

CPD: Activitat 2

Curs 2018-2019, tardor.

Alumne: Kevin Callado Lara | Pablo Aznar Puyalto

Data: 10/09/2018

Número d'activitat: 11/6

Descripció de l'activitat:

Storage: NAS/SAN

Workload: Web 2.0

Budget: 12M€

Descripción general de la solución propuesta	3
Esquema físico de configuración	5
RACKs	5
Servidores	5
Top of Rack	7
Cableado de racks	7
Agregación	8
Switch de agregación	8
Cableado de agregación	8
Arquitectura del sistema de almacenamiento	8
Racks	9
Cableado	11
Redundancia	11
Complejidad de Gestión HW/SW:	12
Capacidad	13
Aspectos genéricos	13
Red	14
Bi-section Bandwidth	14
Oversubscription ratio	14
Almacenamiento	14
Ancho de banda máximo	14
Capacidad de almacenamiento	14
Cuellos de botella entre nodos	14
Datos específicos de los workloads	15
Consumo	16
Costes	19
Coste de la infraestructura	19

Coste por byte	22
Coste por Mhz	22
Coste eléctrico mensual de los equipos de computación y comunicación	22
Escalabilidad	23
Anexo	24
Puntos a considerar	24
Referencias	25

Descripción general de la solución propuesta

Màxim 1 pàgina. En el cas del lliurament final ha de tenir una descripció (pot ser visual) de la solució proposada. En el cas de lliurament setmanal, la descripció serà més aviat de les decisions preses fins el moment, i de coses pendents, així com una petita descripció de les modificacions produïdes des del lliurament anterior. Als lliuraments setmanals, alguns dels punts que hi ha a continuació poden estar en blanc, o tenir més d'una opció disponible en aquell moment. Que us serveixi també de lloc on escriure reflexions, feina a fer, ...

Tenemos que diseñar un CPD para procesar y servir imágenes a nuestros clientes (web 2.0) por lo tanto nos centraremos en diseñar un CPD buscando el máximo rendimiento respecto a peticiones por segundo.

Es muy importante la disponibilidad de los datos que suben los usuarios y la tolerancia a los fallos ya que si cae un nodo el sistema tiene que seguir funcionando para que los usuarios sigan pudiendo acceder a sus datos.

El diseño del CPD busca la máxima eficiencia respecto peticiones por segundo así como la tolerancia a fallos. Hemos diseñado una red de *management* para poder acceder al CPD de forma remota.

Hemos intentado consultar todos los precios en la misma página para que todos los precios estén más unificados porque pueden variar mucho. Así mismo todos los productos de servidores han estado configurados a través de *thinkmate.com* y los switches a través del configurador de *CISCO* por lo que sabemos que son válidos.

Procedemos a describir qué reglas hemos seguido para diseñar el CPD y que nos han ayudado después a escalarlo y a diseñar y decidir los componentes necesarios para su funcionamiento.

Por cada 800 MHz de un procesador somos capaces de servir una petición con 100 ms de latencia, eso nos permite servir 10 peticiones por segundo por cada 800 MHz de un core.

Por lo tanto con un procesador estándar de 4 cores y 3.2 GHz podemos servir:

$$10 \text{ peticiones/s} * 4 \text{ cores} * \frac{3.2 \text{ GHz}}{800 \text{ Mhz}} = 640 \text{ peticiones/s por cada procesador.}$$

Además por cada core del procesador necesitamos 4 GB de RAM.

Por otra parte recibimos 600 bytes por petición y tenemos que enviar 180 KB de respuesta por lo tanto debemos asegurarnos de tener un buen ancho de banda para poder enviar nuestras peticiones. En total consideramos que una petición ocupa 180 KB de ancho de banda ya que los 600 bytes son negligibles.

Asimismo cada petición accede 5 veces a disco consultando 1 KB por acceso por ende, cada petición accede a 5 KB de disco.

En resumen:

- El número de peticiones que podemos servir por procesador es:
 $10 \text{ peticiones/s} * n^{\circ} \text{cores} * \frac{\text{frecuencia}}{800 \text{ MHz}} = N^{\circ} \text{ peticiones/s por procesador.}$
- El número de datos que recibiremos en total será de: $n^{\circ} \text{peticiones} * 600 \text{ bytes.}$
- La cantidad de información que generaremos en total será de: $n^{\circ} \text{peticiones} * 180 \text{ KB.}$
- En general necesitaremos un acceso a los datos de $5 \text{ KB/s} * n^{\circ} \text{peticiones} = x \text{ KB/s}$

Esquema físico de configuración

En aquest apartat caldrà aportar informació sobre la distribució de les màquines en el CPD. Es tracta de fer l'esquema físic de la distribució en l'espai. Es pot realitzar usant eines pròpies o software especialitzat 1 . S'ha d'incloure un diagrama i una especificació de, com a mínim, els següents elements:

- *RACKs: quants i com s'organitzaran*
- *Cablejat per rack i entre racks: incloent capacitat de cada enllaç, ús o no de Link*
- *Aggregation, tipus de tecnologia de xarxa i velocitat, oversubscription rates.*
- *Arquitectura del sistema d'emmagatzematge: Com s'emmagatzemaran les dades i com s'accediran a través de la xarxa (en els casos que apliqui)*
- *Redundància: Quins sistemes estaran redundats? Indicar-ho en l'esquema.*


En este apartado explicaremos primero cada elemento del CPD y luego comentaremos como los unimos todos para darle la forma final.

