

TEHNISKĀ SPECIFIKĀCIJA

HPC mākoņskaitļošanas resursu noma pētniecības projekta ietvaros

Pretendentam tehniskais piedāvājums iepirkumā "HPC mākoņskaitļošanas resursu noma pētniecības projekta ietvaros" jāsagatavo ievērojot nolikuma un tehniskās specifikācijas prasības, nodrošinot visaptverošu un detalizētu informāciju par piedāvāto hiperkonverģēta HPC mākoņskaitļošanas pakalpojumu risinājumu. Tehniskajā piedāvājumā nepieciešams iekļaut izsmēlošu un precīzu aprakstu par visu tehnisko, organizatorisko un funkcionālo prasību izpildi, nodrošināšanu un realizāciju attiecībā uz HPC mākoņskaitļošanas infrastruktūru un ar to saistītajiem pakalpojumiem.

Tehniskā piedāvājuma sagatavošanas procesā pretendents jādemonstrē dziļa izpratne par augstas veiktspējas skaitļošanas specifiku, hiperkonverģēto infrastruktūru, mākoņa pakalpojumu arhitektūras principiem un mākslīgā intelekta risinājumu tehniskajām prasībām.

Tehniskajā piedāvājumā obligāti jāiekļauj visaptverošs un strukturēts apraksts par:

- HPC mākoņskaitļošanas mākoņa infrastruktūras pakalpojuma tehnisko arhitektūru un funkcionalitāti, ietverot detalizētu informāciju par resursu konteinerizāciju, programdefinētu infrastruktūru, automatizētu resursu pārvaldību un elastīgu mērogojamību. Pretendentam jādemonstrē, kā hiperkonverģentā arhitektūra nodrošina optimālu resursu izmantošanu un pārvaldības efektivitāti.
- Pakalpojumu pārvaldības procesu īstenošanas metodiku un organizatorisko struktūru, kas nodrošinās efektīvu un nepārtrauktu hiperkonverģentās sistēmas darbību. Šajā sadaļā nepieciešams aprakstīt automatisko orķestrāciju, resursu optimizāciju, slodzes balansēšanu un dinamisko resursu piešķiršanu.
- Hiperkonverģentās infrastruktūras konfigurācijas, aktivizēšanas un sākotnējās iestatīšanas procesus, ietverot programdefinēto tīklu konfigurāciju, izplatīto glabāšanu un konteinerizācijas slāņu iestatīšanu. Pretendentam jāsniedz detalizēts pakalpojuma uzsākšanas plāns ar konkrētiem etapiem.
- Tehniskā atbalsta organizāciju, kas specializējas hiperkonverģentajās tehnoloģijās, konteinerizācijā un mākoņa pakalpojumu pārvaldībā, ietverot detalizētu informāciju par ekspertīzes līmeni un problēmu risināšanas spējām.

A. HPC mākoņskaitļošanas infrastruktūras resursu nomas pakalpojuma apraksts

1. Hiperkonverģentās arhitektūras funkcionālās prasības

Visiem piedāvātajiem mākoņa resursiem obligāti jānodrošina hiperkonverģēta funkcionalitāte, kas integrē aprēķinu, glabāšanas un tīkla resursus vienotā, programdefinētā platformā.

Hiperkonverģentajai arhitektūrai jānodrošina vienkāršota pārvaldība, automatizēta resursu optimizācija un elastīga mērogojamība atbilstoši mainīgajām darba slodžu prasībām.

Resursu konteinerizācijai jānodrošina abstrakcija no fiziskās aparatūras un citiem operētājsistēmas procesiem, ļaujot dinamiski piešķirt un pārpiešķirt resursus starp dažādiem konteineriem bez

fiziskās infrastruktūras pārkonfigurēšanas. Konteinerizācijas slānim jāatbalsta dažādas izolācijas tehnoloģijas, ieskaitot hiperuzraudzības tehnoloģijas (Kata containers vai citus nozares standartus). Programmdefinētajai infrastruktūrai jānodrošina automatizētu resursu pārvaldību caur programmatūras saskarnēm un orķestrācijas rīkiem, ļaujot infrastruktūru pārvaldīt kā kodu. Tas ietver programmdefinētus aprēķinus, programmdefinētu glabāšanu un programmdefinētus tīklus.

Hiperkonverģentajai platformai jāspēj nodrošināt horizontālu mērogojamību, pievienojot jaunus virtuālos resursus pēc pieprasījuma bez pakalpojuma pārtraukšanas. Mērogojamības mehānismiem jādarbojas automātiski, balstoties uz iepriekš definētām politikām un resursu izmantošanas rādītājiem.

