
Plec de prescripcions tècniques que regulen el subministrament, la implantació i la posada en marxa de l'ampliació del clúster d'alt rendiment Pirineus III (actuació que forma part de la inversió 2 “*Enfortiment de les capacitats, infraestructures i equipaments dels agents del SECTI*” del component 17 del PRTR)

Barcelona, a la data de la signatura

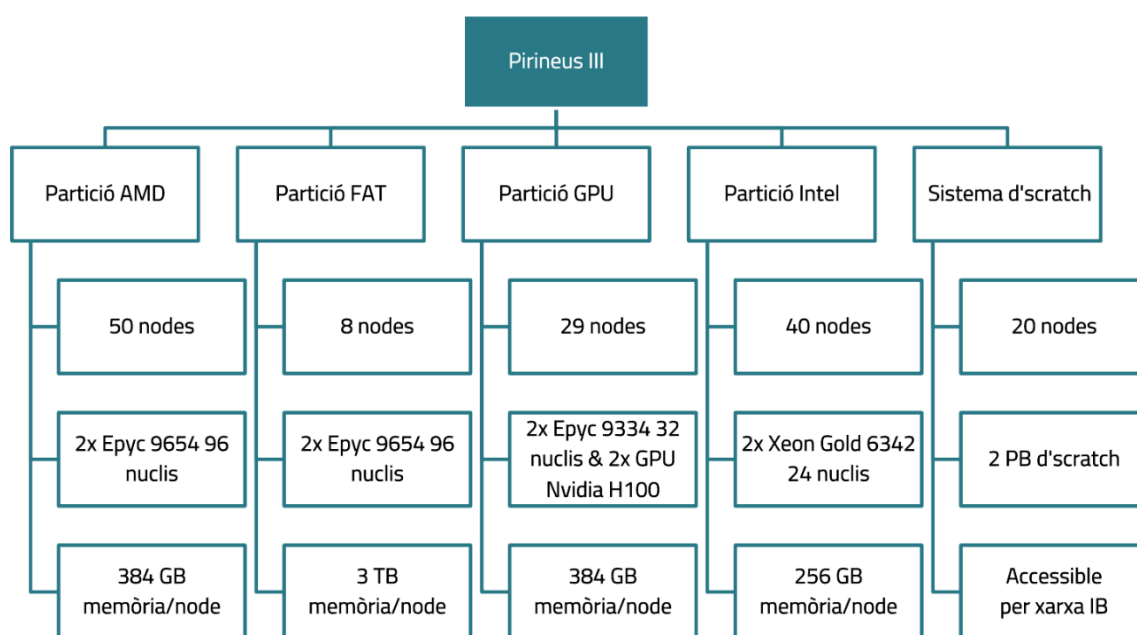
Validat per,	Aprovat per,
--------------	--------------

1.	Antecedents	5
2.	Objectiu i abast del projecte	6
3.	Característiques tècniques de l'equipament.....	6
	Lot 1: Clúster de nodes accelerats amb GPGPU orientat a HPC.....	6
	Lot 2: Clúster de nodes accelerats amb GPGPU orientat a IA	7
4.	Instal·lació i configuració dels equips	9
	4.1. Pla d'acceptació i de proves	10
5.	Documentació i formació	11
6.	Garantia de suport.....	12
7.	Oferiment de prestacions que optimitzen el rendiment dels equips.....	13
8.	Durada del projecte i terminis de lliurament.....	13
9.	Annex I.....	14

1. Antecedents

El Consorci de Serveis Universitaris de Catalunya (CSUC) té una llarga tradició en la prestació de serveis de computació d'altres prestacions a la comunitat de recerca. Aquest servei té com a objectiu millorar l'eficiència de tot el sistema permetent als grups de recerca tenir a la seva disposició un maquinari d'alt rendiment, gestionat per un equip de suport expert i que es manté actualitzat amb les últimes tecnologies disponibles al mercat.

L'any 2023 es va realitzar l'última renovació del maquinari de supercomputació del CSUC, que va entrar en funcionament el 2024. Aquesta renovació va suposar un salt molt important en la capacitat de l'entitat de prestar servei i una actualització dels recursos disponibles que permet, en part, donar resposta a les noves necessitats de la recerca, principalment en l'accés a maquinari orientat al seu ús en aplicacions d'IA, gràcies a l'aposta realitzada en l'adquisició d'una quantitat important de nodes equipats amb unitats de procés gràfic (GPGPU). El maquinari que es va adquirir es presenta de forma esquemàtica a la següent figura:



En el present procediment de licitació es pretén avançar en aquesta línia i incrementar encara més la quantitat de recursos orientats a noves aplicacions amb especial èmfasi en aplicacions pròpies de la computació d'altres prestacions accelerades amb GPGPU (p. ex. dinàmica molecular o problemes de ciència de materials) i d'IA. Per aquest motiu es planteja l'adquisició de maquinari optimitzat per aquests nous usos tal com es planteja a les següents seccions.

