

CPD: Activitat 2

Curs 2018-2019, tardor.

Alumne: Genís Bosch, Oriol Fonollà

Data: 13/09/2018

Número d'activitat: 12/6

Descripció de l'activitat:

Storage: NAS/SAN

Workload: Web 2.0

Budget: 12M€

Descripció general de la solució proposada

La nostra activitat consisteix en crear un CPD on les especificacions donades són:

- 12M € de pressupost
- Workload de tipus Web 2.0
- Storage NAS/SAN

Donat que el workload és de tipus web 2.0, avaluarem la càrrega de treball en peticions per segons que podem resoldre.

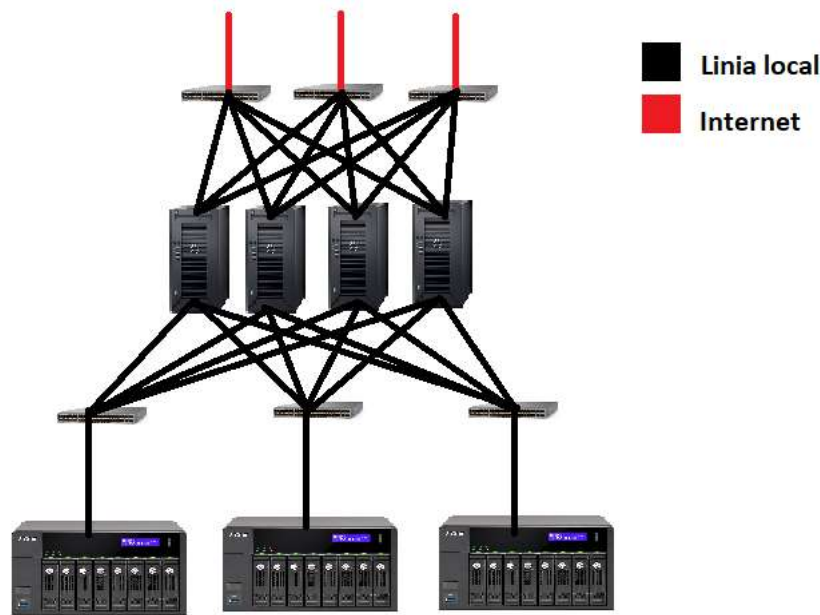


Figura 1. Disseny global simplificat del CPD.

Cóm podem veure en la Figura 1, el disseny global del CPD és de tipus SAN/NAS, i per tant els servidors no són els encarregats de gestionar les dades, sinó que només resolen les peticions fent accessos a discos. Així doncs, contem amb un potent sistema de switchs per reduir els colls d'ampolla entre els servidors i l'emmagatzematge de dades.

Per tant, comptem amb uns servidors connectats, per una banda, a la xarxa d'Internet, per on rebran les peticions i retornaran els resultats, i per l'altre a disc, on accediran per a obtenir les imatges que se'ls han sol·licitat.