Papildus piedāvātajai hiperkonverģentajai HPC mākoņskaitļošanas platformai jānodrošina savietojamība un droša integrācija ar trešo pušu risinājumiem, izmantojot standartizētas API un drošus tīkla protokolus, lai īstenotu hibrīdmākoņa scenārijus māšīnmācīšanās slodžu migrāciju un darībnāšanu publiskajos mākoņos, kā arī liela apjoma datu kopu datu apmaiņu un pakalpojumu nepārtrauktību.

2. HPC skaitļošanas resursu pakalpojuma specifikācijas

HPC mākoņskaitļošanas infrastruktūrai jānodrošina augstas veiktspējas skaitļošanas resursi, kas optimizēti gan tradicionāliem augstas veiktspējas aprēķiniem, gan mūsdienīgiem mākslīgā intelekta un mašīnmācības algoritmiem. Pakalpojumam jāgarantē kopumā vismaz 224 virtuālie centrālā procesora kodoli ar augstas veiktspējas skaitļošanas optimizāciju.

Centrālā procesora resursu pakalpojuma konfigurācijai jābalstās uz mūsdienīgiem x86-64 procesoriem ar atbalstu uzlabotām instrukciju kopām, vektorizācijas iespējām (AVX2, AVX512 u.c.) un aparatūras paātrinātu šifrēšanu. Procesoru kodoliem jānodrošina konsistenta veiktspēja ar garantētiem veiktspējas līmeņiem, minimizējot procesoru veiktspējas svārstības apstrādes laikā.

Operatīvās atmiņas pakalpojumam jānodrošina kopumā vismaz 1536 GB atmiņas ar kļūdu korekcijas atbalstu un optimizētu atmiņas joslas platumu liela atmiņas patēriņai programmaprogrammām. Atmiņas piešķiršanai jābūt elastīgai, ļaujot dinamiski pielāgot atmiņas piešķiršanu atbilstoši darba slodžu prasībām.

Grafikas procesoru aprēķinu pakalpojumam jānodrošina:

(1) Vismaz divi augstas veiktspējas MI akseleratori ar šādām minimālajām prasībām:

- vismaz 141 GB HBM3e vai ekvivalenta augstas joslas platuma atmiņa,
- NVLink / SXM vai ekvivalents augstas veiktspējas GPU–GPU savienojums,
- Tensor Core atbalsts FP8, FP16, BF16, TF32,
- kopējā AI veiktspēja ne zemāka kā NVIDIA H200 vai ekvivalents.

(Piemērs: NVIDIA H200 SXM vai ekvivalents.)

(2) Vismaz divi profesionālie GPU ar šādām minimālajām prasībām:

- vismaz 96 GB GDDR7 atmiņas,
- Tensor Core paātrinājums FP8/FP16/TF32,
- kopējā FP32/TF32/Tensor Core veiktspēja ne mazāka kā NVIDIA RTX 6000 PRO Blackwell vai ekvivalents,
- GPU jābūt ne vecākam par Blackwell paaudzi.

(Piemērs: NVIDIA RTX 6000 PRO Blackwell vai ekvivalents.)

(3) GPU resursiem jābūt pilnībā pieejamiem Kubernetes konteineriem, izmantojot NVIDIA GPU Operator vai ekvivalentu risinājumu, ar minimālu konteinerizācijas slodzi.

Grafikas procesoru resursi jānodrošina ar pilnu aprēķinu spēju un garantētu veikspējas līmeni visā līguma darbības laikā. Grafikas procesoru resursi jānodrošina ar minimālu konteinerizācijas papildu slodzi, garantējot konsistentu veikspēju mākslīgā intelekta un mašīnmācības darba slodzēm

3. Programmdefinētās glabāšanas pakalpojuma prasības

HPC mākoņskaitļošanas infrastruktūrai jāintegrē programmadefinēta glabāšanas sistēma, kas nodrošina elastīgu, mērogojamu un augstas veikspējas datu glabāšanas pakalpojumus ar automātisku redundanci un katastrofu atgūšanas spējām.

Izplatītās glabāšanas arhitektūrai jānodrošina vismaz 840TB glabāšanas kapacitātes ar iespēju programmatūras līmenī definēt redundances līmeni un algoritmus, kā arī datu aizsardzību. Glabāšanas sistēmai jāatbalsta dažādi glabāšanas līmeņi - augstas veikspējas SSD glabāšana MI modeļu ātrajiem aprēķiniem uz tā paša hiperkonverģētā mezglā ("local NVME storage"), SSD augstas veikspējas glabāšana ātrai piekļuvei datiem un ietilpības glabāšana liela apjoma datu kopu glabāšanai.

Augstas veikspējas SSD glabāšanas slānim jānodrošina vismaz 120 TB kapacitātes, izmantojot datu centra klases NVMe SSD diskus (individuālu disku veikspēja nedrīkst būt mazāka kā 14GB/sek. lasīšanai un ne mazāk kā 7GB/sek. rakstīšanai, t.i. jānodrošina jaunākās paaudzes PCI express NVME disk). Glabāšanai jābūt piemērotai intensīviem aprēķiniem un datu apstrādes uzdevumiem ar zemu latentumu.

