

INFORME DE LA CONSULTA PRELIMINAR DE MERCADO PARA LA PREPARACIÓN DEL EXPEDIENTE DE CONTRATACIÓN “SUMINISTROS Y SERVICIOS ASOCIADOS AL GOBIERNO DEL DATO Y FACTORÍA DE ANALÍTICA DE DATOS”

1. INTRODUCCIÓN

La Agencia Digital de Andalucía estudia la posibilidad de contratar “SUMINISTROS Y SERVICIOS ASOCIADOS AL GOBIERNO DEL DATO Y FACTORÍA DE ANALÍTICA DE DATOS”. Con el fin de recabar información de los operadores económicos relativa al objeto de dicha contratación y para ayudar en la posible elaboración futura de los pliegos correspondientes a un expediente de contratación, se abrió con fecha 28 de diciembre de 2021 una consulta preliminar de mercado, que puede ser accedida en la siguiente dirección:

https://www.ceh.junta-andalucia.es/haciendayadministracionpublica/apl/pdc_sirec/perfiles-licitaciones/consultas-preliminares/detalle.jsf?idExpediente=16

Este documento constituye el informe contemplado en el artículo 115.3 de la Ley 9/2017, de 8 de noviembre, de Contratos del Sector Público, por la que se transponen al ordenamiento jurídico español las Directivas del Parlamento Europeo y del Consejo 2014/23/UE y 2014/24/UE, de 26 de febrero de 2014.

2. RESULTADO DE LA CONSULTA

La consulta exponía el deseo de contar con los suministros y servicios necesarios para la constitución del Centro de excelencia de gobierno del dato y la Factoría de analítica de datos de la Junta de Andalucía.

Se solicitaba de los participantes la remisión de un cuestionario con:

- Objetivos y alcance del contrato (apartado 3 de este informe)
- Estimación de perfiles, dedicación y tarifa de la línea base del servicio (apartado 4 de este informe)
- Estimación de perfiles, dedicación y tarifa de los servicios a demanda (apartado 5 de este informe)
- Recomendaciones para el suministro de servicios en la nube (apartado 6 de este informe)

Se estableció el día 20 de enero de 2022 como fecha límite para remitir los cuestionarios cumplimentados a través de la Presentación Electrónica General de la Junta de Andalucía, con destino a la Dirección General de Estrategia Digital de la Agencia Digital de Andalucía.

Durante el periodo entre la publicación de la consulta y el final del plazo, varias empresas contactaron por correo electrónico con el equipo de Analítica y Gobierno del dato con consultas, que fueron respondidas.

Estas son las preguntas recibidas y sus correspondientes respuestas:

1. ¿Qué tipologías de datos se plantean integrar en la plataforma de Analítica de Datos?



| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 1/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

El servicio de Factoría de analítica de datos debe ser capaz de realizar la ingesta, almacenamiento, procesamiento y análisis de datos estructurados, semi-estructurados y no estructurados.

2. ¿Con qué volumen aproximado cuenta cada una de esas tipologías?

El volumen de los objetos de datos del proyecto no es conocido a priori, puesto que el perímetro del servicio lo constituye cualquier fuente de datos perteneciente a la Administración de la Junta de Andalucía, sus agencias administrativas y sus agencias de régimen especial, e incluso la aportación de fuentes de datos de terceros. La plataforma en la nube debe ser escalable a cualquier capacidad de almacenamiento y procesamiento que pueda necesitarse durante el proyecto.

3. ¿Cuáles suponen procesos de análisis en tiempo real y cuáles serían en batch?

El tipo de proceso de análisis de los datos en la nube no es conocido a priori, puesto que el perímetro del servicio lo constituye cualquier fuente de datos perteneciente a la Administración de la Junta de Andalucía, sus agencias administrativas y sus agencias de régimen especial, e incluso la aportación de fuentes de datos de terceros. La plataforma en la nube debe ofrecer la funcionalidad de ingesta y procesamiento de flujos de datos y por lotes, según las necesidades y requerimientos del caso de uso demandado.

4. ¿Qué periodicidad plantean para los análisis en lote? Diaria, semanal, mensual, etc.

La periodicidad no es conocida a priori, siendo dependiente de las necesidades y requerimientos del caso de uso demandado por cualquier organismo de la Administración de la Junta de Andalucía, sus agencias administrativas o sus agencias de régimen especial.

5. Necesitamos que nos confirmen que dentro del servicio de Factoría de analítica de datos queda cubierto el propio desarrollo de los casos de uso analíticos que se demanden

Efectivamente, el servicio de Factoría de analítica de datos cubre el desarrollo de los casos de uso analíticos, incluyendo, entre otras, las tareas de diseño de los procesos de ingesta y transformación de datos (ETL), análisis/consultas sobre los data sets y visualización de resultados (informes, cuadros de mando, ...), así como las tareas de desarrollo de funciones (APIs) para la compartición de resultados con otras aplicaciones.

