

## TÉRMINOS DE REFERENCIA

**Objeto de Contratación:** SERVICIO DE INSTALACIÓN Y CONFIGURACIÓN DE HADOOP EN EL INEC C3 RRAA

**Código del Proceso:** EC-INEC-399778-NC-RFB

**Fecha:** 25-06-2025

### A. DATOS GENERALES

CPC RESPECTIVO A NIVEL 9	DESCRIPCIÓN DEL ITEM	CANTIDAD REQUERIDA	PRECIO UNITARIO	SUBTOTAL
512900021	Servicio de Instalación y configuración de Hadoop en el INEC C3 RRAA	1	\$138.955,00	\$138.955,00

- 1. Plazo de Ejecución:** 70 días calendario, contados a partir del siguiente día de la suscripción del contrato.

Durante este período, el proveedor deberá cumplir con todas las fases establecidas en la metodología de trabajo detallada en este documento. El plazo de ejecución del contrato se divide de la siguiente forma:

**Primera etapa:**

- Instalación y configuración del clúster Hadoop: **20** días calendario, contados a partir del día siguiente de la suscripción del contrato.

**Segunda etapa:**

- Migración: **15** días calendario, contados a partir del siguiente día de la finalización de la instalación y configuración del clúster.
- Pruebas de funcionamiento del clúster: **10** días calendario, contados a partir del siguiente día de la finalización de la migración.
- Transferencia de conocimiento: **15** días calendario, contados a partir del día siguiente de finalización de las pruebas de funcionamiento del clúster, pudiendo iniciarse durante la fase de pruebas de funcionamiento.

**Tercera etapa:**

Corresponde al soporte técnico del servicio posterior a la implementación e incluye:

- El soporte de funcionamiento estable y continuo del clúster Hadoop por un periodo de **10** días calendario, contados a partir de la finalización de la transferencia de conocimientos.
- La utilización de 20 horas de soporte especializado para la atención de requerimientos funcionales, revisiones técnicas y operativas del clúster Hadoop implementado, contados a partir de la finalización de la transferencia de conocimientos.

**2. Vigencia de la Oferta:** 120 días calendario.

**3. Forma y cronograma de entregas:**

- ✓ Parcial

**4. Formas de Pago:**

- ✓ **Otra:** El Instituto Nacional de Estadística y Censos – INEC realizará el pago del contrato, de acuerdo con los siguientes hitos:

**Primer Hito: pago del (50%),** Se realizará una vez culminada la fase de instalación y configuración del clúster Hadoop, contra la presentación de:

- Diagrama de arquitectura del clúster
- Procedimientos de inicio/parada de servicios
- Buenas prácticas de administración del clúster
- Informe técnico de instalación y configuración del clúster
- Manual técnico de administración y mantenimiento del clúster.
- Acta entrega-recepción parcial del servicio
- Factura correspondiente

**Segundo Hito: pago del (40%),** Se efectuará contra la entrega parcial y satisfactoria del servicio, que incluye las fases de migración, pruebas de funcionamiento del clúster y transferencia de conocimiento, debidamente documentadas y respaldadas con los siguientes entregables:

- Informe técnico de migración.
- Informe de resultados de las pruebas de funcionamiento.
- Informe de la transferencia de conocimiento. (Acompañado del registro de asistencia).
- Acta de entrega-recepción parcial del servicio.
- Factura correspondiente.

**Tercer Hito: pago del (10%):** Se realizará contra la presentación del informe técnico correspondiente a la utilización de las 20 horas de soporte técnico especializado, establecidas en la metodología de trabajo como parte del soporte técnico posterior a la implementación.

Para este hito se entregará:

- Informe técnico detallado del uso de horas de soporte técnico especializado, en el que deberá incluir las actividades realizadas, incidencias atendidas, ajustes y mejoras ejecutadas.
- Acta de entrega-recepción definitiva del servicio.
- Factura correspondiente.

#### 5. Tipo de presupuesto: *Inversión*

- ✓ Código único del proyecto "CUP": 31210000.0000.388017
- ✓ Descripción del proyecto de inversión: *Robustecimiento de la Producción Estadística del Ecuador*

#### 6. Garantía de fiel cumplimiento

En el caso de resultar ganador, para seguridad del cumplimiento del contrato y para responder por las obligaciones que contrajeran a favor de terceros, relacionadas con el contrato, el adjudicatario, antes o al momento de la firma del contrato, rendirá garantías o póliza de seguros emitida por una compañía debidamente registrada en el país por un monto equivalente al cinco (5%) por ciento del valor de aquel. Con cargo a la garantía de fiel cumplimiento se podrá efectivizar lo establecido en la normativa legal vigente.

**7. Lugar de entrega:** En la ciudad de Quito en el Edificio del INEC Planta central ubicado en las calles Juan Larrea N15-36 y Jose Riofrio.

### B. REQUERIMIENTO DE LA CONTRATACIÓN

#### 1. ANTECEDENTES:

El Instituto Nacional de Estadística y Censos - INEC dentro de sus procesos de generación de estadísticas oportunas para el país, debe mantener actualizada su estructura informática que permita a todas las unidades productoras del INEC acceder a información de registros administrativos procesados para el uso estadístico de manera oportuna y ágil.