RACKs

Nuestros racks de servidores tendrán 40 servidores y 2 switches que harán de *Top of Rack* para mantener la redundancia.

Servidores

Nuestro servidor es RAX XS4-21S1-10G y se encargará de procesar las peticiones, dejamos una tabla con sus especificaciones y componentes:

RAX XS4-21S1-10G Configured Price: 12.709,93 €	
Selection Summary	
Processor	2x Intel® Xeon® Gold 6148 Processor 20 core 2.40GHz 27.50MB Cache (150W)
Motherboard	Intel® C624 Chipset - 14x SATA3 + 4x U.2 - 1x M.2 - Dual Intel® 10-Gigabit Ethernet (RJ45) - IPMI 2.0 with LAN
Memory	12 x 16GB PC4-21300 2666MHz DDR4 ECC Registered DIMM
Chassis	Thinkmate® RAX-1304 1U Chassis - 4x Hot-Swap 3.5" SATA/SAS3 - 750W Redundant Power
Boot Drive	512GB Samsung 970 PRO M.2 PCIe 3.0 x4 NVMe Solid State Drive
Controller Card	LSI 3108 SAS 3.0 12Gb/s 8-port Hardware RAID Controller with 2GB Cache
Battery Backup	CacheVault Flash Module Protection for LSI 3108 SAS Controller
Network Card	Intel® 10-Gigabit Ethernet Converged Network Adapter X710-DA2 (2x SFP+)
Riser Cards	Thinkmate® 1U Riser Card Left Slot 1x PCIe 3.0 x16 Thinkmate® 1U Riser Card - Right Side WIO - 1x PCIe 3.0 x8
Cables	IEC60320 C13 to C14 Power Cable, 16AWG, 240V/15A, Black - 4'
Operating System	No Operating System
Warranty	5 Year Advanced Parts Replacement Warranty

Top of Rack

Nuestro switch de Top of Rack es el switch de CISCO N3K-C36180YC-R. Este switch dispone de 48 puertos *SFP* donde se conectarán todos los servidores y de 6 puertos *QSFP28* que utilizaremos para conectar los racks con el nivel de agregación.

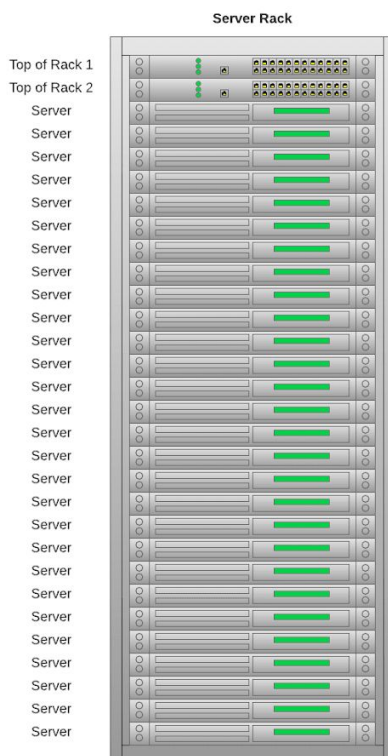
En total disponemos de 8 racks de servidores por lo que tenemos 320 servidores y 16 switches *top of rack*.

Con cada uno de los servidores podemos soportar 1.200 peticiones¹ por segundo. Por lo que en total podemos soportar 384.000 peticiones por segundo.

Cableado de racks

Como cada servidor puede procesar 1.200 peticiones por segundo, en total puede generar 1.6 Gbps de datos². Por eso cada servidor usará un link *SFP* para conectarse con el switch de agregación. Usaremos dos puertos por servidor dado que se conectarán con los dos *ToR* de cada rack. Para conectarlos usaremos un link de 10 Gbps por lo que si redondeamos hacia arriba usaremos 2 Gb para datos y nos quedan 8 Gb de ancho de banda libres.

Los racks se conectan y comunican entre ellos a través del nivel de agregación así que no hay cableado entre ellos.



¹ Siguiendo la fórmula que encontramos en el primer apartado $10 \text{ peticiones/s} * 20 \text{ cores} * 2.4\text{GHz}/800 \text{ MHz} = 1.200 \text{ peticiones}$

² $1.200 \text{ peticiones} * 180\text{KB/peticion} = 1.65 \text{ Gbps}$

Agregación

Switch de agregación

Dispondremos en total de dos switches de agregación del modelo Cisco N9K-X9432C-S, un switch modular que dispone de 32 puertos *QSFP28* por lo que todas las conexiones a agregación se harán mediante links de 100 Gbps.

Como comentamos anteriormente podemos soportar un total de 384.000 peticiones por segundo lo que nos lleva a que en total tenemos que ser capaces de transmitir 553 Gbps³

Cableado de agregación

Como tenemos dos niveles de *ToR*, cada nivel de *ToR* se conectará a un switch de agregación distinto. Por lo que el primer nivel de *ToR* se conectará al switch de agregación 1 y el segundo nivel de *ToR* se conectará al segundo switch de agregación. En total usaremos 8 puertos de cada switch de agregación para hacer estas conexiones.

