Disaster Recovery: l'importanza di una corretta implementazione di un Disaster Recovery Plan

Di Capua Marco

Abstract – *Questo elaborato pone come obiettivo principale l'individuare le procedure corrette per implementare un Disaster Recovery Plan in grado di ristabilire i servizi in tempi adeguati e di evitare la perdita di informazione. Saranno presentati brevemente i requisiti e le metriche utilizzate per l'implementazione di tale piano, le tipologie di replica delle informazioni, l'applicazione di una procedura di DRP con allineamento sincrono ed infine un breve esempio per sottolineare l'importanza di una preparazione adeguata al fine di affrontare un possibile disastro.*

I. Introduzione

Le grandi aziende e compagnie che al giorno d'oggi vanno a ricoprire una vasta parte dell'economia mondiale, devono porre sempre più attenzione nel fornire un servizio continuo e costante ai propri clienti in grado di gestire grandi quantità di dati e di evitare perdite nel caso di incidenti non pianificati.

I tipi di incidenti che possono portare a questi tipi di problemi possono essere catastrofi naturali come ad esempio i terremoti, interruzioni di corrente oppure i sempre più frequenti attacchi informatici. Le aziende dunque devono adottare delle strategie adeguate a ridurre al minimo le perdite causate dagli incidenti e riprendere rapidamente l'erogazione dei servizi principali, in quanto una scarsa preparazione a questi tipi di eventi può portare ad ingenti danni economici oltre che alla perdita irreversibile di dati importanti dei propri clienti.

Tutti questi aspetti hanno portato ad accrescere l'importanza dei concetti di Business Continuity e Disaster Recovery tanto da renderli essenziali per una qualsiasi azienda che fornisce servizi in rete.

II. Business Continuity e Disaster Recovery

La Business Continuity è un processo strategico atto a gestire eventi di qualsiasi scala in grado di interrompere interi processi aziendali, rendendo possibile un ripristino delle funzionalità interrotte ad un livello di servizio minimo predefinito. Gli eventi in grado di influenzare negativamente la normale continuità aziendale possono variare dalla sostituzione di una figura dirigenziale all'interno della azienda, fino alla interruzione di servizi business fondamentali a causa di disastri di tipo naturale.

La Disaster Recovery è un insieme di tecnologie e procedure che determinano la capacità di recuperare e ripristinare i dati e i servizi applicativi forniti da una infrastruttura IT di una organizzazione, riducendo o eliminando il lasso di tempo che intercorre tra un evento catastrofico che determina una interruzione dei servizi e il ripristino degli stessi.

La Disaster Recovery dunque è un sottoinsieme specifico della Business Continuity e tutte le procedure atte al recupero dei dati e al ripristino dei servizi vengono generalmente definite all'interno di un Disaster Recovery Plan.

III. Metriche del Disaster Recovery

Nella stesura del Disaster Recovery Plan le aziende pongono particolare attenzione sia sulla analisi dei possibili eventi scatenanti uno scenario di disastro e delle relative potenziali perdite economiche causate da una interruzione del servizio, sia le vulnerabilità esistenti nel sistema e nello stabilire le priorità dei servizi da tutelare per mantenere la continuità aziendale. Le tecniche utilizzate per effettuare questo tipo di analisi sono rispettivamente il Risk Assesment e la Business Impact Analysis.

Gli studi appena menzionati portano alla definizione di indici che costituiscono le metriche del Disaster Recovery; il Recovery Time Objective (RTO) e il Recovery Point Objective (RPO).

L'RTO esprime in unità di tempo quanto un sistema può restare inattivo senza provocare danni significativi a un'azienda, incluso il tempo necessario al sistema per recuperare il tempo perduto. Questo processo di recupero comprende i passi che l'IT deve intraprendere per riportare le applicazioni e i loro dati nello stato in cui si trovavano prima che si verificasse l'incidente.

L'RPO identifica letteralmente un "*punto di ripristino*" della situazione aziendale, ovvero l'istante in cui, dopo l'essersi verificato il disastro, l'azienda può ripristinare la sua operatività. Di conseguenza il lasso di tempo identificato compreso tra il disastro e il punto di ripristino comporta una possibile perdita di dati; tanto più e lungo questo periodo quanti più dati saranno persi.

