

INFORME DE NECESIDAD

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

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 con 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. BASE LEGAL:

Constitución de la República del Ecuador

“Art. 226. Las instituciones del Estado, sus organismos, dependencias, las servidoras o servidores públicos y las personas que actúen en virtud de una potestad estatal ejercerán solamente las competencias y facultades que les sean atribuidas en la Constitución y la ley.

“Art. 227.- La administración pública constituye un servicio a la colectividad que se rige por los principios de eficacia, eficiencia, calidad, jerarquía, desconcentración, descentralización, coordinación, participación, planificación, transparencia y evaluación”.

“Art. 288.- Las compras públicas cumplirán con criterios de eficiencia, transparencia, calidad, responsabilidad ambiental y social. Se priorizarán los productos y servicios nacionales, en particular los provenientes de la economía popular y solidaria, y de las micro, pequeñas y medianas unidades productivas”.

Ley Orgánica del Sistema Nacional de Contratación Pública

“Art.3.- Contratos Financiados con Préstamos y Cooperación Internacional.-. En las contrataciones que se financien, previo convenio, con fondos provenientes de organismos

multilaterales de crédito de los cuales el Ecuador sea miembro, o, en las contrataciones que se financien con fondos reembolsables o no reembolsables provenientes de financiamiento de gobierno a gobierno; u organismos internacionales de cooperación, se observará lo acordado en los respectivos convenios. Lo no previsto en dichos convenios se regirá por las disposiciones de esta Ley”.

Reglamento General a la Ley Orgánica del Sistema Nacional de Contratación Pública:

“Art. 2.- Contratos financiados con préstamos y cooperación internacional.- En la suscripción de los convenios de crédito o de cooperación internacional se procurará la concurrencia de proveedores nacionales. Las contrataciones previstas en el artículo 3 de la Ley Orgánica del Sistema Nacional de Contratación Pública se aplicarán independientemente de que el financiamiento internacional sea total o parcial, siempre que se observen las condiciones previstas en el convenio.”

“Art 10.- Uso de herramientas informáticas.- Las entidades contratantes deben aplicar de manera obligatoria las herramientas informáticas para los siguientes procedimientos: (...)

4. Procedimientos Especiales: (...)

e) Procedimientos financiados con préstamos de organismos internacionales;(…)

Para los procedimientos financiados con préstamos de organismos internacionales y procedimientos de contratación en el extranjero se realizará la publicación de la Información relevante a través de la herramienta Publicación del Portal COMPRASPÚBLICAS. (...)”

Normativa del Banco Mundial Regulaciones de Adquisiciones para Prestatarios en Proyectos de Inversión, Cuarta edición noviembre 2020

El numeral 1.3, de la Sección I “Introducción”, establece que: “(...) las reglas y los procedimientos específicos de adquisiciones que se apliquen en la ejecución de un proyecto estarán determinados por las circunstancias de cada caso, los Principios Básicos de Adquisiciones (optimización de la relación calidad-precio, economía, integridad, adecuado al propósito, eficiencia, transparencia y equidad) guían de manera general las decisiones del Banco en el marco de estas Regulaciones de Adquisiciones”.

El numeral 2.1 “Aplicabilidad”, de la Sección II “Consideraciones Generales”, establece que: “El Convenio Legal rige la relación jurídica entre el Prestatario y el Banco. Las Regulaciones de Adquisiciones son aplicables a la adquisición de Bienes y la contratación de Obras, Servicios de No-Consultoría y Servicios de Consultoría en las operaciones de financiamiento de proyectos de inversión, tal como se establece en el Convenio Legal (...).”

El numeral 3.2 “Prestatario” de los “Roles y Responsabilidades”, de la Sección III “Gobernanza”, establece que: “El Prestatario tiene la responsabilidad de llevar a cabo las

actividades de adquisición financiadas por el Banco de conformidad con las presentes Regulaciones”.

Normativa del Instituto Nacional de Estadística y Censos

Conforme Resolución No. 040-DIREJ-DIJU-NI-2022, del 22 de noviembre del 2022, el Director Ejecutivo, resuelve: *“Delegar la autorización, expedición y suscripción de actos administrativos, actos de simple administración y documentos oficiales que se requieran para el normal desenvolvimiento de la Gestión Administrativa Financiera del Instituto Nacional de Estadística y Censos, dentro de los Procesos de Contratación en el marco de Cooperación con Organismos Internacional y Multilaterales”, de conformidad con el siguiente articulado:*

Artículo 1.- Delegar al Coordinador/a General Administrativo/a Financiero/a en Planta Central, y a los Coordinadores/as Zonales en su respectiva jurisdicción (...)”