Objektu glabāšanas funkcionalitātei jānodrošina S3 saderīga programmatūras saskarne ar nepārtrauktu integrāciju ar publisko mākoņu objektu glabāšanas servisiem, lai nodrošinātu liela apjoma MI datu kopu datu apmaiņu ar publiskajiem mākoņiem kā arī optimizētu datu glabāšanas izmaksas.

Bloku glabāšanas pakalpojumam jānodrošina augstas veikspējas pastāvīgie sējumi virtuālajām mašīnām un konteineriem ar momentuzņēmumu, klonēšanas un dublēšanas iespējām. Bloku un objektu glabāšana jāintegrē ar konteineru orķestrācijas platformām tādām kā Kubernetes.

Visiem datu glabāšanas pakalpojumiem jābūt šifrētiem (encryption at rest), lai nodrošinātu pētniecībā izmantotās informācijas konfidencialitāti. Pakalpojuma sniedzējs uzņemas atbildību par tā platformā uzglabāto datu (t.sk. medicīniskās informācijas un attēlu datu) drošību un nepieejamību trešajām pusēm. Datu noplūdes incidenta gadījumā, ja tas noticis Pakalpojuma sniedzēja vainas dēļ (piemēram, neievērojot TS noteiktās prasības risinājumam un industrijas labo praksi drošības pārvaldībai), Pakalpojuma sniedzējs atmaksā Pasūtītājam bezstrīdus kārtībā kompensāciju ne mazāku kā nomas maksa par visu līguma darbības periodu.

4. Programmdefinēto tīkla pakalpojumu prasības

HPC mākoņskaitļošanas infrastruktūras hiperkonverģentajai platformai jāintegrē programmadefinēts tīkls, kas nodrošina elastīgu, mērogojamu un drošu tīkla infrastruktūru ar centralizētu pārvaldību un automātisku konfigurāciju.

Virtuālo tīklu pakalpojumam jānodrošina vismaz 200 Gbps efektīva joslas platuma starp resursiem (t.s.k aprēķinu un datu glabāšanas) ar maza kavējuma savienojamību ($< 10 \mu s$) augstas veikspējas skaitļošanas lietojumprogrammām. Tīkla arhitektūrai jāatbalsta mikrosegmentācija, tīkla izolācija un daudznomnieku režīms ar drošu nomnieku atdalīšanu.

Slodzes balansēšanas un datu plūsmas pārvaldības pakalpojumiem jānodrošina automātiska slodzes sadalīšana starp resursiem, pārslēgšanās spējas un pakalpojumu kvalitātes datplūsmas prioritizācija. Slodzes balansētājs jāintegrē ar automātiskās mērogošanas mehānismiem dinamiskajam kapacitātes pielāgojumam. Ārējās savienojamības pakalpojumam jānodrošina vismaz 1 Gbps simetriska interneta savienojamība ar augstu pieejamību un vairāku augšupējo pakalpojuma sniedzēju redundanci.

Jānodrošina iespēja konfigurēt šifrētus VPN savienojumus individuāliem pētniekiem, jābūt iespējai nodrošināt vismaz 25Gbit/s izdalītās līnijas (liela apjoma datu apmaiņai ar augstas veiktspējas histopatoloģisko mikropreparātu skeneri, kurš atradīsies laboratorijā_ Pilsoņu iela 13, Zemgales priekšpilsēta, Rīga, LV-1002 un privātu savienojamību ar Pasūtītāja biroju Balasta dambis 80A, Rīga, LV-1048 ar minimālo simetrisko datu pārraides ātrumu 1Gbit/s un Arhitektu iela 1, Rīga, LV-1050 ar minimālo simetrisko datu pārraides ātrumu 25Gbit/s.

B. HPC mākoņskaitļošanas PLATFORMAS PĀRVALDĪBAS UN ATBALSTA PAKALPOJUMI

5. Mākoņa orķestrācijas un pārvaldības platforma

HPC mākoņskaitļošanas infrastruktūras hiperkonverģentās platformas pārvaldībai jānodrošina mūsdienīga un intuitīva mākoņa orķestrācijas platforma, kas ļauj efektīvi pārvaldīt visus infrastruktūras komponentus caur vienotu lietotāja saskarni. Pārvaldības platformai jābūt balstītai uz atvērtā koda risinājumiem mākoņa orķestrācijas risinājumu.

Lietotāja saskarnei jānodrošina visaptveroša redzamība visiem infrastruktūras komponentiem ar reāllaika uzraudzību, resursu izmantošanas rādītājiem, veiktspējas analītiku un kapacitātes plānošanas rīkiem. Informācijas panelim jābūt atsaucīgam un pieejamam no dažādām ierīcēm un platformām.

