

GOBIERNO DE DATOS Y ANALÍTICA INSTITUCIONAL

(PLATAFORMA TIPO DATA FABRIC)

SERVICIO DE RENTAS INTERNAS
ECUADOR

SOLICITUD DE INFORMACIÓN (RFI)

Fecha de elaboración: 4 de abril de 2023

Tabla de Contenido

1	INTRODUCCIÓN	6
1.1	Objetivo	6
1.2	Cronograma	8
1.3	Forma de presentación	8
1.4	Confidencialidad de la información	8
1.5	Idioma de la respuesta al documento	9
2	ANTECEDENTES	10
2.1	Información general del SRI	10
2.2	Necesidad principal	11
2.3	Necesidades específicas	11
3	REQUERIMIENTOS DE LA SOLUCIÓN	13
3.1	Plataforma Tipo Data Fabric	15
3.1.1	Requisitos Generales	29
3.1.2	Virtualización de Datos	33
3.1.3	Datos Federados	34
3.1.4	Capacidades de Almacenamiento	36
3.1.5	Gobierno del Dato	37
	Gobierno del Dato	37
	Data Catalogue y Gestión del Metadato	39
	Dominios y colecciones de datos	41
	Glosario de negocio	42
	Ontologías	44
	Knowledge Graph	44
3.1.6	Calidad del Dato	45
3.1.7	Linaje	47
3.1.8	MDM – Datos Maestros	48
3.1.9	Data Marketplace	49

3.1.10	Seguridad y Cumplimiento del Dato	51
3.1.11	Data Privacy	52
3.1.12	Capacidades Analíticas	53
3.1.13	Capacidades ETL.....	55
3.1.14	Preprocesamiento de datos para modelos	56
3.1.15	Machine Learning.....	57
3.1.16	Capacidades incluidas aplicaciones operacionales	61
3.1.17	Soporte y capacitación	62
3.2	Migración de Datos y/o reingeniería de flujo de datos.....	62
3.2.1	Información de referencia de volúmenes de datos actuales	62
3.3	Servicios de Gestión.....	64
3.4	Requerimientos Técnicos.....	67
3.4.1	Tecnológicos	67
3.4.2	Monitoreo.....	68
3.4.3	Seguridad.....	69
3.4.4	Sistema de Gestión de Identidades.....	69
3.4.5	Gestión de Eventos e Información de Seguridad.....	70
3.4.6	Interoperabilidad	70
3.4.7	Integración con terceros	70
3.4.8	Implementación.....	71
3.4.9	Soporte y Mantenimiento	71
3.4.10	Licenciamiento o Suscripción.....	71
3.5	Atributos de calidad de la solución.....	72
3.5.1	Disponibilidad	72
3.5.2	Interoperabilidad	73
3.5.3	Mantenibilidad.....	73
3.5.4	Rendimiento	74
3.5.5	Seguridad.....	75
3.5.6	Verificabilidad	76
3.5.7	Usabilidad	77

4	CAPACITACIÓN Y TRANSFERENCIA DE CONOCIMIENTOS	78
5	PREGUNTAS ESPECÍFICAS.....	79
	5.1 Experiencia.....	79
	5.2 Servicio	79
	5.3 Equipo y recursos	80
	5.4 Información de la solución tecnológica	80
	5.5 Estimación de la Inversión	82
6	INFORMACIÓN REQUERIDA POR EL SRI.....	85
7	CONDICIONES GENERALES.....	86
	7.1 Inscripción.....	86
	7.2 Fecha de entrega de la información solicitada por el SRI	86
	7.3 Fechas para la presentación y demos de la solución tecnológica propuesta ...	86
8	ANEXOS.....	87
	Anexo 1. Formato de respuestas para requerimientos	87
	Anexo 2. Situación tecnológica actual de la Institución	89
	Anexo 3. Datos estadísticos institucionales relevantes	97
	Anexo 4. Glosario de términos.....	100

Índice de Gráficos:

Gráfico 1: Esquema General Plataforma Tipo Data Fabric	16
Gráfico 2: Arquitectura referencial del Sistema Integrado de Gestión Tributaria..	29
Gráfico 3: DAMA-DMBOK2 Marco de Referencia de Gestión de Datos.....	66
Gráfico 4: Componentes macro de la arquitectura	89
Gráfico 5: Componentes arquitectura de software	90
Gráfico 6: Arquitectura Datawarehouse macro.....	91
Gráfico 7: Componentes de Seguridad Informática por Ámbito	92
Gráfico 8: Ambientes de los Centros de Datos	93
Gráfico 9: Comunicación entre componentes	94
Gráfico 10: Arquitectura de Comunicaciones (Ámbito y Componentes)	95
Gráfico 11: Componentes interacciones externas.....	96

Índice de Tablas:

Tabla 1: Matriz de cumplimiento.....	87
Tabla 2: Total contribuyentes activos	97
Tabla 3: Número de declaraciones recibidas por año.....	97
Tabla 4: Número de comprobantes electrónicos emitidos por año	98
Tabla 6: Usuarios internos a nivel nacional	99

1 INTRODUCCIÓN

1.1 Objetivo

El Servicio de Rentas Internas (SRI), como entidad pública en la administración y gestión de los impuestos en el Ecuador está interesada en realizar un estudio de mercado con el fin de contratar el diseño, aprovisionamiento, implementación, administración, soporte, mantenimiento y servicios de gestión de una Plataforma bajo la arquitectura tipo Data Fabric¹.

Para lograr este objetivo, el SRI envía a los interesados en participar en un posterior proceso de contratación, el siguiente documento de solicitud de información (RFI por sus siglas en inglés) para que remitan la información necesaria que facilite el análisis de las soluciones disponibles en el mercado y la elaboración de los términos de referencia.

El proceso de contratación formará parte de las inversiones que hace el SRI utilizando recursos del contrato de préstamo con el BID y que, si bien en esta fase se recaba información para robustecer los TDRS, el proceso de selección se regirá por las políticas de adquisiciones del BID.

Las respuestas recibidas, así como las presentaciones y demos que se realicen por parte de los interesados serán solamente para un fin de investigación y análisis de factibilidad, las cuales no tendrán ningún vínculo con el posterior proceso de contratación.

Esta solicitud tiene como alcance conocer a los fabricantes, partners y/o proveedores con experiencia en la prestación de servicios en el suministro e implantación de soluciones informáticas orientadas a la Gobernanza, análisis, extracción, almacenamiento y procesamiento masivo, analítica avanzada de información y visualización de resultados, que cuenten con experiencia demostrada en el diseño e implementación de soluciones

¹ Gartner define el concepto Data Fabric como un concepto de diseño que sirve como una capa integrada (tejido) de datos y procesos de conexión. Una estructura de datos utiliza análisis continuos sobre activos de metadatos existentes, detectables e inferidos para respaldar el diseño, la implementación y la utilización de datos integrados y reutilizables en todos los entornos, incluidas las plataformas híbridas y de múltiples nubes. <https://www.gartner.com/smarterwithgartner/data-fabric-architecture-is-key-to-modernizing-data-management-and-integration>

tecnológicas similares, de tal forma que puedan entregar al Servicio de Rentas Internas, una propuesta completa, cuyo alcance se remita a los siguientes puntos:

- Conocer las funcionalidades sobre los tipos de soluciones o alternativas tecnológicas que ofrecen los fabricantes, partners y/o proveedores para la implementación de la Plataforma tipo Data Fabric en el Servicio de Rentas Internas.
- Obtener información sobre los conocimientos, experiencias, certificaciones y alianzas establecidas por los posibles fabricantes, partners y/o proveedores en la planeación, dirección, ejecución y control de proyectos relacionados con la implementación de plataformas tipo Data Fabric, mediante la exposición de casos exitosos en otras administraciones tributarias, gubernamentales y/o entidades privadas.
- Obtener información sobre los roles, niveles de conocimiento, experiencia y certificaciones requeridas para conformar equipos de trabajo que lleven a cabo las labores requeridas para la implementación y ejecución de la plataforma tipo Data Fabric.
- Conocer el tiempo aproximado de la implementación del servicio requerido por el Servicio de Rentas Internas, en función de los alcances propuestos en este documento y basados en las experiencias exitosas en otras administraciones tributarias, gubernamentales o entidades privadas.
- Conocer el tipo y estándares de seguridad que utiliza la plataforma tipo Data Fabric ofertada y su grado de compatibilidad e integración con sistemas propios del Servicio de Rentas Internas.
- Obtener información sobre metodologías o mejores prácticas adoptadas, tipo de infraestructura de hardware y software que soporte la gestión y prestación de los servicios o solución, arquitectura de datos e infraestructura requerida, esquemas de facturación, licenciamiento, suscripción y costo del servicio o solución, entre otros aspectos.
- Obtener el dimensionamiento de capacidades y costos a nivel de hardware, software y servicios de implementación, para cubrir las necesidades planteadas.

1.2 Cronograma

A continuación, se detallan las fechas planificadas para la presentación del RFI:

Fecha de lanzamiento del RFI: 25 de abril de 2023

Fecha máxima para realizar preguntas: 12 de mayo de 2023

Fechas límite para absolver preguntas por parte del SRI: 19 de mayo de 2023

Fechas límite para envío de respuestas al RFI por parte del oferente: 29 de mayo de 2023

Fechas para realizar demos y presentaciones: desde el 08 de junio de 2023 hasta el 07 de julio de 2023.

El SRI se reserva el derecho de analizar las respuestas de los interesados al RFI y de solicitar las aclaraciones, demos o presentaciones que a su juicio se requieran.

1.3 Forma de presentación

Todas las interacciones entre el SRI y los interesados en participar en este requerimiento deberán ser canalizadas a través del correo programaintax@sri.gob.ec, el cual es gestionado por el equipo del Programa de transformación digital “In Tax” del SRI.

No se aceptarán respuestas al RFI que se entreguen por un medio diferente o que se entreguen en papel en las dependencias del SRI.

Para realizar las preguntas, en el término establecido para el efecto, y enviar la respuesta al RFI, se deberá indicar en el asunto del correo – antes de cualquier referencia – lo siguiente: RFI No. 003-2023- PLATAFORMA TIPO DATAFABRIC.

1.4 Confidencialidad de la información

Toda la información que entregue y solicite el Servicio de Rentas Internas y la que suministren los fabricantes, partners y/o proveedores se considerará como confidencial y no podrá ponerse a disposición de terceros para su uso. Los fabricantes, partners y/o proveedores deberán mantener la confidencialidad sobre toda la información que

suministre el SRI en la presente solicitud de información, no la utilizarán para la presentación de sus servicios en otras organizaciones ni la utilizarán para cualquier otro fin distinto al que acá se pretende.

Los potenciales fabricantes, partners y/o proveedores de los servicios requeridos y el SRI se comprometen a mantener en secreto esta información y deberán adoptar todas las medidas razonablemente necesarias para garantizar que ella no sea revelada o divulgada por sus servidores en cumplimiento del compromiso acá establecido.

1.5 Idioma de la respuesta al documento

Las respuestas a la solicitud de información y toda la correspondencia o documentos relativos a éste deben redactarse en español, incluyendo la información técnica. La información referida a las certificaciones solicitadas se podrá presentar en español o en inglés.

2 ANTECEDENTES

2.1 Información general del SRI

El Servicio de Rentas Internas fue creado el 2 de diciembre de 1997 basándose en los principios de justicia y equidad, como respuesta a la alta evasión tributaria, alimentada por la ausencia casi total de cultura tributaria. Desde su creación se ha destacado por ser una institución independiente en la definición de políticas y estrategias de gestión que han permitido que se maneje con equilibrio, transparencia y firmeza en la toma de decisiones, aplicando de manera transparente tanto sus políticas como la legislación tributaria.

En 2020 el SRI se constituyó en la primera institución ganadora del Premio Ecuatoriano de Calidad y Excelencia en el Servicio Público, consolidándose como un referente nacional en calidad, eficiencia y liderazgo.

Durante los últimos años se evidencia un enorme incremento en la recaudación de impuestos. Entre los años 2000 y 2021 la recaudación tributaria a cargo del SRI ha superado los USD 189 mil millones. La cifra alcanzada por el SRI se debe principalmente a la eficiencia en la gestión de la institución, a las mejoras en procesos de control y asistencia a los contribuyentes, a la automatización de varios servicios que facilitan el cumplimiento tributario voluntario, así como a las acciones para fortalecer la transparencia fiscal y combatir las prácticas evasivas.

El Servicio de Rentas Internas (SRI) para incrementar los ingresos tributarios como porcentaje del Producto Interno Bruto (PIB) y aumentar la competitividad económica con un mejor clima para el comercio exterior y la inversión ha generado el Programa de transformación digital “In Tax” enfocado a la innovación y mejora en los procesos de control tributario, a través de la adquisición (Licenciamiento y/o servicio) de una solución tecnológica integral “Plataforma tipo Data Fabric”, así como, las adquisiciones (Licenciamiento y/o servicio) de componentes y herramientas complementarias que apalancarán dichos objetivos.

2.2 Necesidad principal

Los sistemas de información del Servicio de Rentas Internas, con el paso del tiempo y fruto de la mejora continua, se han convertido en repositorios de información y fuentes de conocimiento en forma de silos, pero que son fundamentales para el apoyo en la toma de decisiones basadas en datos.

Asimismo, los problemas en los que hay que realizar análisis exhaustivos y sopesar gran cantidad de posibilidades se han vuelto transversales, la localización, descripción, recopilación, clasificación, catalogación, disponibilización y gobierno de datos, se hace cada vez más latente y necesaria que nunca, gestionarlos.

En este sentido, el Servicio de Rentas Internas busca fortalecer su Modelo de Gobierno de Información y Analítica Avanzada a través de una metodología y soluciones automáticas o servicios que permitan adoptar una estrategia basada en la recolección, extracción, almacenamiento, procesamiento y análisis de grandes volúmenes de datos para una adecuada toma de decisiones (Plataforma tipo Data Fabric).

2.3 Necesidades específicas

- Reducir notablemente los tiempos de procesamiento de la información que dispone el SRI, considerando que la misma es voluminosa y se encuentran almacenadas en diferentes bases de datos, a fin de mejorar la oportunidad en la predicción de Modelo de Riegos y toma decisiones.
- Disponer de una capa unificada de datos, que represente la única fuente de la verdad, que proporcionen vistas de negocio (ontologías) para que sean conocidos y utilizados por todos los procesos internos de la institución.
- Contar con un módulo de Gobierno de Datos que permita manejar y administrar: dominios de información, diccionario, datos maestros, modelos conceptuales, calidad de datos, roles y responsabilidades (data owner's data Stewart) y la definición de políticas de gobernanza.

- Utilizar fuentes externas disponibles como datos de redes sociales, datos no estructurados como imágenes, documentos, videos y otras fuentes de datos con la finalidad de generar modelos de predicción, clasificación, grafos, redes, segmentación y minería de texto que fortalezcan las competencias propias de la Administración Tributaria.
- Contar con una infraestructura tecnológica especializada en el campo de analítica avanzada.
- Contar con la arquitectura de información, así como repotenciar las herramientas para la extracción, integración, procesamiento y visualización de una forma estandarizada y centralizada.
- Fortalecer las habilidades y capacidades de los servidores con perfiles destacados en gobierno de información y en procesamiento, y análisis de BigData y Analítica Avanzada así como también, empoderar a los "científicos de datos funcionales", servidores públicos que entienden los casos de uso de la Administración Tributaria y los datos, pero que no tienen las habilidades de codificación de los "científicos de datos programáticos".

A fin de evaluar las diferentes alternativas en la implementación de la plataforma tipo Data Fabric, el SRI requiere que las propuestas presentadas consideren los siguientes escenarios:

- Implementación en la infraestructura tecnológica del SRI (On premise)
- Implementación en nube (Cloud)

3 REQUERIMIENTOS DE LA SOLUCIÓN

Para el diseño, aprovisionamiento, implementación, administración, soporte, mantenimiento y servicios de gestión de la Plataforma se han contemplado tres componentes:

1. Solución Tecnológica tipo Data Fabric: Comprende el diseño, aprovisionamiento, la implementación, administración y soporte de los componentes y las herramientas tecnológicas necesarias para soportar los procesos de gestión de datos, flujos de información y analítica, a lo largo del ciclo de vida de los datos, para lo cual el fabricante, partner y/o proveedor deberá:
 - Diseñar e implementar la arquitectura de datos de la plataforma DataFabric, que integre las fuentes de información y que provea datos consistentes a los servicios internos y externos del SRI, garantizando la seguridad en el manejo de los datos e información y cumpliendo con los lineamientos de gobernanza de datos que establezca el SRI.
 - Proveer, implementar y soportar los componentes tecnológicos y procesos para la integración y transformación de datos, de las fuentes identificadas a la Plataforma DataFabric, así como la solución de virtualización de datos para fortalecer y facilitar la disponibilidad de la información.
 - Diseñar, implementar y soportar los componentes tecnológicos y procesos para el almacenamiento de información en soluciones de Data Warehouse, Data Lakes y/o Data Marts, para el uso en la operación del SRI y la analítica de datos.
 - Proveer e implementar herramientas tecnológicas para la visualización, modelado, análisis descriptivo, predictivo y prescriptivo de datos, desarrollo de ML y AI, despliegue, puesta en producción y reentrenamiento de modelos, que permitan apoyar la toma de decisiones.
2. Migración y/o reingeniería de datos o flujos de datos más relevantes: Comprende la definición y ejecución de los planes de trabajo para la migración y/o reingeniería de los flujos de datos de la información existente en el SRI relacionados a los casos

de uso, desde las fuentes de origen hacia la plataforma tipo Data Fabric. Esta migración y/o reingeniería deberá garantizar la integridad y la calidad de los datos, y la disponibilidad de la información tanto para los sistemas y aplicaciones que hoy día consumen y alimentan las diversas fuentes de origen, así como para los nuevos sistemas que se desarrollarán en el marco de la modernización tecnológica para el SRI.

3. Servicios de gestión de datos: Comprende el diseño, ejecución y entrega del modelo de operación y gestión de datos para el SRI en concordancia con las políticas de gobernanza, calidad, seguridad y demás sistemas de gestión que en el SRI se definan. Comprende la definición y ejecución de los procesos, procedimientos y demás instrumentos que sean necesarios basados en el marco de referencia de gestión de datos del DAMA- DMBOK2. El fabricante, partner y/o proveedor deberá:

- Diseñar, proveer, implementar, soportar y mantener una solución de gestión de datos unificada, consistente, efectiva y administrable que permita potenciar el aprovechamiento de los datos de la Entidad para volver más eficientes los servicios institucionales y los procesos tributarios.
- Facilitar el manejo y explotación de la información institucional, así como la gestión de datos históricos de la entidad, a través de la provisión e implementación de herramientas y procesos de gestión de datos maestros (MDM) y de metadatos unificados, garantizando su consistencia a lo largo de los flujos de información.
- Proveer e implementar procesos y procedimientos que generen capacidades que permitan gestionar los datos a lo largo de todo su ciclo de vida, aplicando las mejores prácticas y atendiendo a los lineamientos de gobierno digital del país.

A continuación se indicarán para cada uno de estos componentes las especificaciones mínimas que se deben cumplir y posteriormente se indicarán los requerimientos transversales relacionados con metodologías, calidad y niveles de servicio que aplican a los tres componentes.

3.1 Plataforma tipo Data Fabric

El tipo Data Fabric está diseñado para ayudar a las organizaciones a resolver problemas de datos complejos y de casos de uso mediante la administración de sus datos, independientemente de los diversos tipos de aplicaciones, plataformas y ubicaciones donde se almacenan estos datos.

El objetivo principal de la plataforma es extraer valor a los datos. Para ello, la plataforma debe facilitar tanto la gestión del ciclo de vida de los datos como su posterior exploración, y comprensión de éstos. Tampoco hay que perder de vista que los datos se deben usar para tomar decisiones a partir de ellos por lo que el volumen de datos presente requiere que la plataforma acometa esta tarea de manera unificada y escalable para ayudar a la administración a realizar esta toma de decisiones de la mejor manera posible y en entornos con un volumen de datos elevado. La plataforma tipo Data Fabric debe permitir el acceso y el intercambio de datos en un entorno de datos distribuidos y proporcionar un marco de gestión de datos único y coherente, que establezca el acceso y el procesamiento de datos sin problemas de los diferentes almacenamientos que posea de la Entidad. Este acceso a los diferentes almacenamientos presentes debe permitirse tanto a través de una virtualización del dato como de flujos de copia y sincronización de los datos presentes a otros sistemas de almacenamiento.

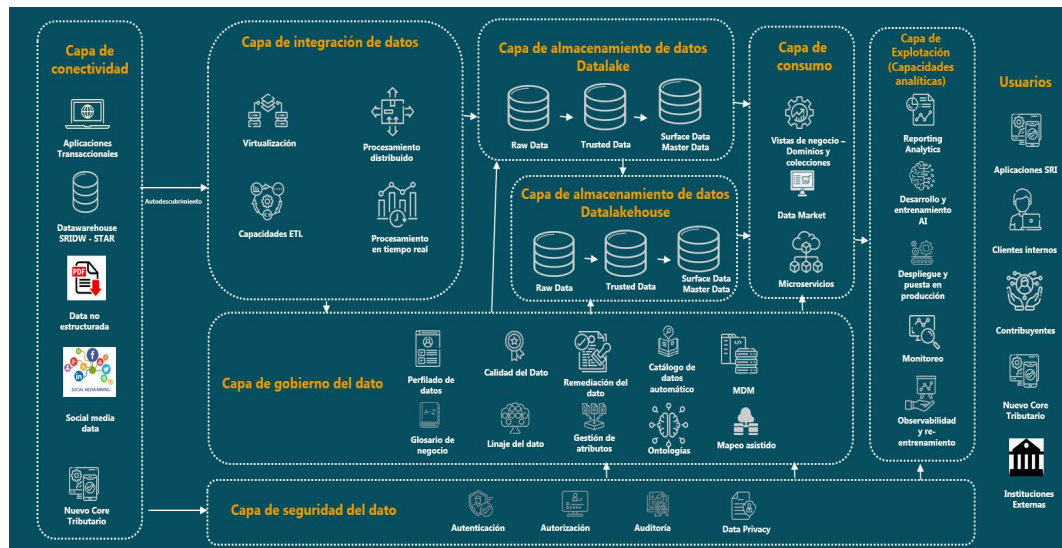
La plataforma deberá permitir a los analistas de negocios puedan usar funciones de arrastrar y soltar para acceder y preparar datos, crear modelos e implementar modelos en producción en un entorno unificado con acceso a datos estructurados y no estructurados, incluido el análisis de texto. Los científicos y analistas de datos pueden usar múltiples algoritmos de aprendizaje automático listos para usar para implementar funciones conocidas y combinar modelos basados en código para proporcionar resultados y experimentos personalizados. Tener científicos de negocios, TI y datos trabajando en la misma plataforma mejora la colaboración y la productividad, lo que conduce a resultados comerciales más rápidos e impactantes.