Para poder ser tolerable a falladas debemos conectar los dos switches de agregación entre ellos, como el máximo de tráfico que vamos a transmitir son 553 Gbps usaremos 6 links *QSFP28* entre los dos switches de agregación.

Por último tenemos que ser capaces de enviar todo el tráfico hacia la red, por lo que además usaremos 6 puertos más de cada switch de agregación para enviar el tráfico a nuestro proveedor de internet.

De estos 6 links que utilizamos para enviar el tráfico, el ancho de banda efectivo es de 553 Gb por lo tanto aún tenemos 47 Gb de ancho de banda disponibles.

Tanto los links que conectan los switches de agregación como los que nos dan acceso a internet són links agregados.

Arquitectura del sistema de almacenamiento

Para el sistema de almacenamiento nos hemos propuesto usar el presupuesto restante que disponíamos después de comprar la parte de computación.

Hemos decidido usar este método ya que haciendo una aproximación con las imágenes que subirán los usuarios y con el número total de usuarios hemos considerado que no sería ningún problema y que nuestro presupuesto restante es lo suficientemente grande para abastecer a todos los usuarios.


El precio dedicado para la parte de almacenamiento son 3 millones de euros con lo que lo utilizaremos para comprar los racks, los servidores, los JBOD, el switch que comunicara la parte de storage con la de computación y el cableado.

³ 384.000 peticiones * 180 KB/peticion = 552.96 Gbps.

Racks

Para los racks nos hemos decantado por utilizar un servidor que gestione 8 JBODs. Cada JBOD ocupa 4 U y el servidor 1 U por lo que cabría todo en un rack con capacidad de 42 U.

Esta es nuestra configuración del JBOD:

JD72-0410-SAS3 Configured Price: 48.964,82 €	
Selection Summary	
Chassis	Supermicro SuperChassis 417BE1C-R1K23JBOD - 4U - 72 x 2.5" SATA - 1200W Redundant
Storage Drive	72 x 1.92TB Intel® SSD D3-S4510 Series 2.5" SATA 6.0Gb/s Solid State Drive
Controller Card	LSI MegaRAID 9380-8e SAS 12Gb/s PCIe 3.0 8-Port Controller with 1GB Cache
Cables	3-Meter External SAS Cable - 12Gb/s to 12Gb/s SAS - SFF-8644 to SFF-8644 IEC60320 C13 to C14 Power Cable, 16AWG, 240V/15A, Black - 4'
Watts	1200W.
Warranty	5 Year Advanced Parts Replacement Warranty

Hemos decidido utilizar discos SSD y no HDD debido a sus propiedades respecto a los discos HDD.

El servidor que utilizaremos para conectar todos los JBOD se encuentra especificado en la siguiente tabla. Este servidor irá conectado al switch de storage que se conectara al de agregación.

RAX XS4-21S1-10G

Configured Price: 5.703,26 €



Selection Summary

Processor	Intel® Xeon® Bronze 3104 Processor 6-core 1.70GHz 8.25MB Cache (85W)
Motherboard	I® C621 Chipset - 8x SATA3 - 1x M.2 - Dual Intel® 1-Gigabit Ethernet (RJ45)
Memory	6 x 8GB PC4-21300 2666MHz DDR4 ECC Registered DIMM
Chassis	Thinkmate® STX-4336 4U Chassis - 36x Hot-Swap 3.5" SATA/SAS3 - 12Gb/s SAS Single Expander - 1280W Redundant Power
Boot Drive	256GB Micron M1100 M.2 SATA 6.0Gb/s Solid State Drive
Controller Card	LSI MegaRAID 9361-8i SAS 12Gb/s PCIe 3.0 8-Port Controller with 1GB Cache
Battery Backup	LSI MegaRAID 9361-8i SAS 12Gb/s PCIe 3.0 8-Port Controller with 1GB Cache
Network Card	Intel® 10/40-Gigabit Ethernet Converged Network Adapter XL710-QDA2 (2x QSFP+)
Cables	IEC60320 C13 to C14 Power Cable, 16AWG, 240V/15A, Black - 4'
Operating System	No Operating System
Warranty	5 Year Advanced Parts Replacement Warranty

El switch que utilizamos para el storage es el siguiente Cisco Nexus 3232C, que tiene 32 puertos de 100 Gbps, utilizamos este switch para simplificar las conexiones con agregación y para una futura escalabilidad.

Cableado

Si todas las peticiones acceden a storage a la vez, como cada petición accede a 5 KB de storage, necesitaremos 1.920.000 KB de ancho de banda, 15.360.000 Kb que son 15,36 Gbps, por lo que le pasaremos como mínimo 15,36 Gbps al switch del storage.