Esquema físic de configuració

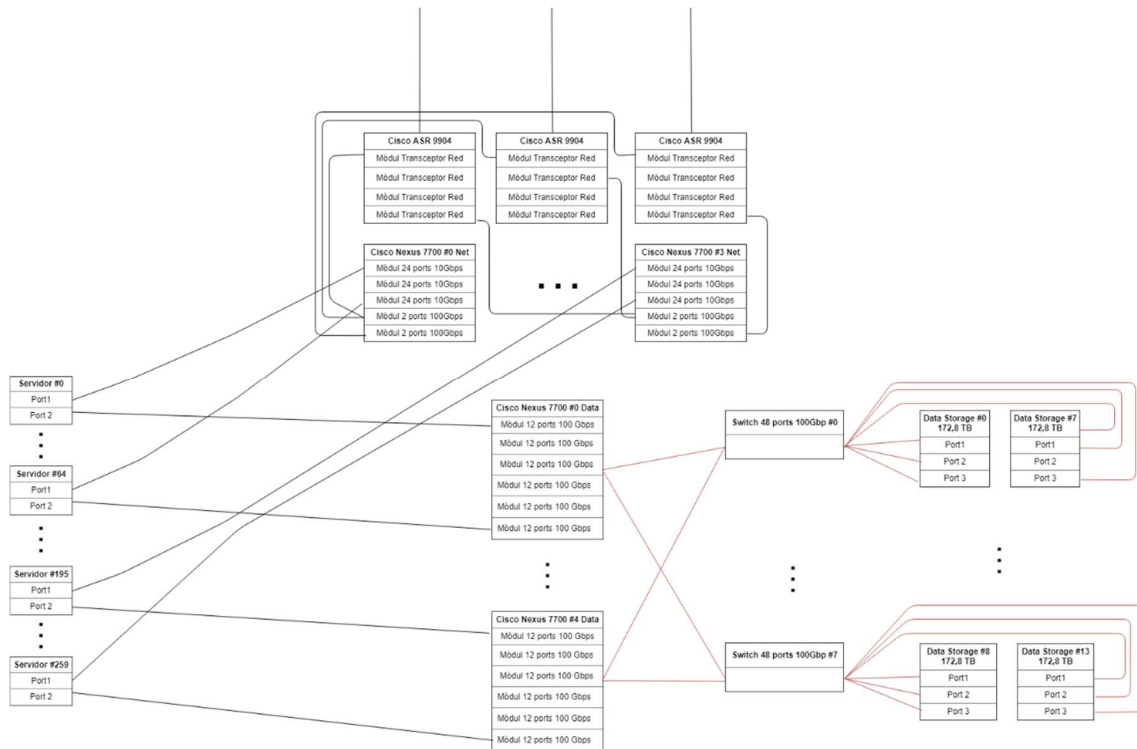


Figura 2. Esquema complet del disseny del CPD

Per al disseny del nostre CPD proposem els següents components:

- **Servidors:** Construïts dinàmicament a la pàgina web d'intel, utilitzarem un xassís Intel Server Chassis R1304WFXML. Sobre aquest, hi posarem un processador Intel Xeon Platinum 8180M, amb 28 cores, cadascun d'ells corrent a una freqüència de 2.50GHz. El preu total per a un servidor és de 18758,32 €. Són de 1U.
- **Data storage:** Utilitzarem servidors d'emmagatzematge construïts sobre les necessitats a thinkmate.com. Aquests seran servidors de 72 discs per a cada un, amb uns discs de 2.4 TB i una velocitat de lectura de 12Gb/s. Això significa que per a cada servidor de dades, obtenim un total de 172 TB d'emmagatzematge, a un preu total de 70467.45 €. Són de 4U.
- **Switchs i altres components:** Seran especificats a mesura que apareixin durant la descripció del CPD.

Així doncs, un cop nombrats els principal components, procedim a l'explicació de l'organització.

Utilitzarem 260 servidors, distribuïts en racks de 42U. Hem de contar també amb l'alimentació, la qual es tracta de 1 U per a cada servidor, així doncs, obtenim 2U per a cada servidor, el que fa un total de 520U. Això són, per tant 13 racks dedicats als servidors.

Aquests servidors tenen dues targetes de connectivitat, una per a connectar-se a internet i l'altre per accedir al data storage. Aquests primers tindran un ample de banda de 10Gbps. Per tant, disposarem de 260 cables de 10Gbps connectant els servidors amb els switchs d'internet.

Aquests switchs, dels quals disposarem de 4, seran construïts a partir d'un Cisco Nexus Chasis 7700 i 5 mòduls. Dos d'aquests mòduls seran de 2 ports de 100 Gbps, ja que hauran de connectar amb les línies d'internet (disposarem de 3) de 100Gbps amb tecnologia CFP-100G. Els altres 3 mòduls seran de 24 ports de 10 Gbps cada un, per així poder connectar 65 servidors cada switch, obtenint el total de 260 servidors al utilitzar 4 Nexus. Cada un d'aquests utilitzarà un transceiver per a cada cable.

Com aquests switchs ocupen 18U cada un, i n'hi hem de contar 6 més de l'alimentació, dedicarem un rack per a cada Nexus, invertint així 4 racks per a aquests.

