



ESCENARI EXEMPLE



David López  
V1.1. Primavera 2021

### Visió gràfica

Tràfic exterior

Servers

Tràfic amb disc

Discs

Temps

Temps

### Visió gràfica

Tràfic exterior

Servers

Tràfic amb disc

Discs

Petició 1KB

Temps

Temps

ESCENARI EXEMPLE

### Visió gràfica

Tràfic exterior

Servers

Tràfic amb disc

Discs

Petició 1KB

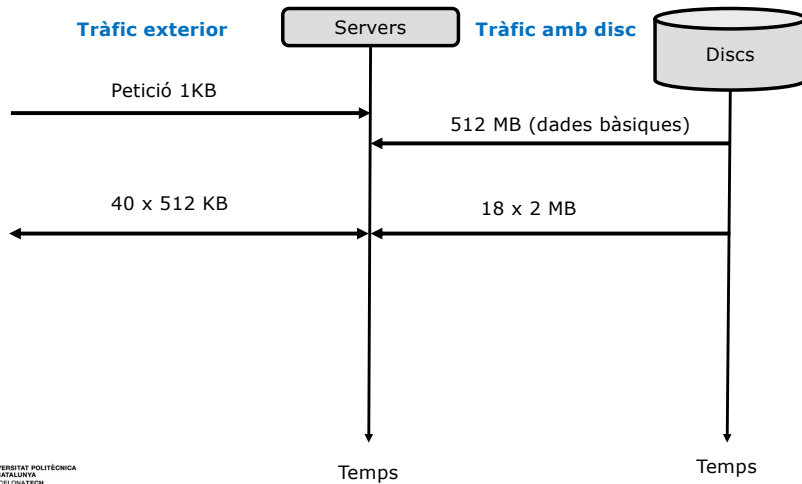
512 MB (dades bàsiques)

Temps

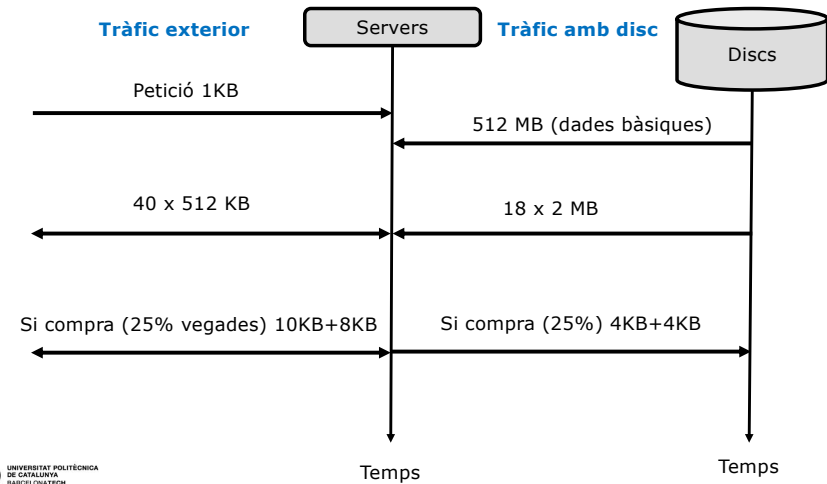
Temps

ESCENARI EXEMPLE

## Visió gràfica



## Visió gràfica



## Pressió per capacitat

Usuaris:  $(400.000 \text{ clients} + 30.000 \text{ propietaris}) \times 6 \text{ MB} = 2.580.000 \text{ MB} = 2.58 \text{ TB}$

