

Activitat EBH

Emmagatzematge, *backup* i *housing*

Estudiant 1 Peralta Farré, Guillem
Estudiant 2 Oitavén de Sousa, Pedro Afonso

Escenari 16

Data: 12/10/2023

1.-Descripció bàsica

TAULA 1: ESCENARI ORIGINAL: EXTRET DE L'ENUNCIAT. OMPLIU EL QUE HI HA EN GRIS.	
Nombre de Us	112U
Alçada Rack (en Us)	42U
Consum	360,kW
Sobreprovisionament d'electricitat	7%
Nombre de servidors	56
Diners Totals	€10.000.000,00
Diners gastats	€7.500.000,00
taula 2: Elements que escolliu vosaltres	
Elements de mirror i backup	
GB a emmagatzemar al backup	3539,52
Dies entre 2 backups	1
Còpies senceres a mantenir	14
Opció Backup (1=M-A; 2=MS3; 3=Cintes)	2
Opció Mirror (0=NO; 1=SI)	1
Sistema de backup on-site? (0=N=; 1=SI)	0
Elements de housing	
Opció escollida (1:MOCOSA, 2: CPDs Céspedes, 3: Mordor)	3
Gestió local de <i>backup</i> ? (0=No, 1=Si)	0
Monitorització? (0=NO; 1=SI)	1
Bandwidth provider	

Tipus de línia (1:10Mbps; 2:100Mbps; 3:1Gbps; 4:10Gbps; 5:100Gbps)	4
Número de línies agregades	2
Segon proveïdor? (0=NO, 1=SI)	1
SAN? (0=no, 1=8Gbps, 2=16Gbps, 3=32Gbps, 4=64Gbps, 5=128Gbps)	2
Cabina de discos	
Opció Disc principal (Entre 1 i 10)	10
Nombre de discos a comprar	10
Opció cabina de discos (Entre 1 i 6)	2
Nombre de Cabines	1
Cabina de discos 2 (cas de fer servir dos tipus)	
Opció Disc (Entre 1 i 10)	0
Nombre de discos a comprar	0
Opció cabina de discos (Entre 1 i 6)	0
Nombre de Cabines	0
Cabina de discos 3 (cas de fer servir tres tipus)	
Opció Disc (Entre 1 i 10)	0
Nombre de discos a comprar	0
Opció cabina de discos (Entre 1 i 6)	0
Nombre de Cabines	0

2.-Anàlisi de necessitats

2.1- Número de GB a emmagatzemar (en cru).

Hem de tenir en compte els GB que ens ocupen tant els usuaris com els productes.

Usuaris (GB): $256\text{KB} \times 20.000 \text{ usuaris} = 5120000 \text{ KB} = 5,120 \text{ GB}$

Productes (GB): $(64\text{KB} + 8\text{KB}) \times 5.000.000 \text{ productes} = 360.000.000 \text{ KB} = 360 \text{ GB}$

Total (GB): $4,8828\text{GB} + 343,3228\text{GB} = \mathbf{365,120 \text{ GB}}$

2.2- Velocitat requerida del sistema de disc (IOPS).

Tràfic total(KBps) = $6338370,4 \text{ Kbps} * 1\text{KB}/8\text{Kb} = 792296,3\text{KBps}$
IOPS = $792296,3\text{KBps} / 4\text{KB} = \mathbf{198074,075 \text{ IOPS}}$

2.3- Tràfic amb el client (entre servers i de server a switch de connexió a xarxa):

MB client: $1\text{KB}(\text{connexió}) + 10*400*8\text{KB}(\text{BD producte}) + 10*80\text{KB} + 30*64\text{KB}(\text{productes vistos}) + 30*80\text{KB}(\text{addicionals per producte vist}) + 0,2*16\text{KB} = 37,1242 \text{ MB}$
Tràfic client: $2.000.000\text{sessions}/1\text{dia} * 1\text{dia}/86400\text{segons} * 37,1242\text{MB}/1 \text{ petició} * 8\text{Mb}/1\text{MB} = \mathbf{6874,8519 \text{ Mbps}}$

2.4- Tràfic amb el disc:

MB disc: $256\text{KB} + 10*400*(8\text{KB}) + (30*64\text{KB}) + 0,2*(256\text{KB}) = 34,2272 \text{ MB}$
Tràfic disc: $2.000.000\text{sessions}/1\text{dia} * 1\text{dia}/86400\text{segons} * 34,2272 \text{ MB}/1 \text{ petició} * 8\text{Mb}/1\text{MB} = \mathbf{6338,3704 \text{ Mbps}}$

2.5- Pressió sobre la xarxa (ample de banda mínim necessito per servir el tràfic de client i disc). M'arriba?:

Tràfic total = Tràfic client + Tràfic disc = $6874,8519\text{Mbps} + 6338,3704\text{Mbps} = 13213,2223\text{Mbps} = \mathbf{13,2132 \text{ Gbps}}$
Degut a que disposem d'una LAN de 10 Gbps no ens arribaria, i per tant podríem desviar el tràfic mitjançant una SAN.