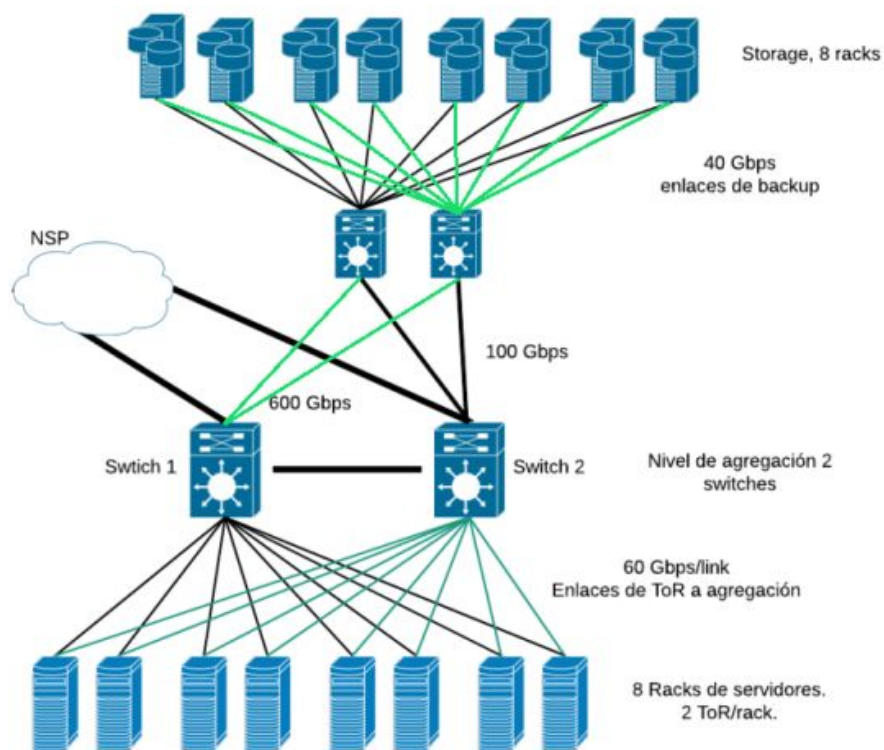
Dado que nuestro switch tiene una capacidad de sobras para este ancho de banda, solo utilizaremos una conexión del servidor al switch, utilizando un cable de 40 Gbps.

Redundancia

Como hemos comentado anteriormente, usamos dos switches de agregación y dos switches de *Top of Rack* por cada rack para ser tolerables a las falladas y poder tener tiempo a reaccionar ante alguna.

De la misma forma pondremos dos switches en storage por si se cae uno poder acceder a los datos utilizando el otro.

Por lo tanto los dos sistemas redundantes son los switches de *Top of Rack*, como hemos comentado y podemos ver en el esquema anterior y los switches del nivel de agregación como podemos observar en el siguiente esquema.



Complejidad de Gestión HW/SW:

Aquest apartat es una valoració qualitativa de la *manageability* del CPD proposat. Inclourà una discussió sobre aspectes com ara la presencia de KVMs⁴ i PDUs⁵ que facilitin la gestió dels recursos hardware.

Para tener una gestión remota de todo el sistema para evitar desplazarnos cuando sea necesario hemos querido diseñar una buena red de *management*.

Con este objetivo hemos usaremos el IPMI de las placas bases de nuestros servidores, el tráfico de control seguirá la política *inBand*, el tráfico de control irá junto el tráfico de datos. De esta forma podemos acceder al IPMI de forma remota y así gestionar los servidores.

Podemos soportar este modelo de red de control porque tenemos ancho de banda de sobras en nuestros servidores, como hemos explicado en los apartados anteriores, siempre tenemos un margen de mínimo de un 10% de ancho de banda disponible aunque trabajemos a máxima capacidad. Dado que el tráfico de control ocupa muy poco ancho de banda y como acabamos de explicar, no usamos toda la capacidad que tenemos disponible, este diseño no nos va a suponer ningún problema.

Por otra parte tenemos el problema de que un servidor se cuelgue y no podamos acceder a las IPMI, por eso hemos instalado enchufes inteligentes (PDU) para no tener que desplazarnos al CPD.

⁴ <http://www.apc.com/products/family/index.cfm?id=319&ISOCountryCode=es>

⁵ <http://www.apc.com/products/family/index.cfm?id=70&ISOCountryCode=es>

Capacidad

Aquesta secció cobreix un extens ventall d'aspectes del CPD, i és la que tractarà la capacitat a l'hora de donar servei als usuaris finals. En concret es tracta d'estimar la capacitat final del CPD a l'hora de córrer el workload que us hagi estat assignat, utilitzant les mètriques que s'hagin suggerit. Caldrà valorar els diferents aspectes que considereu rellevants per al workload en qüestió, així com aspectes generals de capacitat teòrica màxima del CPD. Caldrà cobrir com a mínim:

- Aspectes genèrics
 - Xarxa:
 - Bi-section bandwidth
 - Oversubscription rate
 - Emmagatzematge:
 - Màxim ample de banda amb discos (agregat)
 - Capacitat d'emmagatzematge
 - Colls d'ampolla entre nodes (Disc/CPU/Xarxa) i en quina situació el trobarem (en accés a disc local, en accés a disc remot, en accés a memòria remota, en fase de computació...)
- Dades específiques de cada workload
 - Vegeu la mesura de capacitat i les unitats del vostre workload a l'enunciat general

Aspectos genéricos

Metrica	Valor
Bi-section bandwidth	0,75
Oversubscription rate	Agregación: 1,33:1 Top of rack: 4:1
Máximo ancho de banda con discos	100 Gbps
Capacidad de almacenamiento	8.847,36 TB
Máximo de peticiones/s admitidas	384.000

Red

Bi-section Bandwidth

Si dividimos entre dos nuestro CPD, nos queda un switch de agregación a cada lado conectados entre ellos por un link de 600 Gbps.

