

Activitat 1

Emmagatzematge, *backup* i *housing*

**Estudiant 1 cognoms, nom
Estudiant 2 cognoms, nom**

Escenari Exemple

Data: DD/MM/AAAA

1.-Descripció bàsica

Escenari: Exemple

Copy & Paste del full de càlcul "Resum"

2.-Anàlisi de necessitats

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

Mida de les dades:

- Tenim 400.000 llogaters i 30.000 propietaris i emmagatzemem 6 MB de dades de cadascun, per tant necessitem $430.000 \times 6\text{MB} = 2.580.000 \text{ MB} = 2.58 \text{ TB}$
 - Per altra banda tenim 3.000.000 propietats $\times 75 \text{ MB} = 225.000.000 \text{ MB} = 225 \text{ TB}$
- Total= $2.58+225= 227.58\text{TB}$

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

- En aquest escenari les escriptures són pràcticament nul·les (2 KB de cada 584002 KB, un 0,0037%)
- 5000 peticions/dia $\times 1\text{dia}/86400 \text{ segons} \times 548002 \text{ KB/ petició} = 31713,08 \text{ KBps}$
- Els discos accedeixen en blocs de 4KB, per tant $31713,08 / 4 = 7928,26968 \text{ IOPS}$; arrodonint 7930 IOPS

2.3- Tràfic amb el client (exterior, però també de server a switch de connexió a xarxa):

Cada petició necessita 1KB (sol·licitud) + $40 \times 512\text{KB}$ (comunicacions client) + un 25% dels clients 18KB (lloguen)= $0.75 \times (512 \times 40 + 1) + 0.25 \times (512 \times 40 + 1 + 18) = 20485,5 \text{ KB / petició} = 20,4855 \text{ MB / petició}$

$5000 \text{ peticions/dia} \times 1\text{dia}/86400 \text{ segons} \times 20,4855 \text{ MB/ petició} \times 8 \text{ Mb/ MB} = 9,484 \text{ Mbps}$

2.4- Tràfic amb el disc:

Cada petició necessita 512 MB (inicials) + 18 comunicacions de 2MB + un 25% fan 8 KB addicionals = $548,002 \text{ MB / petició}$

$5000 \text{ peticions/dia} \times 1\text{dia}/86400 \text{ segons} \times 548,002 \text{ MB/ petició} \times 8 \text{ Mb/ MB} = 253,705 \text{ Mbps}$

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

Trafic exterior o entre servidors= 9.484 Mbps

Traffic amb disc = 253.705 Mbps

Tràfic total = $9.484 + 253.705 = 263.189 \text{ Mbps} = 0.263189 \text{ Gbps}$

L'escenari té una xarxa de 1Gbps, ens arriba.

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

Observem que les dades més crítiques són les de clients, i no les de propietats. Per tant decidim tenir dues cabines, una per clients i altra per propietats. Decidim posar els clients en alta replicació i les propietats en RAID 5, ja que tot replicat surt molt car i les dades de clients són crítiques. Els IOPS dels clients són molt petits, i només tenim 2.58TB.

Recalculem Capacitat i IOPS del servidor de clients i de propietats

Servidor de clients:

- Capacitat = 2.58 TB
- Escriitures = 0.000027
- IOPS=

Cada petició= $75\text{MB} + 25\% \times 0.008\text{MB} = 75,002 \text{ MB/petició}$

$(5000 \text{ peticions})/\text{dia} \times (1 \text{ dia})/(86400 \text{ segons}) \times (75,002 \text{ MB})/\text{petició} \times 1000\text{KB}/(1 \text{ MB}) \times (1 \text{ IO})/4\text{KB}$
= 1085 IOPS

Servidor de propietats:

- Capacitat = 225 TB
- Escriitures = 0
- IOPS=

Cada petició= $437\text{MB} + 18 \times 2\text{MB} = 473\text{MB/petició}$

$(5000 \text{ peticions})/\text{dia} \times (1 \text{ dia})/(86400 \text{ segons}) \times (473 \text{ MB})/\text{petició} \times 1000\text{KB}/(1 \text{ MB}) \times (1 \text{ IO})/4\text{KB} =$
6843 IOPS

Després de avaluar les dades, ens decidim per:

