

**Nº Expediente: PA2/2025**

**PLIEGO DE PRESCRIPCIONES TÉCNICAS DEL CONTRATO MIXTO DE SUMINISTRO DE LICENCIAS DE USO DE SOFTWARE Y SERVICIOS ASOCIADOS DE INSTALACIÓN Y SOPORTE PARA LA AMPLIACIÓN DE LA PLATAFORMA BIG DATA ANALYTICS MEDIANTE HIBRIDACIÓN EN CLOUD EN EL MARCO DEL PROYECTO “HEALTHDATA@MAD-R&I” ACORDE AL ESPACIO EUROPEO DE DATOS EN SALUD, DE LA CONVOCATORIA DEL MINISTERIO PARA LA TRANSFORMACIÓN DIGITAL Y DE LA FUNCIÓN PÚBLICA EN EL MARCO DEL PLAN DE RECUPERACIÓN, TRANSFORMACIÓN Y RESILIENCIA- NEXT GENERATION EU (PROGRAMA ESPACIOS DE DATOS SECTORIALES), MEDIANTE PROCEDIMIENTO ABIERTO SUJETO A REGULACIÓN ARMONIZADA Y TRAMITACIÓN URGENTE.**

### **Primero. Justificación y Objeto del Contrato**

---

La presente contratación tiene como objetivo contribuir al fortalecimiento y ampliación de la capacidad de análisis de datos sanitarios para uso secundario con fines de investigación e innovación en salud mediante la adquisición de licencias en modo de suministro, y la implementación de los servicios asociados a su instalación y soporte, en el entorno tecnológico del *DataLake* de HealthData@MAD-R&I. El objetivo final es mejorar la capacidad para utilizar datos existentes y nuevos para la toma de decisiones en tiempo real y el desarrollo de modelos predictivos en el ámbito de la salud.

El proyecto HealthData@MAD-R&I tiene como objetivo general definir, desarrollar, validar y documentar casos de uso relacionados con la investigación y la innovación que permitan validar los servicios del Espacio de datos sanitarios electrónicos de uso secundario. En el marco del mismo, se han previsto cuatro casos de uso en diferentes áreas de investigación que sirven como demostradores del potencial científico-tecnológico y comercial que los datos de salud aportan a la economía del dato en el sector sanitario:

- Caso de uso 1: Optimización de las derivaciones de sujetos con enfermedades reumáticas y musculoesqueléticas entre atención primaria y reumatología
- Caso de uso 2: Trayectorias asistenciales de las mujeres largas supervivientes de cáncer de mama SURBCANMADRID
- Caso de uso 3: Hospitalización No Programada (MAD-HNP.eSTRATA)
- Caso de uso 4: Modelo explicativo en vida real sobre la efectividad de las estatinas en la reducción de eventos cardiovasculares y la mortalidad en la población anciana sin antecedentes de enfermedad cardiovascular de la Comunidad de Madrid.

Para lo anterior, se precisa fortalecer y ampliar las capacidades de la Plataforma HealthData@MAD-R&I que dará soporte a los casos de uso, por lo que resulta necesario la adquisición de licencias que cubran todas las necesidades y requerimiento técnicos.

Por todo lo cual, el **objeto** de esta **contratación** es el suministro de licencias de uso de *software* para la ampliación de licencias en *Cloud*, para expandir la plataforma de *Big Data Analytics*. Esto se conseguirá mediante la hibridación en *Cloud*, lo que permitirá facilitar la incorporación de nuevos datos, desarrollar modelos de Inteligencia Artificial (IA) basados en inferencia NVIDIA y facilitar el intercambio de datos con otros departamentos. La finalidad es extraer valor de los datos sanitarios existentes disponibles, utilizando la información para la toma de decisiones informadas en el ámbito de la salud. Para lograrlo, se fortalecerá y ampliará la capacidad de análisis de datos sanitarios, adquiriendo licencias de la plataforma tecnológica Cloudera. Esto permitirá optimizar el uso de los datos existentes y futuros, facilitando la toma de decisiones en tiempo real y el desarrollo de modelos predictivos en el ámbito de la salud.

Las características técnicas principales de las licencias de uso de software a contratar se especifican en el punto 4 del presente Pliego.

