

BOLETIN DE RESPUESTAS No. 2

PROCESO: REQUEST FOR INFORMATION – RFI

OBJETO DEL PROCESO: GOBIERNO DE DATOS Y ANALÍTICA INSTITUCIONAL
(PLATAFORMA TIPO DATA FABRIC)

CÓDIGO DEL PROCESO: RFI No. 003-2023

FECHA: 29 de mayo de 2023

No.	Pregunta	Respuesta
1	¿Qué certificaciones deberá cumplir el personal técnico para participar en el proyecto?	Forma parte de la propuesta requerida.
2	¿Qué experiencias deberá cumplir la empresa para participar en el proyecto?	5.1 Experiencia
3	Para los costos de implementación, ¿es posible añadir un pago por anticipo del servicio?	Sujeto a análisis, las condiciones definitivas se establecerán en el TDR
4	Para el equipamiento que se use, ¿se necesita adjuntar datasheets o brochures técnicos de hardware y software?	Si
5	Tomando en cuenta que el servicio se debe dar en Territorio Ecuatoriano, dentro de la experiencia de la empresa ¿podrá ser válido aquellos contratos de CLOUD, ANALISIS DE INFORMACION DE VIDEO o DATOS; ¿Siendo el oferente propietario del Datacenter, Seguridades y Comunicaciones?	Sujeto a análisis.
6	¿Aportaría en garantizar el servicio las siguientes certificaciones empresariales? a. ISO 27001 por seguridad de la información de punto a punto b. ISO 22301 por continuidad del negocio o servicio, garantizando la operatividad c. ISO 9001 para garantizar la calidad del servicio d. Certificación del Data Center con certificación TIER IV emitido por una certificadora internacional como el UPTIME INSTITUTE.	Si
7	¿Cuáles son los marcos o frameworks válidos de referencia internacional que necesitan para realizar la transferencia de conocimiento técnico y funcional?	No se restringe a ningún marco o framework, forma parte de la propuesta requerida
8	¿Es posible presentar una respuesta usando la figura de consorcio de varios proveedores en donde: a. Se responde al RFI y al RFP con la sumatoria de las soluciones tecnológicas de cada miembro del consorcio b. Se responde al RFI y al RFP con la sumatoria de experiencia, curriculums, proyectos exitosos, asesoría, consultoría, etc. de cada miembro del consorcio.	Forma parte de la propuesta requerida.
9	¿Para el caso de conexión de APIs de cualquier entidad externa al SRI, el medio de comunicación será por Internet, enlace de Datos dedicados o Red de Datos Gubernamental?	Externo público - anillo gubernamental Externo privado - Enlace dedicado
10	¿Cuántas y Cuales son las políticas de gobernanza del SRI mencionadas en el párrafo siguiente? "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	Se estima levantar 1 política de gobernanza basada en el marco de referencia del DMBOOK

No.	Pregunta	Respuesta
	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”	
11	¿Cuántos y cuáles flujos de datos existen actualmente relacionados a este párrafo? “Migración y/o reingeniería de datos o flujos de datos más relevantes...”	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso. Ver Anexo 1 - Volumetría
12	¿Si la plataforma que se ofrezca por el proveedor no puede instalarse on premise, dicho proveedor queda descalificado de un futuro RFP? “...También deberá poder instalarse completamente on-premise...”	No existe descalificación de proveedores en esta etapa, las respuestas permitirán afinar los TDR
13	¿El siguiente párrafo ubicado en RG-1, descarta alianzas entre varios proveedores tecnológicos que ofrezcan su solución de manera separada: on premises y en la nube, ¿que requieran ser integradas para darle una solución completa al SRI? “...No se permiten componentes alojados en la nube indicada que enlace a módulos fuera de la nube...”	Estamos entendiendo el mercado, se reciben propuestas, Indique % de cumplimiento conforme Anexo 1
14	¿El siguiente párrafo descarta completamente alianzas o consorcios entre proveedores que requieran integrar sus sistemas, para contestar un futuro RFP? “...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...”	Estamos entendiendo el mercado, se reciben propuestas, Indique % de cumplimiento conforme Anexo 1
15	Dado que las regulaciones ecuatorianas no permiten colocar en nubes publicas datos pertenecientes al estado, confidenciales y reservados o por motivos de seguridad nacional, el SRI: a. ¿Considera que los datos que va a manejar caen en alguna de estas categorías? b. ¿Si caen en estas categorías, Uds. consideraran para un futuro RFP incluir a proveedores que solo tengan soluciones en la nube publica que alojarían sus datos fuera del país?	Sujeto a análisis hasta que se elaboren las especificaciones técnicas
16	¿Respecto a la gobernanza del proceso del RFP, se seguirán los procesos y directrices del área de supervisión y auditoría del BID?, encontrados en: https://www.iadb.org/es/acercadelbid/supervision-y-auditoria	No, por favor revisar el objetivo del RFI
17	¿Es posible extender una semana más el plazo de entrega de la respuesta al presente RFI?	Se publicará boletín de enmiendas con el nuevo cronograma actualizado.
18	1. Teniendo en cuenta la diversidad de tecnologías para gobierno de datos y analítica en general, ¿podrían por favor indicar las tecnologías de información con las que cuentan en la actualidad? Lo anterior con el fin de buscar las herramientas más acordes a la infraestructura y arquitectura actual.	Arquitectura de DW: SAP Business Objects 4.1 Pentaho Apache SuperSet Oracle Discoverer SAIKU QlikView
19	2. En cuanto a la capa unificada de datos requerida, podrían por favor indicar en un diagrama o enumerar la composición de las fuentes de datos actuales	Bases de datos Oracle (11g,12c y 19c); archivos de texto (csv, json, XML)

No.	Pregunta	Respuesta
20	3. En cuanto a "generar modelos de predicción, clasificación, grafos, redes, segmentación y minería de texto que fortalezcan las competencias" ¿podrían por favor indicar si ya cuentan con la investigación y formulación de las preguntas de negocio o el modelo en sí, o hace parte del presente alcance la investigación, búsqueda de información y formulación de modelos?	Si
21	4. ¿Podrían por favor indicar si existe un presupuesto para este proyecto o se trata de propuestas abiertas?	Se cuenta con un presupuesto referencial, el cual será afinado conforme las propuestas recibidas.
22	5. En cuanto a "que integre las fuentes de información y que provea datos consistentes a los servicios internos y externos del SRI" ¿Podrían por favor indicar cuales, cuantos y que tecnologías son los sistemas de información con los que se debe interactuar?	Referirse al gráfico 5: Componentes Arquitectura de Software, y se detallan las versiones: JDK 1.6, 1,7 y 1,8, Servidores de Aplicaciones JBOSS EAP 5, 6 y 7, Apache Web Server 2.4
23	6. En el RG-3 se indica "La Solución Software Integral propuesta deberá ser un producto de mercado ya existente y consolidado" ¿Este requisito General es de obligatorio cumplimiento? Lo anterior debido a que debe conocerse para proponer diversos componentes y no instalar algunos innecesarios.	De preferencia sí, debe ser una solución tipo Data Fabric se requiere completa e integrada
24	7. En cuanto al RG-9 Acceso a la plataforma. ¿cuentan con Directorio activo?	Si
25	8. ¿Es únicamente para usuarios internos?	Usuarios internos: Toda la plataforma Usuarios externos: Reportería
26	9. ¿el proyecto se puede realizar remotamente?	Escenario híbrido, dando prioridad a lo presencial.
27	Por favor detallar: • ¿Qué casos de uso debe proveer la funcionalidad de Gobierno de Datos, y • ¿Qué casos de uso deben ser implementados, configurados y probados dentro de la funcionalidad de Gobierno de Datos?	Los casos de uso serán definidos en el TDR, se espera que sobre los casos de uso se implementen tanto las funcionalidades de gobierno como de analítica.
28	Confirmar si el SRI actualmente tiene alguna herramienta de gobierno de datos implementada y de ser afirmativa la respuesta por favor indicar el fabricante, producto y versión empleada	No
29	Confirmar si el SRI actualmente tiene un modelo operativo de gobierno de datos definido y operativo	No
30	¿Cuántas fuentes de datos tienen gobernadas? Por favor detallar el tipo de datos que entregan tales fuentes y si son on-premise o cloud	No existe aún un modelo de Gobierno de Datos
31	¿Cuántos dominios de datos tiene el SRI?	Alrededor de 29
32	¿Cuántos activos de información tiene el SRI?	Existen 1576 conceptos de información levantados
33	Solicitamos describir las iniciativas de Gobierno de Datos que haya tenido en el pasado	No se han realizado iniciativas de gobierno anteriormente
34	¿Qué papel juegan los datos en la estrategia de negocios del SRI?	Constituyen uno de los activos más importantes de la institución pues con ellos permiten soportar la toma de decisiones.

No.	Pregunta	Respuesta
35	Solicitamos describir Objetivos del negocio ya formalizados y que existan por escrito en la organización, y que guarden relación con los datos que serán gestionados por la solución de Data Fabric que es objeto del RFI	<ol style="list-style-type: none"> Incrementar el cumplimiento voluntario de las obligaciones tributarias en la ciudadanía. Incrementar la efectividad en los procesos de control y cobro, así como los legales y demás procesos administrativos que forman parte del ciclo de la recaudación. Incrementar los conocimientos y el desarrollo de capacidades en la ciudadanía en el marco de sus deberes y derechos fiscales. Fortalecer las capacidades institucionales: <ul style="list-style-type: none"> * Incrementar la eficiencia institucional en el SRI. * Incrementar el uso eficiente del presupuesto en el SRI. * Incrementar el desarrollo del talento humano en el SRI
36	Solicitamos describir con abundancia de detalles las prioridades ante cumplimiento de regulaciones de la Ley de Protección de Datos personales.	Las especificaciones serán detalladas en los TDR, envíe su propuesta
37	A nivel de seguridad de información, ¿qué normativas o regulaciones debe de cumplir?	EGSI v2, ISO 27001, LOPDP, LOTAIP, ISO 27000, 270017, 270018 y sus actualizaciones
38	¿Cuáles son los dominios de datos que deberá administrar la solución? Por ejemplo: personas, organizaciones, productos, materiales, ubicaciones, etc. Agradecemos todo el detalle posible	Gestión de Servicios tecnológicos, Productos de Datawarehouse, Gestión de la Documentación, Devolución de Impuestos, Gestión de canales de atención, Registro del Sujeto (pasivo, beneficiario, vehículos), Declaraciones, Anexos, Vehículos, Comprobantes, Auditoría Tributaria, Gestión de Mercancías, Gestión de Infracciones, Gestión de Cobro, Recursos de Revisión, Juicios Tributarios, Recaudación, Declaraciones, Anexos, Registro del Sujeto (pasivo, beneficiario, vehículos), Fraude Fiscal, Acuerdos Internacionales de Intercambio de Información Tributaria, Gestión del Talento Humano, Indicadores de Gestión, Previsiones y Estadísticas, Seguridad de la Información, Gestión de Proyectos, Registro de catálogos transversales, Gestión de Talento Humano, Gestión Financiera, Gestión Administrativa
39	¿Qué porcentaje de trabajos se ejecutarán mediante el procesamiento de datos modificados (%) basado en CDC (change data capture)?	Ver Anexo 1 - Volumetría
40	¿Cuántos datos procesará por lotes por día (GB)?	Ver Anexo 1 - Volumetría
41	¿Cuántos datos perfilará (Data Profile) por día (GB)?	Ver Anexo 1 - Volumetría
42	¿Cuántos datos desduplicará (incremental) por día (GB)?	Ver Anexo 1 - Volumetría
43	Solicitamos detallar la cantidad de sistemas que deben ser técnicamente catalogados. En su respuesta por favor incluir incluso cuáles son las tecnologías correspondientes y la cantidad de metadatos en cada uno de estos sistemas (tablas, vistas, columnas, etc.)	Ver Anexo 1 - Volumetría
44	Solicitamos que se detalle la cantidad de objetos de negocio que deben ser catalogados (términos de Glosario, políticas, procesos, etc.).	Los casos de uso se definirán en los TD, se consideran 5 casos de uso (2 de complejidad alta, 2 complejidad media, 1 complejidad baja). Adicionalmente se considera 1 política, 1 proceso a implementar.
45	Por favor indicar si el SRI tiene actualmente algún ejemplo de modelado de datos (workflows) para de esta forma se pueda entender los casos de uso típicos	Si

No.	Pregunta	Respuesta
46	¿Qué impacto esperan tener en la organización como resultado de la implementación de MDM?, ¿Cuáles son los usos que se dará a MDM y cuáles son los beneficios que el SRI espera obtener?	Se espera elevar el nivel de madurez en la organización orientado hacia la toma de decisiones basada en datos, tener una sola versión de la verdad, facilitar la construcción de modelos de análisis, reflejándose en una mejor detección de casos, así como el incremento recaudatorio
47	¿Cuántos usuarios van a tener accesos al lago de datos?	Tabla 6
48	¿Cuántos usuarios van a tener accesos a las herramientas de gobierno?	Tabla 6
49	¿Cuántos usuarios van a tener acceso a las herramientas de modelado de casos de uso de analítica?	Tabla 6
50	¿Cuántos usuarios van a tener acceso a las herramientas de visualización de los resultados de analítica?	Tabla 6
51	¿Cuáles son los casos de uso que la herramienta de modelado de analítica debe ser capaz de manejar?	Los casos de uso se definirán en los TDR, se sugiere revisar los requerimientos de la Capa de Explotación, Machine Learning y Capacidades analíticas.
52	¿Cuáles son los casos de uso que el oferente deben comprometer dejar implementados y probados en la herramienta de modelado de analítica?	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor.
53	¿Cuál es el volumen actual de datos? a. Que quiere Ingestar. b. A nivel de Gobierno, cual es el volumen de datos que quiere gobernar	Ver Anexo 1 - Volumetría
54	¿Cuál es el incremento diario del volumen de datos que espera ingestar?	Ver Anexo 1 - Volumetría
55	¿Cuál es el incremento mensual del volumen de datos que espera ingestar?	Ver Anexo 1 - Volumetría
56	Solicitamos se detalle el formato, las versiones y el Sistema Operativo en que corren los orígenes de datos; esto a fin de verificar compatibilidades.	Sección 3.2.1 Información de referencia de volúmenes de datos actuales
57	¿Cuántas y cuáles son las fuentes de información que se deberá dejar configuradas? (audio, video, SQLServer, excel, etc.)	Data semiestructurada (json, XML, csv, txt, etc.), Data estructura (DB relacionales ORACLE)
58	¿Cuánta cantidad (TB) de datos tienen previsto llevar a la nube?	Ver Anexo 1 - Volumetría
59	¿Cuánto es el tiempo actual de ejecución de cada uno de los procesos ETL's?	Es variable desde pocos minutos hasta más de 60 horas
60	¿Cómo esperan consumir la información almacenada en este nuevo repositorio unificado?	Analítica; dashboards; Sección 3.1.9 Data Marketplace; Sección de consumo
61	Solicitamos confirmar si el SRI tiene actualmente un DWH y en tal caso solicitamos detallar fabricante, producto, versión, cantidad de tablas y tamaño en TB	Sí. Modelo Físico en la BDD Oracle, Modelo Lógico en SAP Business Intelligence, Oracle Discoverer. Ver Anexo 1 - Volumetría
62	Solicitamos confirmar si el SRI tiene actualmente un DataMarts, y en tal caso solicitamos detallar fabricante, producto, versión, cantidad de tablas y tamaño en TB.	Sí. Modelo Físico en la BDD Oracle, Modelo Lógico en SAP Business Intelligence, Oracle Discoverer. Ver Anexo 1 - Volumetría
63	Solicitamos se detalle el número total de ETL de extracción	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.
64	Solicitamos se detalle el número total de ETL de transformación	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.

No.	Pregunta	Respuesta
65	Solicitamos se detalle el número de tablas de hechos	Como referencia, existen alrededor de 100 productos DWH, que usan al menos una tabla de hechos, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.
66	Solicitamos se detalle el número de tablas de dimensiones	Como referencia, existen alrededor de 150 tablas de dimensiones, que usan al menos una tabla de hechos, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.
67	Solicitamos se informe el volumen de datos (GB) por tabla de hechos / dimensiones o procesos ETL	La más grande 10 TB, más pequeña 2 MB. Tamaño de las dimensiones: 250 GB la más grande, menos de un MB la más pequeña. Ver Anexo 1 - Volumetría
68	Desde un enfoque consultivo, más que de un proveedor de una cierta herramienta, o conjunto de herramientas, consideramos que, dado que es tan amplio el alcance de los requerimientos del RFI, en afán de buscar la mejor solución para el SRI, solicitamos se acepte una solución conformada por más de 1 fabricante, a fin de tomar el mejor componente de cada una de las soluciones funcionales requeridas en el RFI.	De preferencia que sea un solo fabricante que maneje varios componente, presente su propuesta.
69	Solicitamos explicar el objetivo de ingesta de datos de redes sociales	* Análisis de tendencias y patrones (estilo de vida y consumo) * Identificación de actividades económicas no declaradas (comercio electrónico) * Análisis de sentimientos sobre la administración tributaria y sus actuaciones, perspectiva social del contribuyente, entre otras.
70	Solicitamos se definan los casos de uso que el SRI tiene pensado implementar tanto como objetivo del proyecto a contratar (a fin de estimar el esfuerzo en servicios profesionales requerido), como los otros que planea hacerlo a futuro (a fin de estimar las capacidades de las herramientas), en cada una de las etapas del proyecto, como son gobierno, calidad, analítica, etc.	Los casos de uso se definirán en los TDR, sugerimos elaborar su propuesta en base a su experiencia.
71	Solicitamos una reunión presencial antes del plazo para entrega de respuestas al RFI a fin de solventar dudas respecto del alcance que el SRI quiere dar en cuanto a casos de uso de Gobierno y casos de uso de Analítica	La información proporcionada será utilizada como parte del estudio de mercado, las aclaraciones se efectuarán conforme al cronograma publicado
72	El RFI solicita: La herramienta deberá poder generar páginas web de la gestión de maestros con las herramientas mencionadas anteriormente. Al respecto solicitamos que se facilite una explicación más detallada del alcance requerido.	3.1.8 MDM – Datos Maestros

No.	Pregunta	Respuesta
73	<p>El RDI, en su pg. 11, solicita: 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. Al respecto consultamos lo siguiente:</p> <p>a) Al mencionar “la predicción de Modelo de Riegos”, tiene algún caso de uso ya definido. b) ¿Cuáles son los tiempos actuales? c) ¿Cuál son las características técnicas en RAM, procesador, tipo de disco y arquitectura de los equipos en los cuales procesan actualmente la información? Esta información es indispensable para estimar la potencia de procesamiento el cloud que se debe ofertar. d) ¿Cuál es volumen en GB de información que procesa diariamente? e) ¿Cuál es volumen en GB de información que procesa semanal?</p>	<p>a) Si b) Para procesamiento: variable dependiendo las fuentes (entre 1 semana a varios meses); Para generación de modelos: 287 días en modelos de alta complejidad c) c), d) y e) Ver anexo 1 - Volumetría</p>
74	<p>El RFI en su pg. 12 solicita: 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 Al respecto, solicitamos detallar los casos de uso que el SRI requiere sean implementados como parte del presente proyecto, a partir de datos de redes sociales, datos no estructurados como imágenes, documentos, videos y otras fuentes de datos</p>	<p>Los casos de uso se definirán en los TDR, se consideran 5 casos de uso (2 de complejidad alta, 2 complejidad media, 1 complejidad baja) sugerimos elaborar su propuesta en base a su experiencia. Adicionalmente revisar Anexo 1 - Volumetría</p>
75	<p>El RFI en su página 12 solicita: 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:</p> <ul style="list-style-type: none"> • Implementación en la infraestructura tecnológica del SRI (On premise) • Implementación en nube (Cloud) <p>Dado que la mayor parte de los fabricantes de clase mundial (World class) privilegian las soluciones en cloud y en caso de tenerlas disponibles en modalidad on-premise lo hacen con funcionalidades acotadas, solicitamos que sea permitido presentar una propuesta que sólo contemple en nube.</p>	<p>Sujeto a análisis, forma parte de la propuesta requerida.</p>

No.	Pregunta	Respuesta
76	<p>El RFI en su página 13 solicita: 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.</p> <p>Al respecto, tenemos las siguientes consultas: * ¿Qué herramienta de flujos tiene el SRI actualmente y cuál es su versión? * ¿La herramienta de flujos es un standalone o corre clúster? De correr en clúster solicitamos se detalle cuántos nodos tiene. * ¿Cuántos procesos de flujos de datos tiene actualmente? * ¿Cuál es el volumen de datos en GB con estos flujos?</p>	<p>Herramienta de flujos de información (pipelines) no posee la organización, por lo cual debe formar parte de propuesta requerida.</p> <p>Adicional Ver Anexo 1 - Volumetría</p>
77	<p>El RFI, en su página 14, solicita: 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.</p> <p>Al respecto, tenemos las siguientes consultas: * ¿Cuál es el volumen de datos históricos del SRI? * ¿Los datos históricos residen en más de una tabla histórica? De ser así, ¿estas tablas (más de una) mantienen los mismos campos y tipos de datos?</p>	<p>Ver Anexo 1 - Volumetría</p>
78	<p>El RFI en la sección RG17 solicita: La plataforma garantizará la sincronización de permisos/exposición del dato de acuerdo con la configuración de gobierno del dato.</p> <p>Al respecto solicitamos se confirme si esto se refiere a los permisos que se pueden otorgar sobre los diferentes dataset que contendría el datalake</p>	<p>Si</p>
79	<p>El RFI en la sección RG18 solicita: 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.</p> <p>Al respecto tenemos la siguiente consulta: ¿Se refiere una herramienta en la cual se pueda revisar y gestionar los metadatos?</p>	<p>SI</p>

