

PLEC DE PRESCRIPCIONS TÈCNIQUES

I. OBJECTE

L'objecte d'aquest plec és definir els requeriments tècnics de *l'adquisició i instal·lació d'un nou sistema d'emmagatzematge per el sistema de computació d'alt rendiment del Servei Científico-Tècnic d'Informàtica* (Scientific IT core facility, en endavant SIT) de la Universitat Pompeu Fabra.

Aquest sistema ha d'incloure:

- El maquinari necessari
- Les llicències del programari
- Garantia de maquinari
- Suport i actualització del programari
- Cablejat i altres component per a la connexió amb la resta de sistemes existent.
- Instal·lació, posada en marxa i formació al personal tècnic.

Aquesta adquisició està finançada en part pel programa Maria de Maeztu del Departament de Medicina i Ciències de la Salut de la UPF.

II. JUSTIFICACIÓ DE LA CONTRACTACIÓ

El SIT disposa de diferents infraestructures informàtiques que ofereixen servei IT a la comunitat recercadora de la Universitat. Una d'aquestes infraestructures és un sistema d'emmagatzemament d'altres prestacions i capacitats que treballa conjuntament amb el clúster de computació (HPC).

Aquest sistema està arribant al final de la seva vida útil, tant d'hores de funcionament com de suport per part del fabricant.

Així doncs, cal substituir l'actual sistema emmagatzemament per un de nou, de característiques semblants, que per una banda permeti assumir la càrrega actual així com la previsió de servei previst per la seva vida útil. El nou sistema ha de garantir la compatibilitat amb el sistema de computació actual, que queda descrit en l'annex I, a l'hora que permeti guanyar en prestacions per possibles ampliacions d'aquest sistema.

Aquest nou sistema s'haurà d'instal·lar físicament al Campus de la Ciutadella.

III. REQUERIMENTS TÈCNICS

Els licitadors han de presentar una proposta de maquinari i programari que compleixi com a mínim amb els requisits establerts per a substituir l'actual sistema d'emmagatzematge de la infraestructura computacionals del SIT detallada en l'annex II i que a continuació es desenvolupa:

Descripció del sistema d'emmagatzematge

1. Donada la diversitat en el paradigma de la ingesta, emmagatzemament i explotació de dades de l'àrea de la recerca en medicina i ciències de la vida, la infraestructura ha de poder-se configurar i adaptar de forma dinàmica per poder executar i processar correctament:

- 1.1. Aplicacions de l'àmbit de HPC Classic (codis basats en MPI/OpenMP amb alt rendiment d'operacions en coma flotant, xarxa de baixa latència i sistema d'arxius paral·lel d'alt rendiment)
 - 1.2. Aplicacions amb alta demanda d'emmagatzemament utilitzats tan en paradigmes de virtualització o contenidors, com el processament de dades en sistemes HPC i en el processament d'aplicacions d'escalat horitzontal o BigData.
2. Per la diversitat de projectes que la infraestructura ha de poder realitzar i la confidencialitat de les dades biològiques per a la recerca que s'hi ha d'emmagatzemar, cal un sistema d'arxius paral·lel com l'existent que ofereixi les següents característiques:
 - 2.1. Definició de *filesets* (possibilitat de particionar un filesystem en diversos elements amb espai de noms administrativament independents, amb els seus propis usuaris, quotes, punts de muntatge, etc.)
 - 2.2. Sistema integrat de *Informations lifecycle management* que permeti la implementació de polítiques i accions segons un sistema de regles segons diferents paràmetres, com a mínim, tamany del fitxer, temps d'accés i/o modificació, tecnologia de disc-
 - 2.3. Sistema de metadades distribuït i balancejat de dades en l'inode
 - 2.4. Suport de muntatge natiu en màquines virtuals mitjançant protocol IP
 - 2.5. Capacitat de compressió de dades
3. El sistema d'emmagatzematge no ha de tenir cap punt de fallada únic
 - 3.1. El nivell de gestió de l'emmagatzematge ha de comptar amb un mínim de dos servidors idèntics redundants per l'accés a disc. El sistema de redundància ha de permetre que la caiguda d'un dels servidors no aturi el servei.
 - 3.2. El sistema d'emmagatzemament haurà de disposar de sistemes de protecció hardware de fallades habituals en aquests sistemes. RAID amb un mínim de doble paritat, espai d'spare per fer front a la fallada de discs equivalent el 5% de cada una de les tecnologies ofertes.
 - 3.3. Cal dotar del sistema d'un mínim de dos servidors per servir els protocols NFS v4, NFS v3, GPFS, SMB 2, SMB 3.0, POSIX-compliant, HDFS 3.1.0, S3/Swift. Aquests servidors han de poder balancejar la carga i fer front a la caiguda d'algun dels seus components. De la mateixa manera, ha de permetre la seva ampliació en cas saturació en nombre de servidors.
4. El sistema d'emmagatzematge ha de comptar amb un mínim de 900TB de capacitat utilitzable neta i una arquitectura amb un mínim de dos nivells. (Entenem com a capacitat neta la resultant un cop es descompten els sistemes de redundància i espai d'spare. En definitiva, la capacitat màxima en dades i metadades de fitxers que pot fer front el sistema d'arxius)