A cada switch le llegan 8 links de entrada de 100 Gbps. Por lo tanto el Bi-section Bandwidth es $\frac{600}{800} = \frac{3}{4} = 0,75$.

Oversubscription ratio

$$ov = \frac{1}{Th}, Th = \frac{\text{links salida} * Gb \text{ salida}}{\text{links entrada} * Gb \text{ entrada}}$$

Agregación: Links de entrada 8 de 100 Gbps, links de salida 6 de 100 Gbps,

$$\frac{6*100}{8*100} = \frac{3}{4}, ov = \frac{1}{Th} = \frac{4}{3} = 1,33 : 1.$$

Top of rack, tienen 40 entradas de 10 Gbps y sale un link de 100 Gbps.

$$\frac{1*100}{40*10} = \frac{1}{4}, ov = \frac{1}{Th} = \frac{4}{1} = 4 : 1$$

Almacenamiento

Ancho de banda máximo

El máximo ancho de banda con discos que podemos soportar es 100 Gbps.

Esto se debe a que el único link que conecta los switch de storage con los de agregación es un cable de 100 Gbps. Como cada servidor de storage está conectado al switch correspondiente con un link de 40 Gbps, el cuello de botella en este caso es el link storage-agregación.

De todas maneras podemos escalar esta solución en cualquier momento porque tenemos puertos libres en los dos switches y para nuestra solución actual es más que suficiente.

Capacidad de almacenamiento

$$1,92 \text{ TB/disco} * 72 \text{ discos/JBOD} * 8 \text{ JBOD/rack} = 1.105,92 \text{ TB/rack}$$

$$1.105,92 \text{ TB/rack} * 8 \text{ racks} = 8.847,36 \text{ TB storage}$$

Cuellos de botella entre nodos

El cuello de botella entre los nodos de computación y red es el enlace entre switches del *top of rack* y agregación dado que tenemos 40 servidores con enlaces de 10 Gbps y solo conectamos agregación con el switch del *tor* con un link de 100.

El cuello de botella entre los nodos de storage y red, como comentamos en el apartado anterior, es el link de agregación y storage.

De todas formas dado nuestro diseño estos cuellos de botella nunca pasarían a ser efectivos porque ninguna de nuestras máquinas utiliza el ancho de banda total con el que puede

transmitir, siempre tenemos ancho de banda disponible y hemos diseñado el CPD para que sea capaz de trabajar a máxima capacidad sin problemas siguiendo el RSA establecido.

Datos específicos de los workloads

Máximo número de peticiones HTTP por segundo:

Como hemos calculado anteriormente soportamos como máximo 384.000 peticiones/s.

Además como hemos visto no tenemos problemas de ancho de banda para servirlos.

Mayor *working set* que podemos tener,

Si el número máximo de peticiones por segundo es 384.000 peticiones/s y cada petición ocupa 180 KB, nuestro mayor *working set* que podemos tener en un momento dado es:

$$384.000 \text{ peticiones/s} * 180 \text{ KB/petición} = 553 \text{ Gbps}$$

Consumo

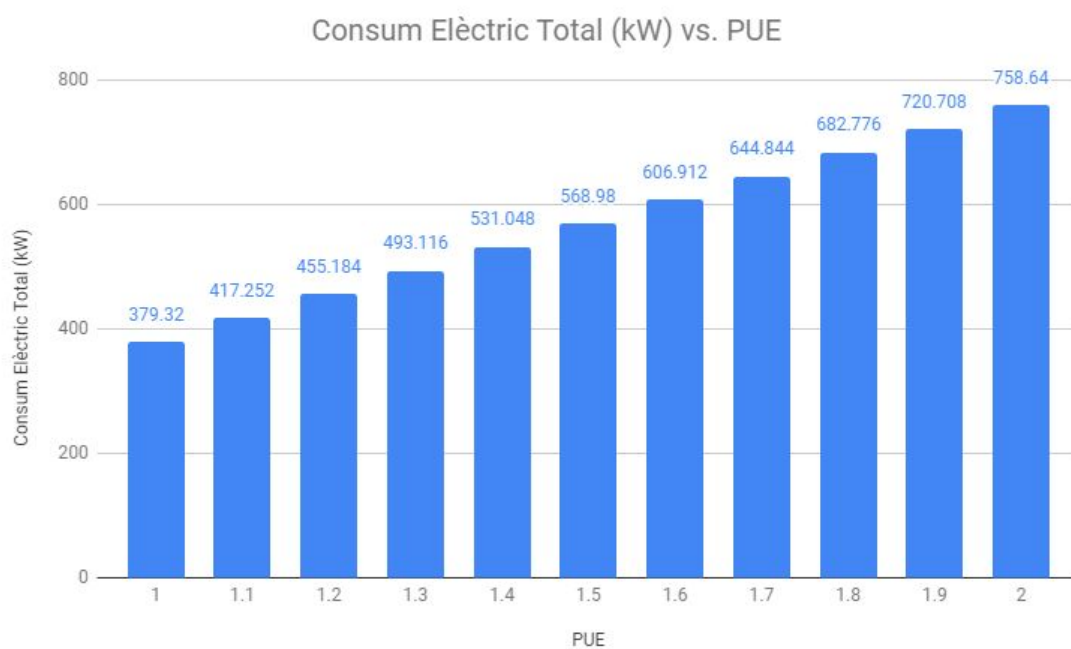
Aquesta secció inclourà un detall del cost computacional de cadascun dels tipus de nodes que hi haurà en el CPD a partir dels seus components. Podeu ajudar-vos del full de càlcul proporcionat junt amb l'enunciat. Assumirem sempre que els consums seran els màxims a l'hora de capacitar el CPD. A part del cost de cada node i equip de comunicació, caldrà estimar el consum mensual global del CPD, així com la calor dissipada quan el CPD està en marxa a un 60% de capacitat (tots els nodes encesos, funcionant cadascun al 60%).