3.-Decisions preses

3.1- Descripció dels elements d'emmagatzematge escollits, en funció de les necessitats. Quants tipus de cabines? (i perquè), RAID escollit a cadascuna d'elles. Nombre de cabines de cada tipus

En primer lloc teníem clar que volíem un disc SSD, ja que tenim una gran quantitat de IOPS, i necessitaríem comprar molts més discs HDD que SSD. A més, aquests últims són més ràpids, més fiables i duren més.

Ni el consum elèctric ni el preu no són un limitant en el nostre cas, ja que tenim molts més diners dels que realment necessitem en el nostre sistema. És per aquest motiu que hem triat discs que fossin Enterprise, degut a la seva fiabilitat, rapidesa i rendiment superior a un disc Consumer.

Entre els SSD Enterprise hem escollit l'opció 10 de forma més o menys arbitrària (pero es l'opció amb el menor consum elèctric, i ja posats a triar millor no desperdiciar innecessàriament). A més,

aquest disc segons l'apèndix 6 té el percentatge de fallades més baix, per tant els discs són més fiables i ens duraran més.

Hem decidit escollir un RAID 61, ja que volem seguretat de dades, pel que hem decidit triar l'opció més segura de mirroring a més de disposar de doble paritat. Degut a que ni els preus ni el consum es diferencien tant hem escollit aquesta opció. A més, com anem sobrats de diners podem triar tranquils, i hem considerat que la seguretat és de les coses més importants a tenir en compte. Tenim pocs discos pel que considerem un sol clúster.

Com tenim poques dades (sense considerar snapshots), concretament 365,12 GB en total, les quals són 5,12GB d'usuaris i 360GB de productes no hem considerat que sigui bona idea separar-los, ja que no té sentit emmagatzemar 5,12 GB en discs de TB. Tot i haver-ho adjuntat, 365,12 GB segueix sent poques, per tant ens hi cap en una sola cabina.

En total degut al raid 61 tenim uns 8 discs, on tenim dos grups de quatre degut al mirroring, i d'aquests quatre tenim dos de dades i dos de paritat. A més, hem considerat afegir dos discs més per poder tenir spare disk, el qual permet detectar si algun disc està a punt de fallar i fer la transferència de dades a un d'aquests spare disk abans que falli del tot. Per sobre aprovisionament no fan falta discs extra, ja que multiplicar per deu la quantitat de dades (ignorant snapshots) o IOPS no ens altera el nombre de discs.

Per la cabina necessitem support RAID 61 i no necessitem una cache SSD, per el que escollim una entre la 2 i la 4 (en principi, per el numero de us i el preu, podem canviar a RAID 51 pero la cabina 1 segueix sent pitjor per no admetre spare disk). Les 36 badies de l'opció 4 son més de les necessàries i el preu és superior, per el que escollim una cabina de tipus 2.

3.2- Es justifica la necessitat d'un SAN? Si la resposta és si, raonar si el cost és assumible o no, i cas de no ser-ho calcular l'impacte sobre el rendiment del CPD

Disposem d'una LAN de 10Gbps, però tenim un tràfic de xarxa de 13,2312 Gbps. És per aquest motiu que hem decidit posar una SAN, per tal de poder dividir el tràfic, dirigint 6,8748519 Gbps per la LAN (la comunicació amb els clients) i la resta per la SAN (per la comunicació amb els discs). També hi ha tràfic en la LAN degut al mirroring, però és negligible. (Veure la secció 3.7). Hem decidit posar una SAN de 16Gbps per tal de tenir marge de creixement. Hem considerat que és molt important tenir accés a la xarxa, i si en un futur creix exponencialment i se'ns satura pot ser un greu problema. Hem calculat el cost que suposa tenir una línia de 8Gbps o una de 16Gbps (els demés casos ni els hem considerat ja que surten fora de les nostres necessitats), i degut a que tenim gran quantitat de diners hem decidit prioritzar que en un futur encara que creixem no ens afectarà. La diferència entre el cost econòmic de les dues opcions es de 200.000 euros.

Mirant el consum elèctric tenim uns 2,51 KW per part de la SAN de 8Gbps, i 5,35 KW la de 16Gbps. Degut a que també hem vist que la diferència en consum es poca, ens hem continuat quedant amb l'opció de 16Gbps.

