

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 repositori 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 missions 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 va posar en marxa el juliol de 2023 ; 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 missions 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 ajudarà a les necessitats del projecte. Necessitarem accés a Google Compute Engine per crear nodes per a diversos propòsits, amb tota la gamma de tipus de màquines per triar mitjançant imatges de Linux (no Windows).
2. Orquestració i contenidors: Es pretén utilitzar Kubernetes i Docker, ja que estan àmpliament disponibles en els núvols comercials i privats. Necessitem accés a Google Kubernetes Engine (GKE) per crear clústers Kubernetes i la possibilitat d'utilitzar-lo diverses vegades per crear diversos clústers per a diversos entorns. Necessitem la capacitat de crear clústers estàndard (no clústers de pilot automàtic). Necessitem la capacitat de crear grups de nodes i nodes per utilitzar-los dins de GKE, la gamma completa de tipus de màquines per triar amb imatges de Linux (no Windows).
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. *Persistent disk* - aquesta és la solució d'emmagatzematge en bloc de Google. Es presenta en diversos modes de rendiment, des de l'estàndard fins a l'extrem (alguns són SSD però no estan connectats localment). Des d'una perspectiva tècnica, ens agradaria el màxim rendiment possible, però per descomptat el rendiment depèn del cost. La categoria d'emmagatzematge persistent "equilibrat" sembla ser un compromís sensat de cost i rendiment.
- b. *Locally Attached Ephemeral SSD* - SSD efímer connectat localment: per a emmagatzematge en calent connectat directament.
- c. *Google Cloud Storage*: és l'emmagatzematge d'objectes a Google Cloud. Els usos seran emmagatzemar conjunts de dades, emmagatzemar artefactes i altres emmagatzematges d'ús general.

A més, la solució hauria de contenir la capacitat de desplegar "notebooks" de 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 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. Francesc Xavier Luri Carrascoso

Catedràtic d'universitat adscrit al Departament de Física Quàntica i Astrofísica de la Universitat de Barcelona

ANNEX 1

Call: HORIZON-CL4-2023-SPACE-01

(STRATEGIC AUTONOMY IN DEVELOPING, DEPLOYING AND
USING GLOBAL SPACE BASED INFRASTRUCTURES, SERVICES,
APPLICATIONS AND DATA 2023)

Topic: HORIZON-CL4-2023-SPACE-01-71

Type of Action: HORIZON-RIA

Proposal number:

101135205

Proposal acronym:

SPACIOUS

1. Objetivos científicos

Gaia ya es la fuente de producción científica más exitosa entre las misiones de la ESA, con más de 8600 artículos publicados basados en sus datos (fuente: ADS). El tamaño de sus publicaciones de datos (DRs) ha ido aumentando de forma constante, siendo la más reciente (DR3) de aproximadamente 10 TB; sin embargo, hasta ahora su explotación se ha basado principalmente en enfoques clásicos de análisis de datos, que consisten en la descarga de datos desde el archivo de la misión. Para aprovechar completamente este volumen de datos, serán necesarios nuevos enfoques como el desarrollado en este proyecto. Además, las próximas publicaciones aumentarán considerablemente el volumen de datos: DR4 (previsto para 2025) alcanzará los 0.5 petabytes. Por otro lado, la misión Euclid, lanzada en julio de 2023) está generando otro conjunto de datos que también requerirá herramientas innovadoras para su explotación eficiente y completa. La única forma de aprovechar con éxito todos estos nuevos datos es cambiar el paradigma del análisis de datos, tal como se propone aquí.

Para aumentar la producción científica basada en los archivos de la ESA y producir productos de datos y métodos de nivel superior, nuestro objetivo es implementar varios proyectos de análisis científico a gran escala (desafíos) en el sistema BDAF utilizando datos de Gaia y Euclid. Estos proyectos se desarrollarán con la colaboración de la comunidad científica, con un paquete de trabajo específico WP200 dedicado a permitir la explotación científica de estas misiones. WP200 se encarga de implementar los proyectos, asegurando que las características y el rendimiento del BDAF se ajusten a las necesidades de los usuarios científicos, y también de brindar apoyo a los equipos externos cuando sea aplicable.