Els aspectes mínims a cobrir seran:

- *Desglossar el consum de potència d'un node per components, així com el consum dels equips de comunicació.*
- *Consum de potència global per a diferents nivells possibles de PUE.*
- *Consum mensual de les màquines – inclou xarxa i emmagatzematge (en kwh), desglossat per node de comunicació, node de computació i sistema centralitzat de disc (si s'escau)*

Para calcular el consumo de nuestras máquinas no añadimos en ningún momento un porcentaje de potencia por riesgo a picos porque siempre hemos calculado el consumo energético basándose en el consumo máximo de las máquinas y no el estimado.

Componente			Cantidad	Consumo	
				Unitario	Total
Computación			1	2,950 kW	275,2 kW
	Red de comunicación				
		Switch top of rack	16	2,2 kW	35,2 kW
	Servidores		320	0,750 kW	240 kW
Storage			1	3,25 kW	85,4 kW
	Red de comunicación				
		Switch	2	1,3 kW	2,6 kW
	Racks				82,8 kW
		JBOD's	64	1,2 kW	76,8 kW
		Servidor	8	0,75 kW	6 kW
Agregación				9,36 kW	18,72 kW
	Red de comunicacion				
		Switch de agregación	2	9,36 kW	18,72 kW
TOTAL					379,32 kW



Componente	Potencia consumida	Consumo mensual
Nodos de computación	240 kW	172.800 kW/h
Nodos de comunicación	56,52 kW	40.694,4 kW/h
Nodos de storage	82,8 kW	59.616 kW/h
TOTAL	375,32 kW	273.110,4 kW/h

Costes

- *Cost total per byte*
- *Cost total per Mhz*
- *Cost elèctric mensual dels equips de computació i comunicació a màxima potència*

Coste de la infraestructura

Coste Switch TOR	Componentes	Cantidad	Precio	Total precio
N3K-C36180YC-R	N3K-C36180YC-R	1	€24,411.02	€24,411.02
Cable corriente	CAB-9K10A-EU	2	€26.57	€53.14
Fuente de alimentación	NXA-PAC-1100W-PI2	2	€733.92	€1,467.84
Ventilador	NXA-FAN-65CFM-PI	2	€143.54	€287.08
transceiver(100Gbps)	QSFP-100G-SR4-S	1	€1,356.66	€1,356.66
transceiver(10Gbps)	SFP-10G-SR	80	€481.96	€38,556.80
Cables (Servidor-TOR)	SFP-10G-AOC5M	40	€195.36	€7,814.40
Cables(TOR-Agreg)	QSFP-100G-AOC20M	1	€1,627.87	€1,627.87
Consumo (W):	2200		TOTAL	€75,574.81
Coste Switch Agregación	Componentes	Cantidad	Precio	Total precio
Linecard / Module	N9K-X9432C-S	2	€20,359.85	€40,719.70
Nexus 9000 Base software	NXOS-703I7.4	1	€0.00	€0.00
Fabric module with 100G sup	N9K-C9508-FM-S	4	€16,967.22	€67,868.88
Fuentes de alimentacion	N9K-PUV2-3000W-B	3	€3,736.03	€11,208.09
Cables (Italia mismo conector	CAB-AC-16A-SG-IT	3	€26.57	€79.71
Supervisor	N9K-SUP-A+	2	€6,789.36	€13,578.72
transceiver(100 Gbps)	QSFP-100G-SR4-S	21	€1,356.66	€28,489.86
Cables 100 Gbps	QSFP-100G-AOC20M	6	€1,627.87	€9,767.22
Consumo (W)	9360		TOTAL	€171,712.18
Extras agregacion (general)	Componentes			
Cables 100 gbps	QSFP-100G-AOC20M	6	€1,627.87	€9,767.22
Chasis	N9K-C9508	1	€10,181.99	€10,181.99
	TOTAL EXTRA			€19,949.21

Para el coste del TOR tenemos 40 transceiver de 10 Gbps que son los que van en la conexión del servidor y otros 40 para que se conecten al TOR. También necesitamos 1 transceiver de 100 Gbps para conectar el servidor con el switch de agregación. Entre el servidor y el TOR necesitamos 40 cables de 10G y entre el TOR y el switch de agregación 1 cable de 100G.

Para el coste del Switch de Agregación tenemos 8 transceivers para conectar TOR, 6 transceivers para conectar el switch hacia el exterior, 6 transceivers para tener conexión entre los switches y 1 transceiver para el storage.