Hasta el año 2015, las consultas y procesamientos de datos tardaban desde 1 hasta 6 días en realizarse. Por esta razón, los técnicos de la Dirección de Registros Administrativos (2015), con una instalación en la infraestructura disponible, realizaron la implementación de dos clústeres de procesamiento de datos basados en Hadoop (código abierto) con el Framework Hortonworks, el cual permite administrar los clústeres de procesamiento de datos de manera centralizada y rápida. Hasta el tiempo actual se tienen varios componentes de código abierto instalados en los clústeres de procesamiento con el framework Hortonworks (HDFS, YARN,

MapReduce, Hive, Pig, Ranger, Spark2) y más componentes que son importantes ya que permiten administrar los clústeres y emitir alertas tempranas. Para realizar esta implementación no se requirió ninguna contratación externa.

En la actualidad, este procesamiento de registros administrativos se realiza de manera masiva con las funcionalidades del software libre Hadoop, lo cual ha contribuido en tener una distribución en el cómputo y almacenamiento dando respuestas oportunas a los requerimientos institucionales y externos. Con Hadoop las consultas responden en cuestión de minutos con bases de datos que llegan a los 700'000.000 registros, pues el sistema de partición de discos designados para el procesamiento mejora la distribución de la carga de información.

Por otro lado, se debe considerar que, en el año 2018, la empresa Hortonworks tuvo una fusión con la empresa Cloudera Inc., lo que ocasionó el bloqueo de los accesos a los links públicos que estaban disponibles de forma libre para instalación de nuevos clústeres basados en Hortonworks; también dejaron de publicar nuevas mejoras tanto en la administración de los clústeres como de las herramientas de procesamiento de información y representación gráfica de tipo big data.

Debido a los grandes cambios expuestos anteriormente, hasta el año 2018 se pudieron utilizar las herramientas de Hortonworks de forma libre, pero al momento no se puede descargar e instalar clústeres de Hadoop para el procesamiento masivo de datos en la institución, por lo cual es necesario contratar un servicio externo con profesionales especializados para instalar, configurar e implementar un clúster de Hadoop que permita distribuir el almacenamiento y paralelizar el procesamiento de grandes cantidades de datos provenientes de fuentes administrativas.

Sin disponer de esta arquitectura distribuida como Hadoop, los tiempos de respuesta en el procesamiento se afectarían seriamente, como se lo llevaba antes del año 2015. No se contaría con la distribución de carga para el procesamiento y los procesos electrónicos y scripts desarrollados por aproximadamente ocho (8) años deberían ser reconstruidos para adaptarse a otras herramientas involucrando uso de recursos adicionales que actualmente no dispone el INEC. Por tal razón la contratación del Servicio de Instalación y Configuración de Hadoop en el Proyecto de "Robustecimiento de la Producción Estadística del Ecuador".

Con documento 102-2022-BM-LC6-EC de 1 de julio de 2022, se notifica al Ministerio de Economía y Finanzas que el 30 de junio de 2022, el Directorio Ejecutivo del Banco Internacional de Reconstrucción y Fomento (BIRF) ha aprobado un Préstamo a la República del Ecuador para el Proyecto "Fortalecimiento del Sistema Estadístico del Ecuador", por un monto de USD \$80,00 millones, el cual será efectivo mediante la firma de un acuerdo de préstamo entre el INEC y Banco Mundial para lo cual el INEC debe cumplir con las condiciones de firma y efectividad.

Mediante Oficio Nro. SNP-SNP-SGP-2022-0111-O, de 14 de septiembre de 2022 la Secretaría Nacional de Planificación emite el dictamen de prioridad al Proyecto: "Robustecimiento de la Producción Estadística del Ecuador", con CUP: 31210000.0000.388017, por el Período Agosto 2022 – Septiembre 2026.

Con oficio MEF-SP-2022-0982, de 19 de octubre de 2022, la Subsecretaría de Presupuesto del Ministerio de Economía y Finanzas comunica al INEC que "(...) durante la ejecución del Presupuesto General del Estado asignará al Instituto Nacional de Estadística y Censos el espacio presupuestario requerido para continuar con el proceso de obtención de dictamen de incremento al PAI 2022 del proyecto "Robustecimiento de la Producción Estadística del Ecuador" con CUP 31210000.0000.388017 por un total de USD 247.936,16 para el año 2022 por USD 61.984,04 y para el año 2023 por USD 185.952,12; es preciso señalar que el organismo y correlativo se asignará una vez que se suscriba el crédito (...)".

Con fecha 26 de octubre de 2022 se firmó el "Loan Agreement" (contrato acuerdo de préstamo) entre la República del Ecuador y el Banco Internacional de Reconstrucción y Fomento (BIRF) para financiar el "Fortalecimiento del Sistema Estadístico del Ecuador". El acuerdo de préstamo consta de cuatro (4) partes o componentes a financiar. El Proyecto: "Robustecimiento de la Producción Estadística del Ecuador" con CUP: 31210000.0000.388017, se financia con la Parte 3 (componente C3) del "Loan Agreement" (página 7): "Part 3: Strengthening the statistical production from Administrative Records and INEC managerial capacity", en español: Robustecimiento de la producción estadística a partir de los registros administrativos y de la capacidad de gestión del INEC."