Clients: Opció 8 de disc (SSD Enterprise) en RAID 51 (cal un mínim de 6 discos)

Els discos de l'opció 8 tenen una capacitat de 1920GB, com estic en RAID 51 només 2 dels 6 discos tenen dades reals, pel que tinc 3,84TB GB disponibles. Permet creixement, però pel preu dels discos afegeixo 2 més, quedant 8 discos en RAID 51 (3 de dades reals = 5.76 TB)

Donat que són dades molt problemàtiques, és interessant disposar de *spare disc*. La cabina 2 és la més adequada, ja que amb 24 badies en tenim prou i no ens cal suport SSD. Afegirem 2 discos de *spare*

Total: 1 cabina de tipus 2 + 10 discos de l'opció 8 (8 + 2 spare)

Propietats:

La majoria de les IOPS són per les propietats, que requereixen 225TB d'espai. Mirant les opcions ens decanem per l'opció 3 de disc en RAID 5.

El mínim de discos requerits per RAID5 són 23, però per cobrir les dades actuals, sense creixement.

Tenint en compte que tinc que fer clústers pel RAID 5, suposo que fem clústers de 6 discos (5 dades + 1 degut al RAID 5). Faria 5 clústers de 5+1 discos, el que dona 30 discos. La cabina més

gran és de 36 badies i voldré afegir discos de spare (puc afegir fins a 4). Es pot observar que amb una cabina pràcticament no tinc creixement possible. Per tant escollim tenir dues cabines. El mínim són 23 discos, mirem de tenir al menys el doble donat que el preu per disc no és tan alt. Una opció serien 46 discos, però farem clústers de 6 discos (5+1) per tant fem 10 clústers de 6 discos (60 discos, 50 per dades i 10 degut al RAID). 5 clústers per cabina. Escollim cabines de 36 badies amb SSD (els discos són HDD, per si hi hagués pics d'escriptura en, per exemple, una recuperació de backup). Com ocupem 30 de les 36 badies, posem 4 discos de spare.

Total: 2 cabines tipus 5 i 68 discos de tipus 3 en RAID 5

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

El tràfic total (extern + disc) és de 263.189 Mbps = 0,263189 Gbps.
Podem créixer un 380% el tràfic total sense saturar la xarxa.

3.3.- Posem un *mirror*?

El cost del mirror és alt (gairebé €200.000/any) però pot reduir molt el temps de recuperació de dades. Tenint en compte que cada hora de downtime és €1.500.000, realment surt a compte. Per això ja sabem que l'empresa Off-site ha de ser l'opció 1 o 2.

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

Tenint en compte l'alt cost per hora de downtime, ens decidim per **Mordor**, que és la única que ens ofereix la possibilitat de contractar una segona línia de dades (i encara que la probabilitat de que falli la línia és petita, el cost de la línia és també molt petit).

3.5- Posem monitorització?

Encara que només sigui perquè comprem cabines de disc amb spare disc, ens interessa monitorització i que ens vigilin quan la tècnica SMART avisi i ens canviïn els discos. Per tant escollim monitorització.

3.6- Opció de backup?

Com tenim *mirror*, només pot ser la opció 1 o 2. La 2 (**Monsoon S3**) és lleugerament més barata en el nostre escenari (la suma de mirror més backup, només el mirror és lleugerament més barat l'altre opció).

Pel que fa als dies entre 2 backups, al no haver pràcticament escriptures i tenint un mirror, el que ens permet recuperar ràpidament dades del mirror, necessitem poca informació a backup, doncs es faria per recuperar versions velles dels fitxers. Per tant, i buscant un preu raonable, farem backup cada 8 dies. Guardarem 14 còpies (2 últimes i una mensual de l'últim any) i no ens caldrà

tenir un sistema de backup on-site ja que el mirròr ens permet recuperació i no cal enviar còpies off-site contractar mirròr i backup amb una empresa externa. Al no tenir backup local posem tant la casella "Sistema de backup on-site" i "gestió local de backup" a 0.

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

Amb una xarxa interna d'1 Gbps, contractem una única línia d'1Gbps. Però amb la possibilitat d'una segona línia.

Si tinguéssim molt de tràfic (moltes escriptures) hauríem de calcular si l'ample de banda del nostre sistema és suficient Això pot fer-nos canviar en la decisió de posar o no SAN.