Aquí solo ponemos los 6 cables de 100G para conectar el switch al exterior debido a que para conectar los switches con los TOR ya lo hemos considerado en el apartado anterior.

Como extra en la parte de agregación tenemos el chasis del switch y los 6 cables para conectar los switches de agregación entre ellos.

Coste Rack Computacion	Componentes	Cantidad	Precio	Total
Servidor	RAX XS4-21S1-10G	40	€12,709.93	€508,397.20
Switch TOR	N3K-C36180YC-R	2	€75,574.81	€151,149.62
Chasis Rack		1	€400.00	€400.00
PDU	AP8953	1	€1,270.50	€1,270.50
TOTAL				€661,217.32
Coste Computacion		Cantidad	Precio	Total
Rack computacion		8	€661,217.32	€5,289,738.56
Switch Agregacion		2	€171,712.18	€363,373.57
				€5,653,112.13

Para el coste del rack en computación tenemos los 40 servidores y los 2 TOR junto con el chasis del rack y la PDU.

Por lo tanto para el tema de computación tenemos los 8 racks y los 2 switches que nos dan un total de 5.653.112,13 euros

Coste Switch de storage	Componentes	Cantidad	Precio	Total
Switch:	N3K-C3232C-B8C	1	€27,209.26	€27,209.26
Fuente alimentación	NXA-PAC-650W-PI	2	€397.39	€794.78
Cables corriente	CAB-9K10A-EU	2	€26.57	€53.14
Ventiladores	NXA-FAN-30CFM-B	4	€111.06	€444.24
transceiver(40)	QSFP-40G-SR4	16	€2,016.94	€32,271.04
transceiver100	QSFP-100G-SR4-S	2	€1,356.66	€2,713.32
Cables 40 gbps	QSFP-H40G-AOC10M	8	€677.51	€5,420.08
Cables 100 Gbps	QSFP-100G-AOC20M	2	€1,627.87	€3,255.74
Consumo	1300		Total	€72,161.60

Para el coste del switch del storage necesitamos 8 transceivers de 40 para los 8 servidores de storage y 8 transceivers de 40 para poderlos conectar al switch. También necesitamos 2 transceivers de 100 en el switch para que se pueda conectar con los 2 switches de agregación.

Usaremos 8 cables de 40 para los servidores y 2 cables de 100 para los switches de agregación

Coste Rack Storage				
Servidor		1	€5,703.26	€5,703.26
JBOD		8	€48,964.82	€391,718.56
Chasis Rack		1	€400.00	€400.00
				€397,821.82
Coste Storage				
Rack Storage		8	€397,821.82	€3,182,574.56
Switch Storage		2	€72,161.60	€144,323.20
				€3,326,897.76

Para el coste del rack del storage hemos tenido en cuenta 8 JBODs y el servidor que gestionará estos JBODs junto con el chasis del rack.

Por lo que como coste del storage tenemos los 8 racks y el switch del storage que nos sale a un total de 3.326.897,76 euros

Coste Trafico 5 años	
Peticiones/Servidor	1200
Coste peticion (KB)	180
Ancho de banda/servidor	1.728
Servidores	320
Total ancho de banda (Gb)	552.96
Cost trafico mensual (€/Gb)	€63.00
	€2,090,188.80

Para el coste del tráfico hemos cogido de peticiones que hace cada servidor por lo que ocupa una petición. Con esto conseguimos el ancho de banda por cada servidor que por la cantidad de servidores que tenemos en total nos da el ancho de banda total. Una vez teniendo esto y partiendo que el coste de trafico mensual nos cuesta 63 euro al mes, tenemos que el coste total por todo el ancho de banda durante 5 años es de 2.090.188,80

Coste CPD Infraestructura	
Storage	€3,326,897.76
Computacion	€5,653,112.13
Trafico	€2,090,188.80
Consumo	€420,352.40
TOTAL	€11,490,551.09

Por lo tanto como coste final de toda la infraestructura tenemos que el storage nos ha ocupado aproximadamente los 3 millones que teníamos planeado, la computación es lo que más nos ha ocupado, luego le sigue el trafico que vendria a ser lo que nos cuesta tener acceso al exterior y finalmente el consumo que es lo que gastan todos los servidores y switches. El total que nos da todo esto es 11.393.721,67 euros que se aproxima bastante a los 12 millones que teníamos como presupuesto.

Hemos utilizado una hoja de cálculo para facilitar todas las operaciones, adjuntamos el link para que se pueda consultar de donde vienen todos los números.

https://docs.google.com/spreadsheets/d/1TYsS5tkBqIt_C_ndO9LnWFX0Cnx6BBLjNrvGVMwdEuA/edit?usp=sharing

Coste por byte

Precio dividido entre Storage total:

$$11,5\text{M€} * \frac{1}{8.847,36 \text{ TB}} * \frac{1 \text{ TB}}{2^{40} \text{ bytes}} = 1,18 * 10^{-9} \text{ €/byte}$$

