

PLEC DE PRESCRIPCIONS TÈCNIQUES PER LA CONTRACTACIÓ D'UN SERVEI DE RECURSOS AL NÚVOL DE GOOGLE I SERVEIS ADDICIONALS PEL PROJECTE "SCIENCE PLATFORM CLOUD INFRASTRUCTURE FOR OUTSIZE USAGE SCENARIOS" (SPACIOUS) DE LA UNIVERSITAT DE BARCELONA (Contracte basat en l'acord marc d'homologació de proveïdors per a la prestació de serveis al núvol, per a les institucions que formen part del grup de compra CSUC 22/26)

EXPEDIENT 2024/151

L'objecte del present document (en endavant PPT) és l'establiment de les prescripcions tècniques que regeixen el procediment de contractació d'un servei de recursos al núvol de Google i serveis addicionals per a poder realitzar el desplegament d'un clúster de càlcul (escalable) basat en una tecnologia de contenidors, connectat a un gran pool d'emmagatzematge amb baixa latència pel projecte SPACIOUS: "*Science Platform Cloud Infrastructure for Outsize Usage Scenarios*" de la Universitat de Barcelona, amb GA 101135205 i finançat per EU Horizon Europe HORIZON-CL4-2023-SPACE-01-71.

1. Antecedents

L'Institut de Ciències del Cosmos de la Universitat de Barcelona (ICCUB) és un centre interdisciplinari dedicat a la recerca fonamental en els camps de la cosmologia, l'astrofísica i la física de partícules. L'institut té un fort programa tecnològic gràcies a la seva participació en col·laboracions internacionals en astronomia observacional i física experimental de partícules. L'Institut es va crear l'any 2006 com a instrument de la Universitat de Barcelona per al suport actiu de la recerca en astrofísica teòrica i física de partícules, prestant especial atenció a la seva sinergia amb la cosmologia, per promoure la física experimental i el desenvolupament d'instruments, permetent una participació significativa de la Universitat de Barcelona en grans col·laboracions internacionals, i per atraure personal científic altament qualificat. L'ICCUB ha experimentat un creixement important, convertint-se en una institució de recerca consolidada amb més de 50 científics de llarga durada, 15 enginyers i 65 investigadors postdoctorals i doctorands.

El grup Gaia de l'ICCUB és un equip interdisciplinari format per enginyers de programari i científics. El nostre grup ha estat implicat en la missió espacial Gaia de l'Agència Espacial Europea (ESA) des del principi. Té una llarga experiència en el tractament massiu de dades i l'explotació de dades científiques que el fa molt adequat per afrontar el repte d'utilitzar els serveis comercials al núvol (CCS) amb finalitats científiques. El nostre personal científic treballarà en estreta col·laboració amb els nostres experts en enginyeria en supercomputació i processament massiu de dades per a la gestió de les infraestructures del núvol. A més, estem en una bona posició per difondre els beneficis de la computació en núvol en la recerca. El nostre equip també és molt actiu en la divulgació de la ciència i la tecnologia. Les nostres activitats inclouen l'organització d'esdeveniments, conferències, exposicions, xarxes socials i pàgines web.

El projecte SPACIOUS pretén impulsar l'explotació científica de les dades de la missió de l'ESA mitjançant el desenvolupament d'un marc computacional per a la investigació astrofísica que requereix que s'abordin les tecnologies de Big Data i Data Mining. Serà d'aplicació immediata als llançaments de dades de la missió Gaia, que han assolit desenes de Terabytes de mida i que arribaran a centenars de Terabytes en un futur proper; també serà d'aplicació immediata a la missió Euclid que es posarà en marxa aquest any i que en breu publicarà les seves primeres dades; i, finalment, marcarà el camí per a una explotació més eficient de l'espai present i futur i conjunts de dades massius sobre el terreny, definint mètodes d'anàlisi de dades que van més enllà de l'enfocament actual de "dades al codi".

Per assolir aquest ambiciós objectiu, SPACIOUS pretén combinar les tecnologies existents en un nou entorn de Big Data i Data Mining (BDAF) per permetre l'anàlisi de les dades de la missió

de l'ESA fins a un nivell que no és possible assolir amb els mitjans d'anàlisi tradicionals. L'ambició de SPACIOUS és obrir la plataforma BDAF a la comunitat, alhora que s'explota les dades disponibles de la missió Gaia i Euclid desenvolupant la nostra pròpia investigació astrofísica (i aprenent de l'experiència), en col·laboració amb equips científics externs. Ens centrarem a abordar qüestions clau de recerca astrofísica per augmentar el coneixement de la nostra galàxia i el cosmos, alhora que contribuïm a posar el sector espacial europeu al capdavant de l'astrofísica de Big Data.

Un dels objectius de SPACIOUS és fer que el sistema BDAF sigui fàcil de portar a una varietat d'entorns d'execució, des de clústers interns fins a grans instal·lacions de supercomputació. Entre els entorns objectiu incloem els núvols comercials, i l'objectiu d'aquesta licitació és proporcionar els recursos al núvol que permetin el desenvolupament i prova d'aquestes opcions al núvol de Google.

