

Activitat EBH

Emmagatzematge, *backup* i *housing*

**Estudiant 1 Rios López, Adrián
Estudiant 2 Vidal Sulé, Guillermo**

Escenari 18

Data: 28/09/2023

1.-Descripció bàsica

Mirar archivo adjunto en la entrega del racó 18.VT1.Rios.Vidal.xlsx

AQUEST APARTAT NOMÉS PEL LLIURAMENT FINAL

IMPORTANT: Pels apartats següents no poseu només el número, justifiqueu el perquè. És més important el perquè que el número en si.

2.-Anàlisi de necessitats

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

30.000 clientes que de cada uno se guardan 4 GB, 1.000.000 de vídeos muy visitados que siempre se guardan cada uno ocupando 5 MB. Durante 10 días se han de guardar todos los vídeos que se suban, contando que solo el 10% de sesiones sube un vídeo y que en un día hay 1.000.000 de sesiones. Esto indica que se han de guardar 100.000 vídeos cada día durante 10 días multiplicamos y nos quedan 1.000.000 de vídeos que se han de almacenar concurrentemente en el peor caso, cada uno ocupa 5 MB.

GB a almacenar = $(30.000 \cdot 4 \text{ GB}) + (1.000.000 \cdot 5 \text{ MB}) + (1.000.000 \cdot 5 \text{ MB}) = 130.000 \text{ GB}$.

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

Cada día se llevan a cabo un total de 1.000.000 sesiones, de las que un 10% son *uploads*. Además, por cada sesión se harán 10 peticiones por defecto de 128 KB cada una. Por último, en el caso de *visualize* se llevan a cabo unas 5 peticiones de media¹. Por tanto, calculamos la presión con la siguiente fórmula:

$$\text{Presión} = \frac{1000000 \text{ peticiones}}{1 \text{ día}} \cdot \frac{1 \text{ día}}{86400 \text{ segundos}} \cdot \frac{24280 \text{ KB}}{1 \text{ petición}} = 281018.52 \text{ KBps}$$

Sabiendo que los discos acceden en bloques de 4 KB:

$$\text{IOPS} = \frac{281018.52 \text{ KB/s}}{4 \text{ KB/operación IO}} = 70254.63 \text{ IOPS}$$

De los cuales un ~2% son escrituras.

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

¹ En caso de ser necesario justificar el cálculo de KB por petición, se ha calculado mediante $128 \text{ KB} \cdot 10 + 5 \text{ MB} \cdot 0.9 + 5 \text{ MB} \cdot 5 \cdot 0.1 = 24280 \text{ KB}$

Por cada sesión el cliente envía 4 KB al servidor y la MV calcula en función de las recomendaciones de los datos obtenidos en el disco y envía 512 KB de datos. Luego, en el caso de upload (100.000 sesiones diarias) el cliente envía un vídeo de 20 MB. En el caso de view (900.000 sesiones diarias) tendríamos que serían 4 KB de petición por cada uno de los 5 vídeos. También necesitamos enviarle el propio vídeo de 20 MB, 5 vídeos de media por sesión.

$$\text{Mbps} = (900.000 * (20 * 5 \text{ MB} + 0,004 * 5 \text{ MB} + 0,512 \text{ MB} + 0,004 \text{ MB})) + (100.000 * (20 \text{ MB} + 0,512 \text{ MB} + 0,004 \text{ MB})) * 8 \text{ bits} / 1 \text{ Byte} * 1 \text{ día} / 86.400 \text{ s} = 8.567,963 \text{ Mbps}$$

2.4- Tràfic amb el disc:

Por cada sesión enviamos 10 posibles recomendaciones al server de 128 KB, luego si se decide upload son 5 MB que se transfieren a discos y si es view son 5 MB de cada vídeo 5 vídeos de media de discos a server.

$$\text{Mbps} = (900.000 * (5 * 5 \text{ MB} + 10 * 0.128 \text{ MB}) + 100.000 * (5 \text{ MB} + 0.128 * 10 \text{ MB})) * 8 \text{ bits} / 1 \text{ byte} * 1 \text{ día} / 86.400 \text{ s} = 2.248,15 \text{ Mbps.}$$

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

La presión sobre la red es la suma del tráfico de discos + tráfico de clientes/servidores.