Todos los componentes y elementos que conformen la plataforma podrían estar basados e implantados en tecnologías abiertas siempre y cuando cuenten con el soporte y niveles de servicio sobre la plataforma. Si la plataforma contuviera algún componente basado en

tecnología propietaria, el adjudicatario deberá proporcionar las licencias presentadas que aseguren el correcto funcionamiento de la plataforma.

El siguiente gráfico muestra la arquitectura referencial para la solución planteada.

Gráfico 1: Esquema General Plataforma tipo Data Fabric



Fuente: Programa IN TAX- Servicio de Rentas Internas

Fecha: 03 de abril de 2023

A continuación se describen cada uno de los componentes de la arquitectura referencial esperada.

Capa de conectividad:

La Plataforma facilitará la integración de información proveniente de diversas fuentes asegurando su homogeneidad, calidad y consistencia a lo largo de su ciclo de vida, esto se hará a partir de la incorporación de herramientas tecnológicas y las mejores prácticas de gestión de datos que permitirán contar con información para mejorar los procesos de la entidad y acrecentar la explotación de información para fines de inteligencia de negocio, analítica y toma de decisiones.

Un Data Fabric es esencialmente una capa operativa de datos que no sólo reúne todos los datos, sino que los transforma y los procesa mediante Machine Learning para descubrir patrones y conocimientos. Sin un Data Fabric, todo esto tiene que suceder en cada aplicación individual, lo cual no es una solución sostenible.

Deberá estar integrada también con otras fuentes disponibles en la institución y deberá ser capaz de integrarse con estas fuentes/información sin requerir altos esfuerzos de implementación, acelerando la implantación y despliegue de Data Fabric.

Además, la plataforma deberá poder gestionar y analizar datos con diferentes formatos, desde datos no estructurados (fotos, videos, textos, pdf, etc.) a datos altamente estructurados como tablas SQL o ficheros CSV. La plataforma podrá extraer información tanto de ficheros planos (por ejemplo, ficheros de texto) u ofimáticos (por ejemplo, fichero con formato Excel) como de sistemas disponibles como bases de datos (ya sean on-premise o en cualquier nube pública).

La plataforma deberá proporcionar la posibilidad de permitir el descubrimiento automático de los datos, infiriendo esquemas (si los hubiere) e identificando los tipos de los datos analizados y relacionándolos con los ya existentes en la plataforma de manera automática. Adicionalmente, la plataforma no estará limitada a un tamaño máximo ni de fuentes de información ni de volumen de datos permitiendo escalar la solución lo que se desee.

Capa de integración:

La plataforma deberá ser lo menos intrusiva posible. Por un lado la plataforma deberá ofrecer la posibilidad de copiar la información a una nueva fuente de datos de una manera totalmente automática con el modelo de gobierno (simplemente eligiendo la cadencia y el origen y destino) en tiempo real o en modo batch siendo la plataforma autónoma en el despliegue de los mecanismos necesarios para esta sincronización (por ejemplo despliegue de una solución CDC automáticamente).

Por otro lado, la plataforma deberá permitir la virtualización del dato para acceder al mismo directamente sobre la fuente origen sin copiarlo, también deberá ser capaz de

conectarse a múltiples orígenes de datos y permitir el acceso a los mismos de una manera homogénea, independiente del gestor de datos al que se conecte y ofreciendo una capa de acceso común y unificado a sus usuarios. Estas fuentes/información pueden ser tecnologías generales (bases de datos relacionales, no-relacionales, ficheros, etc.)

Deberá proporcionar la posibilidad de virtualizar los datos obtenidos de los distintos orígenes de datos y, en el caso que se considere oportuno para el posterior análisis, la réplica de los datos dentro de la propia plataforma. Tanto la capacidad de replicación en tiempo real en un datastore propio como el acceso a la información a través de un virtualizador, son capacidades esenciales que permitirán la adaptación a la problemática oportuna dependiendo del caso de uso. Ambas metodologías de acceso al dato deben estar disponibles tanto sobre el modelo físico como sobre el modelo lógico de los datos, para que ésta sea más sencilla y rápida.

La plataforma deberá ser capaz de sincronizar el catálogo de datos auto descubierto de manera automática junto con los dominios y las vistas de negocio (definidas posteriormente), exponerlo a los usuarios y aplicar las políticas de acceso que apliquen a cada usuario y dato. Todo esto gestionado a través del modelo de gobierno para evitar inconsistencias y duplicación de configuraciones.

Adicionalmente la plataforma deberá contar con la capacidad de replicación de datos de manera automática, es decir, sin ningún tipo de desarrollo o diseño de flujos de ingesta por parte de los usuarios. Simplemente indicando a través de atributos en la herramienta de gobierno la necesidad de replicación, la cadencia, niveles de consistencia, el origen y el destino, la propia plataforma deberá ser capaz de lanzar internamente los procesos de sincronización y copiado de datos entre el origen y destino con la configuración seleccionada. Tanto el acceso a través del virtualizador como la sincronización de la copia de datos deberá ser escalable desde volúmenes pequeños a volúmenes altos de información o combinación de ambas.

La ingesta de datos deberá ser capaz de incluir información estructurada, semiestructurada y no estructurada desde diferentes fuentes tanto de tipo batch como real time y con latencias de entrega de resultados configurables por el usuario que van desde los milisegundos a horas o días.

La plataforma deberá tener capacidad de virtualización, auto descubrir y sincronizar datos procedentes de las fuentes de dato origen. Esto debe ser un proceso totalmente automatizado y la herramienta deberá ser capaz de leer todos los metadatos disponibles para generar un diccionario técnico de datos que aúne todos los metadatos procedentes de las múltiples fuentes de información disponibles. La plataforma deberá soportar múltiples formatos y conectores para diversas bases de datos tanto cloud como on-premise. Además, se incluirá en el diccionario no sólo toda la información de los esquemas de datos sino también toda información adicional que requiera la institución.

Asimismo, la plataforma deberá contar un descubrimiento enriquecido de datos basado en algoritmos de reconocimiento de patrones capaz de identificar criterios y datos personales.

En este sentido la plataforma deberá ser capaz de representar esta información autocontenida en el catálogo de datos y ofrecer la posibilidad al usuario de definir relaciones a nivel de modelo físico de datos entre datos procedentes de la misma o de diferentes bases de datos (tablas o columnas). Esto debe ser recogido por el virtualizador o la capa de acceso al dato para resolver queries que relacionen ambas tablas, simplificando así estas tareas a los usuarios finales.

Capa de gobierno:

La solución tipo Data Fabric incorporará los requisitos tanto de gestión como de tratamiento avanzado del dato, a través de un gobierno del dato embebido, para no sólo poder conseguir una democratización del dato controlada y granular sino también para hacer eficiente su explotación con políticas de calidad, seguridad y gestión de los datos maestros.

La plataforma de gobierno del dato deberá ser una única plataforma software end-to-end que optimice y mejore la gestión del ciclo de vida, basada en contenedores que virtualiza y gobierna la información, disponibilizándola en vistas de negocio para su uso en analítica y algoritmia tanto informacional como operacional, permitiendo de esta manera un gobierno del dato accionable.

La plataforma deberá implementar el paradigma Data Fabric proporcionando utilidades para cada una de los trabajos o etapas relacionadas con el tratamiento de los datos. Estas utilidades deberán incluir procedimientos de optimización y altos grados de automatización que minimicen al máximo la intervención humana para cada uno de los procesos.

La plataforma deberá gestionar los datos de manera holística, es decir deberá tener tanto capacidades de gobierno del dato como de su exposición. Esto quiere decir que la plataforma pueda disponibilizar los datos para las necesidades de la institución los requisitos normativos o de uso que puedan tener las diferentes áreas, a la vez que la propia plataforma, fruto de su actividad, posibilite de manera directa la vuelta de los datos a los sistemas operacionales.

La plataforma deberá proporcionar a las distintas áreas la posibilidad de cumplimentar sus necesidades de gobierno del dato, seguridad, linaje, reglas de calidad, gestión de datos maestros y data Marketplace. A su vez, la plataforma, fruto de la actividad que realiza, deberá dar la posibilidad de proporcionar datos a los sistemas operacionales y analíticos en los casos que sea necesario.

La plataforma deberá proporcionar utilidades para la gestión de la calidad de los datos en el que se permita descubrir valores nulos, valores fuera de rango (outliers), inconsistencias, esta utilidad permitirá crear reglas de control de calidad de datos que podrán ser ejecutadas automáticamente en el momento que se introduzcan los datos dentro de la plataforma, así como llevará un log de auditoría de las reglas de calidad ejecutadas y resultados obtenidos. Como requisito adicional el diseño de estas reglas de calidad debe proporcionar una interfaz gráfica para su diseño y no únicamente programática.

Para que la plataforma pueda ser utilizada por los usuarios de la institución, deberá facilitar la traducción de campos de datos técnicos en términos que puedan ser entendidos y utilizados por el usuario final. Además de la definición estándar de estos términos, la plataforma ofrecerá la posibilidad de generar distintos tipos de información relativos a la institución que sean completamente configurables y adaptadas al contexto

de su uso. Por ejemplo, la herramienta debe ser capaz de crear KPIs con significado para la institución y configurarlos con los campos necesarios para su definición. Toda esta información de negocio debe ser ordenada dentro de la herramienta por jerarquías de varios niveles, también completamente configurables, y controladas por accesos y roles concretos permitiendo un perfilado de grano fino en las acciones que pueda tomar cada usuario.

La plataforma deberá contar con un buscador inteligente que permita buscar términos, conceptos y datos en cualquiera de las ubicaciones y sistemas gestionados, bases de datos, ficheros, metadatos, glosario de negocio, ontologías, atributos, etc.

La plataforma deberá contar con un lugar donde se publiquen los conjuntos de datos obtenidos y que facilite la búsqueda y la comprensión de los datos disponibles para las diferentes personas usuarias de la plataforma. El acceso a estos datos deberá tener diferentes interfaces y formatos, incluyendo API REST y conectores ODBC/JDBC.

La plataforma deberá proporcionar utilidades para la gestión de datos maestros. Herramientas para gestionar la duplicación de información, el editado y la corrección de datos maestros, golden records o el mantenimiento de información de referencia para el resto de sistemas. Además, estas herramientas deberán ser totalmente configurables según el contexto y su uso debe ser orientado al usuario no técnico. La herramienta deberá poder generar páginas web de la gestión de maestros con las herramientas mencionadas anteriormente. Esto quiere decir que la información cruzada de dos bases de datos que contienen información duplicada, la corrección deberá suceder a través de páginas web sencillas exponiendo los esquemas y registros detectados de forma simple y entendible por un usuario sin necesitar ningún tipo de conocimiento técnico en bases de datos o analítica. El producto de datos maestros deberá dar la posibilidad de generación y solución de un golden record. Un golden record es un registro de referencia único y preciso de una entidad dentro de un sistema de gestión de datos maestros (MDM). Este registro incluye toda la información relevante sobre la entidad, verificada y consolidada a partir de múltiples fuentes de datos. El golden record es considerado la versión más autorizada y confiable de la información, y se utiliza como fuente de verdad para tomar decisiones y llevar a cabo actividades críticas.

Capa de seguridad:

Como requisito general la plataforma deberá poder integrarse con los sistemas de autenticación de usuarios existentes pudiendo integrar los accesos con las mismas credenciales y grupos de seguridad disponibles (autorización basada en roles/usuarios).

La plataforma deberá implementar mecanismos de colaboración e intercambio de datos a través del concepto de relaciones contractuales entre el consumidor de datos y el proveedor de estos. A través de "Data Contracts" se podrán gestionar los acuerdos de compartición y uso del dato para conseguir un uso responsable e informado de los datos. Estos contratos entre consumidores y productores deberán ser acompañados de herramientas que posibiliten su gestión, compartición y auditorías de uso con dashboards de control y métricas, revocaciones de permisos, etc. En estos Data Contracts se podrán regular también no sólo el uso y propósito de los datos sino también incorporar regulación pertinente cuando aplique como por ejemplo la Ley de Protección de Datos Personales.

La plataforma permitirá tener flujos de trabajo y configurar la anonimización o la exposición/ocultación de los datos y asegurar el cumplimiento de las diferentes normativas de seguridad aplicables. Esto debe ser configurable a múltiples niveles: tablas, columnas o filas y aplicable directamente o a través de atributos en metadatos (seguridad ABAC) y por rol o perfil (RBAC). La seguridad ABAC debe estar sincronizada entre las políticas de seguridad definidas y los metadatos disponibles en el catálogo de datos. Además, estas políticas deben soportar la autorización por rol o usuario. Es decir, a un usuario analista puede aplicarse una política de acceso ABAC mientras que un usuario administrador puede tener otros permisos de acceso diferentes de tal manera que se consiga configurar a nivel individual cada acceso al dato como sea necesario.

Tanto los mecanismos de compartición como los de seguridad y su integración son vitales para compartir y correlacionar información y deberán ser capaces de publicar información relevante manteniendo siempre el grado de anonimización necesario para cumplir con las normativas vigentes.

En este sentido, la plataforma además de poder accionar estas reglas de acceso deberá ser capaz de aplicarlas en un contexto normativo como por ejemplo la Ley de Protección

de Datos Personales junto con el resto de requisitos. La plataforma deberá ayudar al cumplimiento completo de normativas estándar de tratamiento de datos ya que debe contener los mecanismos de seguridad necesarios para implementar cualquiera de los requisitos estándar como anonimización, control de acceso, auditoría, derecho al olvido, etc.

Capa de almacenamiento:

La plataforma deberá contar con la capacidad de almacenamiento de objetos escalable, robusta, segura y de integración sencilla con el resto de componentes de la plataforma. Dicho almacenamiento contará con vertientes de alto rendimiento y con posibilidad de configurar la frecuencia y concurrencia de acceso desde poca a mucha frecuencia y concurrencia.

La plataforma deberá considerar el almacenamiento de data estructurada, semi estructurada y no estructurada, utilizando para ello diferentes herramientas y componentes tecnológicos, estos incluyen como mínimo: Data warehouse, Data marts, Data lakes, Data hubs.

- Datawarehouse:

Corresponde a la base de datos donde se integra y depura la información proveniente de diversas fuentes para ser procesada y analizada desde múltiples perspectivas. Esta base de datos debe ser optimizada para consultas a gran velocidad.

El producto de Data Warehouse que el fabricante, partner y/o proveedor proponga como parte de la solución deberá ser escalable, flexible, orientado a temas, no volátil, totalmente integrable a los ambientes tecnológicos de datos, control y gestión de seguridad y tener capacidad para habitar y operar en la nube.

Cabe resaltar que la solución de Data Warehouse debe permitir:

- Manejar grandes volúmenes de datos, para múltiples estructuras de datos y formatos. Es importante en todo caso, realizar optimizaciones y gestión

- sobre las cargas que generan las consultas.
 - Soportar la ingesta continua de datos, aunque el cargue masivo o en lotes de información sea lo preferido. Igualmente, debe soportar la actualización continua de información y aun así mantener la optimización para la consulta de información.
 - Procesar y enlazar diferentes tipos de datos, no exclusivamente datos estructurados.
 - Realizar de forma automática la optimización y la gestión de carga de trabajo para consultas repetitivas.
 - Incluir capacidades de administración y gestión tales como actividades basadas en roles, consejeros (advisors), monitoreo de utilización y planeación de capacidad.
- Datamarts:

Los Data Marts son versiones especializadas de un Data Warehouse (enfocadas a un tema, departamento u objetivo específico), cuya estructura de datos contenidos simplifica su organización y acceso. Son subconjuntos de datos que tienen el objetivo de proveer información específica a la organización, para fines de inteligencia de negocio, analítica y toma de decisiones.

Las herramientas de Data Marts que el fabricante, partner y/o proveedor proponga como parte de la solución tecnológica del DataFabric deberán permitir la creación de Data Marts dependientes, independientes e híbridos; integrables a la solución de Data Warehouse propuesta y a los componentes tecnológicos de inteligencia de negocio y analítica de la solución del DataFabric.

Adicionalmente, al ser los Data Marts especializaciones de un Data Warehouse, las características requeridas para las herramientas de Data Warehouse se extiende a las de Data Marts.

- DataLakes:

Son soluciones de gestión de datos no estructurados y semi estructurados, en su formato nativo, sin procesar, para posteriormente ser procesados y llevados a un almacén de datos o bien para su explotación directa como insumo de los procesos de inteligencia de negocio, analítica avanzada y ciencia de datos.

Los servicios de Data Lakes que el fabricante, partner y/o proveedor incluya como parte de la solución tecnológica del DataFabric deberá tener capacidades de flexibilidad, escalabilidad, manejo de metadatos, control y gestión de seguridad y tener capacidad para habitar y operar on premise o en la nube.

- **DataHubs:**

Las herramientas de DataHubs que el fabricante, partner y/o proveedor proponga como parte de la solución tecnológica del DataFabric deberán permitir almacenar datos de manera sencilla para facilitar la toma de decisiones. Y la diferencia clave respecto al datalake y el Datawarehouse está en que el DataHubs se centra en categorizar los datos y hacerlos disponibles para una recuperación fácil y rápida.

Existen adaptadores que permiten la ingesta de datos en el DataHubs, donde se armonizan, limpian y gestionan adecuadamente, y se ponen a disposición de toda la empresa, incluidos los sistemas de origen. El DataHubs pretende dar más autonomía a los usuarios de negocio para trabajar con los datos (autoservicio de datos).

Los servicios de DataHubs que el fabricante, partner y/o proveedor incluya como parte de la solución tecnológica del DataFabric deberá tener capacidades de flexibilidad, escalabilidad, manejo de metadatos, control y gestión de seguridad y tener capacidad para habitar y operar on premise o en la nube

Capa de consumo:

La plataforma deberá proporcionar herramientas para el uso y visualización amigable de ontologías (Definición formal de tipos de datos, propiedades y relaciones entre las entidades que existen en un determinado dominio) y relaciones semánticas entre los

datos. Estas ontologías deberán poder visualizarse como un diccionario de información o mediante gráficos soportando formatos estándar de la industria. Estas vistas de negocio no requerirán una persistencia obligatoria del dato sino que nos permitirán representar el dato sin moverlo. La plataforma tipo Data Fabric deberá ser capaz de acelerar los procesos de implantación con su uso/importación.

Además, estas vistas deberán soportar todas las tipologías de políticas de seguridad disponibles en el acceso. El uso de ontologías/modelos de datos lógicos con el formato OWL/RDF (estándar del consorcio W3C) se considera importante ya que permitirá definir no sólo los propios accesos al dato y su exposición sino que también establecer y visualizar las relaciones entre los datos.

La plataforma deberá ser capaz de generar y sincronizar las vistas definidas en las ontologías y mapearlas al dominio de datos físico con el fin de resolver la relación entre las vistas de negocio y el modelo físico de datos auto descubierto. para posteriormente soportar un proceso de publicación donde se controlen las vistas e información expuesta.

La plataforma también deberá soportar la publicación del catálogo de datos disponible para todos los usuarios y de acuerdo con las políticas de seguridad definidas. Al contar con ontologías y estándares existentes, los catálogos de datos y la capa de explotación del dato se permite exponer un elevado porcentaje de términos desde el inicio, acelerando la implantación y adopción de dichas ontologías/estándares desde el inicio, sin requerir complejas transformaciones o adaptaciones a dichos estándares en fases de implantación. Una vez expuestas éstas serán accesibles y tanto procesos analíticos, consultas, modelos de machine learning o reportes podrán hacer uso de ellas con un lenguaje de consulta estándar como SQL.

La plataforma deberá ofrecer además una capa de gestión de cómo se realiza la exposición del dato al usuario final. Será necesaria la capacidad de poder definir dominios del dato donde se puedan agrupar y visualizar gráficamente, tanto vistas de negocio como tablas técnicas que contengan la información disponible de esos dominios. Además de poder definir roles y permisos de edición sobre los mismos para controlar la capa de consumo del dato por las personas/perfiles adecuados.

La plataforma deberá habilitar diferentes posibilidades de acceso a los datos desde otros sistemas habilitando un acceso de DaaS – Data As a Service (API REST, ODBC/JDBC, etc.) cumpliendo con los estándares de seguridad y acceso definidos. Esto aplica tanto a las tablas físicas del catálogo como a todas las colecciones y vistas creadas por la institución y que se quieran exponer para su consumo.

La plataforma además deberá poder recoger y representar el linaje de los datos y asociarlo con los procesos que usan esos datos ya sean de lectura o escritura. Este linaje se asociará a los datos utilizados de tal forma que se pueda ver de origen a fin cuales han sido sus transformaciones y cambios. Además, este linaje se deberá poder relacionar con el resto de información disponible como colecciones publicadas, data stores con la información auto descubierta, términos de negocio, etc. Y toda esta información se deberá poder representar en forma de grafo para explorar las relaciones entre datos o metadatos.

La plataforma deberá ser capaz de poder diseñar y generar cuadros de mandos y, adicionalmente, deberá posibilitar el uso de los datos gestionados desde la plataforma (ya sean fuentes internas o externas) por herramientas externas (por ejemplo, Power BI o Qlik). La herramienta permitirá la customización de cualquier cuadro de mando, ya sea técnico o de alto nivel, y además deberá contar con plantillas, templates o cuadros de mando ya disponibles, por ejemplo, la monitorización de la calidad del dato o aspectos de gobierno del dato.

Todos los requisitos mencionados anteriormente deben ser capaces de cubrirse desde casos con poco volumen a grandes volúmenes de información.