No.	Pregunta	Respuesta
80	<p>El RFI solicita en la sección GD14: 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). Al respecto tenemos las siguientes consultas: * Solicitamos se nos proporcione contexto de cuáles ERP utilizan y sobre que motor de base de datos y sistema operativos residen. * Confirmar si actualmente alguna de las entidades públicas o privadas con las que se deberá conectar el Data Fabric ya tiene un software Historia Clínica Electrónica (HCE) y de ser así detallar fabricante, producto y versión, y el formato de los datos a extraer de tal sistema. * Confirmar si el Data Fabric deberá tener algún tipo de integración con HCE * Podría darnos contexto del software de Historia Clínica Electrónica (HCE)</p>	<p>Forma parte de la propuesta requerida; que la plataforma disponga de mecanismos de conexión a cualquier tipo de datos;</p> <p>Las versiones se encuentran en la sección 3.2.1 Información de referencia de volúmenes de datos actuales</p>
81	<p>El RFP en la sección GD27 solicita: La plataforma deberá permitir recomendar automáticamente entidades de datos según un patrón de búsqueda. Solicitamos Confirmar si este requerimiento se refiere a: Búsqueda: Tabla*. Resultado: Tabla1, Tabla2, Tabla3</p>	<p>No necesariamente solo tablas, sino también búsqueda de dominios, datos maestros, conceptos de información, etc.</p>
82	<p>El RFI en la sección GD29 solicita: La plataforma deberá permitir ingestar diccionarios de datos físicos directamente desde una fuente (es decir, RDBMS) u otras plataformas de metadatos. Al respecto, solicitamos se cite una fuente como ejemplo</p>	<p>A la fecha la organización no dispone de una herramienta de gestión de fuentes de metadatos, la extracción de esta información es manual, como por ejemplo los diccionarios de datos que arroja una base de datos.</p>
83	<p>El RFI en la sección SC5 solicita: 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. Solicitamos confirmar si se requiere obligatoriamente que una sola herramienta gestione los permisos a nivel de datos tanto a nivel de capa lógica como física, ¿o esto puede ser de manera independiente, al menos en la plataforma de gobierno?</p>	<p>Se espera que, 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. Envíe su propuesta</p>
84	<p>Respecto de las capacidades analíticas, de manera específica tenemos las siguientes consultas y solicitudes de aclaración:</p> <ol style="list-style-type: none"> 1. Para el análisis de grafos, ¿qué profundidad requieren? ¿necesitan modelos analíticos sobre los grafos o análisis exploratorio sobre los mismos? 2. ¿Es necesario que la herramienta pueda correr queries de SQL directo o puede hacerse mediante notebooks de Python que permitan un acceso a bases de datos y correr los queries allí? 3. ¿Se necesita que la herramienta genere anuncios de voz y que puedan desplegarse desde la misma, o es suficiente con que pueda generar modelos de texto a voz, puedan descargarse y productivizarse de forma externa? 4. ¿Qué tipos de modelos son los más relevantes? (ML clásico, Deep Learning, etc.) 	<ol style="list-style-type: none"> 1. Se requieren las dos funcionalidades. 2. Directo 3. No requerido al momento. 4. Al menos Machine Learning, Minería de Texto y Deep Learning. 5. Al menos Data Scientists y Citizen Data Scientists. 6. Al menos 20 7. Al menos 10 8. Ver Anexo 1 – Volumetría 9. Si, conforme a las políticas de seguridad establecidas. 10. Si, conforme a las políticas de seguridad institucionales, nacionales y extranjeras. 11. Si, forma parte de la propuesta requerida. 12. Según lo que se establezca en el TDR. 13. Según lo que se establezca en el TDR. 14. No 15. En proceso

No.	Pregunta	Respuesta
	<p>5. ¿Qué tipos de usuarios usarán la plataforma en cuanto a analítica? (Por ejemplo, Data Scientists, ML Engineers, Citizen Data Scientists, etc.)</p> <p>6. ¿Cuántos usuarios estarán creando modelos analíticos dentro de la plataforma?</p> <p>7. ¿Cuántos usuarios crearán modelos analíticos al mismo tiempo (de forma concurrente)?</p> <p>8. ¿Cuál es la volumetría promedio de las tablas que entrarían en los modelos de machine learning?</p> <p>9. ¿Está permitida la conexión entre la web y la plataforma analítica adquirida?</p> <p>10. ¿Se requiere de alguna certificación particular para la instalación de los servicios?</p> <p>11. La plataforma del fabricante H2O tiene conexión a múltiples fuentes de datos ¿Existe alguna otra que consideren puede ser importante en algún momento y que no fue listada?</p> <p>12. En cuanto al componente analítico, ¿Cuánto tiempo se necesita de parte del proveedor para responder a solicitudes no planificadas o requisitos adicionales?</p> <p>13. ¿Existe algún requerimiento de tiempo de respuesta en producción de los modelos en producción para datos tabulares y no tabulares?</p> <p>14. ¿Existe el equipo humano de SRI encargado de la provisión de ambientes en la nube y el gobierno de los mismos?</p> <p>15. ¿El SRI dispone de personal que maneje Kubernetes?, ¿Con qué nivel de experiencia?</p> <p>16. ¿Qué espera el equipo usuario de SRI respecto de la explicabilidad de los modelos analíticos generados?</p> <p>17. ¿Existe alguna restricción por parte del SRI respecto de qué algoritmo debe ser el que se entregue? (ej. Regresión logística, árboles de decisión, etc.)</p> <p>18. ¿Existe alguna restricción con respecto a correlación entre variables, grado de información (IV), WoE, o algún otro por parte del SRI?</p>	<p>16. Seguridad, Logs, Despliegue en la misma herramienta, Visualización del flujo en un mismo lugar incluida la visualización para los procesos de supervisión, Soporte.</p> <p>17. Que provengan de una fábrica de modelos o si es modelo mixto (open Source) que certifiquen el uso al 100% de la versión del algoritmo.</p> <p>18. No se entiende la pregunta, favor aclarar.</p>
85	<p>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). Solicitamos indicar por parte de la entidad la plataforma y/o soluciones de autenticación con los cuáles cuenta la entidad actualmente y a los cuales la solución de analítica debe integrarse.</p>	IDM y Directivo activo de la institución
86	<p>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).</p> <p>Solicitamos a la entidad informar la estimación de la</p>	Tabla 6

No.	Pregunta	Respuesta
	cantidad de usuarios concurrentes y nombrados que tendrán acceso a la solución.	
87	<p>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.</p> <p>Pueden especificar un poco más acerca de este requerimiento como lo visualizan? ¿cómo? ¿por donde? ¿Cuándo? ¿Hora?</p>	La plataforma debe poder gestionar la asignación de roles y perfiles a través de interfaz de usuario o componente existente en la plataforma. Envíe su propuesta.
88	<p>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.</p> <p>Solicitamos a la entidad confirmar si se refieren a que la plataforma pueda ejecutar cargas analíticas sobre la misma infraestructura sin degradar el desempeño en almacenamiento y rendimiento de la plataforma ni a otros sistemas que ejecuten en la misma plataforma, es decir, ejecutar cargas de analítica y OLTP al mismo tiempo.</p>	Si
89	<p>La plataforma deberá permitir encriptar/descriptar las copias de seguridad de los diferentes repositorios de forma transparente sobre los soportes físicos de estas copias.</p> <p>Solicitamos a la entidad ampliar detalle sobre esta solicitud debido a que es de nuestro entender que la encriptación de datos en línea y en reposo garantiza la integridad y fugas de información</p>	Toda información que se encuentre almacenada por temas respaldo, logs, o copias de seguridad, ya sean tanto a nivel confidencialidad de datos, deben ser encriptadas considerando datos en origen, destino y tránsito, con el objetivo de precautelar la sensibilidad de la información, formule su propuesta.
90	<p>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.)</p> <p>Es nuestro entender que este requerimiento es lograr separar roles de administración como DBAs y Sysadmins de los roles de usuario funcional para garantizar que se accedan los datos por quienes los necesitan y no por alguien con privilegios altos.</p>	Si, forma parte de la propuesta requerida.
91	<p>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)</p> <p>De acuerdo con lo solicitado por la entidad, por favor aclarar qué tipo de datos se deben gestionar? Desde qué tecnología</p>	La solución se encontrará en producción en el centro de datos principal y la réplica en contingencia en el centro de datos secundario (GYE)

No.	Pregunta	Respuesta
	y su versión? ¿Las fuentes de datos están alojadas en versiones actualizadas y soportadas por los fabricantes de las diferentes soluciones?	
92	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. Por favor indicar cuántos casos de uso y flujos se migrarían hacia la nueva plataforma. Qué tecnologías se están involucrando en este requerimiento y en qué lenguaje? Python, R, etc.?	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
93	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. De acuerdo al requerimiento es nuestro entender que la entidad espera que los datos obtenidos y virtualizados puedan ser replicados a elección en la misma plataforma en tiempo real hacia otro destino?	En la misma plataforma
94	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. Por favor especificar, ¿Qué tecnologías y fabricantes están involucrados en las fuentes de datos indicadas por la entidad para que se pueda cumplir este requerimiento?	Principalmente estos tipos de fuentes: data semiestructurada (json, XML, csv, txt), Data estructura (DB relacionales ORACLE); sección 3.2.1 Información de referencia de volúmenes de datos actuales, para los volúmenes actuales y versiones
95	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: Datawarehouse, Data Marts, Data lakes, Data hubs. Data Hub involucra tecnologías de terceros, ¿Hacia qué tecnologías debe funcionar el Data Hub? ¿Se piensa utilizar el Data Hub para orígenes de datos voluminosos? Las anteriores preguntas se realizan debido a el desempeño que esto involucra y/o amerita.	Forma parte de la propuesta requerida
96	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. Es de nuestro entender que para la bodega de datos se espera contar también con alta disponibilidad y redundancia Activo/Activo así como las demás bases de datos	SI

No.	Pregunta	Respuesta
	institucionales para que pueda soportar altas cargas de datos y consulta sobre la bodega?	
97	<p>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 alterno se soportan los ambientes de: contingencia, preproducción, testing, desarrollo y ciertos servicios complementarios de producción.</p> <p>De cuál instancia de base de datos se espera extraer la información? ¿Desde producción o desde la réplica de contingencia?</p>	Producción
98	<p>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.</p> <p>Por favor aclarar en qué capa se debe realizar el descubrimiento de datos.</p>	Sección Data Catalogue y Gestión del Metadato, GD -21
99	<p>.. La plataforma deberá soportar múltiples formatos y conectores para diversas bases de datos tanto cloud como onpremise. 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.</p> <p>¿Qué información debe almacenar diferente a datos?</p>	Descripción, alcance, otros
100	<p>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.</p> <p>Dada la gran cantidad de información que maneja la entidad, es de beneficio que se puedan comprimir los datos con el objetivo de ahorrar espacio de almacenamiento de los datos y que no haya pérdida de desempeño con la compresión?</p>	SI
101	<p>El soporte debe ser íntegramente en español tanto a nivel de tickets y notificaciones como a nivel de soporte telefónico o remote.</p> <p>Se sugiere respetuosamente revisar este requerimiento dado que las grandes compañías de tecnología y proveedores de estos tipos de soluciones están a nivel global. Si lo que se requiere es que el implementador y/o partner hable en idioma español, aclarar el requerimiento.</p>	Aplica para implementadores, proveedores y soporte técnico.

No.	Pregunta	Respuesta
102	<p>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.</p> <p>De acuerdo a lo requerido por la entidad, se conoce que no existe un único proveedor que cumpla o tenga todos los productos que cumplan las funciones requeridas por la entidad, por lo cual respetuosamente se recomienda retirar este requerimiento. ¿Se permite integrar soluciones de terceros o toda la plataforma debe ser desarrollada por un mismo fabricante?</p>	Forma parte de la propuesta requerida
103	<p>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 procesado 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.</p> <p>El requerimiento está orientado a un proponente específico sesgando la pluralidad en la participación del proceso, por lo cual solicitamos respetuosamente a la entidad evaluar el requerimiento.</p>	Envíe su propuesta
104	<p>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.</p> <p>El requerimiento de kubernetes y contenedores está orientado a un proponente específico sesgando la pluralidad en la participación del proceso, por lo cual solicitamos respetuosamente a la entidad evaluar el requerimiento.</p>	Envíe su propuesta
105	<p>La plataforma permitirá el desarrollo y ejecución de sistemas operacionales basados en microservicios haciendo hincapié en el uso de las tecnologías open source y contenerización: Kubernetes y Docker. En este ámbito la plataforma deberá garantizar por diseño que se soporte el aislamiento de servicios y de datos.</p> <p>El requerimiento de contenerización, kubernetes y docker, está orientado a una tecnología que orienta a un solo proponente, lo cual sesga la pluralidad de oferentes en la participación del proceso, por lo cual solicitamos respetuosamente a la entidad evaluar el requerimiento.</p>	Envíe su propuesta

No.	Pregunta	Respuesta
106	<p>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).</p> <p>Si la solución requiere hacer uso de componentes híbridos (OnPremise y Nube), una alternativa híbrida es una opción para la entidad?</p>	Si
107	<p>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.</p> <p>Solicitamos a la entidad evaluar o especificar el requerimiento ya que en el caso de la implementación en nube, los costos de la solución se darán por el uso o consumo que esta tenga durante periodos específicos o generales con lo cual establecer un costo estable durante el tiempo, dependerá de los requerimientos del negocio y su uso o demanda que tendrá la solución en nube.</p>	Forma parte de la propuesta requerida
108	<p>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.</p> <p>Entendemos que el pago por uso predecible aplica también para la alternativa On Premise?</p>	Forma parte de la propuesta requerida
109	<p>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.</p> <p>Dado que la entidad solicita una solución compleja en la que se deben incorporar muchos servicios, aplicaciones o funciones, existen servicios que la entidad solicita que solo apliquen a soluciones On-Premise o Nube, solicitamos a la entidad permitir que se puedan instalar o funcionar de manera híbrida (On Premise / Publico)</p>	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
110	<p>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.</p> <p>La entidad solicita "No se permiten componentes alojados en la nube indicada que enlace a módulos fuera de la nube" pero también solicita que la solución integre información de otros servicios, repositorios o medio de nubes públicas, donde posiblemente se requiera incluir en estos algún mecanismo de extracción, procesamiento u otro en nubes, por lo cual solicitamos respetuosamente a la entidad evaluar el requerimiento.</p>	<p>Se espera que: 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</p>
111	<p>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.</p> <p>Dado los requerimientos, funcionalidades y complejidad de la solución que solicita la entidad, no hay un único producto que cumpla con todos los requerimientos solicitados, por lo cual solicitamos que se permita integrar a múltiples productos en la solución.</p>	<p>Envíe su propuesta</p>
112	<p>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.</p> <p>Dado el inventario por parte del SRI de los sistemas, versiones, y al revisar esta contra la compatibilidad de los productos a usar en la solución requerida por la entidad, si se llegase a requerir la actualización de las versiones de bases de datos u otros componentes para cumplir con la compatibilidad y funcionalidad de la solución ofertada, el SRI está en la capacidad de efectuar estas actualizaciones sobre la infraestructura actual?</p>	<p>Se espera el servicio por parte del proveedor, conforme punto 16 del ítem 5.4 Información de la solución tecnológica del RFI</p>
113	<p>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.</p> <p>Se debe verificar este requerimiento dado que según el análisis realizado, cómo un proponente garantiza la encriptación de los datos si están ubicados en productos de terceros donde no se tiene el control de la información con el esquema de virtualización solicitado por la entidad.</p>	<p>Toda información que se encuentre almacenada por temas respaldo, logs, o copias de seguridad, ya sean tanto a nivel confidencialidad de datos, deben ser encriptadas considerando datos en origen, destino y tránsito, con el objetivo de precautelar la sensibilidad de la información, formule su propuesta.</p>
114	<p>¿El proveedor debe de contar con las opciones de poder brindar este servicio tanto en nube como en premisas para tener la opción de una arquitectura híbrida o poder mover</p>	<p>Si, solución contenerizada</p>

No.	Pregunta	Respuesta
	de en premisas a la nube o viceversa en caso de un cambio de política del país en cuanto a soberanía de datos?	
115	¿El proveedor debe de brindar el sitio alternativo en premisas en caso de que el principal sea en nube y/o viceversa?	Si, forma parte de la propuesta requerida el tipo de solución que se plantea
116	Por el volumen de datos y los tipos de analítica que se van a correr, ¿Cómo medirán que el desempeño de la plataforma propuesta cumpla con las expectativas de la institución? Si los datos no se van a consolidar en una plataforma moderna y escalable y pueden residir en un origen que no está definido, ¿Cómo garantizar que la plataforma actual donde residen los datos pueda cumplir con las metas de desempeño esperado? Sobre todo, si se piensa virtualizar.	La plataforma debe incluir infraestructura y software necesario, se espera 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.
117	Se menciona que la plataforma debe ser compatible e integrada a la plataforma actual de los Servicios de Rentas Internas. ¿Se tiene un inventario de las tecnologías y versiones con las cuales hay que integrarse?	Sección: 3.2.1 Información de referencia de volúmenes de datos actuales
118	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. Se menciona el poder utilizar datos de redes sociales. ¿La extracción de dichos datos debe ser parte de plataforma? Si es así, ¿A cuáles redes sociales se debe de conectar la plataforma?	1. Forma parte de la propuesta requerida. 2. Facebook, Instagram, Twitter, Tik Tok, LinkedIn, etc.
119	Cuando mencionan que se debe de brindar el servicio de reingeniería y migración de datos, ¿cuáles son las plataformas origen de los datos? Se menciona que se debe de garantizar la calidad de los datos, ¿Qué se debe de hacer con los datos que actualmente no cumplen con dichas reglas, por ejemplo, fechas erróneas?	En un 90% Oracle, también existen fuentes en archivos JSON y MySQL Se deben levantar y registrar las reglas de calidad de datos en conjunto con el usuario, así mismo se levantarán las acciones en caso de que no cumplan las mismas, como identificar los datos, no considerar los datos o dejarlos pasar en determinados escenarios
120	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. ¿Cuáles dominios se desean implementar y cuáles son las fuentes de datos externas a las cuáles hay que brindar catálogos y otros servicios de gobernanza?	Los dominios serán identificados en el desarrollo de los casos de uso, así como las fuentes en su totalidad Existen en el SRI alrededor de 29 dominios de información y se deberá documentar aquellos que pertenezcan a los casos priorizados
121	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 onpremise o en cualquier nube pública).	3.2.1 Información de referencia de volúmenes de datos actuales