Presión red = 8.567,963 Mbps + 2.248,15 Mbps = 10.816,113 Mbps = 10.816113 Gbps. Es decir que nos falta 1 Gbps.

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

Hemos decidido poner 3 cabinas de disco, 2 para datos de usuario, que es lo que más ocupa en tamaño, pero lo que menos IOPS "consume" gastando únicamente unas $\text{IOPS}_{\text{Usuario}} = (10 * 128 \text{ KB} * 1.000.000) * 1 \text{ día} / 86.400 \text{ s} * 1 / 4 \text{ KB} = 3704 \text{ IOPS}$ (0% Escritura). Y otra cabina para los vídeos de los dos tipos.

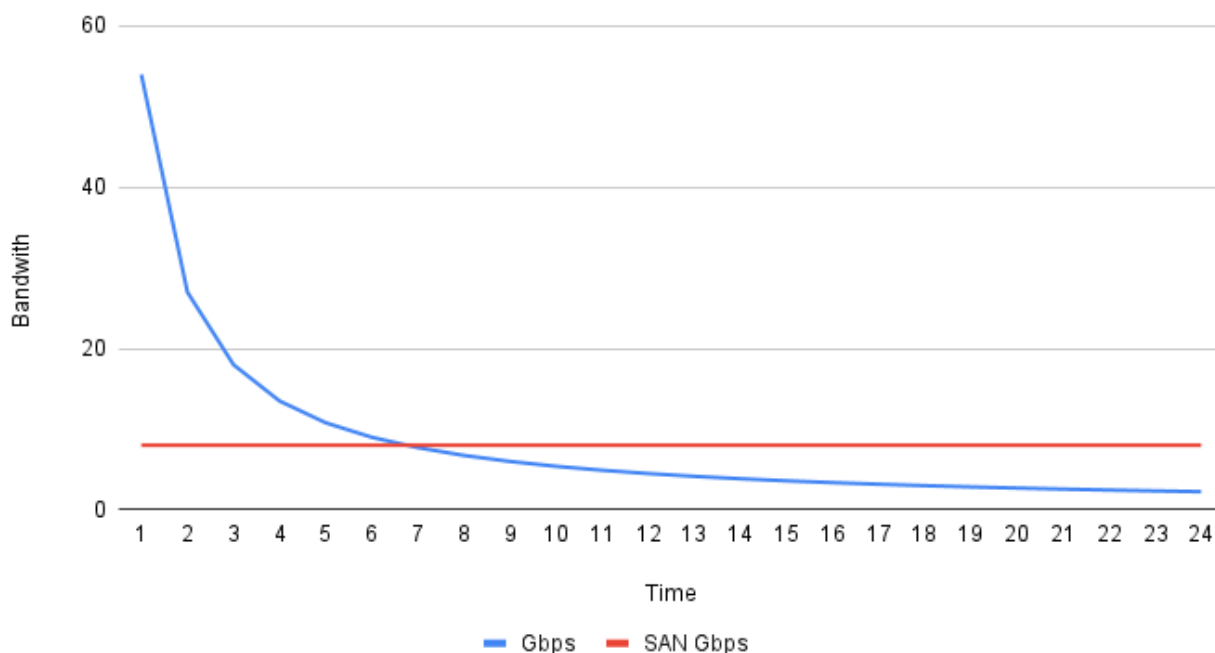
Cabina datos usuarios: Esta cabina ha de ser capaz de almacenar un total de al menos 120 TB y soportar una presión de 3704 IOPS. Además, al ser datos tan importantes haremos RAID 10 para poder tener un mirror de los datos en caso de que un disco fallase tener la copia del disco entera, también añadiremos dos hot spare disks para poder empezar a restaurarlos en cuanto SMART detecte que un disco va a fallar. Al introducir estos datos en las hojas de cálculo nos da que la opción 9, necesitaría 32 discos por su pequeña capacidad de almacenamiento comparado a los hdd, respecto a las cabinas la mejor que podemos comprar es la 4, por su alto número de bahías que nos permite tener 36 discos y como no necesitamos SSD caching pues no elegimos la 5. Esto nos permite tener hasta un total de 17 discos para almacenar datos (17 de mirror + 2 hot

spare disk) por cabina, lo que nos da un almacenamiento total de 65.280 usuarios, es decir podríamos incrementar el número de usuarios en un 171%