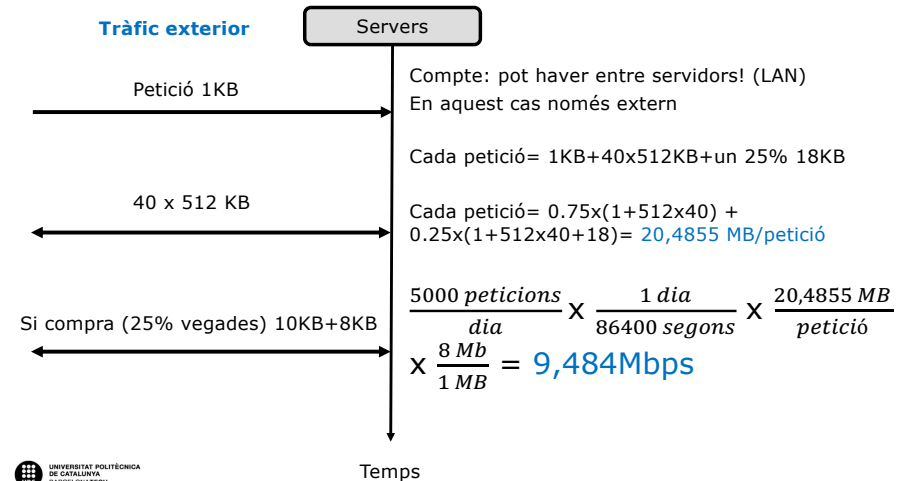
Propietats:  $3.000.000 \text{ propietats} \times 75 \text{ MB} = 225.000.000 \text{ MB} = 225 \text{ TB}$

Total=  $2.58 \text{ TB} + 225 \text{ TB} = 227.58 \text{ TB}$

És molt, però fixeu-vos que no tenen els mateixos requeriments (SLA)

- Podem fer 2 tipus de cabines (per exemple RAID51 en usuaris i RAID 5 en propietats)

## Tràfic amb el client

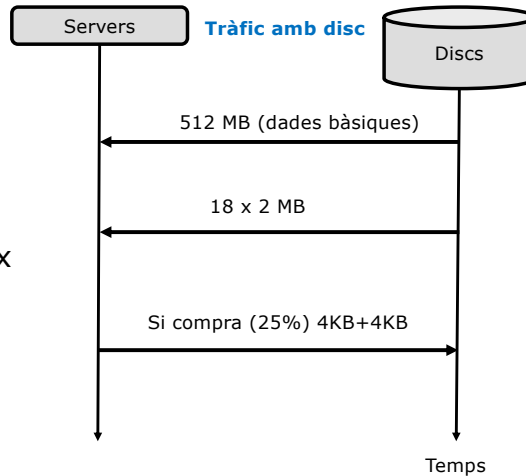


## Tràfic amb el disc

Cada petició= 512MB+18x2MB+  
25%x0.008MB

Cada petició=  $0.75 \times (512 + 18 \times 2) + 0.25 \times (512 + 18 \times 2 + 0.008) = 548,002$   
MB/petició

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



ESCENARI EXEMPLE

9

## Tràfic

Entre servidors o servidors i exterior:

Atenció: en aquest cas només tenim amb l'exterior. No oblideu entre servidors si cal.

Tràfic= 9,484 Mbps

Entre servidors i disc: 253,705 Mbps

Tràfic total= 9,484 Mbps + 253,705 Mbps= 263,189 Mbps = 0,263189 Gbps

Tenim una xarxa de 1Gbps. Tenim de sobres

- Cal dealtar, podríem desviar el tràfic de disc per mitjà d'una SAN

ESCENARI EXEMPLE

10

## Pressió per IOPS

Escriptures: Pràcticament nul·les -> 2KB de cada 584002KB: un 0,00036% (no sempre és així). Considerarem un 0% d'escriptures

$$\text{Pressió: } \frac{5000 \text{ peticions}}{\text{dia}} \times \frac{1 \text{ dia}}{86400 \text{ segons}} \times \frac{548002 \text{ KB}}{\text{petició}} = 31713,07 \text{ KBps}$$

Els discos accedeixen en blocs de 4KB, per tant  $\frac{31713,07 \text{ KB/s}}{4 \text{ KB/operació IO}} = 7928,26968 \text{ IOPS}$   
(arrodonint = 7930 IOPS)

