (Grandfather - Father - Son), che si basa su tre diversi gruppi di backup:

- Mensili (Nonni): backup completo da riporre possibilmente offline dopo l'esecuzione.
- Settimanali (Padri): backup completo o differenziale.
- Giornalieri (Figli): backup incrementali o differenziali.

Nella figura 11 è mostrato un esempio dell'applicazione di tale metodo.

Un' altra strategia è quella chiamata Torre di Hanoi, organizzato in sessioni e livelli di backup.

| Lunedì | Martedì | Mercoledì | Giovedì | Venerdì | Sabato | Domenica |
|--------|---------|-----------|---------|---------|--------|----------|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| 22 | 23 | 24 | 25 | 26 | 27 | 28 |
| 29 | 30 | 31 | 1 | 2 | | |

Figura 11 Un esempio di schema di backup GFS.

Figure 11 A GFS backup scheme example.

| | |
|--|--------|
| | Padre |
| | Figlio |
| | Nonno |

Nella figura 12 è mostrato un esempio che prevede 16 sessioni e cinque livelli di backup. Il livello A è costituito da backup incrementali, i livelli B, C e D sono backup differenziali, mentre il livello E è un backup completo.

In pratica viene conservato solo il backup più recente per ciascun livello, mentre gli altri sono cancellati ed i relativi nastri riutilizzati. L'eliminazione è rimandata nel caso in cui un backup del livello precedente non è ancora stato completato.

| Sessione \ Backup | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|-------------------|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|
| 1 | A | | A | | A | | A | | A | | A | | A | | A | |
| 2 | | B | | | | B | | | | B | | | | B | | |
| 3 | | | | C | | | | | | | | C | | | | |
| 4 | | | | | | | | D | | | | | | | | |
| 5 | | | | | | | | | | | | | | | | E |

Figura 12 La strategia di backup "Torre di Hanoi".

Figure 12 The "Tower of Hanoi" backup strategy.

Presso la sezione di Napoli, attualmente, la rotazione dei nastri segue lo schema GFS, con backup "Nonni" mensili stoccati offline ad intervalli bimestrali, backup differenziali "Padri" a cadenza settimanale e backup incrementali "Figli" a cadenza giornaliera per i dati che subiscono modifiche continue.

È stato scelto tale schema in quanto la gestione dei nastri è più semplice e rapida rispetto ad altri schemi come, ad esempio, la "Torre di Hanoi".

Dal punto di vista strutturale si distinguono diverse tipologie di backup, le principali sono:

- BoP (Backup on-Premises).
- BAAS (Backup As A Service).
- STaaS (Storage as a Service).

BoP è il metodo più classico di backup. I dati sono stoccati su macchine locali oppure supporti fisici di backup, quali nastri, DVD o hard disks; implica l'implementazione di infrastrutture hardware e necessita di numerosi supporti fisici.

A fronte di una elevata disponibilità dei dati ed un basso tempo di recovery, il metodo BoP necessita di una corretta gestione delle macchine e dei supporti per evitare perdite di dati.

BAAS è un concetto simile al SAAS già analizzato per il software. L'infrastruttura hardware è messa a disposizione dal provider, offre una elevata scalabilità ed un rapido accesso ai dati da qualunque postazione collegata alla rete a fronte del pagamento di un canone.

StaaS rappresenta una via di mezzo fra i due metodi descritti sopra. L'infrastruttura hardware appartiene all'organizzazione che sfrutta lo spazio di archiviazione offerto sul cloud da un provider (ad esempio Google Drive). Nella tabella 4 sono riassunti i principali vantaggi e svantaggi delle metodologie BaaS e BoP.

| | Vantaggi | Svantaggi |
|------|--|---|
| BaaS | <p>Nessun costo Hardware. Accesso ai dati da qualunque postazione. Dati protetti in appositi Data Centers. Considerevole riduzione del personale dedicato al processo di backup. Scalabilità verso l'alto e il basso semplificata. Backup e ripristino dei dati integrato e automatico. Rischio di downtime "near zero".</p> | <p>Canone mensile che aumenta con la quantità dei dati gestiti. Nessuno o scarso controllo sui processi di backup. Problemi di riservatezza e sicurezza dei dati, Maggior tempo di recovery.</p> |
| BoP | <p>Gestione diretta Hardware. Sistemi ad hoc. Gestione diretta dei dati. Rapido accesso ai dati. Basso tempo di recovery. Performance migliori. Accesso dati offline.</p> | <p>Costo Infrastruttura Hardware. Personale dedicato. Gestione supporti fisici.</p> |

Tabella 4 BaaS e BoP - Vantaggi e svantaggi.

Table 4 BaaS and BoP – Pros and cons.

La cosiddetta "nuvola" permette di ridurre la complessità del sistema IT proprietario e, nel contempo, di avere sistemi facilmente scalabili. I providers si occupano di garantire la protezione e curare la deduplicazione dei dati. I processi gestionali sono facilitati e velocizzati e l'accesso ai dati è distribuito e può avvenire tramite qualunque postazione.

Bisogna considerare, però, oltre a quelli già citati, un aspetto negativo di non poco conto. Il provider potrebbe sperimentare difficoltà finanziarie e ridurre o cessare l'erogazione di alcuni servizi, o subire attacchi informatici che metterebbero a repentaglio l'integrità e la riservatezza dei dati stoccati sui loro server. Per approfondimenti su tale aspetto si può iniziare consultando: [Y. Izrailevsky and C. Bell, "Cloud Reliability" in *IEEE Cloud Computing*, vol. 5, no. 03, pp. 39-44, 2018. doi:10.1109/MCC.2018.032591615 url: <https://doi.ieeecomputersociety.org/10.1109/MCC.2018.032591615>].

2. Il costo del backup

Ma qual è il costo del processo di backup? Dare una risposta precisa non è facile; molti sono i fattori che concorrono nel calcolo, i principali sono:

- Il costo dei supporti fisici (dischi, nastri, cd);
- Il costo di acquisto dell'hardware;

- Il costo del software;
- Il costo di assistenza ed aggiornamento hardware e software;
- Il costo dell'energia elettrica;
- Il costo per il condizionamento dell'aria nel centro di calcolo.

Nella tabella 5 è indicato il costo dei supporti, basati sui dati provenienti dalla tabella 1.

| | COSTO* Per anno | * Prezzi rilevati a Luglio 2019 su Mepa [https://www.acquistinretepa.it/opencms/opencms/]. Iva esclusa |
|----------------------------------|------------------------|--|
| ULTRIUM 6 (Native) 2,5 TB | € 125,16 | 20,86 €/unità Iva Esclusa C7976A HPE Ultrium RW Data Cartridge - LTO Ultrium 6 |
| ULTRIUM 3 (Native) 400 GB | € 261,24 | 21,77 €/unità Iva Esclusa LTX400GN SONY LTO ULTRIUM 3 |
| BLU-RAY 2,5 GB | € 408 | 3,40 €/unità Iva Esclusa |

Tabella 5 Costo dei supporti fisici.

Table 5 Physical media cost.

Per avere un ordine di grandezza sui costi ipotizziamo una configurazione, che lavori per un intervallo di tempo di 5 anni, composta da:

- 2 Unità a nastro Dell PowerVault TL4000 di cui una impegnata in Full Load e l'altra in Idle;
- 1 server Dell PowerEdge R815 con attività transazionale e carico CPU al 70%;
- Array di dischi Dell PowerVault MD3400 entrambe in Full Load.

