

Facultat d'Informàtica de Barcelona

Departament d'Arquitectura de Computadors

Centres de Processament de Dades

Activitat EBH

Emmagatzematge, *backup* i *housing*

Isábal Pallarol, Alicia

Dai, Natalia

Escenari 014

Data: 3/10/2022

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	50
Diners Totals	€5.000.000,00
Diners gastats	€3.500.000,00

taula 2: Elements que escolliu vosaltres	
Elements de mirror i backup	
GB a emmagatzemar al backup	1504
Dies entre 2 backups	7
Còpies senceres a mantenir	4
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)	0

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)	1
Cabina de discos	
Opció Disc principal (Entre 1 i 10)	8
Nombre de discos a comprar	24
Opció cabina de discos (Entre 1 i 6)	2
Nombre de Cabines	1

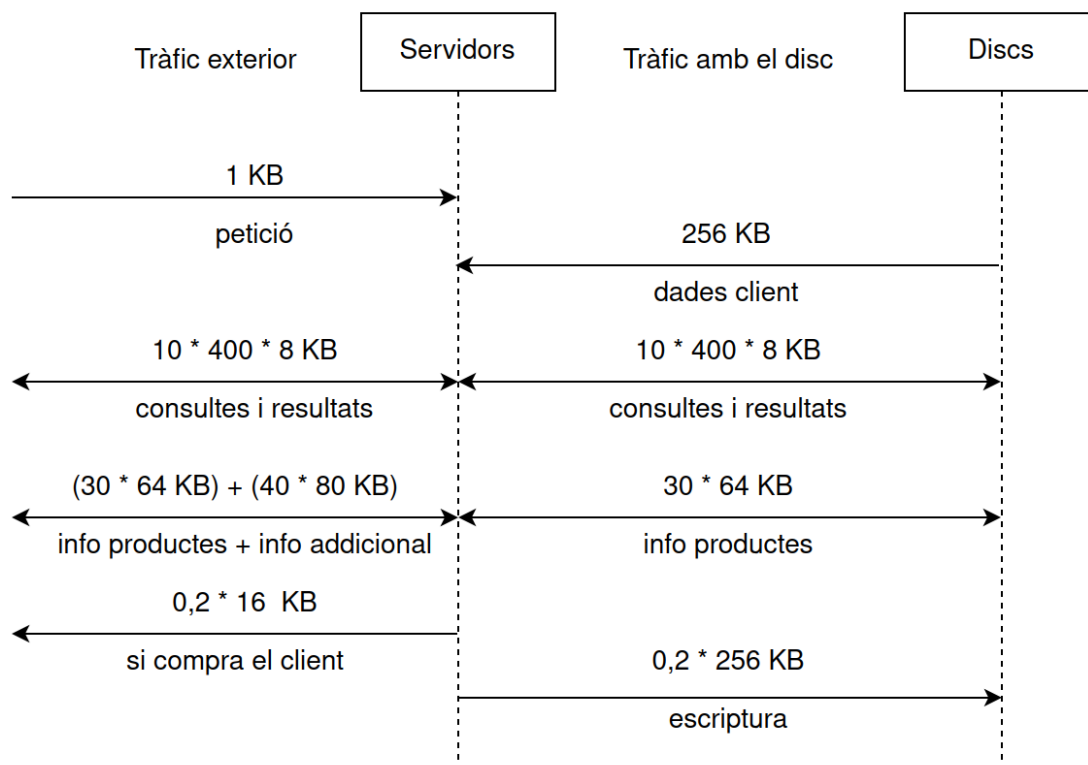
Cabina de discos 2 (cas de fer servir dos tipus)	
Opció Disc (Entre 1 i 10)	0
Nombre de discos a comprar	0
Opció cabina de discos (Entre 1 i 6)	0
Nombre de Cabines	0
Cabina de discos 3 (cas de fer servir tres tipus)	
Opció Disc (Entre 1 i 10)	0
Nombre de discos a comprar	0
Opció cabina de discos (Entre 1 i 6)	0
Nombre de Cabines	0

TAULA 3: OPEX	anual	cinc anys
Consum energètic (hardware només)	€65.364,64	€326.823,20
Empresa de Housing escollida	Mordor	
Cost Housing (inclou electricitat addicional)	€65.804,70	€329.023,48
Off-site: empresa escollida	MonsoonS3 MS3	
Cost mirror	€1.611,74	€8.058,72
Cost backup	€5.025,81	€25.129,03
Cost Bandwidth provider	€75.600,00	€378.000,00