Cabina datos video: En esta cabina hemos de almacenar un total de 10 TB, pero hemos de soportar un total de 66.551 IOPS (2% escritura), por lo que es bastante, entonces usaremos discos SSD, hemos escogido la opción de disco número 9 por su gran capacidad, baja tasa de fallos y rapidez. Respecto a la RAID para no perder datos utilizaremos una RAID 10, y 2 hot spare disk. Hemos escogido la cabina 4 por las mismas razones que antes y que es capaz de cubrir la demanda de electricidad. Es decir, 16 discos útiles que son un total de 24.576.000 videos, ahora solo almacenamos 2 millones, con lo que supone un crecimiento de 1.128,8 %.

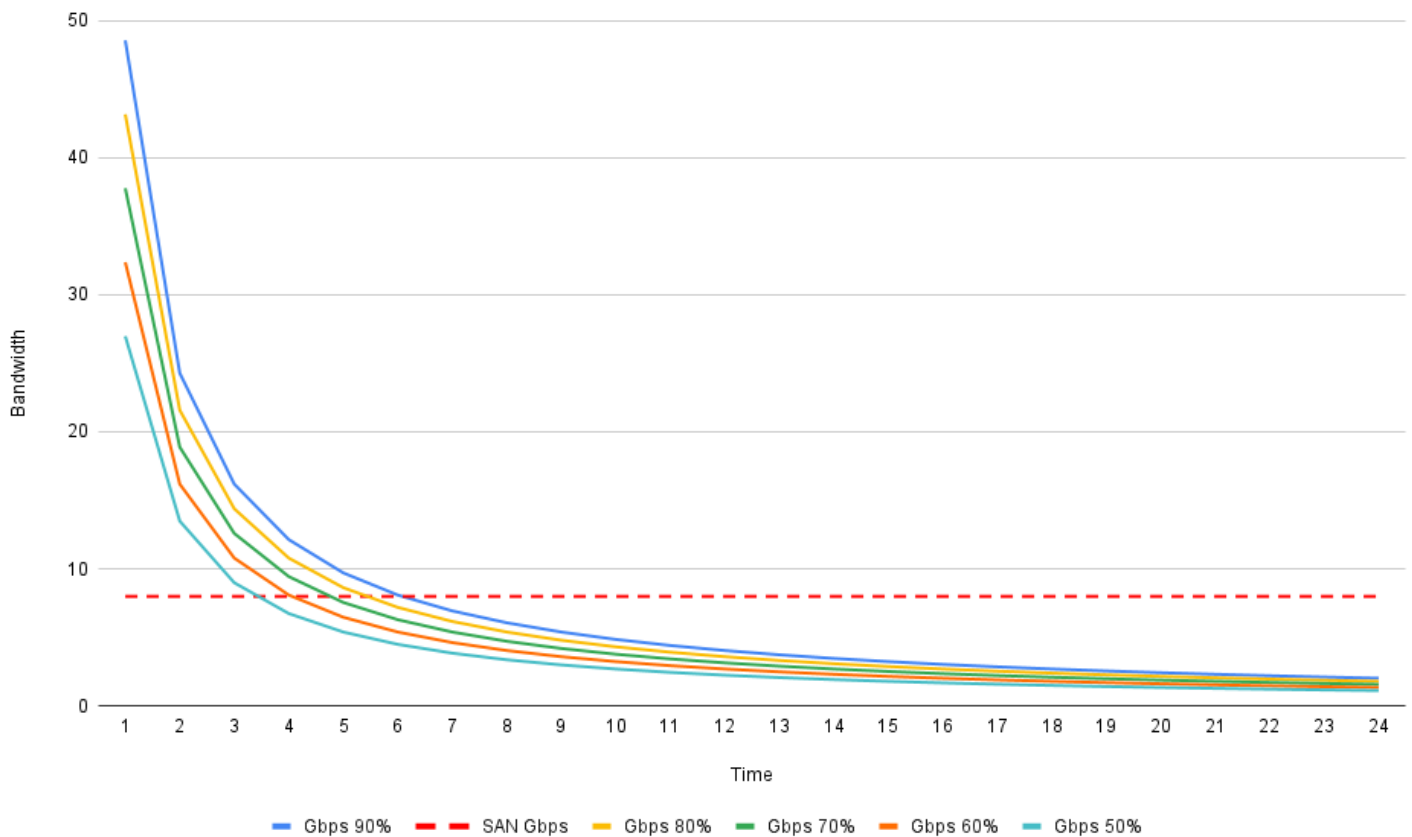
3.2- Es justificada la necessitat d'un SAN? Si la respuesta es si, raonar si el cost és assumible o no, i cas de no ser-ho calcular l'impacte sobre el rendiment del CPD