Questa configurazione differisce da quella implementata e descritta nel paragrafo 1.3 per la sostituzione del modello PowerVault MD3400 al posto dei modelli PowerVault MD3200i e MD1200 che sono fuori produzione e pertanto per essi non esistono dati comparabili.

Il marchio Dell è stato scelto per avere una stima il più possibile aderente alla reale configurazione adottata e descritta nel paragrafo 1.3.

Il periodo di 5 anni è l'MTBF (Mean Time Between Failure) indicato da Dell per la maggior parte dei suoi server, anche se una valutazione precisa è difficile da fare, in virtù dell'estrema variabilità delle configurazioni, carichi di lavoro e condizioni ambientali cui le macchine sono sottoposte. In figura 13 è mostrato lo schema riassuntivo della nostra ipotesi di configurazione. I dettagli relativi alle singole macchine si possono consultare sul sito Dell [<http://dell-ui-eipt.azurewebsites.net/#/>].

In tabella 6 lo schema del costo dell'energia elettrica, desunto dalla figura 10 e corrispondente all'offerta Luglio 2019 di ENEL e-light Impresa [<https://www.enel.it/it/luce-e-gas/luce/offerte/elight-luce-impresa>]. È stata considerata una potenza calcolata come media aritmetica tra la potenza di input e la potenza massima.

La tabella 7 offre uno sguardo di insieme sui costi di acquisto delle macchine, che sono stati divisi per i cinque anni di funzionamento ipotizzati.

I risultati dell'analisi dei costi di gestione sono mostrati nella tabella 8. I dati dei costi di condizionamento dell'aria non sono disponibili; il software non comporta costi in quanto sono state utilizzate solo applicazioni Open Source quali Linux e Bacula.

Il costo di assistenza è una stima basata sugli interventi effettuati negli ultimi due anni.

Figura 13 Riassunto configurazione.

Figure 13 Configuration summary.



| | | |
|---|--------------|--------------|
| Input power | 1028 watts | 3507.7 btu/h |
| Max potential power | 2199.9 watts | |
| Input current | 4.7 amps | |
| Airflow rate | 55.4 CFM | 26.1 l/s |
| Weight | 93.5 kg | |
| Air temperature rise | 54 °C | |
| Sound power level | 7.3 bels | |
| Kilowatt hour cost (ENEL e-light impresa) | 0.0735 €/Kwh | |

Tabella 6 Costo - Energia.
Table 6 Cost - Energy.

| | |
|-------------------------|---------|
| Potenza Input (W) | 1028 |
| Potenza Max (W) | 2199.9 |
| Potenza Media (W) | 1613,95 |
| Energia (kWh) | 1,61 |
| Energia/Anno (kWh/anno) | 14103,6 |
| Costo/anno (€/anno) | 1036,61 |

| Modello | Costo Di Acquisto (€) | Unità | Totale (€) | Costo Annuo (€/anno) |
|-------------------|-----------------------|----------|-----------------|----------------------|
| PowerEdge R815 | 2627,12 | 1 | 2627,12 | 525,42 |
| PowerVault MD3400 | 18737,77 | 2 | 37475,5 | 7495,11 |
| PowerVault TL4000 | 4869,66 | 2 | 9739,32 | 1947,86 |
| TOTALI | | 5 | 49841,94 | 9968,39 |

Tabella 7 Costo annuale macchine.
Table 7 Hardware annual cost.

| | |
|---|-----------------|
| Costo acquisto annuo (€) | 9968,39 |
| Costo Energia annuo (€) | 1036,61 |
| Costo supporti annuo (€) | 125,16 |
| Costo Condizionamento aria annuo (€) | NA |
| Costo del software annuo (€) | 0 |
| Costo di assistenza hard e soft annuo (€) | 1000 |
| Totale | 12130,16 |

Tabella 8 Costo totale del sistema.
Table 8 System total cost.

2.1 Il costo del cloud

Analizziamo quali sono i prezzi di due big competitors del settore:

- Microsoft Azure;
- Google.

Microsoft Azure

Microsoft Azure è un set di servizi cloud professionali. Permette di creare, gestire e distribuire applicazioni su una rete globale di dimensioni elevate. [<https://azure.microsoft.com/it-it/overview/what-is-azure/>].

In tabella 9 l'analisi dei costi per l'implementazione di una macchina virtuale e di uno spazio di archiviazione di 36 TB (si veda tabella 1).

Per informazioni sulle opzioni di replica di Microsoft Azure, consultare:

[<https://docs.microsoft.com/it-it/azure/storage/common/storage-redundancy>].

Google Cloud

Anche Google Cloud, al pari di Microsoft, offre la sua piattaforma di servizi destinata ad utenti business.

In tabella 10 l'analisi dei costi per servizi simili a quelli di Azure.

3. Disponibilità ed affidabilità dei sistemi di storage

Si definisce:

- **Affidabilità:** La probabilità che il sistema compia le funzioni a cui è preposto senza guastarsi per un certo periodo di tempo con un dato livello di confidenza.
- **Disponibilità:** La probabilità che il sistema sia attivo e funzioni correttamente in un dato momento.

Per calcolare la disponibilità e l'affidabilità del nostro sistema, facciamo riferimento ad uno studio condotto dalla Dell su alcuni sistemi-tipo, alcuni dei quali sono sufficientemente simili alla nostra configurazione, permettendo una comparazione.

Per approfondimenti: [http://www.dell.com/content/topics/global.aspx/power/en/ps3q02_shetty?c=us].

Lo studio prende in considerazione componenti disposti secondo alcune configurazioni-tipo.

L'hardware preso in esame è costituito da:

- Dell PowerEdge 6450 server;
- QLogic® QLA 220 Host Bus Adapter (HBA);
- Dell/EMC FC4700 storage array;
- Dell PowerVault™ 56F Fibre Channel switch.

3.1 Le configurazioni analizzate

Di seguito le configurazioni prese in esame.

Non-redundant Direct attach storage (DAS).

Hardware:

- Un server PowerEdge 6450;
- Un HBA;
- Un FC4700 storage array.

Microsoft Azure Estimate

| Service type | Region | Description | Estimated Cost |
|---|-------------|--|--|
| Virtual Machine | West Europe | 1 D2 v3 (2 vCPU, 8 GB di RAM) Windows (solo sistema operativo) Riservata 1 anno 0 dischi del sistema operativo gestiti S4 100 unità di transazione | € 103,76 |
| Storage Accounts | West Europe | Archiviazione BLOB in blocchi Utilizzo generico V2 LRS ridondanza Livello di accesso frequente 36000 GB capacità 100.000 operazioni di lettura 1 operazioni di altro tipo 1000 GB recupero dati 1000 GB scrittura dati | |
| Support | | Support | € 0,00 |
| | | Licensing Program | Microsoft Online Services Program (MOSP) |
| | | Monthly Total | € 699,74 |
| | | Annual Total | € 8396,87 |
| Disclaimer All prices shown are in Euro (€). This is a summary estimate, not a quote. For up to date pricing information please visit https://azure.microsoft.com/pricing/calculator | | | |

Tabella 9 Stima dei costi di Microsoft Azure.

Table 9 Microsoft Azure cost estimate.