Així doncs, l'objecte de la present licitació és la contractació del subministrament de recursos al núvol de Google i els serveis relacionats que permeti combinar les tecnologies BDAF i un anàlisi i explotació massiva de dades que permeti assolir els objectius de recerca del projecte SPACIOUS.

L'empresa adjudicatària quedarà obligada a executar l'objecte del contracte, de conformitat amb les condicions que es determinen al present plec, degut als requeriments experimentals de l'estudi.

2. Requisits tècnics del núvol de Google

Per a la implementació del sistema BDAF com a entorn de Big Data i mineria de dades requerim el desplegament d'un clúster de càlcul (escalable) basat en una tecnologia de contenidors, connectat a un gran pool d'emmagatzematge amb baixa latència.

El subministrament de recursos al núvol de Google i els serveis connexos han de complir amb els següents requisits:

1. Clústers de computació: Hem de desplegar clústers de computació sota demanda. El sistema BDAF en contenidors es desplegarà als nodes del clúster per executar el programari científic. Més d'un clúster pot estar actiu en un moment determinat, i la seva mida (nombre de nodes) s'escalarà sota demanda (des d'un sol node fins a un gran nombre de nodes). No necessitem necessàriament nodes amb GPU, però la seva disponibilitat serà avaluada positivament com a millora.
2. Orquestració i contenidors: Es pretén utilitzar Kubernetes i Docker, ja que estan àmpliament disponibles en els núvols comercials i privats.
3. Emmagatzematge: atès que l'objectiu de SPACIOUS és l'explotació de grans conjunts de dades de missions espacials, els clústers de càlcul necessiten un conjunt de dades gran; l'espai requerit variarà amb el desenvolupament del projecte, amb un valor màxim de 600TB. Aquest emmagatzematge ha de ser accessible amb una latència baixa i un alt rendiment des dels clústers (idealment es troben al mateix centre de dades que els nodes del clúster). Tanmateix, per limitar els costos, caldria la possibilitat de traslladar les dades a un emmagatzematge de menys rendiment però més barat. En altres paraules, tot i que no són necessàries per als càlculs, volem mantenir les dades en un emmagatzematge de baix cost i traslladar-les (tan aviat com sigui possible) a un emmagatzematge d'alt rendiment quan sigui necessari per als càlculs. També s'avaluaran positivament les opcions per limitar el cost de la càrrega i descàrrega de grans quantitats de dades (centenars de terabytes).

A més, la solució hauria de contenir la capacitat de desplegar quaderns Python sense servidor com a SaaS per col·laborar amb un ús fàcil de la CPU.

3. Servei de suport i formació

Es requereix suport tècnic bàsic del proveïdor per a la definició, el desplegament i l'ús de la nostra configuració al núvol. El suport ha de ser especialment proper durant les fases inicials, i ha de permetre definir i configurar un entorn de núvol adequat per al projecte: comunicacions, seguretat, gestió d'usuaris, configuració de recursos, etc. En fases posteriors es requereix suport per a ajustos de la configuració i per a l'optimització dels recursos descrits anteriorment.

Es requereix com a mínim una formació inicial de dos dies de jornada laboral per a la configuració inicial de l'entorn, i un altre període de com a mínim dos dies laboral de formació al cap d'un any per a la revisió de la configuració i optimització dels recursos.

4. Comptabilitat i registre

El subministrament a contractar ha de permetre extreure i contenir informes automàtics dels recursos consumits (o portal web que permeti comprovar-ho), incloent almenys les hores de CPU, emmagatzematge i transferències, acumulades fins a un temps determinat i amb l'historial de consum (per exemple, per dia o per setmana). La latència (temps entre la generació de costos i la seva informació a la eina) ha de ser com a màxim de 24h. Valorarem especialment el suport del proveïdor per a la comptabilitat i el control dels fons disponibles: avisos de despeses, eines comptables, limitacions de costos d'ús de recursos, etc.

Específicament:

- L'eina d'anàlisi ha de proporcionar el cost detallat en temps real (en la plataforma Cloud, i de manera opcional i complementària en altres eines del proveïdor) de tots els elements de cost que s'estiguin fent servir en l'entorn Cloud (no tan sols hores de CPU, emmagatzematge i transferències, sinó de tots aquells serveis operatius en la plataforma Cloud).
- Les dades de cost acumulat han de tenir un historial de consum per dies, setmanes i mesos durant la duració del projecte.

5. Pla de sortida

El licitador notificarà a al responsable del contracte quan s'hagi consumit el 75% dels recursos i s'assegurarà que totes les persones de contacte de l'Entitat rellevants per a aquest contracte seran notificades oportunament.

El Licitador proporcionarà l'assistència suficient per transferir totes les dades relatives al contracte sense cap cost des de l'entorn (núvol) del Licitador a un entorn de tercers (núvol).

Dr. Xavier Luri