TAULA 4: CAPEX	Cost
Diners gastats en servers, xarxa, etc	€3.500.000,00
SAN	€218.106,00
Sistema emmagatzematge	€13.528,00

TAULA 5: AJUST AL PRESSUPOST	
Opex a 5 anys, total	€794.874,43
Capex a 5 anys, total	€3.731.634,00
Despeses totals a 5 anys	€4.526.508,43
Diferència respecte al pressupost	€473.491,57

2.-Anàlisi de necessitats



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

Usuaris -> $250.000 * 256 \text{ KB} = 64 \text{ GB}$ (0.064 TB)

Productes -> $20.000.000 * (8\text{KB} + 64 \text{ KB}) = 1440 \text{ GB}$ (1.44 TB)

Total -> $64 \text{ GB} + 1440 = 1504 \text{ GB}$ (1.504 TB)

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

-> En aquest escenari les escriptures són pràcticament nul·les -> $(0,2 * 256 \text{ KB}) / 34227,2$

(aquest surt de l'apartat 2.4 - cada petició) = 0,0015%, quasi 0% (44,45 IOPS)

-> $300.000 \text{ peticions/dia} * 1 \text{ dia} / 86400 \text{ segons} * 34227,2 \text{ KB/petició} = 118844,45 \text{ KBps}$

-> Els discos accedeixen en blocs de 4 KB, per tant $118844,45 / 4 = 29711,1125$;

arrodonint seria 29712 IOPS en total.

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

Cada petició -> 1 KB (petició) + (10 x 400 x 8 KB) (consulta i resultats) + (30 x 64 KB + 40 x 80 KB) (informació productes amb detalls + addicional) + (0,2 x 16 KB) (en cas de si compra) = 1 KB + 32000 KB + 5120 KB + 3,2 KB = 37,1242 MB

300.000 sessions/dia x 1 dia/86400 segons x 8 Mb/1 MB x 37,1242 MB/sessió =

1031,228 Mbps

2.4- Tràfic amb el disc:

Cada petició -> 256 KB (dades client) + (10 x 400 x 8 KB) (consulta i resultats) + (30 x 64KB) (informació productes amb detalls) + (0,2 x 256 KB) (escriptura) =

256 KB + 32000 KB + 1920 KB + 51,2 KB = 34,2272 MB

300.000 sessions/dia x 1 dia/86400 segons x 8 Mb/1 MB x 34,2272 MB/sessió =

950,76 Mbps

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

Tràfic exterior o entre servidors -> 1031,228 Mbps

Tràfic amb disc -> 950,76 Mbps

Tràfic total -> 1981,988, arrodonint -> **1982 Mbps**

Com l'escenari té una xarxa de 2 Gbps, ens arriba, però amb poc marge, necessitarem 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, hem triat estructurar l'emmagatzematge en RAID 51, ja que ens sembla una opció efectiva a nivell de cost i que ens assegura tenir tolerància a fallades així com ens proporciona redundància. Hem triat els discs de tipus 8 (Kingston SEDC100M), 10 discs en total -> 8 (3 dades + 5 redundància (4 per RAID 1)) + 2 (spare disks).

Aquesta estructura ens permet tenir un creixement que ens sembla apropiat per poder mantenir la infraestructura de l'escenari tal que:

TB = un disc és 1920 GB -> x 3 =5760 GB -> 5.76 TB i utilitzant només 1.504 TB, ens permet créixer 4,356 TB: **2,82** vegades.

IOPS = tenim 29712 IOPS en total -> 29667 L / 45 W -> però en raid 51 seria:

30024 IOPS -----> 29846 L / 178 W (una escriptura es (2R+2W) x2)

Els discs d'aquest tipus ens ofereixen 540.000 L / 205.000 W per disc, veient com clarament estem molt per avall a nivell d'IOPS, tot restant el que tenim al nostre cas, tenim que ens sobren 510154 L / 204822 W per disc, el que són -> **17.1 vegades més de lectures /1150.7 vegades més d'escriptures.**

Tenint en compte la poca quantitat d'informació que emmagatzemem dels usuaris, hem decidit usar únicament una cabina. En concret, hem triat la cabina 2, ja que per les nostres necessitats ens sembla la més adient: no necessitem cabines de major capacitat i al usar SSD no necessitem el suport dels tipus 4 i 5.