Google Cloud Estimate

| Service type | Region | Description | Estimated Cost |
|--|-------------|---|----------------|
| Virtual Machine | West Europe | 2 vCPU 8 GB di RAM Windows (solo sistema operativo) 730 hours per month VM class: regular Instance type: n1-standard-2 Region: Belgium Paid OS Cost: € 51,09 GCE Instance Cost: € 46,76 Total available local SSD space 3x375 GB | € 30,83 |
| Storage | West Europe | Belgium Regional Storage: 36.000 GB | € 839,12 |
| | | Monthly Total | € 869,96 |
| | | Annual Total | € 10.439,52 |
| Disclaimer For up to date pricing information please visit https://cloud.google.com/products/calculator . The estimated fees provided by the Cloud Pricing Calculator are for discussion purposes only and are not binding you or Google. Your actual fees may be higher or lower then the estimate. A more detailed and specific list of fees will be provided at time of sign-up. | | | |

Tabella 10 Stima dei costi di Google Cloud.

Table 10 Google Cloud cost estimate.

La figura 14 mostra la configurazione del sistema.

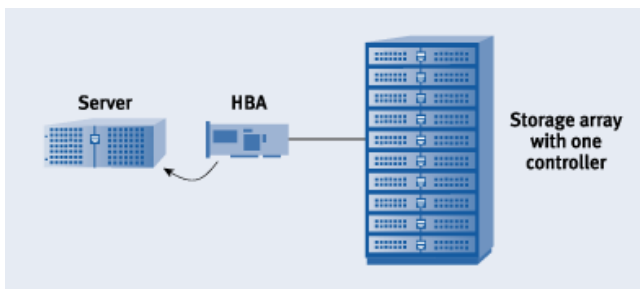


Figura 14 Configurazione Non-redundant DAS.

Figure 14 Non-redundant DAS configuration.

Redundant Direct attach storage (DAS)

Hardware e configurazione (figura 15):

- Un server PowerEdge 6450;
- Due HBA;

Basic cluster

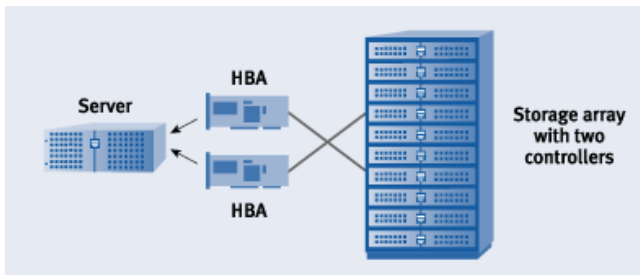


Figura 15 Configurazione Redundant DAS.

Figure 15 Redundant DAS configuration.

Hardware e configurazione (figura 16):

- Due server PowerEdge 6450;
- Due HBA (uno per ogni server);
- Un FC4700 storage array con due controller.

Questa configurazione è stata implementata attraverso Microsoft® Cluster Service [<https://docs.microsoft.com/it-it/windows-server/failover-clustering/create-failover-cluster>] [<https://docs.microsoft.com/it-it/windows-server/failover-clustering/failover-clustering-overview>].

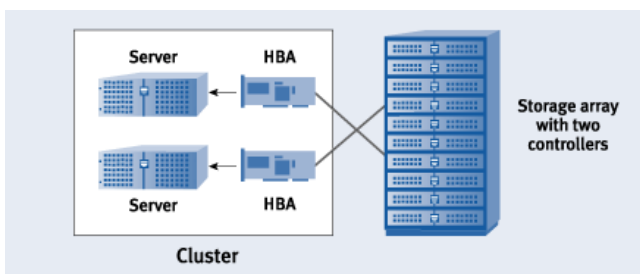


Figura 16 Configurazione Basic Cluster.

Figure 16 Basic Cluster configuration.

High-availability cluster.

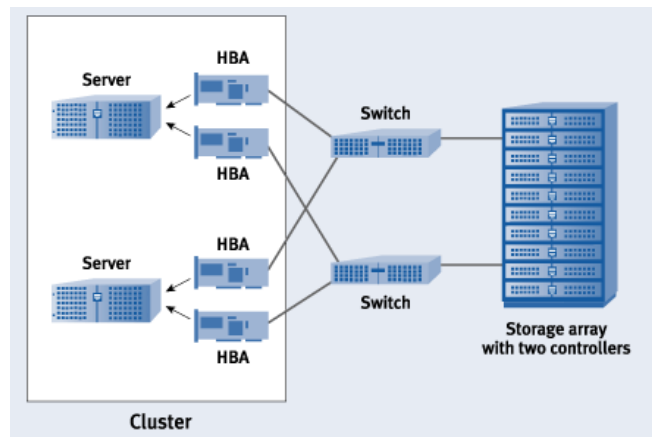
Hardware e configurazione (figura 17):

- Due server PowerEdge 6450;
- Quattro HBA (due per ogni server);
- Un FC4700 storage array con due controller;
- Due PowerVault 56F switch.

Anche in questa configurazione è stato utilizzato Microsoft Cluster Service.

Figura 17 Configurazione High Availability cluster.

Figure 17 High Availability cluster configuration.



Simple SAN (Storage Area Network) (SAN)

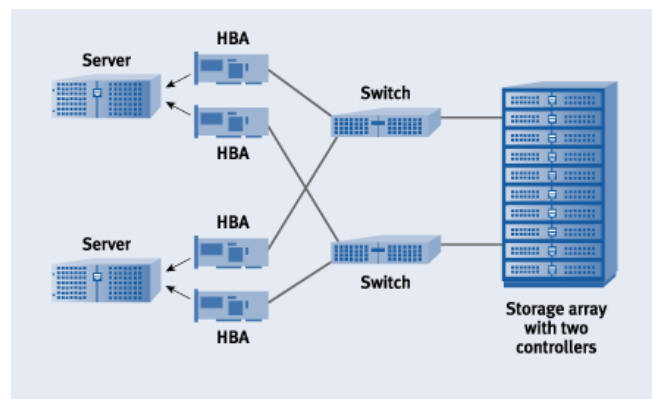
Hardware e configurazione (figura 18):

- Due server PowerEdge 6450;
- Quattro HBA (due per ogni server);
- Un FC4700 storage array con due controller;
- Due PowerVault 56F switch.

SAN (Storage Area Network) è un network ad alte prestazioni in cui i dati sono trasferiti dai server alle risorse di storage. I due server non sono “clustered”.

Figura 18 Configurazione Simple SAN.

Figure 18 Simple SAN configuration.



3.2 Il calcolo dell'affidabilità, disponibilità ed MTBF

In una indagine statistica possiamo definire come frequenza relativa, il valore ottenuto dividendo la frequenza assoluta per la dimensione del campione che stiamo esaminando.

Il valore della frequenza relativa è sempre compreso nell'intervallo [0,1] e la somma delle frequenze relative è sempre 1.

È possibile descrivere il campione con le frequenze cumulative. La frequenza cumulativa relativa è la sommatoria delle frequenze relative.

Se assimiliamo le frequenze relative alla probabilità di accadimento di un evento dato, possiamo definire come **funzione di ripartizione** la funzione che fa corrispondere ai valori di x , le probabilità cumulate $P(X \leq x)$, con X che rappresenta la nostra variabile aleatoria continua.

La Funzione di Ripartizione è:

- definita per ogni valore di x in \mathbb{R} ;
- assume valori compresi fra 0 ed 1;
- la funzione è monotona non decrescente.

