

ESTUDIOS PREVIOS

PROCESO DE CONTRATACIÓN SELECCIÓN ABREVIADA DERIVADO DE LA CELEBRACIÓN DE ACUERDO MARCO DE PRECIOS - NUBE PÚBLICA III – CCE-908-1AMP-2019

OBJETO

**ADQUISICIÓN DE UNA SOLUCIÓN BASADA EN PLATFORM AS A SERVICE (PAAS)
SUMINISTRADA POR CLOUD SERVICE PROVIDERS, PARA EL RECONOCIMIENTO
AUTOMÁTICO ÓPTICO DE CARACTERES Y HERRAMIENTAS DE INTELIGENCIA
ARTIFICIAL DE ANÁLISIS DE INFORMACIÓN NO ESTRUCTURADA PARA LA
DIRECCIÓN DE INFORMACIÓN, ANÁLISIS Y REACCIÓN INMEDIATA.**

10 DE AGOSTO DE 2020

En desarrollo de lo dispuesto por la Ley 80 de 1993 artículo 25, numerales 7 y 12, modificado por el artículo 87 de la Ley 1474 de 2011; el artículo 2 de la Ley 1150 de 2007; así como el artículo 2.2.1.1.2.1.1, se elaboran los Estudios y Documentos Previos que soportan el presente proceso de contratación.

1. DESCRIPCIÓN DE LA NECESIDAD A SATISFACER

La Constitución Política de Colombia en los artículos 117, 119 y 267 establecen que la Contraloría General de la República es un órgano de control, que tiene a cargo la vigilancia de la gestión fiscal de la administración y de los particulares o entidades que manejan fondos públicos respecto de todos los tipos de recursos.

La Contraloría General de la República de manera oportuna, independiente y efectiva, controla y vigila la gestión fiscal de los recursos públicos, garantizando la participación de la ciudadanía mediante la articulación regional, con un enfoque preventivo que contribuya al buen manejo de los recursos públicos y que genere una mejora en la gestión del Estado y calidad de vida de los colombianos.

En su actuar evalúa los resultados obtenidos por las diferentes organizaciones y entidades del Estado, al determinar si adquieren, manejan y/o usan los recursos públicos dentro del marco legal, sujetos a los principios de economía, eficiencia, eficacia, equidad y sostenibilidad ambiental; examina la razonabilidad de los estados financieros de los sujetos de control fiscal y determina en qué medida logran sus objetivos y cumplen sus planes, programas y proyectos; tiene a su cargo establecer la responsabilidad fiscal de los servidores públicos y de los particulares que causen, por acción o por omisión y en forma dolosa o culpable, un daño al patrimonio del Estado; impone las sanciones pecuniarias que correspondan y las demás acciones derivadas del ejercicio de la vigilancia fiscal y procura, igualmente, incluyendo el resarcimiento del patrimonio público.

Para el cumplimiento de la función la Entidad ha sido sometida a una gran transformación llevada a cabo mediante el Acto Legislativo No. 04 de 2019, que reformó el artículo 267¹,

¹Artículo 267 Constitución Política de Colombia: "Artículo modificado por el artículo 1 del Acto Legislativo 4 de 2019". El nuevo texto es el siguiente: La vigilancia y el control fiscal son una función pública que ejercerá la Contraloría General de la República, la cual vigila la gestión fiscal de la administración y de los particulares o entidades que manejen fondos o bienes públicos, en todos los niveles administrativos y respecto de todo tipo de recursos públicos. La ley reglamentará el ejercicio de las competencias entre contralorías, en observancia de los principios de coordinación, concurrencia y subsidiariedad. El control ejercido por la Contraloría General de la República será preferente en los términos que defina la ley.

El control fiscal se ejercerá en forma posterior y selectiva, y además podrá ser preventivo y concomitante, según sea necesario para garantizar la defensa y protección del patrimonio público. El control preventivo y concomitante no implicará coadministración y se realizará en tiempo real a través del seguimiento permanente de los ciclos, uso, ejecución, contratación e impacto de los recursos públicos, mediante el uso de tecnologías de la información, con la participación activa del control social y con la articulación del control interno. La ley regulará su ejercicio y los sistemas y principios aplicables para cada tipo de control.

El control concomitante y preventivo tiene carácter excepcional, no vinculante, no implica coadministración, no versa sobre la conveniencia de las decisiones de los administradores de recursos públicos, se realizará en forma de advertencia al gestor

268, 271, 272 y 274, raíz del cual se expidió el Decreto N°2037 de 2019 “por el cual se desarrolla la estructura de la Contraloría General de la República, se crea la Dirección de Información, Análisis y Reacción Inmediata y otras dependencias requeridas para el funcionamiento de la entidad”.

La Dirección de Información, Análisis y Reacción Inmediata se encuentra adscrita al Despacho del Señor Contralor General de la República, perteneciente por ende al nivel superior de dirección del nivel central de la entidad.

El referido Decreto adicionó en su artículo 3° el artículo 42A al Decreto Ley 267 de 2000 en el que definió las funciones de la Dirección de Información, Análisis y Reacción Inmediata en los siguientes términos:

“Artículo 3. Adicionar el artículo 42A al Decreto ley 267 de 2000, así:

“ARTÍCULO 42A. DIRECCIÓN DE INFORMACIÓN, ANÁLISIS Y REACCIÓN INMEDIATA. Son funciones de la Dirección de Información, Análisis y Reacción Inmediata:

1. Dirigir la formulación de políticas, planes, programas, proyectos y procedimientos relacionados con el acceso, acopio, custodia, seguridad, uso, análisis y aprovechamiento de datos e información, que contribuyan a incrementar la eficiencia, eficacia y resultados con valor agregado de las acciones de vigilancia y control fiscal, con el apoyo de la Unidad de Seguridad y Aseguramiento Tecnológico e Informático y la Oficina de Sistemas e Informática en lo que corresponda.

2. Adoptar el Modelo de Gobierno de Datos y estrategia de datos abiertos en la Contraloría General de la República.

(...)

fiscal y deberá estar incluido en un sistema general de advertencia público. El ejercicio y la coordinación del control concomitante y preventivo corresponde exclusivamente al Contralor General de la República en materias específicas.

La vigilancia de la gestión fiscal del Estado incluye el seguimiento permanente al recurso público, sin oponibilidad de reserva legal para el acceso a la información por parte de los órganos de control fiscal, y el control financiero, de gestión y de resultados, fundado en la eficiencia, la economía, la equidad, el desarrollo sostenible y el cumplimiento del principio de valoración de costos ambientales. La Contraloría General de la República tendrá competencia prevalente para ejercer control sobre la gestión de cualquier entidad territorial, de conformidad con lo que reglamente la ley.

El control jurisdiccional de los fallos de responsabilidad fiscal gozará de etapas y términos procesales especiales con el objeto de garantizar la recuperación oportuna del recurso público. Su trámite no podrá ser superior a un año en la forma en que lo regule la ley.

La Contraloría es una entidad de carácter técnico con autonomía administrativa y presupuestal. No tendrá funciones administrativas distintas de las inherentes a su propia organización y al cumplimiento de su misión constitucional. (...)”

11. Dirigir y adoptar las políticas y protocolos, para compartir, suministrar, integrar o facilitar el acceso, intercambio, integración e interoperabilidad de datos o información de la Entidad con organismos nacionales o internacionales, bajo las directrices del Contralor General de la República.

(...)”

Así mismo, el Decreto 2037 de 2019, creo 3 unidades que integran la Dirección de Información, Análisis y Reacción Inmediata, una de las cuales se denomina Unidad de Información cuyas funciones fueron establecidas en el artículo 4° del Decreto indicado, que adicionó el artículo 42B al Decreto Ley 267 de 2000, como a continuación se indica:

“Artículo 4. Adicionar el artículo 42B al Decreto Ley 267 de 2000, así:

“ARTÍCULO 42B. UNIDAD DE INFORMACIÓN. Son funciones de la Unidad de Información:

- 1. Dirigir y coordinar la implementación del Modelo de Gobierno de Datos y estrategia de datos abiertos en la Contraloría General de la República y la ejecución de las funciones y proyectos que se generen en su aplicación.*
- 2. Orientar, gestionar y controlar la adquisición, acceso, almacenamiento, seguridad, uso y generación de información y datos en la Contraloría General de la República, de conformidad con las políticas y protocolos que se adopten para el efecto.*
- 3. Administrar y actualizar las bases de datos e información para el ejercicio de la función de vigilancia y control fiscal, de conformidad con las directrices y protocolos adoptados por la Dirección de Información, Análisis y Reacción Inmediata.*
- 4. Proponer y participar en la formulación de las políticas de acceso, custodia, almacenamiento, aseguramiento, intercambio y uso de la información adquirida, obtenida o administrada por la Entidad, en coordinación con las dependencias competentes.*
- 5. Evaluar y determinar la conveniencia técnica del intercambio de información e interoperabilidad entre organismos de investigación, vigilancia y control a nivel nacional e internacional, para la programación y desarrollo de operaciones simultáneas o conjuntas, bajo los lineamientos del Director de Información, Análisis y Reacción Inmediata.*
- 6. Compartir, suministrar, integrar o facilitar el acceso, intercambio, integración e interoperabilidad de datos o información de la Entidad con organismos nacionales o internacionales, bajo las políticas y protocolos adoptados para el efecto.*

7. Las demás que le sean asignadas por la ley, o delegadas por el Contralor General de la República, y que correspondan a la naturaleza de la dependencia”.