Mediante RESOLUCIÓN No. 030- DIREJ-DIJU-NI-2024 suscrita el 18 de noviembre de 2024, por el Director Ejecutivo, se resolvió: *“ Expedir la siguiente reforma a la Resolución Nro. 042-DIREJ-DIJU-NI-2022 de 28 de noviembre de 2022, a través de la cual se emitió el “Manual Operativo (MOP) y Procedimientos de Gestión Laboral (PGL), asociados a la ejecución del Proyecto “Fortalecimiento del Sistema Estadístico Nacional del Ecuador”, conforme el siguiente articulado: Artículo 1.- Actualícese el Manual Operativo del Proyecto “Fortalecimiento del Sistema Estadístico Nacional del Ecuador”, por la versión 6 de octubre de 2024; mismo que, pasa a formar parte integrante de la presente resolución.”*

Normas de Control Interno para las Entidades, Organismos del Sector Público y de las Personas Jurídicas de Derecho Privado

“410-04 Plan estratégico y operativo de tecnologías de la información y comunicaciones

La unidad de tecnologías de la información y comunicaciones elaborará e implementará un plan estratégico para administrar y dirigir todos los recursos tecnológicos y de comunicación, el mismo que estará alineado con el plan estratégico institucional y éste con el Plan Nacional de Desarrollo y las políticas públicas del gobierno.

El plan estratégico tendrá un nivel de detalle suficiente para permitir la definición de planes operativos de tecnología de Información y especificará como éste contribuirá a los objetivos estratégicos de la organización; incluirá un análisis de la situación actual y las propuestas de mejora con la participación de todas las unidades de la organización, se considerará la estructura interna, procesos, infraestructura, comunicaciones, aplicaciones y servicios a brindar, así como la definición de estrategias, análisis pertinente de riesgos, cronogramas, presupuesto de la inversión y operativo, fuentes de financiamiento y los requerimientos legales y regulatorios de ser necesario.

La unidad de tecnologías de la información y comunicaciones elaborará planes operativos alineados con el plan estratégico de tecnologías de la información y comunicación y los objetivos estratégicos de la institución, estos planes incluirán los portafolios de proyectos y de servicios, la arquitectura y dirección tecnológicas, las estrategias de migración, los aspectos de contingencia de los componentes de la infraestructura y consideraciones relacionadas con la incorporación de nuevas tecnologías de información vigentes a fin de evitar la obsolescencia.

El plan estratégico de tecnologías de la Información y comunicación y los planes operativos de tecnología de información, así como el presupuesto asociado a éstos serán analizados y aprobados por la máxima

autoridad de la organización e incorporados al presupuesto anual de la misma; se actualizarán, monitorearán y evaluarán, según la naturaleza de la Entidad y volumen de sus operaciones, para determinar su grado de ejecución y tomar las medidas correctiva cuando corresponda.

Las adquisiciones de hardware, software u otros servicios tecnológicos deben responder a los proyectos incluidos en el plan estratégico de tecnologías de la información y comunicación de la organización. Las situaciones de excepción deben ser autorizadas por la máxima autoridad de la entidad o por la instancia que corresponda; y, se someterán al trámite de reforma pertinente.”

Acuerdos

ACUERDO Nro. MINTEL-MINTEL-2024-0009

Reformar el Acuerdo Ministerial No. MINTEL-MINTEL-2022-0032 de 15 de noviembre de 2022, mediante el cual se expidió la “NORMA TÉCNICA QUE REGULA EL PROCEDIMIENTO PARA LA APROBACIÓN DE VIABILIDAD TÉCNICA DE PROYECTOS DE GOBIERNO ELECTRÓNICO Y PARA LA AUTORIZACIÓN DE CRITICIDAD DE SOFTWARE Y SERVICIOS RELACIONADOS AL SOFTWARE”.

“(…) Art. 11.- De los proyectos que aplican al procedimiento simplificado.- Las entidades sujetas al ámbito de aplicación de esta norma, realizarán el procedimiento simplificado mediante la plataforma GOB.EC (…)”.