In generale possiamo affermare che la Funzione di Ripartizione, $F(x)$, per una variabile aleatoria continua X esprime la probabilità che X non superi il valore x

$$F(x) = P(X \leq x)$$

In figura 19 è rappresentata la funzione di ripartizione.

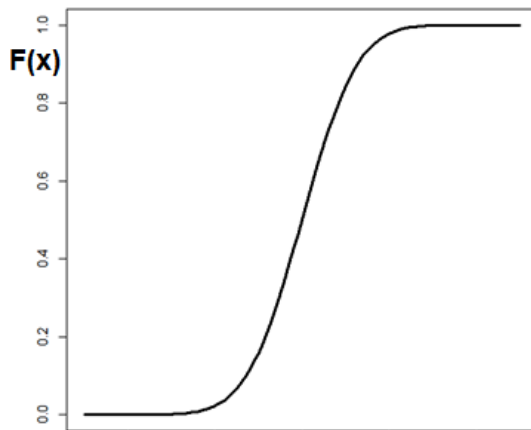


Figura 19 Rappresentazione grafica della funzione di ripartizione $F(x)$.

Figure 19 Distribution function graph $F(x)$.

Nel nostro caso la funzione di ripartizione è $F(t)$ e rappresenta la probabilità (cumulativa) che il tempo del primo guasto sia inferiore od uguale al tempo t .

Sarà:

$$0 \leq F(t) \leq 1$$

Calcolando la derivata della funzione di ripartizione, si ottiene una nuova funzione detta densità di probabilità. Indichiamo tale funzione con $f(x)$.

$$f(x) = F'(x)$$

Tale funzione ha le seguenti proprietà:

- $f(x) > 0$ per qualunque valore di x ;
- l'area totale sottesa alla funzione è 1;
- la probabilità che la variabile aleatoria assuma valori in un intervallo, è uguale all'area sottesa dalla funzione relativa all'intervallo di valori dato.

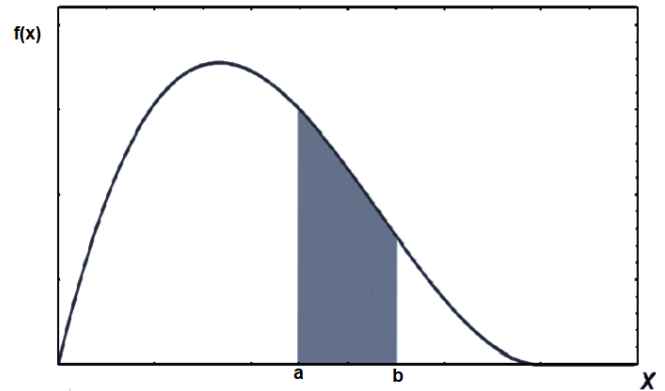
Nella figura 20 la probabilità che la variabile aleatoria assuma valori nell'intervallo $[a,b]$ è data dall'area sottesa alla curva colorata di grigio.

Nel nostro caso la funzione densità di probabilità rappresenta la funzione di densità di primo guasto e sarà:

$$f(t) = \frac{dF(t)}{dt}$$

Figura 20 Grafico della funzione densità di probabilità.

Figure 20 Probability density function graph.



Indichiamo con:

- $N_f(t)$ il numero di componenti funzionanti all'istante t ;
- $N_g(t)$ il numero di componenti guasti all'istante t .

Si dovrà avere per ogni t :

$$N_g(t) + N_f(t) = N_0 \quad (a)$$

Dove N_0 è il numero totale di componenti.

$F(t)$ rappresenta l'inaffidabilità di un componente. È una probabilità ed è la funzione di ripartizione cumulativa della probabilità di guasto.

La probabilità che un certo evento si verifichi è data dal rapporto tra il numero di esiti favorevoli a tale evento ed il numero totale di eventi possibili, si può, quindi, definire:

Inaffidabilità $F(t)$, come la probabilità per il componente di essere guasto al tempo t :

$$F(t) = \frac{N_g(t)}{N_0}$$

Affidabilità (Reliability) $R(t)$, come la probabilità per il singolo componente di essere ancora funzionante al tempo t :

$$R(t) = \frac{N_f(t)}{N_0}$$

Dividendo i due membri della (a) per N_0 si ottiene:

$$F(t) + R(t) = 1$$

e quindi:

$$R(t) = 1 - F(t)$$

A titolo di esempio, calcoliamo la probabilità che nell'intervallo $[0, t]$ non avvengano guasti, cioè l'affidabilità in tale intervallo:

Possiamo scrivere:

$$R(t) = \int_{-\infty}^{+\infty} f(t)dt - \int_0^t f(t)dt = \int_t^{\infty} f(t)dt$$

Definiamo **tasso di guasto (λ) o failure rate**, il rapporto tra la probabilità di prima *failure* e la probabilità di *reliability* al tempo t . Cioè la probabilità che un componente, che al tempo t si trovi in stato di buon funzionamento, si guasti in un intervallo di tempo compreso tra t e $t + dt$. In pratica si può scrivere come il rapporto tra elementi che si guasteranno nell'intervallo $(t; t+dt)$, rispetto al numero di elementi ancora funzionanti al tempo t :

$$\lambda(t) = \frac{1}{N_f(t)} \cdot \frac{dN_g(t)}{dt} = \frac{1}{N_f(t)} \cdot \frac{N_0}{N_0} \cdot \frac{dN_g(t)}{dt}$$

Ma:

$$\frac{N_0}{N_f(t)} = \frac{1}{R(t)}$$

e:

$$\frac{dN_g(t)}{N_0} = dF(t) \Rightarrow \lambda(t) = \frac{1}{R(t)} \cdot \frac{dF(t)}{dt} = \frac{f(t)}{R(t)}$$

L'andamento nel tempo di $\lambda(t)$ segue generalmente la cosiddetta "curva a vasca da bagno" (*bathtub curve*) (figura 21), caratterizzata da tre periodi:

- un primo periodo (mortalità infantile) con $\lambda(t)$ decrescente esponenzialmente. Si riscontra nel periodo di messa a punto e collaudo di nuovi componenti;
- un secondo periodo con $\lambda(t)$ relativamente costante (vita utile);
- un terzo periodo (obsolescenza) con $\lambda(t)$ che aumenta rapidamente a causa dell'invecchiamento dei componenti.