En concordancia con lo anterior, el Gobierno Nacional expidió el Decreto N°403 del 16 de marzo 2020 que en su artículo 593 reglamentó el acceso y análisis de la información requerida para el cumplimiento de los fines institucionales y misionales de la Entidad, por conducto de la Dirección de Información, Análisis y Reacción Inmediata. El título X del Decreto en cuestión contiene disposiciones relativas al acceso a los sistemas de información y bases de datos, entre las cuales resulta pertinente citar las alusivas a la finalidad y condiciones de acceso, previstas respectivamente en los artículos 89 y 90:

*“**Artículo 89. Finalidad.** La Contraloría General de la República tendrá acceso a los sistemas de información o bases de datos de las entidades públicas y de las entidades privadas que dispongan o administren recursos y/o ejerzan funciones públicas. La información solo podrá ser utilizada para los fines y propósitos de la vigilancia y control fiscal. Esta facultad es aplicable a todos los órganos de control fiscal exclusivamente frente a sus sujetos u objetos de vigilancia y control fiscal. “*

*“**Artículo 92. Condiciones de acceso.** Los órganos de control fiscal acordarán con la entidad correspondiente los términos mediante los cuales se hará efectiva la interoperabilidad o el acceso a los sistemas de información o a las bases de datos solicitado, según las alternativas tecnológicas disponibles y con plena garantía de la seguridad y la integridad de la información, a través de la suscripción de protocolos técnicos o memorandos de entendimiento. (...)”* Subrayado fuera del texto.

En igual forma, el Decreto 2037 de 2019, creo 3 unidades que integran la Dirección de Información, Análisis y Reacción Inmediata, una de las cuales se denomina Unidad de Información cuyas funciones fueron establecidas en el artículo 4° del Decreto indicado, que adicionó el artículo 42B al Decreto Ley 267 de 2000, como a continuación se indica:

*“**Artículo 4.** Adicionar el artículo 42C al Decreto Ley 267 de 2000, así:*

²“Por el cual se dictan normas para la correcta implementación del Acto Legislativo 04 de 2019 y el fortalecimiento del control fiscal.”

³ “Artículo 59. Acceso y análisis a la información. La Contraloría General de la República podrá requerir, conocer y examinar todos los datos e información sobre hechos u operaciones, actos, contratos, programas, proyectos o procesos en ejecución, en los que se involucren recursos públicos y/o se afecten bienes o intereses patrimoniales de naturaleza pública, sin oponibilidad de reserva legal, exclusivamente para el ejercicio de sus funciones.

Para lo anterior, la Contraloría General de la República podrá apoyarse en la gestión inteligente de la información, entendida como el uso eficiente de todas las capacidades tecnológicas disponibles, inteligencia artificial, analítica y minería de datos, análisis predictivo y prospectivo, entre otros, para la determinación anticipada de eventos o malas prácticas, con probabilidad significativa de ocurrencia, persistencia o mutación, y que impliquen riesgos de pérdida del patrimonio público. Para el efecto, podrá conformar sistemas de seguimiento en tiempo real y confirmación directa con los gestores y los beneficiarios de dichas transacciones. En consecuencia, cada entidad deberá disponer lo necesario para garantizar el suministro oportuno y en tiempo real de la información requerida por la Contraloría General de la República, por conducto de la Dirección de Información, Análisis y Reacción Inmediata -DIARI-”.

"ARTÍCULO 42C. UNIDAD DE ANÁLISIS DE LA INFORMACIÓN.

Son funciones de la Unidad de Análisis de la Información:

1. Dirigir y coordinar los procesos y acciones relacionados con la minería, cruces y analítica de datos e información, mediante el uso de tecnologías y técnicas especializadas, para la identificación de riesgos de detrimento al patrimonio público, la planeación y focalización de la vigilancia y control fiscal y demás actuaciones a cargo de la Entidad.
2. Elaborar los informes de analítica de datos y liberarlos previa aprobación del Director de Información, Análisis y Reacción Inmediata, de acuerdo con los protocolos que se adopten para el efecto.
3. Asesorar a las dependencias de la Contraloría General de la República en el análisis de datos de fuentes internas o externas -estructurados, semiestructurados o no estructurados- a través de soluciones tecnológicas y técnicas especializadas para el ejercicio de la vigilancia y el control fiscal.
4. Ejercer funciones de policía judicial en los ámbitos relacionados con la gestión de datos e información electrónica, de conformidad con las normas y parámetros inherentes a esta actividad, en especial la preservación de la cadena de custodia.
5. Identificar el ciclo integral de los bienes, fondos o recursos públicos, desde su fuente hasta su ejecución, precisando los beneficiarios finales de las erogaciones presupuestales, y reportarlo a la Dirección de Información, Análisis y Reacción Inmediata.
6. Diseñar metodologías y herramientas de gestión que contribuyan al mejoramiento de la vigilancia y control fiscal y demás procesos de la Entidad.
7. Conformar y dirigir un banco de prácticas de gestión pública e identificar indicios, gastos indebidos, malas prácticas, riesgos, patrones o tendencias relacionadas con la gestión fiscal, que sirva como insumo para los ejercicios de vigilancia y control fiscal que se adelanten por las diferentes dependencias de la Entidad.
8. Informar o divulgar a la Unidad de Reacción Inmediata los resultados de las prácticas y actividades identificadas como riesgosas, para el ejercicio de acciones de reacción inmediata, la construcción de propuestas y recomendaciones para el fortalecimiento de la transparencia y el control fiscal.
9. Conformar, en coordinación con las áreas misionales de la Contraloría General de la República, el centro integrado de análisis de información pública, el cual podrá tener carácter interinstitucional, y podrán participar de manera permanente o transitoria

servidores de otros organismos de control, de la Rama Judicial, de la Policía Nacional, entre otras entidades o autoridades oficiales.

10. Las demás que le sean asignadas por la ley, o delegadas por el Contralor General de la República, y que correspondan a la naturaleza de la dependencia”.

Por lo anteriormente expuesto la Contraloría General de la Republica ha identificado la necesidad de implementar nuevas herramientas tecnológicas que posibiliten el cumplimiento de su misión y las funciones de la Dirección de Información, Análisis y Reacción Inmediata, en concordancia con los objetivos planteados en el documento CONPES 3975 AI y 3920 Big Data.

Así mismo esta necesidad de implementación se alinea con el Plan Nacional de Desarrollo 2018-2022, *“Pacto por Colombia, pacto por la equidad”*, título “Pactos Estructurales”. “I Pacto por la Legalidad”, numeral C. Alianza contra la corrupción: tolerancia cero contra los corruptos.

Adicionalmente, esta necesidad también se alinea con el Plan Estratégico de la Contraloría General de la República cuatrienal denominado *“Una Contraloría para Todos”*, donde se consigna el plan de gobierno de la presente administración, direccionado a una vigilancia y control fiscal en tiempo real, preventivo, con alta tecnología y a la participación activa de la ciudadanía, específicamente el objetivo estratégico N°5 *“Habilitar las capacidades y servicios tecnológicos para impulsar la transformación digital de la entidad por medio de la práctica de arquitectura empresarial”* 5.1. *“Implementar soluciones tecnológicas que permitan la interoperabilidad, el intercambio de información en tiempo real y de forma segura, en toda la organización y con sus socios estratégicos.”* 5.2. *“Adoptar y adaptar nuevas tendencias a partir de la investigación a nivel de tecnologías, mejores prácticas, y marcos de referencia para establecer un ciclo de mejora continua buscando la efectividad.”* 5.3. *“incentivar el uso y apropiación de las herramientas tecnológicas a través de una adecuada gestión del conocimiento.* 5.4. *integrar los diferentes sistemas de información para facilitar la optimización de procesos, la adaptación de los cambios y el suministro de información de manera oportuna y confiable.”* 5.5. *“mejorar la gestión de los servicios tecnológicos, confidencialidad, integridad, disponibilidad y gobernanza”.*

Todo este escenario de transformación se soporta con tecnologías de información requeridos para los procesos misionales y de apoyo, modelos de inteligencia institucional, políticas de seguridad de la información, herramientas de colaboración, transferencia de conocimiento, capacitación, infraestructura de tecnología, comunicaciones, herramientas audiovisuales y equipos entre otros.

Las necesidades en cuanto a información y analítica que busca solventar la DIARI desde sus Unidades de Información y de Análisis presentan retos dentro de un entorno cambiante y dinámico dada su complejidad. Así mismo, las capacidades tecnológicas necesarias para acompañar la solución a estas necesidades debe ser elástica, flexible y dinámica permitiendo robustecer los procesos y resultados que los diferentes modelos de seguimiento y prevención deben continuamente obtener.

Dentro del marco de crecimiento y transformación digital en el que se encuentra la Entidad, la Dirección de Información, Análisis y Reacción Inmediata – DIARI, para sus procesos de analítica avanzada, requiere automatizar y fortalecer el proceso de analítica de información “no estructurada” para poder alimentar los programas de análisis de big data, en donde se realizan los procesos de analítica mediante , que es diferente del almacenamiento, procesamiento e infraestructura en la nube de las diferentes bases de datos e información de la DIARI⁴.

Por información “no estructurada” se entiende e incluyen la diversidad de documentos como contratos, pliegos, etc. que se reciben en formatos estáticos, difíciles de procesar y de organizar como PDF, imágenes, JPG, entre otros. Un ejemplo claro de información “no estructurada” son los repositorios de contratación pública SECOP I y SECOP II, en donde se albergan cerca de 7.5 millones de contratos. Por tal razón, para poder alimentar los modelos analíticos y realizar todos los procesos de análisis que acompañan la gestión de prevención y seguimiento con alertas tempranas, que la DIARI le provee a las delegaturas de la Contraloría General de la Republica, se requiere de un proceso manual que limita la capacidad de análisis y respuesta de manera severa.