Programmatūras saskarņu vadītai pārvaldībai jānodrošina pilnīgu programmatisku kontroli pār visiem infrastruktūras aspektiem, ļaujot īstenot infrastruktūras kā koda prakses ar rīkiem kā Terraform, Ansible vai Chef. Programmatūras saskarnei jābūt REST principu ievērojošai, labi dokumentētai un atpakaļ saderīgai.

Daudznomnieku pakalpojumiem jānodrošina droša resursu izolācija starp dažādiem nomniekiem, ar detalizētu piekļuves kontroli, resursu kvotām un norēķinu atdalīšanu. Nomnieku izolācija jānodrošina visos infrastruktūras līmeņos - aprēķinos, glabāšanā, tīklā un drošībā.

Lomām balstītai piekļuves kontrolei jānodrošina detalizēta atļauju pārvaldība ar integrāciju uzņēmuma identitāšu sistēmām kā Active Directory, LDAP, SAML vai OAuth2.0 identitāšu nodrošinātājiem. Vienreizējās pieteikšanās spēja ir obligāta prasība.

6. Automatizācijas un orķestrācijas pakalpojumi

HPC mākoņskaitļošanas infrastruktūras hiperkonverģentajai platformai jānodrošina uzlabota automatizācijas spējas, kas minimizē manuālu administrēšanu un nodrošina konsistentu, atkārtojamu infrastruktūras darbību ar automatizētu nodrošināšanu, mērogošanu un dzīves cikla pārvaldību.

Automātiskās mērogošanas pakalpojumiem jānodrošina intelegenta resursu mērogošana, balstoties uz reāllaika rādītājiem, paredzošo analītiku un lietotāja definētām politikām.

Mērogošanas operācijas jāveic bez pakalpojuma pārtraukšanas, izmantojot ritošus atjauninājumus un zils-zaļus (blue-green deployment) izvietošanas stratēģijas. Infrastruktūras dzīves ciklapārvaldība jāautomatizē, ietverot automātisku labošanu, programmatūras atjauninājumus, konfigurācijas novirzīšanas noteikšanu un novēršanu. Dzīves cikla politikas jākonfigurē ar uzturēšanas logiem, atpakaļgaitas procedūrām un izmaiņu apstiprinājuma darba plūsmām.

Darba slodžu izvietošanas optimizācijai jānodrošina intelektīti plānošanas algoritmi, kas ņem vērā resursu saistību, veikspējas prasības un izmaksu optimizāciju. Izvietošanas lēmumi jāpieņem reāllaikā, balstoties uz pašreizējo sistēmas stāvokli.

7. Konteinerizācijas un mikroservisu atbalsta pakalpojumi

HPC mākoņskaitļošanas infrastruktūras hiperkonverģentajai platformai jānodrošina pilnīgs atbalsts mūsdienīgām konteinerizācijas tehnoloģijām ar pārvaldītu Kubernetes pakalpojumu, konteineru reģistru, CI/CD kanālu integrāciju un servisu tīkla spējām. Pārvaldītajam Kubernetes pakalpojumam jānodrošina ražošanai gatavs Kubernetes klāsteris ar automatizētu klāstera pārvaldību, mezglu automātisku mērogošanu, drošības pastiprināšanu un integrētu uzraudzību. Kubernetes versiju pārvaldība un atjaunināšanas automatizācija ir obligātas funkcionalitātes.

Konteineru glabāšanas saskarnes integrācijai jānodrošina pastāvīgo sējumu nodrošināšana ar dinamiskām glabāšanas klasēm (sākotnēji jābūt konfigurētam vismaz NVME lokālai glabātuvei, NVME bloka pakalpojumam, ietīpības glabāšanas bloka pakalpojumam), momentuzņēmumu spējām un sējumu paplašināšanu. Jābūt iespējai deklarātīvi specificēt jaunas glabāšanas klases, norādot izmantoto datu nesēju veidu un bojājumnoturības iestatījumus (repliku skaits, dzēšanas kodēšanas iestatījumi, piemēram EC3+2 u.c.).

Jābūt iespējai deklarātīvi pieprasīt sējumus izmantojot jebkuru no definētajām glabāšanas klasēm. Jānodrošina iespēja deklarātīvi definēt un pieprasīt izolētas S3 savietojamas glabātuves un koplietotas datņu sistēmas izmantošanai konkrētām skaitļošanas slodzēm. Glabāšanas veikspēja jāoptimizē konteineru darba slodžu specifikai. Grafikas procesoru resursu pārvaldībai konteineru vidē jānodrošina ar NVIDIA GPU operatoru vai ekvivalentu risinājumu, ļaujot konteineriem piekļūt grafikas procesoru resursiem ar resursu izolāciju un koplietošanas spējām. Grafikas procesoru plānošana jāintegrē ar Kubernetes resursu pārvaldību.