Capa de explotación:

El Data Fabric deberá estar preparado para satisfacer de forma integrada y modular las futuras necesidades de AI y ML automáticamente y en niveles sostenibles. El Machine Learning puede proporcionar los datos y los conocimientos de forma proactiva, lo cual ayudará a los responsables de la toma de decisiones a tener mejores conocimientos e información más oportuna. Los resultados deseables deberán descubrir hechos ocultos de los datos sin que se busquen o soliciten específicamente.

La plataforma deberá trabajar con interfaces gráficas de usuario y diagramación de procesos en flujo, con lo que se logre un ambiente de trabajo visual, amigable, intuitivo e interactivo que potencie el máximo retorno al facilitar la implementación de proyectos de minería de datos en menor tiempo y que priorice el flujo de trabajo sin el uso o conocimiento de código fuente (o con la mínima cantidad de éste), enfocando la herramienta para el uso de usuarios funcionales.

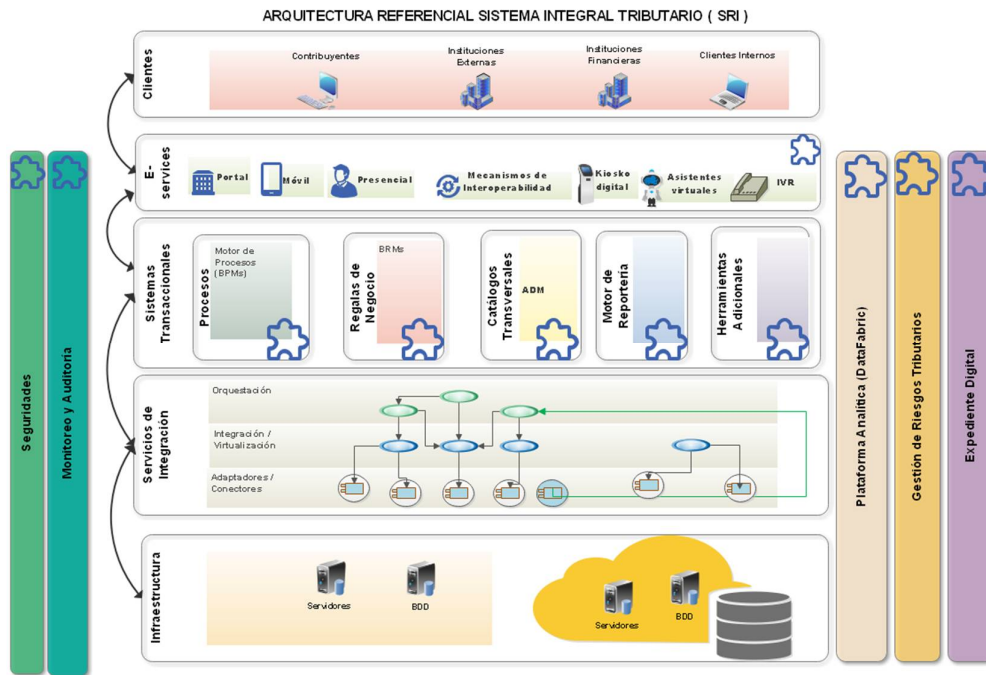
Así mismo, permitirá automatizar procesos e integración con otras herramientas, lo que inicia desde el desarrollo de rutinas en la herramienta, capacidad de integración de múltiples fuentes de datos y diversos formatos; y calendarización de la ejecución de los modelos desarrollados de acuerdo con la periodicidad requerida en cada proyecto, a fin de emitir alertas tempranas para minimizar el riesgo tributario.

Finalmente todos aquellos módulos de la plataforma, que permitan exportar, importar, consultar, etc., deberán disponer API de servicios web. Los cuales deberán estar disponibles para ser consumidos bajo parámetros de control configurados en la herramienta sin necesidad de intervención de código para poder exponer la funcionalidad. Esto con el objetivo de que el acceso al dato sea abierto permitiendo conexión de otras aplicaciones y que no cuente con restricciones para acceder y explotar el dato tanto para usuarios internos, aplicaciones transaccionales existentes, contribuyentes, entidades externas y nuevas aplicaciones que se lleguen a implementar.

Actualmente el Servicio de Rentas Internas tiene previsto implementar un nuevo Sistema Integral de Gestión Tributaria, bajo un esquema que permita fortalecer su interoperabilidad, escalabilidad, sostenibilidad y trazabilidad de sus componentes funcionales y tecnológicos. En dicho diagrama se contempla la capa de Plataforma DataFabric ², la cual es el objeto de estudio en este documento:

² Corresponde al componente Plataforma Analítica dentro de la arquitectura referencial.

Gráfico 2: Arquitectura referencial del Sistema Integrado de Gestión Tributaria



Fuente: Programa IN TAX- Servicio de Rentas Internas

Fecha: 10 de marzo de 2023

Nota: El gráfico de la arquitectura referencial del Sistema Integral de Gestión Tributaria se lo detalla con fines informativos para considerar y velar por la interoperabilidad de la capa o solución de la plataforma DataFabric con los demás componentes que resulten de la implementación de las otras capas.

3.1.1 Requisitos Generales

Por favor indique el cumplimiento de la solución ofrecida con respecto al requerimiento definido en los numerales correspondientes, considerando cada alternativa de implementación (ON PREMISE o CLOUD).

El cumplimiento se debe responder teniendo en cuenta la siguiente clasificación:

T – Total Se tiene la capacidad disponible al 100%

P – Parcial Especificar que funcionalidad no cumple en la columna de justificación.

N – No disponible La capacidad no está disponible (0%), requiere el 100% de adaptación.

Este nivel de cumplimiento y justificación de la funcionalidad se debe reflejar en el diligenciamiento del cuadro de cumplimiento de requerimientos (Anexo 1).

Ref.	Concepto	Descripción
RG-1	Modelos (IaaS, PaaS, SaaS y On-premise)	La plataforma completa podrá instalarse en nubes privadas y nubes públicas, tanto como PaaS, IaaS o SaaS. También deberá poder instalarse completamente on-premise. Así mismo deberá permitir una combinación híbrida en la nube y en local (on-premise). No se permiten componentes alojados en la nube indicada que enlace a módulos fuera de la nube
RG-2	Solución tipo Data Fabric completa e integrada	La plataforma debe ser capaz de cubrir todas las necesidades de Data Fabric sin necesidad de integrar múltiples proveedores o productos diferentes. Todas las funcionalidades deben estar integradas de forma nativa y directa en una única solución
RG-3	Solución tipo Data Fabric existente	La Solución Software Integral propuesta deberá ser un producto de mercado ya existente y consolidado, que no requiera de desarrollos específicos para su puesta en marcha o de integraciones entre diferentes productos. Contando con casos de éxito en implementaciones que cumplan con las premisas indicadas en al menos los últimos 3 años.
RG-4	Escalabilidad	La plataforma debe utilizar Kubernetes y todos los módulos desplegarse en contenedores para ser escalable de manera distribuida tanto en dimensiones como en volumen de accesibilidad, almacenamiento y procesamiento de datos, ofreciendo una capacidad de respuesta eficiente ante un elevado volumen de datos, servicios y usuarios. Tomando como base la característica de los componentes (Kubernetes y Docker) el no requerir de productos licenciados para poder ejecutar ambientes contenerizados
RG-5	Licenciamiento completo e incluido	Las licencias deben incluir número ilimitado de usuarios sin restricciones, no sujeta a coste adicional por usuarios, a los distintos módulos y pudiendo acceder a toda la funcionalidad disponible del tipo Data Fabric
RG-6	Escalabilidad licenciamiento	La plataforma debe ser modular y lo suficientemente flexible para escalar/desescalar con el tiempo

Ref.	Concepto	Descripción
RG-7	Robustez	La plataforma debe ser Fault Tolerant en todos sus elementos y estar construida sobre kubernetes y contenedores con todos los elementos replicados y con arquitecturas y protocolos de actuación que garanticen la gestión eficiente de fallos que puedan afectar a la disponibilidad de la plataforma, de tal forma que se garantice un alto nivel de disponibilidad de esta en cualquier circunstancia.
RG-8	Accesibilidad Web	Todos aquellos módulos y herramientas que componen la plataforma deban ser accesibles vía web. Aun cuando la instalación sea ejecutada de forma on-premise
RG-9	Acceso a la plataforma	La plataforma deberá disponer de diferentes perfiles y roles que garanticen la responsabilidad y visibilidad al acceso a las diferentes funcionalidades y datos.
RG-10	Seguridad	La plataforma deberá contar con las herramientas necesarias que provean una capa de seguridad, con gestión de la autenticación e integración con sistemas externos, y reglas de acceso a los datos, ya sean tanto a nivel confidencialidad de datos, encriptación de datos en origen, destino y tránsito. como en accesos de usuarios y procesos.
RG-11	Garantizar la continuidad de la plataforma	Con el objetivo de que la nueva plataforma tenga continuidad en el tiempo, y siga siendo la misma, independientemente de la alternancia de distintas empresas adjudicatarias para la ejecución de los nuevos expedientes, la plataforma deberá ser por un lado fácilmente escalable, puesto que seguirá creciendo más allá de la vida del contrato y utilizando los frameworks de desarrollo que son estándar para el desarrollo de proyectos alrededor de la gestión del dato (Kubernetes, spark, etc.)
RG-12	Monitorización	Es necesario que la plataforma elegida disponga de herramientas internas integradas, para la monitorización del uso de los recursos facilitando al usuario la configuración de métricas, indicadores, umbrales y alertas, con una visión lo más integrada posible de la solución completa. Esta monitorización debe aportar visibilidad a nivel técnico y funcional de la solución
RG-13	Cuadro de Mandos de monitorización técnica y funcional	La plataforma deberá tener la capacidad de incluir y manejar cuadros de mandos de monitorización técnica y funcional con indicadores de procesamiento, como pueden ser velocidad, volumen, rendimiento, coste, etc.
RG-14	Log centralizado	La plataforma deberá contar con herramientas que recolecten logs para su control, análisis y visualización.

Ref.	Concepto	Descripción
RG-15	Disponibilidad	La plataforma deberá asegurar la disponibilidad 24x7
RG-16	Auditoria y linaje	La plataforma deberá mantener trazabilidad de quién accede, cuándo y cómo se tratan los datos y servicios disponibles, de quien se ha realizado las consultas, así como datos sobre, IP's, nombres de equipos, usuarios, estructuras accedidas, ente otros. La plataforma debe poder garantizar que el dato está gobernado desde el inicio, ya sea virtualización o primera ingesta de información, haya trazabilidad y se pueda tener control del linaje
RG-17	Sincronización de permisos de seguridad	La plataforma garantizará la sincronización de permisos/exposición del dato de acuerdo con la configuración de gobierno del dato.
RG-18	Gobierno del dato accionable y embebido	El gobierno del dato debe ser una herramienta de control del dato sobre el que se puede programar y realizar procesos y analíticas y no algo solamente declarativo o informativo.
RG-19	Gestión y Gobierno de Datos federados	La plataforma debe contener mecanismos para permitir la gestión del dato por diferentes unidades y deberá permitirles compartir/consumir la información entre ellas de manera autónoma cuando sea necesario de una manera sencilla
RG-20	MLOps y Data Ops	La plataforma tendrá la capacidad de implementar las herramientas y procesos necesarios para metodología de DataOps y MLOps. Desde la adquisición del dato hasta la monitorización y el reentrenamiento de modelos cuando estos se degraden o pierdan efectividad
RG-21	Pago por uso predecible	La plataforma debe permitir establecer un costo estable durante el tiempo de duración del contrato con parámetros bien definidos que ayude a la entidad a saber un límite de uso del sistema sin incurrir en costos variables no dimensionados, además no debe contar con costos ocultos que afecten la facturación con micro costos.
RG-22	Orquestación multinube	La plataforma debe tener la capacidad de descubrir, virtualizar y/o ingestar fuentes de datos multinube, bien sea transportando o no los datos dependiendo de las necesidades de la entidad y los requisitos de seguridad de dicha área
RG-23	Replicación y resiliencia	La plataforma debe permitir ser replicada a otra zona de disponibilidad e incluso a otra nube

Ref.	Concepto	Descripción
RG-24	Gestión del ciclo del dato	La plataforma debe comportarse como un DataFabric, que permite automatizar y orquestar las fuentes de datos, brindando facilidad de acceso al proceso de explotación del dato como activo de la organización. Las funciones deben ejecutarse de igual forma tanto en nube pública, privada, híbrida o en on-premise
RG-25	Soporte en español	El soporte debe ser integralmente en español tanto a nivel de tickets y notificaciones como a nivel de soporte telefónico o remoto

3.1.2 Virtualización de Datos

Ref.	Concepto	Descripción
V-1	Virtualización de Datos	La plataforma deberá incluir funcionalidad de virtualización de datos. (Se entiende por Virtualización de Datos al consumo de datos cuando son requeridos por los sistemas sin necesidad de replicarlos en otros repositorios.)
V-2	Múltiple datastores	El virtualizador soportará el acceso a diferentes tipos de fuentes de datos
V-3	Integración con dominios	La plataforma/El virtualizador deberá sincronizarse con los dominios definidos en la herramienta de gobierno automáticamente y sin necesidad de intervención manual
V-4	Exposición de tablas físicas	La plataforma deberá poder acceder a las tablas físicas auto descubiertas en los orígenes de datos y presentar los datos, ejecutar filtros sobre ellos y realizar operaciones
V-5	Exposición de modelo lógico	La plataforma debe ofrecer el catálogo a sus usuarios formado por los datos físicos y por las business views del modelo lógico
V-6	Escalado	La plataforma/El virtualizador deberá poder gestionar volúmenes altos de información en sus queries y observar el plan de ejecución en la misma plataforma
V-7	Soporte sistemas de ficheros	El virtualizador podrá lanzar consultas OLAP sobre sistemas distribuidos de ficheros como HDFS y resolver las queries de manera distribuida
V-8	Seguridad de acceso	La plataforma debe forzar el acceso al dato de acuerdo con las políticas de seguridad definidas tanto en el modelo físico de datos como en el modelo lógico definido

Ref.	Concepto	Descripción
V-9	Copiado de datos batch	La plataforma debe ser capaz de gestionar automáticamente el copiado y sincronización de información entre una fuente origen y un destino seleccionado sin un desarrollo de flujos y de manera automática definiendo simplemente un atributo en la herramienta de gobierno
V-10	Copiado de datos real-time	La plataforma debe ser capaz de sincronizar una fuente origen con una fuente destino en tiempo real y de manera automática definiendo simplemente un atributo en la herramienta de gobierno
V-11	Exposición de datos JDBC/ODBC	Exposición de datos tanto al modelo lógico como al modelo físico de datos a través de una interfaz JDBC/ODBC
V-12	Exposición de datos por HTTP REST	Exposición de datos tanto al modelo lógico como al modelo físico de datos a través de una interfaz HTTP REST, debe permitir realizar la configuración y exponer el componente como un producto del catálogo de componentes para ser solicitado por los consumidores de los datos
V-13	Control de la concurrencia de queries distribuidas	Posibilidad de seleccionar la concurrencia utilizada en el virtualizador para paralelizar en mayor o menor grado las queries al origen
V-14	Queries de fusión de datos entre BBDDs	El virtualizador soportará la mezcla y operaciones de cruce de información entre múltiples fuentes de datos
V-15	Cache	El virtualizador podrá persistir datos a modo cache para optimizar las queries y su performance
V-16	Optimización consultas	El virtualizador podrá optimizar las queries en función de la tecnología de la fuente de datos haciendo "push-down" o trayéndose los datos a memoria para su procesado
V-17	Datos comprensibles	La plataforma deberá permitir realizar consultas entre las fuentes internas y virtualizadas unificando área de negocio y áreas técnicas, las consultas pueden ser usadas con términos de negocio con resultados entendibles por todas las áreas de la institución

3.1.3 Datos Federados

Ref.	Concepto	Descripción
DF-1	Propiedad de fuentes de datos distribuida	Las fuentes de información pertenecen a diferentes propietarios con una visión parcial y éstos pueden ser autónomos gestionando el ciclo de vida de estos. Entendiéndose ciclo de vida como aquel proceso integral por el cual se

Ref.	Concepto	Descripción
		gestiona los datos de la organización, el cual debe incluir el proceso de integración, gobierno y explotación del dato.
DF-2	Propiedad de dominios distribuida	Las fuentes de datos físicos son compartidas pero la propiedad del dato es expuesta en diferentes dominios, conceptos y atributos por lo que la plataforma debe ser capaz de garantizar la gestión del ciclo de vida del dato a nivel lógico y físico.
DF-3	Mecanismos de compartición	Los propietarios de los datos son capaces de publicar y consumir datos de una manera sencilla a través de mecanismos de compartición de la plataforma
DF-4	Protección propiedad del dato	La plataforma debe ser capaz de compartir información sin obligar a copiarla a un espacio propio del consumidor y se hace en tiempo real con las capacidades de virtualización
DF-5	Contratos de uso de datos	La compartición de datos puede ir acompañada de una previa firma de un contrato legal (data contract) donde se estipulan las condiciones de uso y a las que el consumidor se debe subscribir. Una funcionalidad clave a la hora de compartir y maximizar la exposición de datos explotables es la presencia de un data Marketplace. En este tanto convergen consumidores y productores de información donde pueden controlar y aceptar un acuerdo de uso entre ambas partes. El acuerdo incluye información de los medios de acceso, las fechas, los niveles de seguridad y privacidad de los datos y el motivo de acceso a la colección
DF-6	Reutilización de modelos lógicos de datos	La plataforma provee los mecanismos para reutilizar modelos lógicos de datos existentes entre diferentes propietarios locales de datos. Por ejemplo: una unidad realiza un modelo lógico de datos y éste se quiere reutilizar en otra unidad para que todos los activos desarrollados puedan depender del mismo modelo de datos. La compartición, importación y exportación entre propietarios debe ser gráfica y sencilla.
DF-7	Reutilización de activos de analítica	La plataforma permite compartir activos de analítica, reportes, etc. para su uso en diferentes contextos que dependan del mismo modelo de datos lógico

3.1.4 Capacidades de Almacenamiento

Ref.	Concepto	Descripción
AL-1	Datawarehouse, Data Lake y/o Data Lakehouse	La plataforma deberá contar con la capacidad de almacenamiento de objetos escalable, robusta, segura y de integración sencilla con el resto de componentes de la plataforma. Dicho almacenamiento contará con vertientes de alto rendimiento y con posibilidad de configurar la frecuencia y concurrencia de acceso desde poca a mucha frecuencia y concurrencia.
AL-2	Modelos de datos dimensionales	La plataforma deberá contar con la capacidad de almacenamiento de datos rápido y completamente administrado a escala de petabytes que permita el procesamiento masivo en paralelo para la ejecución rápida de consultas.
AL-3	Almacenamiento Estructurado	La plataforma deberá contar con la capacidad de base de datos relacional administrada y escalable que permita seleccionar entre diferentes proveedores de bases de datos.
AL-4	Almacenamiento no Estructurado	La plataforma deberá contar con la capacidad de base de datos no estructurados distribuida y que, de soporte a varios modelos de forma nativa, como: clave-valor, documentos, grafos y columnares.
AL-5	Almacenamiento Cache	La plataforma deberá contar con la capacidad de almacenamiento en caché distribuido, y administrado, basado en memoria que proporciona un almacén de alto rendimiento evitando el uso de sistemas de bases de datos basados en acceso a disco.
AL-6	Consulta	La plataforma deberá contar con la capacidad de consultar de forma interactiva datos almacenados en un sistema de almacenamiento de objetos por medio de soluciones estándar
AL-7	Búsqueda Texto Libre	La plataforma deberá contar con la capacidad de análisis de logs, búsqueda e indexación de texto completo de forma administrada.
AL-8	Escalabilidad horizontal	La plataforma estará preparada para dar respuesta en situaciones de alta demanda y uso crítico, permitiendo añadir nuevos servidores para atender al incremento de la demanda en el almacenamiento, sin impacto en el rendimiento.
AL-9	Tipología de Almacenamiento	La plataforma deberá proporcionar diferentes tipos o clases de almacenamiento para gestionar diferentes tipos de datos (estructurados, no estructurados y semi estructurados) y patrones de acceso
AL-10	Consulta de Datos	La plataforma deberá proporcionar funcionalidades que permitan realizar consultas mediante un interfaz tipo SQL.
AL-11	Procesamiento Flexible	La plataforma deberá proporcionar integración nativa con servicios de procesamiento de datos. Por ejemplo: procesamiento de imagen, ejecución algoritmos de machine learning, soluciones de data Discovery, etc.

Ref.	Concepto	Descripción
AL-12	Volumen y Velocidad	La plataforma deberá ser capaz de soportar ingestas de volúmenes de datos de diferentes fuentes y a diferentes velocidades. Ejemplo: sensores IOT, logs de servidores, OLTP, etc.
AL-13	Virtualización de datos para analítica	La plataforma deberá permitir trabajar con las fuentes de datos desde su origen sin necesidad que trasladar la totalidad de la fuente a la plataforma para su para ejecutar procesamiento sin causar deterioro de rendimiento del modelo o la función analítica a realizar
AL-14	Calidad de datos	La plataforma debe incluir en el proceso de adquisición de datos la aplicación de reglas de calidad con el fin de facilitar y por ende reducir tiempos al momento de iniciar a trabajar con ellos en el proceso analítico
AL-15	Almacenamiento operacional analítico	La plataforma deberá permitir un almacenamiento intermedio para transformar y limpiar los datos entre la capa operacional y la informacional y de esta forma facilite la comunicación entre sistemas operacionales e informacionales
AL-16	Capas de almacenamiento RAW universal y analítica	La plataforma deberá de tener al menos 3 capas de datos diferenciadas, raw (para el dato en bruto), universal (para el modelado del dato) y analítica (para realizar las actividades analíticas) que aseguren un correcto almacenamiento de los datos y la preparación de estos para su explotación
AL-17	Seguridad	La plataforma debe contar, para cada almacén de datos, con segmentación de red y permitir realizar cifrado en reposo
AL-18	Monitoreo	La plataforma debe contar para cada almacén de datos de un sistema de logs y monitoreo
AL-19	Seguridad	La plataforma debe dotar a cada almacén de datos con capacidades de resiliencia

3.1.5 Gobierno del Dato

Gobierno del Dato

Ref.	Concepto	Descripción
GD-1	Lectura de tipos de datos	La plataforma deberá permitir el acceso a cualquier tipo de datos, estructurados como no estructurados.
GD-2	Gestión de datos sensibles	La plataforma deberá tener la capacidad de gestionar diferentes datos sujetos a regulaciones y normativas tanto externas como internas, facilitando su identificación, acceso y análisis.