Buscando resolver la necesidad de proveer a la DIARI con de una plataforma tecnológica que solucione las restricciones en el proceso de extracción de información no estructurada, junto con la capacidad para su posterior análisis y lectura en lenguaje natural (NLP⁵), se ha identificado una solución de referencia que cumpliendo unas condiciones técnicas, que asegure la calidad, flexibilidad y capacidad del proceso y que permita extraer, almacenar, procesar, analizar y disponer la información “no estructurada” de manera eficiente en “información estructurada”.

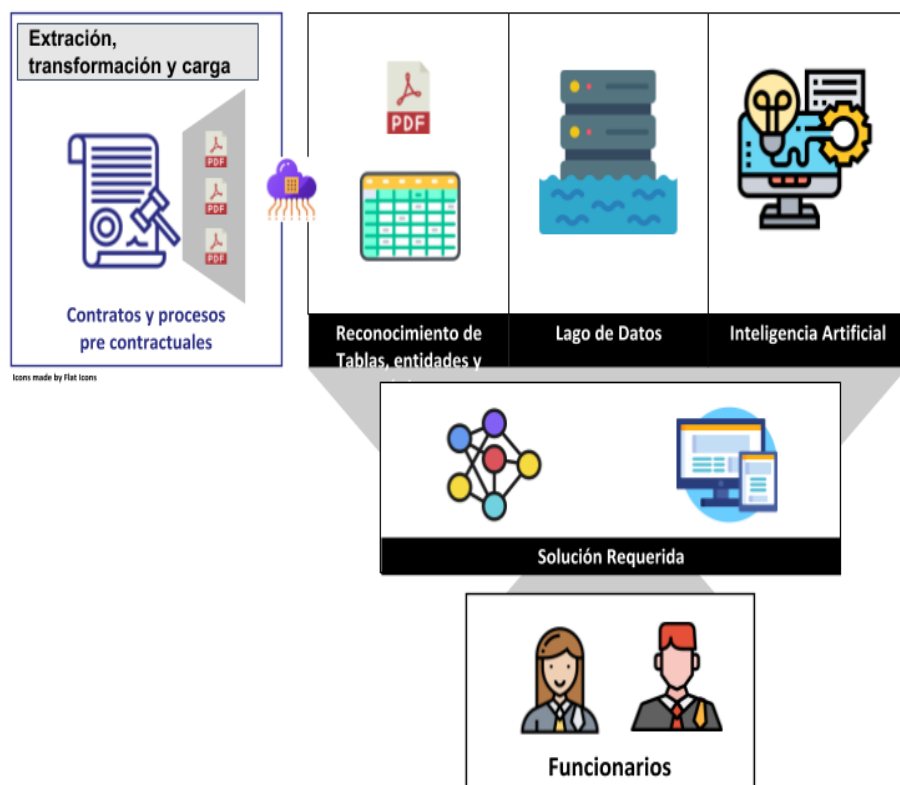
Con el fin de identificar técnica y funcionalmente la solución para esta necesidad, durante un período cercano a tres meses se efectuó un estudio y análisis técnico en herramientas de reconocimiento automático óptico (OCR) de caracteres y herramientas de inteligencia artificial para el análisis de información no estructurada, adelantando pruebas de

⁴ Orden de Compra N°50479 – Objeto: “Adquisición De Tokens De Microsoft Azure Para La Dirección De Información, Análisis Y Reacción Inmediata”, a través de Proceso De Contratación Selección Abreviada Derivado De La Celebración De Acuerdo Marco De Precios - Nube Pública III – CCE-908-1AMP-2019

⁵ Natural language processing: proceso de inteligencia artificial que permite lectura, entendimiento y análisis semántico de información.

extracción con elementos de Microsoft Azure, Amazon Web Services y desarrollos propios en códigos R y Python.

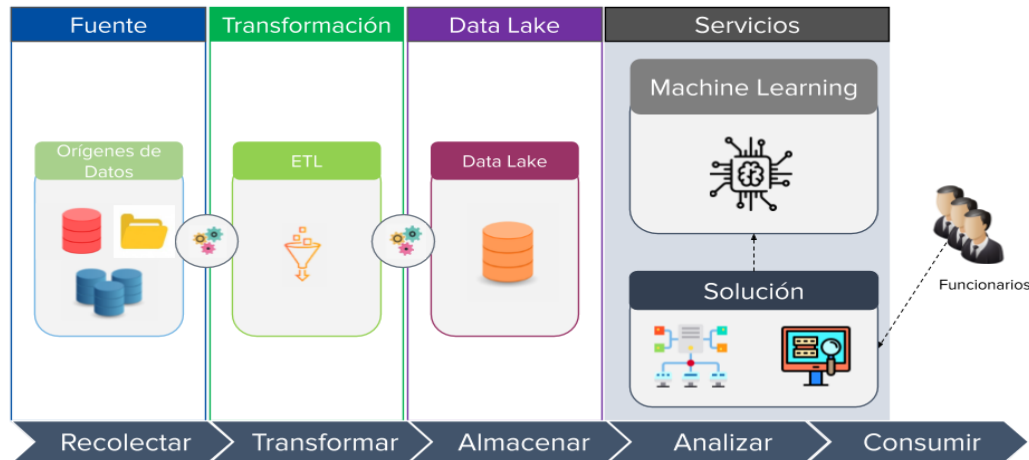
Los resultados del estudio permitieron a la Dirección de Información, Análisis y Reacción Inmediata – DIARI determinar el diagrama de contexto y una arquitectura de referencia para cubrir la necesidad, en donde se involucran varias fases del proceso, como se ilustra a continuación:



Las fases dentro del diagrama de contexto son;

1) *Extracción / recolección.* 2) *Transformación* 3) *Carga / Almacenamiento.* 4) *Análisis: Proceso de inteligencia artificial lineal y tabular de reconocimiento de información (tablas, entidades, unidades, precios), emparejamiento, comparación, referenciación.* 5) *Consumo: Front-end para usuarios / funcionarios.*

De acuerdo a este diagrama de contexto se determina que la arquitectura de referencia de la solución requerida es la siguiente:



- a) Debe contar con un módulo de gestión de análisis que permita:
- Formulario de consulta con la lista de contratos cargados en la aplicación.
 - Cargar el archivo del contrato, para la cual se realizará el análisis.
 - Asociar información específica al contrato
 - Contar con un enlace que dirija al usuario a los sistemas de la entidad para cada una de los contratos analizados. (Integración mediante servicios)
 - Debe permitir descargar el resultado del análisis de un contrato
 - Debe permitir corregir la clasificación del algoritmo.
- b) La aplicación debe contar con un módulo de administración que permita la configuración de:
- Roles, usuarios y permisos.
 - Descriptores y restrictores
 - Parámetros del sistema
- c) La solución debe permitir la construcción de diferentes reportes y/o representaciones gráficas de la información permitiendo el despliegue drill up y drill down tomando diferentes criterios.

Esta sería una arquitectura transversal que permita extraer y clasificar dicha información, es importante resaltar que a la fecha se tienen alrededor de 2.4 TB de información no estructurada y se proyecta que este volumen aumente de acuerdo con las necesidades de la DIARI, y de los repositorios en referencia que son SECOP I y SECOP II que albergan cerca de 40TB de información no estructurada.

Con esta plataforma se busca lograr extraer automáticamente texto y datos de documentos en formatos estáticos de manera masiva, automática y confiable, permitiendo:

- Automatizar rápidamente las actividades manuales de documentos, lo que le permite procesar millones de páginas de documentos en horas.
- Obtener información estructurada de información no estructurada, que permita ejecutar análisis profundos a través de procesos de inteligencia artificial y así poder tomar medidas al respecto dentro de las aplicaciones de los modelos de análisis de contratación pública.
- La gestión necesaria para superar estos procesos manuales, utilizando herramientas que incluyan el aprendizaje automático para leer y procesar instantáneamente cualquier tipo de documento y extraer con precisión texto, formularios, tablas y otros datos sin la necesidad de ningún trabajo manual o código personalizado permitiendo ejecutar las funciones de vigilancia y control de manera masiva y automática.
- El Desarrollo, prueba, implementación y operación de modelos de análisis predictivos.
- La búsqueda inteligente de tópicos (p.e. Representante Legal, Jurisprudencia, Objeto contractual, precios, ítems, entre otros) con capacidad de búsqueda en lenguaje natural por la Contraloría General de la República.
- La inclusión de un catálogo de datos que permita descubrir, comprender y consumir fuentes de datos de la entidad creando índices de búsqueda inteligentes que permita relacionar a través de los tópicos identificados en los documentos, relaciones emergentes entre estos.

De acuerdo con lo expuesto la necesidad radica en contar con los servicios de una plataforma especializada de digitalización avanzada con deep learning OCR, que cuente con la capacidad de analizar los datos en cualquier formato (linear y tabular). Esto permitirá el procesamiento, la transformación y análisis mediante procesos de inteligencia artificial de los datos entregados por las entidades públicas y privadas que manejan recursos públicos. Con esto de manera eficiente, automática y no supervisada se obtendrán resultados que protejan los procesos de contratación pública anticipadamente de eventos o malas prácticas, con probabilidad significativa de ocurrencia, persistencia o mutación, y que impliquen riesgos de pérdida del patrimonio público.

Estimando procesar, transformar y analizar dentro de esta solución un total 14.000.000 de páginas, las cuales entre otras corresponden en promedio a 500.000 contratos y sus anexos.