- 4.1. Un màxim d'un 90% de la seva capacitat neta en discos rotacionals NL-SAS o equivalents per hostatjament de dades.
- 4.2. Un mínim d'un 10% de la capacitat neta estarà ubicat en discos de tecnologia NVMe. Aquesta tecnologia serà utilitzada per a dos propòsits:
 - 4.2.1. Un mínim equivalent al 3% de la capacitat neta total del sistema serà destinada a les Metadades. Les metadades només han d'estar en tecnologia NVMe.
 - 4.2.2. La resta d'espai NVMe es destinarà a l'hostatjament de dades, ja sigui en sistemes d'arxius amb múltiples tecnologies de disc (o hostatjament) o amb sistemes d'arxius únicament NVMe.
- 4.3. Sistema únic per a la gestió d'ambdós tipus de tecnologia de disc.
- 4.4. Els sistemes de redundància RAID o similar han de tenir un mínim de doble paritat.
- 4.5. Disposarà d'un espai en espera per assumir una possible caiguda de disc, anomenem com Espai d'spare, d'un mínim d'un 5% dels discos en cada una de les tecnologies ofertes (NL-SAS i NVMe).
5. El sistema ha de permetre una ampliació mínima de fins a 3 vegades la capacitat proposada en NL-SAS de manera no disruptiva.
6. El nivell de gestió de l'emmagatzematge, format per un mínim de 2 servidors anomenats anteriorment, haurà de complir 3 funcions:
 - 6.1. Emmagatzemar i servir les metadades del sistema d'arxius del primer nivell d'un sistema d'emmagatzemament jeràrquic. (HSM)
 - 6.2. Realitzar tasques de ILM (scan i generació de llistes) sobre les regles definides de migració i recall de HSM
 - 6.3. Crear un sistema d'arxius d'alt rendiment en IOPS on s'emmagatzemaran les dades administratives del sistema HSM (base de dades, recovery logs, etc.)
 - 6.4. Crear altres sistemes d'arxius d'alt rendiment en IOPS on s'emmagatzemaran les dades administratives dels servidors (base de dades recovery logs, etc.)
7. Els sistemes d'arxius han de ser gestionats per software de virtualització de fitxers amb les següents característiques funcionalitats/capacitats:
 - 7.1. Agnòstic (no dependre d'un hardware/fabricant concret) al maquinari sobre el que resideix.
 - 7.2. Namespace únic i sense límit de creixement
 - 7.3. Explotació de les dades a través de múltiples protocols (mínim NFS v4, NFS v3, GPFS, SMB 2, SMB 3.0, POSIX-compliant, HDFS 3.1.0, S3/Swift) només tenint la dada emmagatzemada una vegada al sistema d'arxius
 - 7.4. Ampliable a nivell de còmput o d'emmagatzematge (i agnòstic al HW)