Ref.	Concepto	Descripción
GD-3	Exposición servicios web	Todos aquellos módulos de la plataforma, que permitan exportar, importar, consultar, etc., deberán disponer API de servicios web. Los cuales deberán estar disponibles para ser consumidos bajo parámetros de control configurados en la herramienta sin necesidad de intervención de código para poder exponer la funcionalidad. Esto con el objetivo de que el acceso al dato sea abierto permitiendo conexión de otras aplicaciones y que no cuente con restricciones para acceder y explotar el dato.
GD-4	Inferencia Automática	La plataforma deberá disponer de funcionalidades para inferir automáticamente datos y metadatos. Ejemplo: Tipos de datos como DNI, direcciones u otros datos sociodemográficos.
GD-5	Relación multidominio	La plataforma deberá permitir registrar o asociar datos, entidades y definiciones para los distintos dominios de negocio, conteniendo términos y estándares propios de cada dominio de negocio- Incluyendo jerarquía y relaciones entre dominios, conceptos y atributos
GD-6	Gestión de constantes	La plataforma dispone de la posibilidad de crear y gestionar constantes como variables dentro de los valores usados ya sea en el diccionario, glosario de negocio o cualquier elemento de gobierno. Esto permite la parametrización y adaptación de casuísticas diversas. Estas constantes pueden servir de plantillas para ser reutilizadas en procesos globales del gobierno del dato y evitan la manualidad y proporcionan disminución de esfuerzos en el proceso
GD-7	Gestión de Roles	La gestión de roles y permisos asociados permite la gestión de la autorización de funcionalidades dentro de la herramienta. Esto permite configurar quienes son los responsables de los procesos de aprobación, quién puede editar ciertos valores o quién puede acceder a qué información. El modelo de autorización debe ser customizable por rol y llegar a un grano fino tanto en funcionalidades como en acceso al dato. También debe poder integrarse con el Directorio Activo y herramientas de administración de roles.
GD-8	Registro de usuarios clave	La plataforma deberá permitir registrar los usuarios de la organización y segmentarlos a través de diversos grupos y perfiles, cada uno con sus políticas de acceso personalizadas
GD-9	Linaje y relaciones entre assets	La plataforma deberá permitir el linaje y relaciones visual de términos e indicadores de negocio. El linaje deberá ser tanto técnico (de qué tabla y columna provienen los datos) así como un linaje de negocio (que permita relacionar el glosario, ontologías y colecciones publicadas)

Ref.	Concepto	Descripción
GD-10	Servicios Colaborativos	La plataforma deberá prestar servicios colaborativos de flujos de trabajo y de aprobaciones. Estos flujos de aprobación podrán ser parametrizables a las necesidades de la organización (distintos niveles o etapas de aprobación) y podrán ser aplicado a diversos componentes del gobierno del dato. La plataforma deberá permitir configurar el flujo de trabajo de las actividades del Catálogo de datos (Agregar / Editar / Eliminar / Aprobaciones)
GD-11	Visualización de pasos en los flujos de trabajo	La plataforma deberá permitir mapear, actualizar y automatizar visualmente los pasos de los workflows definidos
GD-12	Agregación de Archivos	La plataforma deberá permitir agregar archivos relacionados con los activos, términos, definiciones, relaciones, etc. para una mejor comprensión (PDF, Visio, Word, etc.)
GD-13	Control de Versiones y Registros de Cambios	Permitir el control de versiones y registro de cambios para asegurar la trazabilidad de cambios dentro de las plataformas, así como visualizar gráficamente estos registros.

Data Catalogue y Gestión del Metadato

Ref.	Concepto	Descripción
GD-14	Conexión a BBDD	La plataforma deberá disponer de mecanismos de conexión a cualquier tipo de bases de datos, sean relacionales, OLAP o Big Data y aplicadas a un contexto sanitarios como ERPs, Historia Clínica Electrónica (HCE)
GD-15	Conexión a fuentes	La plataforma deberá permitir conectarse a cualquier tipo de fuente de datos: RDBMS, sistemas de ficheros, etc.
GD-16	Meta data	La plataforma deberá permitir gestionar y almacenar atributos, tipo de datos, valores permitidos, restricciones, sistema de origen, etc. - Fuente de datos, tablas, nombres de campo - Longitud del campo - Estructura de datos
GD-17	Definición de tipos de metadatos	La plataforma podrá definir distintos tipos de metadatos. Por ejemplo, nombre de columnas, llaves primarias o foráneas, tipo de campo, fecha de actualización, y otros atributos personalizados como data owner's o identificación de datos críticos.
GD-18	Gestión de metadatos/Atributos	La plataforma deberá poder gestionar nuevos atributos, crearlos, asignarlos, modificarlos, etc...
GD-19	Aprobación de los metadatos	La plataforma deberá disponer de herramientas de aprobación de cambios en el registro de los metadatos desde cualquier ámbito dentro de la organización

Ref.	Concepto	Descripción
GD-20	Repositorio centralizado	La plataforma deberá contar con un repositorio centralizado donde almacenará los metadatos y toda la información relacionada.
GD-21	Automatización de recolección de metadatos y autodescubrimiento de Data Stores	La plataforma deberá permitir auto descubrir los esquemas y metadatos técnicos de cualquier fuente de datos automáticamente, de sistemas de bases de datos (relacionales o no), almacenamiento de ficheros, distribuidos o no
GD-22	Difusión y distribución	La plataforma deberá contar con herramientas de difusión como la virtualización del dato y distribución de los metadatos publicados en el repositorio a otros componentes. Esto permitirá mantener estas definiciones de metadatos a lo largo de todos los componentes de la plataforma tecnológica.
GD-23	Acceso perfilado a gestión del dato	La plataforma deberá permitir el acceso al repositorio por perfiles y autenticación de usuarios. Pudiendo identificar por rol cuales son las funcionalidades permitidas
GD-24	Clasificación por Normativas	La plataforma deberá clasificar los metadatos según pautas marcadas por normativas internas y disposiciones legales
GD-25	Modelado del Metadato	La plataforma deberá disponer de mecanismos para especificar claves primarias o relación, reglas de validación
GD-26	Búsqueda Semántica	La plataforma deberá permitir realizar búsquedas semánticas dentro del repositorio de metadatos con una barra tipo Google con corrección de errores y sugerencias en la búsqueda
GD-27	Recomendación de Entidades	La plataforma deberá permitir recomendar automáticamente entidades de datos según un patrón de búsqueda.
GD-28	Desarrollo de conectores específicos	La plataforma deberá permitir, para aplicativos de desarrollo propio, el desarrollo de conectores específicos sin modificación del producto
GD-29	Ingestión de diccionarios de datos externos	La plataforma deberá permitir ingestar diccionarios de datos físicos directamente desde una fuente (es decir, RDBMS) u otras plataformas de metadatos
GD-30	Registro Información Adicional	La plataforma deberá contar con funcionalidades que permitan gestionar anotaciones u observaciones adicionales a los objetos que forman parte del modelo de datos.

Dominios y colecciones de datos

Ref.	Concepto	Descripción
GD-31	Creación de dominios del dato	La plataforma de gobierno debe ser capaz de crear una capa lógica agrupando los datos por dominios para su gestión, exposición y consumo. Esta definición se realiza en la instancia de la generación de las vistas de negocio, en donde se mapea la meta data técnica contra el glosario de términos
GD-32	Gestión de colecciones de Datos	La plataforma podrá gestionar, definir, modificar, eliminar de manera física o lógica, colecciones de datos compuestas por conjuntos de datos físicos (tablas de uno o diferentes datastores) y conjuntos de vistas lógicas de datos
GD-33	Combinación y fusión de datos de diferentes fuentes	La plataforma permitirá combinar diferentes tablas de datos o conjuntos de datos de diferentes fuentes de datos/ data stores dentro de una colección de datos. Los dominios de datos son un conjunto de información relevante para el negocio, los conceptos aquellos términos que permiten definirlo y las colecciones son las agrupaciones de datos para su consumo. Una colección son datos que ya tienen semántica, calidad, seguridad y están listos para ser explotados.
GD-34	Acceso directo a los datos a través de las colecciones de datos	La plataforma permitirá acceder y consumir los datos directamente a través de las colecciones de datos
GD-35	Creación de capa lógica de datos	La plataforma debe proveer de la capacidad de crear una capa lógica de datos con ontologías OWL semánticas que pueda posteriormente mapearse a los datos físicos sin tener que modificarlos
GD-36	Automapping	La plataforma deberá ser capaz de automapear los términos/conceptos ontológicos de la capa lógica con los datos auto descubiertos (tablas, columnas, etc....) y presentes en los datos físicos
GD-37	Funcionalidades avanzadas de creación de datos	La capa lógica de datos permitirá definir nuevos datos con funciones avanzadas en las que se puedan combinar varios datos técnicos y usar funciones matemáticas o funciones SQL, o tomando información de otro tipo de fuentes que contengan un mayor detalle o a través de cargas masivas desde archivos tipo Excel
GD-38	Propagación de metadatos	Los metadatos asignados a las ontologías o conceptos ontológicos se deberán poder propagar a las tablas y columnas donde la capa lógica los haya mapeado. Es decir, una vez mapeada una columna con un concepto ontológico/business terms si se desea se podrán propagar los atributos del concepto ontológico/Business terms a la columna mapeada

Glosario de negocio

Ref.	Concepto	Descripción
GD-39	Glosario de negocio	La plataforma tendrá un módulo de gestión de términos y activos de negocio
GD-40	Gestión de tipos de activos de negocio	La plataforma deberá poder definir nuevos tipos de activos de negocio, mediante plantillas predefinidas o definiendo los campos que requerirán la edición o creación del nuevo tipo de assets de negocio
GD-41	Glosario de Términos e indicadores de Negocio	La plataforma deberá permitir almacenar, agrupar, buscar, etiquetar, relacionar, importar/exportar, términos y definiciones de negocio utilizando estándar de ontologías formato tipo OWL/RFD
GD-42	Categorización de términos de negocio	La plataforma deberá poder ordenar por jerarquías customizadas los términos y elementos de negocio (KPIs, etc.)
GD-43	Monitorización de glosario de términos de negocio	<p>La plataforma podrá visualizar los términos de negocio y controlar los atributos relacionados a través de dashboards predefinidos.</p> <p>ONTOLOGÍA – DOMINIOS Descripción formal explícita de clases en un área de conocimiento, propiedades que describen varias características y atributos de las clases y restricciones (o facetas) en las propiedades. Una ontología junto con un conjunto de instancias individuales de clases constituye una base de conocimiento.</p> <p>CLASE – CONCEPTOS Las clases describen objetos en el área de conocimiento. Por ejemplo, una clase 'vinos' representa todos los vinos. Los vinos específicos son instancias de esta clase. Una clase puede tener subclasses que representan clases que son más específicos que la superclase. Por ejemplo, podemos dividir la clase de todos los vinos en vinos tintos, blancos y rosados.</p> <p>PROPIEDADES DE DATO Una vez que hemos definido algunas de las clases, debemos describir la estructura interna de las clases. Estas propiedades se convierten en elementos adjuntos a las clases. Por lo tanto, la clase 'vino' tendrá las siguientes propiedades: color, cuerpo, sabor y azúcar.</p> <p>RESTRICCIONES / FACETAS Las propiedades pueden tener diferentes facetas que describen el tipo de dato, los valores permitidos, el número de valores (cardinalidad) y otras características de los valores que puede tomar la ranura. Describe el patrón de comportamiento de la propiedad.</p> <p>PROPIEDADES DE OBJETO</p>

Ref.	Concepto	Descripción
		Las propiedades de objeto permiten crear relaciones/verbos entre clases. Por ejemplo la propiedad <se produce en> que relacionaría la clase 'vino' con la clase 'bodega' generando una tripleta: 'El vino se produce en bodegas'.
GD-44	Vinculaciones Externas	La plataforma deberá permitir incluir un enlace externo en una descripción de un término comercial para vincular a páginas web externas como referencia
GD-45	Vinculación a definiciones de negocio	La plataforma deberá permitir vincular metadatos auto descubiertos a términos e indicadores de negocio.
GD-46	Gestión automatizada	La plataforma deberá permitir identificar aquellos cambios sobre un mismo término procedentes de diferentes sistemas, y deberá gestionar cuál prevalece de forma automatizada, determinando reglas de parametrización y configuración.
GD-47	Flujos de trabajo de aprobación	La plataforma deberá permitir la definición de tareas y flujos de aprobación para la gestión de altas, modificaciones y bajas de términos.
GD-48	Parametrización de flujos de trabajo de aprobación	La plataforma deberá permitir que los flujos de trabajo sean configurables con diferentes etapas de acuerdo con las necesidades organizativas
GD-49	Carga Inicial	La plataforma deberá permitir establecer procesos de carga de datos o términos de negocio para la carga inicial de datos de unidades o departamentos nuevos y distribuirlos al resto de sistemas.
GD-50	Exposición de Servicios	La plataforma deberá permitir servicios de exposición de información a través de diferentes interfaces como JDBC/ODBC o HTTP REST. Por ejemplo, para conectar herramientas de BI o Analítica tipo Power BI o Tableau
GD-51	Perfiles de Acceso	La plataforma deberá permitir perfilado de acceso por tipo de maestro, por niveles de jerarquía / clase / familia dentro de un maestro.
GD-52	Administración de Roles	La plataforma deberá permitir la configuración de roles en base a las funciones asignadas a un área o departamento completo, administrador, funciones específicas a un centro de servicios compartidos, quién puede editar o modificar componentes de gobierno, áreas negocio, etc.
GD-53	Establecimiento de restricciones	La plataforma deberá indicar el uso permitido del conjunto de datos y / o campo de datos - Requisitos de seguridad / privacidad - Restricciones contractuales / legales
GD-54	Gestión de dominios del dato. Acceso	La plataforma de gobierno debe ser capaz de crear una capa lógica agrupando los datos por dominios para su gestión, exposición y consumo. Además, éstos deben

Ref.	Concepto	Descripción
	Directo a los datos a través de la capa lógica de datos con ontologías OWL	poder ser implementados rápidamente basándonos en ontologías OWL/RDF que son el formato estándar a nivel internacional para trabajar ontologías
GD-55	Perfilado en dominios del dato	El control y gestión de los dominios debe ser perfilado a determinados roles
GD-56	KPIs	La plataforma deberá permitir albergar lógica / cálculos detrás de los términos negocio. Por ejemplo para el cálculo de indicadores y métricas clave de negocio.
GD-57	Etiquetado	La plataforma deberá permitir "etiquetar" los términos del glosario con palabras clave y categorías de búsqueda.

Ontologías

Ref.	Concepto	Descripción
GD-58	Modelado Gráfico	La plataforma deberá proveer un interfaz gráfico para el modelado conceptual, lógico y físico, incluyendo el de los dominios de información
GD-59	Capacidad de Modelado	La plataforma deberá permitir el modelado de datos sin la necesidad de desarrollos adicionales
GD-60	Plantillas de Modelado	Estas se pueden importar en formato OWL/RDF o Excel

Knowledge Graph

Ref.	Concepto	Descripción
KG-1	Knowledge Graph	La plataforma debe ser capaz de mostrar gráficamente las relaciones entre datos, ontologías y términos de negocio
KG-2	Shortest path	La plataforma deberá poder analizar de manera automática cual es el camino más corto y la relación más directa entre datos, ontologías y términos de negocio
KG-3	Analítica de grafos sobre Knowledge Graph	La plataforma deberá poder realizar analítica de grafos sobre la información disponible para derivar relaciones complejas entre datos
KG-4	Buscador	El Knowledge Graph incorporará un buscador para permitir encontrar fácilmente elementos dentro de cualquier elemento de gobierno, ya sea en el diccionario de datos, dominios, términos de negocio o cualquiera de los elementos disponibles en la herramienta de gobierno del dato.

Ref.	Concepto	Descripción
KG-5	Selección de elementos y Foco	El componente Knowledge Graph permitirá hacer foco y enseñar las relaciones de un elemento con el resto pudiendo examinar en detalle que relaciones tiene y con qué elementos.
KG-6	Visualización avanzada	La herramienta de Knowledge Graph deberá permitir hacer zoom, mover gráficos, seleccionar capas y disponer claramente de una visualización avanzada capaz de mover múltiples capas con miles de objetos
KG-7	Visualización de ontologías OWL/RDF	La herramienta de Knowledge Graph deberá permitir visualizar automáticamente cualquier ontología semántica con formato OWL/RDF
KG-8	Navegación multisalto	La plataforma deberá poder representar varios saltos entre nodos gráficos para gestionar relaciones complejas entre la información

3.1.6 Calidad del Dato

Ref.	Concepto	Descripción
DQ-1	Gestión de Reglas de Calidad	La plataforma tendrá un módulo de calidad para gestionar las reglas de calidad de los datos
DQ-2	Reglas de Calidad zero-code	La plataforma deberá permitir crear/editar/eliminar reglas de calidad a través de interfaces sencillas (zero-code) para la identificación de incidencias (formatos, reglas de identificación de duplicados, propuestas de fusión, etc.)
DQ-3	Reglas de calidad sobre el modelo de gobierno	La plataforma permitirá centralizar y gestionar la calidad como una parte más del modelo de gobierno, permitiendo el link con el diccionario, linaje, metadatos y relaciones entre activos.
DQ-4	Definición de reglas de negocio	La plataforma deberá tener la funcionalidad que permita la definición de reglas asociadas a los datos, pudiendo definir diferentes versiones de la misma regla en función del dominio de información y del instante temporal. Se pueden aplicar reglas de calidad genéricas aplicables a varias ejecuciones de calidad
DQ-5	Perfilado y limpieza	La plataforma deberá permitir el perfilado de datos, limpieza y correspondencia entre sus fuentes y destinos
DQ-6	Reglas de Normalización	La plataforma deberá permitir establecer reglas de normalización de datos contra información contenida en tablas corporativas. Por ejemplo, tipos de vías no estandarizadas, países o localidades, etc.
DQ-7	Discovery de la calidad del dato por dominios	La plataforma deberá proveer de soporte de proyectos independientes por usuarios de tipo analistas de negocio (no técnicos). Esto es un perfilamiento inicial sobre los datos que permita identificar reglas de calidad a ser definidas

Ref.	Concepto	Descripción
DQ-8	Validación de AI y ML	La plataforma deberá permitir la validación de información mediante técnicas de Machine Learning
DQ-9	Validación de reglas de negocio en tiempo real	La plataforma deberá permitir la validación de reglas de negocio en tiempo real. Esta capacidad debe poder ser consumida mediante la consulta de servicios web o de tipo APIs Rest.
DQ-10	Validación de reglas de negocio en procesos batch	La plataforma deberá disponer de funcionalidad que permita validar el cumplimiento de las reglas de negocio asociadas a los datos de un repositorio físico elaborando informes que permitan realizar un seguimiento en el tiempo del grado de cumplimiento.
DQ-11	Detección de incumplimientos de estándares de datos en tiempo real	La plataforma deberá permitir la validación del cumplimiento de estándares asociados a los datos en tiempo real. Esta capacidad debe poder ser consumida mediante la consulta de aplicaciones y servicios web y APIs Rest.
DQ-12	Detección de incumplimientos de estándares de datos en repositorios existentes	La plataforma deberá permitir la validación del cumplimiento de estándares de calidad asociados a los datos en un repositorio existente, elaborando informes que permitan identificar los registros que incumplen los estándares definidos y que permitan hacer un seguimiento en el tiempo del grado de incumplimiento de los estándares definidos.
DQ-13	Ejecución Proactiva de reglas de calidad	Desde el modelo de gobierno se debe poder definir las reglas de calidad que deben cumplir los registros de una tabla y los flujos de escritura en esa tabla deben poder aplicarla de manera automática cada vez que se inserta algún dato
DQ-14	Monitorización de la Calidad del Dato	La plataforma deberá disponer de herramientas de monitorización de la calidad del dato en forma gráfica (cuadros de mando, informes, etc.).
DQ-15	Confidencialidad de los datos	La plataforma deberá permitir clasificar los datos según su confidencialidad conforme a normativas existentes o regulaciones legales (LOPD, GDPR, etc.)
DQ-16	Generación de reglas predefinidas	La plataforma deberá permitir incorporar reglas de calidad de información predefinidas para la identificación de incidencias (formatos de CIF, CNPJ -formatos estándar de calidad-, reglas de identificación de duplicados, etc.)
DQ-17	Generación de nuevas reglas	La plataforma deberá permitir crear reglas de calidad específicas adicionales a las existentes en la herramienta.
DQ-18	Indicadores de Calidad	La plataforma deberá disponer de indicadores de calidad que permita medir el grado de calidad de los datos existentes de forma gráfica.
DQ-19	Gestión de Indicadores de Calidad	La plataforma deberá permitir consultar y medir la evolución de la calidad de los indicadores de calidad.