Connectant amb aquests 4 switchs, i en forma de malla completa, hi haurà 3 xassís de Cisco ASR 9904, un per a cada línia contractada, els quals contindran cada un 4 mòduls transeptors per a rebre l'ample de banda de 100Gbps de cada un dels Nexus 7700. Utilitzaran també transceivers per a cada cable de cada mòdul. Aprofitant que aquests switchs/routers ocupen només 10U, en col·locarem un a cada rack dels Cisco Nexus 7700.

Al arribar a aquest punt, ja hem descrit la part que contempla la unió dels servidors amb la xarxa exterior, internet. Ara ens centrarem en la part interconnectat entre els servidors i els data storages.

Procedim, per tant a descriure els switchs intermitjos entre servidors i data storages:

En primer llocs, utilitzarem xassís de Cisco Nexus 7700. Sobre aquests, hi afegirem 6 mòduls, que és el màxim, de 12 ports d'ample de banda de 100Gbps cada un. N'utilitzarem 4, un per a cada bloc de 65 servidors, els quals es connecten directament al switch, ocupant 65 ports dels 72 disponibles. Els altres 7 ports restants aniran connectats a 7 switchs que agruparan l'ample de banda dels data storages. Cada Nexus 7700 ocupa, com hem dit abans, 18U, així que els hi dedicarem dos racks, que compartiran dos a dos, i sobran 6U a cada rack, ocupats per les fonts d'alimentació i els switchs d'agrupació.

Aquests switchs, força més barats, ens ofereixen 32 ports de 100Gbps. Ocupen 1U, així que caben en els espais restants dels racks ocupats per els Nexus 7700. Per una banda, connectaran amb els 4 Nexus, creant una malla completa. Per l'altre banda, cada switch agruparà 6 cables provinents dels servidors de storage, ja que aquests tenen 3 targetes de connexió.

Aquests últims, els servidors de storage, ocupen 4U, i n'utilitzarem un total de 14, així que els hi haurem de dedicar dos racks més.

Capacitat

Per als càlculs de capacitat de dades i processat, partim de que tenim 260 servidors, amb 28 cores cada un, i amb una freqüència per core de 2.5GHz. Sabem que cada core necessita 4GB de memòria, així que per a fer funcionar els 28 cores, necessitem un total de 112 GB, cosa que hem tingut en compte per a seleccionar el servidor.

Amb aquests números, i contant que a cada petició se li dediquen 800MHz i 100ms, obtenim que un core pot gestionar 3 peticions cada 100ms, i per tant, 30 peticions cada segon. Això, escalat als 28 cores de cada servidor, fan un total de 840 peticions per segon a cada servidor.

Amb una mitjana de 180 KBytes de resposta per petició, obtenim un ample de banda a cada servidor, provinent d'internet, de 864000 Kbps, o el que és el mateix, 843,75Mbps.

Agrupant 65 servidors, obtenim un total de 53.55 Gbps, i agrupant els 4 grups d'aquests, un total de 214.23 Gbps d'ample de banda total cap a internet, que podem dividir en 3 línies de 100 Gbps cada una.

Per altre banda, els accessos a disc són de 5KB per a cada petició, per el que, cada segon, un servidor accedeix a un total de 4.1 MB, o el que és el mateix, obté un ample de banda de 32.8 Mbps. Tot i això, a cada servidor li assignem un cable de 100Gbps per si les imatges sol·licitades en un moment donat són molt més grans.

Cada switch Nexus rep, per tant, un ample de banda acumulat de 32.8Mbps per cada servidor, i en gestiona un total de 65, per el que l'ample de banda total és de 2.08 Gbps.

Llavors, tot i haver-hi una malla completa entre switchs Nexus i els switchs de 32 ports, contant el pitjor dels casos, en un switch, l'ample de banda pot arribar a ser de $7 * 2.08$, és a dir, 14.57 Gbps, per el que amb cables de 100 Gbps en tenim suficient.

Finalment, els tres cables que van a cada servidor de storage tindran, com a molt, un ample de banda de 4.86 Gbps, ja que l'ample de banda es pot dividir per igual entre els 3 cables. Fem, també, per tant, amb els cables de 100 Gbps. Tot i això, hem decidit adquirir uns servidors de storage de 12Gbps d'ample de banda de lectura, per si es carrega tot l'ample de banda en un sol disc.