No.	Pregunta	Respuesta
	¿Se podrían especificar los sistemas disponibles y las bases de datos a las cuales se desea integrar la plataforma?	
122	Tanto en la capa de integración, capa de virtualización y el catálogo es necesario delimitar o enumerar las tecnologías y versiones a las cuales hay que integrar sobre todo en la parte de integración en tiempo real. ¿A qué plataformas de datos y metadatos se debe de conectar la solución de catálogo de datos y de gobernanza?	Forma parte de la propuesta requerida
123	¿La capa de acceso común debe de conectarse a que fuentes de información y cuáles son los métodos de consumo? Esa conexión debe ser por JDBC? ODBC? REST, Drivers?	Si, forma parte de la propuesta requerida el tipo de solución que se plantea
124	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. En el requerimiento se menciona que debe ser una solución basada en contenedores, la cual virtualiza, gobierna y disponibilizar la información. ¿El hecho que este basado en contenedores es mandatorio?	Envíe su propuesta
125	La plataforma de gobierno de datos debe de garantizar que un administrador de la plataforma no tenga acceso a los datos sensibles o poder crear cuentas con privilegios de super usuario? Debe manejar políticas del acceso a datos como, por ejemplo, desde que herramienta o aplicación se puede acceder, desde que ip, rango de fechas y horas? ¿Debe de controlar que tipo de comandos se pueden ejecutar hacia la fuente de datos para evitar que puedan cambiar alguna configuración o dejar la fuente origen en estado vulnerable?	Sección: 3.1.10 Seguridad y Cumplimiento del Dato
126	En la sección de calidad no se especifica que debe de pasar con los datos que no cumplen con las reglas de negocio definidas, por ejemplo si un campo viene en nulo o en blanco debe ser asignado a un usuario para su revisión y corrección mediante colas de trabajo de encargados de Calidad de Datos?	Se deben levantar y registrar las reglas de calidad de datos en conjunto con el usuario, así mismo se levantarán las acciones en caso de que no cumplan las mismas, como identificar los datos, no considerar los datos o dejarlos pasar en determinados escenarios
127	Los servicios de la plataforma para la minería de datos, machine learning, algoritmos analíticos, entre otros deben residir junto con los datos para evitar posibles fugas de información, pérdida de desempeño por tener que transmitir grandes volúmenes de información por red entre la fuente y los sistemas de minería de datos, y sobre todo mantener la encriptación y auditoria de los datos en todo el proceso?	Sección: 3.1.10 Seguridad y Cumplimiento del Dato
128	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	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
	de acceso, las fechas, los niveles de seguridad y privacidad de los datos y el motivo de acceso a la colección. Todos los datos deben de pasar por el Marketplace? Se tiene estimado el impacto de desempeño en analítica de pasar por todas estas capas de virtualización y datamarket?	
129	La plataforma debe disponer de capacidades de texto a voz que permitan implementar sistemas de anuncios basados en voz ¿En cuál de las soluciones de la plataforma se necesita convertir el texto en voz?	Para usuario final
130	En el cuadro de la estimación de la Inversión no está la línea de Suscripción? Se asume que si la solución es on-premise la misma debe ser licenciada a perpetuidad?	Forma parte de la propuesta requerida
131	En caso de ofertar una solución on-premise en qué línea va lo referente al hardware y software de base necesario para instalar la plataforma?	La propuesta debe incluir la infraestructura y software necesario
132	Se solicita que la solución deba utilizar kubernetes. El SRI ya tiene desplegada una solución de Kubernetes? En caso afirmativo, podrían compartir la información de la misma? Sobre qué hipervisor la están ejecutando?	hipervisor. VMWare
133	Solicitamos se extienda el cronograma establecido en 10 días hábiles para la presentación de oferta luego de que el SRI remita las respuestas, toda vez de que se trate de un proyecto complejo.	Se publicará boletín de enmiendas con el nuevo cronograma actualizado.
134	Por favor confirmar si para la opción cloud hay que cotizar los servicios de nube pública (IaaS) necesario para instalar la solución o en su defecto especificar los requerimientos de IaaS.	Forma parte de la propuesta requerida
135	Utiliza actualmente el SRI proveedores de nube pública? Podrían mencionarlos?	Sí: Azure
136	Cuando menciona nube privada, se está refiriendo a algún proveedor local en Ecuador o están pensando que el SRI monte una nube privada para instalar la solución? En caso de que el SRI desee montar su propia nube privada, tiene alguna preferencia de proveedor (Nutanix, VmWare, etc.)?	Forma parte de la propuesta requerida
137	Los documentos llamados brochure, datasheets, documentos técnicos generados por todos los fabricantes están en idioma inglés, razón por la cual solicitamos de la manera más comedida se acepte el idioma inglés ya que los documentos originales y fidedignos en este idioma.	Las respuestas a la solicitud de información deberán ser entregadas en español, cualquier documento técnico adicional podrá ser entregado en inglés y con su traducción no oficial al español.
138	En el pliego se habla de tiempo Real por favor detalle a que llama tiempo real Streaming? Micro batch?	Que los datos se proporcionan para su análisis casi de inmediato una vez que se recopilan
139	Es necesario que el motor de tiempo real realice calidad de datos en vuelo es decir en ejecución?	Forma parte de la propuesta requerida
140	Se requiere que el motor de tiempo real ejecute los modelos de machine learning desarrollados y entrenados en la misma plataforma?	Forma parte de la propuesta requerida
141	Se considera que el motor de Real time sea parte de la misma plataforma donde se diseñan se desarrollan y entrenan los modelos?	Forma parte de la propuesta requerida
142	Qué tipo de integración se requiere para el Motor de tiempo real Sensores?	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
143	Qué tipo de integración se requiere para el Motor de tiempo real Bus de integración?	Forma parte de la propuesta requerida
144	Qué tipo de integración se requiere para el Motor de tiempo real IOT? Como: Modbus. MQTT.OPC-UA. OPC-DA. UVC Camera. PylonCamera. Modbus.OSI PI Historian.BACnet gateway devices. Kafka. Cassandra (adapter only). Boar.dRea. der.	Forma parte de la propuesta requerida
145	Se requiere variedad de Integración en el motor Real time como Apache Camel, Audio data. Axeda. BACNET. Cassandra. Common Event Format (CEF). Database ODBC: s IBM DB2, Oracle, IBM Netezza, Sybase ASE. HDAT Reader. HDFS. MQ. JMS. Kafka. MapR. Modbus. MQTT. Nurego. OPC-UA. OPC-DA. QuasarDB. RabbitMQ. RADAR (Asterix CAT240 format support). REST. Solace. Teradata. TIBCO Rendezvous. Timer.URL. WebSocket. XML/JSON file socket adapter.	Forma parte de la propuesta requerida
146	Qué tipo de integración se requiere para el Motor de tiempo real Bases de datos?	Forma parte de la propuesta requerida
147	La interfaz del motor de tiempo real debe ser drag and drop?	Forma parte de la propuesta requerida
148	El motor Real Time debe permitir diseñar los flujos de forma gráfica low code no code?	Sección. 3.1.13 Capacidades ETL
149	El motor Real Time debe permitir analizar datos no estructurados?	forma parte de la propuesta requerida
150	El motor Real Time debe permitir analizar Video?	Envíe su propuesta
151	El motor Real Time debe permitir analizar imágenes?	forma parte de la propuesta requerida
152	El motor Real Time debe permitir procesamiento en memoria distribuido? Esto con el fin de gestionar la velocidad	Forma parte de la propuesta requerida
153	Se requiere que el motor de tiempo real gestione millones de registros en segundos?	Forma parte de la propuesta requerida
154	Es necesario que el motor Real time cuente con una capa de monitoreo que permita gestionar cada proyecto y su comportamiento?	Forma parte de la propuesta requerida
155	Se requiere que el motor Real time tenga la capacidad de retener algunos datos en memoria que permitan optimizar tiempos?	Forma parte de la propuesta requerida
156	Se considera que el motor de Real time utilice el mismo Motor de procesamiento en memoria y distribuido donde se desarrollan los modelos de Mlyá que esto permite una mayor integración?	Forma parte de la propuesta requerida
157	Se requiere que el motor de Real Time maneje la misma arquitectura sobre kubernetes?	Forma parte de la propuesta requerida
158	Se considera la ejecución de algoritmos linear regresión en tiempo Real?	Forma parte de la propuesta requerida
159	Se considera la ejecución de algoritmos logistic regression en tiempo Real?	Forma parte de la propuesta requerida
160	Se considera la ejecución de algoritmos Support vector machine en tiempo Real?	Forma parte de la propuesta requerida
161	Se considera la ejecución de algoritmos K-means clustering en tiempo Real?	Forma parte de la propuesta requerida
162	Se considera la ejecución de algoritmos t-Distributed Stochastic Neighbor Embedding (t-SNE). en tiempo Real?	Forma parte de la propuesta requerida
163	El motor de Real Time requiere que soporte modelos desarrollados con Python code?	Forma parte de la propuesta requerida
164	El motor de Real Time requiere que soporte modelos desarrollados con C code?	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
165	El motor de Real Time requiere que soporte modelos ONNX?	Forma parte de la propuesta requerida
166	Se requiere que el Motor de Real Time soporte ONNX model inferencing on GPU platforms for edge (on-site) and for cloud (Docker).?	Forma parte de la propuesta requerida
167	El motor de Real time debe tener integración con Open Source como Python?	Forma parte de la propuesta requerida
168	Se requiere la capacidad de Gobierno de Modelos desarrollados en la plataforma Analítica?	SI
169	Se requiere la capacidad de Gobierno de Modelos desarrollados con Open Source?	Siempre y cuando tenga soporte y sea parte de la plataforma integrada
170	Se requiere la capacidad de medir el rendimiento de modelos en un módulo de Gestión analítica?	Si
171	Se requiere que la aplicación de Gobierno de modelos centralice todos los desarrollos analíticos realizados en la plataforma?	SI
172	Los datos transaccionales se van a migrar a los repositorios de almacenamiento del Gráfico 1: Esquema General Plataforma tipo Data Fabric	Si, en función de los casos de uso mismos que se definirán en el TDR. Considerar como parte de la propuesta requerida. Se priorizarán 3 casos de uso a ser implementados por el proveedor y 2 a ser implementados por SRI con acompañamiento del proveedor que
173	Se debe contemplar el almacenamiento de respaldo del transaccional?	NO
174	Cuentas personas por modulo se requiere sean capacitadas?	No definido. Envíe su propuesta
175	la capacitación debe ser presencial o puede ser virtual?	Escenario híbrido, dando prioridad a lo presencial.
176	La aplicación de Machine learning debe permitir la generación automática de resúmenes de comparación de modelos a través de gráficos de umbrales, curvas ROC, estadísticas de concordancia y tablas de clasificación errónea para uno o más modelos.?	Sección: 3.1.15 Machine Learning
177	La aplicación de Machine learning debe permitir la construcción de modelos de Deep learning con técnicas que incluyan redes neuronales de avance profundo (DNN), redes neuronales convolucionales (CNN), redes neuronales recurrentes (RNN) y autoencoders.?	Sección: 3.1.15 Machine Learning
178	La aplicación de Machine learning debe permitir soporte para arquitecturas de red predefinidas, como LeNet, VGG, ResNet, DenseNet, Darknet, Inception, ShuffleNet, MobileNet, YOLO, Tiny YOLO, Faster R-CNN y U-Net.	Forma parte de la propuesta requerida
179	la plataforma analítica debe permitir hacer preprocesamiento del texto: tokenizar, suprimir stop words, mayúsculas/minúsculas, stemming o lematización, k-gramas. ?	Forma parte de la propuesta requerida
180	Para el DataLake existe alguna preferencia alineada con lo que existe actualmente. En caso de estar abiertos a otras marcas existe alguna preferencia tanto on-premise como cloud: Cloudera, Teradata, Azure, Oracle o alguna otra?	Forma parte de la propuesta requerida
181	Cuenta el SRI con algún acuerdo marco con alguna marca proveedora de bases de datos?	Contrato de soporte Oracle
182	Es necesario proveer las licencias de bases de datos o el SRI utilizará licencias que ya dispone?	La nueva solución debe incluir las licencias de base de datos que requiera.
183	Es posible plantear un escenario de SQL Server para el Datawarehouse?	Considerando el volumen de datos a manejar se espera la propuesta del proveedor.
184	Es posible plantear un escenario de hadoop open Source para el Datawarehouse?	Mientras sea una plataforma de administración y gestión centralizada, es factible.

No.	Pregunta	Respuesta
185	EL SRI busca explícitamente un recambio de la infraestructura de almacenamiento o es posible reutilizar lo que ya tienen?	La propuesta debe incluir la infraestructura y software necesario
186	EL SRI busca explícitamente un recambio de las licencias de almacenamiento en Datalake y Datawarehouse o es posible reutilizar lo que ya tienen?	Si, es posible reutilizar el Datawarehouse
187	Que cantidad de Usuarios tienen por ETL?	Los 100 productos DWH, son consumidos por más de 3000 usuarios.
188	Que cantidad de sesiones concurrentes tienen para el ETL?	Actualmente el Usuario/Esquema de Oracle, ejecuta mediante Jobs de BDD, Paquetes PL/SQL, flujos de SAP DataServices, etc. en forma continua durante todo el día.
189	Que peso de los datos en GB tienen por sesión a ser procesados?	Ver Anexo 1 - Volumetría
190	Que cantidad de Usuarios manejan para dashboards?	Al menos 1800 usuarios
191	Que cantidad de sesiones concurrentes tienen para desarrollar Dashboards?	Al menos 150 usuarios concurrentes
192	Que peso de los datos en GB tienen por sesión a ser procesados?	Ver Anexo 1 - Volumetría
193	Que cantidad de usuarios manejan para Visualización?	Tabla 6
194	Que cantidad de sesiones concurrentes tienen que consulten información (usuarios finales)?	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%
195	Que cantidad de usuarios manejan para el resto de las aplicaciones: Machine Learning, Seguridad, Políticas y Gobernanza, Exploración.	Tabla 6
196	Que cantidad de sesiones concurrentes utilizan cada una de estas últimas al día de hoy?	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%
197	Que cantidad de GB manejan en cada una de estas últimas?	Ver Anexo 1 - Volumetría
198	Que cantidad de políticas de gobernanza y seguridad se aplican al día de hoy?	Existe un procedimiento de Gestión de la Información que tiene temas básicos de calidad, baja de información y atención de necesidades de información, adicional existe 1 política de seguridad de la información que se base en el EGSI V2
199	Si la oferta la presenta un partner y es quien firmará el contrato, se aceptará como válido los certificados emitidos por fabricante que solicitan en el punto 5.1 de este RFI????	No, la experiencia certificada es para quien presentará la oferta e implementará el proyecto.
200	¿La conexión al entorno donde se realizarán las labores iría por escritorio remoto o por VPN?	Escenario híbrido, dando prioridad a lo presencial.
201	¿Qué indicadores manejan actualmente?	Estratégicos, tácticos, operativos
202	¿A qué tipo de persona se realiza la presentación de los indicadores?	Proceso de Planificación Institucional - Control de Gestión
203	¿Qué importancia tienen los datos?	Constituyen uno de los activos más importantes de la institución pues con ellos permiten soportar la toma de decisiones.
204	¿Qué tipo de homologaciones realizan?	El negocio utiliza la metodología CRISP-DM
205	¿Microservicios a que hace referencia? ¿Almacenamiento o consumo de datos?	De acuerdo al uso: 3.1.15 Machine Learning; 3.1.16 Capacidades incluidas aplicaciones operacionales
206	¿De dónde provienen los datos que alojan en los Microservicios?	De acuerdo al uso: 3.1.15 Machine Learning; 3.1.16 Capacidades incluidas aplicaciones operacionales
207	¿Tienen una ETL predefinida?	Pentaho SAP DataServices Oracle PLSQL
208	¿Ya cuentan con algún tipo de analítica descriptiva, prescriptiva o predictiva?	Analítica descriptiva y avanzada (SPSS Modeler)

No.	Pregunta	Respuesta
209	¿Basado en la bodega de datos que actualmente tienen, que tipo de análisis realizan actualmente sobre los datos que la alojan?	Analítica con SPSS, Reporteria con SAP Business Objects, Oracle Discoverer, Pentaho, Saiku, Apache SuperSet, Qlik.
210	¿Con que periodicidad analizan los datos?	Depende el caso de uso, puede variar diario, semanal, mensual.
211	¿Todas las tablas que se encuentran en la bodega tienen información similar?	Estructura relacionales en bases de datos Oracle
212	¿Cuáles son las columnas con mayor veracidad para extraer información?	El dominio de recaudación es uno de los que tiene mayor calidad de información. Se espera que con el proyecto exista un ejercicio de calidad de datos.
213	¿Los datos existentes se encuentran con calidad de datos?	La revisión y validación forma parte de la propuesta requerida
214	¿Cuáles son los datos demográficos más relevantes para el análisis que realizan?	Información de catastro y registro civil
215	¿Actualmente tienen un modelo de dashboards para la visualización de los datos? ¿En qué herramienta?	forma parte de la propuesta requerida
216	¿Que buscan analizar en estas visualizaciones?	Desde estadísticas generales hasta análisis de tendencias, patrones, comportamientos, etc.
217	Teniendo en cuenta la Bodega de Datos actual. ¿Tienen el modelo de entidad relación?	Se tiene levantado por productos
218	¿Se cuenta con un modelo de Log para validar la trazabilidad de los procesos?	Se manejan Logs en los procesos de integraciones y ETL's
219	¿Tienen el Mock-Up para el diseño de la aplicación?	Si se refiere al LookAndFeel debe alinearse al diseño institucional realizado por comunicaciones
220	Al tratarse de información considerada sensible, ¿es viable como se plantea en el documento una fábrica de datos montada sobre una infraestructura cloud publica (AWS, Azure, GCP)?	SI
221	¿Es posible que sea a través de una nube privada alojada dentro del territorio nacional?	SI
222	¿Cuáles son las políticas de adquisiciones del BID?	https://projectprocurement.iadb.org/es/politicas
223	¿Cuál es la volumetría esperada?	Datos actuales sección: 3.2.1 Información de referencia de volúmenes de datos actuales. Ver Anexo 1 - Volumetría
224	¿Qué tipo de disponibilidad de necesitan de los datos, Near real time, batch?	Capa de integración
225	Los modelos to be de fábricas de datos son ecosistemas que integran diferentes soluciones de operación, inteligencia de negocios, analítica, calidad y gobierno de datos entre otros, ¿existen preferencias de fabricantes o tecnologías puntuales?	No
226	¿Existen necesidades puntuales (no macro) que deban ser atacadas primero?	Relacionado a los casos de uso, envíe su propuesta
227	¿Existe un listado de fuentes externas de datos que deberían conectarse a la solución tecnológica?	Si (33 convenios con entidades externas y bases de datos públicas ej. INEC, Panama papers, Super Intendencia de Bancos y Compañías, Función Judicial (Juicios), Sistemas de georeferenciación (Geoportal-INEC, Cartografía-IGM, Google maps) etc.)
228	Se habla puntualmente de "soluciones drag and drop", ¿es un limitante? Roles como científicos de datos deberían ver el código de las soluciones de analítica y roles de BI si podrían trabajar en drag and drop	Es un sugerido, si no existe no es un limitante

No.	Pregunta	Respuesta
229	La gestión de datos se soporta en herramientas tecnológicas, se mencionan algunas de las áreas del conocimiento del DAMA, ¿se considera una consultoría en gobernanza de datos que apalanque la estrategia de la organización?	No
230	¿Qué tipos de datos no estructurados y que volumen de ellos se espera procesar y con qué propósitos?	Data semiestructurada (json, XML, csv, txt, etc.), Data estructura (DB relacionales ORACLE), los cuales son utilizados para los análisis propios de la Administración Tributaria. Ver Anexo 1 - Volumetría
231	¿Qué avances existen dentro de modelos de gobernanza de datos, existen procesos, políticas, Cloe, dominios de datos y demás?	Existe un procedimiento de Gestión de la Información que tiene temas básicos de calidad, baja de información y atención de necesidades de información; además se han identificado los dominios de información que deberán ser validados y ajustados por el proveedor, existen conceptos de información que deberán ser actualizados según el alcance de los casos de uso, no existen aún asociados atributos y demás componentes del gobierno
232	Los KPI se piden de generación automática, ¿estamos hablando realmente de KPI o de métricas de desempeño?, los indicadores de negocio deben ser definidos por los especialistas del negocio	No se especifica generación automática en los KPIs favor aclarar pregunta.
233	¿Existe control de accesos por directorio activo?	Si
234	¿Existen catálogos de datos, meta data técnica y de negocio?	Existen diccionarios de datos
235	¿Se pueden considerar modelos como Lake-house o data mesh?	3.1.4 Capacidades de Almacenamiento
236	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. Agradecemos ilustrar con un ejemplo que tipo de datos son los que se deben capturar desde la solución para hacer Update sobre los sistemas transaccionales.	Dependerá de los escenarios planteados en los caso de uso, por ejemplo: 1. Al aplicar reglas de calidad y determinar datos erróneos por ejemplo error en el formato de fecha la plataforma deberá corregir en el sistema transaccional el dato correcto. 2. Si un contribuyente en una declaración envía un dato inconsistente o atípico (basado en su información histórica) la plataforma deberá enviar una alerta hacia el contribuyente para que revise su información y la corrija en tiempo real.
237	¿Cuál es el número de usuarios que accederán a la plataforma?	Tabla 6
238	¿Cuántas son, de manera inicial, las fuentes de datos que deberán integrarse?	Data semiestructurada (json, XML, csv, txt, etc.), Data estructura (DB relacionales ORACLE), datos de redes sociales y fuentes externas.
239	¿Existe ya un gobierno de datos definido por la entidad?	No
240	¿Existe una estimación de volumetría de los actuales datos a integrar, así como su estimación de crecimiento?	Datos actuales sección: 3.2.1 Información de referencia de volúmenes de datos actuales. Ver Anexo 1 - Volumetría
241	¿Cuáles son los principales desafíos o problemas actuales que enfrenta en la gestión tributaria y cómo espera que esta solución tecnológica los resuelva?	Detallado en los ítems 2.2 Necesidad principal y 2.3 Necesidades específicas y 3. Requerimientos de la solución.
242	¿Qué herramientas externas y sistemas existentes necesita que la solución tecnológica pueda integrar y trabajar en conjunto?	Sistemas transaccionales propios de la administración tributaria
243	¿Cuáles son las expectativas en términos de rendimiento y tiempos de respuesta para consultas y análisis de datos?	Se espera tiempo real o cercano al tiempo real

No.	Pregunta	Respuesta
244	¿Cuáles son los plazos y fechas límite para la implementación de esta solución tecnológica?	A partir de la fecha de adjudicación del contrato, se estima un horizonte de implementación aproximado de 2 años y medio
245	¿Qué modelos de licenciamiento o estructura de costos se ajustarían mejor a sus necesidades?	Aún no está definido, envíe su propuesta.
246	¿Existen requisitos regulatorios o normativos específicos que debamos tener en cuenta al diseñar la propuesta?	Ley de Protección de Datos Personales, LOTAIP
247	¿Cuáles son las fuentes de información específicas que desea integrar en la plataforma de conectividad?	Data semiestructurada (json, XML, csv, txt, etc.), Data estructura (DB relacionales ORACLE), datos de redes sociales y fuentes externas.
248	¿Qué nivel de escalabilidad requiere la solución de conectividad?	Escalable horizontal y vertical
249	¿Existen requisitos especiales en términos de formatos de datos o gestores de bases de datos que debamos tener en cuenta?	Se espera que la solución incluya tipos de datos: XML, Json, archivos planos, pdf, estructurados.
250	¿Cuáles son las necesidades de sincronización y replicación de datos? ¿Se requiere una replicación en tiempo real o en modo batch? ¿Existen consideraciones particulares en cuanto a la consistencia y la cadencia de los datos replicados?	Actualmente se manejan intervalos cortos de 5 minutos, por hora, diaria, semanal y mensual. Se espera el menor tiempo de sincronización y replicación.
251	¿Cuál es el alcance esperado de la ingesta de datos en la plataforma? ¿Se espera que la plataforma admita datos estructurados, semiestructurados y no estructurados? ¿Cuáles son las fuentes de datos potenciales y las latencias de entrega de resultados esperadas?	Que tenga la capacidad de integrar datos de diferentes orígenes y formatos, estructurados, semiestructurados y no estructurados. Sistemas transaccionales propios del SRI, BDD Datawarehouse, fuentes externas como bases de datos de dominio público (ej. INEC), sitios web especializados (Panamá papers), etc. Se espera contar el menor tiempo de latencia.
252	¿Cuál es la importancia de la gestión de datos maestros para su organización? ¿Qué tipo de herramientas y funcionalidades desean tener para gestionar la duplicación de información, editar y corregir datos maestros y mantener la información de referencia?	La Gestión de Datos maestros es fundamental para poner a disposición de los usuarios de manera que permita crear modelos analíticos y toma de decisiones a partir de los datos. Con la implementación del Gobierno se podrán identificar y optimizar fuentes duplicadas, adicional a las reglas de calidad de datos que permitan identificar duplicidad, deben ser herramientas de fácil manejo para usuarios no técnicos y gráficas con reportes que permitan revisar las acciones realizadas sobre los datos
253	¿Cuáles son los principales riesgos o desafíos de seguridad a los que se enfrenta su organización en relación con la gestión y protección de los datos?	Fugas, divulgación no autorizada, mal uso de información, filtración masiva de datos.
254	¿Cuáles son los requisitos normativos y de cumplimiento que deben tenerse en cuenta al gestionar los datos en su organización?	Se debe considerar el cumplimiento de la Política de Seguridad de la Información y Política de Datos Personales, así mismo el cumplimiento de estándares y buenas prácticas como ISO 27001 y DMBOOK y otras
255	¿Existen políticas internas de seguridad de la información en su organización? Si es así, ¿cuáles son los aspectos clave de estas políticas?	Si, cumple con la ISO 27002:2013
256	¿Cuáles son los volúmenes de datos con los que trabajarán en la plataforma?	Ver Anexo 1 - Volumetría
257	¿Cuál es la preferencia en cuanto a la frecuencia y concurrencia de acceso a los datos? ¿Es necesario gestionar altas cargas de consulta y actualización continua de información?	SI
258	¿Cuáles son las principales funcionalidades que se requieren para la visualización y uso amigable de ontologías y	Permita definir términos de negocio, establecer relaciones, jerarquizarlos, definir conectores de términos de negocio, atributos de negocio.