Ref.	Concepto	Descripción
DQ-20	Conexión con Fuentes Externas	La plataforma deberá permitir conectar contra fuentes externas de información para la validación de la información.
DQ-21	Gestión de registros fallidos	Posibilidad de separar automáticamente registros que no cumplen todas las reglas de calidad definidas
DQ-22	Flexibilidad en la corrección del dato	La plataforma deberá permitir la corrección de la calidad del dato desde la misma herramienta en el sistema origen
DQ-23	Validación de datos	La plataforma deberá permitir evaluar el dato y proponer cambios en tiempo real antes del almacenamiento en las fuentes de datos. Esto se debe realizar utilizando las capacidades de virtualización del dato
DQ-24	Proposición de Correcciones	La plataforma deberá permitir proponer correcciones en base a los indicadores definidos y pueden ser gestionados por los propietarios de datos y responsables de calidad.
DQ-25	Remediación integrada	La plataforma deberá permitir definir workflows para correcciones automática sobre los datos
DQ-26	Trazabilidad de los procesos de validación	La plataforma deberá permitir flexibilidad a la hora de dar de alta y registrar cualquier tipo de regla que implique la transformación del dato. A través de los flujos de aprobación se definen las etapas para aprobar la regla de calidad o el proceso de remediación sobre el dato resultante del proceso
DQ-27	Auditoria de reglas de calidad	La plataforma deberá permitir la auditoria de las reglas de calidad ejecutadas sobre el dato y mostrar gráficamente los registros. Y esta información añadirla al linaje global de la plataforma

3.1.7 Linaje

Ref.	Concepto	Descripción
LJ-1	Linaje	La plataforma deberá registrar el linaje del dato y asociarlo al dato para su posterior análisis. El linaje debe representar el origen del dato y su destino incluyendo todas las relaciones entre ellos
LJ-2	Linaje Técnico y De Negocio	La Plataforma deberá poder mostrar el linaje del dato mostrando tanto los esquemas, tablas, columnas y metadato técnico, y los términos de negocio/ontologías asociadas con los metadatos técnicos
LJ-3	Linaje en grafos	La plataforma deberá poder representar en grafos, las relaciones entre datos (metadato técnico), términos de negocio, Ontologías en grafos que permitan visualizar correctamente dominios y atributos de los datos.

LJ-4	Análisis de Cambios	La plataforma deberá permitir el análisis de impacto en los cambios realizados a un dato a través de la trazabilidad de cambios entre datos y procesos de origen a destino de manera gráfica.
------	---------------------	---

3.1.8 MDM – Datos Maestros

Ref.	Concepto	Descripción
MDM-1	Revisiones y análisis	La plataforma deberá disponer de herramientas de revisión y aprobación de los datos maestros en todo el ciclo de vida del dato.
MDM-2	Soporte de multidominio	La plataforma deberá proveer soporte de multidominio funcional, es decir, cubrir diferentes tipos de áreas funcionales de datos simultáneamente como gestión de recursos humanos, contabilidad, gestión médica, citas, etc.
MDM-3	Soporte localizaciones	La plataforma deberá proveer soporte de localización para gestionar múltiples zonas geográficas según cuales sean los datos maestros y sus referencias. Esto para permitir la estandarización de los datos de acuerdo a una nomenclatura común para definir ubicaciones geográficas.
MDM-4	Limpieza de datos	La plataforma deberá tener la capacidad de ejecutar la limpieza de datos, detectando duplicidades, acrónimos, etc. Esto se dará como resultado del proceso de evaluación de calidad
MDM-5	Estandarización de datos	La plataforma deberá ser capaz de estandarizar los datos maestros según reglas preestablecidas, por ejemplo para la desduplicación de datos.
MDM-6	Gestión del ciclo de vida	La plataforma deberá ofrecer mecanismos de gestión del ciclo de vida del dato maestro y sus referencias con el contexto del dominio de información específico
MDM-7	Tratamiento de datos no estructurados	La plataforma deberá permitir el tratamiento de datos no estructurados, y permitir la gestión de maestros de información que proviene de videos si previamente han sido clasificados y estructurados.
MDM-8	Histórico de Cambios	La plataforma deberá disponer de registro histórico de cambios en cualquiera de sus campos
MDM-9	Repositorio de datos maestros	La plataforma deberá contemplar el almacenamiento de los datos maestros, estructuras jerárquicas, referencias y cualquier información relevante de los datos maestros en un repositorio
MDM-10	Visualización de Datos Maestros	La plataforma deberá ofrecer una visualización sencilla y accesible para usuarios no técnicos
MDM-11	Generación y customización de pantallas	El módulo MDM deberá poder generar las pantallas de gestión de datos y ofrecer la posibilidad de customizarlas con un look&feel adecuado de acuerdo con las necesidades y dominios de información específico para la organización

Ref.	Concepto	Descripción
MDM-12	Estandarización del Dato Maestro	La plataforma deberá ser capaz de realizar operaciones de estandarización como por ejemplo la desduplicación de datos.
MDM-13	Desduplicación del Dato Maestro	La plataforma deberá ser capaz de realizar operaciones de identificación y desduplicación del dato
MDM-14	Trazabilidad de los Cambios	La plataforma deberá registrar y dejar trazados los cambios aplicados en la resolución de incidencias / conflictos en el MDM.
MDM-15	Valores Omitidos	La plataforma deberá tener capacidad de reemplazar valores omitidos desde origen en función de reglas preestablecidas
MDM-16	Trazabilidad dato maestro	Permitir el control de versiones y registro de cambios para asegurar la trazabilidad de cambios dentro de las plataformas. La plataforma deberá permitir evaluar el dato y proponer cambios en tiempo real antes del almacenamiento en las fuentes de datos
MDM-17	Workflows de correcciones datos maestros	La plataforma deberá permitir definir workflows tipo ETL para correcciones (cambios que afecten a negocio) o en su defecto acciones a realizar según el grado de impacto
MDM-18	Gestión de la calidad de datos maestros	La plataforma deberá tener una gestión y resolución de incidencias de calidad que permita resolver tanto individual como masivamente.
MDM-19	Trusted Data	El repositorio de datos maestros es capaz de verificar si un registro ha sido modificado o no

3.1.9 Data Marketplace

Ref.	Concepto	Descripción
DMP-1	Data Marketplace	La plataforma debe permitir el consumo y publicación de datos entre unidades, en función de roles y accesos asignados.
DMP-2	Publicación de colecciones de datos	Los usuarios pueden publicar los datos de manera segura, sencilla y gráfica
DMP-3	Consumo de colecciones de datos	Las colecciones se encuentran en un catálogo disponible y pueden consumirse directamente desde el catálogo sin necesidad de exportar a ficheros u hojas Excel
DMP-4	Catálogo de colecciones de datos disponibles	Las colecciones se muestran en un catálogo común de datos
DMP-5	Atributos asociados a las colecciones	Las colecciones permiten tags y atributos para su clasificación

Ref.	Concepto	Descripción
DMP-6	Búsqueda de colecciones de datos disponibles	Las colecciones junto con sus metadatos se buscarán semánticamente por diferentes atributos y filtros
DMP-7	Configuración de colecciones consumo público	Las colecciones se podrán crear de varias maneras, cuando se publiquen para uso público se pueden consumir por cualquier usuario de la plataforma siempre que al mismo le den acceso
DMP-8	Configuración de colecciones de datos de consumo privado	Las colecciones de datos publicadas requieren la petición de acceso por parte del consumidor y la previa aprobación para su acceso
DMP-9	Protección de datos a nivel de colecciones y su interoperabilidad	La plataforma permite aplicar requisitos de cumplimiento de protección de datos personales entre los productores y consumidor. Por ejemplo, reglas de anonimización u ofuscamiento de datos
DMP-10	Definición de propósito de uso de los datos	La plataforma permite aplicar requisitos de cumplimiento de protección de datos personales entre los productor y consumidor. Por ejemplo, reglas de anonimización u ofuscamiento de datos
DMP-11	Temporalidad del "data contract"	Definición automática y notificación de fecha de caducidad del contrato.
DMP-12	Revocación de accesos automática	Cuando un data contract deje de tener validez, la plataforma debe automáticamente configurar las colecciones para dejar de estar accesibles
DMP-13	Mecanismos de aprobación de accesos	Aprobación de acceso a datos publicados de manera privada
DMP-14	Mecanismos de revocación de accesos	Revocación de permisos de acceso a datos privados previamente autorizados para su uso.
DMP-15	Automatización de permisos de acceso	La plataforma debe ser capaz de accionar de manera automática el acceso a las colecciones de datos una vez se haya aprobado el acceso
DMP-16	Analítica de acceso y consumo de datos/colecciones	La Plataforma permitirá hacer analítica del consumo de las colecciones y los datos en reportes y dashboards configurables de manera sencilla por los usuarios
DMP-17	Identificación y priorización automática de las colecciones y datos más importantes	El Data Market será capaz de indicar las colecciones y datos más utilizados, y priorizarlos o señalarlos según varios criterios y fórmulas que se podrán configurar según desee cada empresa

3.1.10 Seguridad y Cumplimiento del Dato

Ref.	Concepto	Descripción
SC-1	Seguridad Centralizada	La plataforma deberá tener un módulo de seguridad centralizada para poder gestionar toda la seguridad de acceso al dato de manera centralizada configurable por usuario, dominios y por atributos ABAC.
SC-2	Configuración flexible	La plataforma deberá ser parametrizada para mejorar su seguridad o corregir vulnerabilidades
SC-3	Integración con Directorio Activo/Gestión de Identidades	La plataforma deberá integrarse con el sistema de Gestión de Identidades y/o del Directorio Activo.
SC-4	Seguridad de acceso al dato	La seguridad de acceso al dato se podrá garantizar en el origen de acceso para cualquier uso del dato sin tener que configurar cada herramienta
SC-5	Seguridad por dominios de datos	La plataforma podrá perfilar la seguridad de acceso al dato sobre los dominios expuestos tanto sobre la capa lógica como la capa física del dato
SC-6	Seguridad de grano fino	La plataforma soportará seguridad de grano fino (tablas y columnas) sobre cualquier acceso a datastores (interno o externo)
SC-7	Seguridad ABAC	La plataforma soportará seguridad de grano fino (tablas y columnas) por atributos asignados al catálogo de datos
SC-8	Encriptación del dato	La plataforma deberá disponer para el almacenamiento, métodos de encriptación de datos más comunes del mercado y esta encriptación deberá ser transparente a las aplicaciones y usuarios.
SC-9	Enmascaramiento del dato	La plataforma deberá permitir enmascarar los datos.
SC-10	Ocultación del dato	La plataforma podrá esconder datos no autorizados a los usuarios
SC-11	Encriptación de Red	La plataforma deberá garantizar que los datos viajan encriptados, para los canales de comunicación de las aplicaciones con los diferentes repositorios de datos.
SC-12	Segmentación de red	La plataforma deberá tener los sistemas de aplicación y las bases de datos en subredes independientes
SC-13	Hasheo de la información	La plataforma deberá ser capaz de hashear la información cuando se consuma on-the-fly para ocultar datos
SC-14	Cruce de información hasheada	La plataforma debe ser capaz de cruzar datos hasheados sin enseñarlos al usuario en ningún momento

Ref.	Concepto	Descripción
SC-15	Encriptado de copias de seguridad	La plataforma deberá permitir encriptar/desencriptar las copias de seguridad de los diferentes repositorios de forma transparente sobre los soportes físicos de estas copias.
SC-16	Integridad de los datos	La plataforma deberá disponer de mecanismos que garanticen la integridad de los datos en los canales de comunicación detectando paquetes perdidos y modificados.
SC-17	Gestión de acceso al dato	La plataforma deberá restringir el acceso a los datos por medio de perfiles de acuerdo con las disposiciones internas o regulaciones legales (LOPD, GRPD, ley de protección de datos, etc.)
SC-18	Separación de funciones	La plataforma deberá permitir una separación de funciones clara entre los administradores de los repositorios y los responsables de gestionar la seguridad de los datos.
SC-19	Gestión descentralizada de los procesos de encriptación	La plataforma deberá proporcionar herramientas para la gestión centralizada, por parte del responsable de seguridad, de los procesos de encriptación y sus claves asociadas.
SC-20	Accesos y comportamiento	La plataforma deberá permitir la monitorización de accesos por medio de herramientas gráficas
SC-21	Clasificación y Confidencialidad del Dato	La plataforma deberá permitir clasificar los datos en función de la confidencialidad del dato según normativas internas/externas aplicables
SC-22	Auditorías del dato	La plataforma deberá proveer de mecanismos para realizar auditorías de accesos, de seguridad y vulnerabilidades
SC-23	Notificación de alertas	La plataforma deberá incluir medios de alertas o avisos en caso de detección de riesgo de violación de las políticas establecidas
SC-24	Regulación GDPR	La plataforma deberá gestionar el cumplimiento de la regulación de protección de datos GDPR

3.1.11 Data Privacy

Ref.	Concepto	Descripción
DP-1	Identificación datos sensibles	La plataforma deberá identificar de forma automática datos definidos como sensibles en base a unas reglas definidas previamente
DP-2	Clasificación de datos sensibles	La plataforma deberá clasificar de forma automática estos datos en función de sus diferentes niveles de sensibilidad

3.1.12 Capacidades Analíticas

Ref.	Concepto	Descripción
A-1	Analítica	La plataforma debe incluir herramientas para minería y análisis de datos.
A-2	Procesamiento Distribuido	La plataforma deberá contar con la capacidad de procesamiento distribuido, siempre teniendo en cuenta que han de ser servicios administrados
A-3	Acceso a Múltiples fuentes	La plataforma será capaz de ingestar datos procedente de diversas fuentes de datos, tanto estructuradas, semiestructuradas y no estructuradas. Tanto internas como externas: BBDD, buses de datos, sistemas de ficheros, dispositivos IOT, logs, webs, aplicaciones, redes sociales etc.
A-4	Interoperabilidad y extensibilidad	La plataforma debe tener la capacidad para extender las capacidades de interoperabilidad con el fin de integrar la mayor cantidad de datos provenientes de fuentes externas relevantes a través de interfaces estándares como json.
A-5	Ingesta de datos en Tiempo Real	La plataforma deberá contar con la capacidad de ingesta de datos en tiempo real, o cerca de tiempo real, que cuente al menos con las funcionalidades de distribución y/o carga de datos a otros destinos, transformación de los datos y almacenamiento temporal configurable.
A-6	Procesado de datos en Streaming	La plataforma deberá contar con la capacidad de procesamiento de datos en tiempo real pudiendo realizar análisis y agregaciones de los datos en tiempo real, o cercano a tiempo real. Al menos deberá contar con la capacidad de realizar los análisis con lenguajes basados en SQL.
A-7	Ingesta y procesado de datos por lotes	La plataforma deberá contar con la capacidad de ingestar, procesar y trasladar datos entre diferentes servicios de proceso y almacenamiento, así como orígenes de datos locales/remotos a intervalos especificados o en modo ad-hoc
A-8	Datos	La plataforma tiene capacidades de virtualización de datos.
A-9	Basado en eventos o planificado	La plataforma deberá contar con la capacidad de ejecución de procesos/código en respuesta a eventos o programaciones sin necesidad de aprovisionar ni administrar servidores.
A-10	Ingesta de ficheros	La plataforma deberá ser capaz de capturar incrementalmente cambios producidos en ficheros que están siendo simultáneamente modificados por otros aplicativos.
A-11	Ingestas masivas	La plataforma debe disponer de utilidades que simplifiquen la ejecución de cargas masivas, por ejemplo, cargas iniciales de datos, etc.
A-12	Acceso a datos vía API	La plataforma deberá permitir el procesamiento de datos provenientes directamente de servicios web seguros y APIs (REST).

Ref.	Concepto	Descripción
A-13	Acceso vía mensajería	La plataforma tendrá la capacidad de que los sistemas externos, puedan suscribirse a los sistemas de mensajería de la plataforma.
A-14	Transferencia Bidireccional	La plataforma debe contar con comunicación bidireccional, permitiendo a la capa de ingesta importar, alimentar, así como exportar o consultar los datos que en ella residen.
A-15	Velocidad de Ingesta de datos	La plataforma deberá permitir diferentes velocidades de ingestas: tiempo real, microbatch, batch, etc.
A-16	Flujos programáticos	La plataforma deberá permitir crear flujos de trabajo de forma programática y o gráfica, y también planificarlos y monitorizarlos de forma centralizada.
A-17	Tipología de Datos	La plataforma deberá permitir el tratamiento de datos estructurados, semiestructurados y no estructurados.
A-18	Modelo lógico de datos	La plataforma deberá permitir el desarrollo de flujos de datos dependiendo únicamente del modelo de datos lógico definido sin depender del modelo físico existente
A-19	Change data Capture	La plataforma deberá permitir capturar cambios producidos en BBDD y procesar la información asociada, bien disparando procesos de transformación, bien replicando estos cambios en otros repositorios
A-20	Consistencia de las copias	La plataforma deberá ser capaz de garantizar la consistencia de los datos en el destino
A-21	Copia de datos automática	La plataforma deberá ser capaz de gestionar la copia y sincronización de datos entre origen y destino de manera automática sin bloquear el origen de datos y de manera totalmente automática (sin desarrollo de workflows de ningún tipo)
A-22	Promoción entre entornos	La plataforma deberá disponer de mecanismos que permitan el control de versiones de procesos y la promoción de estos entre entornos de desarrollo, prueba, producción, etc.
A-23	Acceso plataforma vía API	La plataforma deberá exponer sus procesos vía API permitiendo el lanzamiento de estos procesos desde otros aplicativos. Siendo capaz la plataforma de crear, publicar, monitorear y autenticar fácilmente y a cualquier escala
A-24	Visualización básica	La solución deberá contar con características que permitan explorar los datos a través de la generación de estadísticos descriptivos básicos, análisis de datos duplicados, faltantes, atípicos. Además de la presentación de tablas personalizadas para valorar según los criterios definidos por el usuario referente a la calidad de los datos, además, permitirá visualizar los datos y sus resultados analíticos gráficamente, en mapas, pasteles, diagramas de barras.

Ref.	Concepto	Descripción
A-25	Visualización avanzada	La plataforma permitirá la implementación de visualizaciones avanzadas de datos a través de distribuciones, histogramas, diagramas de caja, diagramas de malla, dispersiones, líneas, barras, circulares, áreas, mixtos, etc.
A-26	Análisis de datos	Deberá permitir preparar datos para modelos en series de tiempo y modelos de corte transversal
A-27	Integración	La solución permitirá importar/exportar modelos basados en lenguajes estándares mínimo PMML (Predictive Model Markup Language).

3.1.13 Capacidades ETL

Ref.	Concepto	Descripción
ETL-1	Herramienta Low-code	La plataforma permitirá el diseño gráfico de flujos de datos que automáticamente puedan ejecutarse tanto sobre altos como bajos volúmenes de datos
ETL-2	Colas de ejecución	La plataforma deberá disponer de capacidades que permitan definir prioridades en la ejecución de los procesos
ETL-3	Permisos de accesos	La plataforma deberá permitir gestionar los permisos necesarios para la ejecución de los procesos.
ETL-4	Tipos de integración	La plataforma deberá permitir diferentes paradigmas de integración de datos: basado en eventos, basado en APIs, mensajería, colas de mensajes confiables y de publicación/suscripción etc.
ETL-5	Independencia	La plataforma deberá funcionar tanto de forma autónoma (diseño y lanzamiento de procesos desde sus propias interfaces) como integrada en la plataforma de Big Data
ETL-6	SDK conectores	La plataforma debe disponer de un frameworks de desarrollo que permita crear nuevos conectores y modificar los existentes
ETL-7	Orquestador de Procesos	La plataforma debe mantener un log del proceso de carga que permita consultar y gestionar los errores que se presenten en los procesos ETL
ETL-8	Orquestador de Flujos	La plataforma deberá contar con la capacidad de orquestar la integración de los datos entre los entornos on-premise y las herramientas y productos de la plataforma BigData, que permita integrar ambos entornos, estableciendo los flujos de datos necesarios.
ETL-9	Triggers de lanzamiento de procesos	La plataforma deberá permitir el lanzamiento de procesos ETL batch o streaming a partir de eventos producidos en sistemas de mensajería

Ref.	Concepto	Descripción
ETL-10	Acceso a repositorio de datos	La plataforma tendrá la capacidad de que los sistemas externos, puedan conectarse a los diferentes repositorios de la plataforma, por ejemplo, almacenamiento distribuido, BBDD en memoria, etc.,
ETL-11	Desarrollos personalizados	La plataforma debe tener la capacidad de desarrollar sobre ella componentes personalizados para ser incluidos en los flujos, este desarrollo debe soportar como mínimo 3 los siguientes lenguajes SCALA, PYTHON o PySpark.