Servisu tīkla integrācijai jānodrošina uzlabota tīklošanas spējas mikroservisu arhitektūrām ar datplūsmas pārvaldību, drošības politikām, novērojamību un izklīdētu izsekošanu. Jānodrošina augstas veikspējas mikroservisu tīklojums, kurš izmanto augstas tīkla veikspējas tehnoloģijas (piemēram Cilium Service Mesh ar eBPF, vai Calico VPP dataplane ar Istio, u.c.). Nav pieļaujams izmantot mikroservisu tīklojumu, kurš nav piemērots augstas veikspējas skaitļošanas vajadzībām un nenodrošina noteiktās latentuma un joslas platuma prasības.

8. Pakalpojuma uzturēšana un optimizācija

Piegādātājam nomas perioda laikā pēc Pasūtītāja pieprasījuma jāveic visu sistēmas komponentu programmatūras versiju atjaunināšana, nodrošinot atbilstību jaunākajām drošības prasībām un veikspējas standartiem. Pēc katras programmatūras atjaunināšanas obligāti jāveic sistēmas integritātes testi, lai pārliecinātos par visu komponentu savstarpējo saderību un stabilu darbību.

Pēc integritātes testu veikšanas piegādātājam jāveic sistēmas darbības veikspējas uzlabošana (performance tuning), optimizējot konfigurācijas un parametrus (t.sk. mezglu operētājsistēmas kodola parametrus), lai sasniegtu maksimālu sistēmas efektivitāti un nodrošinātu atbilstību vai pārsniegtu ražotāju publicētos veikspējas rādītājus.

C. PAKALPOJUMU LĪMEŅA GARANTIJAS UN ATBALSTA STANDARTI

9. Pakalpojumu pieejamības un veiktspējas garantijas

HPC mākoņskaitļošanas infrastruktūras pakalpojumam jānodrošina augsts pieejamības līmenis ar 99,5% darbības garantiju, kas tiek mērīta mēneša griezumā un ietver gan plānotos, gan neplānotos pārtraukumus, izņemot iepriekš saskaņotos uzturēšanas darbus.

Veiktspējas garantijām jānodrošina, ka visi piedāvātie resursi - centrālie procesori, operatīvā atmiņa, grafikas procesori, glabāšana un tīkls - konsekventi nodrošina vismaz 90% no deklarētajiem veiktspējas parametriem standarta etalonu testa apstākļos. Veiktspējas pazemināšanās zem šī sliekšņa uzskatāma par pakalpojumu līmeņa pārkāpumu.

Grafikas procesoru veiktspējas garantijām jānodrošina, ka katrai grafikas procesoru kartei jānodrošina vismaz 95% no ražotāja specificētās maksimālās aprēķinu veiktspējas un atmiņas joslas platuma. Regulāri (vismaz reizi mēnesī) jāveic etalonu testi ar standartizētām mākslīgā intelekta un mašīnmācības darba slodzēm, lai pārbaudītu veiktspējas atbilstību.

Glabāšanas veiktspējas garantijām NVME klases bloku glabātuves pakalpojumam jānodrošina vismaz 10 GB/s apkopotas caurlaidības centrālajai glabāšanas sistēmai un vismaz 500 000 IOPS nejaušas rakstīšanas un lasīšanas veiktspēju, glabāšanas kavējums nedrīkst pārsniegt 1 milisekundi parastajām operācijām.

Glabāšanas veiktspējas garantijām HDD bloku glabātuves pakalpojumam jānodrošina vismaz 3 GB/s apkopotu secīgas lasīšanas un 1,5 GB/s secīgas rakstīšanas caurlaidību, kā arī ne mazāk kā 3 000 IOPS nejaušām 4 KiB lasīšanas un rakstīšanas operācijām. Vidējam 4 KiB nejaušu operāciju kavējumam nedrīkst pārsniegt 15 milisekundes, 95. percentiles kavējumam – 30 milisekundes.

Vietējās NVMe glabātuves servisa veiktspējas garantijām katram aprēķinu mezglam jānodrošina PCIe 5.0 NVMe diska ierīces ar ražotāja deklarēto secīgas lasīšanas veiktspēju ne mazāku kā 14 GB/s un rakstīšanas veiktspēju ne mazāku kā 7 GB/s. No konteineriem pieejamajai lokālajai glabātuvei jānodrošina vismaz 8 GB/s secīgas lasīšanas un 4 GB/s secīgas rakstīšanas caurlaidība, kā arī vismaz 500 000 IOPS nejaušām 4 KiB lasīšanas un rakstīšanas operācijām. Vidējam kavējumam nejaušām 4 KiB operācijām nedrīkst pārsniegt 0,5 ms, 95. percentiles kavējumam – 1 ms parastajās darbības slodzēs.

Tīkla veiktspējas garantijām jānodrošina vismaz 200 Gbps efektīva caurlaidība starp aprēķinu resursiem ar kavējumu zem 10 mikrosekundēm. Pakešu zudumu attiecība nedrīkst pārsniegt 0,01% normālas darbības apstākļos.