Como nuestra LAN es de 10 Gbps y ya tenemos una presión sobre la red de más de eso necesitamos un SAN, suponiendo que duplicásemos el tráfico a discos tendríamos sobre unos 5 Gbps, es decir con la opción 1 nos sobra que son 8 Gbps (para el tráfico actual) . El coste es muy asumible con nuestro elevado presupuesto y gestión inteligente de los discos y cabinas. Ahora nos preguntamos ¿Cuándo el SAN llegará a ser el cuello de botella si hay un pico en el tráfico? Bien, si miramos el siguiente gráfico, hemos condensado el 100% de los usuarios en un periodo de tiempo determinado por el eje X, es decir, en vez de asumir que todos los usuarios se distribuyen de manera igualitaria en 24 horas, asumimos que todos deciden entrar en ese período de tiempo.



En este caso vemos que el cuello de botella empieza a partir de tener todos los usuarios metidos en 7 horas, lo que es bastante razonable, por lo tanto, dándonos un crecimiento del 242.86%.

Luego veremos si esto se puede soportar por parte de los discos en sí. También en el siguiente gráfico vamos a ver lo mismo pero con el 90% del tráfico de usuarios, 80%...



3.3.- Posem un *mirror*?

Sí, pondremos un sistema de mirror y backup externo contratado al proveedor Monsoon S3, porque nos reduce el recovery point objective a la vez que no tenemos que gestionar ningún tipo de backup de cintas.

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

En lo referente al housing realmente no necesitamos ningún tipo de seguridad extra para los datos que nosotros guardamos, no hay SLA, por lo tanto, podemos ahorrar sin necesidad de tener las líneas de internet duplicadas, con esta información podemos descartar MORDOR. Ahora solo queda decidir entre 1 y 2. Realmente no necesitamos un sistema de seguridad alto y ambos ofrecen monitorización y cambio de spare disk, entonces vamos a escoger el más barato en este caso es la opción 2.

En los siguientes apartados haremos un análisis de los posibles riesgos.

3.5- Posem monitorització?

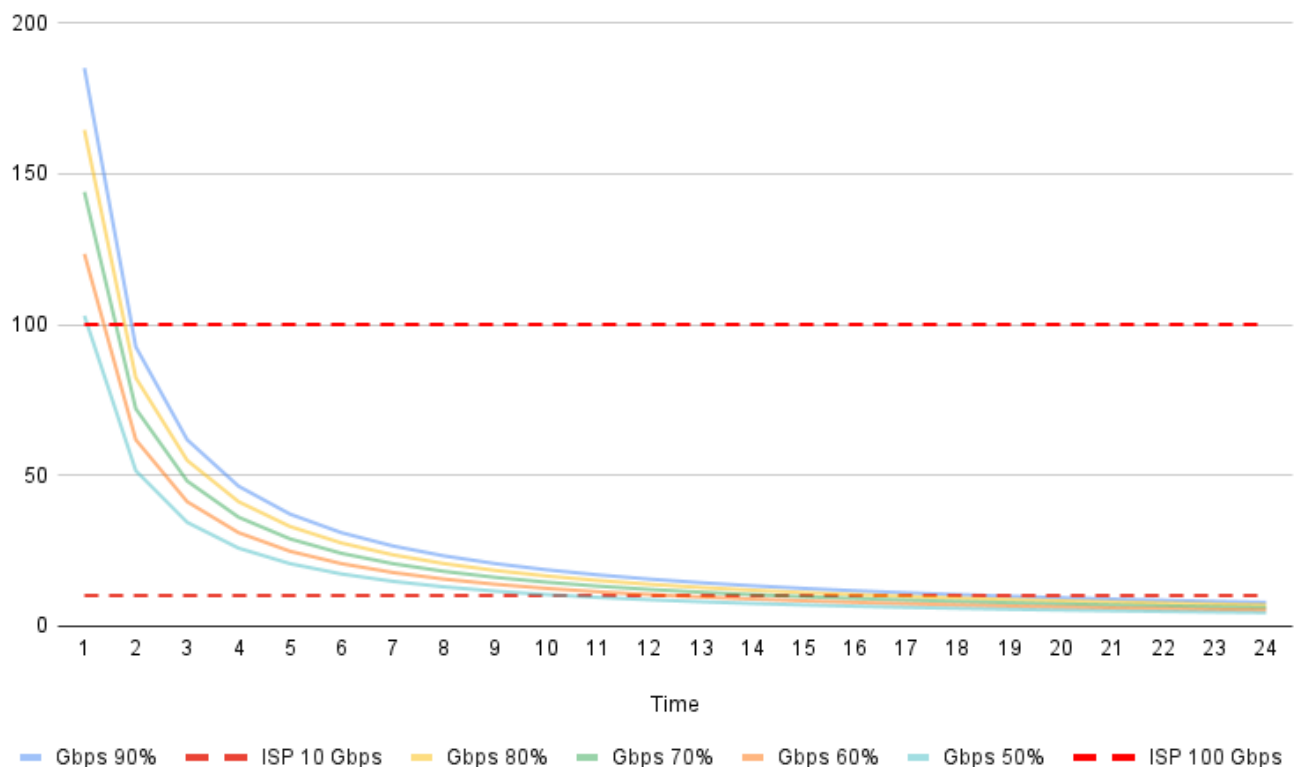
Sí, necesitamos monitorización para aprovechar los spare disks y no perder datos, además que nuestro presupuesto nos lo permite.