Todo ello con el fin de cumplir las nuevas funciones y el objetivo principal de la Contraloría

General de la República, teniendo en cuenta las potestades atribuidas en el Artículo 59 el cual señala: *“Acceso y análisis a la información. La Contraloría General de la República podrá requerir, conocer y examinar todos los datos e información sobre hechos u operaciones, actos, contratos, programas, proyectos o procesos en ejecución, en los que se involucren recursos públicos y/o se afecten bienes o intereses patrimoniales de naturaleza pública, sin oponibilidad de reserva legal, exclusivamente para el ejercicio de sus funciones. Para lo anterior, la Contraloría General de la República podrá apoyarse en la gestión inteligente de la información, entendida como el uso eficiente de todas las capacidades tecnológicas disponibles, inteligencia artificial, analítica y minería de datos, análisis predictivo y prospectivo, entre otros, para la determinación anticipada de eventos o malas prácticas, con probabilidad significativa de ocurrencia, persistencia o mutación, y que impliquen riesgos de pérdida del patrimonio público. Para el efecto, podrá conformar sistemas de seguimiento en tiempo real y confirmación directa con los gestores y los beneficiarios de dichas transacciones. En consecuencia, cada entidad deberá disponer lo necesario para garantizar el suministro oportuno y en tiempo real de la información requerida por la Contraloría General de la República, por conducto de la Dirección de Información, Análisis y Reacción Inmediata -DIARI-.”*

La nueva capacidad de cómputo definida, se encuentra contenida en los servicios de Nube Pública de los siguientes Cloud Service Providers:

Amazon Web Services
Google Cloud
Microsoft Azure
Oracle Cloud

La necesidad se puede suplir con la oferta del Catálogo del Acuerdo Marco de Precios Nube Pública III CCE-908-1-AMP-2019, consignado en el siguiente link:

<https://www.colombiacompra.gov.co/tienda-virtual-del-estado-colombiano/tecnologia/nube-publica-iii>,

Especialmente en el documento de descarga hoja de cálculo denominada Servicios CSP, que pueden ser consultados en el link:

Segmento 1. AWS

<https://calculator.s3.amazonaws.com/index.html#r=IAD&s=EC2&key=files/calc-f44a3af7f646dd6c017ca3767505c9fcf87cd196&v=ver20191021xG>

Segmento 2. Google

<https://cloud.google.com/products/calculator/>

Segmento 3. Microsoft

Segmento 4. Oracle

https://cloud.oracle.com/en_US/cost-estimator

Para ello se procederá de conformidad con el Acuerdo Marco de Precios señalado y la Guía para comprar en la Tienda Virtual del Estado Colombiano a través del Acuerdo Marco para la prestación de Servicios de la Nube Pública III, teniendo en cuenta la necesidad planteada y las especificaciones técnicas se definirá mediante la ejecución del RFI⁶ en el que se determinará el Cloud Service Providers que sea más adecuada para suplir la necesidad descrita.

Finalmente, y de acuerdo a la necesidad expuesta, la Dirección de Información Análisis y Reacción -DIARI -Inmediata amparada en el proyecto de inversión - *“Fortalecimiento del ejercicio del control fiscal con eficiencia en la gestión de datos de la Contraloría General de la Republica Nacional”* (código BPIN 2020011000002), objetivo *“Modernizar los sistemas de información existentes en el ciclo de vida de los datos”, producto “Servicio de Implementación Sistemas de Gestión”-*, encuentra necesario para fortalecer y optimizar el control fiscal, que ejecuta en el ejercicio de sus competencias, adquirir servicios en la Nube de soluciones PaaS y servicios complementarios que permita gestionar los altos volúmenes de información no estructurada que se encuentra en repositorios de información; de tal forma que permita la captura, transmisión, análisis, explotación, almacenamiento y aseguramiento de los datos en una solución combinada e integrada bajo el modelo de servicio PaaS (Platform as a Service) y servicios complementarios.

En atención a lo expuesto la Dirección de Información, Análisis y Reacción Inmediata – DIARI encuentra necesario adelantar el presente proceso de contratación con el fin de suplir la necesidad señala.

2. OBJETO – ALCANCE DEL OBJETO O ESPECIFICACIONES TÉCNICAS Y CONDICIONES GENERALES.

2.1. Objeto

Adquisición de una solución basada en Platform AS a Service (PAAS) suministrada por Cloud Service Providers, para el reconocimiento automático óptico de caracteres y herramientas de inteligencia artificial de análisis de información no estructurada para la Dirección de Información, Análisis y Reacción Inmediata. (801114-43)

De acuerdo a la Codificación de Bienes y Servicios del Código Estándar de productos y servicios de Naciones Unidas, los servicios a contratar se encuentran codificados, hasta el

⁶ La Solicitud de Información es más conocida por sus siglas del término anglosajón, RFI (*Request for Information*)

tercer nivel de la siguiente manera:

Código Segmento	Nombre Segmento	Código Familia	Nombre Familia	Código Clase	Nombre Clase
81000000	Servicios Basados en Ingeniería, Investigación y Tecnología	81110000	Servicios Informáticos	81112000	Servicios de datos

Tabla 1: Códigos UNSPSC

El presente proceso se encuentra incluido en el Plan Anual de Adquisiciones de la Entidad vigencia 2020.

Código	Objeto Actual	Fecha estimada de inicio de proceso de selección	Duración estimada del contrato	Modalidad de selección	Fuente de los recursos	Valor total estimado	Valor estimado en la vigencia actual	Se requieren vigencias futuras	Datos del contacto responsable
81112000	Adquisición de una solución basada en platform as a service (paas) suministrada por cloud service providers, para el reconocimiento automático óptico de caracteres y herramientas de inteligencia artificial de análisis de información no estructurada para la dirección de información, análisis y reacción inmediata.	ago-20	3 meses	Acuerdo Marco	Nación	\$ 1.650.000.000	\$ 1.650.000.000	NO	OSWALDO ZAPATA - JEFE DE UNIDAD DE ANÁLISIS DE LA DIARI HOSLANDER ADLAI SAENZ BARRERA - JEFE DE UNIDAD DE INFORMACIÓN DE LA DIARI

2.2. Alcance al Objeto y Especificaciones Técnicas.

2.2.1. Alcance del Objeto:

Se estima dentro de esta fase lograr procesar, transformar y analizar dentro de esta solución un total 14.000.000 de páginas, las cuales entre otras corresponden en promedio a 500.000 contratos y sus anexos.

CONTRATOS Y PROCESOS A PROCESAR POR LA CONTRALORÍA GENERAL DE LA REPÚBLICA		
Total Contratos SECOP I Y SECOP II	7.000.000	Contratos disponibles
Total Contratos Fase 1	500.000	Contratos fase 1 Cada contrato se conforma de 4 anexos y en promedio cada documento cuenta con 7 páginas aproximadamente, para un promedio total aproximado de 28 paginas

Total paginas Fase 1= 28* 500.000 = 14.000.000

Peso Contratos Fase 1

2.4 TB

- Tamaños del documento que puede varían entre 2MB y 10MB.
- La información fuente para conformar el lago de datos deberá tomarse de datos abiertos del SECOP I, SECOP II y Repositorio de la CGR.

2.2.2. Requerimientos técnicos

En el marco de la identificación de la necesidad de la adquisición de nuevas capacidades de cómputo, la Dirección de información, Análisis y Reacción Inmediata -DIARI-, requiere contratar los servicios de computación en la Nube Pública bajo la modalidad de entrega “*Servicio autogestionado*”, en concordancia con el Acuerdo Marco de Precios Nube Pública III CCE-908-1-AMP-2019, bajo el modelo de servicios en la nube tipo PaaS⁷.

⁷ **“2.4.2 Plataforma como Servicio (Platform as a Service – PaaS)** Este modelo de servicio proporciona al consumidor la posibilidad de desplegar en la infraestructura de nube aplicaciones creadas por el mismo consumidor (o adquiridas a un tercero) utilizando lenguajes de programación, bibliotecas, servicios y herramientas soportadas por el proveedor de nube. El consumidor no gestiona ni controla la infraestructura subyacente de la nube, servidores, sistemas operativos o almacenamiento, pero tiene control sobre las aplicaciones desplegadas y posiblemente configuraciones para el entorno de hospedaje de aplicaciones. El proveedor de PaaS es responsable del mantenimiento, soporte y operación de las plataformas dispuestas como servicio.

Esta capacidad no excluye necesariamente el uso de lenguajes de programación compatibles, bibliotecas, servicios y herramientas de otras fuentes.

A continuación, algunos ejemplos de servicios tipo PaaS [5]:

- *Inteligencia de Negocios: Plataformas para la creación de aplicaciones como paneles, sistemas de informes y análisis de datos.*
- *Base de datos: Servicios que ofrecen soluciones de base de datos relacionales escalables o almacenes de datos no SQL escalables.*
- *Desarrollo y pruebas: Plataformas para el desarrollo y los ciclos de pruebas de desarrollo de aplicaciones, que se expanden y se contraen según sea necesario.*
- *Integración: Plataformas de desarrollo para la construcción de aplicaciones de integración en la nube y dentro de la empresa.*

Implementación de aplicaciones: Plataformas adecuadas para el desarrollo de aplicaciones de uso general. Estos servicios proporcionan bases de datos, entornos de ejecución de aplicaciones web, entre otros”.