Tenim 24 badies amb els que farem 2 grups: 12 discs -> (5 discs dades + 1 paritat **RAID5**) x 2 **RAID1**; 10 discs -> (4 discs dades + 1 paritat **RAID5**) x 2 **RAID1**

TOTAL = 9 discs en total de dades, 13 de redundància, 2 de spare disc = 24.

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

Com hem vist a la secció anterior, tenim 1.98Gbps de tràfic i disposem d'una LAN de tan sols 2 Gbps, això limita moltíssim el nostre creixement (podem créixer un 1%) pel que sigui quin sigui el cost de la SAN l'hem de poder assumir necessàriament, per això hem decidit posar la de 8Gbps, que costa poc més de 218k €, per ajustar-nos millor al nostre pressupost.

3.3.- Posem un *mirror*?

La nostra política de backups inclou un backup setmanal de 1504 gb, que, al cap i a la fi, no és una quantitat excessiva de dades, amb un mirror el pressupost ens segueix quadrant perquè la diferència amb no tenir-lo no és tan gran. Per això la nostra proposta sí inclou un mirror i no backup onsite, la qual cosa ens facilitarà un risc menor de tenir temps de downtime degut a la restauració de backups. Amb això podem triar entre les empreses 1 o 2. Hem triat la segona en ser una solució més efectiva a nivell de cost.

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

Havent analitzat les opcions de Housing i hem vist que la diferència entre l'opció 2 amb monitorització i l'opció 3, amb monitorització inclosa, tenen una diferència de preu de mil euros, per això hem triat la 3, ja que el PUE és menor que el de la 2 (1,15 opció 3 vs 1,4 opció 2) i el downtime és de màxim 1,6 hores a l'any mentre que l'opció 2 el downtime és d'un màxim de 22 hores a l'any. Pel que fa a la resta de característiques, són semblants i/o poc significatives pel nostre cas, a més amb aquesta opció podem tenir una segona línia gratis. Finalment, backup local no en farem, considerem que la política de backups que hem dissenyat ja és suficient.

3.5- Posem monitorització?

Ens interessa la monitorització per quan hi hagi potencials fallades de disc (tècnica SMART) i ens puguin canviar els discs. Ja ve inclosa en la nostra opció de Housing.

3.6- Opció de backup?

En tenir mirror només podem triar entre les opcions 1 o 2, hem acabat triant la 2 ja que la diferència de preu és de 3000€ i ofereixen el mateix, no tenint gairebé escriptures (0.0015%) i tenint mirror, ens permet recuperar ràpidament dades del mirror, necessitem poca informació pel backup que, onsite, no ens cal.

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

Contractem 2 línies d'10 Gbps, una d'elles amb un 2n proveïdor, en cas que el primer tingués algun problema. A més, per part de l'empresa de Housing del mirror ens ofereixen una segona línia gratuïtament. Aquesta decisió està basada en el fet que tenim diners suficients per aquesta capacitat de les línies.

Actualment, el nostre sistema compta amb poques escriptures, si això canviés hauríem d'augmentar el nombre de línies, per sort disposem de superàvit al nostre pressupost que permetria fer aquest augment en qualsevol moment.

El nostre bandwidth amb l'exterior és afegit pel mirror i pel backup tal que tenim:

Tenim = 51'2 KB → 256KB x 0.2 compra

$$Bandwidth = \frac{300.000 \text{ peticions}}{1 \text{ dia}} \times \frac{1 \text{ dia}}{86400 \text{ s}} \times \frac{51,2KB}{1 \text{ petició}} \times 8 \text{ bits} = 1422,22 \text{ Kbps}$$

El nostre tràfic total (calculat a l'apartat 2.5) és de 1982 Mbps + 1,422Mbps = 1983,4 Mbps.

Possible creixement = 20000 Mbps / 1983,4 Mbps = 10,08x

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?**

Podem recuperar versions de l'última setmana, l'últim de l'any, l'últim del trimestre i l'últim del mes, a més que fem backup cada setmana (7 dies). Aquest no és el millor escenari per a qüestions d'errors o corrupció, per la qual cosa comptant amb que tenim sobreprovisionament de disc, podem pensar a implementar snapshots.

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

No hauria d'afectar al rendiment del negoci ja que com tenim RAID 51, si es trenca un o fins i tot 2 discos, sempre tenim una còpia.

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

No hauríem de tenir, ja que tenim RAID 51.

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

Tenim contractat Mordor, que ens ofereix dues línies d'entrada d'electricitat i dues línies de connexió de xarxa, llavors tenim redundància per si de cas cau una línia elèctrica. A més, té un generador dièsel amb capacitat per aguantar la potència pic durant 72 hores, així que no hauria d'haver-hi problema.