Se prevén dos conjuntos de desafíos. Primero, los desafíos internos, llevados a cabo por los propios equipos de SPACIOUS, que también se utilizan para impulsar los requisitos de diseño del sistema, sirviendo como casos de prueba y demostradores tecnológicos del marco BDAF. En una segunda fase de SPACIOUS, abriremos el sistema a la comunidad a través de una Convocatoria Abierta de propuestas, y seleccionaremos los desafíos más relevantes (tanto científica como técnicamente) presentados para su implementación. Los equipos científicos externos podrán utilizar el marco BDAF sin costo durante la duración del proyecto, a través de las Convocatorias Abiertas.

El concepto básico de SPACIOUS es proporcionar un entorno de análisis de datos que combine: 1) La disponibilidad de una amplia variedad de conjuntos de datos en un único repositorio; 2) Un marco fácil de implementar que proporcione acceso a este conjunto de datos bajo un paradigma de Big Data, permitiendo una asignación escalable de recursos de computación adaptada a las necesidades de procesamiento; 3) Un conjunto de herramientas de análisis de datos adaptadas a este marco, facilitando la explotación científica de los datos.

Este sistema permitirá el uso de los conjuntos de datos existentes y futuros de las misiones Gaia y Euclid, en combinación con conjuntos de datos de instalaciones terrestres, de formas que no son soportadas por el paradigma usual (y actualmente dominante) basado en bases de datos relacionales y descarga de datos. Este enfoque a veces se describe como "llevar el código a los datos en lugar de los datos al código". De este modo, hemos identificado la falta de una plataforma de código a datos para la explotación de los datos de las misiones espaciales de la ESA que pretendemos cubrir. También es importante señalar que la propia ESA ha reconocido esta necesidad y está analizando internamente opciones potenciales para este tipo de servicio.

Esto mejorará significativamente la explotación científica de estos conjuntos de datos y conducirá a productos y resultados que no podrían obtenerse utilizando los enfoques actuales. También permitirá que una mayor parte de la comunidad científica pueda realizar análisis de datos masivos y profundos de los conjuntos completos de datos de los Archivos de Ciencia Espacial de la ESA, aumentando su impacto científico. Durante el proyecto, utilizaremos el marco y las

metodologías desarrolladas para implementar una serie de desafíos científicos basados en el uso de los datos de las misiones Gaia y Euclid.

2. Cronograma

La duración del proyecto SPACIOUS es de 36 meses, divididos en tres etapas.

1. **La etapa preparatoria:** los primeros 6 meses (enero – junio 2024) se dedicarán principalmente a actividades preparatorias, incluyendo la publicación de ofertas de trabajo, coordinación, reuniones de lanzamiento y contratación de personal. Sin embargo, una parte del desarrollo ya comenzará en esta etapa.
2. **La etapa principal de desarrollo:** los siguientes 16 meses (julio 2024 – octubre 2025) se dedicarán al desarrollo de la infraestructura (plataforma BDAF), lo cual incluye el diseño, configuración y pruebas de la(s) plataforma(s) de Big Data en el Centro de Supercomputación de Barcelona y en la nube. Además, se llevarán a cabo los desafíos internos, como prueba del sistema y guía para su desarrollo. Durante los últimos 4 meses, abriremos las convocatorias a la comunidad científica para la explotación de los datos a través de la Convocatoria Abierta de Grandes Trabajos (Big Jobs Open Call).
3. **La etapa operativa:** durante los últimos 14 meses (noviembre 2025 – diciembre 2026), la plataforma BDAF estará disponible para la comunidad y el equipo interno para analizar los datos mediante la implementación de los desafíos y generar los resultados.

Desarrollaremos el Marco de Análisis de Big Data (BDAF) que se utilizará para implementar los desafíos de SPACIOUS. Cuatro de estos desafíos se implementarán internamente en el consorcio SPACIOUS; estos son los desafíos internos que se describen más adelante. La implementación de los desafíos internos mejorará y validará la plataforma. Posteriormente, abriremos el uso de BDAF a equipos científicos externos para que puedan implementar sus propios desafíos. Finalmente, dejaremos plantillas en la nube y un entorno virtualizado de BDAF como legado al final del proyecto.