El servicio de computación en la nube requerido debe poseer las siguientes características:

Características del Servicio	Características
Auto-servicio por demanda	Un consumidor puede unilateralmente aprovisionar capacidades de cómputo, tales como tiempo de servidor y almacenamiento en red, según sea necesario y de manera automática sin necesidad de interacción humana con cada proveedor de servicios.
Amplio acceso desde la red	Los servicios proporcionados deben poder ser accesibles sobre la red y a través de mecanismos estándares que promuevan el uso desde plataformas heterogéneas del cliente (por ejemplo: computadores, teléfonos móviles o tabletas).
Reservas de recursos en común	Los recursos computacionales son puestos a disposición de los consumidores, los cuales comparten diferentes recursos físicos y virtuales asignados dinámicamente y por demanda. Hay un sentido de independencia de la localización en la que el usuario no tiene un estricto control del lugar exacto en el que se encuentra su información o de los servicios contratados, aunque sí debe poder especificar un ámbito mínimo de actuación (por ejemplo: un país, una región o un centro de proceso de datos concreto). Ejemplos de recursos incluyen almacenamiento, procesamiento, memoria y ancho de banda.
Rapidez y elasticidad	Los recursos proporcionados deben poder crecer o decrecer en cualquier momento, en algunos casos automáticamente, con el fin de escalar rápidamente y responder a la demanda de los usuarios.
Servicio supervisado	Los sistemas en la nube automáticamente controlan y optimizan el uso de los recursos dotándose de capacidades para medir su rendimiento en un nivel de abstracción suficiente para la naturaleza del servicio proporcionado. Además, dicho control debe permitir ser monitoreado y reportado de manera transparente tanto al proveedor del servicio como al consumidor del mismo.

Las condiciones de la prestación del servicio, se acogen a lo establecido en el Acuerdo Marco de Precios CCE-908-1-AMP-2019, las cuales pueden ser consultadas en el documento “Anexo 3 - Condiciones transversales” allegado como anexo en el presente estudio previo o en la página de Colombia Compra eficiente asociado al link antes señalado.

2.2.2.1. Requerimientos técnicos entorno privado.

- Debe contar con un módulo de gestión de análisis que permita:
 - A. Formulario de consulta con la lista de contratos cargados en la aplicación.
 - B. Cargar el archivo del contrato, para la cual se realizará el análisis.
 - C. Asociar información específica al contrato
 - D. Contar con un enlace que dirija al usuario a los sistemas de la entidad para cada una de las contratos analizados. (Integración mediante servicios)
 - E. Debe permitir descargar el resultado del análisis de un contrato
 - F. Debe permitir corregir la clasificación del algoritmo
- La aplicación debe contar con un módulo de administración que permita la

Fuente: Servicios Tecnológicos Guía Técnica, Guía de Computación en la nube v1.1 pagina 32. MINTIC.
<https://www.mintic.gov.co/arquitecturati/630/w3-article-75554.html>

configuración de:

- a) Roles, usuarios y permisos.
 - b) Descriptores y restrictores
 - c) Parámetros del sistema
- La solución debe permitir la construcción de diferentes reportes y/o representaciones gráficas de la información permitiendo el despliegue drill up y drill down tomando diferentes criterios.

2.2.2.1.1. Requerimientos proceso de clasificación de tópicos o conceptos de referencia

- a) Definición y asociación de descriptores y restrictores anidados, es decir que se apliquen para cualquier proceso cargado en el sistema. La configuración de descriptores y restrictores debe tener un permiso asociado a un rol específico.
- b) Módulo de entrenamiento del sistema para que comprenda conceptos jurídicos y conceptos contractuales específicamente junto con emparejamiento de conceptos extraídos contra conceptos de referencia, como precios, unidades de medida, especificaciones, etc.
- c) Es deseable contar con funcionalidades que permitan el aprendizaje desde un conjunto de datos etiquetados previamente definidos mediante la búsqueda de patrones en dicho conjunto de datos.
- d) Debe contar con un servicio de procesamiento de lenguaje natural (NLP) para descubrir ideas y relaciones debe permitir identificar el idioma del texto; extraer frases, nombres de lugares, personas, marcas o eventos clave; comprender el grado de positividad o negatividad del texto; analizar el texto mediante tokenización y categorías gramaticales; y organizar automáticamente una colección de archivos de texto por tema.
- e) Este servicio de procesamiento de lenguaje natural (NLP) para descubrir ideas y relaciones debe permitir organizar automáticamente por términos o temas relevantes, debe poder usar los temas para ofrecer contenido personalizado o para proporcionar una búsqueda y una navegación más ricas.

2.2.2.1.2. Requerimientos Técnicos de la nube pública solicitada dentro del PaaS.

- a) Se requiere la plataforma especializada de Digitalización avanzada con deep learning que cuente con la capacidad de analizar los datos en cualquier formato (linear y tabular)
 - Los servicios utilizados para implementar la solución no deben ser versiones de

prueba, beta o vistas preliminares.

- La solución debe incluir un servicio que extraiga automáticamente texto y datos desde documentos escaneados.
- El servicio debe permitir identificar el contenido de campos de formularios e información almacenada en tablas.
- El servicio debe usar aprendizaje automático para "leer" documentos a fin de extraer texto y datos de forma precisa sin necesidad de cambios manuales ni utilizar código personalizado.
- El servicio debe permitir automatizar rápidamente flujos de trabajo con documentos
- El servicio debe permitir usar los datos extraídos dentro de aplicaciones. Además, puede crear índices de búsqueda inteligente y flujos de trabajo de aprobación automatizados.
- El servicio debe permitir detectar automáticamente el diseño de un documento y los elementos clave en la página, entender las relaciones de datos en cualquier formulario o tabla integrados, y extraer todo con su contexto.

b) Adicionalmente, se requiere sobre del CSP la capacidad de análisis de lenguaje Natural (NLP) y búsqueda semántica.

- El servicio de búsqueda elástica gestionada debe ser compatible con SQL sin afectar la capacidad de hacer búsquedas en textos complejos o que contengan puntuación.
- El servicio de búsqueda elástica gestionada debe permitir la encriptación en tránsito y en reposo
- El servicio de búsqueda elástica gestionada debe permitir el cifrado entre nodos implementando Transport Layer Security (TLS) en todas las comunicaciones que se realizan entre los recursos de cómputo que integran el clúster del servicio de búsqueda elástica.
- El servicio de búsqueda elástica gestionada debe contar con alta disponibilidad y debe garantizar un SLA de 99,9%
- El servicio de búsqueda elástica gestionada debe permitir generar snapshots para mitigar riesgos asociados a la falla de un nodo.
- El servicio de búsqueda elástica gestionada debe permitir hacer actualizaciones en los clústeres que soportan el servicio sin generar tiempos de caída del servicio.
- El servicio de búsqueda elástica gestionada debe permitir que de forma automática ante el fallo de un nodo la carga de servicio se distribuya en los demás nodos y que se creen nodos nuevos para reemplazar a los que fallaron.
- El servicio de búsqueda elástica gestionada debe permitir configurar el servicio

para acceso a través de una VPN

- El servicio de búsqueda elástica gestionada debe permitir monitorear el rendimiento de varias etapas de una operación a través de registros lentos de dos tipos:
 - Registros lentos de índice: suministran información acerca del proceso de indexación y se pueden utilizar para ajustar la configuración del índice.
 - Registros lentos de búsqueda: suministran información acerca de la velocidad con la que se realizan consultas y recuperaciones. Estos registros ayudan a ajustar el rendimiento de cualquier tipo de operación de búsqueda en el servicio.
- c) Se requiere personal idóneo y certificado por Cloud Service Provider (CSP - Proveedor de servicios en la nube) para la extracción, transformación y cargue de manera segura la información suministrada por la Entidad a la nube pública, y para todas las demás actividades y entregables detallados en la descripción de la necesidad y de los entregables.
- d) Se debe garantizar que la Entidad cuente con el número de créditos y recursos necesarios en la nube pública para almacenar, procesar, analizar, consumir y desplegar la información suministrada.
- e) Las herramientas de gestión de los servicios en la nube deberán permitir registrar los detalles de los cambios en sus recursos para proporcionar un historial de configuración; obtener detalles de cómo se veía la configuración de un recurso en cualquier momento en el pasado.
- f) Soportar conexiones dedicadas privadas desde las oficinas de la Entidad hacia el centro de datos del proveedor de nube con el entendido de que este servicio se cobra de forma independiente.
- g) Soportar conexiones VPN para la integración de las oficinas de la Entidad con la plataforma en la nube.
- h) Permitir el aprovisionamiento de certificados SSL, o soluciones similares, para la protección de los balanceadores de carga que forman parte de la arquitectura de nube.
- i) Permitir monitorear la infraestructura para identificar su comportamiento, así como patrones de uso anormales.
- j) Contar con una bitácora de actividades y operaciones realizadas en la nube,

disponible en todo momento para su revisión y análisis de forma manual o con herramientas automatizadas.

- k) Contar con mecanismos para permitir el análisis del tráfico de la red, subredes e instancias habilitadas en la nube.
- l) Contar con capacidad de visualizar los logs de las instancias (sistema operativo) desde la consola web para prevenir la alteración de dichos registros.
- m) Disponer de una herramienta que con base a su ambiente cloud, haga recomendaciones para mejorar la seguridad.
- n) Disponer de herramientas para el reporte del estatus de los sistemas operativo y aplicaciones con el fin de identificar riesgos de seguridad en sistema operativo o problemas con las aplicaciones instaladas contra una base de datos de riesgos y problemas conocidos.
- o) Contar con documentación sobre los procesos de seguridad, así como las mejores prácticas para la implementación de ambientes en la nube.
- p) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe poder acceder a los demás servicios y recursos que conforman la solución.
- q) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe poder ejecutar funciones en balanceadores de carga de aplicaciones.
- r) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe permitir gestionar de forma centralizada código que está distribuido en múltiples funciones.
- s) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe permitir ensamblar e implementar arquitecturas usando aplicaciones anidadas compatibles con el modelo de aplicaciones sin servidor con el fin de lograr un bajo acoplamiento, poder crear aplicaciones más potentes, evitar el trabajo duplicado y ayudar a garantizar la coherencia.
- t) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe permitir el uso de Step Functions para combinar múltiples funciones del servicio en aplicaciones y microservicios sin servidor y con capacidad de respuesta, sin tener que escribir código para la lógica del flujo de trabajo, los procesos paralelos, el manejo de errores, los tiempos de espera o los

reintentos.