3.6- Opció de backup?

Para la cantidad de datos que nosotros necesitamos almacenar nos sale más rentable tener un backup offsite de 130 TB en la misma empresa con la que hemos contratado el mirror. Haremos un backup cada 10 días, lo que da aproximadamente unos 36 backups a almacenar siempre.

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

Para contar el tráfico con el exterior simplemente vemos que necesitaremos una velocidad adicional, contando que las escrituras a disco también se han de realizar al mirror, $5 \text{ MB} * 100.000 * 1 \text{ dia} / 86400 \text{ s} = 5.79 \text{ Mbps}$ que no tienen prácticamente impacto sobre nuestro SAN y tampoco sobre la conexión al exterior. Aunque si bien lo podemos soportar ya de por sí nuestra conexión con el exterior va un poco al límite, y en caso de que hubiese crecimiento de sesiones o picos o se concentrase el X % en Y horas, probablemente no podríamos aguantar solo con 10 Gbps porque consumimos 8.567,963 Mbps. Como podemos ver en este gráfico, la diferencia del cuello de botella es abrumadora.



Como vemos en las 24 horas apenas nos llega para el 90% de usuarios. En cambio, nos podríamos permitir un pico del 90% accediendo en 2.5 horas con la de 100 Gbps, y además nos lo podemos permitir si miramos el presupuesto, por eso pondremos 100 Gbps.

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?**

Tenemos un mirror externo, así que podemos recuperarlo completamente desde la última escritura, si también está corrupto en el mirror tenemos un backup con 36 versiones realizadas cada 10 días.

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

Si se rompe un disco hay un 70% de probabilidades de que antes de que se rompa se detecte y se copie a un spare disk. De todas maneras tenemos un RAID 10 en ambas cabinas con lo que no se pierde ningún dato si solo se rompe un disco y no hace falta reconstruirlo porque simplemente se puede copiar del mirror, además tenemos el mirror externo por si fallasen justamente los 2 que hacen mirror.

En caso de que fallasen los de la cabina de datos de usuario o vídeos se tardaría en el peor caso (si está completamente lleno) aproximadamente 8 horas en copiarlo por completo, pero no tendríamos que parar el servicio porque los IOPS son 467k de lectura y 65k de escritura, es decir quedarían unos 462.000 IOPS de lectura y los datos de usuario solo necesitan 3800 IOPS.