10. Tehniskā atbalsta pakalpojumu organizācija

Tehniskā atbalsta pakalpojumiem jābūt organizētiem pēc daudzlīmeņu atbalsta modeļa ar specializētu ekspertīzi hiperkonverģentajās tehnoloģijās, mākoņa infrastruktūrā un konteinerizācijas risinājumos. Atbalsta komandai jābūt sertificētiem speciālistiem ar pierādītu pieredzi līdzīgu infrastruktūru pārvaldībā.

Atbalsta pieejamībai jābūt 16 stundas dienā darba dienās (pirmdiena-piektdiena, 6:00-22:00 Latvijas laiks) ar kvalificētu tehnisko personālu, kas spēj risināt gan ikdienas operacionālus jautājumus, gan sarežģītas tehniskas problēmas. Nedēļas nogales ārkārtas atbalsts jābūt pieejams

kritiskām problēmām. Reakcijas laika garantijām jābūt ne ilgāk kā 1 stunda no incidenta ziņojuma saņemšanas brīža.

Kritiskās problēmas, kas ietekmē pakalpojuma pieejamību vai drošību, jāsāk risināt nekavējoties ar nepārtrauktu darbu līdz atrisināšanai. Komunikācijas valodai jābūt latviešu valodā visos atbalsta līmeņos, vai arī ar tulka pakalpojumiem, kas pieejami specializētām piegādātāja konsultācijām, ja nepieciešams. Visa tehniskā dokumentācija jānodrošina latviešu vai angļu valodā.

12. Uzraudzības un analītikas pakalpojumi

HPC mākoņskaitļošanas infrastruktūras hiperkonverģentajai platformai jāintegrē visaptverošs uzraudzības un analītikas risinājums, kas nodrošina reāllaika redzamību visiem infrastruktūras komponentiem un pakalpojumu veikspējas rādītājiem ar paredzošās analītikas spējām.

Infrastruktūras uzraudzības pakalpojumam jābūt rādītāji no visiem slāņiem: aparatūras izmantojums, konteinerizācijas veikspēja, lietojumprogrammu reakcijas laiki, tīkla datplūsmas modeļi un drošības notikumi. Uzraudzības dati jāapkopo centralizētā informācijas panelī ar pielāgojamiem skatiem.

Brīdinājumu sistēmai jābūt konfigurētiem intelligentiem sliekšņiem ar mašīnmācības balstītu anomāliju noteikšanu, kas samazina viltus pozitīvos brīdinājumus un nodrošina rīcībspējīgus paziņojumus. Brīdinājumu korelācija un pamatcēloņu analīzes spējas ir obligātas. Veikspējas analītikas pakalpojumiem jānodrošina vēsturisko tendenču analīze, kapacitātes plānošanas ieteikumi, optimizācijas iespēju identificēšana un izmaksu analīze. Paredzošā analītika jāizmanto, lai prognozētu nākotnes resursu vajadzības.

Ziņojumu pakalpojumiem jābūt automatizēti ziņojumi par pakalpojumu izmantošanu, veikspējas rādītājiem, pieejamības statistiku un izmaksu sadali. Ziņojumiem jābūt pielāgojamiem un pieejamiem vairākos formātos (PDF, CSV, programmatūras saskarne).

Žurnālu pārvaldības pakalpojumiem jānodrošina centralizēta žurnālu vākšana, korelācija, meklēšanas spējas un uzglabāšanas politikas.

D. LĪGUMA NOTEIKUMI UN PAKALPOJUMA STANDARTI

13. Nomas periods un līguma struktūra

HPC mākoņskaitļošanas resursu nomas periods tiek noteikts 24 kalendāra mēnešu garumā, sākot no pakalpojuma tehniskās pieņemšanas datuma. Šis periods nodrošina pietiekamu laiku sarežģītu pētniecības projektu īstenošanai un resursu plānošanas stabilitāti. Līguma pagarināšanas iespēja paredz papildu 12 mēnešu periodu pēc pamata perioda beigām, ar nosacījumu, ka pagarinājums tiek saskaņots rakstveidā vismaz 30 kalendāra dienas pirms sākotnējā līguma termiņa beigām.

Pagarinājuma nosacījumiem jāpaliek identiskiem sākotnējā līguma noteikumiem. Pasūtītājam nav pienākums veikt pakalpojumu nomas pasūtīšanu uz papildu periodu.

Pasūtītājam ir tiesības izbeigt līgumu jebkurā laikā pēc minimālā sešu (6) mēnešu perioda, rakstiski paziņojot par to ne vēlāk kā 30 (trīsdesmit) kalendārās dienas iepriekš. Priekšlaicīga līguma izbeigšana nedrīkst radīt datu zudumu, un visi datu migrācijas un drošas dzēšanas pakalpojumi jānodrošina Pasūtītājam bez papildu maksas.