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

Tenim contractades 2 línies de xarxa, si cau una tenim l'altra, encara que d'un proveïdor diferent per si el problema vingués de l'operador. Amb una línia en realitat ens va bé igualment de velocitat amb la SAN que posarem.

- **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**

Com tenim totes les dades en RAID 51, és molt difícil que hi hagi necessitat de recuperació d'un 1% de les dades (1,504 TB en total), ja que és poc probable que es corrompeixi simultàniament un disc i el ser mirror (redundància).

- **Pèrdua / corrupció de la totalitat de les dades**

Velocitat de recuperació, $(205K (W) / 4 \text{ escriptures reals}) * 22 \text{ discos} * 4kB/W = 4510kBps$

—> 4.510 MBps

Amb una ocupació de $1,504TB / 9 * 1,92TB \text{ (capacitat disc)} * 100 \rightarrow 8,7\% \text{ ocupat}$.

$9 \text{ discs de dades} * 1504GB * 0,087 * 1 \text{ (totalitat de dades)} = 1177.632 \text{ GB}$.

Temps total de recuperació: $(1177.632 * 10^3) / 4.510 = 261.116 \rightarrow / 60 = \mathbf{4.35 \text{ min}}$

Com tenim el mirror contractat, són 0.05 € per GB i tenim 1504 GB -> **75,2 €**

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 complert, calcular quant perdo en diners per tenir-lo aturat i quan em costaria evitar aquesta situació.

Caiguda de la xarxa de dades:

Com sabem:

- La probabilitat de downtime d'una línia està entre el 0,017% i el 0,036% .
- La probabilitat de downtime de dues línies està entre el 0,00034% i el 0,00071%

Com el nostre SLA és de 40.000 euros per hora de downtime:

- Si és una línia, seria entre 7,45 h i 15,77 h (arrodonint 8 i 16 hores), i hauríem de pagar una penalització de entre 320.000 - 640.000 euros .
- Si són dues línies seria solament 1 hora de downtime -> 40.000 euros de penalització.

Fallada de disc

Com tenim implementat RAID 51, si falla un disc puc copiar-lo del mirròr, no cal reconstruir. Tot i així, cal fer la còpia del disc, Si es pot predir la fallada per SMART (70%) es farà la còpia quan el clúster estigui inactiu. A més, tenim spare discs. Si falla es podrà atendre els accessos sense problema.

Tenim 24 discos (9 de dades i 13 de RAID 51). Són discos SSD Enterprise TLC (Kingston SEDC100M), per tant la probabilitat de fallada és un 0,45 anual. D'aquestes fallades, com tenim SMART, sols tenim que reconstruir el 30%. Per tant, si tenim 24 discos x 0,45% fallada x 30% fallades amb reconstrucció, els discos tenen un 3,24% de probabilitat de fallar per any, i un 16,2% de que falli un en 5 anys.

De totes formes, els IOPS estan molt sobredimensionats (apartat 4.3), per tant, si necessito copiar un disc en un altre és ràpid.

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 (sobreprovisionament)? 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, ...)

Tenim 24 discos per tot com s'havia esmentat:

24 discos -> necessitem 30024 IOPS -> 29846 L / 178 W (una escriptura es $(2R+2W) \times 2$)

ens donen -> 540.000 L / 205.000 W per disc

x 24 discos -> 12.960.000 L / 4.920.000 W

es pot créixer un **434,23x en lectures i 27640,45x en escriptures en IOPS**,

així que no ens preocuparia un 20% més en el volum de negoci.

9 discos de dades -> capacitat = $9 \times 1920 \text{ GB} = 17280 \text{ GB}$ i requerim 1504 GB, és a dir, es

pot créixer un **11,49x en dades**, així que no ens preocuparia un 20%

més en el volum de negoci.

La pressió sobre la xarxa és de 1.98 Gbps i tenim una de 2 Gbps, sabem que hi haurà

problemes ja que té solament un 1,01x de creixement, i per això proponem posar una

SAN de 8 Gbps. És a dir, podem créixer un **4,04x a la pressió sobre la xarxa**, així que no

ens preocuparia un 20% més en el volum de negoci.

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...

Com hem esmentat abans, necessitem una SAN i podríem augmentar la freqüència dels backups (ja que tenim la capacitat d'emmagatzemar-los) de 7 dies a menys dies. També seria interessant augmentar els nostres productes i així tenir més clients.