Això és pressió mínima. Ara cal calcular quina arquitectura de disc utilitzo

ESCENARI EXEMPLE

11

## Prendre decisions

Total= 227.58

- Usuaris= 2.58 TB
- Propietats = 225 TB

IOPS= 7930 IOPS

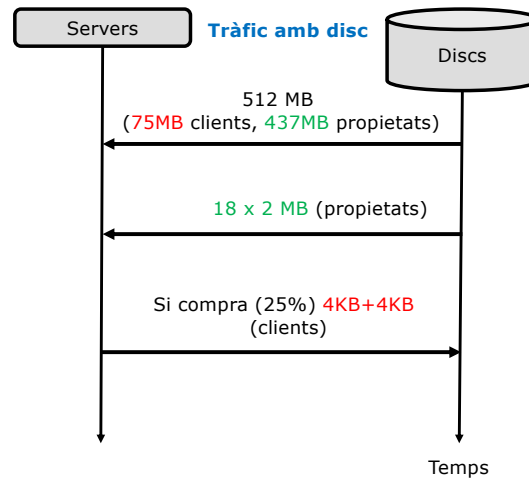
- Si fem una cabina i volem alta seguretat és molta pressió (perquè m'interessa Enterprise i, per exemple RAID51)
  - Per sobre de 50000€ només els discos
- Però tenen diferents necessitats: Puc fer, per exemple, dues cabines, una per clients (RAID 51) i altra per propietats (RAID5)
  - Cal recalculer tràfic usuaris i propietats

NO HI HA UNA ÚNICA SOLUCIÓ (simplement solucions raonades)

ESCENARI EXEMPLE

12

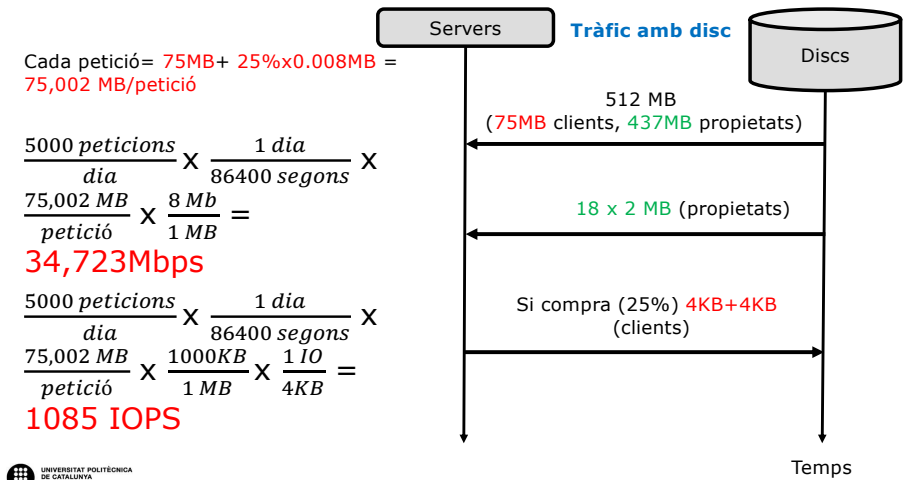
## Tràfic amb el disc



ESCENARI EXEMPLE

13

## Tràfic amb el servidor de clients



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

$$\frac{5000 \text{ peticions}}{\text{dia}} \times \frac{1 \text{ dia}}{86400 \text{ segons}} \times \frac{75,002 \text{ MB}}{\text{petició}} \times \frac{8 \text{ Mb}}{1 \text{ MB}} =$$

34,723Mbps

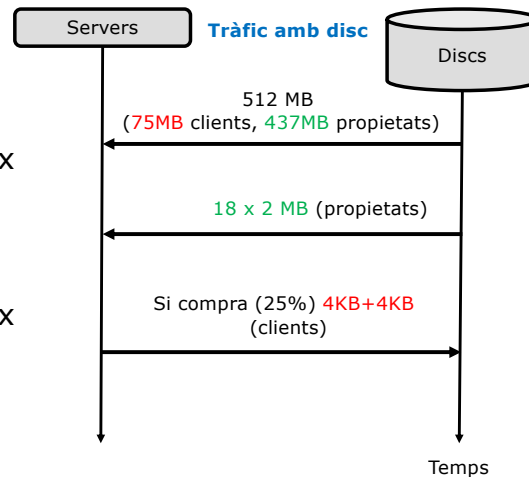
$$\frac{5000 \text{ peticions}}{\text{dia}} \times \frac{1 \text{ dia}}{86400 \text{ segons}} \times \frac{75,002 \text{ MB}}{\text{petició}} \times \frac{1000 \text{ KB}}{1 \text{ MB}} \times \frac{1 \text{ IO}}{4 \text{ KB}} =$$

1085 IOPS

ESCENARI EXEMPLE

14

## Tràfic amb el disc



Cada petició= 437MB+18x2MB=473MB/petició

$$\frac{5000 \text{ peticions}}{\text{dia}} \times \frac{1 \text{ dia}}{86400 \text{ segons}} \times \frac{473 \text{ MB}}{\text{petició}} \times \frac{8 \text{ Mb}}{1 \text{ MB}} =$$