3. MECANISMO DE CONTRATACIÓN SELECCIONADO:

El procedimiento seleccionado en función de la normativa legal vigente es: Solicitud de Ofertas.

4. DETALLE DE LA NECESIDAD (JUSTIFICACIÓN):

El INEC, como ente rector de la estadística pública nacional, enfrenta una creciente demanda por procesar grandes volúmenes de datos provenientes de registros administrativos, encuestas y otras fuentes no estructuradas, que requieren tecnologías robustas de cómputo distribuido. Esta necesidad ha motivado el uso de plataformas de big data, en particular el ecosistema Hadoop, cuya arquitectura permite una escalabilidad horizontal, tolerancia a fallos, y una significativa reducción en los tiempos de respuesta al ejecutar procesos complejos de análisis y transformación de datos.

Desde el año 2015, el INEC ha trabajado con el framework de Hortonworks sobre Hadoop, permitiendo atender los requerimientos institucionales con tiempos de respuesta competitivos. No obstante, este entorno ya se encuentra discontinuado y carente de soporte oficial, dado que la empresa Hortonworks fue absorbida por Cloudera Inc. en 2018, lo que bloqueó el acceso público a sus repositorios y suspendió el desarrollo de actualizaciones y parches de seguridad. Este hecho ha generado una obsolescencia técnica progresiva del actual clúster, lo cual pone en riesgo la operatividad de los sistemas analíticos del INEC.

Si bien la implementación original fue realizada con talento interno, actualmente el INEC no cuenta con personal que posea el nivel de especialización requerido para desplegar e integrar un clúster moderno de Hadoop. Esta afirmación se respalda en el comunicado emitido por la Dirección de Administración de Recursos Humanos, en la que se indica que, tras revisar la

base de servidores de la Institución, no se identifican perfiles con la experiencia técnica específica ni las certificaciones especializadas necesarias para llevar a cabo la instalación, configuración y administración de una arquitectura Hadoop de nueva generación. Esta brecha de conocimiento técnico —que abarca componentes como Apache NiFi, Apache Atlas, Apache Ranger, Spark 3, Hive 3 con LLAP, entre otros— constituye un factor crítico que limita la capacidad institucional para ejecutar la migración de forma segura, eficiente y conforme a las mejores prácticas de la industria.

La escasez de profesionales especializados en tecnologías Hadoop en Ecuador ha sido evidenciada tanto en el sector público como en el privado. Universidades y centros de formación superior no incluyen en sus mallas curriculares programas integrales sobre arquitectura e ingeniería de datos con Hadoop, y las certificaciones internacionales ofrecidas por entidades como Cloudera, Hortonworks (en su momento) o Databricks tienen costos elevados, y deben ser cursadas mayoritariamente en el exterior o en modalidad online, lo cual implica una inversión considerable y plazos prolongados de formación.

Además, la implementación funcional de un clúster moderno de Hadoop conlleva una alta complejidad técnica. Es necesario coordinar múltiples servicios distribuidos, configurar entornos de alta disponibilidad, establecer rigurosas políticas de seguridad y gobernanza de datos, orquestar flujos de información en tiempo real y asegurar la interoperabilidad con otras plataformas analíticas. Estos desafíos incrementan significativamente el riesgo de fallas operativas. Dada la complejidad inherente a este tipo de instalaciones, resulta indispensable contratar un servicio especializado externo.

Adicionalmente, es importante señalar que con la puesta en marcha del nuevo clúster Hadoop, el INEC se convertirá en la primera institución del sector público ecuatoriano en operar un entorno de procesamiento masivo de datos a gran escala con arquitectura big data actualizada, lo cual refuerza el rol del INEC como referente en el uso de tecnologías emergentes para la producción estadística basada en registros administrativos.

Desde el punto de vista financiero y administrativo, cabe destacar que este servicio se enmarca dentro del proyecto “Robustecimiento de la Producción Estadística del Ecuador” (C6 - RRAA), en concordancia con el marco normativo nacional y de las políticas de racionalización del gasto público. Se prioriza el uso de soluciones basadas en software libre, como Apache Hadoop y su ecosistema, alineando el proceso con los principios de eficiencia, sostenibilidad del Estado y optimización de los recursos públicos, sin comprometer la calidad ni la continuidad operativa de los servicios estadísticos institucionales, en concordancia con los principios de eficiencia y sostenibilidad del Estado.