## **Segundo. Entorno Tecnológico Actual**

---

HealthData@MAD-R&I formará parte del Espacio de Datos de la Comunidad de Madrid, que actualmente utiliza la plataforma Cloudera Data Platform, en Cloud Amazon Web Services (AWS), con los componentes que se describen en el punto 4 del presente pliego. Dicha plataforma centraliza datos y herramientas para iniciativas de analítica descriptiva y avanzada

## **Tercero. Características de las prestaciones**

---

Con respecto al objeto del Contrato, se admiten licencias de uso de software en modalidad de nube, con los requisitos funcionales detallados a continuación:

### **1. Requisitos funcionales de las licencias de uso a suministrar**

La Comunidad de Madrid, dentro del Plan de Salud Digital de la Consejería de Digitalización, tiene en curso varias iniciativas que persiguen la definición, diseño y desarrollo de nuevos sistemas analíticos con modelos predictivos que permitan extraer valor de los datos existentes, sanitarios y no sanitarios, del Servicio Madrileño de Salud; como por ejemplo: optimizar la gestión global y local de la capacidad, alerta temprana de variabilidad en la práctica clínica, predicción, seguimiento y actuación en el ámbito de la cronicidad, detección temprana de desajustes y alertas epidemiológicas, todo ello para su aplicación en la toma de decisiones.

Esencialmente, para gestionar y procesar volúmenes ingentes de datos y para detectar patrones en los mismos, se utilizan tecnologías denominadas como *Big Data*, las cuales son capaces de utilizar dichos datos para extraer conclusiones de valor.

Actualmente se dispone de un repositorio central a la fecha, con más de 13.000 millones de registros (*Data Lake*) que se sustenta en una infraestructura de *Big Data* basada en Cloudera (Data Platform private Cloud (CDP)) para el tratamiento de información, en el CPD Central de la DGSD Athene@ en el Hospital 12 de octubre. Está estructurado modularmente para que los procesos de alimentación e ingesta, así como de explotación de datos, permitan interconexión a otros subsistemas de análisis de datos. Además, este *Data Lake* permitirá la interoperabilidad de informaciones con la Historia Clínica del Sistema Nacional de Salud, proyecto que promueve el Ministerio de Sanidad para mejorar el acceso a la información.

Por tanto, la actual infraestructura está compuesta de varios nodos que disponen implementados los distintos roles en nodos diferenciados.

Debido a las necesidades de crecimiento en función de los organismos y departamentos que consuman y realicen modelos de IA avanzada en la plataforma, se ve necesario hibridar la plataforma de Cloudera con su versión Cloud. En este sentido, se utilizará la nube de AWS en el entorno seguro que la misma tiene para los servicios de Madrid Digital.

Esta hibridación permite el crecimiento puntual para consumo de datos y desarrollo de analítica y analítica avanzada, con la capacidad de inferencia de modelos de IA y IA generativa, así como la orquestación de agentes de IA.

Todo desarrollo realizado en la plataforma *on premise* puede ejecutarse en la plataforma Cloud y viceversa, pudiendo además sincronizar los datos necesarios entre entornos.

Para ello, se van a desplegar en nube las siguientes funcionalidades de Cloudera en *Cloud*:

- **Datalake** en modalidad Like Duty. Para almacenar los datos y los servicios comunes de Cloudera on Cloud.
- **Cloudera Data Flow** para gestionar las cargas de ingesta y transformación sencilla usando Apache Nifi sobre Kubernetes.
- **Cloudera Data Engineering**, para realizar las cargas ETL o jobs de spark orquestados a través de una interfaz visual de Apache Airflow.
- **Cloudera Data Warehouses** para tener motores de analítica aislados por cores y memorias para Apache HIVE, Apache IMPALA y TRINO.
- **Cloudera AI Workspaces** para disponer de un entorno de analista e IA donde desarrollar, entrenar, experimentar y desplegar AMPs (ejemplos de IA de más de 61 proyectos github documentados y listos para usar) así como el uso de los múltiples runtimes y librerías incluyendo el uso de GPU y diferentes notebooks. Además están disponibles los AI Studios como: RAG Studio, Synthetic Data Studio, Agent Studio y Fine Tuning Studio.
- **Cloudera AI Inference** para desplegar los modelos usando el entorno NVIDIA NIM con las librerías optimizadas de NVIDIA para los modelos más comunes.