218.981Mbps

$$\frac{5000 \text{ peticions}}{\text{dia}} \times \frac{1 \text{ dia}}{86400 \text{ segons}} \times \frac{473 \text{ MB}}{\text{petició}} \times \frac{1000 \text{ KB}}{1 \text{ MB}} \times \frac{1 \text{ IO}}{4 \text{ KB}} =$$

6843 IOPS

ESCENARI EXEMPLE

15

## Cabina clients

### Discosrequerits.xlsx

- 2.58 TB
- 1085 IOPS
- 0.000027 escriptures

Mirem les opcions (millor disc Enterprise)

L'opció 8 permet RAID 51 per uns 2232€ (6 discos)

- És un disc SSD, no calen discos suport SSD a la cabina
- Posarem Spare Disc (augmenta fiabilitat): dos discos més
- Però és suficient?
  - 6 discos RAID 51 IOPS= 540Kx6 R / 205Kx6 W = 3,24 M/ 1,23 M
    - Aproximadament 2986 vegades superior
  - 6 discos RAID 51 Capacitat = reals dos de dades 1920GBx2 = 3,84 TB
    - 1,48 vegades superior
  - Pel preu (372€/disc) podem posar 8 discos (3 reals = 5.76TB)

ESCENARI EXEMPLE

16

#### Cabina clients

##### Cabina 2 (té 24 badies, sense suport SSD)

- 1 cabina és suficient
- 8 discos en RAID 51 (només 3 de dades)= capacitat real de 5.76TB
- Total per posar al full de càlcul= 1 cabina tipus 2 + 10 discos opció 8 (8 per configurar RAID51 + 2 *spare*)

##### Creixement

- IOPS: gairebé 4000x (sense problemes)
- Clients: tenim 5.76TB-2.58TB = 3.18TB
  - Cada client 6MB -> caben total 960.000 clients (530.000 clients més)
  - Tinc 430.000. Ocupo un 45% de l'espai (creixement de 123%)

#### Cabina propietats (225 TB)

##### Discos requerits.xlsx

- 225 TB
- 6843 IOPS
- 0% escriptures

Les opcions 4-10 molt cares (més orientades a IOPS)

##### Opció 3 de disc en RAID 5 (capacitat= 10 TB - enterprise)

- Pressió: 10 discos per IOPS, **23** per Capacitat
- Volem créixer i és barat -> al menys dupliquem discos = 46 discos
- S'han de dividir en clústers (per exemple, de 5 discos + 1 de raid -> 50 discos de dades)
- 10 clústers de 6 discos (5 dades 1 RAID) = 60 discos

#### Cabina propietats (225 TB)

- 60 discos

##### Cabina 5 (té 36 badies, amb suport SSD)

- 36 badies, 30 discos per cabina
- Posarem 4 *Spare Disc* per cabina (augmenta fiabilitat)
- 2 cabines
- 60 discos -> IOPS= 710 IOPS/disc x 60 disc = 42600 IOPS
  - requerits= 6843, puc créixer un 623%
- 50 discos de dades -> capacitat= 500 TB
  - requerits = 225TB, puc créixer un 222%
- Total per posar al full de càlcul= 2 cabines tipus 5 + 68 discos opció 3

#### Més decisions

##### Posem una SAN?

- No. El tràfic total (extern + disc) és de 263.189 Mbps = 0,263189 Gbps. Tenim una xarxa d'1Gbps
- Podem créixer gairebé un 380% en necessitats de xarxa sense necessitar més

##### Posem Monitorització?

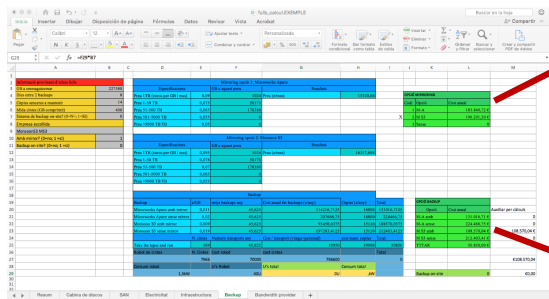
Sol ser sempre interessant, però si posem discos *spare disc* cal monitoritzar per saber quan canviar-los.