Nella pratica i componenti sono utilizzati durante il periodo di vita utile. Poichè in questo periodo il valore λ può essere considerato costante, l'espressione dell'affidabilità può essere riscritta:

$$\lambda(t) = \frac{1}{N_f(t)} \cdot \frac{dN_g(t)}{dt} = \frac{1}{N_f(t)} \cdot \frac{d[N_0 - N_f(t)]}{dt} = \frac{d[N_f(t)]}{N_f(t)dt}$$

$$\Rightarrow \int_{\delta}^{\delta} \lambda(t)dt = \int_{\delta}^{\delta} \frac{d[N_f(t)]}{N_f(t)dt} dt = \int_{\delta}^{\delta} \frac{d[N_f(t)]}{N_f(t)} = - \int_{\delta}^{\delta} \frac{1}{N_f(t)} d[N_f(t)]$$

Considerando che:

$$N_f(0) = N_0 \Rightarrow \lambda(t) = -[\ln N_f(t) - \ln N_0] = - \ln \left[\frac{N_f(t)}{N_0} \right]$$

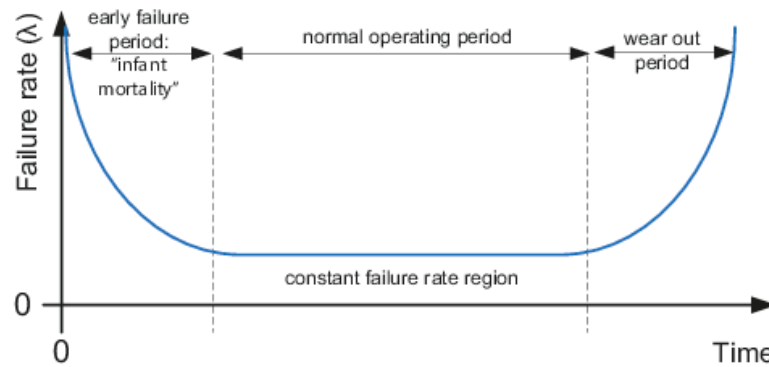


Figura 21 Curva a “vasca da bagno.” [https://www.researchgate.net/figure/Typical-failure-rate-curve-over-lifetime_fig2_299356707].

Figure 21 Bathtub curve. [https://www.researchgate.net/figure/Typical-failure-rate-curve-over-lifetime_fig2_299356707].

Cioè:

$$e^{-\lambda(t)} = \frac{N_f(t)}{N_0} = R(t) \quad (b)$$

I parametri utilizzati per valutare l'affidabilità e la disponibilità dei componenti sono numerosi. Oltre quelli descritti precedentemente, i principali sono:

- MTTF (Mean Time To Failure);
- MTTR (Mean Time To Repair);
- MTBF (Mean Time Between Failure);
- MUT (Mean Up Time);
- MDT (Mean Down Time).

MTTF rappresenta il tempo medio fra l'istante zero, in cui il componente è funzionante, e l'istante del suo primo guasto: Di solito è usato per sistemi non riparabili.

MTTR esprime il tempo medio che intercorre tra l'insorgenza di un guasto ed il completamento della sua riparazione.

MTBF è che il tempo medio che intercorre tra due guasti ed è ovviamente applicabile solo a componenti riparabili.

MUT è il tempo medio di funzionamento.

MDT è il tempo medio di indisponibilità di un dato componente ed è dato dalla somma del tempo di scoperta del guasto, del tempo di riparazione e del tempo di rimessa in servizio. In figura 22 sono riassunti i principali parametri di affidabilità e disponibilità.

Calcoliamo l'MTTR del componente i-esimo ($MTTR_i$), attraverso i seguenti passaggi:

La disponibilità di un componente i-esimo (A_i) può essere definita come:

$$A_i = \frac{MUT_i}{MUT_i + MDT_i} \quad (\text{si veda figura 22})$$

Ma:

$MUT + MDT = MTBF$ di conseguenza:

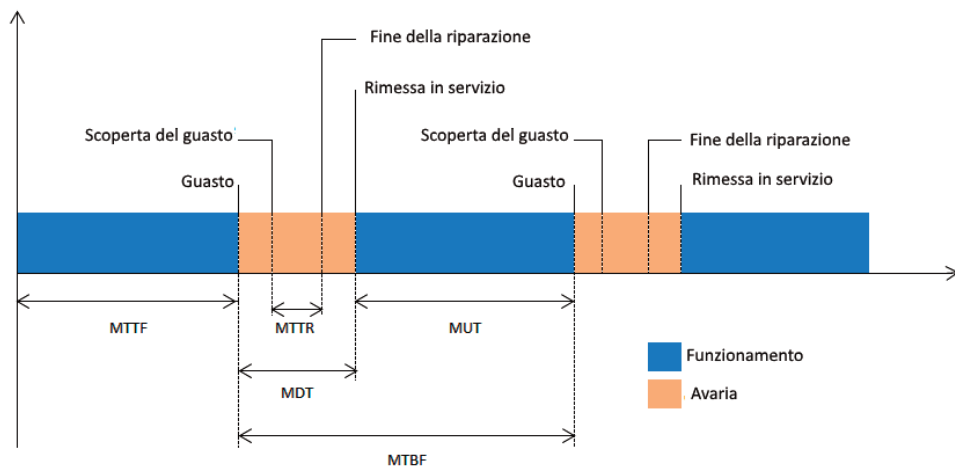


Figura 22 Principali parametri di affidabilità e disponibilità, da [http://jackadit.com/index.php?page=indus3].

Figure 22 Reliability and availability main parameters [http://jackadit.com/index.php?page=indus3].

$$A_i = \frac{MUT_i}{MUT_i + MDT_i}$$

sommando a tutti i termini la quantità MDT e considerando che, nel caso di sistemi altamente affidabili e/o riparabili rapidamente, si ha $MDT \cong MTTR$:

$$A_i = \frac{MUT_i + MDT_i}{MTBF_i + MDT_i} = \frac{MTBF_i}{MTBF_i + MTTR_i} \tag{c}$$

Il tasso di guasto per ora può essere definito anche come l'inverso dell'MTBF:

$$\lambda = \frac{1}{MTBF} \tag{d}$$

Esempio: se un apparato ha $\lambda = 10^{-5}$ guasti/ora, avrà un $MTBF = 100.000$ ore.

Sostituendo nella (b), otteniamo:

$$A_i = \frac{1}{1 + \lambda_i MTTR_i}$$

Esplicitando rispetto ad $MTTR_i$ otteniamo:

$$MTTR_i = \frac{1 - A_i}{A_i \lambda_i}$$

O l'equivalente:

$$MTTR_i = \frac{MTBF_i - A_i MTBF_i}{A_i} \tag{e}$$

Dove: $MTBF_i$ è l'MTBF del componente i-esimo; A_i è la disponibilità del componente i-esimo.

MTTR ed A per l'intero sistema ($MTTR_{sys}$ ed A_{sys}) sono, ovviamente, dati dalle formule:

$$MTTR_{sys} = \sum_{i=1}^n MTTR_i$$

$$A_{sys} = \sum_{i=1}^n A_i$$

Attraverso le formule (b), (c), (e), si ottengono i valori riassunti nelle tabelle 11, 12 e 13.

Le figure di seguito sono tratte dal sito Dell. [http://www.dell.com/content/topics/global.aspx/power/en/ps3q02_shetty?c=us].

Tabella 11 MTTR e indice di guasto per i singoli componenti del sistema.

Table 11 MTTR and failure index for the system individual components.

| | MTTR _i (hours) | Failure rate λ _i (failures/hour) | MTTR _i × failure rate (failures) |
|------------------------------|------------------------------|--|--|
| FC4700 storage array | 3.99997 | 1.4501E-5 | 5.8003E-5 |
| PowerVault 56F switch | 7.66 | 3.91E-6 | 3.00009E-5 |
| QLA 2200 HBA | 8.00 | 3.95E-6 | 3.16E-5 |
| PowerEdge 6450 | 4.75 | 2.18E-5 | 0.000104011 |
| Sum | | 4.4233E-5 | 0.00022369 |

Tabella 12 Affidabilità dei singoli componenti del sistema.