Questi due parametri sono di fondamentale importanza per stabilire una corretta procedura di Disaster Recovery valutando i costi dell'indisponibilità dei servizi e rapportandoli con i costi delle soluzioni di recovery. In base al tipo di applicazione in analisi i costi di indisponibilità possono essere più o meno alti; ad esempio una applicazione real-time resa non disponibile da un disastro causa enormi perdite economiche dopo pochi minuti, mentre per le applicazioni di back-office il costo aumenta dopo diverse ore/giorni. Il grafico nella *Figura 1* evidenza il rapporto che intercorre tra costo di soluzione di recovery e tempo di indisponibilità per le applicazioni real-time e di back-office. Lo stesso grafico può essere utilizzato anche per valutare i costi dell'RPO, tuttavia in questo caso non si tratta di più di un rapporto costo-tempo ma di un rapporto costo-dati persi.

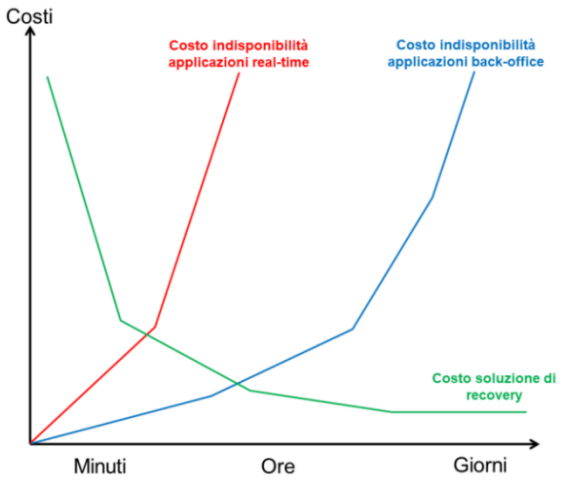


Figura 1 RTO e costi

Esistono diverse tecnologie utilizzate per la soluzione dei problemi legati a RTO e RPO. In base alla quantità di tempo di inattività che si è disposti a perdere, le diverse tecnologie per risolvere i problemi legati all'RTO sono:

- Back-up su Tape off-site (RTO di giorni)
- Electronic Tape Vaulting (RTO di parecchie ore)
- Back-up Disk to Disk (RTO di alcune ore)
- Remote DB Logging (RTO di alcune ore)
- Remote Disk Copy Asincrono (RTO di poche ore)
- Remote Disk Copy Sincrono (RTO di minuti)

L'utilizzo di configurazioni di cluster geografico o metropolitano può influire notevolmente sul valore di RTO, in quanto

velocizzano notevolmente il ripristino dei servizi dell'azienda.

Le tecnologie utilizzate per risolvere i problemi legati all'RPO sono:

- Back-up (RPO di circa 24 ore)
- Remote DB Logging (RPO di ore/minuti)
- Remote Disk Copy Asincrono (RPO di pochi minuti/secondi)
- Remote Disk Copy Sincrono (RPO tendente a zero)

In questo caso la distanza geografica che intercorre tra i vari data center influisce notevolmente sul valore di RPO.

IV. Disaster Recovery Site

Al fine di applicare la procedura descritta in un Disaster Recovery Plan è necessario predisporre almeno un data center adibito alla procedura di Disaster Recovery e identificare una posizione geografica adatta dello stesso.

In base alle esigenze e alle dovute valutazioni del rapporto costi-benefici è possibile identificare quattro tipologie di Disaster Recovery Site:

1- Hot Site: questo tipo di sito dispone di tutti i sistemi hardware/software, infrastrutture di supporto, personale di supporto e sono operativi 24h al giorno 7 giorni su 7. Evidentemente questo tipo di soluzione è il più completo ma anche il più dispendioso a livello economico, tuttavia le attrezzature presenti in questo sito consentono di poterlo utilizzare non solo come un Disaster Recovery Site, ma anche come un sito per il Load Balance.
2- Mirrored Site: sito che identifica una esatta copia del sito primario in tutti gli aspetti tecnici, operativi e applicativi. Questo tipo di soluzione possiede il più alto livello di availability ottenibile rispetto alle altre soluzioni in questo elenco, in quanto i dati sono processati e memorizzati simultaneamente sia nel sito primario che nel Mirrored Site.
3- Warm Site: il sito è parzialmente attrezzato e dispone di alcune o tutte le infrastrutture IT necessarie per ripristinare i servizi. Le componenti tipicamente presenti all'interno di un warm site sono i server, alcuni clients, gli applicativi e i collegamenti alla rete.
4- Cold Site: questa soluzione è la più economica e meno efficiente dell'elenco. Identifica un sito con le infrastrutture di base ed un numero di servizi ridotti. Tipicamente questo tipo di siti sono in grado di supportare per un breve lasso di tempo alcune delle operazioni IT.