Con esta contratación, el INEC podrá continuar brindando servicios críticos de producción estadística basados en registros administrativos, los cuales son de vital importancia tanto para la ciudadanía como para los organismos de planificación y toma de decisiones a nivel nacional. Esta solución tecnológica garantizará la continuidad operativa, permitirá escalar las

capacidades analíticas institucionales y aportará al cumplimiento de los objetivos estratégicos del INEC en materia de procesamiento masivo de información.

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

Tabla 1. Componentes de Hadoop instalados con sus respectivas versiones en cada servidor

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 datos.

5. DESCRIPCIÓN DEL SERVICIO ESPERADO

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

REQUERIMIENTO DEL CLÚSTER HADOOP
INSTALACIONES Y CONFIGURACIONES MINIMAS
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 prerrequisitos 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.
TRANSFERENCIA DE CONOCIMIENTOS
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.
SOPORTE TÉCNICO DEL SERVICIO
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.

6. ANÁLISIS

5.1 Análisis de Beneficio

La contratación del servicio de instalación y configuración de Hadoop permitirá al INEC actualizar completamente su ecosistema de big data, migrando desde una plataforma obsoleta y sin soporte hacia un entorno moderno, robusto y sostenible basado en software libre. Este salto tecnológico no solo permitirá mantener y mejorar el rendimiento del procesamiento de grandes volúmenes de información, sino que también preservará la inversión realizada durante años en el desarrollo de procesos analíticos específicos para Hadoop.

Además, con esta contratación se logrará:

- La optimización del rendimiento y escalabilidad de la infraestructura adquirida recientemente.
- La integración de nuevos componentes que aumenten la capacidad de gestión, monitoreo y gobernanza de datos.
- La reducción del riesgo tecnológico y operativo por fallas en el sistema actual desactualizado.
- La eliminación de dependencias con software propietario costoso y restrictivo.

La implementación de este nuevo clúster Hadoop en el INEC también permitirá el cumplimiento de los objetivos institucionales estratégicos, garantizando la continuidad operativa de los productos estadísticos que actualmente ofrece la institución. Estos productos son de vital importancia para la ciudadanía en general, así como para los distintos niveles de gobierno y la cartera de Estado, quienes los utilizan como insumos clave para la toma de decisiones y la formulación de políticas basadas en evidencia.

La institución contará así con una base tecnológica moderna que permitirá generar productos estadísticos con mayor calidad, trazabilidad y velocidad de respuesta, habilitando incluso nuevas líneas de investigación e innovación estadística.

5.2 Eficiencia o efectividad

El servicio a contratar permitirá asegurar la continuidad operativa y analítica de uno de los activos tecnológicos más estratégicos del INEC. Un clúster Hadoop moderno, correctamente instalado y configurado, es esencial para garantizar eficiencia computacional, distribución de carga, procesamiento paralelo de tareas, y reducción de tiempos de respuesta, incluso con bases de datos que superan los 700 millones de registros.

Sin este entorno, los tiempos de procesamiento volverían a niveles pre-2015 (de 1 a 6 días por consulta), afectando la calidad, oportunidad y eficiencia de los productos estadísticos institucionales.

Este servicio permitirá:

- Maximizar el aprovechamiento de la nueva infraestructura disponible.
- Reducir los tiempos de procesamiento y respuesta en hasta un 90% en comparación con arquitecturas tradicionales.
- Prevenir la duplicación de esfuerzos y recursos al no tener que rediseñar todos los procesos desarrollados en los últimos 8 años.

7. CONCLUSIONES

- Es necesario instalar versiones estables y disponibles del software de los componentes de Hadoop y nuevos componentes en reemplazo a los que ya no cuentan con soporte.
- Es necesario migrar el servicio de procesamiento (datos y funcionalidades) al nuevo servidor, para mantener la disponibilidad del servicio, tanto interno como externo, que brinda el INEC, incrementando las capacidades tecnológicas de almacenamiento y de procesamiento.

8. RECOMENDACIÓN

Se recomienda llevar a cabo la contratación del “SERVICIO DE INSTALACIÓN Y CONFIGURACIÓN DE HADOOP EN EL INEC C3 RRAA” para contar con versiones estables y disponibles del software de los componentes de Hadoop y mantener la disponibilidad de los servicios que brinda el INEC.

9. 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	