No.	Pregunta	Respuesta
	relaciones semánticas en la plataforma de consumo de datos?	
259	¿Cuáles son las opciones disponibles para acceder a los datos desde otros sistemas? ¿Se contempla el uso de API REST, ODBC/JDBC u otros estándares de acceso?	Si, capa de gobierno, capa de consumo
260	¿Cuáles son los principales desafíos o problemas que esperan resolver mediante el uso de inteligencia artificial y aprendizaje automático en su organización?	Identificar comportamientos de contribuyentes que permitan reducir la evasión y elusión, así como el incremento en la recaudación debido a la determinación de casos de control con mayor precisión
261	¿Tienen previstas integraciones con otras herramientas o sistemas existentes en su organización? ¿Qué tipo de integraciones les gustaría establecer y qué beneficios esperan obtener de ellas?	Modelos de riesgo, reportería. Extracción de información y tableros
262	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. Por favor explicar mejor la dinámica que se necesita porque se puede entender como recomendaciones en tiempo real son sugerencias que provee la plataforma o el implementador. En cuanto al punto momento antes del almacenamiento por favor expandir ya que los datos que se evalúan ya están almacenados y si se quiere hacer una limpieza las operaciones que se van a realizar deben estar validadas por un encargado para no generar impacto negativo en las DB.	Forma parte de la propuesta requerida
263	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. Por favor explicar mejor si es una regla de negocio o calidad o si ambas significan lo mismo en el contexto de SRI. De igual forma señalar la expectativa del alcance en cuanto a flujos, aprobación y proceso de remediación.	Son reglas de calidad Se espera que la herramienta permita identificar el cumplimiento de las reglas de calidad del dato, que tenga flujos de aprobación, así como generar reportes de cumplimiento, estadísticas de ejecución de reglas y cambio de datos en el origen si amerita
264	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. Por favor explicar mejor la expectativa de la gráfica de los registros y la conexión de calidad con el linaje.	El linaje mostrará como el dato fue calculado, así mismo en este reporte se mostrará si se ha aplicado alguna regla de calidad para la transformación, de forma visual
265	En el pliego se especifica ejecución en tiempo real, nos pueden por favor detallar el tiempo de ejecución al que se denomina Tiempo Real?	Los datos se proporcionan para su análisis casi de inmediato una vez que se recopilan
266	Las ejecuciones de Tiempo Real deben ser de millones de registros?	Dependerá del caso de uso, pero si existen casos de uso que manejan este volumen de información (comprobantes electrónicos)
267	Cuántas Transacciones por segundo debe ejecutar el motor de tiempo real?	No se detalla por segundo sin embargo solicitamos revisar el Anexo 1 - Volumetría
268	La aplicación de tiempo Real debe tener capacidad de ejecutar diferentes proyectos al mismo tiempo?	La solución debe permitir tiempo real

No.	Pregunta	Respuesta
269	El Requerimiento de Tiempo Real debe permitir realizar filtros y operaciones en ejecución?	La solución debe permitir tiempo real
270	La ejecución de tiempo Real debe permitir ejecutar Machine Learning en streaming?	La solución debe permitir tiempo real
271	Qué tipo de capacidades analíticas deben ejecutarse en Tiempo Real, por favor detalle los algoritmos?	Modelos analíticos en línea de Clasificación y Segmentación. Pronósticos, árboles de decisión, regresiones, redes neuronales, modelos lineales generalizados, clústeres, detección de anomalías etc.
272	La ejecución en tiempo Real debe integrarse por medio de REST o qué tipo de integración?	La solución debe permitir tiempo real
273	La aplicación de tiempo real debe permitir integración con Open Source?	La solución debe permitir tiempo real
274	La interfaz de la aplicación de tiempo Real debe estar integrada con la interfaz de la plataforma analítica?	La solución debe permitir tiempo real
275	El desarrollo de los proyectos de Real Time debe ser por medio de una interfaz gráfica arrastrar y soltar?	La solución debe permitir tiempo real
276	La aplicación de Tiempo Real se requiere que ejecute reglas de calidad de datos?	La solución debe permitir tiempo real, calidad de datos
277	la aplicación Real time debe ejecutar Video, Imágenes?	De los casos de uso identificado no existe este tipo de fuente de información (Video, imágenes), pero es deseable que la plataforma si considere el manejo de estos tipos, para casos de uso futuros.
278	Se considera dentro del alcance la inclusión de un motor de Decisión analítico?	Si
279	Se considera dentro de las capacidades de la plataforma un motor de decisión que integre los modelos analíticos desarrollados en la plataforma analítica?	Si
280	Se contempla la ejecución de los modelos analíticos en un motor de decisión que mezcle reglas de negocios con el resultado analítico?	Si
281	Se contempla que la ejecución de los modelos analíticos sea dentro de un Work Flow de Negocio gráfico que permita gestionar decisiones?	Si
282	Se contempla un módulo de gestión de modelos analíticos?	Debe ser el mismo módulo de gobierno
283	Se tiene un esquema de gobierno de modelos analíticos?	Debe ser el mismo módulo de gobierno
284	Se contempla la capacidad de gestionar un único repositorio de modelos para la puesta en producción?	Si
285	Se contempla que la gobernanza permita automatizar el ciclo de vida del modelo mediante un enfoque CI/CD?	Forma parte de la propuesta requerida
286	Se contempla que la integración del Módulo de gobernanza de modelos mediante un enfoque CI/CD tenga múltiples integraciones de aplicaciones utilizando API REST?	Forma parte de la propuesta requerida
287	los modelos analíticos se encuentran en un formato migrable estándar o se requiere rehacerlos?	Los modelos Analíticos propiamente se encuentran en SPSS de IBM. Los Cubos de Información: Modelo Físico en la BDD Oracle, Modelo Lógico en SAP Business Objects y Oracle Discoverer.
288	Cuántos informes en la capa de visualización se contemplan en el proyecto	Depende de los casos de usos los cuales se especificarán en los TDR. Considere un número estimado a implementar dentro de su propuesta.
289	Cuántas tablas por Fuente de información se desean automatizar la carga en la fábrica de datos	Ver Anexo 1 - Volumetría
290	Cuántos Casos de Uso analíticos se contemplan en el proyecto	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor.

No.	Pregunta	Respuesta
291	Clasifique el número de modelos analíticos en complejidad Alta media y baja	Complejidad baja: 2 Complejidad media: 5 Complejidad alta: 4
292	La capacitación puede ser 100% virtual?	Escenario híbrido, dando prioridad a lo presencial.
293	Por favor detallar el número de usuarios por modulo que requieren ser capacitados?	Capa de gobierno: Al menos 10 Capa de Explotación: Al menos 20 Capa de Integración: Al menos 8 Capa de consumo: Al menos 40 Capa de conectividad: Al menos 5 Capa de seguridad: Al menos 10
294	Por favor detallar los usuarios por aplicación y sesiones concurrentes	Se tienen alrededor de 1.800 usuarios en la herramienta actual con una concurrencia del 10 al 15%. Ver Tabla 6.
295	por favor detallar el tamaño de los datos a procesar por aplicación EJ ETL, Visualización, Machine Learning	Ver Anexo 1 - Volumetría
296	La infraestructura para soportar toda la arquitectura para la opción On Premise, Cloud o Híbrido, debe ser considerada en la propuesta?	La propuesta debe incluir la infraestructura y software necesario
297	Hay que presentar todos los costos de infraestructura o los requerimientos de infraestructura los colocará el SRI? a. Implementación en la infraestructura tecnológica del SRI (On premise) b. Implementación en nube (Cloud) – para este caso en el tema de Almacenamiento, hay que considerar toda la data que tiene el SRI, con BDD de 300TB? La pregunta es por el siguiente punto: b.1. 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.	La propuesta debe incluir la infraestructura y software necesario
298	Se debe migrar toda la información a la nueva arquitectura de almacenamiento?	solamente lo relacionado a los casos de uso
299	La plataforma debe permitir ser replicada a otra zona de disponibilidad e incluso a otra nube, eso se debe cotizar para el esquema Cloud? Se debe manejar alta disponibilidad, contingencia?	SI
300	Actualmente el SRI tiene varios motores de BDD donde almacena grandes volúmenes de información, este proyecto en su arquitectura tiene la capa de almacenamiento que sería adicional a lo que ya tiene el SRI en almacenamiento y motores de BDD? Hay alguna preferencia en motores de BDD como ORACLE? o dependiendo de si es relacional o no relacional?	No se tiene preferencia, se espera se especifique en la propuesta requerida
301	Describa el modelo de facturación y pago de la solución incluyendo los diferentes componentes por cada tipo de implementación aplicado (ON PREMISE / CLOUD)? La consulta es si la infraestructura debe ser cotizada, así como las licencias de motores de BDD para almacenamiento en ambos esquemas - ON PREMISE y CLOUD, considerando los volúmenes de información indicados? En el esquema CLOUD será como servicio mensual la facturación a presentarse?	La propuesta requerida deberá incluir infraestructura y software necesario para la plataforma esta puede ser en nube, local o híbrida)
302	Para esquema CLOUD, también se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia?	SI

No.	Pregunta	Respuesta
303	Tanto para la cotización en On-premise como Cloud de la solución, la infraestructura será provista por SRI? o el proponente deberá incluir infraestructura dentro de su propuesta?	La propuesta puede considerar una oferta incluyendo la solución con o sin la infraestructura, siempre que considere el soporte para que la institución contrate el servicio y la instalación
304	Sobre los procesos de migración de los desarrollos existentes del SRI, nos podrían proporcionar mayor detalle?, de qué tipo son, su volumetría y consumo.	La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría
305	El precio de licenciamiento de la plataforma debe ser anual?	No es un requisito obligatorio, presente su propuesta.
306	Sobre los servicios profesionales, qué tipo de actividades se esperan realizar en la implementación, solicitamos mayor detalle?	Incluya las actividades como parte de su propuesta.
307	Hay definido un marco de tiempo por el cual el SRI requiera el acompañamiento de los servicios profesionales?	No definido. Envíe su propuesta
308	Para los desarrollos de analítica será necesario utilizar la volumetría completa que se describe en las fuentes de información o un porcentaje?	Toda la información para dejar definido el DataLake y según los casos de uso a desarrollar solo en esos casos se utilizar lo que se requiera
309	Cuántos usuarios concurrentes habrán de desarrollo y cuántos de consulta?	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%. 10 usuarios de ingeniería de datos y 20 científicos de datos.
310	Cuál es el presupuesto asignado?	El presupuesto se asignará en función del estudio de mercado
311	Cuál es la estimación de crecimiento a nivel de datos?	Ver Anexo 1 - Volumetría
312	¿Cuáles son todos los requisitos de seguridad y cumplimiento normativo que se debe cumplir?	EGSI v2, ISO 27001, LOPDP, LOTAIP, ISO 27000, 270017, 270018 y sus actualizaciones
313	¿Cantidad de GB por hora a mover?: Cantidad de datos desde un sistema transaccional transferido al Datawarehouse. Teniendo en cuenta un tamaño medio es de 600GB por hora.	Ver Anexo 1 - Volumetría
314	1. Cantidad de Querys simples (menores a 1 minuto), moderados (entre 1 y 10 minutos) y complejos (mayores a 10 minutos) por hora de todas las bases a obtener la información. Teniendo en cuenta que un tamaño medio es: - 4000 querys simples (<1min) - 50 querys moderados (1 - 10 min) - 10 querys complejos (>10min)	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso. El tiempo de ejecución del ETL puede variar de entre pocos minutos hasta dos o tres días.
315	2. ¿Cuántas bases de datos se van a consultar? Teniendo en cuenta que un tamaño medio es 10.	Sección 3.2.1 Información de referencia de volúmenes de datos actuales; bases de datos actuales, tamaños y versiones
316	3. ¿Cuántos usuarios concurrentes? Teniendo en cuenta que un tamaño medio es de 20.	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%
317	Consultamos si la institución ha llevado a cabo un proceso para medir el nivel de madurez de este modelo, es decir si ya tiene un enfoque más formal para la gobernanza y la mayordomía. Para su dimensionamiento es necesario saber lo siguiente:	Si se llevó a cabo una consultoría y el nivel de madurez es iniciado
318	1. Cantidad de usuarios concurrentes, es decir conectados simultáneamente. Teniendo en cuenta que un tamaño medio es 200.	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%
319	2. ¿Cuántas tablas desean catalogar? Teniendo en cuenta que un tamaño medio es 100.000	Ver Anexo 1 - Volumetría
320	3. ¿Cuántas columnas en promedio tienen sus tablas? Teniendo en cuenta que un tamaño medio es 40 por tabla.	Ver Anexo 1 - Volumetría

No.	Pregunta	Respuesta
321	4. ¿Cuántos registros en promedio tienen sus tablas? Teniendo en cuenta que un tamaño medio es de 1 millón por tabla.	Ver Anexo 1 - Volumetría
322	5. En lo referente a artefactos de gobierno de información, cantidad de: - Categorías: Una categoría es una entidad que contiene términos de negocio relacionados, por ejemplo, categoría clientes, ventas, contabilidad, teniendo en cuenta que un tamaño medio es 300. - Términos en el glosario: Teniendo en cuenta que un tamaño medio es 5.000 - Conjunto de datos de referencia: Un conjunto de datos de referencia es un catálogo, por ejemplo, de provincias con su código, códigos de tipo de producto, etc., que son utilizados en el perfilamiento de datos para identificar clases de datos. Teniendo en cuenta que un tamaño medio es de 50. - Clases de datos: Las clases de datos ayudan a identificar por ejemplo emails, cédulas, código de producto, ciudades, provincia y cualquier otro dato dentro de los orígenes de información, pueden ser identificadas por diversos mecanismos (expresiones regulares, lista de valores válidos, conjunto de datos de referencia). Teniendo en cuenta que un tamaño medio es de 500. - Clasificaciones: Una clasificación permite clasificar activos de información como información que identifica a una persona, información sensible, información confidencial o cualquier otra categoría que se desee crear. Teniendo en cuenta que un tamaño medio es de 30.	Existe un levantamiento de Dominios de información que son alrededor de 29, conceptos de información que son 1576, aún no se han asociado atributos, categorías, datos de referencia, etc. En relación con la clasificación de la Información se tiene: Pública, Reservada Alta, Reservada Media, Reservada Baja y Dato Personal Sensible. Se deberá realizar la validación y actualización de esta clasificación para los objetos de los casos de uso
323	6. ¿Cuántas columnas se esperan perfilar diariamente? Teniendo en cuenta que un tamaño medio es de 10.000 por día	Ver Anexo 1 - Volumetría
324	7. ¿Cuántas reglas de calidad se esperan ejecutar diariamente? Teniendo en cuenta que un tamaño medio es 1.000 por día.	Depende del levantamiento de reglas en la implementación de los casos. Ver Anexo 1 - Volumetría
325	8. ¿cuántas columnas nuevas se desean importar diariamente? Teniendo en cuenta que un tamaño medio es de 20.000 por día.	Ver Anexo 1 - Volumetría
326	9. ¿Cuántas interacciones de linaje se esperan ejecutar diariamente? Teniendo en cuenta que un tamaño medio es de 50 por día.	Ver Anexo 1 - Volumetría
327	10. ¿Cuántos objetos con información de linaje (activos) se espera tener?: Una importación de linaje importa tablas, procedimientos almacenados, reportes, consultas. La pregunta hace referencia a un valor total aproximado de cantidad de todos estos objetos, un valor único, no hace falta total de tablas o procedimientos, sino un valor total. Teniendo en cuenta que un tamaño medio es de 10.000 activos.	Ver Anexo 1 - Volumetría
328	11. ¿Cuántos científicos o analista de datos concurrentes van a hacer preparación de datos desde la plataforma?: La preparación de datos es equivalente a modificar, crear, borrar columnas de Excel, utilizar fórmulas, entre otros. Teniendo en cuenta que un tamaño medio es de 3 concurrentes.	10 Ingenieros de datos; 20 científicos de datos
329	1. Cantidad de TB a almacenar? Teniendo en cuenta que un tamaño medio es de 17 TB sin compresión	Ver Anexo 1 - Volumetría

No.	Pregunta	Respuesta
330	1. ¿Cantidad de usuarios viewer, teniendo en cuenta que un tamaño medio es 500?: Los miembros pueden leer el contenido público. Por ejemplo, pueden suscribirse a los reportes y ver dashboards.	Entre 800 y 1300 usuarios
331	2. ¿Cantidad de usuarios analytics, teniendo en cuenta que un tamaño medio es de 200?: Los miembros pueden crear nuevos reportes, dashboards, exploraciones, historias, nuevos trabajos, conexiones de servidor de datos y orígenes o módulos de datos. Pueden ejecutar informes, responder a solicitudes y cargar archivos.	Entre 100 y 200 usuarios
332	3. ¿Cantidad de usuarios explorador, teniendo en cuenta que un tamaño medio es 100?: Super usuario: los miembros pueden crear nuevos modelos (cubos), los miembros tienen los mismos permisos que los usuarios analytics. También pueden acceder a: - Exploration - Planning Analytics para Microsoft Excel - Cognos Framework Manager - Dynamic Query Analyzer - Jupyter Notebook - Transformer	Al menos 50 usuarios
333	4. ¿Cantidad de usuarios administradores, teniendo en cuenta que un tamaño medio es 2?: Los miembros tienen los mismos permisos de acceso que los usuarios exploradores. También pueden acceder a: - IBM Software Development Kit - Componentes del menú Gestionar, incluido IBM Cognos Administration.	Mínimo 2
334	5. ¿Concurrencia, % de usuarios conectados simultáneamente? Teniendo en cuenta que un tamaño medio es de 2% para reportes y 6% dashboards.	Se tiene 1.800 usuarios de consulta de los cuales entre el 10% y 15% son concurrentes. Se estiman 20 científicos de datos y 10 ingenieros de datos, al menos 10 científicos de datos de forma concurrente.
335	6. ¿Cantidad de reportes, teniendo en cuenta que un tamaño medio es 400?	Los que se consideren en los Casos de Uso que serían planteados
336	7. ¿Número de horas de ventana de tiempo para procesar todos los reportes en la noche, teniendo en cuenta que un tamaño medio es 8 horas?	La necesidad es realizarlo en tiempo real.
337	1. Cantidad de analistas de negocio, teniendo en cuenta que el tamaño medio es 3.	Tabla 6 - Capa de exploración
338	1. ¿El SRI proveerá la capacidad tecnológica requerida para implementar la plataforma según lo dimensionado por el ofertante, teniendo en cuenta las respuestas al dimensionamiento solicitado?	La propuesta debe considerar la capacidad tecnológica para implementar la plataforma
339	2. ¿El SRI proveerá el hardware con los sistemas operativos y software base sobre el cual se haría la implementación, refiriéndose a software base todo aquello que no contemple la plataforma propuesta?	Se debe definir una vez se tenga la propuesta, en su propuesta podría incluso considerar SQ base para su solución
340	3. ¿En nuestro caso la solución propuesta requiere de un hardware especializado, que es un appliance optimizado para la gestión de datos, cómo debería formularse este costo en la oferta?	Se debe definir una vez se tenga la propuesta, en su propuesta podría incluso considerar SQ base para su solución
341	1. Este servicio tendría que ser obligatoriamente en una nube en el Ecuador? Si la respuesta es NO las siguientes preguntas solo aplican en	Aún no definido, se evalúan opciones.

No.	Pregunta	Respuesta
	el caso de que se prefiera una nube local, ya que en nuestro caso XXXX oferta a nivel mundial estos servicios en su nube con todo lo requerido. Si la respuesta es SI por favor leer las siguientes preguntas:	
342	2. ¿El SRI proveerá en una nube la capacidad tecnológica requerida para implementar la plataforma según lo dimensionado por el ofertante, es decir el servicio de la nube sería mediante un contrato directo entre el SRI y el proveedor de la nube?	No, el costo del servicio de nube debe estar incluido en la propuesta del ofertante.
343	3. ¿Si es el caso en que el ofertante también debiera ofrecer la capacidad tecnológica requerida para implementar la plataforma en la nube, es decir todo como un servicio, como sería el plan de pago ofrecido por la institución? Teniendo en cuenta lo siguiente:	Mensual, por uso consumido
344	3.1 ¿El ofertante tendría que ajustarse a las condiciones estrictas de pago de la empresa proveedora del servicio de la nube, lo cual puede llegar a ser económicamente inviable, como tiene previsto manejar esto?	Presente su propuesta para su análisis y factibilidad
345	3.2 ¿El servicio en la nube es contratado según demanda y facturado de acuerdo con el consumo y ambas cosas son algo que solo se podría analizar y pronosticar después de cierto tiempo de uso, ustedes tienen alguna experiencia al respecto o como pensarían sobre llevar esta realidad?	Puede ser con soporte del proveedor o la propuesta puede considerar capacitación
346	3.3 ¿En nuestro caso la solución propuesta requiere de un hardware especializado, que es un appliance optimizado para la gestión de datos, por lo tanto de ser adjudicados nos veríamos obligados a realizar la compra de este equipo para instalarlo en la infraestructura del proveedor de la nube, pero el tiempo propuesto del servicio de arrendamiento (3 años) no sería suficiente para depreciar el equipo y por lo tanto la inversión no sería factible, ante este escenario cómo pensarían establecer el pago de estos costos?	Puede venir el costo como parte de la propuesta para que la institución adquiera el equipo, siempre que sea on premise.
347	No nos queda claro si el volumen máximo de datos que se ha manejado hasta la fecha	Sección 3.2.1 Información de referencia de volúmenes de datos actuales. Ver Anexo 1 - Volumetría
348	Si se trabajara bajo el mandato de master data management, catálogo de datos ¿o cómo sería el requerimiento?	Se abordará a partir de las necesidades de información para la toma de decisiones, a partir de ahí se administrarán los dominios de información, atributos, reglas de calidad, etc.
349	No se especifica explícitamente si el proyecto se realizará en la nube o es on premise	Forma parte de la propuesta requerida que tipo de solución se plantea (cloud / onpremise)
350	En la página 12. Donde reza "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". Es un proyecto de big data. ¿Se tiene contemplado que el alcance inicie sobre arquitectura de big data o se plantea un data lake que evolucione, pero donde inicialmente lleven los datos que tienen más tangibles?	Sección 3.1 - Capa de almacenamiento
351	"Dice A fin de evaluar las diferentes alternativas en la implementación de la plataforma tipo Data Fabric, el SRI	Si, forma parte de la propuesta requerida el tipo de solución que se plantea

No.	Pregunta	Respuesta
	<p>requiere que las propuestas presentadas consideren los siguientes escenarios"</p> <p>Con esto damos por entendido que se puede presentar una arquitectura hibrida o una arquitectura cloud o una arquitectura onpremise</p>	
352	<p>Donde hace referencia a la "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" nos preguntamos ¿Cuántos flujos existen en el mapeo que quieren migrar? Es prudente ver de qué complejidad son. ¿Estos flujos de datos están en ejecución en la actualidad sobre que tecnología son ejecutados?</p>	<p>La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría</p>
353	<p>Cuando hablan de capa de almacenamiento, ¿existe problema en que la fábrica de datos tenga un alto volumen de datos redundantes?</p>	<p>Sección 3.1.6 Calidad del Dato</p>
354	<p>Cuando hablan de Datawarehouse, ¿Que granularidad de los datos se espera? ¿el procesamiento se espera en bash o existen procesos que se realizan transaccional?</p>	<p>Granularidad Detallada, procesamientos para la integración en batch, actualmente dependiendo de los casos, existen procesos programados a intervalos cortos de 5 minutos.</p>
355	<p>Donde se habla de las versiones de bases de datos Oracle, nos gustaría saber si se piensa migrar a la última versión estable de base de datos en un futuro cercano. En el mismo cuadro se hace referencia la replicación, pero no especifica si se hace con Golden gate, dataguard u otra metodología de replicación.</p>	<p>Réplica al centro de datos alterno por volúmenes de datos</p>
356	<p>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.</p> <p>Confirmar el volumen de información actual.</p>	<p>La solución debe manejar tiempos adecuados, indicar tiempos esperados según la experiencia actual... 3.2.1 Información de referencia de volúmenes de datos actuales</p>
357	<p>Contar con un módulo de Gobierno de Datos que permita manejar y administrar.</p> <p>Cuales son la fuentes de datos actuales</p>	<p>Data semiestructurada (json, XML, csv, txt, etc.), Data estructura (DB relacionales ORACLE). Sección 3.2.1 Información de referencia de volúmenes de datos actuales y Anexo 1 - Volumetría</p>
358	<p>2. Migración y/o reingeniería de datos o flujos de datos más relevante</p> <p>Indicar magnitud de datos a migrar.</p>	<p>La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría</p>
359	<p>La plataforma completa podrá instalarse en nubes privadas y nubes públicas, tanto como PaaS, IaaS o SaaS. También deberá poder instalarse completamente on-premise.</p> <p>Se deben instar en cloud y On premise? Aclarar por favor</p>	<p>La oferta puede considerar cada escenario</p>
360	<p>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.</p> <p>¿Qué porcentaje del proyecto se permitirá desarrollar?</p>	<p>EL desarrollo solo debería ser la personalización o ajustes de parte de la solución que no se tenga en el software del ofertante ese valor debería estimarse con esa brecha.</p>

No.	Pregunta	Respuesta
361	La plataforma debe utilizar Kubernetes y todos los módulos desplegarse en contenedores para ser escalable de manera distribuida. ¿Aplica a pesar de que la solución pueda tener arquitectura diferente, compuesta por múltiples componentes nativos de nubes?	Puede remitirse esa aclaración en la propuesta
362	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. Favor confirmar cantidad aproximada de usuarios	Tabla 6
363	La plataforma debe permitir ser replicada a otra zona de disponibilidad e incluso a otra nube. Al mencionar otra nube, ¿Se refiere a cambiar de proveedor de nube?	SI?, considerar si esto aplica, ya que cada nube es diferente y tiene diferentes herramientas
364	La plataforma/El virtualizador deberá sincronizarse con los dominios definidos en la herramienta de gobierno automáticamente y sin necesidad de intervención manual. Este apartado está mencionando independencia de las plataformas de Gobierno y la de gestión del dato. ¿Indicar si las funcionalidades mencionadas, hablan de una única plataforma o herramientas complementarias a esta, como el módulo de gobierno?	Se habla de una única plataforma o herramienta
365	La plataforma deberá permitir agregar archivos relacionados con los activos, términos, definiciones, relaciones, etc. para una mejor comprensión (PDF, Visio, Word, etc.) Aclarar el tema de archivos.	Archivos de definiciones para gobierno de dato en diferentes formatos
366	La plataforma deberá permitir, para aplicativos de desarrollo propio, el desarrollo de conectores específicos sin modificación del producto. Favor aclarar este punto, brindando mayor nivel de detalle.	Se debe entregar el código fuente para que el SRI haga ajustes sobre la solución o nuevos desarrollos?
367	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. Pueden confirmar volumen de datos.	Los casos de uso serán definidos en el TDR, se espera que sobre los casos de uso se dimensione a mayor detalle los volúmenes de datos
368	la plataforma deberá permitir conectar contra fuentes externas de información para la validación de la información. Detallar los tipos de fuentes externas a considerarse.	Indicar cuales son los tipos de interfaces que se requiere manejar, otras BDD, archivos planos, conexión vía FTP, SFTP, APIS de servicios, WS?
369	La plataforma deberá permitir la corrección de la calidad del dato desde la misma herramienta en el sistema origen. ¿Se refiere a realizar cambios sobre la fuente de datos origen, base de datos origen? Aclarar este punto.	Es deseable, una remediación desde el origen.