3. Marco de Análisis de Big Data (Big Data Analysis Framework, BDAF)

El Marco de Análisis de Big Data (BDAF) que se desarrollará se basa en la experiencia adquirida en el proyecto GENIUS FP7 para ofrecer un marco de Big Data productivo y exportable. Este marco se desarrollará y operará en dos infraestructuras complementarias diseñadas y presupuestadas como parte de proyecto:

Recursos de BSC:

- **Computación y Almacenamiento:** Los recursos de computación y almacenamiento para el intercambio y explotación de datos por parte de la comunidad investigadora utilizando la plataforma BDAF serán proporcionados por el BSC. Esto incluirá hasta 5 millones de horas de CPU de los sistemas de computación del BSC y 1 PB en el contexto de la Red Española de Supercomputación (RES, www.res.es/en). La solicitud de recursos de almacenamiento ya ha sido concedida bajo el esquema de pre-reserva de acceso RES, para permitir su uso durante toda la duración del proyecto SPACIOUS, para el despliegue de la plataforma y explotación de los casos de uso del proyecto, además de dos años después del proyecto para asegurar su sostenibilidad. La necesidad de 5 millones de horas de CPU se ha estimado a partir de la experiencia del equipo de UB en la ejecución en el BSC de elementos del pipeline de procesamiento de datos de Gaia que realizan un procesamiento completo de uno de los tipos de datos de Gaia (astrometría). El tamaño de almacenamiento de 1 PB se basa en el tamaño esperado del Gaia DR4 (0.6 PB) más una reserva para conjuntos de datos científicos adicionales, datos intermedios y productos finales.

Servicios de Nube Comercial (CCS):

- **Recursos:** Utilizando el presupuesto de SPACIOUS, queremos contratar recursos de almacenamiento y cómputo en uno de los CCS (Google) donde desarrollaremos, desplegaremos y operaremos el sistema BDAF. Es importante señalar que haremos un esfuerzo especial en diseñar el BDAF para que cumpla con la Nube Europea de Ciencia Abierta (EOSC), con el objetivo de sus futuros usos más allá del final del proyecto. Esta infraestructura de Big Data se basará en Yarn, Apache Hadoop y Apache Spark, que son motores poderosos para la gestión de grandes cantidades de datos. El presupuesto asignado se ha definido a partir de la experiencia del equipo de [Gaia-UB en el proyecto OCRE](#), donde se han ejecutado varios proyectos científicos (asimilables a los desafíos de menor escala de SPACIOUS) en la nube de Google.

En cada una de estas infraestructuras construiremos un repositorio de datos que contendrá los conjuntos de datos de misión de la ESA más relevantes que se utilizarán durante el proyecto, junto con una selección de otros conjuntos de datos de encuestas terrestres que aportan valor añadido para su explotación. Este repositorio de datos será la base para el despliegue del sistema BDAF y se complementará (a través de espacios de usuario o mediante adiciones a la base) con conjuntos de datos necesarios para la implementación de los desafíos seleccionados.

Basado en este marco (BDAF + repositorio de datos), se proporcionarán una variedad de soluciones de minería de datos para la explotación científica de los datos. La infraestructura BDAF será compatible con bibliotecas de Aprendizaje Automático como MLib de Apache Spark, por lo que el usuario tendrá acceso directo a algoritmos de Aprendizaje Automático tales como clasificación, regresión, árboles de decisión, recomendación, agrupamiento, modelado de temas y conjuntos frecuentes, entre otros. Además, proporcionaremos utilidades para flujos de trabajo de Aprendizaje Automático y otros métodos estadísticos como el análisis de componentes principales y pruebas de hipótesis. Queremos enfatizar que estas herramientas de minería de datos estarán especialmente adaptadas para grandes cantidades de datos y también tendrán la capacidad de confrontar modelos con datos. La provisión de una infraestructura de Big Data con las herramientas y métodos complementarios, combinada con la disponibilidad conjunta de una variedad de datos de los archivos de la ESA y observatorios terrestres, representará un gran avance en la capacidad de análisis de datos con respecto a las capacidades de las plataformas existentes que albergan datos de misiones espaciales.

Para la implementación de los desafíos científicos de SPACIOUS (internos y externos), planeamos diferentes formas de beneficiarnos del entorno descrito anteriormente, en los dos entornos descritos (BSC y CCS):