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?**

Faig un backup en cinta cada 10 dies. Per tant els fitxers a recuperar poden ser un problema si hi ha corrupció o esborrat erroni.

En canvi, tinc bastant sobreprovisionament de disc, de manera que puc pensar en implementar snapshots per si haig de recuperar

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

Les dades de clients estan en RAID51. El trencament d'un disc a la part de clients no afecta al rendiment per haver replicació. És fàcilment recuperable amb la còpia, i el fet de tenir-lo en RAID 5 garanteix que tampoc passa res si es trenquessin dos discos. Més de dos implicaria recuperació del mirròr extern.

Les dades de propietats estan en RAID5. La recuperació es pot fer per mitjà de recuperació de RAID a costa de rebaixar l'ample de banda amb disc.

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

En cas de les cabines de clients no hi ha problema al haver RAID 51, però en les cabines de dades de propietaris al recuperar discos RAID5 podríem tenir una penalització a les IOPS.

El disc escollit (opció 5) té 6TB de capacitat i 118 IOPS. Hem de cobrir 7930 IOPS. Tenim 68 discos de dades (la resta dels comprats són pel RAID5 i *spare*)

- 68 discos x 118 IOPS són 8024 IOPS que puc servir.
- N'hi ha tres cabines. Si fallés un disc, només una de les cabines tindria penalització per reconstrucció del disc. En realitat serveixen $7930/3 = 2643.3$ IOPS per cabina
- La cabina penalitzada estaria amb problemes 24 hores (penalització de 4 hores per TB, discos de 6 TB) a meitat de velocitat

- Per tant, durant 24 hores serviria 2 cabines al complet (2643.3×2) i una cabina la meitat (1321.6), o sigui 6608,33 IOPS

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

Mordor ens ofereix dues connexions a xarxa elèctrica i un generador dièsel que aguanta 72 hores de funcionament. No n'hauria d'haver problema.

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

Hem contractat la segona línia de xarxa. Tot i això la probabilitat no és nul·la (veure Següent punt)

- **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**

Si hem de recuperar dades per corrupció ens centrem en les de propietats, doncs un RAID51 en les de clients fa difícil que hi hagin necessitat de recuperació d'aquest volum de dades. En el nostre cas, al recuperar del mirròr/ backup extern, la limitació està en la xarxa interna: tenim 1 línia d'1 Gbps amb el proveïdor de mirròr/ backup i una xarxa interna d'1Gbps.

Com tinc una pressió a la xarxa de 0,263189 Gbps, tinc disponible $1 - 0,263189 \text{ Gbps} = 0.736811 \text{ Gbps} = 736.811 \text{ Mbps} / 8 = 92,1 \text{ MBps}$.

Tinc 225 TB de dades (225.000.000MB)

Recuperar la totalitat seria $225.000.000 \text{ MB} / 92,1 \text{ MBps} = 678.6 \text{ hores}$ (o sigui, una mica més de 28 dies)

Perdre un 1% de les dades significa 6.786 hores (6h47')

Recuperar un disc al complet (6TB) costaria 18 hores, mentre que reconstruir un RAID 5 serien 12 hores a menys IOPS.

Recuperar una propietat perduda - esborrat accidental- de 75MB és menys d'1 segon.

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:

Segons les dades que tenim (apèndix 6) , una línia cau 1 hora cada 18 mesos, entre 1 i 3 hores (o sigui, $2 \pm 1 \text{ h}$) cada 3 anys (36 mesos) i entre 3 i 9 hores ($6 \pm 3 \text{ h}$) cada 6 anys (72 mesos)

Per tant la probabilitat de caiguda mensual és de $1h/18m + (2h \pm 1h)/36 + (6 \pm 3h)/72m = (4+4 \pm 2+6 \pm 3 h)/72m = 14 \pm 5 \text{ hores} / 72 \text{ mesos}$

Un mes té en mitja $365,25 / 12 \text{ dies de } 24 \text{ hores (compto l'any de traspàs)} = 730,5 \text{ hores}$

$72 \text{ mesos} \times 730,5 = 52.596 \text{ hores}$

Una línia està penjada entre 9 i 19 hores ($14 \pm 5 \text{ hores}$) de cada 52.596, o sigui que la possibilitat de downtime es de entre $9/52596$ (0,017%) i $19/52596$ (0,036%).