No.	Pregunta	Respuesta
370	La plataforma debe permitir el consumo y publicación de datos entre unidades, en función de roles y accesos asignados. Se considera marketplace como un componente integrado de la plataforma o como un módulo adicional externo.	Como parte de la plataforma
371	La plataforma soportará seguridad de grano fino (tablas y columnas) sobre cualquier acceso a datastores (interno o externo). Especificar a que datastores se refiere. A repositorios dentro del almacén de datos analítico (Lakehouse) u otro. Definir... que datos estarían exponiéndose o compartiendo.	Seguridad sobre todo datastore (roles, perfiles)
372	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. ¿Manejan algún tipo de restricción o preferencia para los algoritmos de encriptación?	Forma parte de la propuesta requerida
373	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. Es posible utilizar una interfaz nativa de las diferentes nubes hacia los repositorios DL? O se refieren a una interfaz de la plataforma Data Fabric.	Interfaz plataforma DataFabric
374	¿Cuál sería la fecha que tienen estimado para dar inicio a la ejecución de la solución?	Fecha estimada de adjudicación de contrato mayo 2024
375	Existe algún SLA establecido para la ejecución de soporte al servicio	El soporte sobre el servicio debe ser 24X7, que será considerado posteriormente en el contrato.
376	En el RG-1 se dice: que se permite una combinación híbrida en la nube y en local (on-premise), pero que no se permiten componentes alojados en la nube indicada que enlace a módulos fuera de la nube. Por favor aclarar.	Si se permite incluso el esquema SAAS eso confirma que todo puede estar en la nube, se aclarará esta inconsistencia en los TDR's.
377	¿Definitivamente es un criterio de exclusión lo indicado en el RG-2?, pues es poco probable que haya en el mercado una sola plataforma que cubra todos los requerimientos. ¿Sería posible proponer integrar más de una tecnología? Por ejemplo una para almacenamiento, una para gobierno y otra para la capa analítica?	Si es posible, se analizaría en la propuesta
378	Respecto al RG-3, cuenta como desarrollo los pipelines que habrá que construir para conectarse a las fuentes de datos, transformarlos y llevarlos luego a la capa de almacenamiento?	Debe quedar como plantilla,
379	En el RG-6; ¿A qué se refiere exactamente la flexibilidad para escalar/desescalar, teniendo en cuenta los posibles entornos (on premise, nube e híbrido)	Subir o bajar recursos conforme el uso y concurrencia, estacionalidad
380	1. ¿Cuentan con políticas, estándares y programa de gobierno y calidad de datos?	No
381	2. ¿Su organización ha definido algún tipo de dominio de datos y metadata (data catalog, data glossary, y data dictionary)?	SI un diccionario de datos que está compuesto por dominios y conceptos de información

No.	Pregunta	Respuesta
382	3. ¿Están definidos los roles y perfiles para las personas que tienen interacción directa con los datos y la información?	Si
383	4. ¿Hay algún requisito legal o regulatorio que deba considerarse en la implementación del programa de gobierno de datos?	Ley de Protección de Datos Personales, LOTAIP, Seguridad del dato 27001 e ISO relacionadas
384	En A-2, ¿qué quiere decir con que los servicios deben ser Administrados?	Desarrollo, monitoreo, programación de ejecución de servicios.
385	Como parte del alcance del proyecto se requiere migrar todo el Data Warehouse actual de SRI a la nueva capa de almacenamiento? ¿Esto implicará migrar también todos los modelos analíticos asociados a estos datos? ¿cuántos modelos analíticos se tendrían que migrar?	En primera instancia el DWH actual es una fuente más de la Plataforma de Analítica, pero dependiendo de los casos de uso se puede requerir dicha migración de datos o procesos específicos.
386	¿Para el caso de soluciones on premise deberá especificarse en valor del hardware requerido?	Si, como parte de la oferta
387	Sobre RG-1, Un arquitectura en nube bajo modelo PaaS o SaaS, es un grupo de servicios propios de proveedor de nube, que en su conjunto conforman la totalidad de la solución, por lo que por definición no aplica que estos servicios puedan instalarse en modalidad On Premises. ¿Es impedimento o limitante para la participación el que la arquitectura de la solución sea generada con servicios propios de la nube que no pueden instalarse en modalidad On Premises?	La propuesta puede ser en nube, onpremise o hibrida,
388	Sobre RG-3, "La Solución Software Integral propuesta deberá ser un producto de mercado ya existente y consolidado..." Al ser Data Fabric un concepto de diseño orientado a la generación de arquitecturas tecnológicas con la operación de los datos como eje central; y que los servicios proporcionados por los proveedores de nube, tanto de forma individual como en conjunto, para análisis y gestión de datos, soluciones de mercado probadas y consolidadas. ¿Aplican estos servicios de nube, tanto de forma individual como en conjunto, como soluciones tipo data fabric existentes?	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. Se espera la propuesta en este sentido.
389	Sobre RG-4, "La plataforma debe utilizar Kubernetes y todos los módulos desplegarse en contenedores..." Sobre RG-7, "La plataforma debe ser Fault Tolerant en todos sus elementos y estar construida sobre kubernetes y contenedores..." Se establece que toda la solución debe ser virtualizada e instalada en contenedores, limitando o impidiendo con esto el uso de servicios (serverless) en nube bajo modalidad SaaS, para la generación de una arquitectura data fabric que consuma los servicios a medida de las necesidades y requerimientos ¿Es Correcto esta conclusión?	Enviar propuesta

No.	Pregunta	Respuesta
390	¿Para la demo, puede el SRI indicar el o los casos de uso para demostrar?	Los casos de uso se definirán en los TDR, la demo puede considerar casos de uso similares de administraciones tributarias, orientados a la prevención del fraude, detección de comportamientos, identificación de riesgos tributarios, etc.
391	¿Para la demo, puede el SRI indicar las funcionalidades o escenarios que quisiera validar?	Si
392	¿Puede el SRI indicar si para la demo proporcionara datos de prueba o si el proponente debe utilizar datos propios?	El ofertante deberá utilizar datos propios (reales o de prueba) considerando la volumetría estimada que posee el SRI y los objetivos a resolver (Detección de fraude, comportamiento, identificación de riesgos, etc.)
393	¿Se utilizará la base de datos Analítica para tener un uso en temas transaccionales (upserts)?	Existe caso en los que los insights de analítica son de valor para el transaccional, por lo cual si se debe contemplar integración.
394	¿Cuál es su política de retención de los datos (1 año, 2 años, 3 años, etc.)?	Depende de los conceptos de información, en su mayoría cumplen la política de retención documental y caducidad de los procesos determinativos de 7 años
395	Crecimiento estimado de la base de datos año con año (en GB, TB)	Ver Anexo 1 - Volumetría
396	Si la empresa ya cuenta con una arquitectura de información, indicar cuál esquema utiliza actualmente (estrella, snowflake, data vault, otro)	Ver Diagrama de arquitectura de DW
397	Concurrencia planeada de consultas (5 usuarios, 10 usuarios, 100 usuarios, etc. obteniendo datos en el mismo segundo)	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%. 10 usuarios de ingeniería de datos y 20 científicos de datos.
398	Para la información que viene desde IoT o Logs, pf. indicar qué tipos de datos vienen en el archivo resultante, tipo de datos: JSON, XML, CSV, AVRO otro	No se ha identificado en los casos de uso fuentes de tipo IoT, pero es deseable si fuese el caso tener todos los indicados.
399	Qué tipo de consultas realiza actualmente (% del Total) 1. Simples:80% 2. Complejas: 20% 3. Espaciales: 0% 4. JSON:0% 5. series temporales:0% 6. agregados:0%	(% del Total) 1. Simples:69% 2. Complejas Agregadas: 30% 3. Espaciales: 0% 4. JSON:1% 5. series temporales:0%
400	¿Cuál es la frecuencia de carga actual?	Desde intervalos cortos: 5 minutos, por hora, diaria, semanal, mensual.
401	¿La Analítica Avanzada será procesada en tiempo real o se planifican modelos para ejecución periódica?	Forma parte de la propuesta requerida
402	1. En el punto 3.4.4 se solicita compatibilidad con el sistema de gestión de identidades del SRI, pero no se especifica cuál es el sistema de gestión de identidades. ¿Es posible conocer los detalles de esta solución? ¿En caso de que no exista, es necesario incluir una solución de gestión de identidades para esta solución?	Se cuenta con IBM (IDM) on premise, sin embargo puede presentar su propuesta
403	2. En el punto 3.4.5 se solicita especificar el esquema para la gestión de eventos de seguridad. La solución puede integrarse con sistemas para la gestión de eventos de seguridad (siem). ¿Esta integración estará a cargo del equipo implementador del data fabric? ¿Es posible saber cuál es el nombre de la solución?	Se encuentra normado el proceso de gestión de eventos de seguridad, no se posee una herramienta informática implementada.
404	3. Con respecto a la seguridad de información y protección de datos personales, es necesario que la solución cumpla con algún estándar específico? ¿Es necesario implementar	Si, ISO 27002:2013 e ISO 27002:2022 e ISO 27701 (on premise) y la ISO 27017 e ISO 27018 (nube)

No.	Pregunta	Respuesta
	controles o herramientas para proteger los datos almacenados en la solución?	
405	4. Con respecto al control de acceso, la solución deberá contar con algún tipo de control de acceso basado en riesgo o implementar un segundo factor de autenticación?	Dentro de las medidas de seguridad la solución debe contar al menos con doble factor de autenticación, y que permita la integración con el Directorio Activo integrado con Microsoft Azure del SRI
406	1. De acuerdo al ítem 1.2 Cronograma, se indica que la fecha límite para entregar las respuestas al RFI, es el 29 de mayo. Se solicita extender el plazo de entrega considerando la complejidad y factores que serán confirmados hasta el 19 de mayo por parte del SRI. De ser posible, considerar como fecha de entrega, el viernes 9 de junio. Atentos a su consideración y respuesta.	Se publicará boletín de enmiendas con el nuevo cronograma actualizado.
407	2. Favor confirmar que las respuestas como oferentes, son todas aquellas incluidas en el ítem 5, 6 y Anexo1. En caso de requerir respuesta de un ítem adicional, favor indicar.	Si, hay que considerar que el anexo 1 también recoge la validación de funcionalidades de los ítems 3 y 4
408	3. A lo largo del documento se habla de una "plataforma", compuesta por varios módulos/productos que satisfacen, de forma independiente con los requerimientos del SRI reflejados y divididos en ítems (3.1.1 al 3.1.17), de ahí que, se solicita evaluar el uso del término "plataforma" en la descripción de la necesidad, respetando la funcionalidad que tiene cada módulo/producto que formará parte de la plataforma de gobierno y analítica de datos, con el fin de evitar posibles errores de interpretación en todos los procesos contractuales e inclusive, posteriores a ellos (auditorías). Confirmar que nuestra solicitud será analizada.	Si es posible, se analizaría en la propuesta
409	4. En el ítem 2.3 Necesidades específicas, se menciona que se considere un escenario en nube. Favor confirmar que la nube ofertada sea una nube pública con sus centros de datos en el exterior o en territorio ecuatoriano, tanto para las capacidades de procesamiento como para los datos.	En el TDR se detallarán las condiciones definitivas de implementación. Envíe su propuesta considerando tal escenario.
410	5. De acuerdo al ítem "Requerimientos de la Solución" punto 3 – Servicios de gestión de datos - se solicita: "Facilitar el manejo y explotación de la información institucional, así como la gestión de datos históricos de la entidad", favor confirmar ubicación, formato y capacidad aproximada de los datos históricos	Sección: 3.2.1 Información de referencia de volúmenes de datos actuales; esta sección incluye versiones de base de datos. Ver Anexo 1 - Volumetría
411	6. Según ítem 3.4.1 que indica "detallar la arquitectura tecnológica", favor confirmar que, a) el proveedor debe indicar los recursos de infraestructura requeridos para la implementación local de la solución propuesta, o que, b) se incluyan los recursos tecnológicos como parte de la propuesta económica.	b) se debe incluir como parte de la propuesta la infraestructura tecnológica
412	7. En el ítem 6: Tecnológicos, se indica "debe detallar en su propuesta la arquitectura tecnológica para su implementación, sea esta on premise o cloud", favor confirmar que al referirse al término "cloud", hace referencia a cloud pública en territorio ecuatoriano o fuera de él	Si, nube pública
413	8. En el ítem 3.4.1 se indica "Adicionalmente la solución debe tener un esquema de alta disponibilidad", favor confirmar que para cumplir con este requerimiento se puede contar con espacio y/o recursos tecnológicos en los centros de datos Quito y Guayaquil del SRI.	Se debe incluir como parte de la propuesta la infraestructura tecnológica

No.	Pregunta	Respuesta
414	9. En el ítem 3.4.1 se indica "Adicionalmente la solución debe tener un esquema de alta disponibilidad y alto rendimiento", favor confirmar que a) el SRI cuenta con infraestructura tecnológica de alto rendimiento o b) el proveedor incluirá, además del sistema para gobierno de datos y analítica, equipamiento de alto rendimiento.	b); se debe incluir como parte de la propuesta la infraestructura tecnológica
415	10. En el ítem 3.4.1 se indica "Se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia", favor confirmar que son ambientes idénticos entre sí en funcionalidad y capacidad	Contingencia es réplica de producción; pruebas es ambiente de certificación y pruebas de carga previo producción y desarrollo es similar en estructuras de datos pero inferior en capacidad.
416	11. En el ítem 3.4.1 se indica "Se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia", favor confirmar que el ambiente de Contingencia protegerá únicamente el ambiente de Producción.	Contingencia es la réplica de producción
417	12. En el ítem 3.4.1 se indica "Se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia", favor confirmar el nivel de contingencia al que hace referencia: a) plataforma, b) datos, c) plataforma + datos.	Replica idéntica de producción en plataforma y datos
418	13. En el ítem 3.4.1, se indica "Los ambientes deben estar 100% aislados el uno del otro (física o lógicamente)", favor confirmar que la arquitectura de segmentación física debe ser provista por el contratista o será provista por el SRI.	El SRI dispone de un centro de datos principal en UIO y un centro de datos alterno en GYE, teniendo en UIO producción y en GYE contingencia, preproducción y desarrollo
419	14. El ítem 3.4.9 cita "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	SI
420	15. Se menciona que se tendrán FUENTES de datos internos y externos. Favor confirmar.	Confirmado
421	16. Con respecto al punto 15, ¿se tienen identificadas y es posible contar con una cuantificación de estas fuentes de datos?	Fuentes internas: BDD Transaccionales, BDD DW, Repositorios, Frontera Fuentes externas: BDD o información de dominio público, páginas web, redes sociales, etc. La conexión o integración de fuentes será escalable conforme los casos de uso planteados.
422	17. Confirmar cuáles son los DOMINIOS que se identifican para las actividades de gobierno de datos (calidad, refinera, transformación, etc.)	La Gestión de Datos maestros es fundamental para poner a disposición de los usuarios de manera que permita crear modelos analíticos y toma de decisiones a partir de los datos. Con la implementación del Gobierno se podrán identificar y optimizar fuentes duplicadas, adicional a las reglas de calidad de datos que permitan identificar duplicidad, deben ser herramientas de fácil manejo para usuarios no técnicos y gráficas con reportes que permitan revisar las acciones realizadas sobre los datos
423	18. En el ítem GD-36, entendemos que se solicita el levantamiento de un modelo ontológico, este levantamiento implica un servicio consultivo que se recomienda sea realizado de forma independiente a la implementación del modelo de gobernanza, ¿es posible tener esta consideración en cuenta?	No

No.	Pregunta	Respuesta
424	19. ¿Para la implementación de las plataformas está considerado proponer una arquitectura objetivo? a. ¿El SRI dicta la arquitectura objetivo? b. ¿Se espera que el proveedor plantee la arquitectura objetivo? c. ¿Dentro de las fases de proyecto se tiene consideración de tener arquitecturas transitorias de acuerdo a las fases de implementación?	Se espera b) y c)
425	20. No se muestra una estrategia, ¿metodología donde se muestre una secuencia de hitos, duración y a su vez cuales entregables irán con cada hito? ¿Se considera que el proveedor proponga las etapas, hitos y mapa de ruta para el proyecto?	Si, se espera como parte de la propuesta requerida
426	21. En la capa de Gobierno, se contempla crear y/o programar las reglas de control de calidad de datos, bajo código, así como otro lenguaje gráfico por bloques? a. El SRI definirá las reglas de calidad que apliquen a cada dataset, cada campo (concepto, umbrales) y el equipo del proveedor programaría esas reglas en las plataformas. ¿Es correcto?	Si
427	22. ¿Se tiene considerado obtener indicadores y/o estadísticos de calidad de los datos durante el proceso de calidad? ¿Cuáles serían los resultados esperados? a. ¿Por ejemplo tener una línea base de un 70% de calidad y llegar un 95% en los datos durante el ciclo de calidad?	Se pretende determinar una línea base de calidad a partir de los casos de uso a implementar, posterior se implementarán las metas a conseguir anualmente
428	23. ¿Se espera que toda la conformación de arquitectura holística sea de un solo fabricante o marca? ¿O se considera que sean de integrados varios productos y/o marcas distintas integrados?	Se espera de un solo fabricante
429	24. En la capa de Gobierno se contempla tener ciclos de refinera de datos, ¿es decir que se pueda considerar hacer desde limpieza, transformación, rectificación o unificación de datos? (Ciclos de calidad) a. ¿Cuántos ciclos de calidad se tienen contemplados	Si, el número será determinado en la implementación de los casos de uso priorizados
430	25. En el ítem 3.2 se solicita 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", en ese sentido favor indicar la siguiente información:	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
431	Información de ambiente A) Cantidad de objetos definidos en el DW actual (tablas) B) Cantidad de procesos a manera de SP C) Cantidad de reportes D) Cantidad de modelos / cubos E) Proyección	Como referencia, existen alrededor de 100 productos DW en su carga intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como caso de uso.