Table 12 System individual components reliability.

| | Mission reliability | | |
|------------------------------|---------------------|---------|---------|
| | 1 year | 2 years | 3 years |
| FC4700 storage array | 0.8807 | 0.7756 | 0.6831 |
| PowerVault 56F switch | 0.9662 | 0.9336 | 0.9022 |
| QLA 2200 HBA | 0.9659 | 0.9329 | 0.9011 |
| PowerEdge 6450 | 0.8257 | 0.6818 | 0.5630 |

Tabella 13 Stima di Disponibilità (A_i) ed MTBF dei singoli componenti del sistema.

Table 13 System individual components Availability (A_i) and MTBF.

| | FC4700 storage array | PowerVault 56F switch | QLA 2200 HBA | PowerEdge 6450 server |
|--|-------------------------|--------------------------|-----------------|--------------------------|
| Availability (A_i) | 0.999942 | 0.9999 | 0.99996 | 0.9998 |
| MTBF (hours) | 68961 | 255358 | 252550 | 45753 |

Le figure 23, 24 e 25 mostrano i risultati finali.

La nostra configurazione può essere assimilata a quella "Basic cluster".

Come ci si poteva aspettare il tempo medio tra i guasti, l'affidabilità e la disponibilità dei sistemi decresce nel corso dei tre anni di mission, ma risulta anche che la nostra configurazione appare sufficientemente equilibrata rispetto alle altre considerate.

Ciò che appare chiaro è che affidabilità e disponibilità dei sistemi aumenta in rapporto all'aumento della ridondanza dei sottosistemi sia della loro "clusterizzazione".

Ovviamente tali aumenti generano un incremento dei costi di gestione.

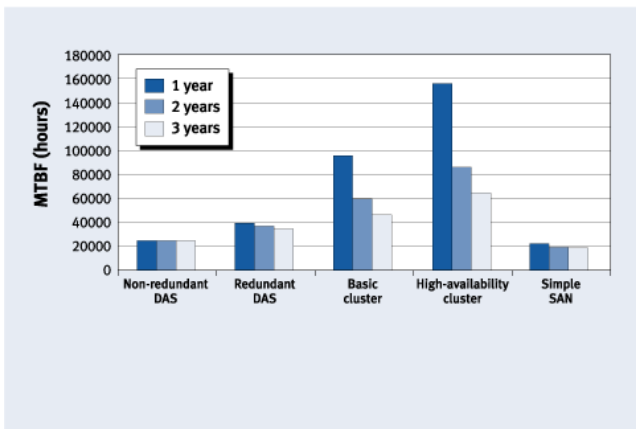


Figura 23 MTBF dei differenti sistemi.

Figure 23 Systems MTBF.

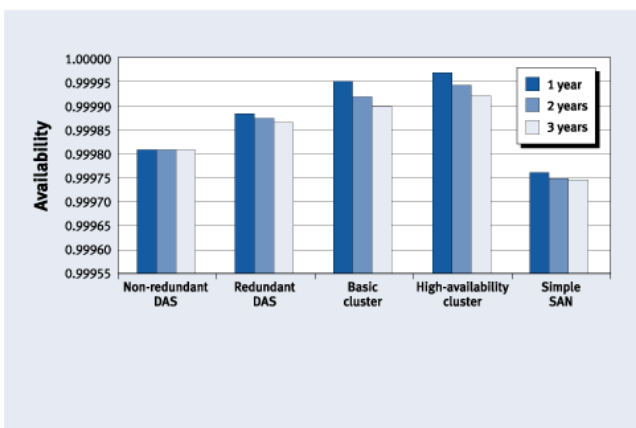


Figura 24 Disponibilità dei differenti sistemi.

Figure 24 Systems Availability.

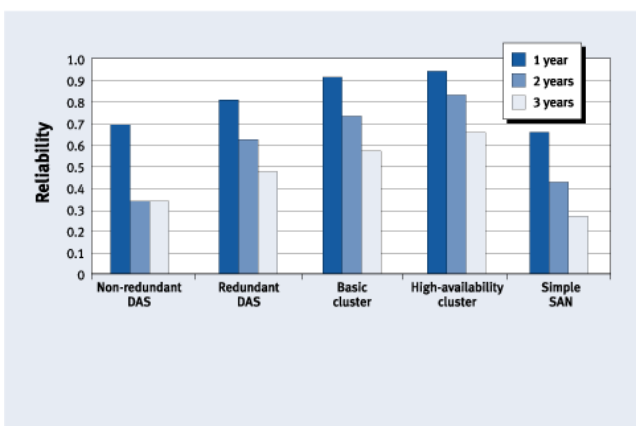


Figura 25 Affidabilità dei diversi sistemi.

Figure 25 Systems Reliability.

4. La soluzione ideale

Alla luce di tutte le considerazioni fatte in precedenza, esiste una configurazione ideale?

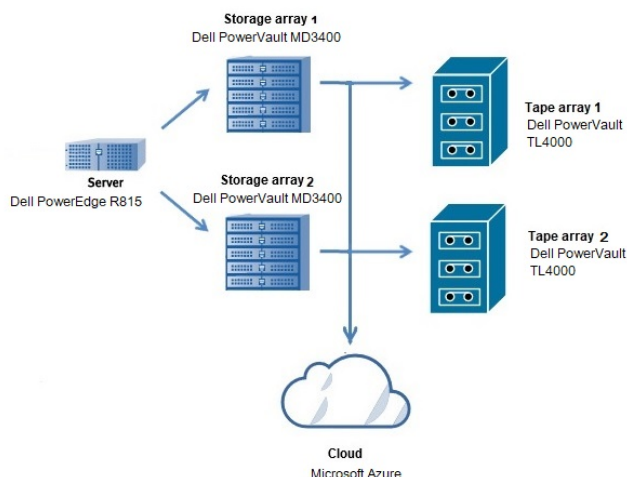
In realtà una soluzione che ben si adatti ad ogni contesto non esiste.

Avendo ben presente ciò e considerando che non è facile trovare una soluzione che sia un buon compromesso tra affidabilità, costi di acquisto, gestione e semplicità d'uso, possiamo disegnare una configurazione che potrebbe essere adottata nel nostro caso (figura 26).

Un server Dell Poweredge R815 controlla il processo di salvataggio dei dati su due array di dischi Powervault MD3400, i quali copiano i dati parallelamente su due Tape library Powervault TL4000 e sul cloud, rappresentato da Microsoft Azure, più economico rispetto a Google Cloud. Il costo annuo della soluzione è rappresentato in tabella 14. I dati sono ricavati dal raffronto della tabella 8 con la tabella 9. Sono stati specificati marca e modello delle macchine al solo scopo di avere una stima dei costi sufficientemente realistica.

Figura 26 Una configurazione ideale.

Figure 26 An ideal configuration.



| | |
|--|------------------|
| Costo Totale annuo acquisto e gestione (€) | 12130,16 |
| Costo Totale annuo Microsoft Azure (€) | 8396,87 |
| TOTALE (€) | 20.527,03 |

Tabella 14 Costo totale annuo della configurazione "ideale".

Table 14 Ideal configuration total annual cost.

5. Upgrading del sistema

Nel prossimo anno è previsto l'aumento della capacità di stoccaggio del sistema attraverso l'acquisto di nuovi sistemi di array di dischi. Una terza libreria a nastri affiancherà le due già implementate, aumentando così la dimensione complessiva dei nastri "on-line". Ciò permetterà di aumentare la frequenza di backup, garantendo al contempo una maggiore velocità di esecuzione dello stesso.

Una parte dei dati, inoltre, sarà salvata anche sul cloud. È attualmente in fase di sperimentazione il backup su storage S3 di Amazon.