V. Repliche e mirroring

La protezione dei dati è un elemento chiave nella procedura di Disaster Recovery, in quanto l'informazione contenuta nei dati, se persa, è irrecuperabile e nega l'erogazione del servizio ad essa associato. Il processo adibito alla protezione dei dati è la replica dei dati. La replica è un processo di creazione di più istanze dello stesso database allineate fra di loro.

La replica può essere utilizzata sia in maniera sincrona che asincrona. Nel primo caso si cerca di aggiornare tutte le repliche in contemporanea tramite diversi metodi (es. protocollo Read Once Write All). Nel secondo caso il meccanismo è diverso, infatti prima si completa la transazione e si aggiorna il database primario e in un secondo momento si effettua l'aggiornamento delle repliche.

La replica sincrona obbliga ad aggiornare due o più storage contemporaneamente e nel caso in cui non fosse possibile si effettua una operazione di rollback della transazione. Come mostrato in *Figura* 2, l'operazione di replica sincrona tra due basi di dati prevede l'invio periodico di un file di log

transazionale al DB Server di backup, il quale, dopo aver aggiornato le transazioni all'interno del suo storage, invia un ACK di avvenuta scrittura al DB Server primario che completa la transazione con successo.

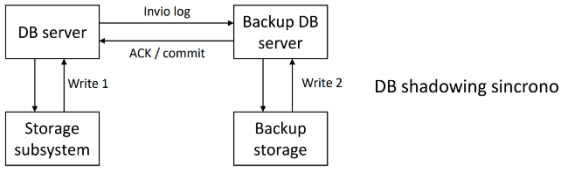


Figura 2 replica sincrona

Al fine di migliorare le prestazioni e di ottenere un valore di RPO tendete a zero, si rivelano particolarmente utili le soluzioni di mirroring. Le strategie di mirroring adoperate nelle operazioni di Disaster Recovery sono:

- Logical volume mirroring: soluzione che si basa sul software del sistema operativo del DB server principale. Implementa delle copie di transazioni sincrone che prevedono la scrittura contemporanea direttamente sui due sistemi di storage replicati. Questo tipo di soluzione è completamente sincrona ma può causare dei problemi di ritardo.
- Physical volume mirroring: lo storage principale si occupa sia della scrittura dei dati nella sua memoria, sia dell'eseguire una copia della transazione sullo storage secondario.

La differenza sostanziale tra logical volume mirroring e replica sincrona menzionata precedentemente, può essere facilmente denotata nella *Figura 3* dove si può notare che il Backup DB server non svolge alcuna operazione nel processo di replica dei dati.

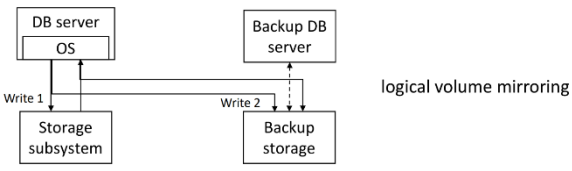
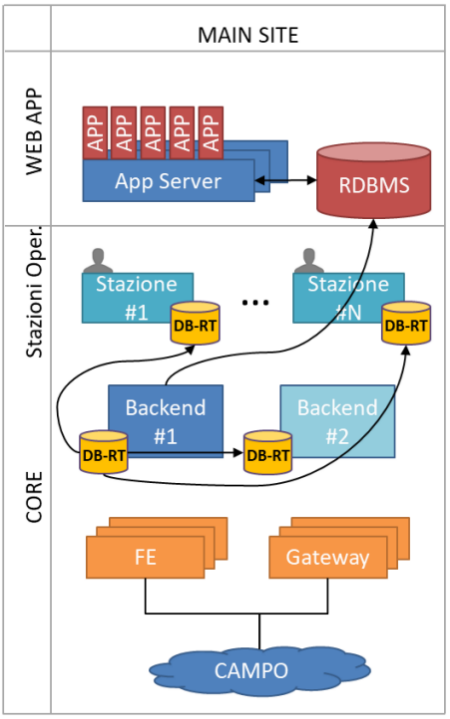


Figura 3 logical volume mirroring

VI. Applicazione del DRP con allineamento sincrono

Verrà ora presentato uno scenario di applicazione del DRP che utilizza una tecnica di mirroring logico per ripristinare le funzioni di un data center colpito da un disastro. L'architettura del sistema in questione è quella descritta in *Figura 4*.



Questa architettura è basata su tre livelli funzionali: acquisizione da campo, elaborazione centrale e applicazioni web di supporto ai processi aziendali. Il data center adibito alla funzione di Disaster Recovery è una esatta copia del data center primario appena descritto, definendo così un Mirrored Site collegato tramite un allineamento sincrono di logical volume mirroring.