Pakalpojumu līmeņa saistības jānodrošina nemainīgā apjomā un kvalitātē visā līguma darbības laikā, nepieļaujot pakalpojumu kvalitātes pazemināšanos vai funkcionalitātes samazināšanu. Visi

pakalpojumu uzlabojumi un jaunievedumi, kas tiek ieviesti līguma darbības laikā, jānodrošina Pasūtītājam bez papildu maksas.

14. Datu centra infrastruktūras prasības

Datu centram jāatrodas Latvijas Republikas teritorijā un tam jāatbilst mūsdienu infrastruktūras standartiem, nodrošinot augstu pieejamību, nepieciešamo redundanci un vides kontroli kritiski svarīgu operāciju nepārtrauktai darbībai.

Fiziskās drošības pasākumiem jāietver 24/7 personāla drošība, video uzraudzība ar minimums 90 dienu uzglabāšanu un apmeklētāju pavadīšanas politikas. Drošības incidentu žurnālu un ziņošanas procedūras jābūt dokumentētām. Vides kontrolei jānodrošina redundantas dzesēšanas sistēmas, nepārtraukta barošana, rezerves ģenerators ar minimums 24 stundu degvielas kapacitāti, ugunsdzēsības sistēmas un vides uzraudzība ar brīdinājumiem.

E. TEHNISKĀ PIEDĀVĀJUMA SAGATAVOŠANAS PRASĪBAS

15. Obligātie piedāvājuma komponenti

Pretendentam tehniskajā piedāvājumā jāiekļauj visaptveroša dokumentācijas pakete, kas ļauj objektīvi novērtēt piedāvāto hiperkonverģentā mākoņa pakalpojuma risinājumu un pretendenta piegādes spējas.

15.1 HPC mākoņskaitļošanas infrastruktūras hiperkonverģentās arhitektūras apraksts

Detalizēts tehniskās arhitektūras dokuments ar HPC HCI platformas specifikāciju, komponentu integrācijas pieeju, resursu konteinerizācijas metodiku un mērogojamības dizainu. Arhitektūras diagrammas jāiekļauj ar datu plūsmas, tīkla topoloģijas un drošības robežu vizualizāciju.

Pakalpojumu piegādes modeļa apraksts ar resursu nodrošināšanas procesiem, automatizācijas spējām, orķestrācijas rīkiem un pārvaldības saskarnēm. Integrācijas punkti ar ārējām sistēmām, programmatūras saskarnēm un trešo pušu pakalpojumiem jāprecizē skaidri. konteinerizācijas tehnoloģijas aprakstā jāiekļauj informācija par izvēlēta hipervizora izvēli un pamatojumu, veikspējas optimizācijas risinājumiem, resursu pārvaldības politikām un daudznomnieku vides īstenošanu.

Glabāšanas arhitektūras specifikācija ar izklaidētās glabāšanas tehnoloģiju, redundances mehānismiem, veikspējas raksturlielumiem un dublēšanas/atgūšanas procedūrām. Dažādi glabāšanas līmeņi un to lietošanas gadījumi jāapraksta detalizēti.

Tīkla arhitektūras dizains ar programmdefinētu tīklu (SDN) īstenošanu, datplūsmas pārvaldību, drošības kontrolēm un ārējās savienojamības iespējām. Tīkla segmentācija, pakalpojumu kvalitātes politikas un uzraudzības spējas jādokumentē visaptveroši.

15.2 Pakalpojumu pārvaldības un atbalsta dokumentācija

Pakalpojumu pārvaldības ietvara dokumentācija ar procesu īstenošanu, organizatorisko struktūru, lomu un atbildību matricām un eskalācijas procedūrām. Procesu plūsmas jāvizualizē ar joslu diagrammām vai līdzīgiem rīkiem.

Tehniskā atbalsta organizācijas apraksts ar komandas struktūru, prasmju sertifikācijām, pieredzes vēsturi un pieejamības grafikiem. Atbalsta rīki, pieteikumu sistēmas un zināšanu pārvaldības procesi jāapraksta detalizēti.

Kvalitātes nodrošināšanas procedūras ar veikspējas uzraudzības metodēm, pakalpojumu līmeņa mērīšanas paņēmieniem, ziņošanas mehānismiem un nepārtrauktu uzlabošanas procesiem.

15.3 Drošības un atbilstības dokumentācija

Informācijas drošības politika un īstenošanas detaļas ar piekļuves kontroli, šifrēšanas standartiem, ievainojamības pārvaldību un incidentu reaģēšanas procedūrām. Drošības arhitektūras diagrammas jāiekļauj ar uzticības robežām un kontroles punktiem.

GDPR atbilstības dokumentācija ar datu apstrādes līgumiem, privātuma ietekmes novērtējumiem, datu subjektu tiesību procedūrām un pārrobežu datu pārsūtīšanas garantijām.