- u) El servicio gestionado para ejecutar código sin aprovisionar ni administrar servidores, debe permitir que cada función del servicio tenga una versión única y actual del código. El servicio debe permitir llamar a una versión específica u obtener la última implementación.

2.2.2.1.3. Requerimientos de diseño para la arquitectura

Durante el RFI se espera que cada uno de los proponentes presenten arquitecturas que incorporen las siguientes buenas prácticas:

A. Escalabilidad

Se espera que los sistemas de la Entidad crezcan con el tiempo y por lo tanto deben construirse sobre una arquitectura escalable. Dicha arquitectura debe soportar el crecimiento de usuarios, tráfico o tamaño de datos. El escalamiento debe poder hacerse de forma horizontal y vertical. Adicionalmente, el proponente debe incorporar en lo posible las siguientes buenas prácticas de diseño en la nube:

- **Aplicaciones sin estado:** Cuando sea conveniente y posible las aplicaciones no deben requerir conocimiento de interacciones previas y no deben almacenar información de sesión. Por ejemplo, una aplicación que, dada la misma entrada, proporcione la misma respuesta a cualquier usuario final.
- **Distribución de carga en múltiples nodos:** Cuando sea conveniente y posible se debe distribuir la carga de trabajo a múltiples nodos en el entorno de nube, seleccionando un modelo push o pull según el caso.
- **Componentes sin estado:** Cuando sea conveniente y posible se debe hacer que una parte de estas arquitecturas sea sin estado al no almacenar nada que deba persistir durante más de una solicitud en el sistema de archivos local.
- **Componentes con estado:** Inevitablemente, habrá capas de la arquitectura en las que no es posible implementar componentes sin estado. En estos casos, cuando sea conveniente y posible se debe buscar que la arquitectura permita escalar esos componentes horizontalmente distribuyendo la carga a múltiples nodos con afinidad de sesión.
- **Afinidad de sesión:** Cuando sea conveniente y posible el balanceador de carga de aplicaciones debe intentar usar el mismo servidor para ese usuario durante la sesión. Adicionalmente, cuando sea conveniente y posible se debe diseñar la lógica del usuario para que cuando se detecte la falta de disponibilidad del servidor, los dispositivos se vuelvan a conectar a otro servidor con poca interrupción para la aplicación.
- **Procesamiento distribuido:** Cuando sea conveniente y posible, en los casos de

uso que involucran el procesamiento de grandes cantidades de datos, cualquier cosa que no pueda ser manejada por un solo recurso de cómputo de manera oportuna, debe ser gestionada usando un enfoque de procesamiento distribuido.

B. Recursos temporales en lugar de servidores fijos

Cuando sea conveniente y posible, el enfoque de la arquitectura debe considerar los servidores y otros componentes como recursos temporales. Se deben iniciar tantos recursos como la Entidad necesite y apruebe. Adicionalmente dichos recursos deben ser usados solo durante el tiempo que los necesite.

- **Creación instancias de recursos informáticos:** Cuando se esté implementando el entorno para realizar pruebas o aumentar la capacidad de un sistema se debe, cuando sea conveniente y posible, usar un proceso automatizado y repetible que evite largos plazos de entrega y que sea propenso a errores humanos.
- **Bootstrapping:** Cuando sea conveniente y posible se deben ejecutar acciones automatizadas de arranque, que son secuencias de comandos que instalan software o copian datos para llevar ese recurso a un estado particular.

Adicionalmente, cuando sea conveniente y posible, se deben poder parametrizar detalles de configuración que varían entre diferentes entornos (como producción o prueba) para que pueda reutilizar los mismos scripts sin modificaciones.

- **Golden images:** Cuando sea conveniente y posible, los recursos de nube deben ser iniciados desde una Golden image, que es un snapshot de un estado particular de ese recurso.
- **Contenedores:** Cuando sea conveniente y posible se debe construir e implementar aplicaciones distribuidas dentro de contenedores de software.
- **Infraestructura como código:** Cuando sea conveniente y posible se deben aplicar técnicas, prácticas y herramientas de desarrollo de software para hacer que toda la infraestructura sea reutilizable, mantenible, extensible y comprobable.

C. Automatización

Cuando sea posible y conveniente se deben introducir automatización en la arquitectura de las aplicaciones para garantizar una mayor capacidad de recuperación, escalabilidad y rendimiento.

- **Administración e implementación sin servidor:** Cuando sea posible y conveniente se deben adoptar patrones sin servidor administrados por el CSP, donde el enfoque operativo este en la automatización de la implementación.
- **Gestión de infraestructura e implementación:** Cuando sea posible y conveniente se deben usar servicios que manejen automáticamente todos los detalles, como el aprovisionamiento de recursos, el equilibrio de carga, el escalado

automático y la supervisión.

Adicionalmente, cuando sea posible y conveniente, se deben utilizar servicios que permitan monitorear y generar alarmas sobre los recursos de cómputo para recuperarlos automáticamente en caso de que se deterioren.

De igual forma, cuando sea posible y conveniente, se debe poder recopilar automáticamente el inventario de software, aplicar parches del sistema operativo, crear imágenes del sistema para configurar los sistemas operativos Windows y Linux y ejecutar comandos arbitrarios. El uso de estos servicios debe simplificar el modelo operativo y garantizar la configuración óptima del entorno.

Finalmente, cuando sea posible y conveniente, los recursos de nube deben implementar esquemas de auto-escalamiento que permitan que la capacidad aumente o disminuya de acuerdo con las condiciones definidas por la Entidad.

- **Alarmas y eventos:** Cuando sea posible y conveniente se debe contar con un servicio de alarmas que notifiquen a través de mensajes de texto cualquier variación que se desvíe de los umbrales establecidos por la Entidad.

Adicionalmente, las alarmas generadas deben permitir iniciar acciones o eventos de respuesta cuando sea conveniente y posible.

Por último, el diseño debe permitir, cuando sea necesario y posible, la creación de reglas personalizadas y específicas de la aplicación que bloqueen los patrones de ataque comunes que pueden afectar la disponibilidad de la aplicación, comprometer la seguridad o consumir recursos excesivos

D. Acoplamiento suelto

A medida que aumenta la complejidad de la aplicación, un atributo deseable de la arquitectura propuesta es que se pueda dividir en componentes más pequeños y poco acoplados. Esto significa que los sistemas de TI deben diseñarse de manera que reduzcan las interdependencias; un cambio o una falla en un componente no debe conectarse en cascada a otros componentes.

- **Interfaces bien definidas:** Para reducir las interdependencias en un sistema, cuando sea pertinente y posible, se debe permitir que los diversos componentes interactúen entre sí solo a través de interfaces específicas e independientes de la tecnología, como las API RESTful.
- **Descubrimiento de servicios:** Las aplicaciones que se implementan como un conjunto de servicios más pequeños dependen de la capacidad de esos servicios para interactuar entre sí. Debido a que cada uno de esos servicios puede ejecutarse en múltiples recursos informáticos, es necesario que, cuando sea necesario y posible, haya una forma de abordar cada servicio.
- **Integración asincrónica:** Cuando sea pertinente y posible, el diseño debe permitir la integración asincrónica como otra forma de acoplamiento flexible entre servicios.

Este modelo es adecuado para cualquier interacción que no necesite una respuesta inmediata y donde sea suficiente un reconocimiento de que se ha registrado una solicitud.

- **Mejores prácticas de sistemas distribuidos:** Cuando sea pertinente y posible se deben implementar fallos controlados (Por ejemplo, para las interfaces de usuario, es posible proporcionar contenido alternativo o en caché en lugar de fallar por completo).

E. Servicios, no servidores

Cuando sea necesario y posible, la arquitectura en la nube debe aprovechar los servicios disponibles usando alguna de las alternativas que a continuación se presentan. La primera es a través de servicios administrados que incluyen bases de datos, aprendizaje automático, análisis y más. La segunda forma es reducir la complejidad operativa de ejecutar aplicaciones a través de arquitecturas sin servidor.

F. Bases de datos

La arquitectura de nube debe proponer la tecnología de base de datos adecuada para cada carga de trabajo. Para tal fin se deben tener en cuenta los siguientes criterios:

- ¿Es esta una carga de trabajo de lectura pesada, escritura pesada o equilibrada?
- ¿Cuántas lecturas y escrituras por segundo vas a necesitar?
- ¿Cómo cambiarán esos valores si aumenta el número de usuarios?
- ¿Cuántos datos necesitará almacenar y por cuánto tiempo?
- ¿Qué tan rápido crecerá esto?
- ¿Hay un límite superior en el futuro cercano?
- ¿Cuál es el tamaño de cada objeto (promedio, mínimo, máximo)?
- ¿Cómo se accederá a estos objetos?
- ¿Cuáles son los requisitos en términos de durabilidad de los datos?
- ¿Este almacén de datos será la "fuente de verdad" de la entidad?
- ¿Cuáles son sus requisitos de latencia?
- ¿Cuántos usuarios simultáneos necesita admitir?
- ¿Cuál es el modelo de datos de la Entidad y cómo va a consultar los datos?
- ¿Son sus consultas de naturaleza relacional?
- ¿Podría desnormalizar su esquema para crear estructuras de datos más planas que sean más fáciles de escalar?
- ¿Qué tipo de funcionalidad requiere?
- ¿Se requieren fuertes controles de integridad o se está buscando más flexibilidad (por ejemplo, almacenes de datos sin esquema)?
- ¿Se requiere informes sofisticados o capacidades de búsqueda?
- ¿Los desarrolladores están más familiarizados con las bases de datos relacionales que NoSQL?
- ¿Cuáles son los costos de licencia de tecnología de base de datos asociados?
- ¿Estos costos consideran la inversión en desarrollo de aplicaciones, el almacenamiento y los costos de uso a lo largo del tiempo?
- ¿El modelo de licencia admite el crecimiento proyectado?
- ¿Podría usar motores de base de datos nativos de la nube para obtener la

simplicidad y la rentabilidad de las bases de datos de código abierto?