6. Conclusioni

Nella stesura di una strategia di backup bisogna tenere conto di diversi aspetti e raggiungere un compromesso accettabile tra costi, sicurezza dei dati e gestibilità dell'infrastruttura al fine di rendere efficace e sicuro il processo di salvataggio dei dati.

I passi da seguire per implementare un corretto piano di backup implicano diversi step, i principali sono:

- Mappare la rete: Quali sono le macchine, di che tipo sono, quale sistema operativo implementano.
- Tracciare i flussi di dati, la loro consistenza ed il livello di criticità.
- Progettare il sistema hardware e software necessario.
- Scelta dei supporti di storage. Bisogna valutare l'affidabilità, i costi di acquisto e gestione. Inoltre sarebbe meglio valutare l'adozione di politiche di archiviazione su cloud, per delocalizzare almeno una parte dei dati.

È importante sapere che qualunque soluzione si adotti, ci sono vantaggi e svantaggi di cui tenere conto nella stesura del piano di backup.

Volendo riassumere possiamo dire che il cloud rappresenta una soluzione estremamente flessibile e facilmente scalabile, su cui l'utente finale esercita un controllo minimo, di contro la soluzione On-Premises ha poca flessibilità e scalabilità, ma un grande controllo sull'infrastruttura hardware e software.

Adottare una buona politica di backup significa proteggere i dati e poterne nuovamente disporre in tempi ragionevolmente brevi, in caso di problemi.

Sitografia

Frost & Sullivan. <https://ww2.frost.com/>

Sito Dell. <http://www.dell.com>

Server rack R720 PowerEdge. <https://www.dell.com/it-it/work/shop/povw/poweredge-r720>

Dell PowerVault MD3 1GB iSCSI SAN Storage Array Series.

<https://www.dell.com/it-it/work/shop/productdetailstxn/powervault-md3200i>

PowerVault MD1200 Direct Attached Storage.

<https://www.dell.com/en-us/work/shop/povw/powervault-md1200>

Libreria a nastro PowerVault TL4000.

<https://www.dell.com/it-it/work/shop/cty/pdp/spd/powervault-tl4000?~ck=anav>

Bacula manuals. <http://blog.bacula.org/documentation/documentation>

Sito Bacula. <http://blog.bacula.org/>

Sito Webmin. <http://www.webmin.com>

Sito Postgres. www.postgresql.org

Sito Mysql. <https://www.mysql.com/it/>

Kern Sibbald, "Bacula Main Reference". <http://www.bacula.org/9.2.x-manuals/en/main/main.pdf>

Dell EMC Enterprise Infrastructure Planning Tool. <http://dell-ui-eipt.azurewebsites.net/#>

Enel E-light Impresa. <https://www.enel.it/it/luce-e-gas/luce/offerte/elight-luce-impresa>

Environmental and shipping specifications for LTO tape cartridges.

https://www.ibm.com/support/knowledgecenter/sv/STCMML8/com.ibm.storage.ts3500.doc/ipg_3584_meesi.html

Guida all'acquisto DCIG 2016-2017 per gli array di storage unificati di fascia media.

<https://www.hpe.com/it/it/pdfViewer.html?resource=/content/hpe/country/it/it/resources/storage/analyst-report/4AA6-8201ENW>

HPE StoreEver MSL2024 and MSL4048 Tape Libraries.

<https://www.hpe.com/h20195/v2/getpdf.aspx/c04154359.pdf?ver=54>

Santosh Shetty (August 2002). Determining the Availability and Reliability of Storage Configurations. http://www.dell.com/content/topics/global.aspx/power/en/ps3q02_shetty?c=us

Introduction au concept FMD: Fiabilité - Maintenabilité - Disponibilité.

<http://jackadit.com/index.php?page=indus3>

Reliability Prediction Basics. Copyright 2007, ITEM Software, Inc..

<http://www.reliabilityeducation.com/ReliabilityPredictionBasics.pdf>

Alberto Bettanti, *Gestione della Qualità. Corso di Laurea in Ingegneria Elettrica, Informatica e Biomedica.*

I principi della teoria dell'affidabilità.

http://www-3.unipv.it/gdq/Approfondimenti/affidabilita_GdQ.pdf

Creare un cluster di failover. <https://docs.microsoft.com/it-it/windows-server/failover-clustering/create-failover-cluster>

Clustering di failover in Windows Server. <https://docs.microsoft.com/it-it/windows-server/failover-clustering>

clustering/failover-clustering-overview

Susan Stanley. Copyright © 2011 IMC Networks. MTBF, MTTR, MTTF & FIT Explanation of Terms <http://www.bb-elec.com/Learning-Center/All-White-Papers/Fiber/MTBF,-MTTR,-MTTF,-FIT-Explanation-of-Terms/MTBF-MTTR-MTTF-FIT-10262012-pdf.pdf>

Sito Azure. <https://azure.microsoft.com/it>

Replica di Archiviazione di Azure.

<https://docs.microsoft.com/it-it/azure/storage/common/storage-redundancy>

Data center on-premise e cloud computing a confronto. <https://www.hpe.com/it/it/what-is/on-premises-vs-cloud.html>

Pier Luca Maffettone. *Metodi di analisi dei dati sperimentali 2009-2010.*

http://wpage.unina.it/p.maffettone/Didattica/Mads/Lezione_2.pdf

Augusto Gandini. *Lezioni di Ingegneria del Nocciolo. Elementi di analisi probabilistica Università di Roma "La Sapienza". Corso di Laurea in Ingegneria Nucleare.*

Hilt, Attila & Jaro, Gabor & Bakos, Istvan. (2016). *Availability Prediction of Telecommunication Application Servers Deployed on Cloud. Periodica Polytechnica, Electrical Engineering.* 60. 72-81. 10.3311/PPee.9051.

Military Handbook. (1991) *Reliability Prediction of Electronic Equipment.* MIL-HDBK-217F.

Jeremy S. (2017). *How Long Do Hard Drives Last? Lifespan and Signs of Failure.*

<https://www.prosofteng.com/blog/how-long-do-hard-drives-last/>

Lynda Stadtmueller. (2018). *Protecting Your Company's Greatest Assets with Intelligent Data Protection.* <https://h20195.www2.hp.com/v2/getdocument.aspx?docname=a00056041enw>

Andy Klein. (2019). *Backblaze Hard Drive Stats Q2 2019.* <https://www.backblaze.com/blog/hard-drive-stats-q2-2019/>

HP LTO Ultrium Media.

https://www8.hp.com/uk/en/pdf/hp-ltoultrium-datasheet_tcm_183_1268652.pdf

Y. Izrailevsky and C. Bell. *Cloud Reliability* in *IEEE Cloud Computing*, vol. 5, no. 03, pp. 39-44, 2018. doi:10.1109/MCC.2018.032591615.

<https://doi.ieeecomputersociety.org/10.1109/MCC.2018.032591615>

QUADERNI di GEOFISICA

ISSN 1590-2595

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/quaderni-di-geofisica.html/>

I QUADERNI DI GEOFISICA (QUAD. GEOFIS.) accolgono lavori, sia in italiano che in inglese, che diano particolare risalto alla pubblicazione di dati, misure, osservazioni e loro elaborazioni anche preliminari che necessitano di rapida diffusione nella comunità scientifica nazionale ed internazionale. Per questo scopo la pubblicazione on-line è particolarmente utile e fornisce accesso immediato a tutti i possibili utenti. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi. I QUADERNI DI GEOFISICA sono presenti in "Emerging Sources Citation Index" di Clarivate Analytics, e in "Open Access Journals" di Scopus.