3.3.- Posem un *mirror*?

Tenir un mirror extern és especialment important si hi ha algun desastre, ja sigui de desastres naturals, incendis o qualsevol altre incident que impliqui la destrucció total o parcial de les dades dels discs, evitant perdre totalment les dades emmagatzemades en ells i poder-les recuperar mitjançant aquest mirror.

Per aquest motiu hem decidit si tenir un mirror.

3.4- Empresa de *housing* escollida i perquè (relació entre el que ofereix, el que necessito i el que costa)

Hem escollit l'empresa de Mordor. El primer que ens ha cridat l'atenció sobre aquesta opció es el fet de tenir el millor en seguretat, i tenir la monitorització ja inclosa en el preu, tot i que aquest és un detall menys rellevant.

També ens hi hem fixat amb el % de downtime, el qual es baix, cosa que ens alegra. Tot i que no tenim un SLA especialment elevat, hem considerat que com millors condicions tinguem, millor. Ens ha interessat molt l'opció de que tinguí una segona línia de connexió de xarxa, per tal de que si cau la xarxa que no ens quedem sense servei, i com la possibilitat que caiguin les dos línies alhora és molt baixa, hem decidit per aquesta empresa de housing.

Hem descartat la segona opció ja que té un %downtime superior al de l'opció 3 i no ofereix segona línia. Pot semblar més barata però el fet de tenir major percentatge de downtime fa que el cost pugi per sobre del de l'opció 3. A més a més, el servei de monitoritzatge ve a part.

Finalment, hem decidit la opció 3 sobre la 1 perquè en aquesta t'obliga a comprar contenidors que hi caben fins a 19 racks de 42U, però al tenir tants pocs discs no ens surt a compte, ja que no creixen tant com per aprofitar aquesta inversió d'espai, i pensem que és millor invertir-los en seguretat. A més a més, és consumirà més. A part d'això té molts bons requisits, sobretot en termes de seguretat, cosa que ha fet que la decisió entre les opcions no fos una tasca fàcil. Però, finalment hem pensat que l'opció 3 ja ens va bé pel nostre volum de dades, i que amb l'opció 1 no acabem d'explotar els recursos que ens ofereixen.

Una dels altres aspectes que ens ha agradat de Mordor és que la ubicació de la infraestructura és secreta, augmentant encara més la seguretat de les dades.

3.5- Posem monitorització?

Degut a què hem triat l'opció 3, Mordor, aquesta ja inclou l'opció de monitorització, però en cas que no la portés l'haguéssim afegit ja que considerem que es important monitoritzar l'estat i correcta funcionament, i com hem comentat abans, al tenir bastants diners no ens afecta gaire aquesta decisió.

3.6- Opció de backup?

Hem decidit triar l'empresa MonSoonS3 MS3. L'hem triat ja que té les mateixes prestacions que l'opció 1, Microworks Apure, però els preus són més barats.

Segons el full de càlcul tenim 30400 GB en discs de memòria, degut a la replicació de raid 61 tenim 7600 GB efectius per utilitzar, que es un $\frac{1}{4}$ dels GB totals. Després, rentant-li 365,12GB ocupats per dades normals, ens queden 7234,88 GB lliures. Cada dia escrivim a memòria $0,2 \times 256 \text{KB} / \text{sessió} \times 2.000.000 \text{ sessions} = 102,4 \text{ GB}$ escrits per dia. Dividim 7234,88GB entre 102,4GB ens queda 70,65 dies, que són 70 dies snapshots diaris.

Podríem guardar 70 dies en el pitjor dels casos, però hem decidit tenir snapshots per 31 dies (per recuperar qualsevol instant dels últims 6 mesos a través dels backups), deixant espai per creixement d'usuaris, o per fer més snapshots en cas que sigui necessari. Això fa que la quantitat total de dades de les que fer backup siguin $365,12\text{GB} + 31 \times 102,4\text{GB} = 3539,52\text{GB}$.

3.7- Tràfic amb l'exterior afegit pel sistema de *backup/mirror* escollit. Quin *bandwidth* caldria?

Considerant les escriptures diàries de l'apartat anterior, el bandwidth necessari pel backup/mirror es $102,4 \text{ GB} / \text{dia} \times 8 \text{b/B} \times 1 \text{ dia} / 86400 \text{s} = 0,00948... \text{ Gbps}$. El tràfic total es de $6,87485 \text{ Gbps}$ (amb els clients) + $0,00948... \text{ Gbps}$ (mirroring) = $6,88433... \text{ Gbps}$.