Mediante memorando Nro. INEC-INEC-2024-0952-M del 23 de diciembre de 2024, a través del cual se indica (...) *"procede a aprobar el Plan de Dirección de Proyecto " C6 - Aprovechamiento de Registros Administrativos para la Producción Estadística" 2025; a fin de que prosigan con el trámite respectivo, en el ámbito de sus competencias de acuerdo a la normativa y más disposiciones emitidas para el efecto" (...).*

Mediante Memorando Nro. INEC-INEC-2025-0048-M del 30 de enero de 2025, el Director Ejecutivo del INEC emitió la aprobación del Plan Estratégico de Tecnologías de la Información (PETI), a ejecutarse en el año 2025, a fin de que prosiga con el trámite respectivo en el marco de las disposiciones y normativa emitida para el efecto. En cual consta el proceso de instalación y configuración de Hadoop en el INEC.

Mediante correo electrónico de 25 de abril de 2025, Banco Mundial señala lo siguiente: *"(...) Hemos recibido el borrador del Plan de Adquisiciones (PA) del proyecto anteriormente mencionado, enviado al Banco el 2025/04/23. Sobre la información proporcionada, el Banco no tiene objeción a las actividades marcadas como "Aprobada" en el Plan de Adquisiciones. (...)"*.

Mediante correo electrónico de 28 de mayo de 2025, la Dirección de Registros Administrativos, solicita a DITIC: *"(...)*

- *La revisión técnica del contenido del documento adjunto.*
- *La validación de la factibilidad y pertinencia técnica de los requerimientos establecidos.*
- *Y, en caso de no existir observaciones, la suscripción del documento como constancia de conformidad (...)*

Mediante correo electrónico de 29 de mayo de 2025, la Dirección de Tecnologías de la Información y Comunicación del INEC dio respuesta al pedido del área requirente de validación de las características técnicas y componentes requeridos para el proceso de prestación de servicio de instalación y configuración de Hadoop en el INEC C3 RRAA, indicando lo siguiente:

*“Se informa que la Gestión de Infraestructura de TI no se tiene el conocimiento y las competencias sobre “aspectos técnicos requeridos para garantizar el fortalecimiento de las operaciones de procesamiento de grandes volúmenes de datos en el entorno analítico del INEC”. Por lo que no se puede emitir observaciones o validar el documento enviado.”*

Con fecha 11 de junio de 2025, Se emitió una solicitud de validación de al menos tres perfiles con conocimientos técnicos a la Dirección de Administración de Recursos Humanos del INEC. Dicha solicitud se enmarcó en la necesidad de asegurar la disponibilidad de talento humano interno con conocimientos sólidos y experiencia comprobada en instalación y configuración de Hadoop, para la implementación del “Servicio de Instalación y configuración de Hadoop en el INEC C3 RRAA”.

En respuesta a esta solicitud la Directora de la Dirección de Administración de Recursos Humanos del INEC indicó lo siguiente:

*“La Dirección de Administración de Recursos Humanos del Instituto Nacional de Estadística y Censos-INEC certifica que una vez revisada la base de servidores de la Institución, no existen servidores que posean un nivel de especialización y experiencia técnica específica para la instalación y configuración de un clúster Hadoop, con el siguiente nivel de experticia:*

- *Arquitectura y componentes avanzados de Hadoop (HDFS, YARN, MapReduce, Hive, Spark, etc.).*
- *Instalación, configuración y puesta en marcha de clústers Hadoop a escala de producción, incluyendo aspectos de seguridad, rendimiento y alta disponibilidad.*
- *Gestión, optimización y resolución de problemas complejos en entornos distribuidos Hadoop.*
- *Conocimiento de las mejores prácticas y estándares de la industria para implementaciones de Big Data.*
- *Poseer Certificación en administración de plataformas Big Data en: HDP y HDF en hortonworks y CDP en Cloudera.”*

Mediante memorando Nro. INEC-DIPLA-2025-0504-M del 16 de junio de 2025, la Dirección de Planificación y Gestión Estratégica emitió la Socialización de la Programación Anual de la Planificación - PAP al 15 de junio de 2025, donde consta el servicio a ser contratado.

## 2. OBJETIVO

Contratar la prestación del servicio de instalación, configuración, migración, puesta en marcha y soporte especializado de un clúster de procesamiento de datos basado en Hadoop, en un entorno virtualizado del nuevo servidor de procesamiento adquirido recientemente por el Instituto Nacional de Estadística y Censos – INEC, que permita modernizar la Plataforma Analítica institucional, mediante una arquitectura escalable, tolerante a fallos y de alto rendimiento, con el fin de garantizar la disponibilidad, continuidad y soporte al crecimiento de los servicios analíticos y transaccionales que el INEC ofrece para la generación de productos estadísticos con valor agregado, dirigidos tanto al ámbito público como privado.

## 3. ALCANCE

El servicio comprende la instalación, configuración, migración, puesta en marcha y horas de soporte especializado de un clúster Hadoop en un entorno virtualizado del servidor de procesamiento del INEC, incluyendo su integración con Oracle, PostgreSQL y Pentaho Data Integration, la validación técnica mediante pruebas de funcionamiento, la transferencia de conocimientos al personal designado y la entrega de la documentación técnica correspondiente, garantizando de esta manera la continuidad del servicio de procesamiento de datos en un ambiente optimizado y de alto rendimiento.