3.1.14 Preprocesamiento de datos para modelos

Ref.	Concepto	Descripción
PD-1	Exploración de datos	Disponer de herramientas gráficas que permitan explorar los datos con el objetivo de al menos identificar la importancia de las variables, redundancia en las variables, presentar medidas de variabilidad de los datos (varianza, desviación estándar, rango, etc.), medidas estadísticas de tendencia central (promedio, mediana, moda)
PD-2	Análisis de datos atípicos	Debe proveer herramientas que permitan identificar datos atípicos y su tratamiento
PD-3	Análisis de datos nulos	Debe describir las características de las variables respecto a valores nulos (porcentajes de nulos por columnas, si los nulos son aleatorios o determinísticos) y tener estrategias de su tratamiento
PD-4	Transformaciones de datos	Debe detectar variables con alta cardinalidad, alta curtosis, alto sesgo (skewness), baja entropía, bajos índices de variación. Luego de la identificación, debe tener estrategias de transformación de datos que traten estos problemas (por ejemplo aplicar alguna función matemática a los datos, estandarizar datos, etc.)
PD-5	Enriquecimiento de datos	Permite agregar nuevas variables de datos no estructurados, por ejemplo datos texto en logs, pdf, variables que datos tipo descripción, etc.
PD-6	Filtrado de datos	Debe tener la característica de detectar datos con valores raros, extremos y atípicos; y su tratamiento.
PD-7	Selección de variables	Dispones métodos supervisados y no supervisado para seleccionar las variables de mayor importancia.
PD-8	Extracción de características	Se debe tener la posibilidad de transformar las variables de los datos a un espacio de menor dimensión.
PD-9	Clúster de variables	Determinar la colinealidad o redundancia de las variables

PD-10	Cálculo de variables	Deberá permitir preparar datos para modelos en series de tiempo y modelos de corte transversal
-------	----------------------	--

3.1.15 Machine Learning

Ref.	Concepto	Descripción
ML-1	Capacidades analíticas	La plataforma deberá contar con la capacidad de realizar estudios de analítica descriptiva. Dichos entornos tendrán en cuenta las librerías estándares de mercado, tanto de procesamiento de datos, estadística y visualización.
ML-2	Capacidades analíticas	La plataforma pondrá a disposición de los científicos de datos un catálogo predefinido de algoritmos para realizar estimaciones, clasificaciones y ranking usando técnicas supervisadas y no supervisadas. Por ejemplo, que se incluya algoritmos de patrones de reconocimiento, minería de texto (para fuente de datos en idioma español e inglés), árboles de decisión, redes neuronales, etc.
ML-3	Capacidades analíticas	La plataforma deberá contar con la capacidad de entrenar, implementar, automatizar y administrar modelos de aprendizaje automático y minería de datos.
ML-4	Capacidades analíticas	La plataforma deberá disponer de capacidades de análisis de grafos.
ML-5	Capacidades analíticas	La plataforma deberá tener la capacidad de procesamiento de lenguaje, información textual no estructurada, documentos, minería de textos, etc.
ML-6	Capacidades analíticas	Deberá incluir al menos las siguientes técnicas de Minerías de Datos: clasificación, segmentación o clustering, asociación, predicción, grafos y redes, reducción de dimensiones, modelos de optimización y minería de texto y cuyos métodos específicos consideran: Máquina de Soporte vectorial, Redes neuronales Base Radial y Multicapa, Redes Bayesianas y Clasificadores Naive-Bayes, Regresiones Lineales Ordinarias y Generalizadas, Regresiones Logit - Probit binarias y multi - categoría, Análisis Discriminante, Análisis de Supervivencia COX (Weibull), Clasificadores por vecinos próximos, Clasificadores incrementales, A priori predictivo y no predictivo, Carmaon, Identificación de patrones secuenciales, Conglomeración Kohonen, Vecino más próximo, k-medias, Conglomeración Jerárquica, Análisis Factorial, Árboles de Decisión, Series de tiempo ARIMA, Detección de Anomalías, modelos de optimización y simulación y minería de texto con procesamiento de lenguaje natural (NLP).
ML-7	Integración	Capacidades de reutilización de código y artefactos implicados en el ciclo de vida del modelo

Ref.	Concepto	Descripción
ML-8	Integración	La plataforma deberá disponer de mecanismos que permitan el acceso desde las soluciones de BI con mayor implantación en el mercado sin necesidad de realizar desarrollos, como por ejemplo MicroStrategy, PowerBI, Tableau
ML-9	Integración	Se debe tener la capacidad de agregar algoritmos y herramientas de machine learning de Python y R
ML-10	Integración	La plataforma dispondrá de herramientas que permitan diseñar, compartir y versionar diferentes análisis en notebooks (opcional)
ML-11	Integración	Capacidades de integración con git
ML-12	Proyectos o áreas de trabajo	La plataforma debe permitir la creación de proyectos para construir uno o varios modelos
ML-13	Proyectos o áreas de trabajo	La plataforma debe permitir configurar los recursos que establecen los máximos de ejecución en Cores, RAM que serán asignados a los proyectos que contienen los modelos.
ML-14	Proyectos o áreas de trabajo	La plataforma de forma integrada debe permitir establecer una política de ejecución de forma periódica de backup sobre un determinado proyecto.
ML-15	Proyectos o áreas de trabajo	La plataforma debe permitir definir pipelines / flujos de entrenamiento para la creación de modelos de machine learning
ML-16	Proyectos o áreas de trabajo	Los proyectos deben permitir la reutilización de modelos previamente creados en la misma plataforma
ML-17	Proyectos o áreas de trabajo	Cuando se requiera, se disponga de un ambiente de trabajo para entrenar modelos de manera gráfica y con una herramienta zero-code
ML-18	Proyectos o áreas de trabajo	Capacidades para cubrir el ciclo de vida completo del preprocesamiento, modelado, evaluación y despliegue
ML-19	Fuente de datos	La plataforma debe permitir conectarse a fuentes de datos locales, bases de datos, fuentes no estructuradas y fuentes en streaming.
ML-20	Fuente de datos	Capacidades para visualizar el datalake de datos haciendo agnóstico al usuario el origen de los datos, siendo un acelerador al evitar tener que estarse conectando a los orígenes, suponiendo facilidades en el entendimiento.
ML-21	Fuente de datos	Capacidades de forma integrada de la visualización de los datalake de todos los orígenes de datos de manera estándar, permitiendo realizar sentencias tipo SQL sobre los datos, siendo agnóstico el origen, permitiendo guardar esas sentencias y sus resultados que pueden ser exportados en diferentes formatos como json, csv, ficheros tipo HDFS entre otros.
ML-22	Fuente de datos	Capacidades para cruzar de manera directa datos de diferentes orígenes siendo agnóstico para los usuarios

Ref.	Concepto	Descripción
ML-23	Fuente de datos	La plataforma debe tener la funcionalidad de forma integrada de la visualización de los datalake de todos los orígenes de datos de manera estándar, configurando a través de políticas de acceso que le permiten a los usuario ver y/o modificar determinados datos en función de su perfil
ML-24	Preprocesamiento de datos	La plataforma debe contar de forma integrada con la generación datos de pruebas para permitir su rápida integración con los modelos que se puedan generar.
ML-25	Preprocesamiento de datos	La plataforma debe contar con diferentes estrategias para la creación de datos de pruebas tales como texto plano, archivos con datos, queries SQL a bases conectados al entorno entre otros
ML-26	Preprocesamiento de datos	Técnicas de selección de datos para entrenamiento, validación y pruebas en base a estrategias, por ejemplo RandomSplit o StratifiedSampling, balanceo de eventos
ML-27	Entrenamiento	La plataforma deberá disponer de capacidades de tolerancia a fallo en caso de que algún proceso de tratamiento del dato o modelado, o parte del mismo falle
ML-28	Entrenamiento	Capacidades para obtener directamente los metadatos de un modelo para datos de entrenamiento, evaluación y test.
ML-29	Entrenamiento	Capacidades para generar y guardar directamente los parámetros, métricas y demás sucesivas reevaluaciones de un modelo, quedando registradas junto con él, así como sus métricas, EDA, y el metadato. Incluyendo poder guardar como meta datado de un modelo todas las métricas asociadas a cada ejecución de este, e ir analizando de forma temporal si existe deriva en las métricas, o en la calidad del dato entrante.
ML-30	Entrenamiento	La plataforma, para las fases de entrenamiento de los modelos, permitirá el uso de recursos distribuidos basados en diferentes tipos de Hardware, por ejemplo, CPUs, GPUs, así como unidades de procesamiento especializadas en IA. -El uso de estos recursos podrá aprovisionarse bajo demanda
ML-31	Entrenamiento	La plataforma debe permitir la construcción de modelos con la opción de configuración automáticos de parámetros cuando se requiera entrenar un modelo y establecer un desempeño inicial.
ML-32	Entrenamiento	La plataforma debe permitir crear modelos usando notebooks de forma integrada a la plataforma

Ref.	Concepto	Descripción
ML-33	Análisis de modelos	Capacidades para mostrar métricas gráficas del proceso de entrenamiento de modelos usando el conjunto de entrenamiento, evaluación y pruebas.
ML-34	Análisis de modelos	Capacidades para evaluar el desempeño de modelos y seleccionar el mejor en base a una condición. Por ejemplo, métricas de precisión, recall, curvas ROC, matrices de confusión, F1 score, etc.
ML-35	Análisis de modelos	La plataforma deberá proveer herramientas y técnicas de interpretación de los modelos de forma agnóstica al tipo de técnica de machine learning usado. Por ejemplo, uso de técnicas de interpretación local y global.
ML-36	Despliegue de modelos	La plataforma debe ofrecer la posibilidad de gestionar el despliegue y el uso de los modelos desde el catálogo/repositorio de modelos ya sea a través de microservicios o a través de procesos analíticos
ML-37	Promoción de entornos	Permitir que los activos generados, inclusive los modelos puedan ser promocionados entre entornos.
ML-38	Promoción de entornos	La pipeline incluirá pasos necesarios antes de hacer el despliegue, como por ejemplo ejecución de pruebas o análisis de código antes de promocionar
ML-39	Modelos desplegados	Una vez el modelo es desplegado en cualquier contexto, deberá contar con la capacidad de ser monitoreado a nivel de métricas técnicas del modelo o a nivel de dato, permitiendo validar su desempeño y ejecutar tareas de seguimiento y afinamiento en caso de ser requerido
ML-40	Modelos desplegados	La plataforma debe permitir visualizar la información de actividad de los modelos analíticos para saber cuántas veces se han lanzado, estado de finalización, tiempo que ha durado, uso de los recursos usados y los usuarios que lo han lanzado
ML-41	Modelos desplegados	La plataforma deberá tener una arquitectura que proporcione modelos empaquetados y expuestos a través de APIs, las cuales se pueda acceder programáticamente a la salida del análisis y pueda ser consumida por herramientas/servicios externos.
ML-42	Modelos desplegados	La plataforma debe permitir diferentes formas de ejecución de los modelos tales como: contexto batch, híbrido, streaming y microservicios con el fin de usarlos de acuerdo con la necesidad del negocio
ML-43	Repositorios Versiones	La plataforma deberá tener un sistema de almacenamiento que permita disponer de un repositorio de versiones de los distintos lenguajes de programación, librerías, código base, etc.

Ref.	Concepto	Descripción
ML-44	Repositorio de modelos	La plataforma deberá tratar a los modelos como un activo con versiones, metadatos asociados y métricas por lo que debe ofrecer las herramientas necesarias para la gestión del ciclo de vida
ML-45	Cuadro de Mando	Los resultados de los modelos deben mostrarse en tableros de mando o dashboards. Los tableros deben actualizarse en tiempo real o bajo demanda para que los usuarios de negocio sean autosuficientes para visualizar los resultados de los análisis sobre los datos.

3.1.16 Capacidades incluidas aplicaciones operacionales

Ref.	Concepto	Capacidades incluidas aplicaciones operacionales
OP-1	Acceso a los datos de uso científico	La plataforma debe, además de tener capacidades de virtualización, poder compartir el acceso a conjuntos de datos (data sets) debidamente anonimizados y securizados para disponibilizarlos en la plataforma a los equipos investigadores.
OP-2	Acceso a los datos securizado	La plataforma debe proveer de un acceso securizado, perfilado y auditado a los datos pseudoanonimizados
OP-3	Soporte para sistemas de anuncios basados en voz.	La plataforma debe disponer de capacidades de texto a voz que permitan implementar sistemas de anuncios basados en voz
OP-4	Capacidades operacionales	La plataforma permitirá el desarrollo y ejecución de sistemas operacionales basados en microservicios haciendo hincapié en el uso de las tecnologías opensource y contenerización: Kubernetes y Docker. En este ámbito la plataforma deberá garantizar por diseño que se soporte el aislamiento de servicios y de datos
OP-5	Capacidades operacionales	Los servicios operacionales deberán poder adaptar los recursos de computación necesarios para cubrir la carga del sistema automáticamente optimizando así la gestión de los recursos para los usuarios.
OP-6	Capacidades operacionales	Los servicios operacionales deberán poder integrarse por medio del estándar OAUTH2 y la integración con @Firma
OP-7	Capacidades de integración técnica de redes	La plataforma deberá soportar integración por BGP y podrá asumir que se le deleguen zonas de DNS que permita una gestión automática de dominios.

3.1.17 Soporte y capacitación

Ref.	Concepto	Soporte y capacitación
SYC-1	Soporte en español	La plataforma cuenta con soporte completo en idioma español
SYC-2	Soporte 24x7 y 8x5 en horario local	El proponente cuenta con servicio de soporte 24x7 para entornos productivos y 8x5 para entornos de pruebas y desarrollo en horario local
SYC-3	Certificación y formación	El proponente cuenta con un programa de certificación y formación para el uso de la plataforma tecnológica

3.2 Migración de Datos y/o reingeniería de flujo de datos

Comprende la definición y ejecución de los planes de trabajo para la migración y/o reingeniería de los flujos de datos más relevantes en el SRI relacionados a los casos de uso que se planteen, desde las fuentes de origen hacia la plataforma tipo Data Fabric. Esta migración y/o reingeniería deberá garantizar la integridad y la calidad de los datos, y la disponibilidad de la información tanto para los sistemas y aplicaciones que hoy día consumen y alimentan las diversas fuentes de origen, así como para los nuevos sistemas que se desarrollarán en el marco de la modernización tecnológica para el SRI.

Además comprende la migración de los flujos de trabajo de los proyectos de minería de datos actuales desarrollados en la Institución, con el acompañamiento de técnicos certificados por el fabricante, partner y/o proveedor.

3.2.1 Información de referencia de volúmenes de datos actuales

Los sistemas transaccionales del Servicio de Rentas Internas en la actualidad soportan su persistencia en tres motores de bases de datos relacionales: siendo el más relevante Oracle y en mínima proporción SQL Server y MySQL.

Infraestructura actual (Datawarehouse)					
No.	Descripción	Información actual y/o características mínimas requeridas			
1	Fabricante y versión del producto	BASE DE DATOS	VERSIÓN		
		CEL	Oracle 19.16		
		CELPAS	Oracle 19.16		
		SRIDW	Oracle 11.2.0.3		
		STAR	Oracle 19.17		
2	Detalle infraestructura	BASE DE DATOS	PROCESADOR	MEMORIA	SISTEMA OPERATIVO
		CEL	Nodo 1: 5 cores	512Gb	Sun 11.4.45
			Nodo 2: 3 cores	240 Gb	
		CELPAS	3 cores	200Gb	Sun 11.4.45
		SRIDW	Nodo 1: 3 cores	92Gb	Sun 11.4.45
			Nodo 2: 3 cores	92Gb	
STAR	6 cores	148Gb	Sun 11.4.45		
3	Almacenamiento (capacidad, tecnología, esquema de conectividad, tipo de protección)	Almacenamiento NetApp: AFF-A700 – FAS9000 Tipo de discos SAS – NLSAS – SSD Tipo RAID 6. Conectividad SAN. Capacidad física: • SAS: 800TB • NLSAS: 360TB • SSD:84TB			
4	Esquemas de HA o replicación	BASE DE DATOS	HA	REPLICACIÓN	
		CEL	Si	Si	
		CELPAS	NO	NO	
		SRIDW	Si	Si	
		STAR	NO	NO	
5	Usuarios creados por base de datos	BASE DE DATOS	# USUARIOS ESQUEMA		
		CEL	4.497		
		CELPAS	2.052		
		SRIDW	3.825		
		STAR	3.176		

Infraestructura actual (Datawarehouse)																	
No.	Descripción	Información actual y/o características mínimas requeridas															
6	Número estimado de tablas	<table border="1"> <thead> <tr> <th>BASE DE DATOS</th> <th># TABLAS</th> <th># TABLAS ÚTILES</th> </tr> </thead> <tbody> <tr> <td>CEL</td> <td>896</td> <td>358</td> </tr> <tr> <td>CELPAS</td> <td>751</td> <td>300</td> </tr> <tr> <td>SRIDW</td> <td>778</td> <td>311</td> </tr> <tr> <td>STAR</td> <td>2295</td> <td>918</td> </tr> </tbody> </table>	BASE DE DATOS	# TABLAS	# TABLAS ÚTILES	CEL	896	358	CELPAS	751	300	SRIDW	778	311	STAR	2295	918
BASE DE DATOS	# TABLAS	# TABLAS ÚTILES															
CEL	896	358															
CELPAS	751	300															
SRIDW	778	311															
STAR	2295	918															
7	Tamaño estimado de la base de datos más grande	<table border="1"> <thead> <tr> <th>BASE DE DATOS</th> <th>TAMAÑO ACTUAL</th> </tr> </thead> <tbody> <tr> <td>CEL</td> <td>54TB</td> </tr> <tr> <td>CELPAS</td> <td>341TB <i>(base de datos más grande)</i></td> </tr> <tr> <td>SRIDW</td> <td>37TB</td> </tr> <tr> <td>STAR</td> <td>62TB</td> </tr> </tbody> </table>	BASE DE DATOS	TAMAÑO ACTUAL	CEL	54TB	CELPAS	341TB <i>(base de datos más grande)</i>	SRIDW	37TB	STAR	62TB					
BASE DE DATOS	TAMAÑO ACTUAL																
CEL	54TB																
CELPAS	341TB <i>(base de datos más grande)</i>																
SRIDW	37TB																
STAR	62TB																

Los datos expuestos son de carácter informativo y no representan la totalidad de la información que el SRI dispone para los diferentes tipos de análisis que ejecuta.

3.3 Servicios de Gestión

La implementación de la nueva arquitectura tecnológica de datos en el SRI tiene el objetivo de fortalecer la gestión de los datos para optimizar el aprovechamiento y uso de estos, tanto en la prestación de los servicios como en la toma de decisiones en la entidad, enmarcados en la gestión efectiva del ciclo de vida de los datos.

Comprende el diseño, ejecución y entrega del modelo de operación y gestión de datos para el SRI en concordancia con las políticas de gobernanza, calidad, seguridad y demás sistemas de gestión que en el SRI se definan.

Comprende la definición y ejecución de los procesos, procedimientos y demás instrumentos que sean necesarios basados en el marco de referencia de gestión de datos del DAMA- DMBOK2. El fabricante, partner y/o proveedor deberá:

- Diseñar, proveer, implementar, soportar y mantener una solución de gestión de datos unificada, consistente, efectiva y administrable que permita potenciar el aprovechamiento de los datos de la Entidad para volver más eficientes los servicios institucionales y los procesos tributarios.
- Facilitar el manejo y explotación de la información institucional, así como la gestión de datos históricos de la entidad, a través de la provisión e implementación de herramientas y procesos de gestión de datos maestros (MDM) y de metadatos unificados, garantizando su consistencia a lo largo de los flujos de información.

Proveer e implementar procesos y procedimientos que generen capacidades que permitan gestionar los datos a lo largo de todo su ciclo de vida, aplicando las mejores prácticas y atendiendo a los lineamientos de gobierno digital del país.

Los componentes mínimos del modelo de gestión de datos a prestar por el fabricante, partner y/o proveedor deben incluir cómo mínimo los relacionados en el marco de referencia de gestión de datos del DAMADMBOK2, que se ilustra a continuación:

Gráfico 3: DAMA-DMBOK2 Marco de Referencia de Gestión de Datos



DAMA-DMBOK2 es una guía y un marco metodológico orientado a la Gestión de datos, que describe conceptos e identifica ampliamente los objetivos de gestión de datos, funciones y actividades, entregables, roles, principios, tecnología y cuestiones de organización y cultura que conforman el ecosistema de los datos.

Es además un compendio de buenas prácticas en constante evolución y adaptación a los dinámicos escenarios tecnológicos, por esta razón se ha integrado a la arquitectura digital del SRI, el uso de este marco de referencia.

El fabricante, partner y/o proveedor deberá implementar y presentar al SRI procesos y flujos estandarizados para la gestión de requerimientos, de incidentes y problemas, a fin de garantizar la atención de estos.

3.4 Requerimientos Técnicos

3.4.1 Tecnológicos

La solución debe acoger las mejores prácticas de la industria para lo cual el proveedor debe detallar en su propuesta la arquitectura tecnológica para su implementación, sea esta on premise o cloud. Adicionalmente la solución debe tener un esquema de alta disponibilidad y alto rendimiento.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

Se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia; donde se provea todo el ciclo de construcción, de ser necesario y por condiciones de agilidad, facilidad o estrategia podrían requerirse ambientes adicionales. Los ambientes deben estar 100% aislados el uno del otro (física o lógicamente) de tal forma que el acceso a cada uno de ellos pueda estar debidamente segmentado y se pueda configurar por grupos de usuarios, quienes tendrían acceso a cada ambiente.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

La información en ambiente de desarrollo y pruebas debe estar encriptada y despersonalizada.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

El proveedor deberá indicar la metodología de liberaciones que utilizará, en la cual se deberá evidenciar un manejo del control de versiones del código.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

El proveedor debe evaluar también la posibilidad de entregar al SRI el código fuente de los desarrollos personalizados que se requieran para la integración.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

Las interfaces que sean de uso de los contribuyentes deberán cumplir con las normas de accesibilidad contempladas en la normativa ecuatoriana (Normas de Accesibilidad Nivel AA).

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.2 *Monitoreo*

Se debe detallar el esquema de monitoreo preventivo y correctivo, considerando que la solución debe mantener un nivel de servicio en un esquema 7x24x365. Se requiere conocer el esquema de atención de incidentes y requerimientos.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.3 Seguridad

Detallar los elementos, certificaciones, parámetros, procesos y políticas sobre los cuales se apoya la solución para garantizar la seguridad informática y de información. Se debe considerar el cumplimiento de la norma de seguridad contemplada en la normativa ecuatoriana (EGSI³).

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.4 Sistema de Gestión de Identidades

El control de acceso de la solución se debe integrar al sistema de gestión de identidades y Directorio Activo de la Institución. Especificar con que soluciones del mercado son compatibles.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

³ El Esquema Gubernamental de Seguridad de la Información (EGSI) versión 2.0 registrado en el Registro Oficial N° 228 del 10 de enero del 2020 se encuentra publicado en el siguiente link: <https://www.gobiernoelectronico.gob.ec/egsi/>

3.4.5 Gestión de Eventos e Información de Seguridad

Detallar el esquema para la gestión de eventos e incidentes de seguridad.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.6 Interoperabilidad

El proveedor deberá indicar el mecanismo de integración entre sus módulos y herramientas existentes en la Institución.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.7 Integración con terceros

Para la integración con instituciones del Estado, la solución debe permitir intercambiar datos e información electrónica que son necesarios en los trámites y servicios ciudadanos que presta el Servicio de Rentas Internas y otras instituciones del Estado con las cuales mantiene convenios de intercambio de información.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.8 Implementación

Se requiere conocer la propuesta de implementación que el proveedor ejecutará en la Institución, considerando los flujos de trabajo, integraciones, orquestación y estabilización de la solución. Se espera una acción coordinada e iterativa. Su propuesta debe considerar que existirá integración con aplicaciones que están fuera del alcance de la contratación y existentes en la Institución.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.9 Soporte y Mantenimiento

El proveedor deberá indicar el mecanismo para realizar el mantenimiento correctivo y preventivo, incluyendo los ajustes que se deriven de los cambios normativos con plazos máximos de implementación, de todos los componentes tecnológicos, herramientas y procesos de gestión de datos de la solución, así como de las herramientas para su administración, operación y monitoreo, para garantizar y cumplir los niveles de servicio.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.4.10 Licenciamiento o Suscripción

En el caso que aplique un licenciamiento, en lo referente a los elementos licenciados y entregados por el proveedor, se deben entregar con una garantía mínima de tres años a partir de la entrega, cubriendo el servicio de soporte y mantenimiento preventivo y evolutivo, así como el derecho a las actualizaciones del software licenciado, durante el mismo periodo de tres años. Así mismo, los elementos proporcionados por el proveedor deberán contar con un certificado de EOL (End of Life) expedido por el fabricante, no

inferior a cinco (5) años para la versión entregada.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5 Atributos de calidad de la solución

3.5.1 Disponibilidad

Se refiere a la característica o propiedad, que tiene un software, de estar disponible cuando se lo requiere, especialmente luego de que ocurre una falla. La falla debe estar reconocida o prevenida de tal manera que el sistema pueda responder de alguna manera. La respuesta que se desea obtener dependerá de la criticidad de la aplicación (módulo, componente, funcionalidad) y del tipo de falla. La respuesta debe estar entre las opciones de "ignorar el error" hasta "mantenerse funcionando como si nada hubiera pasado"

Las tácticas para la disponibilidad se categorizan entre:

- Detectar fallas
- Recuperarse de las fallas
- Prevenir fallas

Las tácticas de detección dependen básicamente de la detección de señales de funcionamiento (signos de vida, transacciones de prueba) de varios componentes. Las tácticas de recuperación son una combinación del reintento de una operación (transacción) o el mantener datos o servicios redundantes. Las tácticas de prevención incluyen la eliminación de puntos únicos de fallo de un servicio y la utilización de mecanismos para limitar el alcance de las fallas.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.2 Interoperabilidad

Se refiere a la propiedad o característica que tiene un componente, de intercambiar información de manera útil; su objetivo es el de intercambiar información o rechazar la petición de intercambio de información. Estos componentes deben ser construidos con la intención de:

- Intercambiar información a sistemas existentes.
- Intercambiar información a sistemas futuros identificados
- Proveer servicios generales incluso sin tener el conocimiento de los detalles de los sistemas que van a utilizar dichos servicios.