- 7.5. Comprensió i xifrat de les dades
- 7.6. Esborrat segur
- 7.7. ILM (Information Lifecycle Management - Gestió del cicle de vida de la dada) transparent i automàtica
- 7.8. Es valorarà el tipus de tiers oferts, que millorin la possibilitat de desbordament a pools externs, cintes o cloud, sense cost de llicència.
- 7.9. Ubiquïtat - Integració amb contenidors a través del driver CSI certificat i integrat a l'Openshift.
- 7.10. Capacitat d'objectivació de la dada. Útil per poder utilitzar la dada incorporada en modus NFS com una dada de tipus objecte i poder utilitzar-la en aplicacions de nova generació (aplicacions de mòbil, per exemple)
- 7.11. Capacitat de poder fer Snapshots a nivell de sistema d'arxius sense penalització.
- 7.12. Capacitat per crear múltiples sistemes d'arxius
- 7.13. Ingesta de dades en el sistema d'arxius via Cloud des d'una Cloud externa (Amazon, Azure, IBM Cloud Object Storage)
- 7.14. Sistema d'arxius dimensionable en calent
- 7.15. Accés concurrent en modus escriptura a diferents usuaris
- 7.16. Accés accelerat (WAN caching) a recursos basats en sistemes d'arxius via protocols NAS des d'altres CPDs, seus remotes
- 7.17. Immutabilitat de la dada. Capacitat d'arxivat "Worm"
- 7.18. Auditing, Logging & Accounting, per poder tenir control de fitxers i els seus usos i/o accessos. Inclòs el control de quotes per poder posar límit de capacitats o poder visualitzar la capacitat ocupada per un usuari o grup
- 7.19. Integració amb REST API
- 7.20. Suport del mateix fabricant que proveeix la infraestructura de disc en modus appliance sobre la solució desplegada
- 7.21. Versions de programari potencialment desplegables en mode cloud sobre AWS i/o IBM cloud
- 7.22. Callhome integrat
- 7.23. Certificació T10 DIF per integritat de la dada
- 7.24. Suport a algorismes de comprensió LZ4, zfast, aphae, alphh per dades genòmiques i de recerca.

8. Comunicacions

8.1. Infiniband

El sistema actual es disposa de:

- 1 switch Mellanox IB-2 SB7800 36 EDR ports (Infiniband EDS 100Gb/s).
- 17 dels ports estan lliures.

Per tal de connectar els nous servidors a la xarxa de baixa latència i per compatibilitat tant amb l'actual switch com el maquinari computacional existent i client de la infraestructura actual, es demana:

8.1.1. El maquinari subministrat disposi de compatibilitat per connectar amb aquest switch i els nodes clients actuals.

8.1.2. Proporcionar el cablejat necessari per a la seva connexió.

8.1.3. És valorarà com a millora l'actualització d'aquest equipament. Només s'acceptaran millores que compleixin amb:

- a. El sistema haurà de ser el més actual de la seva família, compatible amb la infraestructura restant i dimensionat per a substituir l'actual.
- b. Fonts d'alimentació redundants i ventiladors hot-swap.
- c. Que permeti la connexió amb el model Mellanox IB-2 SB7800 36 PORTS o assumeixi la seva carga actual (infiniband EDR-100Gb/s).
- d. Ha d'incloure els cables per de connexió entre les dos switchos (tenint en compte que la distància prevista entre ells pot ser de fins a d'uns 15m)

8.2. Ethernet

La infraestructura actual està connectada a:

- 1 switch Cisco Catalyst 4500
- 21 1GB baseT lliures

Aquest switch està connectat amb el troncal de la UPF a través de dues connexions 10G Multi-Mode per fibres de 50 micres.

Els actuals equipaments del troncal, on connecta el switch de la solució actual són els switchos següents:

- 2 switchos Cisco Fex-101 Nexus2348TQ
- 15 ports 10GbE lliures a cada switch.

D'aquesta manera es garanteix l'alta disponibilitat en cas de caiguda d'un dels troncats.

Per tal de connectar els nous servidors a la xarxa troncal i per compatibilitat tant amb els switchos anteriorment descrits com el maquinari existent i clients de la infraestructura actual, es demana:

- 8.2.1. El maquinari subministrat disposi de compatibilitat per connectar amb aquest switch i els nodes clients actuals.
- 8.2.2. Proporcionar el cablejat necessari per a la seva connexió, tenint en compte la prioritització de connectivitat en fibra Multi-Mode per velocitats superiors a 1Gbps.
- 8.2.3. És valorarà com a millora l'actualització de equipament Cisco Catalyst 4500. Només s'acceptaran millores basades en switchos amb ports 10G o superiors amb les següents característiques amb el mínim de característiques operatives del switch actual i plenament compatible amb els switchos Cisco Fex-101 Nexus2348TQ i les següents característiques mínimes:
 - a. El sistema proposat ha de ser el més actual de la seva família, compatible amb la infraestructura restant i dimensionat per a substituir l'actual.
 - b. 2 Ports 10G per la connexió amb el troncal, amb les corresponents fibres Multi-Mode de 50 micres de 2m i els seus SFP
 - c. un mínim de 24 ports 10G pels servidors de la solució proposada.
 - d. Fonts d'alimentació i ventilació redundades (hot-swap)