## 4. METODOLOGÍA DE TRABAJO

El desarrollo del servicio se realizará en fases claramente definidas, conforme a los objetivos planteados y en coordinación con el administrador del contrato designado por el INEC. A continuación, se describen las fases del trabajo:

Nota: Toda la documentación presentada como entregable en cada una de las fases a continuación descritas deberá ser revisada y aprobada por el administrador del contrato, como condición necesaria para validación de cada fase y el avance correspondiente del servicio.

El contratista proporcionará un líder de proyecto que trabajará con el administrador del contrato designado por el INEC, con quien coordinará el detalle de las actividades, modalidad de la transferencia de conocimientos, número de participantes, recursos, fechas, horarios, entre otros aspectos operativos.

El contratista deberá contar con personal técnico que diseñe la arquitectura óptima del clúster Hadoop, validando los requisitos del sistema y pre-requisitos de instalación, considerando su implementación en un entorno virtualizado. Se deberá analizar la capacidad de los nodos master y esclavo, el tamaño estimado de los datos a procesar o

exportar, y con base en ello, ajustar las necesidades de almacenamiento, procesamiento y administración.

### 1. Fase de instalación y configuración del clúster Hadoop

Durante 20 días calendario, a partir del día siguiente a la suscripción del contrato, el contratista deberá:

- Instalar y configurar versiones estables y disponibles del software de cada componente de Hadoop.
- Instalar y configurar los servicios y componentes descritos en el apartado “6. SERVICIO ESPERADO”.
- Incluir librerías, controladores y componentes necesarios para la interacción del ecosistema Hadoop.
- Instalar y Configurar seguridades del clúster.
- Instalar y Configuración de auditorías.
- Implementar una solución para inventario periódico de bases de datos sin costos/licencias adicionales.
- Implementar una solución de monitoreo de métricas de los componentes implementados sin costos/licencias adicionales.

Entregable de esta fase gestionada por contratista:

- Diagrama de arquitectura del clúster
- Procedimientos de inicio/parada de servicios
- Buenas prácticas de administración del clúster
- Informe técnico de instalación y configuración del clúster.
- Manual técnico de administración y mantenimiento del clúster.

### 2. Fase de migración del ecosistema Hadoop actual

Durante 15 días calendario, posterior a la instalación y configuración, se deberá realizar la migración de:

- Datos de los clústeres actuales (aproximadamente 8 TB).
- Funciones Definidas por el Usuario (UDF). Basados en lenguaje de programación en Java.
- ETLs existentes, asegurando conectividad con el nuevo clúster.

Entregable de esta fase gestionada por contratista:

- Informe técnico de migración.

#### 2.1 Fase de pruebas de funcionamiento

Durante 10 días calendario, el proveedor realizará pruebas y ajustes necesarios para garantizar el correcto funcionamiento del clúster.

Entregable de esta fase gestionada por contratista:

- Informe de resultados de las pruebas de funcionamiento.

## 2.2 Fase de transferencia de conocimientos

Una vez finalizadas las pruebas del correcto funcionamiento del clúster, durante 15 días calendario, se impartirá la transferencia de conocimiento, según el plan presentado en la oferta. Esta fase podrá iniciarse parcialmente durante la fase de pruebas, siempre que no interfiera en los resultados. La modalidad, participantes, fechas y horarios serán previamente definidos en coordinación con el administrador del contrato.

Entregable de esta fase gestionada por contratista:

- Informe de transferencia de conocimientos. (Acompañado del registro de asistencia).

## 3. Soporte técnico del servicio posterior a la implementación

Durante 10 días calendarios, contados a partir de la finalización de la transferencia de conocimientos, se iniciará la etapa de soporte técnico posterior a la implementación, la cual tiene como finalidad garantizar la estabilidad operativa del clúster hadoop implementado incluyendo servicios y componentes instalados y brindar soporte técnico especializado post entrega.

Durante este periodo el proveedor deberá proporcionar un total de 20 horas de soporte técnico especializado, destinadas a la atención de requerimientos funcionales, revisiones técnicas, operativos y/o de configuración relacionados al clúster implementado, y podrá ser utilizado por el INEC de manera flexible, según las necesidades.

El soporte técnico, se prestará bajo un esquema de atención 7x24x365 y podrá ser solicitado en modalidad remota y/o presencial, según necesidad que defina el administrador de contrato, de acuerdo con lo detallado en el numeral 6 del presente documento, específicamente en el apartado "SOPORTE TECNICO DEL SERVICIO".

Entregable de esta fase gestionada por contratista:

- Informe técnico detallado del uso de horas de soporte técnico especializado, incluyendo las actividades realizadas, incidencias atendidas, ajustes y mejoras ejecutadas.