Alcanzar una interoperabilidad entre componentes involucra que éstos se puedan localizar entre ellos y posteriormente administrar las interfaces respectivas de tal manera que puedan intercambiar información.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.3 Mantenibilidad

Se refiere a la propiedad o característica de un componente, que interfiere directamente con el tiempo o dinero que representa hacer un cambio, así como el grado en el que dicho cambio afecta a otros componentes o atributos de calidad.

Los cambios a un componente pueden ser realizados por desarrolladores, instaladores, usuarios finales, por lo que debe haber una preparación para realizar dichos cambios. Existe un costo en la preparación y en la ejecución de un cambio. Las tácticas de mantenibilidad se diseñan de tal manera que permitan preparar los cambios posteriores.

Las tácticas para reducir el costo de ejecutar un cambio pueden incluir el desarrollar componentes más pequeños, incrementar la cohesión y reducir el acoplamiento. Diferir la dependencia entre componentes lo más tarde que sea posible también es una táctica para reducir el costo de ejecutar un cambio.

La reducción del acoplamiento entre y en los componentes es un estándar bastante conocido y general, mismo que incluye una serie de patrones o buenas prácticas a considerar como el encapsulamiento, el uso de intermediarios, restringir las dependencias, colocar cercanamente las responsabilidades relacionadas, refactorizar y abstraer los servicios comunes.

El incremento de la cohesión es otro estándar que tiene por objetivo el separar las responsabilidades que tienen cada uno de los componentes, de tal manera que dos o más componentes no trabajen sobre un mismo propósito.

Diferir la dependencia entre componentes es una categoría de tácticas que afectan directamente al tiempo de construcción, tiempo de carga, tiempo de inicialización y tiempo de ejecución.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.4 Rendimiento

Se trata de la administración de los recursos de un componente respecto a determinados tipos de demanda que permiten alcanzar un comportamiento aceptable con respecto al tiempo.

El rendimiento puede ser medido en términos de “throughput” y latencia, tanto para sistemas transaccionales como sistemas embebidos de tiempo real.

El rendimiento se puede mejorar reduciendo la demanda o administrando los recursos de una manera más apropiada, sin embargo, se debe tomar en cuenta que reducir la demanda podría reducir la fidelidad y la respuesta a algunas peticiones de servicio. El manejo apropiado de recursos podría ser alcanzado mediante la calendarización, replica o generalmente incrementando la cantidad de recursos disponibles.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.5 Seguridad

Los ataques que se realizan en contra de un componente o sistema son considerados ataques sobre la confidencialidad, integridad o disponibilidad de dicho componente o sistema y sus datos.

La confidencialidad es una característica que permite aislar la información de recursos que no deben tener acceso, mientras que otorga el acceso para aquellos recursos que están permitidos. La integridad tiene que ver con la capacidad de no tener modificaciones o eliminaciones no autorizadas de datos, mientras que la disponibilidad tiene que ver con la capacidad de que un sistema esté disponible para las entidades autorizadas.

El alcanzar la seguridad se relaciona con la definición de las diferentes clases de actores. La identificación, autenticación y autorización de los actores son tácticas que permiten saber qué usuarios o sistemas tienen derechos a qué clase de acceso sobre un sistema.

Se debe asumir que ninguna táctica de seguridad es a toda prueba y que el sistema siempre está comprometido, de ahí que las tácticas existen para limitar el alcance de un ataque, o para reaccionar y recuperarse de un ataque.

Recuperarse de un ataque incluye las tácticas utilizadas en el atributo de calidad de disponibilidad, y en general, incluyen el retorno del sistema a un estado consistente previo al ataque sufrido.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.6 Verificabilidad

Es la capacidad o característica que posee un componente (sistema) y que indica cuán verificable es éste. Asegura que un componente sea verificable tiene recompensas tanto en el costo de las pruebas como en la confiabilidad de este.

Para saber si un componente es verificable existen varios instrumentos, mismos que son sistemas que encapsulan recursos para las pruebas, como por ejemplo casos e infraestructura de pruebas que facilitan las pruebas de regresión a lo largo de cada iteración. Otro instrumento que permite incrementar la capacidad de un sistema de ser verificable es el uso de casos de pruebas (ATDD y TDD) previo al desarrollo de un componente.

El controlar y observar el estado de un componente es la principal táctica de verificación de un componente. Para poder controlar y observar el estado de un componente se puede asegurar que el componente tenga la habilidad de: hacer inyección de fallas (errores), registrar el estado de un sistema en puntos claves del mismo, aislar el sistema de su ambiente, abstraer varios recursos.

Los sistemas complejos son difíciles de probar debido a su tamaño en la mayoría de las ocasiones y a las interconexiones que existen entre los componentes del sistema, por esta razón diseñar y construir un componente de la manera más simple posible es otra táctica que apoya a la verificabilidad.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

3.5.7 Usabilidad

Es la propiedad o característica que representa cuán fácil es para el usuario cumplir con una tarea específica y el tipo de ayuda que el sistema provee al usuario. El enfocarse en la calidad del aplicativo permite tener una de las maneras más fáciles y baratas de mejorar la calidad de un sistema, al menos desde el punto de vista de los usuarios. La usabilidad comprende:

- El aprendizaje de las características del sistema
- Usar el sistema de manera eficiente
- Minimizar el impacto de los errores
- Adaptar el sistema a las necesidades del usuario
- Incrementar la confianza y satisfacción

Existe una relación bastante fuerte entre la usabilidad, específicamente el proceso de diseño de la interfaz de usuario y el atributo de calidad de mantenibilidad. Esta relación se promueve mediante el uso de patrones que fuercen la separación de responsabilidades asociadas a la interfaz de usuario y el resto del sistema, por ejemplo, el patrón modelo-vista-controlador.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

4 CAPACITACIÓN Y TRANSFERENCIA DE CONOCIMIENTOS

El fabricante, partner y/o proveedor deberá contar con estrategias orientadas a la capacitación técnica y funcional de la Plataforma DataFabric en coordinación con el equipo del SRI, para ello se establecerá un “Acuerdo de Transferencia de Know-How Técnico” y un “Acuerdo de Transferencia de Know-How Funcional” orientado a facilitar la transferencia de conocimiento.

El entrenamiento técnico y funcional deberá realizarse considerando enfoques metodológicos en marcos de referencia como e-learning u otras que estén internacionalmente reconocidas, con el objetivo de garantizar se desarrollen las competencias necesarias en el SRI para dar soporte y mantener la solución tecnológica, además, deberá garantizar la migración de los flujos de trabajo de los proyectos de minería de datos actuales desarrollados en la Institución, con el acompañamiento de técnicos certificados por el fabricante, partner y/o proveedor.

Se espera la entrega de manuales técnicos, memorias técnicas, y transferencia de conocimiento en idioma español.

Cumplimiento (ON PREMISE) (F, P, N)	Cumplimiento (CLOUD) (F, P, N)	Descripción de la solución y comentarios

INSTRUCCIONES:

Las firmas participantes en este RFI deberán señalar si su solución tecnológica cumple en forma total (T), parcial (P) o ninguna (N) con las funcionalidades y alcances señalados para cada producto señalado en los apartados 5, 6, 7 y 8. Para lo anterior deberá llenar el formulario **anexo 1** al presente RFI con esa respuesta, señalando adicionalmente el porcentaje estimado de cubrimiento en caso de seleccionar la opción (P).

5 PREGUNTAS ESPECÍFICAS

Instrucciones: Favor de responder cada una de las preguntas incluyendo toda la información relevante para que SRI cuente con los elementos necesarios para su valoración. En su caso, podrá adjuntar información complementaria (en formato PDF).

5.1 Experiencia

1. ¿Ha realizado proyectos en implementación de plataformas tipo Data Fabric en otras Administraciones Tributarias, entidades gubernamentales o sociedades privadas? Describa los proyectos mencionando los aspectos más relevantes incluido el tiempo de implementación.
2. ¿Ha brindado asesoría, consultoría o implementación en temas relacionados con implementación del Modelo de Gobierno de Información?
3. ¿Ha participado en proyectos de consultoría, asesoría o implementación cuyo financiamiento provenga de parte de organismos multilaterales? Indicar por favor, un sumario del o los principales proyectos.
4. Indique por favor, si su compañía aplicará para este proyecto a través de Casa Matriz, subsidiaria o filial del grupo empresarial.
5. Sírvase indicar si su compañía posee representantes, subsidiarias o filiales en Latinoamérica y/o Ecuador.
6. Indique por favor algún otro elemento que considere relevante en relación con la experiencia de la compañía.

5.2 Servicio

1. Indicar la capacidad de la firma para organizar equipos de trabajo de alto desempeño y calificaciones para llevar a cabo proyectos a gran escala, comente sobre los equipos que logró conformar.
2. Debido que el diseño de la institución requiere conformidad con la normativa legal de Ecuador. ¿Cuál sería su estrategia para asegurar el cumplimiento de los requerimientos legales del diseño de la institución?

3. Mencionar las estrategias en fases o etapas para la implementación de la plataforma DataFabric en el Servicio de Rentas Internas
4. ¿Qué factores considera claves para que la implementación de una plataforma DataFabric y la adopción del Modelo de Gobierno de Información cumpla con los objetivos planteados? Sírvase indicar una breve descripción de cada uno.
5. Describa el modelo de negocio: licenciamiento / servicio bajo solución (si existen componentes bajo este modelo).
6. Describa el modelo de soporte y mantenimiento que provee, preventivo, correctivo y evolutivo.
7. Describa si su solución cuenta con código fuente propietario.
8. Describa el modelo de transferencia de conocimiento que se provee para su producto.

5.3 Equipo y recursos

1. ¿Cuál sería la estructura del equipo de trabajo para cada una de las etapas?
2. ¿Cuáles serían los perfiles requeridos dentro de la estructura con el respectivo número de participantes? Indicar una descripción general de los roles y perfiles requeridos (años de experiencia, formación académica, etc.) y si su participación es de tiempo completo o parcial.

5.4 Información de la solución tecnológica

1. ¿Es el fabricante de los componentes que presenta?
2. Indique el nombre de las soluciones y página web descriptiva de la solución.
3. Descripción detallada que incluya los siguientes aspectos:
 - Plataforma DataFabric; documento sobre la solución que describa el o los productos, sus principales atributos operativos y los componentes tecnológicos que utiliza para cumplir con los requerimientos del numeral 3.1
4. Descripción de la arquitectura general, de negocio, de aplicaciones, de datos, de interoperabilidad, de infraestructura, de seguridad.
5. Tipo de solución:
 - Software comercial (COTS - Commercial Off-the-Shelf)
 - Software como servicio (SaaS - Software as a Service)

- Software de código abierto (OPEN)
- Plataforma como servicio (PaaS – Platform as a Service)
- Infraestructura como servicio (IaaS)
- Software implementado sobre plataforma de gestión de procesos de negocio (BPMS - Business Process Management Suite)
- Software a la medida (LDSW - Locally Developed Software)

COTS	SaaS	OPEN	IaaS	PaaS	BPMS	LDSW

Nota: En caso de que sea una solución híbrida (mixta), seleccionar las opciones que correspondan en la tabla.

6. Idioma de la solución y documentación.
7. Por favor indique la lista completa de componentes de la solución y describa si es software propio del interesado o se integra de un tercero y el tipo de licenciamiento ofrecido.

Nombre del Componente	Descripción general del componente	Propiedad del componente		Tipo de solución	Esquema de actualización del componente
		Propio del Interesado	De un tercero		

8. Describa el modelo de licenciamiento bajo solución de código abierto (si existen componentes bajo este modelo).
9. Indique las características técnicas y/o funcionales complementarias de la solución y que el SRI aún no ha contemplado dentro de los requerimientos formulados.
10. ¿Cuál es la metodología de soporte de la solución y los correspondientes ANS?
11. Capacidad de adaptación/personalización de la solución.
12. Dependiendo de la solución ofrecida cómo integra las adaptaciones/personalizaciones del software a su producto estándar?

13. ¿Su solución está diseñada para ser desplegada en la nube? ¿Con qué estrategia de aprovisionamiento?
14. En su caso, describa en cuáles servicios de nube pública puede operar su solución.
15. Describa los tipos de modelos de despliegue para cada componente de la solución:

Componente	Descripción	Modelo (PaaS, IaaS, SaaS)

16. Describa la estrategia de migración y/o reingeniería de flujos de datos incluyendo aspectos físicos como: herramientas, procesos de automatización, arquitectura general, recursos involucrados.
17. ¿Qué mecanismos utilizará para asegurar la calidad de los datos virtualizados?
18. ¿Considera que hay algún aspecto que deba considerar el SRI de manera adicional a los requerimientos planteados?
19. Describa el modelo gestión de datos que propone incluyendo como mínimo procesos, procedimientos, metodologías, recursos.
20. ¿Qué metodología recomendaría para gestionar grandes volúmenes de datos de diversas y disímiles fuentes?
21. ¿Cómo priorizaría entre los diferentes requerimientos de gestión de datos?

5.5 Estimación de la Inversión

1. Describa el modelo de facturación y pago de la solución incluyendo los diferentes componentes por cada tipo de implementación aplicado (ON PREMISE / CLOUD)
2. Está dispuesto a aceptar otro modelo de facturación y pago (Si/No cuál?)
3. Con respecto al licenciamiento que aplique para la solución en los diferentes componentes:
 - Ofrece descuentos en licenciamiento por volumen (Si/No)
 - Ofrece licenciamiento ilimitado para algún tipo de usuario (Si/No)
 - Cuenta de depósito del código fuente "escrow account" (Si/No)

4. Costo Total. Diligencia y desglose en el siguiente cuadro:

1. Para soluciones ofrecidas bajo el modelo On premise (IaaS, PaaS), sírvase indicar un estimado del costo neto total (sin impuestos) de la solución tecnológica, considerando todos los elementos necesarios para su funcionamiento, conforme al cuadro siguiente:

N.º	Rubro / servicio	Costo en el proyecto (USD)	Renovación anual * (USD)
1	Implementación de la Plataforma		
2	Licenciamiento de componentes de Plataforma (describir)		
3	Ambientes de pruebas y preproducción		
4	Migración de datos y/o reingeniería de flujos de datos		
5	Licenciamiento de herramientas de migración (ETL o virtualización)		
6	Servicios de gestión de datos		
7	Capacitación y transferencia del conocimiento		
8	Soporte y mantenimiento preventivo		
9	Otros costos (describir)		
	Total		

**La renovación se considera posterior a que ha finalizado el contrato para el proyecto.*

2. Para soluciones ofrecidas bajo el modelo de implementación NUBE (modelos IaaS, PaaS o SaaS), sírvase indicar un estimado del costo neto total (sin impuestos) de la solución tecnológica, considerando todos los elementos necesarios para su funcionamiento, conforme al cuadro siguiente

N.º	Rubro / servicio	Costo en el proyecto (USD)	Renovación anual * (USD)
1	Implementación de la Plataforma		
2	Licenciamiento de componentes de Plataforma (describir)		
3	Suscripción de componentes de plataforma (describir)		
4	Ambientes de pruebas y preproducción		
5	Servicios de nube (Ambiente de producción)		
6	Migración de datos y/o reingeniería de flujos de datos		
7	Licenciamiento de herramientas de migración (ETL o virtualización)		
8	Suscripción de herramientas de migración (ETL o virtualización)		
9	Servicios de gestión de datos		
10	Capacitación y transferencia del conocimiento		
11	Soporte y mantenimiento preventivo		
12	Otros costos (describir)		
	Total		

**La renovación se considera posterior a que ha finalizado el contrato para el proyecto.*

6 INFORMACIÓN REQUERIDA POR EL SRI

Información acerca del proveedor y/o representante:

INFORMACIÓN GENERAL	
Nombre de la compañía:	
Número de Identificación Tributaria:	
Dirección / Medios de contacto / Ciudad / País:	
Actividad económica principal:	
Número de empleados:	
Nombre Representante Legal:	
Cargo Representante Legal:	
Nombre Persona Contacto:	
Medios de contacto:	
Tipo de empresa (pública, privada, mixta):	
¿Es el fabricante de los componentes que presenta? (SI / NO):	
¿Tiene oficinas de representación en Ecuador? (SI / NO):	
Número de empleados en Ecuador:	
Describa bajo cuál figura jurídica se presentaría a un eventual proceso de contratación en Ecuador (directamente, unión temporal, consorcio, subcontratista, otra):	

7 CONDICIONES GENERALES

7.1 Inscripción

Los proveedores interesados en presentar la información solicitada en este documento deben informarlo enviando un correo electrónico a las direcciones enunciadas en el párrafo “Forma de presentación y contactos SRI”, manifestando su intención de ser parte de este análisis.

El presente documento no obliga contractualmente a ninguna de las partes con la información que se solicita y se suministra, dado que corresponde a un proceso de investigación de mercado.

7.2 Fecha de entrega de la información solicitada por el SRI

La fecha límite para la entrega de la información solicitada, será de 10 días hábiles a partir de la publicación del RFI en los portales habilitados; no obstante, la información puede ser entregada antes de esta fecha. La entrega se realizará únicamente mediante correo electrónico.

7.3 Fechas para la presentación y demos de la solución tecnológica propuesta

Posterior a la fecha de entrega de las respuestas del RFI, el Servicio de Rentas Internas podrá solicitar las aclaraciones, demos o presentaciones que a su juicio se requieran, para lo cual se remitirá por correo electrónico las agendas y convocatorias correspondientes.

8 ANEXOS

Anexo 1. Formato de respuestas para requerimientos

INSTRUCCIONES:

Las firmas participantes en este RFI deberán señalar si su solución tecnológica cumple en forma total (T), parcial (P) o ninguna (N) con las funcionalidades y alcances señalados para cada producto señalado en los apartados 3 y 4.

Para lo anterior deberá llenar el formulario **siguiente** con esa respuesta, señalando adicionalmente el porcentaje estimado de cubrimiento en caso de seleccionar la opción (P).