2. Objectiu i abast del projecte

Pels motius exposats a la secció anterior es planteja l'ampliació de Pirineus III a través de dues línies similars però clarament diferenciades: nodes orientats al seu ús en computació d'alt rendiment (HPC per les seves sigles en anglès) i nodes orientats al seu ús en aplicacions d'IA. Els detalls tècnics, de subministrament, implantació i posada en marxa del sistema es descriu als apartats següents d'aquest document.

El procediment de licitació consta de dos lots:

- El lot 1 planteja l'adquisició d'un clúster format per nodes de còmput accelerats amb GPGPU orientats a l'ús principal en aplicacions d'HPC, per aquest motiu aquests nodes hauran d'estar accelerats amb quatre unitats de procés gràfic per node.
- El lot 2 planteja l'adquisició d'un clúster format per nodes de còmput accelerats amb GPGPU orientats a l'ús per aplicacions d'IA, per aquest motiu aquests nodes hauran d'estar accelerats amb vuit unitats de procés gràfic per node.

3. Característiques tècniques de l'equipament

En aquest apartat es descriuen les característiques tècniques del maquinari que es pretén adquirir en els dos lots, els requisits mínims demanats, les condicions d'instal·lació i configuració dels equips, els requisits pel pla de proves i acceptació, els continguts mínims per al pla de formació i les condicions de la garantia.

Lot 1: Clúster de nodes accelerats amb GPGPU orientat a HPC

El maquinari ofert en aquest lot ha de reunir les següents característiques tècniques:

- Rendiment punta global mínim de tot el clúster de 1200 Tflop/s en FP64.
- Els nodes han de tenir dos processadors.
- Els nodes han de tenir un mínim de 64 nuclis.
- Han d'estar basats en processadors d'arquitectura x86 de 64 bits, de la darrera generació disponible del fabricant.
- Els processadors han de tenir una freqüència de, com a mínim, 2.4 GHz.
- Cada node ha de comptar amb quatre GPGPU.

Pel que fa a les GPGPU han de reunir les següents característiques tècniques:

- Compatibles amb *CUDA* i *OpenACC*.
- Han de tenir un rendiment en operacions de coma flotant de 64 bits de, com a mínim, 30 Tflop/s.
- Han de tenir una memòria mínima de 140 GB.

Requisits per a la memòria principal per node:

- Ha de ser com a mínim de 1 TB.
- La memòria haurà de ser, com a mínim, de tipus DDR5.
- El sistema de memòria s'haurà de configurar perquè proporcioni el màxim rendiment, és a dir, tots els bancs de memòria hauran d'estar ocupats. En aquest sentit el licitador haurà de descriure en detall el subsistema de memòria, les seves possibles configuracions, i les latències i amplada de banda a memòria.

Amb relació a les xarxes es requerirà que cada node disposi de quatre interfícies de xarxa independents: la de càlcul, la de dades, la d'accés a serveis i la d'administració dels mateixos nodes i accés a la gestió via processadors de servei fora de banda. Els requisits mínims per cadascuna de les xarxes són els següents:

- La xarxa de càlcul haurà de ser *Infiniband* amb un amplada de banda mínim de 200 Gbps. La configuració de la xarxa haurà de ser tal que asseguri que l'amplada de banda mínima entre dos nodes qualsevol sigui de 100 Gbps (incloent-hi el clúster de *Beegfs*).
- La xarxa de dades haurà de ser de tipus *Gigabit Ethernet* a 25 Gbps.
- La xarxa d'accés als serveis i la d'administració dels nodes hauran de ser de tipus *Gigabit Ethernet* amb un mínim de 25 Gbps.
- L'oferta haurà d'incloure els commutadors i tots els elements de xarxa necessaris per poder integrar el sistema a l'arquitectura de xarxa del CSUC actual detallada en l'annex 1 garantint la seva compatibilitat i redundància.

Amb relació al programari a incloure es requereix que cada node tingui una llicència de RHEL.