9. Armari RACK

- 9.1. Armari de 19" standard de 42U, per a hostatjar el material ofert.
- 9.2. S'ha de lliurar muntat i llest per a col·locar al CPD.
- 9.3. S'han d'incloure tantes PDUs com siguin necessàries per a connectar tots els equips de la proposta amb doble línia elèctrica.
- 9.4. Les PDUs aniran connectades al sistema de distribució elèctric amb entrada a l'armari de corrent. Actualment el CPD disposa de doble SAI amb línies monofàsiques de 16 i 32A i trifàsiques de 16 i 32A lliure en cada un d'ells. Caldrà especificar el nombre i el tipus de línies de corrent necessàries per la seva instal·lació.
- 9.5. Els cables de corrent de servidors i switchos han d'estar inclosos en l'oferta i s'han de connectar a aquestes PDUs, fent servir longituds adients però mínimes.

IV. CONDICIONS DEL LLIURAMENT

1. Configuració i validació de la xarxa, tant ethernet com infiniband
2. Creació, configuració i validació de rendiment dels filesystems, homes, scratch i projects tenint en compte l'actual sistema d'autenticació i autorització (LDAP).
3. Instal·lació, configuració i validació de rendiment els nodes de protocols, incloent la configuració de directoris a compartir.
4. Verificació de funcionament dels sistemes de seguretat establerts que han de garantir el funcionament en cas de caiguda d'un dels seus components

5. Formació: Versarà sobre el funcionament de l'eina, la seva configuració i les principals tasques de manteniment i gestió. Serà presencial en diverses sessions amb una durada estimada total de 15 hores i per a un nombre d'alumnes d'entorn a 8 tècnics informàtics de la Universitat. El material de la formació es lliurarà al final de les jornades.
6. Aportar la documentació tècnica de la instal·lació i tots els components, parametrització, instruccions de manteniment del fabricant, certificacions de l'equipament

V. GARANTIA

El termini de garantia serà de 5 anys. En aquest termini l'adjudicatari serà responsable de les faltes de conformitat que existeixin en el moment del lliurament i es manifestin dins del termini.

L'adjudicatari estarà obligat a prestar dins de la garantia les següents prestacions:

1. Treballs de manteniment que garanteixin el correcte funcionament de l'equipament durant el termini i considerats per mal funcionament que no sigui per un mal ús, incloent els recanvis, mà d'obra, desplaçaments, transport i dietes.
2. Suport usuari, amb eina de tíquet que permeti carregar incidències, consultes, veure l'històric de les incidències de manteniment i poder consultar les actualitzacions amb un temps màxim de resposta del següent dia laborable a l'obertura d'un tíquet.
3. Actualitzacions de programari.

La garantia començarà a comptar a partir de la data de posada en marxa de l'equipament, prèvia acceptació per escrit de la Universitat Pompeu Fabra.

L'empresa adjudicatària haurà de garantir la disponibilitat de peces de recanvi originals o equivalents (directament o a través d'altres agents autoritzats) per tot el cicle de vida previst de l'equip que, com a mínim serà de 5 anys més addicionals a la garantia.

El responsable d'Informàtica Científica del Servei Científic Tècnic.

Carles Perarnau Sabés

Barcelona, 23/9/2022

Annex I: Sistema actual de computació.

Resum descriptiu del maquinari actualment en producció del sistema HPC sistema d'emmagatzematge d'altres prestacions actualment en producció.