Pel què fa a la capacitat, tenim un total de 14 servidors d'emmagatzematge, amb una capacitat de 172TB cada un, és a dir, un total de 2.35 PB.

Consum

El consum del nostre centre de processament de dades correspondria a la suma dels consums dels diversos elements. Així doncs, tindríem que els 230 servidors, consumint cadascun 750 W gastarien un total de 172.5 KW. A això hi hauríem de sumar els diversos switchs, 8 dels quals consumeixen 9000 W i els 7 restants 550 W. Per tant, a falta de el consum dels data storages, el nostre centre de processament de dades ja gasta 248.35 KW. Contant que els 14 data storage consumeixen 3600 W cadascun el consum total s'eleva fins a 273.55 KW.

Per assegurar-nos que sempre tindrem el consum desitjat contractat i podrem fer possibles ampliacions o connexions de diferents elements, hem decidit contractar un 9% mes de tarifa de consum, la qual s'eleva fins als 300 KW i té un cost de 67.370,98€ al any, 336.854,5€ al llarg dels 5 anys.

Complexitat de Gestió HW/SW

Per a administrar el hardware, és viable l'ús d'uns switch KVM, ja que tant servidors com data storage formen part de la mateixa xarxa, i amb trobar un punt on connectar aquests, l'accés pot ser a qualsevol màquina del CPD.

En quant a la protecció del CPD, utilitzar un sistema de PDU's monitoritzables aniria bé de cara a apagar servidors en moments del dia en els que el workload no és tant gran, ja que segurament la càrrega no és balancejada.

Costos

El cost total del CPD és de 10.405.449€. Queden desglossats de la següent forma:

	Quantitat	Preu	Preu parcial
Servidors	260	18800	4888000
Cables 100 Gbps	345	440	151800
Cables 10 Gbps	260	115	29900
Transcivers	603	700	422100
Data storage	14	70500	987000
Swich basic	7	7936	55552
Xassís Nexus	8	67000	536000
Mòdul Nexus 12 p 100 gb (server a switch)	24	58000	1392000
Mòdul Nexus 24 p 10 Gbps	12	58000	696000
Mòdul Nexus 2 p 100 Gbps	8	49000	392000
Mòdul Tranceptor xarxa	12	1300	15600
Chasis Cisco asr 9904	3	103530	310590
Internet	3	37800	113400
Energia	1	336855	336855
Rack swich ASR	1	2668	2668
Rack servers	8	2668	21344
Power supply server	520	45	23400
Power supply nexus	48	45	2160
Rack nexus	8	2668	21344
Rack storage	2	2668	5336
Power supply cisco ASR 9904	3	800	2400
Preu Total			10405449

Contant que cada servidor té 28 cores, i cada un va a una velocitat de 2.5GHz, cada servidor acumula un total de 70GHz, i amb els 260 servidors, això són 18.200 GHz, i per tant, es tracta de de $10.405.449 / 18.200$, 571,72 €/GHz, que són 0,57€/MHz.

Per a cada byte, partint de 2.35PB, ens surt a un cost de 3×10^{-9} € byte.

Aquests servidors estan alimentats amb 750W, per tant el total de potència consumida per aquests, és de 195 kW. Això, de forma mensual són 3601€ mensuals.

Els equips de commutació o switches, consumeixen un total de 51,5 kW, que aplicant-hi la tarifa energètica, surt a 960,41€ mensuals.

Escalabilitat

Per a aconseguir aquest pressupost, hem iterat sobre la construcció base d'un bloc de 65 servidors, i això es pot seguir augmentant, però s'han d'incorporar, també, més Cisco Nexus 7700 per a gestionar l'ample de banda, així com també més línies d'internet. Per el que fa a l'emmagatzematge, només cal ampliar-lo quan l'ample de banda sigui superior a la velocitat de lectura dels discs, que és de 12 Gbps.

Aquest sistema base pot ser escalat tantes vegades com ports dels switchs de emmagatzematge quedin lliures, per tant, contant que només ocupem 10 ports, podem replicar-ho encara 22 vegades més, però això si, augmentant el nombre de mòduls ASR i per tant, també, el nombre de línies d'internet.