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

Si fallan dos discos y coincide que uno es el mirror del otro se tendría que traer del mirror externo, lo que tendría una latencia agregada y se tardaría 8 horas en copiar los datos, aunque podrían comenzar inmediatamente después de que fallen, y como veremos luego es muy poco probable que esto ocurra. El problema que tendremos si esto ocurre es que no se podrán enviar recomendaciones de usuario o en el caso de vídeos ese vídeo en específico no se podrá servir hasta que no se haya reconstruido.

La cantidad de discos que fallarán al año se obtiene multiplicando la probabilidad de fallo por el número de discos: $108 * 0,0045 = 0,486$ discos al año, es decir, podríamos simplemente copiar del mirror (remoto) al spare disk y en menos de 10 horas lo tendríamos y podríamos seguir operando con el mirror de RAID10 (esto tendría un impacto en la seguridad del sistema si tuviésemos una mayor probabilidad de fallo, ya que podrían fallar más de un disco).

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

Si nos quedamos sin electricidad nuestro colocation center ofrece un sistema alternativo basado en un generador diésel, este nos garantiza que solamente tendremos como mucho 22 horas de down time al año, con mucha mala suerte no funcionará durante algo menos de un día al año. No tenemos ningún SLA así que realmente no es un problema grave y en caso de que nos preocupe mucho se podría solucionar con un colocation tier 3.

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

No podremos servir peticiones durante el tiempo que esté caída.

- **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 perdemos el 1% (1.3 TB) de los datos podemos consultar al mirror externo, cuando llegan a la red se encuentran con un SAN de 8 Gbps que tardaría unos 36 minutos en transportar los datos al centro y luego teniendo en cuenta los IOPS se tardaría 1.35 horas en copiarlo, es decir casi 2 horas en poder volver a un servicio normal.

Si perdemos la totalidad de los datos seguimos teniendo el mirror externo se tardaría una hora en transmitir los datos por el cuello de botella del SAN y luego 8 horas en restaurar cada disco (se puede hacer en paralelo).

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:

Lo que se gane por las visitas de clientes durante ese tiempo. Contratar a Mordor como colocation y una segunda línea. Invertir 32.287,34 € euros más al cabo de 5 años.

Fallada de disc

Lo que se gane por las visitas de clientes durante ese tiempo. Realmente no se puede prevenir, ya hemos hecho todo lo posible.

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, ...)

Podemos tener un crecimiento de usuarios hasta un 171%, lo que implicaría tener hasta 65.280 usuarios en total, en referente a vídeos podemos almacenar hasta 24.576.000, lo que sería un crecimiento de 1.128,8 %. Respecto a la red tenemos la conexión externa de 100 Gbps y la SAN de 8 Gbps, como hemos podido ver en los gráficos anteriores podemos permitirnos un gran crecimiento, concretamente la conexión externa hasta un 11,5 veces más de tráfico, y respecto a la SAN serían hasta 3,5 veces más de tráfico.

Es decir que nuestro factor limitante es el crecimiento de usuarios.

Un crecimiento de un 20% en todo no implicaría gran cosa, como hemos demostrado el cuello de botella es el número de usuarios, y este nos permite crecer hasta un 171% que es mucho más que el 20%. La única implicación relevante sería la tarifa del backup y mirror, la cantidad de datos aumenta hasta los 156 TB, si lo ponemos en la hoja de cálculo esto nos implicaría gastar 30.381 € anuales, si hacemos el cálculo de OPEX a 5 años implicaría gastar 151.905 € en total.

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...

Hemos diseñado un sistema acorde a las especificaciones propuestas por la parte del cliente y además hemos garantizado el soporte de un gran crecimiento de los datos que se pueden almacenar y de una manera segura. Hemos sido capaces de hacerlo únicamente usando menos de la mitad del presupuesto. Si quisiéramos tener un CPD aún más robusto, nuestras recomendaciones serían las siguientes, aunque repito, no son necesarias ahora mismo:

Cambiar de empresa de housing al colocation Tier 3, esto nos asegurará menos horas de caída por electricidad y al poner otra línea de internet con otro ISP obtendríamos más redundancia.

Aumentar el bandwidth del SAN para poder gestionar más peticiones diarias aunque ahora podríamos gestionar hasta 3.55x de las sesiones que tenemos diariamente, pero a futuro se podría invertir más conforme se necesiten más discos para almacenar más datos y la cantidad de discos que fallen por probabilidad aumente y así tardar menos en transportar los datos del mirror offsite.