No.	Pregunta	Respuesta																																																				
432	<p>Procesos</p> <p>F) Promedio de ejecución de procesos concurrentes</p> <p>G) Categorización porcentual de procesos</p> <ul style="list-style-type: none"> • % de procesos simple (hasta 500 líneas) • % de procesos medio (mayor a 500 hasta 1500 líneas) • % de procesos alto (mayor a 1501 líneas) <p>H) Línea base de tiempos de ejecución de procesos</p> <p>I) Tiempo de ejecución ordinario</p> <p>J) Tiempo de ejecución picos (quincena / fin de mes / o cuando el cliente considere picos de procesamiento)</p>	<p>G)</p> <ol style="list-style-type: none"> 1. Simples:69% 2. Complejas Agregadas: 30% 3. JSON:1% <p>Al menos 7 instancias de BBD donde están las fuentes. Existen alrededor de 1.800 usuarios de consulta en la plataforma actual, con concurrencia del 10 al 15% H), I, J) Ver anexo 1 - Volumetría</p> <p>Se espera ejecución en tiempo real</p>																																																				
433	<p>Usuarios</p> <p>K) Usuarios concurrentes estimados durante la hora pico?</p> <p>L) Qué tipo de conexión es la más usada? (JDBC, ODBC, .Net)</p>	<p>K) Se tienen alrededor de 1.800 usuarios de consulta con una concurrencia del 10 al 15%</p> <p>L) Se tiene con JDBC y ODBC.</p>																																																				
434	<p>Ambiente de trabajo</p> <p>M) Herramientas actuales</p> <table border="1"> <thead> <tr> <th>Nombre</th> <th>Versión</th> <th>Conector</th> <th>Descripción de uso</th> </tr> </thead> <tbody> <tr> <td>Visualización</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Integración</td> <td></td> <td></td> <td></td> </tr> <tr> <td>Análítica</td> <td></td> <td></td> <td></td> </tr> <tr> <td>DS</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Nombre	Versión	Conector	Descripción de uso	Visualización				Integración				Análítica				DS				<p>SAP Business Objects, Oracle Discoverer, Qlik, Apache Superset, Saiku</p> <table border="1"> <thead> <tr> <th>Uso</th> <th>Nombre</th> <th>Versión</th> <th>Conector</th> </tr> </thead> <tbody> <tr> <td>Visualización/Capa lógica/ ETL (Data Services)</td> <td>SAP Business Objects</td> <td>14.1.4.1327</td> <td>cliente Oracle</td> </tr> <tr> <td>Visualización/Capa lógica</td> <td>Oracle Discoverer</td> <td>10g</td> <td>cliente Oracle</td> </tr> <tr> <td>Visualización</td> <td>Saiku</td> <td>2.6.1</td> <td>jdbc</td> </tr> <tr> <td>Visualización</td> <td>Qlik</td> <td>12.1</td> <td>cliente Oracle</td> </tr> <tr> <td>Dashboards</td> <td>Apache Superset</td> <td>8,3</td> <td>cx_oracle</td> </tr> </tbody> </table>	Uso	Nombre	Versión	Conector	Visualización/Capa lógica/ ETL (Data Services)	SAP Business Objects	14.1.4.1327	cliente Oracle	Visualización/Capa lógica	Oracle Discoverer	10g	cliente Oracle	Visualización	Saiku	2.6.1	jdbc	Visualización	Qlik	12.1	cliente Oracle	Dashboards	Apache Superset	8,3	cx_oracle								
Nombre	Versión	Conector	Descripción de uso																																																			
Visualización																																																						
Integración																																																						
Análítica																																																						
DS																																																						
Uso	Nombre	Versión	Conector																																																			
Visualización/Capa lógica/ ETL (Data Services)	SAP Business Objects	14.1.4.1327	cliente Oracle																																																			
Visualización/Capa lógica	Oracle Discoverer	10g	cliente Oracle																																																			
Visualización	Saiku	2.6.1	jdbc																																																			
Visualización	Qlik	12.1	cliente Oracle																																																			
Dashboards	Apache Superset	8,3	cx_oracle																																																			
435	<p>N) Carga de trabajo (casos de uso): Del 100% de la capacidad de uso de la plataforma, como se distribuirá su uso y aprovechamiento por tipo de carga</p> <table border="1"> <thead> <tr> <th>Tipo de carga</th> <th>Porcentaje de uso</th> <th>Horario (diaria / semanal / mensual)</th> <th>Tipo de ejecución (Batch / Streaming / Near real time)</th> </tr> </thead> <tbody> <tr> <td>BI</td> <td></td> <td></td> <td></td> </tr> <tr> <td>INGENIERIA DE DATOS</td> <td></td> <td></td> <td></td> </tr> <tr> <td>MINERIA DE DATOS</td> <td></td> <td></td> <td></td> </tr> <tr> <td>MACHINE LEARNING</td> <td></td> <td></td> <td></td> </tr> <tr> <td>IA</td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Tipo de carga	Porcentaje de uso	Horario (diaria / semanal / mensual)	Tipo de ejecución (Batch / Streaming / Near real time)	BI				INGENIERIA DE DATOS				MINERIA DE DATOS				MACHINE LEARNING				IA				<p>Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja). Ver Anexo 1 - Volumetría. Presente su propuesta</p> <table border="1"> <thead> <tr> <th>Tipo de carga</th> <th>% uso</th> <th>Horario</th> <th>Tipo de ejecución</th> </tr> </thead> <tbody> <tr> <td>BI</td> <td>20</td> <td>Diario</td> <td>Near real Time</td> </tr> <tr> <td>Gobierno de Datos</td> <td>15</td> <td>Diario</td> <td>Near real Time</td> </tr> <tr> <td>Ingeniería Datos</td> <td>15</td> <td>Diario</td> <td>Streaming/Batch</td> </tr> <tr> <td>Minería Datos</td> <td>20</td> <td>Diario</td> <td>Batch</td> </tr> <tr> <td>Machine Learning</td> <td>15</td> <td>Diario</td> <td>Batch</td> </tr> <tr> <td>IA</td> <td>15</td> <td>Diario</td> <td>Near real Time</td> </tr> </tbody> </table>	Tipo de carga	% uso	Horario	Tipo de ejecución	BI	20	Diario	Near real Time	Gobierno de Datos	15	Diario	Near real Time	Ingeniería Datos	15	Diario	Streaming/Batch	Minería Datos	20	Diario	Batch	Machine Learning	15	Diario	Batch	IA	15	Diario	Near real Time
Tipo de carga	Porcentaje de uso	Horario (diaria / semanal / mensual)	Tipo de ejecución (Batch / Streaming / Near real time)																																																			
BI																																																						
INGENIERIA DE DATOS																																																						
MINERIA DE DATOS																																																						
MACHINE LEARNING																																																						
IA																																																						
Tipo de carga	% uso	Horario	Tipo de ejecución																																																			
BI	20	Diario	Near real Time																																																			
Gobierno de Datos	15	Diario	Near real Time																																																			
Ingeniería Datos	15	Diario	Streaming/Batch																																																			
Minería Datos	20	Diario	Batch																																																			
Machine Learning	15	Diario	Batch																																																			
IA	15	Diario	Near real Time																																																			
436	<p>O) Proyección de crecimiento de datos año a año para los próximos 5 años</p> <p>P) Se tiene proyecto incluir datos adicionales</p> <ul style="list-style-type: none"> • Tipo de datos / Tamaño proyectado 	<p>Ver Anexo 1 - Volumetría</p>																																																				

No.	Pregunta	Respuesta
437	<p>26. Para efectos de licenciamiento de productos en modalidad OPN PREMISES es necesario contar con la siguiente información:</p> <p>a. ETL</p> <p>i. Cantidad de horas para mover los datos (Ventana de mantenimiento)?</p> <p>ii. Cantidad de procesadores (ETL Actual)</p> <p>1. Favor especificar la cantidad de procesadores utilizados actualmente</p> <p>iii. Cantidad de jobs actuales?</p> <p>iv. Tiempo que toma actualmente correr todos los jobs ETL?</p>	<p>Actualmente existen desarrollos de flujos mediante SAP Data Services, Oracle Warehouse Builder, Paquetes PL/SQL que se ejecutan mediante Jobs de BDD, Chrones, Paquetes PL/SQL, flujos de SAP DS, etc. en forma continua durante todo el día. Periodicidad de ejecución: desde intervalos cortos: 5 minutos, por hora, diaria, semanal, mensual. Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso. El tiempo de ejecución del ETL puede variar de entre pocos minutos hasta dos o tres días. LA cantidad actual de procesadores en la parte DWH es de 12.</p>
438	<p>b. CIENCIA DE DATOS</p> <p>i. Cantidad de horas de procesamiento al mes para entrenamiento y desarrollo de modelos</p> <p>ii. Tamaño de ambiente a usar</p>	<p>i. Al menos 900 horas</p> <p>ii. Ver Anexo 1 - Volumetría</p>
439	<p>c. GOBIERNO</p> <p>i. Cantidad de usuarios concurrentes, es decir conectados simultáneamente</p> <p>ii. ¿Cuántas tablas desean catalogar?</p> <p>iii. ¿Cuántas columnas en promedio tienen sus tablas?</p> <p>iv. ¿Cuántos registros en promedio tienen sus tablas?</p> <p>v. En lo referente a artefactos de gobierno de información, cantidad de: categorías, términos en el glosario, conjuntos de datos de referencia, clases de datos y clasificaciones.</p> <p>vi. ¿Cuántas columnas se esperan perfilar diariamente?</p> <p>vii. ¿Cuántas reglas de calidad se esperan ejecutar diariamente?</p> <p>viii. ¿Cuántas columnas nuevas se desean importar diariamente?</p> <p>ix. ¿Cuántas interacciones de linaje se esperan ejecutar diariamente?</p> <p>x. ¿Cuántos objetos con información de linaje(activos) se espera tener? (CONSIDERANDO)</p> <p>1. Cuántos objetos con importación de linaje se esperan tener: una importación de linaje importa tablas, procedimientos almacenados, reportes, consultas. La pregunta hace referencia a un valor total aproximado de cantidad de todos estos objetos, un valor único, no hace falta total de tablas o de procedimientos, sino un valor total.</p>	<p>Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%. 10 usuarios de ingeniería de datos y 20 científicos de datos.</p> <p>Ver anexo 1 Volumetría.</p>
440	<p>xi. ¿Cuántos científicos o analistas de datos concurrentes van a hacer preparación de datos desde la plataforma? (CONSIDERANDO)</p> <p>1. La preparación de datos es equivalente a modificar, crear, borrar columnas de Excel, utilizar formulas, entre otros.</p>	<p>Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%. 10 usuarios de ingeniería de datos y 20 científicos de datos.</p>
441	<p>d. BI</p> <p>i. SRI requiere una solución de BI, o va a utilizar sus herramientas actuales como Qlik o Power BI</p> <p>ii. Cantidad de Usuarios Viewer</p> <p>iii. Cantidad de Usuarios Analytics</p> <p>iv. Cantidad de Usuarios Explorador</p> <p>v. Cantidad de Usuarios Administradores</p> <p>vi. concurrencia, % de usuarios conectados simultáneamente</p> <p>vii. Cantidad de reportes</p>	<p>Se esperan propuestas de soluciones. Alrededor de 1800 usuarios viewer, Entre 100 y 200 analytics, 50 exploradores, 2 administradores. Se esperaría no tener ventanas de tiempo para el procesamiento de reportes. La cantidad de reportes se plantearía en los casos de uso. Alrededor del 10 al 15% de usuarios concurrentes.</p>

No.	Pregunta	Respuesta
	viii. Número de horas de ventana de tiempo para procesar todos los reportes en la noche	
442	e. VIRTUALIZACION DE DATOS i. Cantidad de Querys simples (menores a 1 minuto), moderados (entre 1 y 10 minutos) y complejos (mayores a 10 minutos) por hora de todas las bases a obtener la información. ii. Cuantas Bases de datos se van a consultar iii. Cuantos usuarios concurrentes	(% del Total) 1. Simples:69% 2. Complejas Agregadas: 30% 3. JSON:1% Al menos 7 instancias de BBD donde están las fuentes. Existen alrededor de 1.800 usuarios de consulta en la plataforma actual, con concurrencia del 10 al 15%
443	f. MDM i. Qué dominios necesita implementar? ii. ¿Cantidad de registros por dominio? iii. ¿Cantidad de fuentes a cruzar para obtener el golden record y cuáles? iv. Cantidad de transacciones por segundo a consultar los datos	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja). Ver Anexo 1 - Volumetría. Los dominios son: Gestión de Servicios tecnológicos, Productos de Datawarehouse, Gestión de la Documentación, Devolución de Impuestos, Gestión de canales de atención, Registro del Sujeto (pasivo, beneficiario, vehículos), Declaraciones, Anexos, Vehículos, Comprobantes, Auditoría Tributaria, Gestión de Mercancías, Gestión de Infracciones, Gestión de Cobro, Recursos de Revisión, Juicios Tributarios, Recaudación, Declaraciones, Anexos, Registro del Sujeto (pasivo, beneficiario, vehículos), Fraude Fiscal, Acuerdos Internacionales de Intercambio de Información Tributaria, Gestión del Talento Humano, Indicadores de Gestión, Previsiones y Estadísticas, Seguridad de la Información, Gestión de Proyectos, Registro de catálogos transversales, Gestión de Talento Humano, Gestión Financiera, Gestión Administrativa
444	g. LINAJE i. ¿Desde cuantas y cuales fuentes se debe extraer metadatos para armar el linaje?	El número será determinado en la implementación de los casos de uso priorizados
445	h. MRM i. Cantidad de modelos a monitorear	En principio 5, 3 priorizados y 2 implementados con acompañamiento del proveedor, posterior deberán ser todos los modelos a implementarse
446	i. REPLICACION CDC i. Cantidad de fuentes y cuantos Procesadores de cada fuente a sincronizar	Ver Anexo 1 - Volumetría.
447	j. PIPELINES / FLUJOS i. Cantidad de flujos concurrentes a ejecutarse	Ver Anexo 1 - Volumetría.
448	k. GENERAL i. ¿Se necesitan ambientes no productivos? ¿Cuáles y cuantos?	Si, ambientes de desarrollo, preproducción, contingencia y producción
449	27. Para efectos de licenciamiento de productos en modalidad CLOUD es necesario contar con la siguiente información: a. DATAWAREHOUSE i. Cantidad de horas al día que se necesita la plataforma activa	24 x7

No.	Pregunta	Respuesta
450	b. ETL i. Cantidad de procesadores (ETL Actual) ii. Cantidad de jobs actuales iii. Tiempo que toma actualmente correr todos los jobs ETL	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso. El tiempo de ejecución del ETL puede variar de entre pocos minutos hasta dos o tres días.
451	c. CIENCIA DE DATOS i. Tamaño de ambiente a usar	Al menos 900 horas
452	d. DATALAKE i. Cantidad de GB de información no estructurada a almacenar ii. Cantidad de horas de procesamiento requeridas	Ver Anexo 1 - Volumetría.
453	e. GOBIERNO i. En lo referente a artefactos de gobierno de información, cantidad de: categorías, términos en el glosario, conjuntos de datos de referencia, clases de datos y clasificaciones, CONSIDERANDO: 1. Una categoría es una entidad que contiene términos de negocio relacionados, por ejemplo categoría clientes, ventas, contabilidad, etc.) 2. Un conjunto de datos de referencia es un catálogo, por ejemplo de provincias con su código, códigos de tipos de producto, etc. que son utilizados en el perfilamiento de datos para identificar clases de datos. 3. Las Clases de datos ayudan a identificar por ejemplo emails, cédulas, códigos de producto, ciudades, provincias, y cualquier otro dato dentro de los orígenes de información, pueden ser identificadas por diversos mecanismos (expresiones regulares, listas de valores válidos, conjuntos de datos de referencia) 4. Una clasificación permite clasificar activos de información como información que identifica a una persona, información sensible, información confidencial o cualquier otra categoría que se desee crear.	<p>Existe un levantamiento de Dominios de información que son 29, conceptos de información que son 1576, aún no se han asociado atributos, categorías, datos de referencia, etc. En relación con la clasificación de la Información se tiene: Pública, Reservada Alta, Reservada Media, Reservada Baja y Dato Personal Sensible.</p> <p>Se deberá realizar la validación y actualización de esta clasificación para los objetos de los casos de uso</p>
454	ii. Cantidad de horas de procesamiento para actividades de perfilamiento y preparación de datos.	Baja complejidad: Al menos 16 horas Media complejidad: Al menos 500 horas Alta complejidad: Al menos 1000 horas
455	f. BI i. SRI requiere una solución de BI, o va a utilizar sus herramientas actuales como Qlik o Power BI	Si se requiere
456	ii. Cantidad de Usuarios Viewer 1. Los miembros pueden leer el contenido público. Por ejemplo, pueden suscribirse a los reportes y ver dashboards. Link: https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=roles-license	Entre 1200 y 1800 usuarios
457	iii. Cantidad de Usuarios Analytics 1. Los miembros pueden crear nuevos reportes, dashboards, exploraciones, historias, nuevos trabajos, conexiones de servidor de datos y orígenes o módulos de datos. Pueden ejecutar informes, responder a solicitudes y cargar archivos. Link: https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=roles-license	Entre 100 y 200 usuarios

No.	Pregunta	Respuesta
458	iv. Cantidad de Usuarios Explorador 1. Super Usuario: Puede crear nuevos modelos (cubos) Los miembros tienen los mismos permisos de acceso que los usuarios de Analytics. También pueden acceder a Exploration, Planning Analytics para Microsoft Excel, Cognos Framework Manager, Cognos Cube Designer, Dynamic Query Analyzer, Jupyter Notebook y Transformer. Este usuario no puede crear nuevos modelos Link: https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=roles-license	Al menos 50 usuarios
459	v. Cantidad de Usuarios Administradores 1. Los miembros tienen los mismos permisos de acceso que los Usuarios Exploradores. También pueden acceder a IBM Software Development Kit; y los componentes del menú Gestionar, incluido IBM Cognos Administration. Link: https://www.ibm.com/docs/es/cognos-analytics/11.1.0?topic=roles-license	Mínimo 2
460	vi. concurrencia, % de usuarios conectados simultáneamente vii. Cantidad de reportes viii. Número de horas de ventana de tiempo para procesar todos los reportes en la noche	vi. Se tiene alrededor de 1.800 usuarios de consultas con una concurrencia del 10 al 15% vii. Depende de los que se definan por caso de uso. viii. La necesidad es realizarlo en tiempo real
461	g. VIRTUALIZACION DE DATOS i. Cantidad de Querys simples (menores a 1 minuto), moderados (entre 1 y 10 minutos) y complejos (mayores a 10 minutos) por hora de todas las bases a obtener la información. ii. Cuantas Bases de datos se van a consultar iii. Cuantos usuarios concurrentes	(% del Total) 1. Simples:69% 2. Complejas Agregadas: 30% 3. JSON:1% Al menos 7 instancias de BBD donde están las fuentes. Existen alrededor de 1.800 usuarios de consulta en la plataforma actual, con concurrencia del 10 al 15%
462	h. MDM i. Qué dominios necesita implementar?	Gestión de Servicios tecnológicos, Productos de Datawarehouse, Gestión de la Documentación, Devolución de Impuestos, Gestión de canales de atención, Registro del Sujeto (pasivo, beneficiario, vehículos), Declaraciones, Anexos, Vehículos, Comprobantes, Auditoria Tributaria, Gestión de Mercancías, Gestión de Infracciones, Gestión de Cobro, Recursos de Revisión, Juicios Tributarios, Recaudación, Declaraciones, Anexos, Registro del Sujeto (pasivo, beneficiario, vehículos), Fraude Fiscal, Acuerdos Internacionales de Intercambio de Información Tributaria, Gestión del Talento Humano, Indicadores de Gestión, Previsiones y Estadísticas, Seguridad de la Información, Gestión de Proyectos, Registro de catálogos transversales, Gestión de Talento Humano, Gestión Financiera, Gestión Administrativa
463	i. LINAJE i. ¿Desde cuantas y cuales fuentes se debe extraer metadatos para armar el linaje?	El número será determinado en la implementación de los casos de uso priorizados
464	28. En el ítem "Capa de Integración" se indica "Tanto la capacidad de replicación en tiempo real en un datastore propio como el acceso a la información a través de un virtualizador..." Favor confirmar que, al solicitar la capacidad de replicación en tiempo real, se refiere al procesamiento de datos en su fuente origen sin necesidad de moverlo al repositorio.	Revisar requerimiento V-1

No.	Pregunta	Respuesta
465	29. En el ítem 3.4.8 se indica "Su propuesta debe considerar que existirá integración con aplicaciones que están fuera del alcance de la contratación y existentes en la Institución". Para cumplir con el requerimiento, favor mencionar todas las aplicaciones con sus respectivas versiones y ediciones que están fuera del alcance de esta contratación y existen en el SRI.	Integración de datos y no de aplicaciones
466	30. En el ítem 3.4.8 se indica "Su propuesta debe considerar que existirá integración con aplicaciones que están fuera del alcance de la contratación y existentes en la Institución". Para cumplir con el requerimiento, favor mencionar todas las aplicaciones con sus respectivas versiones y ediciones que están dentro del alcance de esta contratación y existen en el SRI.	Integración de datos y no de aplicaciones
467	31. En el ítem 3.4.9 se menciona "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", entendiéndose que los ajustes solicitados pueden variar en complejidad y plazo, favor confirmar si el modelo de contratación está abierto para que, de acuerdo al alcance de servicios, orientados netamente para cambios derivados de modificaciones normativas u otros, se consideren como una adenda al contrato principal, con su respectiva definición de alcance y pago, al tratarse de posibles modificaciones sustanciales al sistema y sus componentes de apoyo. En el caso que la opción no sea una adenda al contrato, favor especificar la modalidad de cubrir con los esfuerzos (servicios profesionales) invertidos por el proveedor.	Se espera que se envíe en la propuesta
468	32. En el ítem 3.4.10 se indica "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...", favor confirmar el tiempo máximo después de la firma de contrato para la entrega/recepción de licencias.	Se pretende contar con garantía mínima de 3 años a partir de la entrega/recepción del último entregable o módulo implementado.
469	33. En el ítem 3.4.10 se indica "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". De acuerdo a nuestro conocimiento, EOL es un certificado expedido para dispositivos de hardware mas no para software. Favor confirmar para qué ítems aplica el certificado EOL solicitado.	Se refiere al hardware y al software.