Estos componentes se desplegarán en los servicios y unidades de tiempo y cómputo indicados a continuación a nivel mensual, que habría que multiplicarlos por 12 para tener las capacidades anuales equivalentes:

## DATALAKE (LAGO DE DATOS)

Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.
FreeIPA Service	m5.large	2	730
	gp3	100	730
IDBroker Service	t3.medium	1	730
	gp3	200	730
DataLake Master Node (CDP cluster)	r5.2xlarge	1	730
	gp3	712	730
RDS DB Instance (PostgreSQL)	db.m5.large-PostgreSQL-Single-AZ	1	730
Default Compute Cluster	m5a.xlarge	2	730
	gp3	40	730

## CLOUDERA DATA FLOW (FLUJO DE DATOS DE CLOUDERA)

Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.			
AWS Kubernetes Cluster	AmazonEKS	1	730			
Liftie Infrastructure Node	m5a.xlarge	2	730			
	gp3	40	730			
RDS DB Instance (PostgreSQL )	db.m5.large-PostgreSQL-Multi-AZ	1	730			
Data Flow - Compute Nodes **	c5.4xlarge	3	730			
	gp3	1,137	730			
Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.	CCUs/hr	vCores/hr	RAM (GB)/hr
Use Case 02	small	2	730	2.0000	6	12

Teniendo en cuenta que “resource type small” equivale a 3 vcores y 6GB memoria.

## CLOUDERA DATA ENGINEERING (INGENIERÍA DE DATOS DE CLOUDERA)

Purpose Objetivo	Virtual Cluster Type Tipo de clúster virtual	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.
AWS Kubernetes Cluster		AmazonEKS	1	730
Liftie Infrastructure Node		m5a.xlarge	2	730
		gp3	40	730
Liftie DEX Base Node		m5.2xlarge	1	730
		gp3	100	730
RDS DB Instance (MySQL)		db.m5.large-MySQL-Multi-AZ	1	730
NFS for Virtual Clusters		AmazonEFS	500	720
Liftie DEX App Node		m5.2xlarge	3	730
		gp3	100	730
Virtual Cluster 1 - Compute	Core	r5.2xlarge	3	400

Nodes		gp3	100	400
-------	--	-----	-----	-----

## CLOUDERA DATA WAREHOUSE (ALMACÉN DE DATOS DE CLOUDERA)

Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.
AWS Kubernetes Cluster	AmazonEKS	1	730
Database Catalog - Shared Services Nodes	m5.2xlarge	3	730
	gp2	45	730
RDS DB Instance (PostgreSQL )	db.r5.large-PostgreSQL-Multi-AZ	1	730
NFS for CDW	AmazonEFS	50	730
Virtual Warehouse 1 - Compute Nodes	r5d.4xlarge	2	400
	gp2	40	730
Virtual Warehouse 2 - Compute Nodes	r5d.4xlarge	2	400
	gp2	40	730

## CLOUDERA AI WORKSPACES (ESPACIO DE TRABAJO DE CLOUDERA)

Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.
AWS Kubernetes Cluster	AmazonEKS	1	730
Liftie Infrastructure Node (Platform Services)	m5a.xlarge	2	730
	gp3	40	730
Liftie Infrastructure Node (CML Services)	m5.2xlarge	2	730
	gp3	564	730
Model Registry Infrastructure Node (Platform)	m5a.large	2	730
	gp3	40	730
Model Registry Infrastructure Node (MR Services)	m5.2xlarge	1	730
	gp3	129	730
NFS for Workspace	AmazonEFS	1,000	730
CPU - Compute Nodes	m5.4xlarge	5	400