Es requereix que el bastidor o bastidors on s'instal·lin aquests nodes estiguin refrigerats per aigua mitjançant un sistema de portes fredes. Actualment, el CSUC disposa d'un bastidor estàndard de 19 polzades i 42U (200 cm x 60 cm x 120 cm) on s'hauran d'instal·lar una part d'aquests nodes. Aquest bastidor està equipat amb una porta *Vertiv Liebert DCD* amb una capacitat de dissipació de 35 kW. Si es requereix la provisió d'un (o més) bastidors addicionals aquests hauran d'estar equipats amb portes compatibles amb la instal·lació del sistema de refrigeració actual del mateix model *Vertiv Liebert DCD* però amb una capacitat de dissipació de 50 kW.

Lot 2: Clúster de nodes accelerats amb GPGPU orientat a IA

El maquinari ofert en aquest lot ha de reunir les següents característiques tècniques:

- Rendiment punta global mínim de tot el clúster de 1200 Tflop/s en FP32.
- Els nodes han de tenir dos processadors.
- Els nodes han de tenir un mínim de 48 nuclis.

- Han d'estar basats en processadors d'arquitectura x86 de 64 bits, de la darrera generació disponible del fabricant.
- Els processadors han de tenir una freqüència de, com a mínim, 2.0 GHz.
- Els nodes han de comptar amb vuit GPGPU.

Pel que fa a les GPU han de reunir les següents característiques tècniques:

- Compatibles amb *CUDA* i *OpenACC*.
- Han de tenir un rendiment en operacions de coma flotant de 32 bits de, com a mínim, 80 Tflop/s.
- Han de tenir una memòria mínima de 190 GB.

Requisits per a la memòria principal per node:

- Ha de ser com a mínim de 2 TB.
- La memòria haurà de ser, com a mínim, de tipus DDR5.
- El sistema de memòria s'haurà de configurar perquè proporcioni el màxim rendiment, és a dir, tots els bancs de memòria hauran d'estar ocupats. En aquest sentit el licitador haurà de descriure en detall el subsistema de memòria, les seves possibles configuracions, i les latències i amplada de banda a memòria.

Es requerirà que els nodes estiguin equipats amb disc local amb un mínim de 10 TB de tipus NVMe de tecnologia *enterprise*.

En relació a les xarxes es requerirà que cada node disposi de quatre interfícies de xarxa independents: la de càlcul, la de dades, la d'accés a serveis i la d'administració dels propis nodes i accés a la gestió via processadors de servei fora de banda. Els requisits mínims per cadascuna de les xarxes són els següents:

- La xarxa de càlcul haurà de ser *Infiniband* amb una amplada de banda mínim de 200 Gbps. La configuració de la xarxa haurà de ser tal que asseguri que l'amplada de banda mínima entre dos nodes qualsevol sigui de 100 Gbps (incloent-hi el clúster de *Beegfs*).
- La xarxa de dades haurà de ser de tipus *Gigabit Ethernet* a 25 Gbps.
- La xarxa d'accés als serveis i la d'administració dels nodes hauran de ser de tipus *Gigabit Ethernet* amb un mínim de 25 Gbps.
- L'oferta haurà d'incloure els commutadors i tots els elements de xarxa necessaris per poder integrar el sistema a l'arquitectura de xarxa del CSUC actual detallada en l'annex 1 garantint la seva compatibilitat i redundància.

Amb relació al programari a incloure es requereix que cada node tingui una llicència de RHEL. Es requerirà també per aquest lot que s'incloguin llicències d'*Nvidia AI Enterprise* vàlides per cinc anys per totes les GPGPU.

Es requereix que el bastidor o bastidors on s'instal·lin aquests nodes estiguin refrigerats per aigua mitjançant un sistema de portes fredes compatibles amb la instal·lació del sistema de

refrigeració actual del mateix model *Vertiv Liebert DCD* però amb una capacitat de dissipació de 50 kW.