No.	Pregunta	Respuesta
470	<p>34. En el ítem 4 se indica -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".</p> <p>Favor especificar el alcance del acuerdo en términos generales de contenido, cantidad de horas empleadas, número de personas que participan, lugar/es donde se dictarán las capacitaciones, perfil esperado del instructor/es.</p>	Incluir el detalle como parte de la propuesta.
471	<p>35. En el ítem 4 se indica "El entrenamiento técnico y funcional deberá realizarse considerando enfoques metodológicos en marcos de referencia como e-learning u otras que estén internacionalmente reconocidas...", favor confirmar que el marco de referencia esperado esté guiado por las buenas prácticas de la industria y del fabricante/s incluidos en la solución ofertada.</p>	<p>DAMA – Gobierno</p> <p>CRISP - DM Analítica</p>
472	<p>36. En el ítem 4 se indica "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." Favor confirmar que el entregable de este requerimiento es el procedimiento empleado en la migración de flujos de trabajo de los proyectos de minería de datos actuales (proyectos cuya información se solicita).</p>	<p>La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría</p>
473	<p>37. En el ítem 3.1.17 Ref SYC-3 se indica "El proponente cuenta con un programa de certificación y formación para el uso de la plataforma tecnológica".</p> <p>Favor confirmar si el programa de certificación y formación hace referencia al personal del SRI. En el caso que la respuesta sea asertiva, favor especificar que la certificación/formación sean impartida por el proveedor</p>	<p>Se requiere validar si es el oferente quien cuenta con un programa de certificación y formación para el uso de la plataforma tecnológica que se proponga para el personal del SRI.</p>
474	<p>38. En el ítem 3.5.1 se indica "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." Entendiendo que la plataforma para Gobierno de Datos y Analítica Institucional estará conformada por varios componentes de software y hardware (on site o cloud), se entiende que la característica requerida, hace referencia a la Disponibilidad de la plataforma, favor confirmar nuestro entendimiento.</p>	Sí
475	<p>39. En el ítem RG-3 se indica "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." Favor confirmar que para sustento de este requerimiento es posible:</p> <p>a. Considerar implementaciones a nivel internacional b. Considerar implementaciones de la solución integral o</p>	<p>La solución tipo Data Fabric se requiere que sea completa e integrada, presente su propuesta.</p>

No.	Pregunta	Respuesta
	parcial c. Considerar implementaciones del fabricante/s o del proveedor (o proveedores en caso de asociación).	
476	40. En el ítem RG-4 se indica "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 procesado 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". Favor confirmar que si Openshift está soportado dentro de la matriz de compatibilidad/interoperatividad entre los módulos/productos que conforman la plataforma de gobierno-analítica de datos y su base de contenedores, se lo puede incluir como parte de la propuesta.	Cualquier similar a kubernetes (contenedores)
477	41. En el ítem RG-6 se indica "La plataforma debe ser modular y lo suficientemente flexible para escalar/desescalar con el tiempo". Favor confirmar que el factor de escalamiento flexible está sujeto a consideraciones de facturación variada, ya sea en modalidad cloud como en modalidad on premise.	Confirmado
478	42. En el ítem RG-6 se indica "La plataforma debe ser modular y lo suficientemente flexible para escalar/desescalar con el tiempo", favor indicar los módulos/capacidades de escalamiento flexible con su respectiva unidad de medida.	Forma parte de la propuesta requerida en función de los casos de uso Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
479	43. En el ítem RG-12, se indica "Esta monitorización debe aportar visibilidad a nivel técnico y funcional de la solución", favor indicar el formato de visualización de los datos de monitoreo.	Sección 3.4.2 Monitoreo; se espera en la propuesta requerida
480	44. En el ítem RG-13 se indica "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." Favor confirmar que los cuadros de mando se construyan/configuren como parte de esta contratación, de ser así ampliar la información del alcance ya que es una variable indispensable para consideración de esfuerzos profesionales.	Se hace referencia al cumplimiento de la funcionalidad, de requerir construcción de cuadros de mando serán especificados en el TDR, se sugiere enviar costos y estimación e implementación dentro de la propuesta
481	45. En el ítem RG-16, asociada a auditoría y linaje, se hace referencia a varios puntos de control que dependen de módulos/productos distintos, favor confirmar que este requerimiento sea cubierto desde los criterios de control de cada uno de ellos.	Confirmado

No.	Pregunta	Respuesta
482	46. En el ítem RG-23 se solicita "La plataforma debe permitir ser replicada a otra zona de disponibilidad e incluso a otra nube", favor confirmar que esta sea una funcionalidad deseable o debe ser implementada como parte de este proyecto.	Se requiere contar por lo menos con 4 ambientes: Desarrollo, Pruebas, Producción y Contingencia; el SRI cuenta con un centro de datos principal en UIO para producción y un centro de datos alternativo en GYE para los ambientes de contingencia, preproducción y desarrollo
483	47. En el ítem RG-25 se solicita "El soporte debe ser íntegramente en español tanto a nivel de tickets y notificaciones como a nivel de soporte telefónico o remoto". Favor confirmar que el requerimiento aplique al proveedor local.	Se refiere a quien efectúe el soporte de la herramienta
484	48. En el ítem V-2 se solicita "El virtualizador soportará el acceso a diferentes tipos de fuentes de datos". Favor confirmar cantidad y tipo de las fuentes de datos (Motor de base de datos, file Systems, y demás)	Data semiestructurada (json, XML, csv, txt), Data estructura (DB relacionales ORACLE). Sección 3.2.1 Información de referencia de volúmenes de datos actuales para el tamaño actual y versiones. Ver Anexo 1 - Volumetría.
485	49. En el ítem V-4 se solicita "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", favor confirmar si las operaciones sobre las tablas físicas son únicamente de consulta.	Si, de consulta
486	50. En el ítem V-13 se solicita "Posibilidad de seleccionar la concurrencia utilizada en el virtualizador para paralelizar en mayor o menor grado las queries al origen", favor aclarar que la concurrencia sea una característica configurable para todo el catálogo o de forma individual para cada queri.	Se esperaría propuesta de la mejor alternativa
487	51. En el ítem AL-1 se solicita "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.", favor confirmar que la integración de componentes a la que se refiere es netamente a los propuestos en este proyecto, en el caso de existir otros, favor indicar cuáles son.	Si, ver cuadro Arquitectura referencial
488	52. En el ítem AL-3 se solicita "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.", favor ampliar información ya que, cumpliendo con su solicitud, se entregará un sistema de data Lakehouse con la capacidad de base de datos relacional, por ende, no se comprende el requerimiento adicional de seleccionar entre diferentes proveedores de bases de datos. Favor ampliar el detalle del requerimiento.	La solución debe permitir gestionar datos estructurados y no estructurados así como conectarse a diferentes bases de datos: Oracle, SQL Server, MySQL, etc.
489	53. En el ítem AL-6 se solicita "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", favor confirmar que al referirse a soluciones estándar, se refiere a consultas SQL o REST.	Si, " El acceso a estos datos deberá tener diferentes interfaces y formatos, incluyendo API REST y conectores ODBC/JDBC."
490	54. En el ítem AL-8 de escalabilidad horizontal, se entiende que su cumplimiento es obligatorio respecto del impacto al rendimiento, lo que no ocurre con la inversión para los nuevos servidores y demás componentes de aplicación, gestión, seguridad, disponibilidad para su correcto funcionamiento, tanto modalidad nube como on premise.	RG-6: Escalabilidad licenciamiento: La plataforma debe ser modular y lo suficientemente flexible para escalar/desescalar con el tiempo (al menos 3 años de soporte y mantenimiento)

No.	Pregunta	Respuesta
491	55. En el ítem AL-9 se solicita "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", favor confirmar las características del patrón de acceso esperado.	Data semiestructurada (json, XML, csv, txt), Data estructura (DB relacionales ORACLE)
492	56. En el ítem AL-14 respecto de la Calidad de datos, se entiende que el flujo de adquisición de datos deberá incluir la ejecución de las reglas de calidad de datos determinadas por el SRI, de tal manera que el personal responsable del proceso analítico solo consumirá los datos que cumplieron con dichas reglas. Confirmar nuestro entendimiento.	Se deben levantar y registrar las reglas de calidad de datos en conjunto con el usuario, así mismo se levantarán las acciones en caso de que no cumplan las mismas, como identificar los datos, no considerar los datos o dejarlos pasar en determinados escenarios
493	57. En el ítem AL-18 se solicita "La plataforma debe contar para cada almacén de datos de un sistema de logs y monitoreo", se entiende que el alcance de este requerimiento está limitado al sistema de almacenamiento de este proyecto y nada fuera de él, favor confirmar nuestro entendimiento.	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.
494	58. En el ítem A-4, Al referirse a "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." Favor indicar cuál es el caso/s de uso que se pretende ejecutar. Esta información permitirá tomar varias condiciones de diseño, favor extender la información lo más detallada posible.	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
495	59. Respecto del ítem A-5 para la ingesta de datos en tiempo real, Al referirse a "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." Favor confirmar que el SRI hará las modificaciones sobre las aplicaciones que se incluirán en el esquema de procesamiento en tiempo real o casi en tiempo real, como por ejemplo: habilitar CDC a nivel de base datos, modificación de capa aplicativa, instalación de conectores o plug ins, entre otros, modificaciones que serán responsabilidades enteramente del SRI.	Forma parte de la propuesta requerida
496	60. Respecto del ítem A-7 para la Ingesta y procesado de datos por lotes, al solicitar a "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", se está haciendo referencia al ETL, confirmar nuestro entendimiento.	Virtualización y en ciertos casos ETL
497	61. Respecto del ítem A-8 se solicita que "La plataforma tiene capacidades de virtualización de dato" hace énfasis en el módulo específico de virtualización de datos, confirmar nuestro entendimiento.	Sección 3.1.2 Virtualización de Datos

No.	Pregunta	Respuesta
498	62. Respecto del ítem A-9 donde se solicita que "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.", confirmar que dichos eventos sean actualizaciones, mantenimientos, soporte, que se programan y/o planifican para su ejecución.	Se refiere a la sección 3.1.12 Capacidades Analíticas
499	63. En el ítem ML-1 respecto de las capacidades analíticas, considerando que uno de los objetivos del proyecto busca "Contar con una infraestructura tecnológica especializada en el campo de analítica avanzada", favor confirmar si se requiere integrar las capacidades de analítica predictiva y prescriptiva, además de analítica descriptiva.	Si
500	64. En el ítem ML-2 respecto de las capacidades analíticas y considerando que uno de los objetivos del proyecto busca "Contar con una infraestructura tecnológica especializada en el campo de analítica avanzada", favor confirmar si es posible proponer Librerías analíticas (scikit-learn, pandas, numpy, ploty, R entre otras) y lenguajes especializados para ciencias de datos (Spark, Python, Spark-ly, R, Escala, entre otros).	Si, forma parte de la propuesta requerida el tipo de solución que se plantea
501	65. En el ítem ML-3 se solicita "La plataforma deberá contar con la capacidad de entrenar, implementar, automatizar y administrar modelos de aprendizaje automático y minería de datos", se entiende que para cumplir con esas capacidades, se requiere incluir además los modelos de analítica predictiva y analítica prescriptiva, favor confirmar.	Si
502	66. En el ítem ML-14 se solicita "La plataforma de forma integrada debe permitir establecer una política de ejecución de forma periódica de backup sobre un determinado proyecto", se entiende que la plataforma debe permitir la integración con herramientas propias de respaldos provista por el SRI, favor confirmar nuestro entendimiento.	SI
503	67. En el ítem ML-20 se solicita "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"; entendiendo que La arquitectura de virtualización permitirá cumplir con el agnosticismo solicitado, favor confirmar la aceptación de la funcionalidad a través del módulo de virtualización	Si, forma parte de la propuesta requerida el tipo de solución que se plantea
504	68. En el ítem ML-21 se solicita "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"; entendiendo que La arquitectura de virtualización permitirá cumplir con el agnosticismo solicitado, favor confirmar la aceptación de la funcionalidad a través del módulo de virtualización.	Si, forma parte de la propuesta requerida el tipo de solución que se plantea
505	69. En el ítem ML-22 se solicita "Capacidades para cruzar de manera directa datos de diferentes orígenes siendo agnóstico para los usuarios"; entendiendo que La arquitectura de virtualización permitirá cumplir con el	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.

No.	Pregunta	Respuesta
	agnosticismo solicitado, favor confirmar la aceptación de la funcionalidad a través del módulo de virtualización.	
506	70. En el ítem ML-30 se solicita "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". El uso de recursos bajo demanda, es un beneficio exclusivo en nube e implica una facturación variable del servicio. Favor confirmar si será posible la modalidad de facturación variable, es decir, que los valores no serán iguales sino basados en el consumo de cada período de tiempo.	Si
507	71. En el ítem ML-37 se solicita "Permitir que los activos generados, inclusive los modelos puedan ser promocionados entre entornos." Confirmar si el entorno se refiere a la plataforma de fábrica de datos u otra. De ser otra, favor especificar.	Se refiere a la plataforma
508	72. En el ítem ML-43 se solicita "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." ¿Se espera una integración con un sistema de almacenamiento de librerías, repositorio de librerías y versiones de código base para uso, repositorio de programas generados por el SRI? Favor confirmar	Si
509	73. En el ítem OP-2 se solicita "La plataforma debe proveer de un acceso securizado, perfilado y auditado a los datos pseudoanonimizados", favor confirmar el ámbito, es decir dentro de la plataforma requerida o, dentro y fuera de ella.	Dentro de la plataforma
510	74. En el ítem OP-4 se solicita "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", favor confirmar que es posible considerar a Openshift para cubrir este requerimiento.	Si
511	75. En el ítem OP-5 se solicita "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." El cumplimiento de este requerimiento se limita a la modalidad nube con facturación adicional de acuerdo a las cargas adicionales procesadas, favor confirmar que, conociendo estas condiciones, el requerimiento se mantiene y es exclusivo para la modalidad nube.	La propuesta puede incluir, nube, onpremise o híbrido con el detalle respectivo
512	76. En el ítem 5.5 numeral 4.1 se tiene la tabla de cantidades y precios de modalidad on premise, donde se solicita el precio de la "implementación de la plataforma" y "ambientes de pruebas y preproducción", favor confirmar que se hace referencia al precio de los servicios de implementación del ambiente de producción para el primer caso, y de los servicios de implementación de los ambientes	si

No.	Pregunta	Respuesta
	de pruebas/preproducción para el segundo. De no ser así, favor explicar el contenido de cada ítem.	
513	77. En el ítem 5.5 numeral 4.1 se tiene la tabla de cantidades y precios de modalidad on premise, donde se solicita el precio del "licenciamiento de componentes de plataforma", favor confirmar que se hace referencia al precio de licencias de todos los ambientes: producción, desarrollo y pruebas. De no ser así, favor aclarar.	Se espera que la propuesta detalle los costos por componente y de todos los ambientes requeridos
514	78. En el ítem 5.5 numeral 4.1 se tiene la tabla de cantidades y precios de modalidad on premise, donde se solicita el precio del "licenciamiento de componentes de plataforma" y "licenciamiento de herramientas de migración (ETL o virtualizador)", favor confirmar que de todos los módulos/productos que componen la plataforma de analítica, se aislen ETL/Virtualizador, cuyos precios se ubicarán en otro casillero.	Se espera que la propuesta detalle los costos por componente
515	79. En el ítem 5.5 numeral 4.1 se tiene la tabla de cantidades y precios de modalidad on premise, donde se solicita el precio de los "servicios de gestión de datos", favor indicar el alcance de este servicio en términos de entregables para que la entrega/aceptación no sea ambigua. Se sugiere establecer estos entregables en función de tareas específicas, actividades específicas, roles especializados específicos, lapsos, factores de aceptación, entre otros.	Forma parte de la propuesta sugerida
516	80. En el ítem 5.5 numeral 4.2 se tiene la tabla de cantidades y precios de modalidad nube, donde se solicita el precio del "licenciamiento de componentes de plataforma" y "licenciamiento de herramientas de migración (ETL o virtualizador)", entendiéndose que en un modelo de renta no es posible considerar valores de licencias, confirmar que es posible descartar estos campos de la tabla.	La propuesta puede incluir, nube, onpremise o híbrido con el detalle respectivo
517	81. En el ítem 5.5 numeral 4.2 se tiene la tabla de cantidades y precios de modalidad nube, donde se solicita el precio del "soporte y mantenimiento preventivo", entendiéndose que se trata de un modelo de servicio, el soporte y mantenimiento ante eventos no controlados de degradación del rendimiento de uno o varios módulos/productos de la plataforma, forma parte del servicio	Si, se espera que se especifique el nivel del servicio en incidentes y requerimientos, en la propuesta requerida
518	1. Sería correcto afirmar que toda la información estructurada y no estructurada recolectada en este nuevo sistema, debe contribuir a ampliar y detallar el perfil tributario de cada uno de los ciudadanos/negocios del Ecuador? Es decir, la llave de la información será el identificador tributario RUC?	Si, para los casos de uso que involucre al contribuyente
519	2. Según el Gráfico 1 y Gráfico 2, ¿son estas las únicas fuentes de información o plataformas que van a proveer los datos hacia el MDM? ¿Se puede categorizar en alguno de los siguientes conectores? Plataformas de industria, pj. Salesforce, SAP S/4, Hubspot... etc. Plataformas tecnológicas: IBM DB2, ORACLE DB, MS SQL, Desarrollos a la medida vía integración: SOAP, API REST	Se espera este detalle en la propuesta requerida

No.	Pregunta	Respuesta
520	3. Sobre datos No Estructurados: a. ¿Estos donde se encuentran? b. ¿Espacio esperado en disco? c. ¿Algunos ejemplos de casos de uso que se espera darles?	a. La mayoría en la base de datos Oracle y en mínima proporción repositorios documentales. b. Ver anexo 1 – Volumetría c. Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
521	4. "RG-15 Disponibilidad 24x7" Por favor detallar lo esperado.	Si. Se espera disponibilidad 24x7 de la plataforma
522	5. Actualmente tienen una plataforma tipo "Datawarehouse" ¿Esta se espera ser reemplazada o integrada a la nueva?	Integrada a la nueva
523	6. Sobre "RG-5: Las licencias deben incluir número ilimitado" el requerimiento es sólo a nivel de Usuarios o usuarios y registros (un registro es un ciudadano/empresa)?	El detalle de la forma de licenciamiento se espera en la propuesta
524	7. Una vez el registro esté normalizado en el MDM, ¿cómo se espera que ese Registro Maestro sirva a otras plataformas y sea reutilizado por las plataformas origen?	Revisar requerimiento GD-51
525	Amablemente solicitamos aclarar cuáles plataformas y herramientas se emplean actualmente para los flujos de datos, teniendo en cuenta que las especificaciones técnicas mencionan mejora o reingeniería de los mismos.	Actualmente existen desarrollos de flujos mediante SAP Data Services, Oracle Warehouse Builder, Paquetes PL/SQL que se ejecutan mediante Jobs de BDD, Chrones, Paquetes PL/SQL, flujos de SAP DS, etc en forma continua durante todo el día.
526	¿Actualmente cuentan con alguna herramienta de visualización de datos? ¿Cuál?	SAP Business Objects, Oracle Discoverer, Qlik, Apache Superset, Saiku
527	Si bien describen una arquitectura de referencia, ¿Cuál es la arquitectura de solución actual? ¿Tienen avances en inteligencia de negocio?	Diagrama de arquitectura de DW
528	¿Cuáles son las fuentes de información del SRI?	Data semiestructurada (json, XML, csv, txt, repositorios documentales), Data estructura (DB relacionales ORACLE)
529	¿Actualmente el catálogo de metadatos se encuentra actualizado? Si es positivo, ¿Cuentan con una política para la gestión metadatos?	No
530	Por favor, confirmar que Data Fabric, al ser un concepto de diseño según Gartner, no se refiere a una herramienta puntual sino hacia cómo debe diseñarse la solución y/o los beneficios esperados	Si
531	Por favor confirmar que la solución debe considerar un ecosistema híbrido/multicloud.	Deseable
532	Por favor, confirmar que, en la página 12, cuando se refieren a implementación en nube, se refieren a una nube pública tipo Azure, AWS y GCP?	Si
533	Por favor, confirmar que no es mandatorio que la solución esté alojada en la nube de CNT	No es mandatorio
534	Por favor, confirmar la fecha estimada de contratación de la solución	Fecha estimada de adjudicación de contrato mayo 2024
535	Por favor, confirmar que en la pregunta RG3 al referirse a "un producto existente y consolidado" es indispensable que conste en el cuadrante mágico de Gartner y Gartner Critical Capabilities	Es deseable
536	Por favor, confirmar que la respuesta a la pregunta RG2 admite múltiples productos. No existe en el mercado un producto que satisfaga completamente la amplia gama de necesidades que SRI requiere.	Debería ser un solo fabricante que maneje varios componentes

No.	Pregunta	Respuesta
537	Por favor, confirmar que repuesta a la pregunta RG4 es aceptable en la medida en que la plataforma tenga la capacidad de escalar de manera distribuida tanto en dimensiones como en volumen de accesibilidad, almacenamiento y procesado de datos, ofreciendo una capacidad de respuesta eficiente ante un elevado volumen de datos, servicios y usuarios. Sin necesidad de utilizar Kubernetes	Cualquier similar a kubernetes (contenedores)
538	Por favor, confirmar en la medida en que la plataforma propuesta sea Fault Tolerant en todos sus elementos 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, es aceptable sin necesidad de utilizar Kubernetes	Si se requiere que la plataforma sea tolerante a fallos y puede ser con una opción diferente a kubernetes.
539	Por favor, confirmar que para la solución on premise, se deberá proponer una infraestructura de Hardware	Si, debe incluir en la propuesta
540	Por favor informar cuáles son los tiempos de procesamiento de la información en SRI?	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso. El tiempo de ejecución del ETL puede variar de entre pocos minutos hasta horas. Se espera tiempo real
541	Por favor, confirmar si la plataforma debe permitir que los analistas comerciales usen 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.	Si
542	Por favor, confirmar que se va a migrar las actuales bases de datos a la solución de Datawarehouse para evitar los silos de información.	No se pide una solución de DWH, se solicita un plataforma de analítica, en la que una fuente de información es nuestro DWH actual.
543	Por favor, confirmar cuántas serán las sesiones concurrentes para ETL?	Actualmente el Usuario/Esquema de Oracle, ejecuta mediante Jobs de BDD, Paquetes PL/SQL, flujos de SAP DataServices, etc. en forma continua durante todo el día.
544	Por favor, confirmar ¿Cuál es el tamaño de los datos de entrada en el ETL?	Referirse a la sección 3,2,1 Información de referencia de volúmenes de datos actuales.
545	Por favor, confirmar cuántas serán las sesiones concurrentes para Calidad de Datos?	10 Ingenieros de datos; 10 científicos de datos
546	Por favor, confirmar cuál es el tamaño de los datos de entrada en Calidad de Datos?	Ver Anexo 1 - Volumetría.
547	Por favor, confirmar cuántas serán las sesiones concurrentes para Analítica Básica?	10 Ingenieros de datos; 15 científicos de datos
548	Por favor, confirmar cuál es el tamaño de los datos de entrada en Analítica Básica?	Ver Anexo 1 - Volumetría.
549	Por favor, confirmar cuántas personas desarrollan esos procesos analíticos?	Aproximadamente 10 Ing. De datos y 10 científicos
550	Por favor, confirmar cuántas serán las sesiones concurrentes para Machine Learning?	10 Ingenieros de datos; 10 científicos de datos
551	Por favor, confirmar cuál es el tamaño de los datos de entrada en Machine Learning?	Ver Anexo 1 - Volumetría.
552	Por favor, confirmar cuántas personas deben desarrollar los dashboards?	Entre 100 y 200 usuarios