	gp3	512	730
GPU - Compute Nodes	g4dn.2xlarge	4	400
	gp3	512	730

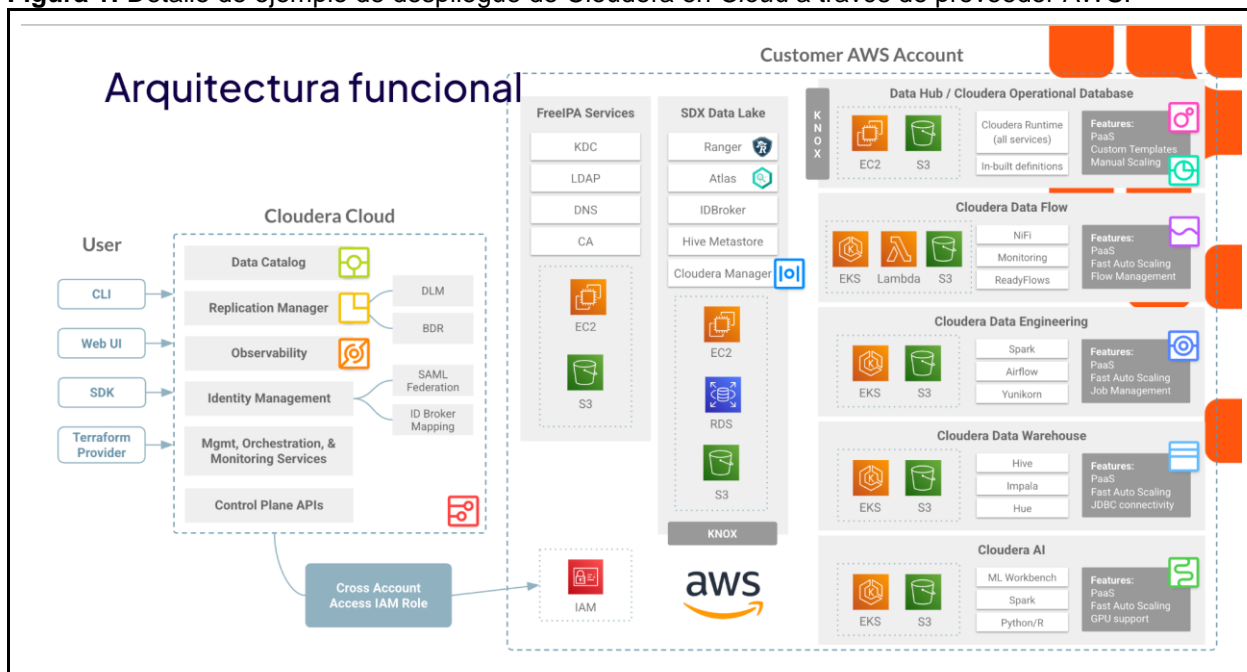
## CLAUDERA AI INFERENCE (INFERENCIA DE CLAUDERA AI)

Purpose Objetivo	Resource Type Tipo de recurso	Resource Count Nº recursos	# of Hours Nº de hrs.
AWS Kubernetes Cluster	AmazonEKS	1	730
Liftie Infrastructure Node (Platform Services)	m5a.xlarge	2	730
	gp3	40	730
Liftie Infrastructure Node (CML Services)	m5.xlarge	2	730
	gp3	64	730
CPU - Compute Nodes	m5.4xlarge	3	730
	gp3	512	730
GPU - Compute Nodes	g4dn.4xlarge	3	730
	gp3	512	730

Estos recursos de AWS tendrán que estar gestionados mediante técnicas de *finops* / reservas, y otros mecanismos de eficiencia de recursos, para reducir el impacto de los consumos de los recursos en AWS.

La siguiente figura proporciona una imagen orientativa de la arquitectura funcional

**Figura 1:** Detalle de ejemplo de despliegue de Cloudera en *Cloud* a través de proveedor AWS.



Fuente: elaboración propia

## **2. Requisitos de seguridad de las licencias de uso de software en la nube**

No se dispone de la categorización de los sistemas de información.

En todo caso, AWS cuenta con el nivel más alto del Esquema Nacional de Seguridad (ENS).

## **3. Servicios de instalación avanzada y soporte a proporcionar por el adjudicatario**

### **Servicios asociados de instalación avanzada del contrato:**

Los servicios de instalación avanzada incluirán:

- Diseño de la implantación de los productos adaptados a los requisitos tecnológicos del SERMAS.
- Instalación lógica, de todos los elementos suministrados, configuración y las pruebas necesarias para su puesta en funcionamiento en las instalaciones del Servicio Madrileño de Salud en base a los requisitos de diseño acordados.
- Entrega de Documentación.
- Transferencia de conocimiento al personal de administración del SERM

### **Servicio de soporte de los licencias de uso de software a suministrar**

## **4. Requisitos de los perfiles profesionales**

Perfil	Certificación
RSA	- CDP Administrator, Public Cloud



	- CDP Data Analyst
--	--------------------

- Interacción con diversos equipos de la Consejería de Digitalización, intermediario entre los administradores y los usuarios de la plataforma
- Trato directo con el cliente para validación de posibles soluciones a problemas de la plataforma, desarrollo de soluciones adhoc, propuestas de solución
- Atención a incidentes causados por la plataforma, atención a usuarios, seguimiento de incidentes, pruebas de concepto, validación del buen funcionamiento de la plataforma
- Habilidades y conocimientos necesarios sobre AWS para coordinar toda la planificación de la hibridación del sistema.