El cuadro deberá reflejar cada uno de los elementos señalados en el numeral 3.1 conforme a la siguiente estructura:

Tabla 1: Matriz de cumplimiento

Módulo	Producto	Descripción	Cumple On Premise (T, P, N)	Cumple Nube (T, P, N)	Características no contempladas
3.1.1	RG-1				
3.1.1	RG-2				
3.1.1	RG-3				
3.1.2	V-1				
3.1.2	V-2				
3.1.2	V-3				

El cuadro deberá reflejar cada uno de los elementos señalados en el numeral 3.4 conforme a la siguiente estructura:

Nombre	Cumple On Premise (T, P, N)	Cumple Nube (T, P, N)	Características no contempladas
Tecnológicos			
Monitoreo			

El cuadro deberá reflejar cada uno de los elementos señalados en el numeral 3.5 conforme a la siguiente estructura:

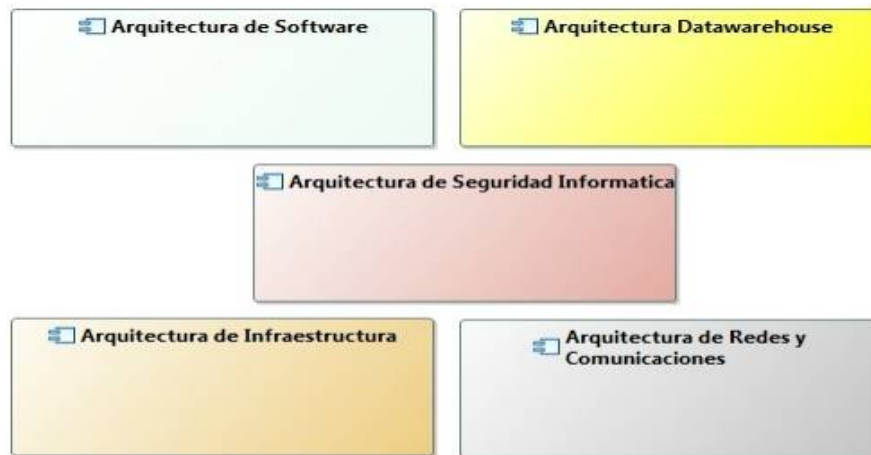
Nombre	Cumple On premise (T, P, N)	Cumple Nube (T, P, N)	Características no contempladas
Disponibilidad			
Interoperabilidad			
Mantenibilidad			

Anexo 2. Situación tecnológica actual de la Institución

Arquitectura:

La arquitectura tecnológica institucional actual está conformada por los siguientes componentes macro según su ámbito de aplicación los componentes son los siguientes:

Gráfico 4: Componentes macro de la arquitectura



Fuente: Arquitectura Tecnológica General - Servicio de Rentas Internas

Fecha: Enero 2019

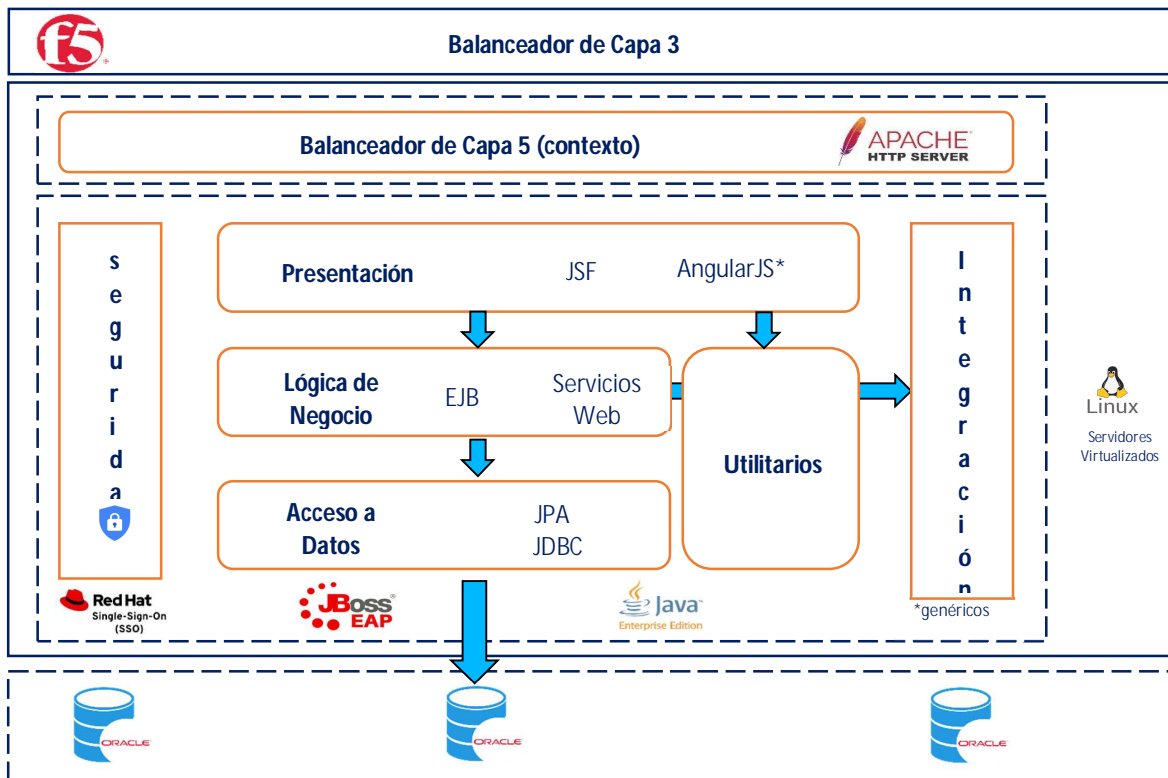
1. *Arquitectura de Software*: Describe la arquitectura que deben cumplir los sistemas de software en la institución desde el ámbito de tecnologías y estándares para la construcción y despliegue de software transaccional.
2. *Arquitectura de Datawarehouse*: Describe la arquitectura que deben cumplir los sistemas de software en la institución desde el ámbito de tecnologías y estándares para la construcción de software de análisis de datos.
3. *Arquitectura de Seguridad Informática*: Describe la arquitectura para aseguramiento de los sistemas de la institución desde el ámbito de tecnologías, políticas, estándares y procedimientos para asegurar los sistemas informáticos.

4. *Arquitectura de Infraestructura:* Describe la arquitectura que requieren los sistemas de software en la institución desde el ámbito de tecnologías de hardware y software base para la operación.
5. *Arquitectura de Redes y Comunicaciones:* Describe la arquitectura para la comunicación de los sistemas de software de la institución desde el ámbito de tecnologías para la especificación de los componentes físicos de una red.

Arquitectura de software:

A continuación, se describe los componentes macro de la arquitectura de software de los sistemas institucionales:

Gráfico 5: Componentes arquitectura de software



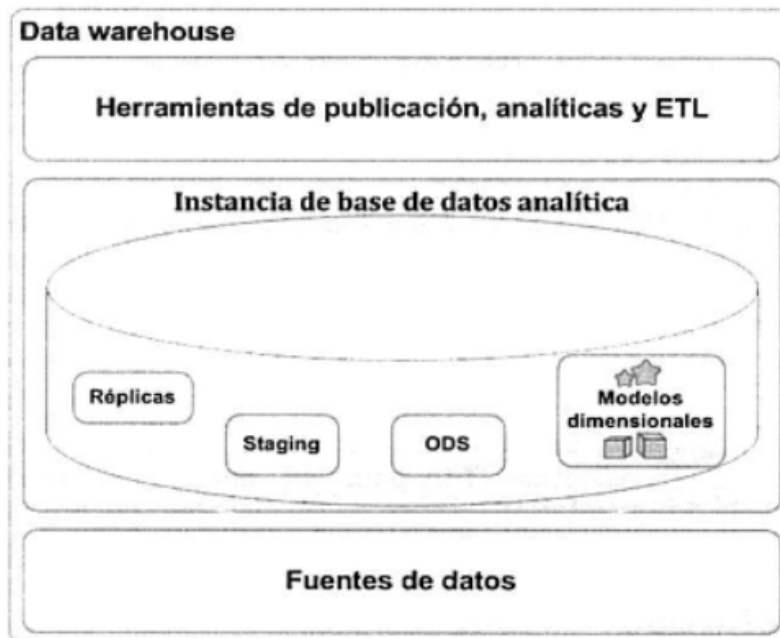
Fuente: Informe Arquitectura Resiliente - Servicio de Rentas Internas

Fecha: Marzo 2023

Arquitectura de Datawarehouse:

En el siguiente diagrama se detallan los componentes macro de la arquitectura de Datawarehouse:

Gráfico 6: Arquitectura Datawarehouse macro



Fuente: Arquitectura Tecnológica Datawarehouse - Servicio de Rentas Internas

Fecha: marzo 2023

Arquitectura de Seguridad Informática:

La Arquitectura de Seguridad Informática está diseñada en función de los ámbitos definidos por los requerimientos de seguridad informática de la Institución. A continuación, se describen los componentes de seguridad informática implementados en cada ámbito.

Gráfico 7: Componentes de Seguridad Informática por Ámbito



Fuente: Arquitectura Tecnológica General Seguridad Informática - Servicio de Rentas Internas

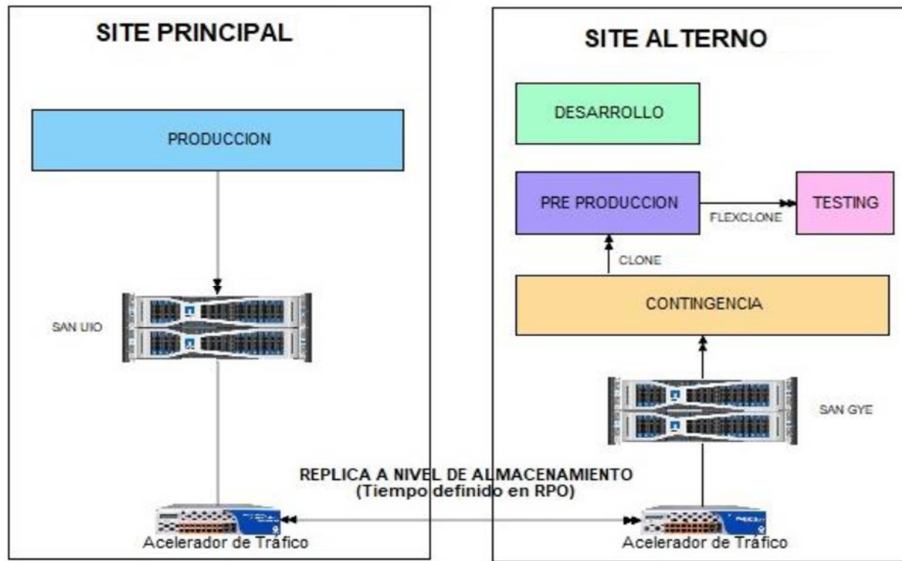
Fecha: 03 de diciembre de 2020

Arquitectura de Infraestructura:

La arquitectura de infraestructura define el hardware, software base, métodos de acceso y protocolos usados a través de los sistemas.

La Institución posee dos centros de datos en el diagrama siguiente se presenta de manera macro el esquema utilizado para la definición de ambientes en los dos Data Center:

Gráfico 8: Ambientes de los Centros de Datos



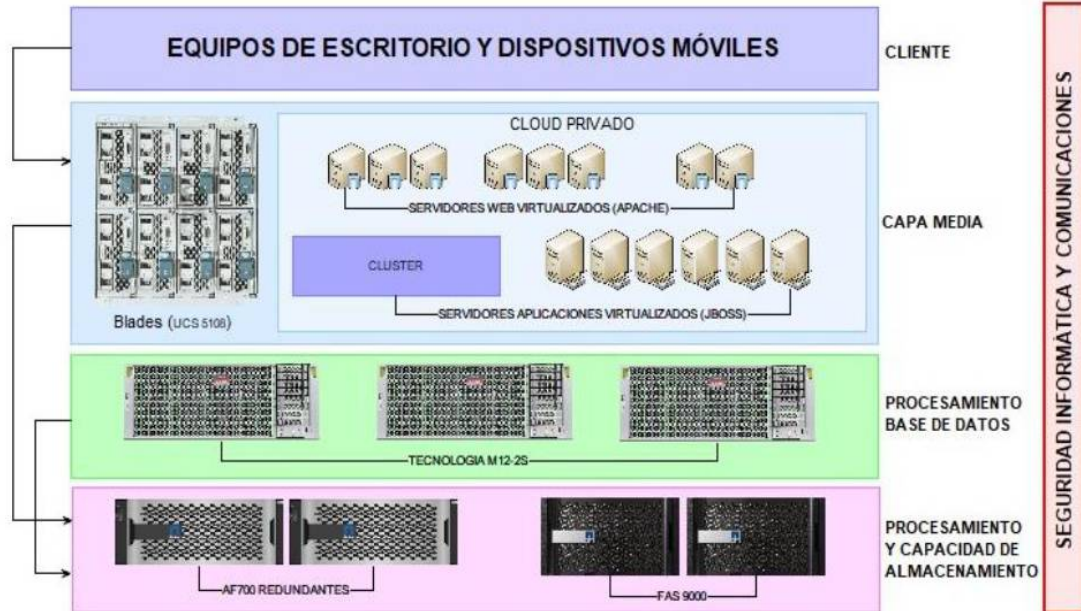
Fuente: Arquitectura Tecnológica General de Infraestructura - Servicio de Rentas Internas

Fecha: 15 de septiembre de 2020

En el centro de datos principal se encuentra toda la infraestructura necesaria para soportar al ambiente de producción de las aplicaciones de la institución, mientras que en el centro de datos alternativo se soportan los ambientes de: contingencia, preproducción, testing, desarrollo y ciertos servicios complementarios de producción.

En el siguiente diagrama se muestran los componentes de la arquitectura de infraestructura y su interacción:

Gráfico 9: Comunicación entre componentes



Fuente: Arquitectura Tecnológica General de Infraestructura - Servicio de Rentas Internas

Fecha: 15 de septiembre de 2020

- La capa media es el componente que soporta las aplicaciones y herramientas, a nivel de hardware (chasis Blade) está formado de varios servidores (cuchillas del chasis) que poseen memoria, disco, CPU, tarjeta gráfica y tarjeta de red. De manera general se cuenta con varios servidores Linux virtualizados que soportan la capa media de aplicaciones conformada por servidores de aplicación Red Hat JBoss Enterprise Application Platform (JBoss EAP) y servidores web Apache; por otro lado, existen servidores Windows virtualizados destinados para herramientas propietarias de gestión de servicios soportan servicios como Active Directory, servidor de correo, servidor de respaldos, servidor FTP, antivirus, servidores de monitoreo de la infraestructura y administración, entre otros.
- El procesamiento base de datos corresponde a la infraestructura y software que soportan los DBMS y datos institucionales. Las bases de datos altamente transaccionales o considerados CORE, están instaladas en servidores con las

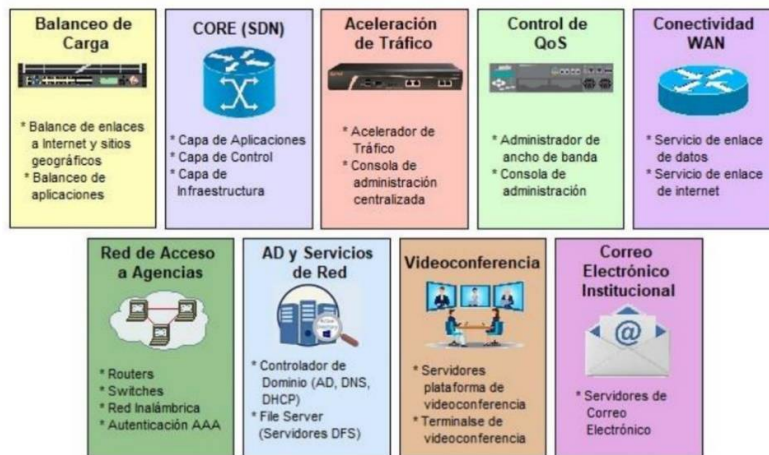
siguientes consideraciones:

- o Características RAS (Reliability, Availability and Serviceability). Los servidores de base de datos poseen características RAS highend (disponibilidad de 99,999%) y de acuerdo con la necesidad las principales bases de datos se han configurado en RAC (Oracle Real ApplicationClusters).
- o Capacidad de particionamiento y/o virtualización.
- o Software base. Los productos implementados en la arquitectura a nivel del motor de base de datos son: Oracle Database Enterprise Edition, Real Application Clúster (RAC), Diagnostics Pack, Partition, Tuning Pack
- Capacidad de almacenamiento corresponde a un conjunto de dispositivos de almacenamiento en los cuales se encuentra almacenada la información estructurada y no estructurada de la institución. Estos dispositivos de almacenamiento pueden ser: storage, bóveda de respaldos y cintas magnéticas.

Arquitectura de redes y comunicaciones:

En el siguiente diagrama se muestran los componentes de la arquitectura de comunicaciones:

Gráfico 10: Arquitectura de Comunicaciones (Ámbito y Componentes)



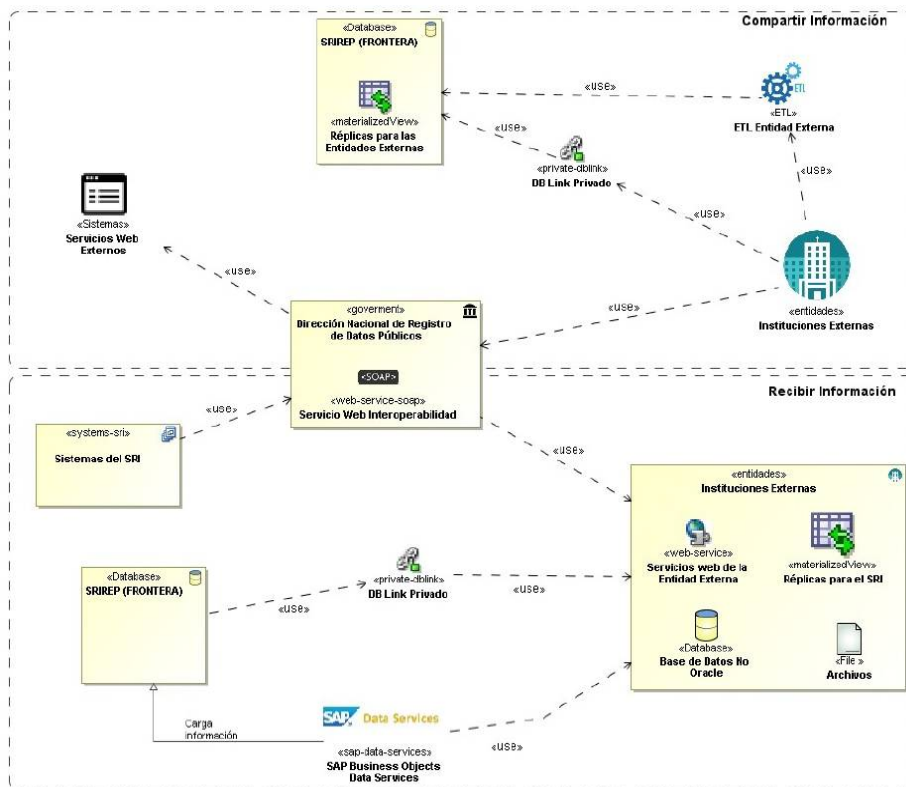
Fuente: Arquitectura Tecnológica General de Redes y Comunicaciones - Servicio de Rentas Internas

Fecha: 30 de noviembre de 2020

Esquema de integración con terceros

Actualmente la interoperabilidad se implementa a través del uso de diferentes estrategias de integración, pudiendo ser a través de: réplicas de datos, servicios web administrados por la DINARDAP y/o archivos (txt, csv, excel), cuyas principales interacciones externas se pueden ver en el siguiente diagrama de componentes:

Gráfico 11: Componentes interacciones externas



Fuente: Arquitectura Tecnológica de la Solución. Convenios Interinstitucionales de Intercambio de Información
 - Servicio de Rentas Internas
 Fecha: 28 de junio de 2019

Versiones de BDD:

Oracle Enterprise Edition versiones:

- 19.16
- 12.2
- 11.2

Anexo 3. Datos estadísticos institucionales relevantes

Número de contribuyentes inscritos:

Tabla 2: Total contribuyentes activos

TIPO DE SUJETO	TOTAL
PERSONAS NATURALES	2.195.896
SOCIEDADES	226.157
TOTAL (*)	2.422.053

Fuente: Base Catastro - Servicio de Rentas Internas

Fecha: 22 de marzo de 2023

Número de declaraciones recibidas:

Tabla 3: Número de declaraciones recibidas por año

Mes Presentación	Año		
	2021	2022	2023
Enero	2.186.542	2.587.271	1.659.418
Febrero	1.602.105	1.408.455	1.063.299
Marzo	2.084.182	1.823.189	2.284.222
Abril	1.439.306	1.296.259	
Mayo	1.265.240	1.072.928	
Junio	1.272.581	1.007.551	
Julio	2.060.794	1.490.565	
Agosto	1.324.717	1.067.342	
Septiembre	1.299.943	1.085.700	
Octubre	1.490.675	972.420	
Noviembre	1.339.192	985.443	
Diciembre	1.321.670	960.771	
TOTAL	18.688.968	15.759.916	5.000.939

Fuente: Base Entrega y Recepción - Servicio de Rentas Internas

Fecha: 31 de marzo de 2023

Número de comprobantes electrónicos emitidos:

Tabla 4: Número de comprobantes electrónicos emitidos por año

Mes	Año		
	2021	2022	2023
Enero	185.872.502	208.074.669	246.533.567
Febrero	179.711.046	198.140.920	227.479.913
Marzo	201.837.901	223.322.297	237.738.291
Abril	187.793.426	222.238.536	
Mayo	185.033.931	231.357.234	
Junio	197.360.902	207.962.584	
Julio	207.663.094	231.624.262	
Agosto	210.306.258	238.069.046	
Septiembre	207.673.180	232.508.199	
Octubre	214.270.625	237.825.976	
Noviembre	211.470.630	235.788.676	
Diciembre	235.999.456	271.094.114	
TOTAL	2.424.994.972	2.738.008.535	711.751.771

Fuente: Base Facturación - Servicio de Rentas Internas

Fecha: 31 de marzo de 2023

Potenciales usuarios de la Plataforma tipo DataFabric por capa:

Tabla 5: Usuarios internos a nivel nacional

DIRECCIÓN	Usuarios Capa de Gobierno	Usuarios Capa Explotación	Usuarios Capa de Integración	Usuarios Capa de Consumo	Total Usuarios
DIRECCIÓN NACIONAL	30	40	10	660	740
ZONA 1	-	-	-	115	115
ZONA 2	-	-	-	60	60
ZONA 3	-	2	-	260	262
ZONA 4	-	-	-	195	195
ZONA 5	-	-	-	80	80
ZONA 6	-	-	-	150	150
ZONA 7	-	-	-	165	165
ZONA 8	-	2	-	405	407
ZONA 9	-	2	-	480	482
TOTAL	30	46	10	2.570	2.656

Fuente: Distributivo - Servicio de Rentas Internas

Fecha: 31 de marzo de 2023

Volúmenes referenciales:

Información Transaccional *	652.95 TB
Bóveda de respaldos	691.23 TB
BDD - SRIDW	23.09 TB
BDD - STAR	52.29 TB

* Incluye volumen de información de comprobantes electrónicos (CEL Y CELPAS)

Anexo 4. Glosario de términos

- Anonimización: Técnica de tratamiento de datos que elimina o modifica los datos personales identificables para obtener datos anónimos que no se pueden asociar con ninguna persona.
- Virtualizador: programa para incrustar un sistema operativo abstracto en otro. Las particiones, la memoria y el control de entrada y salida de los datos está en una máquina virtual diseñada.
- Dashboards: representaciones visuales de la información más importante que se necesita para lograr uno o más objetivos, consolidada y dispuesta en una sola pantalla para que la información se pueda controlar de un vistazo.
- Datastore: una base de datos de documentos NoSQL creada a fin de proporcionar ajuste de escala automático, alto rendimiento y facilidad para el desarrollo de aplicaciones.
- Hashear: Generar la aplicación de un algoritmo matemático que transforma el conjunto de datos de entrada en una expresión alfanumérica que tiene una longitud predeterminada, ofrece más seguridad que el cifrado, ya que los valores hash no pueden volver a convertirse en valores originales sin una clave.
- RandomSplit: Función que se utiliza para dividir aleatoriamente un conjunto de datos (dataset o dataframe) en dos o más subconjuntos con una proporción específica.
- Stratified Sampling: es una forma de representación estadística que muestra cómo se comporta una característica o variable en una población a través de hacer evidente el cambio de dicha variable en subpoblaciones o estratos en los que se ha dividido. Consiste en la división previa de la población de estudio en grupos o clases que se suponen homogéneos respecto a característica a estudiar y que no se solapan.