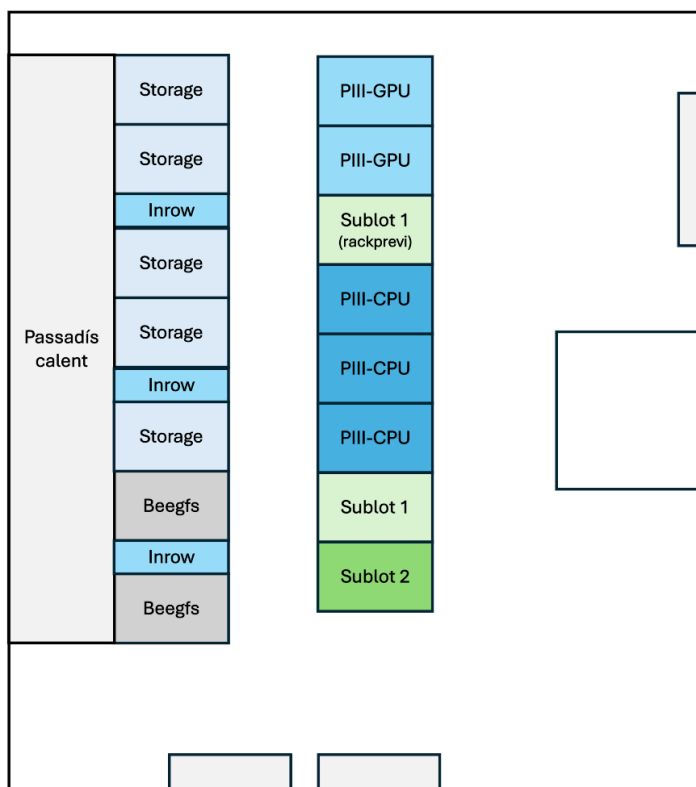
4. Instal·lació i configuració dels equips

Sempre que sigui possible, s'espera que l'equip sigui muntat a fàbrica de manera que un cop arribi a les instal·lacions del CSUC els diferents components del maquinari hagin superat els tests de maquinari bàsics. En cas que això no sigui possible, els elements bàsics de l'equip s'hauran de muntar en fàbrica, a l'hora que es podrà incloure la instal·lació *in situ* de certs elements.

Un cop s'entreguin els equips, l'adjudicatari haurà de proporcionar l'assistència tècnica necessària per a la instal·lació i posada en marxa del sistema, per a la configuració i optimització i per a la integració dels equips a Pirineus III (sistema de gestió del clúster, integració amb els sistemes d'emmagatzematge, serveis, etc). Per últim, haurà de participar en la realització del pla d'acceptació que es descriu a la secció següent.

El sistema ofert haurà d'ubicar-se a la sala 2 del centre de procés de dades (CPD) del CSUC. El maquinari dels lots 1 i 2 haurà d'incloure per a la seva instal·lació els bastidors que siguin necessaris. Aquests hauran de ser de mida estàndard de 19 polzades i 42 U (200 cm x 80 cm x 120 cm).

A continuació s'inclou un esquema de la sala on s'ubicaran els nous equipaments especificant la seva ubicació:



Els licitadors hauran de tenir en compte la ubicació dels bastidors que allotjaran el maquinari del lot 1 de cara a dimensionar adequadament tot el cablejat requerit. La ubicació dels nodes es realitzarà des de la part inferior del bastidor (U1) cap a la part superior (U42). La part que no estigui amb equipament de l'armari haurà d'estar tapat per la part frontal amb panells cecs. En relació amb el cablejat es requerirà que la longitud de tots aquests elements sigui la suficient per a poder emplenar els bastidors des de la part més baixa dels mateixos.

Caldrà tenir en compte que els equipaments estaran instal·lats sobre un terra tècnic de les següents característiques amb un pes màxim suportat 2.000 kg/m² i 500 kg de càrrega puntual.

Pel que fa a l'equipament de xarxa, les connexions de xarxa es realitzaran per la part posterior del bastidor, considerant que la part frontal de l'equipament (la que es veurà des de la porta) serà la que disposa de les connexions elèctriques, i la part posterior (la que es veurà des de la porta freda) la que disposa de les connexions de xarxa. El flux d'aire d'aquests equipaments serà des de porta el bastidor, cap a la porta freda del mateix.

Pels dos bastidors nous s'haurà d'ampliar la rejiband que es troba en el sostre (amplada 150 mm).

L'adjudicatari haurà de proporcionar tots els components i material necessari per a la instal·lació i connexió del sistema: PDU trifàsiques de 32A *Vertiv VP8886* intel·ligents que permetin la mesura del consum elèctric (que hauran d'estar redundades), juntament amb dos sensors (temperatura i humitat) per cada porta (frontal i posterior) del bastidor (en total, 4 sensors per bastidor), cablejat fins als quadres elèctrics i les mesures de protecció dimensionades per aquestes PDU (PIA + diferencial), etc. La instal·lació i configuració de les PDU haurà d'estar finalitzada abans de la connexió elèctrica dels nodes per a equilibrar la intensitat de les fases.