## 5. Modalidad de licenciamiento

Programa	Derechos durante la vigencia				
Creditos CCU para uso en Cloudera Cloud	<ul style="list-style-type: none"><li>• Derecho de uso: Recursos destinados a la Plataforma (Cores, Memoria, almacenamiento)</li><li>• Derecho de actualización: Actualización de versiones y configuraciones que resulten por el fabricante.<ul style="list-style-type: none"><li>◦ Actualización de versiones debido al mantenimiento correctivo, evolutivo, perfectivo y adaptativo del software.</li><li>◦ Acceso a parches, actualizaciones menores y “hot-fixes”.</li></ul></li><li>• Derecho Acceso multicanal: telefónico, e-mail, web de soporte.</li><li>• Derecho de Interacción on-line con técnicos especializados.</li><li>• Derecho de acceso a cuadro de mando con el estado de los productos y su uso.</li><li>• Derecho de acceso a documentación: a una base de conocimiento para la resolución de problemas.</li><li>• Derecho de realizar un número ilimitado de accesos al servicio de apertura de incidencias</li><li>• Derecho de consulta al fabricante (soporte del fabricante):<ul style="list-style-type: none"><li>◦ Horario: Soporte ininterrumpido 24x7</li></ul></li><li>• Derecho de Tratamiento específico para aquellas incidencias con prioridad muy alta.</li></ul>				
	Nivel de gravedad	1 Crítico	2 Alto	3 Medio	4 bajo
		El sistema de producción está inactivo	Fallo de función o característica principal	Fallo de función o Característica menor	Problema menor
Descripción del nivel de gravedad	El sistema de producción está inactivo. El producto no funciona, por lo que se ha generado una disrupción total de la actividad. No existe solución alternativa	Fallo de funcionalidad principal. Las operaciones están muy Restringidas, aunque la actividad puede continuar, pero de forma limitada. Existe	Fallo de funcionalidad menor. El producto no funciona como debiera, por lo que se produce una pérdida de uso menor. Es posible que haya una	No se ha producido pérdida de servicio. Por ejemplo, puede tratarse de una solicitud de documentación , información general o una solicitud de	



		temporal disponible.	una Solución alternativa temporal.	solución Alternativa temporal.	mejora de Software
<b>Objetivos de Tiempos de respuesta</b>					
Asistencia Empresarial 24x7	2 horas	8 horas	12 horas	1 día habil	

## 6. Cobertura de la garantía extendida del adjudicatario

La garantía extendida que debe prestar el adjudicatario durante todo el periodo de vigencia de las licencias se rige por:

- Soporte de nivel 1 y nivel 2 prestado por el adjudicatario a petición del organismo destinatario, en los términos descritos en el PPT;
- Soporte del adjudicatario al organismo para el acceso a la garantía del fabricante (acceso al soporte de nivel 3).
- Soporte a la instalación de actualizaciones.

Horario de contacto: Asistencia Empresarial, 8x5

### Cuarto. Plazo de entrega de las licencias

El plazo máximo de entrega será de diez (10) días naturales contados a partir de la fecha de inicio de ejecución del contrato. No se admiten entregas parciales.

### Quinto. Plazo de ejecución del contrato

El contrato tendrá una duración estimada de once (11) meses, dado que se prevé su formalización el 1 de julio de 2025. No obstante, con independencia de la fecha efectiva de formalización, el contrato deberá finalizar, en todo caso, el 15 de junio de 2026, por lo que la duración definitiva del mismo se ajustará a dicho plazo límite. La eventual reducción del plazo efectivo de ejecución no conllevará una disminución del importe de adjudicación, dado que el presupuesto del contrato se ha calculado en función del cumplimiento de los resultados esperados, con independencia del tiempo requerido para su consecución.

### Sexto. Coordinación del Contrato

A efectos de lograr una correcta ejecución del Contrato, se designa como coordinador del mismo a la responsable del proyecto de HealthData@MAD-R&I en la Fundación para la Investigación e Innovación Biosanitaria de Atención Primaria.

En Madrid, a fecha de última firma electrónica  
La Presidencia del Patronato

Fdo.: Ana Isabel González González  
Jefe de Proyecto.