G. Administrar volúmenes crecientes de datos

Para la administración de grandes volúmenes crecientes de datos se debe implementar una arquitectura de lago de datos cuando sea necesario y posible.

H. Eliminar puntos únicos de falla

La arquitectura debe soportar la falla de un componente individual o de múltiples componentes, como discos duros, servidores y enlaces de red. Para tal fin se deben implementar sistemas con alta disponibilidad o establecer formas de automatizar la recuperación y reducir las interrupciones en cada capa de su arquitectura cuando sea necesario y posible.

- **Redundancia:** Cuando sea necesario y posible se debe incluir redundancia en los componentes de nube; es decir, se debe contar con múltiples recursos para la misma tarea. La redundancia se puede implementar en modo de espera o activo según la necesidad.
- **Detección de fallas:** Cuando sea necesario y posible se debe construir la mayor automatización viable tanto en la detección como en la reacción ante fallas.
- **Almacenamiento de datos duradero:** Cuando sea necesario y posible la arquitectura debe proteger tanto la disponibilidad como la integridad de los datos.

I. Resiliencia automatizada de múltiples centros de datos

Las aplicaciones críticas para el negocio deben contar con protección contra escenarios de interrupción con una baja probabilidad, pero un gran riesgo de impacto, como una catástrofe natural que derriba toda la infraestructura durante mucho tiempo. En tales casos, la arquitectura debe contar con esquemas de resiliencia automatizada en múltiples centros de datos.

J. Optimización de costos

Mediante una metodología iterativa se deben crear arquitecturas con costos optimizados usando los siguientes principios de diseño:

- **Tamaño correcto:** se debe comparar el entorno de la aplicación y seleccionar el tipo de servicio en la nube correcto; es decir, dependiendo de cómo es la carga de trabajo se debe seleccionar la CPU, RAM, red, tamaño de almacenamiento, E / S, etc. apropiados.
- **Aprovechamiento de las opciones de compra:** Se debe seleccionar las opciones de compra más costo-efectivas disponibles para la adquisición de los recursos de nube.

K. Caching

Cuando sea necesario y posible se debe usar caching para mejorar el rendimiento de la aplicación y aumentar la rentabilidad de una implementación. Si es posible y necesario se debe aplicar en múltiples capas de una arquitectura de TI.

2.2.2.1.4. Requerimientos no funcionales

- La solución debe incluir un servicio que extraiga automáticamente texto y datos desde documentos escaneados.
- La extracción de texto debe permitir identificar el contenido de campos de formularios e información almacenada en tablas.
- El servicio de extracción de texto debe usar aprendizaje automático para "leer" documentos a fin de extraer texto y datos de forma precisa sin necesidad de cambios manuales ni utilizar código personalizado.
- El servicio de extracción de texto debe permitir automatizar rápidamente flujos de trabajo con documentos
- El servicio de extracción de texto debe permitir usar los datos extraídos dentro de aplicaciones. Además, puede crear índices de búsqueda inteligente y flujos de trabajo de aprobación automatizados.
- El servicio de extracción de texto debe permitir detectar automáticamente el diseño de un documento y los elementos clave en la página, entender las relaciones de datos en cualquier formulario o tabla integrados, y extraer todo con su contexto
- Debe contar con tecnologías OCR e ICR para procesar los documentos de las providencias que el usuario adjunte.
- Se debe incluir servicios web REST y/o SOAP para garantizar la consulta de demandas, metadatos y transcripción, los cuales serán expuestos a través del bus de servicios de la entidad.
- Dado que se debe garantizar su consulta en el futuro, la solución debe permitir la conservación de los documentos electrónicos en el formato en el que hayan sido elaborados, transformados, enviados recibidos, y preferentemente en un formato correspondiente a un estándar abierto que preserve a lo largo del tiempo la integridad del contenido de los documentos y de los metadatos que la acompañan. Los formatos propuestos son:

A. Archivos texto en formato pdf, docx, doc, txt, rtf.

- Debido a la gran cantidad de archivos que se deben almacenar y que estos deben permanecer en el tiempo, se debe optimizar su consumo de almacenamiento.
- Se debe entregar documentación técnica (arquitectura del sistema y manuales

técnicos) y manuales de usuario.

- Se debe realizar actividades de transferencia de conocimiento a nivel técnico y funcional.
- Se debe hacer uso de buenas prácticas y patrones de desarrollo de Software para garantizar que el sistema tenga un rendimiento óptimo, sea seguro, robusto, escalable, esté disponible en el momento adecuado y en los tiempos esperados.
- La arquitectura de la solución debe garantizar que cuando se realicen cambios en cualquiera de sus componentes afecten de manera mínima la integración y funcionamiento de sus otros componentes.
- Se debe entregar un informe de la ejecución de pruebas de stress sobre el sistema para garantizar los tiempos de respuesta de la aplicación.
- Se espera no tener vulnerabilidades nivel medio o superiores al realizar un análisis con Owasp Top 10.
- Soportar pentest (pruebas de penetración).
- El sistema debe soportar certificado para sitio seguro.
- Incluir mecanismos de seguridad como autenticación segura, autorización, auditoría y encriptación de los datos en tránsito.
- Mecanismos y opciones de llevar registros de auditoría de ingreso y utilización de los diferentes módulos del sistema y reportes de auditoría.
- Incluir una cláusula de optimización de costos para dejar abierta la opción de usar instancias reservadas.
- La base de datos relacional nativa en la nube, debe soportar una arquitectura sin servidor que sea escalable bajo demanda con el fin de optimizar el uso y costo de los recursos.
- El servicio de búsqueda elástica gestionada debe permitir hacer actualizaciones en los clúster que soportan el servicio sin generar tiempos de caída del servicio.
- El servicio de búsqueda elástica gestionada debe permitir que de forma automática ante el fallo de un nodo la carga de servicio se distribuya en los demás nodos y que se creen nodos nuevos para reemplazar a los que fallaron.
- El servicio de búsqueda elástica gestionada debe permitir configurar el servicio para acceso a través de una VPN
- El servicio de búsqueda elástica gestionada debe permitir monitorear el rendimiento de varias etapas de una operación a través de registros lentos de dos tipos:

- Registros lentos de índice: suministran información acerca del proceso de indexación y se pueden utilizar para ajustar la configuración del índice.
- Registros lentos de búsqueda: suministran información acerca de la velocidad con la que se realizan consultas y recuperaciones. Estos registros ayudan a ajustar el rendimiento de cualquier tipo de operación de búsqueda en el servicio

2.3. Condiciones Generales.

Las partes y en especial el proveedor se obliga a cumplir con el objeto del contrato, así como todas las obligaciones previstas en el Acuerdo Marco de Precios -CCE-908-1-AMP-2019, que se encuentra consignado en el siguiente link:

https://www.colombiacompra.gov.co/sites/cce_public/files/cce_documentos/49_minuta_que_incorpora_adenda_2_nube_publica_iii.pdf

Nota: El Acuerdo Marco de Precios - CCE-908-1-AMP-2019 al que se acogerá la Contraloría General de la República como Entidad Compradora, a través del presente proceso de contratación, tuvo como origen la operación principal efectuada por Colombia Compra Eficiente la que llevó a cabo mediante el proceso de Licitación Pública N°CCENEG-015-1-2019 en cumplimiento del artículo 2.2.1.2.1.2.10 del Decreto 1082 de 2015 y el numeral 7 del artículo 3 del Decreto 4170 de 2011, proceso que permea el Acuerdo y por ende las operaciones que registre a través de él.

3. MODALIDAD DE CONTRATACIÓN

La Contraloría General de la República, en atención a la naturaleza de los bienes y servicios objeto de este proceso, considera que la modalidad de contratación aplicar es la Selección Abreviada para la adquisición de bienes y servicios de características técnicas uniformes por Catálogo derivado de la celebración de Acuerdo Marco de Precios para la Prestación de Servicios de Nube Pública III (CCE-908-1-AMP-2019), de conformidad con el literal a) del numeral 2 del artículo 2 de la Ley 1150 de 2007 y el artículo 2.2.1.2.1.2.7 del Decreto 1082 de 2015. Dentro de los beneficios que trae la incorporación al Acuerdo Marco de Precios CCE-908-1-AMP-2019, se pueden resaltar los siguientes:

- Las Entidades compradoras adquieren mejor tecnología con mayor valor por dinero en menor tiempo.
- Se genera la dinamización del mercado, tanto para los proveedores como para las entidades.
- Se Mejora el servicio al ciudadano apoyado en las TI.
- Se Optimiza los tiempos y esfuerzos de contratación.
- Las entidades se pueden concentrar en definir las necesidades que le agreguen valor.

- Se reduce el número de procesos de contratación.
- El Estado actúa como un único comprador, unificando términos y condiciones para el suministro de bienes y servicios para el Estado.
- Permite a la entidad manejar inventarios con mayor flexibilidad.
- Libera tiempo del comprador público para destinar al cumplimiento de la misión de la entidad.
- Ayuda a obtener mayor valor de los recursos públicos.