Un cop els equips es van instal·lant en els bastidors (no en funcionament) l'adjudicatari haurà de retirar tot l'embalatge associat a l'equipament instal·lat de les instal·lacions del CSUC aquell mateix dia.

L'equipament ha d'estar instal·lat i en producció com a màxim el 15 de juny de 2026.

4.1. Pla d'acceptació i de proves

Un cop finalitzada la instal·lació i configuració dels equips al CSUC i per a garantir que el sistema que s'entrega funciona de forma correcta i òptima, l'adjudicatari haurà de participar en la realització del pla d'acceptació.

Perquè el sistema sigui acceptat, caldrà que el licitador executi els següents **tests d'operació** orientats a demostrar que el sistema funciona correctament en l'àmbit d'operació bàsica dels equips i que la configuració i la integració amb l'entorn de treball és la correcta. El llistat final dels tests a realitzar es concretarà amb l'adjudicatari en la fase de planificació del projecte, però aquests inclouran com a mínim el següent:

- Verificació de la seqüència d'aturada i arrencada, incloent-hi aturada elèctrica, demostrant que el sistema arrenca amb tots els serveis i funcionalitats necessàries.
- Demostració del correcte funcionament de l'administració remota incloent la creació i desplegament d'imatges.
- Verificació de la correcta integració del sistema amb el servidor d'autenticació d'usuaris *OpenLDAP*, amb els servidors de fitxers de l'entorn *NAS* i *BeeGFS*, amb el gestor de cues *Slurm* instal·lat a Pirineus III i amb la resta de serveis del clúster.

5. Documentació i formació

L'adjudicatari haurà de proporcionar la documentació i un pla de formació adient per al personal del CSUC que inclourà com a mínim els següents conceptes:

- Manual d'administració dels equips, a lliurar en format digital, que serveixi al personal del CSUC com a guia per exercir les tasques necessàries de configuració, explotació i suport posterior. Aquest manual contindrà almenys la següent informació:
 - Descripció general del sistema i dels seus components i arquitectura.
 - Esquemes gràfics detallats de la distribució dels diferents components i la seva interconnexió.
 - Descripció dels principals paràmetres utilitzats per a la configuració del sistema (paràmetres del *kernel*, configuració de xarxes, etc...).
 - Descripció dels principals procediments bàsics per a l'administració i explotació del sistema: aturada i posada en marxa de tots els components, incloent-hi l'aturada elèctrica; creació de fitxers per al depurat en cas de caiguda del sistema, etc.
- Pla de formació per als tècnics del Centre sobre el funcionament i configuració dels equips, tant del maquinari com del programari, i que inclourà com a mínim un curs d'operació en el que caldrà incloure com a mínim els següents aspectes:
 - Introducció a les característiques generals del sistema: arquitectura, tecnologia dels diferents components, etc.
 - Eines d'administració i gestió.
 - Principals procediments d'operació.
 - Monitoratge del sistema i optimització.
 - Seguretat dels equips.
 - Ús en HPC (configuració gestor de cues, compilació i execució d'aplicacions)
- Es requerirà que el licitador ofereixi formació, tant als tècnics del CSUC com als usuaris del centre interessats en l'ús de les GPGPU. Aquesta formació haurà d'estar centrada en l'ús d'aplicacions i en com treure el màxim profit de l'ús d'aquest tipus d'acceleradors en ell. Aquesta formació ha de comptar amb algun tipus de certificació o reconeixement per part del fabricant de les GPGPU.

6. Garantia de suport

Els equips oferts hauran d'incloure un període mínim de **garantia de suport de cinc anys** per a tot el maquinari i programari subministrat.

El nivell de servei no podrà ser inferior al següent:

- Temps de resposta màxim de 4 hores després de la comunicació de la incidència en horari laboral.
- Temps de resolució màxim de 12 hores i reposició de material el següent dia laborable¹, incloent peces, mà d'obra i desplaçament del personal tècnic. En cas que la incidència sigui per fallida de maquinari, no es considerarà el temps que l'adjudicatari necessiti per disposar de la peça o peces avariades per a la seva reposició.

L'incompliment d'aquests terminis es penalitzarà segons allò disposat a l'apartat O del Quadre de característiques del Plec de Clàusules Administratives Particulars.