En cas de dues línies, la caiguda de les dues simultàniament és $1/(18 \times 18) + 2 \pm 1 / (36 \times 36) + (6 \pm 3/72 \times 72) = 20 \pm 7h$ de cada 5184 mesos, o sigui $20 \pm 7 \text{ hores}$ de cada 3.786.912. Per dues línies, la probabilitat d'estar downtime està entre 0,00034% i 0,00071%

NOTA: els càlculs anteriors serveixen per tots els casos, el que pot variar en cada escenari és el que costa. En el nostre cas €1.500.000 per hora de downtime.

En cas de tenir una línia, en 5 anys (43830 hores) tenim entre un 0,017% i 0,036% de possibilitats de downtime, o sigui entre 7.45 (8 hores) i 15.77 (16 hores). Per tant la penalització que pagarem als clients estarà entre 12 i 24 milions d'Euros (impossible de pagar)

En canvi, al tenir dues línies el màxim en 5 anys seria 1 hora de downtime, o sigui 1,5 milions d'euros.

Fallada de disc

Per reconstruccions de disc puc tenir problemes d'accés en IOPS.

Cabina de clients: estem en RAID 51. Si falla un disc puc copiar-lo del mirror, no cal reconstruir. Cal, però, fer la còpia del disc. Si es pot predir la fallada per SMART (70%) es farà la còpia quan el clúster estigui inactiu. Si falla es podrà atendre els accessos sense problema.

Tenim 8 discos (3 de dades i 5 per RAID). És un disc SSD enterprise TLC, per tant la probabilitat de fallada és un 0,45% anual (veure apèndix 6). De les fallades, només el 30% requereixen reconstrucció, la resta es preveuen per SMART. Per tant $8 \text{ discos} \times 0,45\% \text{ fallada} \times 30\% \text{ fallades amb reconstrucció}$, els discos tenen un 1,08% de probabilitat de fallar per any, un 5,4% de probabilitat de que falli un disc en 5 anys.

De totes formes, els IOPS estan sobredimensionats x4000. Per tant, si necessito copiar un disc en un altre és ràpid.

Cabina de propietats: El disc escollit falla cada 2,84% anual (HDD Enterprise ≥ 10000 rpm), tinc 60 discos (30 per cabina més els spare). Fallen $60 \times 0,0284 = 1,704$ discos per any $\times 5$ anys = 8.52 discos.

El 70% de les fallades es poden predir per SMART, per tant les que s'han de reconstruir són el 30%, $8.52 \times 0,3 = 2.556$ discos en tots (posem 3)

Cada cop que reconstrueixo un disc en RAID 5 triguem 4 hores per TB. Els discos són de 10TB = 40 hores de reconstrucció. Durant aquest temps els discos d'aquell clúster van a la meitat de velocitat. Tinc 10 clústers de 6 discos, 9 a tot funcionament i 1 a mig funcionament. Cada clúster té 6 discos a 710 IOPS = 4260 IOPS per clúster $\times 9,5$ clústers en funcionament = 40470 IOPS, l'escenari demana 6843 IOPS, sense problemes

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, ...)

Al final s'han comprat 60 discos (per dades) per propietats. A 10TB per disc, són 500 TB de dades, com cada propietat són 75MB de dades caben 6.666.666 propietats (un 222% de creixement)

Pel que fa a IOPS, amb 60 discos podem atendre fins a 42600 IOPS i l'escenari bàsic demana 6843 (puc créixer un 632%)

Pel que fa als clients, els IOPS és gairebé 4000 vegades el necessari. En dades, tenim 3 discos de dades amb 1,92TB per disc = 5,76TB. Tenim 30.000 propietaris i 400.000 clients que ocupen 6MB cadascú, per tant podem tenir 5.760.000 MB/ 6MB per client = 960.000 clients, per tant podem créixer en 530.000 clients addicionals (entre propietaris i llogaters), o sigui un 123%.

La pressió sobre la xarxa és de 0,26 Gbps i tenim una d'1 Gbps. No hi ha problema (en aquest escenari).

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...