Coste por Mhz

Mhz totales en procesador entre el coste total. Tenemos servidores con procesadores de 2,4 Ghz, 20 cores cada uno y los servidores son *Dual-Socket*. En total tenemos 320 servidores y el coste total de nuestro CPD es aproximadamente 11,5M € por lo tanto el coste por Mhz es:

$$11,5\text{M€} * \frac{1 \text{ core}}{2,4 \text{ Ghz}} * \frac{1 \text{ Ghz}}{1000 \text{ Mhz}} * \frac{1 \text{ procesador}}{20 \text{ cores}} * \frac{1 \text{ servidor}}{2 \text{ procesador}} * \frac{1}{320 \text{ servidor}} = 0,374348958333 \text{ €/Mhz}$$

Coste eléctrico mensual de los equipos de computación y comunicación

DADES DE CONSUM DELS EQUIPS DE COMPUTACIÓ I COMUNICACIÓ						
Càlcul aproximat de factura, per una línia d'alta tensió (20kV) en mode tarifari 6.1A						
	Períodes					
	P1	P2	P3	P4	P5	P6
Cost €/kW contractat i any	€17,683102	€8,849205	€6,476148	€6,476148	€6,476148	€2,954837
Cost €/kWh consumit	€0,075697	€0,056532	€0,030124	€0,014992	€0,009682	€0,006062
Sobreprovisionament de potència	0%	-> Contracto potència per sobre del màxim consum esperat per seguretat				
kW consumits per computació i comunicació	379kW	-> Potència màxima que espero consumir basat en HW				
Tarifas vigentes de electricidad a partir del 1 de enero de 2013, publicadas en el BOE de 27 de diciembre de 2012.						
Caselles a modificar per l'estudiant						
PUE	Consum Elèctric Total (kW)	kW contractats necessaris	Cost "Término Potencia"	Cost "Término Energía"	Cost Total Energia Anual	Sobrecost
1.00	379.32	379	€18.554,66	€65.515,82	€84.070,48	0,00%

Coste anual: 84.070,48€.

$$\text{Coste mensual} = \frac{84.070,48}{12} = 7005,87 \text{ €}$$

Escalabilidad

Aquesta secció contindrà una valoració qualitativa i quantitativa de les oportunitats d'expansió del CPD i el seu cost associat. Caldrà realitzar projeccions de com els costos i la capacitat escalarien amb possibles expansions del CPD. Aquesta és una secció oberta en què es valorarà la creativitat de l'estudiant a l'hora de proposar millores.

Como hemos diseñado el CPD basándonos en la capacidad de computación de nuestros nodos, las mejoras se basan en el incremento de nodos de computación y las ampliaciones en las infraestructuras de redes.

Con la solución actual por cada dos racks que añadimos necesitamos ampliar un puerto de agregación para la interconexión y un puerto de agregación para el NSP.

Actualmente tenemos 12 puertos libres en agregación y una ampliación de añadir dos racks nos costaría 4 puertos, dos para conectar los racks a agregación y 2 adicionales por switch para conectarlos entre ellos y con el NSP.

Por otra parte no tenemos en cuenta el storage en estas ampliaciones porque tenemos ancho de banda y capacidad suficiente para ampliaciones.

Según los cálculos de arriba podríamos ampliar 3 veces nuestro CPD, por lo que sería añadir 6 racks nuevos y 3 enlaces de agregación al exterior junto con 3 enlaces entre switches de agregación para soportar todo el tráfico, con sus respectivos transceivers.

Esta ampliación nos costaría 9.827.299,35 € y nos permitiría enviar un total de 672000 peticiones, lo que es un aumento del 42,85%.

Para más ampliaciones deberíamos de modificar la arquitectura de red y cambiar los switches de agregación para poder utilizar más puertos. Como nuestro chasis de agregación aún tiene espacio para más switches podríamos añadir nuevos switches a este nivel sin problemas.

Anexo

Puntos a considerar

A último momento y con la solución ya diseñada nos hemos dado cuenta de que el servidor de Storage usa links y transceivers de 40 Gbps pero el switch de storage tiene solo links de 100 Gbps y por lo tanto no son compatibles. Dado a este problema deberíamos de cambiar este switch o buscar otra solución como puede ser utilizar cables de 25 Gbps y splitters el switch para juntarlos todos, lo que implicaría tener que buscar nueva tarjeta de red para el servidor o un nuevo servidor que tuviese tarjetas de red de 25 Gbps en el caso de no encontrar para el seleccionado actualmente.

Referencias

Switch de TOR:

<https://www.cisco.com/c/en/us/products/switches/nexus-3000-series-switches/models-comparison.html#~tab-nexus3600>

Switch de agregación:

<https://www.cisco.com/c/dam/en/us/products/switches/nexus-9000-series-switches/nexus-9500-100GE-modules-comparison.html>

Switch de Storage:

<https://www.cisco.com/c/en/us/products/switches/nexus-3000-series-switches/models-comparison.html#~tab-nexus3200>

Transceivers compatibles:

https://www.cisco.com/c/en/us/products/collateral/switches/nexus-3000-series-switches/data_sheet_c78-729483.html

Rack

<https://www.pccomponentes.com/armario-rack-19-42hu-600x800>

Páginas para buscar los cables

10Gbps:

https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/data_sheet_c78-455693.html

100 Gbps:

<https://www.cisco.com/c/en/us/products/collateral/interfaces-modules/transceiver-modules/datasheet-c78-736282.html>