I due sistemi di storage, localizzati in aree geografiche differenti, vengono combinati all'interno di un Volume Group gestito da un Geographic Logical Volume Manager. Le informazioni sono dunque trasmesse tramite protocollo TCP/IP e quindi, in caso di disastro, è garantita una corretta procedura di Disaster Recovery che consente al Disaster Recovery Site di subentrare in azione in tempi brevi,

evitando grazie al GLVM una possibile perdita di dati.

VII. Caso GitLab

Si presenta ora un esempio di operazione di Disaster Recovery non andato a buon fine, con lo scopo di sottolineare l'importanza di tale procedura e le possibili conseguenze causate da una cattiva gestione dell'emergenza. Nel 2017 a causa di un incidente e di diversi problemi di troubleshooting i dati di diverse ore presenti all'interno dei sistemi di storage della compagnia GitLab sono andati perduti. GitLab è un sito web che gestisce in maniera pubblica o privata un intero ciclo DevOps di progetti, fornendo ai suoi utenti la possibilità di utilizzare funzioni di wiki, issue-tracking, continuos integration e molte altre. Il 31/01/2017 i tecnici GitLab notano che alcuni malintenzionati, dopo aver ottenuto l'accesso al database primario, stavano effettuando un attacco utilizzando degli snippets, rendendo così il sistema instabile. L'attacco ha portato il sistema non essere più disponibile dopo solo tre ore. I tecnici per contrastare l'attacco bloccano il sistema e cancellano tutte le richieste di scrittura massiva eseguite dagli spammer. Il passo successivo dei tecnici consisteva nel ripristinare il sistema al momento precedente all'attacco utilizzando il database secondario. Durante questa fase i tecnici si accorgono che, a causa di un invio massivo di operazioni di scrittura da parte del DB primario verso il DB secondario per effettuare la replica dei dati, quest'ultimo risultava non rispondere alle interazioni. Per far fronte a questo problema si decide di cancellare le ultime ore di replica all'interno del DB secondario al fine di ripristinarne la funzionalità. A causa di un errore da parte del tecnico la base di dati ad essere cancellata è stata la base di dati primaria e non la secondaria. A questo punto gli operatori decidono di utilizzare una delle copie di backup per ripristinare i dati persi ma, a causa di una procedura di replica fragile e tassellata di errori, non è stato possibile eseguire tale procedura. In conclusione gli operatori sono riusciti ad ottenere una copia di backup risalente a 6 ore antecedenti l'attacco e, grazie ad essa, sono riusciti a recuperare gran parte dei dati persi. Il danno procurato da questo disastro ammonta a più di 5000 progetti, 5000 commenti e 700 utenti persi.

VIII. Conclusioni

L'esempio appena mostrato sottolinea l'importanza della preparazione di un Disaster Recovery Plan all'interno di una azienda che fornisce dei servizi in rete. Le fasi di preparazione, di individuazione delle criticità, di valutazione dei costi e rischi e di stesura di un DRP devono essere eseguite tutte con la massima attenzione, senza tralasciare alcun minimo dettaglio. Le conseguenze catastrofiche causate da una scarsa attenzione rivolta a questo tipo di procedure possono comportare enormi perdite economiche tali da minare la stabilità della azienda stessa.

IX. Bibliografia e sitografia

Eric Conrad,

Editor(s): Eric Conrad,

Eleventh Hour CISSP (Second Edition), Syngress, 2014

Susan Snedaker, Chris Rima,

Editor(s): Susan Snedaker, Chris Rima,

Business Continuity and Disaster Recovery Planning for IT Professionals (Second Edition), Syngress, 2014

Calogero Gandolfo,

Seminario generale sulle architetture e sulle

tecnologie adottate nelle soluzioni di

Disaster Recovery e Business Continuity, 2013

URL:
http://www.diag.uniroma1.it/~ciciani/files/Disaster%20Recovery%20e%20Business%20Continuity%20Maggio2013.pdf

Lorenzo Camerini, Paolo Manià, Alberto Zironi,

Disaster Recovery per infrastrutture critiche:

mirroring tra due DSO, 2017

URL:
http://www.forumtelecontrollo.it/allegati/12.30_siemens_inrete_acegas_camerini_merged.pdf

GitLab,

GitLab.com database incident, 2017

URL:
https://about.gitlab.com/blog/2017/02/01/gitlab-dot-com-database-incident/

IBM,

Geographic Logical Volume Manager

IBM PowerHA SystemMirror for AIX, Enterprise Edition, Version 7.2, 2017

URL:

https://www.ibm.com/docs/en/SSPHQG_7.2/glvm/hacmpgeolvm_pdf.pdf