## 5. INFORMACIÓN QUE DISPONE LA ENTIDAD

El Instituto Nacional de Estadística y Censos - INEC cuenta y pondrá a disposición del proveedor un servidor físico, con las siguientes características:

Características del Hardware y licenciamiento disponible para la instalación y configuración del clúster Hadoop en un entorno virtualizado	
Memoria RAM disponible:	1024 GB (16 x 64 TruDDR5 Performance+ 5600 MHz RDIMM)
Almacenamiento disponible:	Cinco (5) Discos de Estado Sólido (ThinkSystem 2,5" S4520 7,68 TB, lecturas intensivas, SATA 6 Gb de intercambio en caliente). <b>Capacidad Usable Aproximada de 22.1TB</b>
Procesadores instalados:	dos (2) Procesadores Intel Xeon Gold 6526Y 16C 195 W 2,8 GHz
Núcleos disponibles:	32 cores físicos
Virtualización disponible:	VMware estándar edición con licenciamiento y soporte vigente

A continuación se listan los componentes que están instalados actualmente en el ecosistema Hadoop del INEC, con las respectivas versiones:

COMPONENTE	SERVIDOR HP	SERVIDOR KOSTAT	DESCRIPCIÓN
	VERSIÓN		
HDFS	2.7.3	2.7.3	Sistema de archivos distribuido Apache Hadoop
YARN	2.7.3	2.7.3	Uno de los componentes principales de Apache Hadoop, es responsable de asignar recursos del sistema a las distintas aplicaciones que se ejecutan en un clúster Hadoop y de programar tareas que se ejecutarán en diferentes nodos del clúster
MapReduce2	2.7.3	2.7.3	Sistema de procesamiento de datos paralelo y distribuido
Tez	0.7.0	0.7.0	Es el marco de procesamiento de consultas Hadoop de próxima generación escrito sobre YARN
Hive	1.2.1000	1.2.1000	Sistema de almacenamiento de datos para consultas y análisis ad-hoc de grandes conjuntos de datos y servicio de gestión de almacenamiento y tablas
HBase	1.1.2	1.1.2	Una base de datos distribuida no relacional, más Phoenix, una capa SQL de alto rendimiento para aplicaciones de baja latencia
Pig	0.16.0	0.16.0	Plataforma de secuencias de comandos para analizar grandes conjuntos de datos
Sqoop	1.4.6	1.4.6	Herramienta para transferir datos masivos entre Apache Hadoop y almacenes de datos estructurados como bases de datos relacionales

Oozie	4.2.0	4.2.0	Sistema de coordinación del flujo de trabajo y ejecución de trabajos de Apache Hadoop. Esto también incluye la instalación de la consola web Oozie opcional que se basa en la biblioteca ExtJS y la instalará
ZooKeeper	3.4.6	3.4.6	Servicio centralizado que proporciona una coordinación distribuida altamente confiable
Ambari Metrics	0.1.0	0.1.0	Un sistema para la recopilación de métricas que proporciona capacidad de almacenamiento y recuperación para las métricas recopiladas del clúster
Ranger	0.7.0	0.7.0	Es un marco para habilitar, supervisar y administrar la seguridad integral de los datos en toda la plataforma Hadoop
SmartSense	1.4.5.2.6.2.0-155	1.4.5.2.6.2.0-155	Ayuda a recopilar rápidamente la configuración, las métricas y los registros de los servicios HDP comunes que ayudan a solucionar rápidamente los casos de soporte y recibir recomendaciones específicas del clúster
Spark2	2.3.0	2.3.0	Es un motor rápido y general para el procesamiento de datos a gran escala
Zeppelin Notebook	0.7.0	0.7.3	Permite crear documentos atractivos, interactivos y basados en datos con SQL, Scala, R o Python directamente en su navegador
Slider	0.92.0	0.92.0	Un marco para implementar, administrar y monitorear aplicaciones distribuidas existentes en YARN

A continuación se describe las funciones definidas por el usuario UDF (desarrolladas en el INEC) y que se requieren sean migradas:

UDF	Descripción del proceso
Validación y Corrección de Cédula	validación y corrección
Validación y Corrección de RUC	validación y corrección
Validación y Corrección de Fecha completa	validación y corrección
Validación y Corrección de número con decimal	validación y corrección
Validación y Corrección de Números enteros	validación y corrección
Validación y Corrección de una cadena	validación y corrección
Validación de Correo Electrónico	validación y corrección
Validación de Letras y Números	validación y corrección
Función formatvar	Estandarización
Función formatvarcs	Estandarización
function tokensCache	Limpieza de datos
function limpiarCache	Limpieza de datos
function algoritmo.CompSimilitud	Para precision

function algoritmo.JaroWinkler	Para precision
function algoritmo.Similarity	Para precision

El INEC requiere contratar el “SERVICIO DE INSTALACIÓN Y CONFIGURACIÓN DE HADOOP EN EL INEC C3 RRAA” para contar con actualizaciones estables y disponibles del software de cada uno de los componentes que tiene actualmente instalados, así como también requiere que se instalen nuevos componentes que ayuden a optimizar el procesamiento masivo de información.

Al contar con actualizaciones estables y disponibles del software de cada uno de los componentes, se podrá solventar y soportar el crecimiento y la transaccionalidad de los servicios que produce y proporciona el INEC para generar productos estadísticos con valor agregado que serán utilizados en el ámbito público y privado.