Adicionalmente se cuenta con la garantía que existe una participación del Administrador del Acuerdo Marco de Precios para:

- Verificar el cumplimiento del tiempo de aprovisionamiento, alcance y recursos contratados.
- Validar el cumplimiento de los Acuerdos de Nivel de Servicio (ANS).
- Realizar pruebas técnicas que soporten la validación de los ANS.
- Servir de aval para la aceptación de los servicios y productos contratados.
- Actualizar el catálogo para incluir los ajustes de las fichas técnicas y precios o cualquier modificación a que haya lugar.
- Revisar periódicamente los precios del mercado para asegurar que las condiciones de los AMP sean competitivas.
- Imponer las sanciones que correspondan, incluso la exclusión de proveedores del catálogo, de forma temporal o definitiva.
- Administrar las garantías.
- Prorrogar la vigencia de los AMP.
- Manejar las peticiones, quejas, reclamos y solicitudes relacionadas con los AMP.

Aunado a lo anterior, el Acuerdo Marco de Precios para la Prestación de Servicios de Nube Pública III CCE-908-1-AMP-2019, se encuentra vigente para colocar órdenes de compra hasta el 25 de octubre de 2021, y para ejecutar las órdenes de compra hasta el 25 de octubre de 2022, conteniendo en su catálogo los servicios que necesita la Dirección de Información, Análisis y Reacción Inmediata -DIARI-

3.1. Identificación del contrato a celebrar

Se suscribirá un contrato de Suministro de Servicios regulado por la Ley 80 de 1993, y el Código de Comercio.

4. PRESUPUESTO OFICIAL

La Contraloría General de la República estimó el presupuesto oficial para el presente proceso de contratación en la suma de *MIL SEISCIENTOS CINCUENTA MILLONES DE PESOS M/CTE (\$1.650.000.000)*.

La estimación del presupuesto se realizó teniendo en cuenta dos factores. El primer factor se realizó utilizando las referencias de adquisiciones en la nube pública con base en los valores de contratos con objetos similares celebrados por otras entidades⁸ dentro del acuerdo marco de la nube pública III. El segundo factor se obtuvo con la construcción de la arquitectura de referencia, el diagrama de contexto y el alcance de uso definido en los 14,000,000 de hojas objetivo, donde se identifican y se estimaron los componentes y los costos más relevantes de la solución. Como los servicios están expuestos al riesgo

cambiarlo se realizó el cálculo teniendo en cuenta la volatilidad y el nivel actual de la tasa de cambio. La tasa de cambio de referencia utilizada en el cálculo de la estimación del presupuesto fue COP/USD 4.000. Este presupuesto estimado es respaldado con el Certificado de Disponibilidad Presupuestal No. 96320 de 27 de julio de 2020, expedido por la Dirección de Recursos Financieros.

El proceso interno de cálculo para estimar el presupuesto inicial, teniendo en cuenta que el RFP deberá tener la aproximación real del costo, es:

Componentes	Cantidad	Costo Estimado	Costo Total
Herramientas de extracción, transformación OCR	14.000.000	0,015	203.000
Data Lake (storage, disponibilidad y carga)	20 TB	1.000	20.000
Machine Learning (Frontend - Backend) Usuer Go	14.000.000	0,0135	189.000
Total USD			412.000
TRM Estimada	4.000	Total COP	1.648.000.000

Los cálculos son estimaciones internas de la DIARI.

Una vez autorizada la contratación por la Junta de Adquisiciones de la Contraloría General de la República se procederá con el trámite respectivo ante la tienda virtual y una vez sea elegida la CSP se descargará el simulador y, se le tendrá en cuenta el valor de la TRM, que corresponda para ese momento, siguiendo el procedimiento establecido en el guía para la compra en la tienda virtual, emitida por Colombia Compra Eficiente.

El valor final de la orden de compra deberá estar expresado en pesos colombianos e incluir el impuesto del IVA y podrá superar el valor estimado en el presupuesto asignado.

⁸ Fuente Colombia Compra Eficiente – (https://www.colombiacompra.gov.co/tienda-virtual-del-estado-colombiano/ordenes-compra?page=1&number_order=&state=&entity=&tool=Nube%20Pública%20III&date_to=&date_from=) SUPERINTENDENCIA DE SOCIEDADES orden de compra 43474/ FONDO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES orden de compra 51730 / ICFES orden de compra 44219

El valor de la orden de compra respecto del Cloud Service Provider encuentra exenta del IVA, de conformidad con el numeral 21⁹ del artículo 476 del Estatuto Tributario en el marco de la aplicación de la exclusión de IVA para los servicios de computación en la nube, hosting, suministro de páginas web y mantenimiento a distancia de programas y equipos, concordante con los siguientes conceptos de la DIAN: i) el Concepto Unificado 017056, del 25 de agosto de 2017, y ii) el Concepto 100202208-1444, del 22 de diciembre de 2017.

De acuerdo con lo anterior la Contraloría General de la República entiende que la exclusión mencionada se aplicaría en el valor de la eventual orden de compra. En cuanto a los componentes de Servicios complementarios estos serían objeto de IVA.

5. FORMA DE PAGO.

El valor de la orden de compra será cancelado en *un solo pago* de conformidad con las disposiciones establecidas en el Acuerdo Marco de Precios Nube Pública III clausula 10.

Para que se lleve a cabo el pago el proveedor deberá presentar a la Contraloría General de la República, la factura correspondiente, acompañada de la certificación de cumplimiento a satisfacción expedida por el supervisor del contrato y del certificado del Representante Legal o del Revisor Fiscal; en el que se deje constancia de que el contratista se encuentra a paz y salvo por concepto de seguridad social y parafiscales, de conformidad con el artículo 50 de la Ley 789 de 2002 y el artículo 23 de la Ley 1150 de 2007.

La factura debe contener los requisitos establecidos en el Estatuto Tributario y demás normas que lo modifiquen, adicionen o aclaren. En todo caso, el pago estará sujeto a la aprobación del PAC (Programación Anual Mensualizada de Caja) y a las retenciones y deducciones de la ley a que hubiera lugar.

6. PLAZO

El plazo de ejecución de la orden de compra, no podrá exceder los dos (2) meses calendario, contados a partir del cumplimiento de los requisitos de perfeccionamiento y ejecución de la orden de compra.

⁹ Art. 476. Servicios excluidos del impuesto sobre las ventas. **Nota 1.** Se exceptúan del impuesto los siguientes servicios y los bienes relacionados explícitamente a continuación:

(...)

21. Suministro de páginas web, servidores (hosting), computación en la nube (cloud computing).

7. LUGAR DE EJECUCIÓN

El lugar de ejecución será en la ciudad de Bogotá D.C. en la sede nivel central de la Contraloría General de la República del nivel central, esto es, en la Carrera 69 No. 44.-35.

8. SUPERVISIÓN

La supervisión será ejercida por el Jefe de la Unidad de Información y el Jefe de la Unidad de Análisis de Información de la Dirección de Información, Análisis y Reacción Inmediata o el quien designe el ordenado del gasto¹⁰.

Los supervisores podrán designar personal de apoyo para el cumplimiento de la labor encomendada y, en todo caso, sin delegar su responsabilidad, el cual deberá ceñirse al reglamento de la entidad, la Ley 1474 de 2011 y demás normas concordantes.

9. GARANTÍAS.

De conformidad con el numeral 19.2 del Acuerdo Marco de Precios Nube Pública III CCE-908-1-AMP-2019, el proveedor deberá constituir la garantía única a favor de esta Entidad, de la siguiente manera:

Amparo	Suficiencia	Vigencia
Cumplimiento del contrato	15% del Valor de la Orden de Compra	Duración de la Orden de Compra y un (1) año más
Calidad de los Bienes	20% del Valor de la Orden de Compra	Duración de la Orden de Compra y un (1) año más
Pago de salarios, prestaciones sociales legales e indemnizaciones laborales	5% del Valor de la Orden de Compra	Duración de la Orden de Compra y (3) tres años más

Tabla 4: Suficiencia de la garantía a favor de las Entidades Compradoras - Fuente Colombia Comprar Eficiente Acuerdo Marco de Precios Nube Pública III

El valor de los amparos de la garantía de cumplimiento es calculado de acuerdo con el valor de la Orden de Compra.

¹⁰ De conformidad con la Resolución N°191 de 2015 modificada por la Resolución N°0718 de 2019.

La vigencia de la garantía y sus amparos debe iniciar desde la colocación de la Orden de Compra. Los Proveedores deberán ampliar la garantía dentro de los tres (3) días hábiles siguientes a la fecha en la que la Orden de Compra sea modificada, adicionada y/o prorrogada. La vigencia de la garantía debe ser ampliada por los plazos señalados en la tabla anterior. En todo caso de conformidad al Decreto 1082 de 2015 la garantía de cumplimiento debe estar vigente hasta la liquidación.

En caso de declaratoria de incumplimiento que afecte la garantía de cumplimiento o alguno de sus amparos, el Proveedor deberá ajustar la suficiencia de la garantía, en los amparos respectivos, de forma tal que cumpla con lo señalado en la Tabla después de haber sido afectada.

Anexos del Estudio Previo:

- Anexo N°1 Ficha técnica
- Anexo N°2 Simulador para estructurar la compra –(Este es un documento interno que se publicará durante el proceso teniendo en cuenta el resultado del RFI)
- Anexo N°3 Condiciones transversales del servicio
- Anexo N°4 CDP 96320
- Anexo N°5 Ficha EBI
- Anexo N°6 Formato de Justificación de elección de CSP para Entidades Estatales (Este es un documento interno que se publicará durante el proceso y se diligencia con base a las cotizaciones de los proveedores)
- Anexo N°7 Formato RFI para estructurar la compra.



Hoslander Adlai Saenz Barrera
Jefe de Unidad
Unidad de Información
DIARI



Oswaldo Juan Zapata Quijano
Jefe de Unidad
Unidad de Análisis
DIARI

Elaboró: Alejandro Falla
Elaboró: Paul Ordosgoitia A
Revisó: Oswaldo Zapata Quijano
Revisó: Hoslander Adlai Saenz Barrera