Marca	Model	#	CPU	Sistema Operatiu	Cor es	Memò ria	Ports Ether	Ports Infiniband	
Dell	PowerEdge R430	12	Xeon E5-2630 v3 @ 2.40GHz	CentOS 7.8.2003	16	128GB	4x1Gb BaseT	1xFDR	ConnectX-3 VPI adapter; single-port QSFP; FDR IB (56Gb/s) and 40GbE; PCIe3.0 x8 8GT/s
Dell	PowerEdge R430	13	Xeon E5-2683 v4 @ 2.10GHz	CentOS 7.8.2003	32	256GB	4x1Gb BaseT	1xFDR	ConnectX-3 VPI adapter; single-port QSFP; FDR IB (56Gb/s) and 40GbE; PCIe3.0 x8 8GT/s
Dell	PowerEdge R730	1	Xeon E5-2630 v3 @ 2.40GHz	CentOS 7.8.2003	16	128GB	4x1Gb BaseT	1xFDR	ConnectX-3 VPI adapter; single-port QSFP; FDR IB (56Gb/s) and 40GbE; PCIe3.0 x8 8GT/s
Dell	PowerEdge R830	1	Xeon E5-4640 v4 @ 2.10GHz	CentOS 7.8.2003	48	512GB	4x1Gb BaseT	1xFDR	ConnectX-3 VPI adapter; single-port QSFP; FDR IB (56Gb/s) and 40GbE; PCIe3.0 x8 8GT/s
Dell	PowerEdge R830	1	Xeon E5-4660 v4 @ 2.20GHz	CentOS 7.8.2003	64	512GB	4x1Gb BaseT	1xFDR	ConnectX-3 VPI adapter; single-port QSFP; FDR IB (56Gb/s) and 40GbE; PCIe3.0 x8 8GT/s
SIE	LADONoS 7	2	Xeon Gold 6248R	CentOS 7.8.2003	48	1,536GB	2x1Gb BaseT	1xEDR	ConnectX-5 VPI adapter card; EDR IB (100Gb/s) and 100GbE; single-port QSFP28; PCIe3.0 x16
IBM	x3850 X5	1	Xeon E7- 4850 @ 2.00GHz	CentOS 7.8.2003	20	256GB	2x1Gb BaseT	2xFDR	Mellanox ConnectX-3 VPI Dual-port QSFP QDR/FDR10/10GbE HCA

Annex II: Descripció del sistema actual

Resum descriptiu del sistema d'emmagatzematge d'altres prestacions actualment en producció.

Marca	Model	Discs	CPU/Disk Type	Cores	Disk brut (TB)	Spare Disks	Raid	Ports Ether	Ports Infiniband	SAS Interfaces
Scratch										
		Filesystem especialitzat en lectura/escriptura (R/W) amb una capacitat de 343TB NL-Sas +6,2TB SSD i metadades en SSD. Tiers dinàmic segons us, tamany dels fitxers i capacitat de desbordement al Cloud								
DELL	PowerEdge R630		E5-2630 v3	64				2x1Gb BaseT	1xFDR	4x12 Gbps SAS
DELL	PowerEdge R630		E5-2630 v3	64				2x1Gb BaseT	1xFDR	4x12 Gbps SAS
DELL	PowerVault MD3460	60x4TB	NL-SAS	-	240	5	Raid-6			4x12 Gbps SAS
DELL	PowerVault MD3060e	60x4TB	NL-SAS	-	240	5	Raid-6			
DELL	Dell Storage SCv2020	24x480GB	SSD	-	9,2	1	Raid-10			4 x 12Gb SAS
DELL	Dell Storage SC120	12x480GB	SSD	-	4,6	1	Raid 5-9			
Projects										
		Filesystem especialitzat en lectura (R) amb una capacitat de 550TB NL-Sas i compressió de fitxers								
DELL	PowerEdge R630		E5-2630 v4	128				2x1Gb BaseT	1xFDR	4x12 Gbps SAS
DELL	PowerEdge R630		E5-2630 v4	128				2x1Gb BaseT	1xFDR	4x12 Gbps SAS
DELL	PowerVault MD3460	60x10TB	NL-SAS	-	600	5	Raid-6			4x12 Gbps SAS
DELL	PowerVault MD3060e	60x4TB	NL-SAS	-	240	5	Raid-6			4x12 Gbps SAS
Protocols										



E-300700/03-06-22

		Serveix els continguts dels filesystem, o part d'ells, en protocols NFSv3, NFSv4, SMB, Objects,...								
IBM	x3650 M3		Xeon CPU X5650	48				2x1Gb BaseT	1xQDR	
IBM	x3650 M3		Xeon CPU X5650	48				2x1Gb BaseT	1xQDR	
Xarxa										
Mellanox	IB-2 SB7800 36 PORTS	17 ports lliures							36xEDR	
Cisco	Catalyst 4500	21 ports lliures							48x1Gb BaseT	
Programari										
1.098	IBM Spectrum Scale Data Management Edition Per Terabyte Annual SW Subscription & Support Renewal 12 Months									
	Versió actualment en producció: 5.1.1									