## 6. SERVICIO ESPERADO

A continuación se detalla de las características del servicio a contratarse:

<b>REQUERIMIENTO DEL CLÚSTER HADOOP</b>
<b>INSTALACIONES Y CONFIGURACIONES MINIMAS</b>
El Clúster de procesamiento de datos Hadoop deberá ser instalado y configurado con parámetros de optimización (tuning) orientados al desempeño sobre un entorno virtual mínimo de 12 nodos. La configuración debe considerar balanceo de carga, garantizando el rendimiento, la estabilidad y la alta disponibilidad requeridos para el entorno Hadoop.
En el presente servicio No se incluye la adquisición de licencias de software ni hardware adicional. Todo el trabajo deberá realizarse utilizando los recursos tecnológicos disponibles en la infraestructura del INEC.
Versión requerida: Apache Hadoop 3.4.1 o superior
Versión Java requerida: Java Development Kit (JDK) versión 8, 11 o 17, compatible con la versión de Hadoop a instalar (preferiblemente Open JDK).
El clúster de procesamiento de Hadoop, deberá ser instalado sobre nodos que operen con sistemas operativos Linux de 64 bits (open source), preferentemente versiones estables y que garanticen el correcto funcionamiento del clúster implementado. La implementación se realizará sobre la infraestructura virtual convergente del INEC, basada en VMWare, la cual será provista y habilitada por la institución.
El clúster de procesamiento de datos Hadoop deberá garantizar la integración funcional y operativa con los sistemas de gestión de bases de datos Oracle y PostgreSQL, así como con la herramienta de integración de datos Pentaho Data Integration (PDI), permitiendo el intercambio, procesamiento y orquestación eficiente de datos entre estas plataformas.
La instalación y configuración del clúster debe basarse en buenas prácticas, validar requerimientos del sistema y verificar prerequisites de instalación.

Se deberá incluir la instalación, configuración y validación de todas las librerías, controladores y componentes necesarios para asegurar la interacción entre los distintos servicios del ecosistema Hadoop. Así como también instalar y configurar seguridades auditorías del clúster.	
La plataforma debe poseer la capacidad de permitir la administración por consola, configuración y monitoreo de los principales servicios como HDFS, YARN, MapReduce, Hive y Spark.	
Se deberá implementar una solución para inventario periódico de bases de datos sin costos/licencias adicionales.	
Se deberá implementar una solución de monitoreo de métricas de los componentes implementados sin costos/licencias adicionales.	
DISTRIBUCION DE NODOS	
Máquina de Administración/Supervisión	una (1) máquina virtual con:  4 GB de RAM, 52 GB de almacenamiento, 2 Cores de procesamiento
Máquinas Maestras (Master Nodes)	dos (2) máquinas virtuales: Cada una con:  9 GB de RAM, 100 GB de almacenamiento, 3 Cores de procesamiento
Máquinas Esclavas (Worker Nodes)	nueve (9) máquinas virtuales: Cada una con:  100 GB de RAM, 1.732 GB de almacenamiento, 5 Cores de procesamiento
COMPONENTES Y SERVICIOS MINIMOS	
Servicios mínimos:	HDFS
	YARN
	MAP REDUCE
	Instalar Hive
	Instalar Tez
	Instalar Zookeeper
	Instalar Spark
	Instalar HUE
	Instalar Zeppelin
	Instalar Sqoop
	Instalar Ranger
Instalar Solr	
REQUERIMIENTO GENERAL PARA EL CLÚSTER HADOOP	

El proveedor debe realizar los servicios de configuración, pruebas y puesta en producción del Clúster, bajo los estándares de los fabricantes de software.
El proveedor realizará la configuración de red y comunicación interna entre las máquinas virtuales del clúster.
El proveedor debe verificar la Optimización de parámetros de rendimiento del clúster para uso intensivo de datos.
El proveedor debe realizar la Configuración de servicios principales de Hadoop, tales como HDFS, YARN, MapReduce, y otros que se requieran.
El proveedor debe realizar la migración de los datos que se encuentran en los clústeres de Hadoop actuales (Tamaño Aproximado de 8 TB en ambos clústeres), incluyendo las funciones definidas por el Usuario – UDF (Basados en lenguaje de programación Java) y Actualización y conectividad de los ETLs actuales con nuevo clúster.
El proveedor deberá realizar Pruebas de funcionamiento y desempeño.
Los costos extras que demanden la instalación, configuración y puesta en funcionamiento del clúster Hadoop implementado, así como de la movilización del personal técnico y de servicios adicionales serán asumidos por el proveedor sin costo adicional para el INEC.
<b>TRANSFERENCIA DE CONOCIMIENTOS</b>
El proveedor brindará transferencia de conocimiento, para el personal tanto permanente u ocasional, encargado de la operación del clúster a ser implementado.
Las fechas y horas correspondientes las coordinarán en conjunto el proveedor y el administrador de contrato. El proveedor como parte de la oferta deberá adjuntar el Plan de transferencia de conocimientos.
El proveedor deberá incluir transferencia de conocimiento de 30 horas para mínimo (3) funcionarios del INEC. Con material entregado en formato digital.
La transferencia de conocimientos deberá ser realizada de manera presencial en la ciudad de Quito en las instalaciones del INEC (planta central). No obstante, el administrador de contrato podrá definir que esta se lleve a cabo de manera virtual.
<b>SOPORTE TÉCNICO DEL SERVICIO</b>
El proveedor deberá proporcionar un total de 20 horas de soporte técnico especializado, destinadas a la atención de requerimientos funcionales, revisiones técnicas, operativos o de configuración relacionados al clúster implementado, y podrá ser utilizado por el INEC de manera flexible, según las necesidades.