Audita un atbilstības ziņošanas spējas ar žurnālu pārvaldību, uzraudzības brīdinājumiem, tiesību ekspertīzes analīzes rīkiem un regulējošo ziņojumu automatizāciju.

Uzņēmējdarbības nepārtrauktības un katastrofu atgūšanas plānošana ar atgūšanas laika mērķu definīcijām, dublēšanas procedūrām, pārslēgšanās mehānismiem un testēšanas grafikiem.

15.4 Finansiālie un līguma dokumenti

Piegādātājam jāsniedz detalizēta cenu tabula ar pakalpojumu izmaksām. Jāiesniedz kopējo izmaksu aprēķins uz 24 mēnešiem.

Līguma noteikumi un nosacījumi ar pakalpojumu piegādes saistībām, pakalpojumu līmeņa definīcijām, sodu klauzulām, intelektuālā īpašuma tiesībām un atbildības noteikumiem.

Šī tehniskā specifikācija nodrošina pilnīgu regulatīvo ietvaru hiperkonverģentās mākoņa infrastruktūras nomas iepirkumam, uzsverot pakalpojumorientētu pieeju ar uzsvāru uz funkcionalitāti, veikspēju un atbalsta kvalitāti.

F. TEHNISKĀ RISINĀJUMA PREZENTĀCIJAS PRASĪBAS

16. Obligātā tehniskā risinājuma prezentācija

Visiem pretendentiem, kuru tehniskais piedāvājums atbilst minimālajām kvalifikācijas prasībām un ir atzīts par atbilstošu, pēc pasūtītāja pieprasījuma būs jāveic detalizēta tehniskā risinājuma prezentācija Pasūtītāja izveidotās komisijas priekšā. Prezentācija ir neatņemama iepirkuma procedūras daļa un tiek izmantota tehniskā piedāvājuma pilnīgai novērtēšanai.

16.1 Prezentācijas organizatoriskie nosacījumi

Tehniskā risinājuma prezentācijai jāietver prezentācijas laiku un jautājumu un atbilžu sesiju. Prezentācija jāveic klātienē Pasūtītāja norādītajā vietā Pasūtītāja birojā, un tā jānotur latviešu valodā ar iespēju izmantot angļu valodu tehnisko terminu skaidrošanai. Prezentācijas datums un laiks tiks saskaņots ar pretendentu ne vēlāk kā 5 darba dienas pirms prezentācijai noliktā laika.

16.2 Prezentācijas obligātais saturs

Tehniskā risinājuma prezentācijā obligāti jāapskata šādi temati ar norādīto minimālo laika sadalījumu:

Hiperkonverģentās arhitektūras demonstrācija:

- Detalizēts arhitektūras pārskats ar vizuālām diagrammām
- Komponentu integrācijas un datu plūsmu demonstrācija
- konteinerizācijas slāņa un resursu abstrakcijas skaidrojums
- Programmdefinētās infrastruktūras (SDI) implementācijas demonstrācija
- Mērogojamības un elastības mehānismu skaidrojums

Tehnisko specifikāciju atbilstības demonstrācija:

- CPU, RAM, GPU, storage un network resursu detalizēta specifikācija
- Veiktspējas rādītāju demonstrācija
- Garantēto SLA parametru skaidrojums un monitoringa rīku demonstrācija

Datu centra infrastruktūras prezentācija:

- Datu centra fiziskās infrastruktūras pārskats (var izmantot video materiālus)
- Sertifikāciju un standartu demonstrācija
- Drošības pasākumu un vides kontroles sistēmu apskats

Pārvaldības un uzraudzības rīku demonstrācija :

- Live demonstrācija ar pārvaldības platformu (var būt demo vide)
- Lietotāja saskarnes funkcionalitātes demonstrācija
- Uzraudzības, alerting un reporting spēju demonstrācija

Prezentācijas neveikšanas sekas

Ja pretendents atsakās no prezentācijas veikšanas vai neierodas noteiktajā laikā bez pamatota iemesla, viņa tehniskais piedāvājums tiek uzskatīts par neatbilstošu un izslēgts no turpmākās vērtēšanas.

Pamatoti iemesli prezentācijas pārcelšanai ir: force majeure apstākļi, nopietna komandas locekļa saslimšana ar medicīnisko atzinumu, vai citi objektīvi apstākļi. Prezentācijas pārcelšana jāaskaņo vismaz 48 stundas iepriekš. Prezentācijas kvalitāte un satura atbilstība ir obligāts priekšnosacījums līguma noslēgšanai. Ja prezentācija neiztur novērtēšanu vai tiek konstatēta būtiska neatbilstība starp iesniegto tehnisko piedāvājumu un demonstrēto risinājumu, pretendents var tikt izslēgts no iepirkuma.