##### Posem un mirròr? Quin tipus de backup?

Son decisions lligades. **Influeix en seguretat i temps de recuperació**

## Mirror?

- El mirror redueix el temps de recuperació de dades, com SLA=1.500.000€ és molt interessant



Codi	Opció	Cost anual
1	M-A	183.840,72 €
2	M S3	196.291,30 €
3	Sense	0

Opció	Cost anual
M-A amb	131.016,71 €
M-A sense	224.466,75 €
M S3 amb	108.570,04 €
M S3 sense	212.403,41 €
TTIAR	50.850,00 €

## Backup

- Com tenim mirror només l'opció 1 o 2
- Detalls:
  - Poques modificacions i tenim mirror
  - Plantejarem un backup cada 8 dies.
  - Guardarem 14 còpies (les dues últimes i una per mes l'últim any)
  - Bastant extrem aquest entorn sense gaires modificacions.

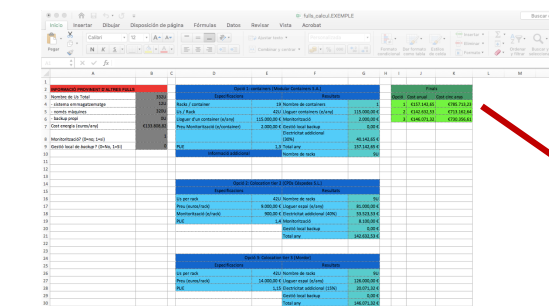
Codi	Opció	Cost anual
1	M-A	183.840,72 €
2	M S3	196.291,30 €
3	Sense	0

Opció	Cost anual
M-A amb	131.016,71 €
M-A sense	224.466,75 €
M S3 amb	108.570,04 €
M S3 sense	212.403,41 €
TTIAR	50.850,00 €

## Empresa de housing

Mireu les característiques de cada sistema!

- En aquest cas Mordor és l'únic que ens ofereix 2a connexió elèctrica (gratis) i segona connexió a xarxa (pagant)
- La diferència de preu no és tan gran



Opció	Cost anual	Cost cinc anys
1	€157.142,65	€785.713,23
2	€142.632,53	€713.162,64
3	€146.071,32	€730.356,61

## Compte!

L'escenari exemple té un nombre d'escriptures molt petit

Les transparències d'aquí endavant no són necessàriament de l'escenari exemple, sinó diferents problemes per il·lustrar

#### Tràfic afegit per tenir mirròr

##### Enviar les escriptures al proveïdor té un cost en la connexió externa

- Imaginem que tenim un mirròr
- Cada escriptura al nostre disc s'ha d'enviar al mirròr
- Com es gestiona el mirròr NO és de la nostra incumbència.

Imaginem que:

- Tinc una LAN de **1 Gbps**
- Tinc un tràfic exterior / entre servidors de 150 Mbps
- Tinc un tràfic totes amb el disc de 500 Mbps (300 Mbps R, 200 Mbps W)

Estic ocupant 650 Mbps de 1 Gbps (puc créixer, no considero posar una SAN)

**PERO!**

Si poso un mirròr augmento el tràfic extern en 200 Mbps, passo a ocupar 850 Mbps de 1 Gbps -> hauria de considerar una SAN

#### Quan costa recuperar el 100% de les dades?

##### Cas 1: RAID 0 (barat)

Recuperem del mirròr. Suposem:

- 1 cabina tipus 3 amb 24 discos tipus 3 (8TB) en RAID0
- Cabina ocupada al 100% de discos, però ocupació al 48% de dades
- Tots els discos de dades (RAID 0), capacitat=  $8TB \times 24 = 192TB$ ; ocupat al 48%  
->  $192TB \times 0,48 = 92.16TB$  a recuperar

Els discos tipus 3 poden fer 640 IOPS. És RAID 0 i són 24 discos que poden escriure en paral·lel  $640 IOPS \times 24 \text{ discos} \times 4 \text{ KB/IOPS} = 61440 \text{ KBps}$

- La comunicació augmenta en  $61440 \text{ KBps} \times 8 \text{ Kbps/1KBps} = 419,52 \text{ Mbps}$
- Tinc que recuperar 92.160.000.000 KB i recupero 61440 KBps, necessito 1,500,000 segons (més de 17 dies)