És per aquest motiu que hem triat un bandwidth provider de 10 Gbps, que al tenir dos línies serien d'uns 20 Gbps. Hem decidit triar aquesta opció per assolir els requeriments de xarxa degut a tràfic d'Internet i tenir un marge per si creixem.

4.-Recomanacions als inversors

4.1.- Anàlisi de Riscos (*Risk Analysis*)

Quines desgràcies poden passar i com les hem cobert?

Al menys s'han de cobrir els següents casos:

- **Hi ha pèrdua d'un fitxer (per error o corrupció). De quan puc recuperar versions?**

Hem decidit tenir 14 backups, 7 per cada dia de la setmana anterior, 6 per cada mes anterior i un de l'any anterior. Podem recuperar versions desde 31 dies abans del backup de l'any anterior (fent ús dels snapshots) fins al dia anterior (i qualssevol punt intermig degut a que es fan backups cada mes i els snapshots son de 31 dies).

- **Es trenca un disc (es perden dades? quan trigo en recuperar-me? el negoci s'ha d'aturar?)**

Degut a disposar d'un RAID 61, si es trenca un disc no cal aturar el servei, ja que l'únic que hem d'anar fent és utilitzar el disc mirror que funciona i quan es pugui anar col·locant les dades allà

per tal de tenir el que teníem al principi abans que es trenqués el disc. A més disposem de spare disk, pel que en el moment de trencar-se o fer-se malbé un sector és possible que ja tinguem les dades en aquest spare disk. Per tant, no hauríem de perdre dades ni aturar el negoci. L'únic que cas que podem tenir es que es trenquin dos dels discs, l'original i el mirròr, que no tinguem disponibles els spare disk, i que es facin malbé quatre discs, dos per paritat, per tal de no poder recuperar les dades, però és un escenari que molt difícilment ens trobarem degut a que hem fet molt èmfasis en la seguretat de les dades i el nostre servei.

- **Puc tenir problemes de servei si falla algun disc?**

Veure apartat anterior.

- **Cau la línia elèctrica. Què passa?**

Al contractar el sistema de housing de Mordor tenim un generador de 72 hores de dièsel de repòs, per tant disposarem d'aquestes hores (o algunes mes, perquè pot fer re-ompliment en calent) per recuperar el servei. En cas que no puguem recuperar-ho, llavors perdrem molts diners. En cas que hi hagi corrupció de dades per estar enmig d'una escriptura quan cau la xarxa, podem recuperar les dades de disc (pero no de la RAM) degut a que tenim backups offsite amb l'empresa MoonSoonS3 MS3, tenint mirroring per tal de recuperar les dades ràpidament.

Si podem recuperar l'electricitat abans de les 72 hores (o més) llavors no haurem perdut res, com si no hagués caigut la línia. A més Mordor ens garanteix un 99,999% uptime, el qual són menys de 30 minuts en 5 anys, cosa que ens podem permetre i és acceptable.

- **Cau una línia de xarxa. Què passa?**

Degut a que utilitzem el sistema de housing de Mordor aquest ens ofereix doble línia de xarxa, per tant podem continuar funcionant correctament, a menys que també falli la segona, cosa bastant poc probable.

- **En cas de pèrdua o detecció de corrupció de dades no ens podem permetre seguir treballant fins que recuperem les dades correctes. Calculeu temps i costos de recuperació en cas de**

- **Pèrdua/ corrupció d'un 1% de les dades**
- **Pèrdua/ corrupció de la totalitat de les dades**

Pèrdua 100% dades:

Sense considerar costos per electricitat o internet normal.

Tenim 3539,52GB a recuperar.

Els discos tipus 10 poden fer 511K R / 82 K W IOPS, limiten les escriptures.

Amb RAID 61 cada operació d'escriptura es en realitat 6R+6W, podem fer 82K/6 escriptures reals (W) per disc * 8 discos (ignorant els spare disk) = 109333,33... W/s * 4KB/W = 437333,33... KBps = 437,33... MBps.

Tenim que recuperar 3.539.520.000 KB i recuperem 437.333,33... KBps, necessitem 8093,41... segons (2h15m).

El cost extra de Monsoon S3 per la recuperació de dades amb el mirròr contractat es de 0,05 euros/GB*3539,52GB = 176,976 euros.

Degut al downtime, els diners perduts són 2,248...h * 40000 euros/h = 89926,77... euros.

Pèrdua 1% dades:

Ara fem el 1% dels valors anteriors.

Necessitem 8093,41.../100 segons = 80,9341... segons (1m21s).

El cost per enviar les dades es de 1,77 euros.

Per downtime es de 899,26... euros.