El soporte técnico podrá ser solicitado en modalidad remota y/o presencial, según necesidad que defina el administrador de contrato, y deberá cubrir:

- Atención a incidentes y errores de configuración.
- Asistencia en tareas de administración y ajustes del clúster.
- Recomendaciones de optimización según evolución del uso.
- Fallas Técnicas atribuidas al proveedor. - El proveedor deberá corregir, sin costo cualquier problema técnico del clúster implementado como errores de configuración, instalación incorrecta, integración fallida entre componentes del ecosistema.
- Corrección de errores y fallas. - El proveedor deberá corregir, sin costo cualquier problema derivado de defectos en la configuración o implementación del entorno durante toda la vigencia del contrato.
- Estabilidad del servicio. - El proveedor debe garantizar que la instalación entregue un entorno estable, escalable, capaz de ser administrado por el equipo del INEC tras su entrega.

El soporte técnico se prestará bajo un esquema de atención 7x24x365 (7 días a la semana, 24 horas al día, 365 días al año). El detalle de requerimientos de tiempos de atención por cada nivel de criticidad está especificado en el SLA a continuación detallado, que es de cumplimiento obligatorio para el proveedor. Estos tiempos se consideran luego de haber sido reportado el incidente y contar con un número de ticket:

Severidad	Inicial (Hardware y Software)	Modalidad
Critica	Máximo 1 hora posterior a la comunicación, bajo la modalidad 24x7x365	En sitio o remoto
Alta	Máximo 2 horas posterior a la comunicación, bajo la modalidad 24x7x365	En sitio o remoto
Media	Máximo 3 horas posterior a la comunicación, bajo la modalidad 24x7x365	En sitio, remoto o teléfono
Baja	Máximo 4 horas posterior a la comunicación, bajo la modalidad 24x7x365	En sitio, remoto o teléfono

El proveedor mantendrá informado al INEC sobre el estado/progreso en la resolución del caso reportado vía telefónica o correo electrónico, utilizando el número de ticket con el que se levantó el incidente.

### C. REQUISITOS MÍNIMOS

Para la verificación del cumplimiento de los requisitos mínimos se utilizará la metodología “PASA O NO PASA”, de acuerdo al siguiente cuadro:

PARÁMETRO	PASA	NO PASA	OBSERVACIONES
Presentación de Formularios			
Capacidad Financiera			
Experiencia específica mínima			
Cumplimiento de Término de Referencia			
Personal técnico mínimo requerido			
Experiencia personal mínimo			

**1. Presentación de formularios:** El oferente deberá presentar todos los formularios previstos en la Solicitud de Oferta de acuerdo a la normativa del Banco Mundial y requisitos mínimos contenidos en los documentos de la licitación (DDL).

**2. Capacidad Financiera:** El Licitante deberá proporcionar prueba documental que demuestre que cumple los siguientes requisitos financieros:

Presentación de declaraciones de impuesto a la renta de los últimos tres (3) años (2022-2024), debidamente presentados ante el Servicio de Rentas Internas, deberá demostrar que en la sumatoria de sus ventas de prestación de servicios tuvieron al menos del 50% del presupuesto referencial sin incluir impuestos.

**3. Experiencia específica del oferente:** Dentro de los últimos 5 años, el oferente deberá demostrar experiencia en la prestación de servicios relacionados con instalación, configuración o gestión de tecnologías orientadas a Big Data mediante la instalación y configuración de Hadoop ó herramientas con funcionalidades similares; ó, migración, gestion o respaldo de grandes volúmenes de datos de al menos 20TB; ó, proyectos informáticos en general que guarden relación con los requerimientos establecidos en los presentes Términos de Referencia.

Para tal efecto, deberá presentar al menos una (1) copia de acta entrega-recepción definitiva o certificado en el que se encuentre detallado la experiencia, monto, periodo y estar debidamente suscrito.

El ó los proyectos presentados deberán sumar al menos el 50% del presupuesto referencial establecido para este proceso, (valor que no incluye IVA), y deberá evidenciar claramente que el servicio ejecutado se enmarca dentro del alcance técnico requerido.

**4. Cumplimiento de Término de Referencia:** El INEC verificará que los oferentes en la oferta presentada, den cumplimiento expreso y preciso a los Términos de Referencia y demás lineamientos descritos claramente en este documento.

