Activitat AS 04

Data protection (classes 1 i 3 de març)

DATA LÍMIT DE LLIURAMENT: Dilluns 15 de març, a mitjanit

DANIEL DONATE DURÁN:

PREGUNTES:

1) Descriu amb les teves paraules els següents conceptes:
    a. Definició RPO i RTO, diferències amb RA i BIA
    b. Defineix *hot spare disk* i la seva utilitat
    c. Problemes del backup: *frozen data*, temps de recuperació, perquè es fa en cintes majoritàriament?
    d. Definició *full backup* i *synthetic backup*
    e. Descriu la idea bàsica *de Shadow copy, snapshots i continuous data protection* (bàsica, un parell o tres de línies)

RESPOSTES:

1-a) El RPO (Recovery Point Objective) és una mesura del volum de dades que poden perdre's i que una empresa/organització considera tolerable. En altres paraules, el RPO es correspon a les transaccions que estem disposats a perdre (o a haver de re-introduir al sistema) en un temps determinat, sense que això sigui crític.

D'altra banda, el RTO (Recovery Time Objective) és el temps el qual l'organització considera tolerable la falta de funcionament dels seus serveis, sense que aquest temps posi en risc la continuïtat del seu negoci.

Aquests dos conceptes (RPO i RTO) són conceptes tècnics. Amb una perspectiva més de negoci tenim altres dues mètriques: RA i BIA.

El RA (Risk Analysis) és un anàlisi de les coses que podrien anar malament en el nostre sistema (fallada d'un cert nombre de discs, caigudes de xarxa, atacs terroristes, etc).

El BIA (Business Impact Analysis) és un anàlisi per intentar establir l'impacte econòmic associat a una interrupció d'alguna de les parts del nostre sistema.

1-b) Els *hot spare disks* són discs que s'introdueixen en una cabina de discs, però que inicialment no estan *actius*: no contenen dades i no realitzen cap operació. El seu propòsit és *substituir* un disc que està a punt de fallar (tecnologia SMART), traslladant les dades del disc al *hot spare disc*, o començar a reconstruir automàticament un disc que ha fallat sobtadament (en un sistema RAID5, per exemple). Això ens permet actuar de forma molt ràpida, evitant haver d'interrompre el sistema.

1-c) Quan estem fent un *backup* d'una base de dades (la idea es pot "generalitzar" a *backup* d'un disc amb altres propòsits, però el de les BBDD em sembla l'escenari més natural on explicar-ho) sorgeix un problema que s'anomena de quiescència. És possible que en un moment donat es produeixi una operació que afecti dos *records* de la base de

dades, un dels quals ja s'hagi fet còpia, i l'altre no, de manera que estem produint una còpia inconsistent de les dades. És per això que es parla de la *necessitat de congelar* les dades. Hi ha diferents tipus de solucions per atacar aquest problema.

El temps de recuperació és el temps que passa entre que nosaltres, com a empresa, sol·licitem/comencem una restauració de dades després d'haver tingut algun tipus de caiguda/pèrdua dels nostres discs fins al moment en el que les dades s'han restablert de forma completa i satisfactòria i tornem a tenir els serveis en funcionament. En conseqüència amb el que hem dit a la pregunta 1-a, aquest temps de recuperació hauria de ser inferior al RTO per anar bé...

Les cintes magnètiques es fan servir en els *backups* dels CPDs (on, lògicament, es treballa amb quantitats elevadíssimes de dades, que a més s'han de copiar amb una alta freqüència) degut al seu preu econòmic.

1-d) El *full backup*, com el seu nom suggereix, és el nom que es dona a una còpia total de les dades dels discos d'un sistema: tot allò que s'ha de copiar es copia d'una sola vegada, des dels discos. És força evident que fer un *full backup* deu ser un procés considerablement lent. Per això existeix una alternativa al *full backup*: el *synthetic backup*, que consisteix en una màquina dedicada i fora del nostre sistema 'principal' que agafa, d'una banda, l'últim *full backup* que tenim disponible, i d'altra banda, tots els canvis que s'han anat produint i tenim guardats (*backup* incremental), per crear, d'aquesta manera, un *backup* sintètic, en el sentit que no es fa llegint directament les dades dels discs.
1-e) La idea de *shadow copy* (o *Business Continuity Volume*) és tenir una còpia dels nostres discs originals mitjançant *hardware* addicional que només guarda dades, no té cap tipus de redundància. És a dir, en lloc de duplicar els discs, tenim només aquells discs que necessitem per copiar les dades originals (podria ser un sistema molt simple, com un JBOD). Cada vegada que fem una escriptura en un disc original, hem de fer també l'escriptura en la *shadow* copy, per mantenir la coherència, fins al moment en que els discs estan sincronitzats, que es quan *congelem* la còpia (i fem un *backup* en cinta).

Els snapshots són una espèicie d'instantània del nostre sistema de dades que es basa en CoW (*Copy on Write*). La idea és guardar tots els enllaços/punters a discs del nostre sistema de fitxers per a poder, en cas de voler modificar/sobre-escriure un bloc en algun moment posterior, poder tornar llegir les dades que hi havia abans, gràcies a CoW i a saber a quina regió del disc s'apuntava abans d'actualitzar el bloc. Notem que la durada dels *snapshots* és, generalment, bastant limitada.

El *continuous data protection* és un *backup* 'extern' a l'empresa. El que es fa és contractar a una empresa externa que manté una espècie de *mirror* o, més habitualment, un *journal* amb les diferents operacions que hem compartit amb l'empresa. D'aquesta manera podem, en un moment donat, realitzar un *rollback* i tornar a un estat anterior a l'actual, si ens interessa. En principi, no tenim per què passar a l'empresa externa tota la informació dels discs, només allò que de veritat volem guardar (p. ex. els fitxers temporals ens els podem "estalviar").