QUADERNI DI GEOFISICA (QUAD. GEOFIS.) welcome contributions, in Italian and/or in English, with special emphasis on preliminary elaborations of data, measures, and observations that need rapid and widespread diffusion in the scientific community. The on-line publication is particularly useful for this purpose, and a multidisciplinary Editorial Board with an accurate peer-review process provides the quality standard for the publication of the manuscripts. QUADERNI DI GEOFISICA are present in "Emerging Sources Citation Index" of Clarivate Analytics, and in "Open Access Journals" of Scopus.

RAPPORTI TECNICI INGV

ISSN 2039-7941

<http://istituto.ingv.it/it/le-collane-editoriali-ingv/rapporti-tecnici-ingv.html/>

I RAPPORTI TECNICI INGV (RAPP. TEC. INGV) pubblicano contributi, sia in italiano che in inglese, di tipo tecnologico come manuali, software, applicazioni ed innovazioni di strumentazioni, tecniche di raccolta dati di rilevante interesse tecnico-scientifico. I RAPPORTI TECNICI INGV sono pubblicati esclusivamente on-line per garantire agli autori rapidità di diffusione e agli utenti accesso immediato ai dati pubblicati. Un Editorial Board multidisciplinare ed un accurato processo di peer-review garantiscono i requisiti di qualità per la pubblicazione dei contributi.

RAPPORTI TECNICI INGV (RAPP. TEC. INGV) publish technological contributions (in Italian and/or in English) such as manuals, software, applications and implementations of instruments, and techniques of data collection. RAPPORTI TECNICI INGV are published online to guarantee celerity of diffusion and a prompt access to published data. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

MISCELLANEA INGV

ISSN 2039-6651

http://istituto.ingv.it/it/le-collane-editoriali-ingv/miscellanea-ingv.html

MISCELLANEA INGV (MISC. INGV) favorisce la pubblicazione di contributi scientifici riguardanti le attività svolte dall'INGV. In particolare, MISCELLANEA INGV raccoglie reports di progetti scientifici, proceedings di convegni, manuali, monografie di rilevante interesse, raccolte di articoli, ecc. La pubblicazione è esclusivamente on-line, completamente gratuita e garantisce tempi rapidi e grande diffusione sul web. L'Editorial Board INGV, grazie al suo carattere multidisciplinare, assicura i requisiti di qualità per la pubblicazione dei contributi sottomessi.

MISCELLANEA INGV (MISC. INGV) favours the publication of scientific contributions regarding the main activities carried out at INGV. In particular, MISCELLANEA INGV gathers reports of scientific projects, proceedings of meetings, manuals, relevant monographs, collections of articles etc. The journal is published online to guarantee celerity of diffusion on the internet. A multidisciplinary Editorial Board and an accurate peer-review process provide the quality standard for the publication of the contributions.

Coordinamento editoriale e impaginazione

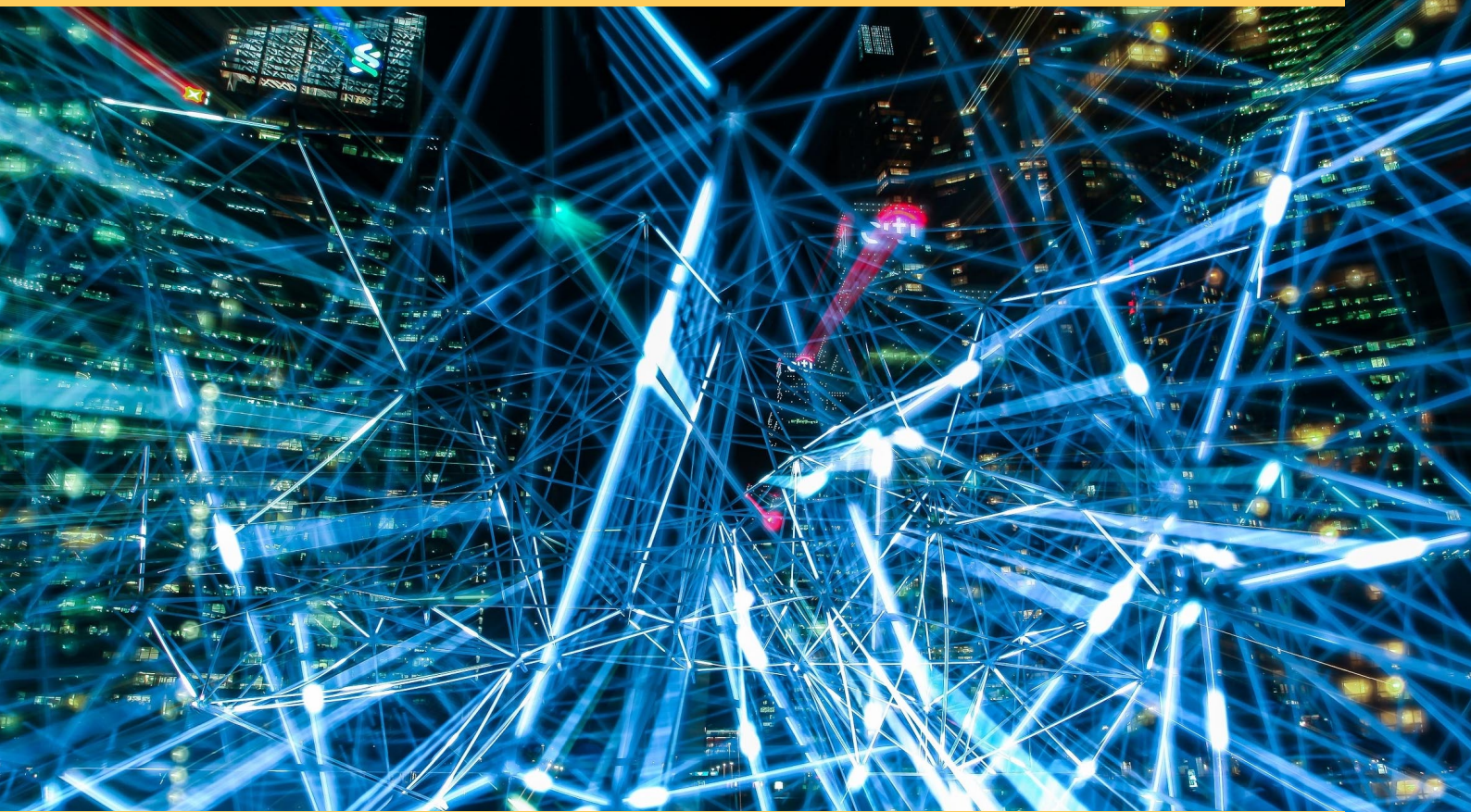
Francesca DI STEFANO, Rossella CELI
Istituto Nazionale di Geofisica e Vulcanologia

Progetto grafico e impaginazione

Barbara ANGIONI
Istituto Nazionale di Geofisica e Vulcanologia

©2020
Istituto Nazionale di Geofisica e Vulcanologia
Via di Vigna Murata, 605
00143 Roma
tel. +39 06518601

www.ingv.it



ISTITUTO NAZIONALE DI GEOFISICA E VULCANOLOGIA