1. **Pequeños Trabajos:** Los equipos científicos podrán lanzar trabajos pequeños en la infraestructura BSC/CCS. Cada usuario tendrá algunos recursos limitados para ejecutar trabajos pequeños (restringidos a un corto período de tiempo de ejecución continua). Esta opción se puede utilizar para aprender a usar el sistema y para el desarrollo preliminar.
2. **Grandes Trabajos:** Los equipos científicos que trabajen en los desafíos seleccionados podrán ejecutar "Grandes Trabajos" en el BSC bajo un procedimiento planificado supervisado por los expertos de SPACIOUS. El apoyo brindado por nuestros expertos incluye orientación sobre el diseño de los algoritmos y el uso óptimo de la infraestructura BDAF.
3. **Soporte Directo:** Para un subconjunto de desafíos seleccionados con alto impacto, también proporcionaremos apoyo directamente en su implementación y en la codificación de algoritmos. Estos serán casos donde las necesidades de procesamiento lleven el sistema a sus límites (permitiéndonos así aprender lecciones importantes de la experiencia) y requieran una profunda experiencia que solo el equipo de SPACIOUS puede proporcionar.

Queremos enfatizar que uno de los puntos clave del BDAF será su portabilidad. El BDAF será diseñado para ser altamente portátil, tanto para alentar su uso por equipos externos como para proporcionar un sistema legado que pueda ser utilizado en la mayor cantidad de áreas posibles. Complementando esto, el proyecto documentará adecuadamente para facilitar su adopción por parte de usuarios externos. Por lo tanto, no excluimos la posibilidad de que durante la fase operativa de SPACIOUS algunos equipos externos puedan adoptar el sistema BDAF para sus propios proyectos de explotación científica.

Por otro lado, dado que el BDAF se desplegará en una infraestructura de computación en la nube comercial, se beneficiará de toda la flexibilidad administrativa inherente a los entornos virtuales, escalabilidad, flexibilidad, legado y longevidad. Esto incluye autoescalado para hacer frente a picos en la demanda, gestión de recursos, aislamiento de procesos, seguridad, autenticación y recuperación ante desastres, entre otros. Por lo tanto, uno de los principales entregables del proyecto será una plantilla de nube de código abierto (por ejemplo, Terraform) capaz de desplegar una plataforma BDAF en Servicios en la Nube (Google Cloud). Esta característica también puede ser utilizada por equipos externos para realizar sus propios despliegues para la explotación científica durante la fase operativa de SPACIOUS o después de su finalización. Nuevamente, destacamos que la infraestructura BDAF será cuidadosamente diseñada para ser compatible con los requisitos de la Nube Europea de Ciencia Abierta.

4. Desafíos científicos externos

Como se mencionó anteriormente, los desafíos científicos externos que se implementarán como parte de SPACIOUS serán un componente clave del proyecto. Estos desafíos servirán a un doble propósito: por un lado, aumentarán la conciencia de la comunidad científica sobre los desarrollos de SPACIOUS. Las Convocatorias Abiertas y la publicidad de los resultados obtenidos serán excelentes canales para difundir el proyecto. Por otro lado, abrirán la explotación de los sistemas SPACIOUS, y especialmente de los datos de Gaia y Euclid, a áreas más allá de la experiencia de sus socios, involucrando equipos adicionales con antecedentes muy diversos. Estos desafíos serán apoyados para su ejecución en los entornos del BSC y CCS:

- Las plataformas desarrolladas estarán disponibles para la comunidad durante los últimos 14 meses del proyecto para desarrollar su análisis de los datos. También se puede considerar el apoyo para el despliegue del sistema en clústeres locales/nubes de los equipos solicitantes.
- Como se mencionó anteriormente, estarán disponibles 5 millones de horas de CPU de los sistemas de computación de investigación del BSC y 1PB en el contexto de la Red Española de Supercomputación (RES, www.res.es/en). Además, se asignarán recursos en Servicios de Nube Comercial para este propósito.
- El apoyo técnico para su implementación y ejecución será proporcionado por los equipos de SPACIOUS.

El procedimiento para la selección de los desafíos comenzará con una Convocatoria Abierta para propuestas a la comunidad. Esta convocatoria proporcionará información detallada sobre los sistemas disponibles y el apoyo que se brindará. También prevemos la posibilidad de proporcionar "cuentas de invitado" para la evaluación del sistema en esta etapa. Las propuestas deben describir el caso científico y estimar las necesidades de procesamiento de datos. Su viabilidad técnica y los requisitos de recursos serán evaluados por el equipo técnico de SPACIOUS. Las propuestas y la evaluación técnica se remitirán a un Comité de Asignación de Recursos que las priorizará. Basado en estas prioridades, los recursos solicitados y la cantidad total de recursos disponibles, se definirá la lista de desafíos externos a implementar, y luego se implementarán en la última fase del proyecto.