#### Quan costa recuperar el 100% de les dades?

##### Cas 2: RAID 51 (1)

Recuperem del mirròr. Suposem:

- 1 cabina tipus 2 amb 20 discos tipus 10 (3.8TB) en RAID51
- Ocupació al 45% de dades
- 8 dels 20 discos de dades (RAID 51) capacitat=  $8 \text{ discos} \times 3.8 \text{ TB} \times 0,45 = 13.68 \text{ TB}$  a recuperar

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

- RAID 51, cada operació escriptura son en realitat 4R+4W, podem fer 82K /4 escriptures reals (W) per disc  $\times 20 \text{ discos} = 410.000 \text{ W/s} \times 4 \text{ KB/W} = 1.640.000 \text{ KBps} = 1640 \text{ MBps}$
- La comunicació augmenta en  $1640 \text{ MBps} \times 8 \text{ Mbps/1MBps} = 13.120 \text{ Mbps}$
- Tinc que recuperar 13.680.000.000 KB i recupero 1.640.000 KBps, necessito 8342 segons (2h19') -> sempre i quan tingui una bona xarxa

#### Quan costa recuperar el 100% de les dades?

##### Cas 2: RAID 51 (2)

Suposem que tinc una xarxa de 1 Gbps

- La comunicació ha augmentat en **13.120 Mbps**
- Tinc que recuperar 13.680.000.000 KB i recupero COM A MOLT 1 Gbps  
 $1 \text{ Gbps} / 8 \text{ Gbps} = 0,125 \text{ Gbps} = 125.000 \text{ KBps}$ , necessito 109440 segons (**una mica més de 30h**)

Recomanacions al client: potser ampliar la xarxa LAN (encara que no formi part del que podem escollir a l'escenari; l'important és detectar febleses)

Es resol de manera anàloga si es perd un 1% de les dades

Cau la xarxa (1)

- Segons les dades que tenim (apèndix 6), una línia cau 1 hora cada 18 mesos, entre 1 i 3 hores (o sigui,  $2 \pm 1h$ ) cada 3 anys (36 mesos) i entre 3 i 9 hores ( $6 \pm 3h$ ) 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  $365,25 / 12$  dies de 24 hores (compto l'any de traspàs) = 730,5 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  $20 \pm 7$  hores de cada 3.786.912. Per dues línies, la probabilitat d'estar downtime està entre 0,00034% i 0,00071%

Cau la xarxa (2)

- Els números anteriors són generals i serveixen per qualsevol escenari, l'important són els números a continuació**
- Una línia: probabilitat downtime entre 0,017% i 0,036%
- Dues línies: probabilitat downtime entre 0,00034% i 0,00071%

Si SLA = 1,5M€ per hora downtime

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

Cau la línia elèctrica

- Estem hostatjats a MODOR? Segona línia elèctrica i SAI, garanteix 99.999% de uptime (menys de 30' en 5 anys)

Falla un disc

Tornem a l'escenari exemple.

**Cabina de clients:**

- Estem en RAID 51. Si falla un disc puc copiar-lo del mirròr, no cal reconstruir. Si es pot predir la fallada per SMART (70%) es farà la còpia quan el clúster estigui inactiu.
- Tenim 8 discos (3 de dades i 5 per RAID). Disc SSD enterprise TLC(probabilitat de fallada 0,45% anual - veure apèndix 6).
- De les fallades, 30% requereixen reconstrucció, la resta SMART.
- $8 \text{ discos} \times 0,45\% \text{ fallada} \times 30\% \text{ fallades amb reconstrucció} = 1,08\%$  de probabilitat de fallar/ any, 5,4% probabilitat 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.

Falla un disc

Tornem a l'escenari exemple.

**Cabina de propietats:**

- El disc falla 2,84% anual (HDD Enterprise  $\geq 10000 \text{ rpm}$ ); 60 discos (30 per cabina més els spare). Fallen  $60 \times 0,0284 = 1,704$  discos per any x 5 anys = 8.52 discos.
- El 70% fallades es poden predir per SMART, s'han de reconstruir el 30%  $\Rightarrow 8.52 \times 0,3 = 2.556$  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 x 9,5 clústers en funcionament = 40470 IOPS, l'escenari demana 6843 IOPS, sense problemes