5. Personal técnico:

Personal técnico mínimo

CANTIDAD	FUNCIÓN	NIVEL DE ESTUDIO	TITULACIÓN ACADÉMICA
Uno (1)	Líder de Proyecto	Cuarto Nivel	<ul style="list-style-type: none"> <li>Título de cuarto nivel: En Ciberseguridad ó en Inteligencia de Negocios</li> </ul> <p><b>Medio de verificación:</b></p> <ul style="list-style-type: none"> <li>Copia de título de cuarto nivel o certificado de registro de título emitido por la SENESCYT.</li> </ul>
Uno (1)	Especialista en Big Data	Cuarto Nivel	<ul style="list-style-type: none"> <li>Título de cuarto nivel: Maestría en Visual Analytics ó Big Data o Inteligencia Artificial</li> <li>Certificación en administración de Plataformas Big data en:               <ul style="list-style-type: none"> <li>HDP y HDF en Hortonworks</li> <li>CDP en Cloudera</li> </ul> </li> </ul> <p><b>Medio de verificación:</b></p> <ul style="list-style-type: none"> <li>Copia de título de cuarto nivel o certificado de registro de título emitido por la SENESCYT.</li> <li>El proveedor deberá presentar copia legible de la certificación emitida por una entidad reconocida en tecnologías Big Data, como Apache Hadoop, Cloudera, Hortonworks o equivalente, la cual deberá estar a nombre del profesional propuesto al cargo.</li> </ul>
Uno (1)	Especialista Arquitecto de datos y Migración de Datos	Cuarto Nivel	<ul style="list-style-type: none"> <li>Título de cuarto nivel: Sistemas de Información ó Analytics ó Big Data o Inteligencia Artificial</li> <li>Al menos 60 horas de Capacitación en Administración y Herramientas BIG Data, administración e implementación de Clúster</li> </ul> <p><b>Medio de verificación:</b></p> <ul style="list-style-type: none"> <li>Copia de título de cuarto nivel o certificado de registro de título emitido por la SENESCYT.</li> <li>El proveedor deberá presentar copia legible del certificado sobre la capacitación</li> </ul>

			requerida, la cual deberá estar a nombre del profesional propuesto al cargo.
Uno (1)	Especialista Entorno de Instalación	Cuarto nivel	<ul style="list-style-type: none"> <li>Título profesional de cuarto nivel en carreras como Ingeniería en Sistemas, Redes, Ciencias de la Computación, Tecnologías de la Información o afines.</li> <li>Certificación en administración de sistemas Linux, fundamentos de Big data (IBM, Cloudera), o plataformas relacionadas que respalden su conocimiento en entornos de datos o infraestructura.</li> </ul> <p><b>Medio de verificación:</b></p> <ul style="list-style-type: none"> <li>Copia de título de cuarto nivel o certificado de registro de título emitido por la SENESCYT.</li> <li>El proveedor deberá presentar copia legible de la certificación, la cual deberá estar a nombre del profesional propuesto al cargo.</li> </ul>

#### 6. Experiencia Mínima del personal Técnico:

FUNCIÓN	EXPERIENCIA SIMILAR EN:
Líder de Proyecto	Dentro de los últimos 10 años haber participado en al menos cinco (5) proyectos, como Gerente, Director o Líder, en la gestión de proyectos tecnológicos, preferiblemente en la implementación de clúster Hadoop, para lo cual deberá presentar mínimo un (1) certificado u otro documento que acredite claramente la experiencia requerida en el cargo y cantidad de proyectos participados (5), pudiendo ser diferentes proyectos con el mismo empleador.
Especialista en Big Data	Acreditar al menos 10 años de experiencia como Especialista en Big Data o Arquitecto de Datos, participando en al menos dos (2) proyectos relacionados con la implementación de herramientas basadas en Hadoop, migración de procesos hacia entornos Hadoop, capacitación en el uso de tecnologías Big Data, así como en la administración de clústeres y gestión de flujos de datos (Data Flow), para lo cual deberá presentar mínimo un (1) certificado u otro documento que acredite la experiencia requerida en el cargo y cantidad de proyectos participados mínimo (2), pudiendo ser diferentes proyectos con el mismo empleador.

<p>Especialista Arquitecto de datos y Migración de Datos</p>	<p>Dentro de los últimos 10 años haber participado en al menos dos (2) proyectos, como Especialista en: Arquitecto de datos o Migración de Datos, en la gestión de proyectos tecnológicos, preferiblemente en la implementación de clúster Hadoop, para lo cual deberá presentar mínimo un (1) certificado u otro documento que acredite la experiencia requerida en el cargo y cantidad de proyectos participados mínimo (2), pudiendo ser diferentes proyectos con el mismo empleador.</p>
<p>Especialista Entorno de Instalación</p>	<p>Dentro de los últimos 10 años debe haber participado en al menos cinco (5) proyectos, demostrando participación en al menos uno (1) en capacitación o implementación de procesos hacia un entorno Hadoop. Y al menos (4) proyectos como ingeniero de desarrollo en software, para lo cual deberá presentar mínimo un (1) certificado u otro documento que acredite la experiencia requerida en el cargo y cantidad de proyectos participados mínimo (5), pudiendo ser diferentes proyectos con el mismo empleador.</p>

En el caso de que presente **certificado laboral** no se aceptará que el certificado sea suscrito por la misma persona de la cual se avala su experiencia, en ese caso deberá ser firmado por talento humano o quien haga sus veces.

**Entre todos los proveedores que cumplan con los requisitos mínimos, se procederá a comparar entre los mismos los precios ofertados y se adjudicará al que tenga el precio más bajo.**

#### D. FIRMAS DE RESPONSABILIDAD

	Nombre:	Cargo:	Firma:
Elaborado por:	Carlos Sotomayor C.I. 1712331212	Analista del sistema de registros Administrativos 3	
Revisado por:	Ángel Chiluisa C.I. 1717593626	Responsable Gestión de Transformación e Integración de Registros Administrativos	
Aprobado por:	Paúl Benavides C.I. 1713015459	Director de Registros Administrativos	