4.2.- Anàlisi de l'impacte al negoci (*Business Impact Analysis*)

En funció de l'anàlisi de riscos anterior i del que costa estar amb la màquina aturada o no donar el servei complet, calcular quant perdo en diners per tenir-lo aturat i quan em costaria evitar aquesta situació.

Caiguda de la xarxa de dades:

Tenim un SLA de 40.000 euros per hora, i en 5 anys, 43830 hores, tenim entre un 0,00034% i 0,00071% de possibilitats de downtime degut a que tenim dues línies de xarxa. És a dir, que tindrem entre 0,15 i 0,31 hores de downtime. Per tant, pagarem entre 6.000 i 12.400 euros als clients durant aquests 5 anys.

No se'ns ocorre ninguna forma raonable d'evitar aquesta situació sense fer grans canvis, per tant hem d'assumir aquest cost.

Fallada de disc

Els nostres discs són els de tipus 10, els quals són Enterprise, SSD i TLC, per tant, la probabilitat anual que falli un disc d'aquests és de 0,45% anual. Degut a que tenim un RAID 61 podem fer la reconstrucció en calent, però anem a la meitat de IOPS. Segons SMART el 70% de les fallades es poden predir, i degut a que tenim spare disks només ens hem de preocupar per l'altre 30%.

Per tant si tenim 8 discos (10 en total tenint en compte que tenim spare disk), tenim:

$8 * 0,0045 * 0,3 = 0,0108 = 1,08\%$ probabilitat de fallada/any, el qual són 5,4% probabilitat falli en 5 anys. Cada cop tardem 4 hores per TB, com tenim discs de 3,8 TB, necessitaríem 15,2 hores per reconstruir un disc sencer.

No perdem diners per downtime.

No es beneficia evitar aquesta situació, ja que hauríem de reduir el nombre de discs, és a dir, reduint el RAID a aplicar, però això comporta altres problemes molt pitjors, com per exemple pitjor tolerància a fallades.

4.3.- Creixement

Si creix el nombre de clients/ màquines/ dades (depèn de l'escenari), hem d'estar preparats.

Quin creixement (en nombre de clients, etc...) podem assumir sense canviar el sistema (sobre aprovisionament)? Quin és el recurs que s'esgota abans? Feu un informe de les implicacions que suposaria un increment d'un 20% en el volum de negoci (tot, clients, dades, ...)

Estem suposant que reduir el nombre de snapshots es un canvi al sistema.

Creixement de sessions per dia

Bandwidth de internet: Tenim 20Gbps, utilitzem 6,88433... Gbps, es pot multiplicar per 2,9

Bandwidth de la LAN: Tenim 10Gbps, utilitzem 6,88433... Gbps, es pot multiplicar per 1,45

Bandwidth de la SAN: Tenim 16Gbps, utilitzem 6,33837... Gbps, es pot multiplicar per 1,57

Memòria per a snapshots: Tenim 7234,88 GB lliures (veure apartat 3.6) per a snapshots, tenim $31 \times 102,4 \text{ GB} = 3174,4 \text{ GB}$ ocupats per snapshots, es pot multiplicar per 2,279...

Bandwidth de discs: No limitant.

Les sessions per dia poden augmentar un 45% fins que ens quedem sense bandwidth a la LAN. Un augment del 20% no implica canviar res.

Creixement del nombre de productes o clients

Suposant que no hi ha un augment del nombre de sessions per dia i per tant de tràfic de internet.

Tenim 7600 GB d'espai efectiu per a dades, la memòria dels snapshots no augmenta i amb 31 dies de snapshots ens quedem amb $7600 \text{ GB} - 3174,4 \text{ GB} = 4425,6 \text{ GB}$ per a dades no replicades. Suposant que el nombre de productes i clients augmenta en la mateixa proporció simultàniament, tenim 365,12GB ocupats, es pot multiplicar per 12,12. Un augment del 20% no implica canviar res.

4.4.- Inversions més urgents

Donat el CPD resultant és possible que no haguem escollit la millor opció per manca de diners. El CPD no és nostre, nosaltres només ho dissenyem, així que al final s'hauria de fer un informe als que posen els diners de en què valdria la pena invertir per millorar rendiment, seguretat o...

En el nostre cas particular teníem molts diners en comparació amb les dades i sistema a gestionar, ja que tenim molt poques escriptures i les dades d'usuaris i productes ocupen molt poc. És per aquest motiu que hem pogut triar lliurement el que hem volgut sense preocupar-nos massa pel pressupost.

Hem pogut triar el millor en seguretat, housing i requisits que s'adeqüen perfectament al nostre sistema.