Entenem com a temps d'atenció el temps des de que es notifica una incidència fins l'inici de les tasques necessàries per a la seva resolució i temps de resolució com el temps invertit en la seva resolució des del moment que s'atén fins que realment queda resolta. Tanmateix, es considera una incidència crítica si afecta a més del 50% de la disponibilitat del servei i greu si l'afectació de la disponibilitat és menor.

Adicionalment, també es requereix:

- Cobertura horària 9x5, de 9h a 18h de dilluns a divendres excepte dies festius.
- Suport telefònic davant incidències i consultes amb resposta immediata.
- La inclusió d'un sistema de seguiment d'incidències i consultes. Si el licitador no disposa d'aquesta eina es farà servir una proporcionada pel CSUC.
- Suport correctiu: totes aquelles prestacions necessàries per mantenir els equips objecte d'aquest plec en uns nivells de disponibilitat i funcionament òptims, així com l'ajustament dels elements que permetin mantenir aquests nivells. Inclou el cost de totes les reparacions i operacions de manteniment necessaris, recanvis i/o reposicions.
- Actualització als últims nivells de sistema operatiu, *firmware*, millores d'enginyeria, etc., que es realitzarà d'acord a la planificació establerta amb el personal tècnic del CSUC.

Els licitadors hauran d'especificar en la seva oferta els perfils dels tècnics que portaran a terme el servei de suport, i la seva experiència prèvia en instal·lacions i serveis de suport de característiques similars a les de l'objecte del present plec.

¹ En totes les referències a "dia laborable", es considera el calendari laboral de la ciutat de Barcelona.

En cas de les incidències crítiques, el CSUC podrà demanar a l'adjudicatari informes puntuals que n'expliquin les causes, la resolució i les mesures correctives preses. Aquests informes hauran d'ésser lliurats pel proveïdor al CSUC en com a màxim 3 dies laborables.

El licitador haurà de descriure en detall els tipus de suport disponibles i les condicions de contractació de cadascun d'ells un cop finalitzat el període de garantia, així com les seves condicions contractuals, que hauran de quedar detallades a l'oferta econòmica del sobre C. Concretament, caldrà especificar la política del licitador respecte a avaries i reparacions de l'equipament, la jornada laboral per a incidències i el temps de resposta previst de resolució.

Cal fer notar que durant el període de garantia, tots els costos de manteniment han d'ésser inclosos en el preu ofert per l'equipament.

7. Oferiment de prestacions que optimitzen el rendiment dels equips

Els licitadors podran oferir prestacions que optimitzen el rendiment dels equips i que caldrà incorporar al sobre que correspongui en cada cas. Aquestes són

Pel lot 1:

- Nombre total de nodes del sistema ofert. (Sobre B)
- Nombre màxim de nuclis per processador. (Sobre B)
- Freqüència dels processadors (sense mode turbo activat). (Sobre B)
- Memòria principal (sense tenir en compte la de les GPU) dels nodes oferts. (Sobre B)

Pel lot 2:

- Nombre total de nodes del sistema ofert. (Sobre B)
- Nombre màxim de nuclis per processador. (Sobre B)
- Freqüència dels processadors (sense mode turbo activat). (Sobre B)
- Memòria principal (sense tenir en compte la de les GPU) dels nodes oferts. (Sobre B)

Les ofertes rebudes pels dos lots es valoraran de manera conjunta i de forma ponderada tal com s'especifica al Plec de Clàusules Administratives Particulars.

8. Durada del projecte i terminis de lliurament

Els licitadors presentaran una proposta de cronograma que distingirà clarament cadascuna de les fases previstes per a l'execució del projecte pel que fa al lliurament i posada en marxa de tot el maquinari i el programari, que inclourà com a mínim les següents fites:

- Revisió de les instal·lacions tècniques.
- Lliurament dels equips.
- Instal·lació i configuració del sistema.
- Pla d'acceptació i proves de l'equipament.
- Pla de formació

El nou equipament haurà d'estar plenament operatiu i en explotació en un termini màxim de 6 mesos començant a comptar des de la data de la signatura del contracte.

9. Annex I

Equipament de xarxa actual de l'entorn de supercomputació:

- Xarxa de càlcul (*Infiniband*) - Mellanox QM8790
- Xarxa de dades (25 Gbps) - Huawei CE16804
- Xarxa d'accés als serveis i la d'administració (25 Gbps) - Huawei CE16804
- Xarxa IPMI o gestió - FS S5800-48TS4S-PE