No.	Pregunta	Respuesta
553	Por favor, confirmar cual es el monto anual de recaudación del SRI	Revisar el siguiente link: https://www.sri.gob.ec/estadisticas-sri
554	Por favor, confirmar si se cuenta con proyectos de analítica que deban ser migrados en un formato estándar?	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
555	Por favor detallar el número de usuarios concurrentes por aplicación	Se tienen alrededor de 1.800 usuarios de consulta en la herramienta actual con una concurrencia del 10 al 15%. 10 usuarios de ingeniería de datos y 20 científicos de datos.
556	Por favor, confirmar los casos de uso que necesita abordar SRI y su complejidad	Los casos de uso se definirán en los TDR, se espera que el proveedor implemente 3 casos de uso y 2 casos de uso lo haga la administración con acompañamiento del proveedor. (2 de complejidad alta, 2 de complejidad media y 1 de complejidad baja)
557	Por favor, informar que procedimientos estadísticos utilizan o trabajan en sus proyectos actualmente? Ejemplo Regresión lineal, arboles de decisión, etc.?	Procedimientos de modelos de segmentación y clasificación.
558	Por favor informar si actualmente desarrollan procedimientos de machine learning? Cuáles algoritmos?	Sí, Clasificación y Segmentación
559	Por favor, informar si los modelos analíticos están consumidos en un motor de decisión? cuál es?	Forma parte de la propuesta requerida
560	Por favor, confirmar que se requiere la inclusión de un motor de decisión analítico?	Forma parte de la propuesta requerida
561	Por favor, informar si se contempla un motor de decisión que permita la ejecución de los modelos analíticos en un esquema de reglas de negocio?	Forma parte de la propuesta requerida
562	Por favor, confirmar que se contempla la capacidad de incluir un módulo de gobernanza de modelos analíticos dentro de la misma plataforma analítica	El módulo de gobierno debe permitir documentar los dominios de información, atributos, linaje, calidad, enfocada al análisis de información
563	Por favor, confirmar que se contempla la capacidad de tener un único repositorio de modelos en un módulo de gobernanza?	Sí
564	Por favor, confirmar que se contempla que el gestor de modelos permita automatizar el ciclo de vida del modelo mediante un enfoque CI/CD?	Forma parte de la propuesta requerida
565	Por favor, confirmar si la plataforma ejecutará modelos analíticos en Real Time y qué modelos?	Forma parte de la propuesta requerida
566	Por favor, confirmar qué tipos de modelos se contemplan ejecutar en Real Time?	Forma parte de la propuesta requerida
567	Por favor, confirmar si contempla la ejecución de algoritmos como máquinas de soporte vectorial, K-means, t-Distributed Stochastic Neighbor Embedding (t-SNE), regresión lineal o ONNX en tiempo Real?	Forma parte de la propuesta requerida
568	Por favor, confirmar si el engine de Real time debe tener integración con Open Source como Python?	Forma parte de la propuesta requerida
569	Por favor, confirmar si requiere que el Motor de Real Time soporte ONNX model?	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
570	Por favor, confirmar el número de personas que se capacitarán en las diversas funcionalidades	Capa de gobierno: Al menos 10 Capa de Explotación: Al menos 20 Capa de Integración: Al menos 5 Capa de consumo: Al menos 40 Capa de conectividad: Al menos 5 Capa de seguridad: Al menos 10
571	Por favor, confirmar que se requiere una interfaz gráfica que permita a los usuarios no técnicos o funcionales la parametrización y uso de algoritmos de NLP?	Forma parte de la propuesta requerida
572	Por favor, confirmar si la plataforma permite modelar y pronosticar cada serie en la jerarquía individualmente, pero también debe permitir realizar reconciliación jerárquica, ajustando los pronósticos en múltiples niveles de la jerarquía.	Forma parte de la propuesta requerida
573	Por favor, confirmar si la plataforma de analítica debe proveer un solver para optimización?	Es deseable que sí.
574	Por favor, confirmar que la plataforma analítica debe proveer algoritmos lineales de optimización como: Primal and dual simplex. Sifting. Network simplex. Interior point with crossover. Concurrent solve capability.?	Forma parte de la propuesta requerida
575	Por favor, informar los procesos que requieren el control de versiones?	Se debe manejar un control de versiones en todos los modelos y flujos construidos en la plataforma.
576	Por favor, informar cuál es la plataforma de streaming y mensajería de SRI?	La organización no posee a la fecha las plataformas indicadas.
577	Por favor, informar cual es la razón de tener un leguaje detallado como estos?	La plataforma se requiere el manejo de lenguajes estándar en la industria (Python, SCALA, PySpark)
578	Por favor, informar los tipos de modelos analíticos se tiene en Python y en R?	Análisis de Redes, Modelos Económicos y Minería de Texto.
579	Por favor, informar cuál es la periodicidad y/o la política de backup?	Backups diarios
580	Por favor, informar cuáles son las herramientas/ servicio externo?	Integración con proveedores externos (convenios 32)
581	Por favor, confirmar si es necesario que tenga capacidades de pago por uso así como capacidad fija para garantizar el control financiero de la solución	Forma parte de la propuesta requerida.
582	Por favor confirmar que se debe contemplar la migración de los volúmenes de datos totales que están detallados incluyendo el de respaldo?	La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría
583	Por favor, confirmar si el detalle de la implementación de los servicios es parte de otro contrato?	No, es parte del mismo contrato
584	Por favor, informar si se tiene cuantificado los modelos de analítica que deben ser migrados con complejidad alta media o alta? Si la respuesta es positiva podrían facilitarnos el número de cada uno de ellos?	Se priorizarán 3 casos de uso a ser implementados por el proveedor y 2 a ser implementados por SRI con acompañamiento del proveedor que contenga información de facturación electrónica, declaraciones, catastro y anexos, que tengan componentes analíticos avanzados, predictivos y componentes de IA
585	Por favor, informar el alcance inicial de los servicios de gestión de datos. Linaje? Catálogo? Cuantas reglas de calidad?	El número será determinado en la implementación de los casos de uso priorizados
586	Por favor, informar el número de reportes que se quieren implementar en un eventual proyecto?	Depende de los casos de usos los cuales se especificarán en los TDR. Considere un número estimado a implementar dentro de su propuesta.

No.	Pregunta	Respuesta
587	Por favor, informar el número de tablas por Fuente de información se desean automatizaren los ETL	Como referencia, existen alrededor de 100 productos DWH, en su carga continua intervienen procesos ETL, sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.
588	Por favor, informar si se tiene cuantificado los reportes y tableros que deben ser migrados con complejidad alta media o alta? Si la respuesta es positiva podrían facilitarnos el número de cada uno de ellos?	Depende del caso de uso, y se debe ir levantando conforme se desarrolla el caso de uso.
589	Por favor confirmar si se debe contemplar la migración de modelos analíticos?	Se priorizarán 3 casos de uso a ser implementados por el proveedor y 2 a ser implementados por SRI con acompañamiento del proveedor que contenga información de facturación electrónica, declaraciones, catastro y anexos, que tengan componentes analíticos avanzados, predictivos y componentes de IA
590	Por favor, informar si se tiene cuantificado los procesos de transformación que deben ser migrados con complejidad alta media o alta? Si la respuesta es positiva podrían facilitarnos el número de cada uno de ellos?	La migración de flujos de trabajo dependerá de los casos de uso que se definan en el TDR, se plantean 5 casos (2 de complejidad alta, 2 complejidad media y 1 complejidad baja). Ver Anexo 1 - Volumetría
591	Por favor, informar qué herramientas de transformación de datos se están utilizando actualmente?	Actualmente existen desarrollos de flujos mediante SAP Data Services, Oracle Warehouse Builder, Paquetes PL/SQL que se ejecutan mediante Jobs de BDD, Chrones, Paquetes PL/SQL, flujos de SAP DataServices, etc. en forma continua durante todo el día. Sin embargo en el proyecto se usarán solo ciertos productos como casos de uso.
592	Por favor, informar qué herramientas de visualización se están utilizando actualmente?	SAP Business Objects, Oracle Discoverer, Qlik, Apache Superset, Saiku
593	Por favor, informar qué herramientas de analítica se están utilizando actualmente?	Los modelos Analíticos propiamente se encuentran en SPSS de IBM. Los Cubos de Información: Modelo Físico en la BDD Oracle, Modelo Lógico en SAP Business Intelligence y Oracle Discoverer.
594	Por favor informar la metodología se encuentran usando para la implementación de proyectos de datos?	CRISP-DM y Metodología AD-HOC
595	Por favor, confirmar si de los servicios de implementación se consideran realizar procesos desde cero o se pretende actualizar y/o mejorar los procesos actuales?	Desde cero
596	Por favor, confirmar si el SRI posee un Plan de Gobierno de Datos o se debe generar un Plan de Gobierno de Datos e implementarlo?	Debe Generar una Política de Gobierno e implementarlo en los casos de uso a ser priorizados
597	Por favor, informar la herramienta de gobierno de datos se están utilizando actualmente?	Ninguna
598	Por favor, informar si los datos fuentes necesitan un proceso de homologación?	Si, dentro del alcance de los casos de uso priorizados
599	Por favor, confirmar si como parte del perfilamiento, se considera algún reporte o informe para comunicar el estado de las tablas perfiladas.	Si
600	Por favor confirmar si como parte de la ejecución de las reglas de calidad, se considera un tablero o informe para el control de ejecución por ciclo.	Si
601	Por favor informar que para la implementación de flujo de revisión y aprobación por roles, cuantas reglas y términos de negocio máximo se van a considerar.	El número será determinado en la implementación de los casos de uso priorizados
602	Por favor informar que para la implementación de términos de negocio, cual es el número máximo a implementar.	El número será determinado en la implementación de los casos de uso priorizados

No.	Pregunta	Respuesta
603	Por favor informar que para la implementación de las reglas de calidad, cual es el número máximo a implementar.	El número será determinado en la implementación de los casos de uso priorizados
604	Por favor informar que para el perfilamiento cual es el número máximo de tablas que se van a considerar	El número será determinado en la implementación de los casos de uso priorizados
605	Por favor, informar que la plataforma deberá soportar de manera costo eficiente sin auto escalar, alta concurrencia de consultas y cargas en sus ambientes	La plataforma debe soportar autoscaling y de manera costo eficiente.
606	Por favor, informar que la solución de DWH deberá soportar alta complejidad de queries, alta demanda y volúmenes grandes de datos sin que esto implique un crecimiento exponencial de los costos asociados	Es lo requerido.
607	Por favor, informar si existe algún Requerimiento específico de RTO o RPO en el ambiente de contingencia?	Si. RPO 24 min; RTO 4horas
608	Considerando la importancia del sistema, SRI asignará una mejor calificación a servicios en la nube en donde el provisionamiento, instalación, configuración, administración y operación de la solución sea hecha por el mismo proveedor del software?	No, se evaluará las dos modalidades
609	Por favor, informar el sistema de gestión de identidades utiliza SRI? Donde está ubicado el AD?	Es una aplicación on premise, actualmente implementada es IBM security identity manager, Se solicita al proveedor proponer el mejor esquema de gestión de usuarios. El SRI se encuentra en el proceso de implementar el Directorio Activo integrado con Microsoft Azure
610	Por favor, informar si se requiere los servicios administrados de la plataforma.	Se requiere un servicio de plataforma integro y administrado.
611	Por favor confirmar si se requiere alta disponibilidad en la solución?	Si, es un requisito.
612	Por favor, confirmar si se requiere tener una solución de disaster recovery?	Si
613	1. En la sección "2.3 Necesidades específicas" indican "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)" a. Por favor ¿pueden aclarar si para la instalación On-Premise, la entidad proveerá toda la infraestructura tecnológica necesaria?	Si
614	b. Para la instalación en Cloud, ¿cuál es el proveedor preferido por la entidad?	Forma parte de la propuesta requerida
615	c. Entendemos que para la instalación en Cloud se hará uso del contrato, servicios y productos que tiene actualmente contratados la entidad, por tanto estos costos no se incluirán dentro de la respuesta a este RFI. Por favor confirmar.	No, se debe incluir todos los costos
616	d. Para la instalación On-Premise, entendemos que la entidad entregará toda infraestructura tecnológica preparada y configurada para hacer la instalación y despliegue de la plataforma de Data Fabric para los cuatro (4) ambientes que están solicitando. Por favor confirmar.	La propuesta debe especificar que requieren en la solución ON PREMISE y de ser plataformas fuera de las manejadas por la institución debe considerar inclusive su configuración.

No.	Pregunta	Respuesta
617	2. En la sección "3 REQUERIMIENTOS DE LA SOLUCIÓN" se indica "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" a. Por favor pueden aclarar si ¿la entidad además de la implementación de la plataforma Data Fabric, también espera que se presten servicios de consultoría para la implementación de "procesos para la integración y transformación de datos, de las fuentes identificadas a la Plataforma DataFabric"?	Sección 4. Capacitación y transferencia de datos
618	b. Si la respuesta a la pregunta anterior es afirmativa, ¿por favor pueden indicar a qué transformaciones se refieren? ¿Cuántas integraciones y con qué plataformas se deben desarrollar?	Forma parte de la propuesta requerida
619	3. En la sección "3 REQUERIMIENTOS DE LA SOLUCIÓN" se indica "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" a. Por favor, ¿pueden aclarar si la entidad espera que se desarrollen modelos analíticos como parte del proyecto? Si la respuesta es afirmativa, por favor ¿pueden proveer información de la cantidad, requerimientos y características de los modelos que se deben desarrollar?	Si, en función de los casos de uso mismos que se definirán en el TDR. Considerar como parte de la propuesta requerida. Se priorizarán 3 casos de uso a ser implementados por el proveedor y 2 a ser implementados por SRI con acompañamiento del proveedor que contenga información de facturación electrónica, declaraciones, catastro y anexos, que tengan componentes analíticos avanzados, predictivos y componentes de IA. Adicionalmente revisar el anexo 1 - Volumetría
620	b. Por favor, ¿pueden aclarar si la entidad espera que se desarrollen modelos de ML y/o AI? Si la respuesta es afirmativa, por favor, ¿pueden proveer los requerimientos, características de las fuentes de información, cantidad y tipos de modelos de ML y AI que se deben desarrollar?	Depende del caso de uso, pero si se busca que la plataforma soporte desarrollo de modelos ML y AI.
621	4. En la misma sección se indica "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" a. Por favor, ¿pueden proveer las fuentes de información que se debe migrar, y sus características, es decir, motor de base de datos, número de tablas, tipos de datos, volumen de datos?	Sección 3.2.1 Información de referencia de volúmenes de datos actuales. Ver Anexo 1 - Volumetría
622	b. ¿Cuántos y cuáles son los flujos de datos más relevantes para la entidad que deben ser migrados?	Los que se consideren en los Casos de Uso que serían planteados.
623	c. Por favor, ¿pueden aclarar cómo se espera que sea el manejo para los requerimientos de migración de datos?	Forma parte de la propuesta requerida

No.	Pregunta	Respuesta
624	5. Se indica "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" a. Entendemos que lo que espera la entidad es que además de implementar la plataforma de Data Fabric, la firma consultora debe dar soporte y mantener a la misma plataforma, ya sea en esquema Cloud o en esquema On-Premise, por un tiempo de un (1) año. Por favor confirmar.	Si, por tres años
625	b. Para tener mayor claridad, por favor, ¿pueden aclarar si la entidad espera que la firma consultora haga la instalación y configuración de la herramienta y posteriormente entregue la información necesaria para que la administración la haga el equipo de tecnología de la entidad?	Si,
626	c. Por favor, ¿pueden indicar cuál es la expectativa de tiempo de duración del contrato?	A partir de la fecha de adjudicación del contrato, se estima un horizonte de implementación aproximado de 2 años y medio
627	6. Se menciona "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." a. Por favor, ¿pueden proveer mayor detalle de lo que espera la entidad cuando menciona "procesos de gestión de datos maestros (MDM) y de metadatos unificados"?	Se pretende documentar los dominios de información, atributos, reglas de calidad, linaje en una sola plataforma de manera que permita identificar fácilmente las afectaciones en caso de cualquier modificación de los datos fuentes que afecten a los modelos analíticos
628	b. Por favor pueden confirmar si, ¿la entidad ya tiene un catálogo de datos maestros organizacional? Si es así, pueden indicar la cantidad de datos maestros que tiene la entidad identificados que deben ser implementados?	Se han identificado 29 dominios de información que contienen internamente atributos y tablas aún no asociadas
629	c. Por favor pueden aclarar, ¿qué alcance espera la entidad cuando se refiere a "Facilitar el manejo y explotación de la información"?	Que los usuarios puedan acceder fácilmente a la información (única fuente de la verdad) y que la explotación y procesamiento de esta sea oportuna.
630	7. 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 y costos de: 1) Implementación en la infraestructura tecnológica del SRI (On premise), 2) Implementación en nube (Cloud)	Si, se requiere en las 2 modalidades
631	8. Por favor pueden aclarar si, ¿los requerimientos de la sección "3 REQUERIMIENTOS DE LA SOLUCIÓN" son requerimientos que debe cumplir la herramienta o son requerimientos que se deben ejecutar como parte de las actividades del proyecto?	Son requerimientos y funcionalidades que se espera que cumpla la herramienta.
632	9. Respecto del requerimiento "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", por favor pueden aclarar si se refieren a software open Source?	Si se refiere a software open Source. Siempre y cuando cumpla con todas las especificaciones.

No.	Pregunta	Respuesta
633	10. En la sección "3.2 Migración de Datos y/o reingeniería de flujo de datos" se indica "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." a. Por favor, ¿pueden aclarar cómo es el manejo propuesto por la entidad para los requerimientos que se hagan como "acompañamiento"?	Forma parte de la propuesta requerida en función de los casos de uso definidos en el TDR
634	b. Para este servicio, la entidad planea contar con un número determinado de horas de la firma consultora para la atención y ejecución de estos requerimientos?	Forma parte de la propuesta requerida en función de los casos de uso definidos en el TDR
635	c. ¿Cómo plantea la entidad hacer el pago de estos servicios?	Forma parte de la propuesta requerida en función de los casos de uso definidos en el TDR
636	11. En la sección "3.3 Servicios de Gestión" se menciona "Los componentes mínimos del modelo de gestión de datos a prestar por el fabricante, partner y/o proveedor deben incluir como mínimo los relacionados en el marco de referencia de gestión de datos del DAMA DMBOK2," a. Dado que el marco de referencia de DAMA DMBOK2 cuenta con 11 funciones o áreas de conocimiento, por favor pueden aclarar qué espera la entidad en el modelo de gestión para cada una de las áreas del DAMA DMBOK2?	1. Gobierno de Datos: Mantener en una herramienta la documentación de Dominios de información, linaje, ontología, afectaciones en caso de cambios 2. Datos Maestros y de Referencia: Documentación de datos maestros y de referencia, administración, actualización, control de accesos para su administración. 3. Metadatos: Documentación y enlace a las capas de datos maestros, linaje, calidad 4. Calidad de Datos: Administración de calidad, levantamiento de reglas, auditorías de calidad, ejecuciones programadas o a demanda, reportes de cumplimiento de reglas
637	b. Por favor, ¿pueden aclarar si actualmente la entidad ya tiene un modelo de gobierno de datos organizacional? Si la respuesta es afirmativa, por favor ¿pueden proveer información del modelo de datos actual?	No
638	c. Particularmente para el área de conocimiento de "Documentos y contenido", por favor, ¿pueden proveer mayor detalle de lo que espera la entidad que se desarrolle?	No está clara la pregunta
639	d. Si como parte del proyecto la entidad espera que se diseñe el modelo de gobierno de datos organizacional, por favor, ¿pueden indicar cuál es la prioridad para el diseño del modelo de gobierno de datos, por ejemplo, los procesos Core de negocio, o un proceso en particular, o un dato maestro como "contribuyente"?	Se implementarán 3 casos de uso a cargo del proveedor y 2 casos implementados por la entidad con acompañamiento del proveedor, el modelo de gobierno se implementará sobre estos casos
640	• ¿La arquitectura de aplicaciones requerida debe ser cubierta en su totalidad por la plataforma a presentar? ¿Puede cubrir componentes parciales?	Debe ser cubierta en su totalidad - RG 2
641	• ¿Para la plataforma de solución esperada se permite que sea compuesta por componentes de diferentes fabricantes? ¿En ambientes híbridos (cloud, on premise)?	Se espera que cumpla RG2 y RG3, puede enviar su propuesta considerando varios fabricantes y/o ambientes híbridos.
642	• ¿Cuál es el tipo de soporte que esperan por parte del proveedor y fabricante?	En la sección 3.4.9 Soporte y Mantenimiento se detalla. Se solicita que en la propuesta se indique el tipo de soporte del proveedor y fabricante para validación
643	• ¿Los proveedores pueden participar parcialmente en cuanto a las soluciones requeridas? Dado que actualmente no cumplimos con el 100% de lo requerido.	No, la propuesta debe cumplir en la totalidad
644	• ¿Pueden participar proveedores extranjeros? ¿Cuál sería el proceso a ejecutar para llevar a cabo el contrato en caso de que sí sea posible la participación?	Si pueden participar proveedores extranjeros conforme a las políticas del BID y para la ejecución contractual se regirá a lo establecido en el contrato y la legislación nacional.

No.	Pregunta	Respuesta
645	<ul style="list-style-type: none"> Agradecemos realizar aclaración o dar mayor información con respecto al ítem de Transferencia de conocimiento a 3 años, dado que no comprendemos el alcance de este y lo que ustedes esperan de esta parte del servicio. 	<p>Contempla la transferencia del conocimiento durante la ejecución del proyecto respecto al manejo y uso de la herramienta, fortalecimiento de capacidades para los usuarios, certificaciones (si aplica), etc.</p> <p>Forma parte de la propuesta requerida.</p>

Para constancia de lo acordado se suscribe el presente boletín:

RAMIRO ESPIN
GERENTE PROGRAMA INTAX
SERVICIO DE RENTAS INTERNAS ECUADOR

RAMIRO ESPINOSA
PATROCINADOR PROYECTO GOBIERNO DE DATOS
Y ANALITICA INSTITUCIONAL
SERVICIO DE RENTAS INTERNAS ECUADOR

PAULINA CORREA
GERENTE DE PROYECTO GOBIERNO DE DATOS Y
ANALITICA INSTITUCIONAL
PROGRAMA INTAX
SERVICIO DE RENTAS INTERNAS ECUADOR

ADRIANA MARTIN
PO PROYECTO GOBIERNO DE DATOS Y ANALÍTICA
INSTITUCIONAL
PROGRAMA INTAX
SERVICIO DE RENTAS INTERNAS ECUADOR

ANEXO 1 – VOLUMETRÍA

Se realiza una volumetría de un posible caso de uso de alta complejidad.

Cuadro dimensionamiento estructuras de datos fuentes para el caso de uso de alta complejidad # 1

Nro.	Modelo Datos (DWH/STAR) Corresponde a los modelos de datos construidos en el DWH del SRI que son fuente para el modelo final de variables de riesgos	Tipo dato	# Campos Mod. Dat.	# Registros del Mod. Dat. (millones)	Periodos Información	Tamaño Mod. Dat. (GB)	Estructuras de datos fuente del modelo de datos de DWH				
							Base de datos	# Tablas fuente del Mod. Dat.	# Campos de las tablas fuente	Tamaño tablas fuente (GB)	# Registros tablas fuente (millones)
1	Modelo 1 (Transaccionalidad)	Estructurado	124	105	2018 - 2023	34	BDD fuente 1	16	656	1.136	5.500
							BDD fuente 2	43	2.896	661	3.000
2	Modelo 2 (Comprobantes Electrónicos)	Estructurado	96	7.000	2018 - 2023	4.485	BDD fuente 3	20	293	9.129	50.000
							BDD fuente 3	56	434	8.236	127.000
Total mod y estructura datos:						4.519				19.162	
Total general:						23.681					

Notas:

Tipos de datos de las tablas fuente: Number, Date, Varchar2

Crecimiento de las tablas fuente de la "BDD fuente 3" es del 12% anual.