6. Orígenes de datos.

6.1 ¿Qué tipo de datos serán parte del perímetro del proyecto?: el servicio de Factoría de analítica de datos debe ser capaz de realizar la ingesta, almacenamiento, procesamiento y análisis de datos estructurados, semi-estructurados y no estructurados, tanto mediante flujos (stream) como en procesos por lotes (batch), según las necesidades y requerimientos del caso de uso demandado.

6.2 ¿Qué tecnología de bases de datos existen?: el servicio de Factoría de analítica de datos debe ser capaz de realizar la ingesta de datos desde cualquier tecnología de base de datos de la Administración de la Junta de Andalucía (relacional, clave-valor, documental, orientada a grafos, ...) según las necesidades y requerimientos del caso de uso demandado.

6.3 ¿Cuántos dominios de datos existen?: no disponemos a priori de un inventario de dominios de datos del proyecto, puesto que su perímetro lo constituye cualquier fuente de datos

| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 2/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

perteneciente a la Administración de la Junta de Andalucía, sus agencias administrativas y sus agencias de régimen especial, e incluso la aportación de fuentes de datos de terceros.

6.4 ¿Se precisa incorporar fuentes en tiempo real?: según las necesidades y requerimientos del caso de uso demandado, es posible que sea necesario incorporar fuentes de datos en tiempo real.

7. Dimensionamiento.

7.1 ¿Nos pueden indicar el volumen de tablas/columnas por tabla de datos existentes?: el volumen de objetos de datos del proyecto no es conocido a priori, puesto que su perímetro lo constituye cualquier fuente de datos perteneciente a la Administración de la Junta de Andalucía, sus agencias administrativas y sus agencias de régimen especial, e incluso la aportación de fuentes de datos de terceros.

8. Gobierno del dato.

8.1 ¿Se considera necesario establecer un maestro de datos?: como parte de las tareas del Centro de excelencia de gobierno del dato, se deberá realizar la definición e implantación de los procesos asociados a datos maestros. Dichos procesos podrán requerir del establecimiento de tablas de datos maestros comunes a todos los data sets que sean procesados y analizados.

9. ¿Dentro de los perfiles de factoría hay que considerar también aquellos necesarios para la gestión de la infraestructura de la plataforma analítica?

Efectivamente, todas las labores de gestión, despliegue y operación de la infraestructura de la plataforma tecnológica corresponden al proveedor del servicio de Factoría de analítica de datos.

10. ¿Qué ámbitos competenciales (a nivel Consejerías y otras entidades) formarán parte del perímetro del servicio, tanto la línea base como los servicios a demanda?

El ámbito de aplicación de este proyecto lo constituye la Administración de la Junta de Andalucía, sus agencias administrativas y sus agencias de régimen especial, las cuales podrán demandar la implementación de casos de uso de analítica descriptiva, predictiva y prescriptiva.

11. ¿Se prevé que los casos de uso de analítica y analítica avanzada se ejecuten sólo con equipo del servicio o con otros equipos / recursos de las Consejerías que ya dispongan medios para ello?

Aunque no sea el modo habitual de prestación de los servicios, es viable que las tareas de desarrollo de casos de uso, incluyendo el diseño de procesos de transformación de datos, análisis/consultas y/o visualización sean realizados por equipos técnicos de los propios Organismos de la Administración de la Junta de Andalucía. En este caso, desde la Factoría de analítica de datos se prestará el servicio de plataforma tecnológica que soporte el pipeline de implementación de dichos casos de uso.

Finalizado el plazo de remisión, se recibieron respuestas formales de diecisiete (17) empresas:

- Accenture, S.L. Sociedad Unipersonal.
- AMAZON WEB SERVICES EMEA SARL, SUCURSAL EN ESPAÑA
- Aplicaciones y Tratamientos de Sistemas SA

| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 3/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- ATOS IT SOLUTIONS AND SERVICES IBERIA SL
- AYESA ADVANCED TECHNOLOGIES, S.A. + HIBERUS IT DEVELOPMENT SERVICES SL
- Deloitte Consulting S.L.U.
- DXC TECHNOLOGY SERVICIOS ESPAÑA S.L.
- FUJITSU TECHNOLOGY SOLUTIONS, S.A.U.
- HITACHI VANTARA, S.A.U.
- Indra Soluciones Tecnologías de la Información,S.L.U.(“MINSAIT”)
- Oracle Ibérica, S.R.L.
- ORANGE ESPAGNE
- Redes System Consulting & Solutions S.L.
- SOLTEL IT SOLLUTIONS, S.L.U
- EY TRANSFORMA SERVICIOS DE CONSULTORIA, S.L. (EY) y Solutia Digital Health S.L. (Solutia)
- TELEFONICA SOLUCIONES DE INFORMATICA Y COMUNICACIONES DE ESPAÑA SA
- VIEWNEXT S.A

3. OBJETIVOS Y ALCANCE DEL CONTRATO

Teniendo en cuenta los objetivos / tareas / servicios demandados con el objetivo de crear el Centro de excelencia de gobierno del dato y la Factoría de analítica de datos de la Junta de Andalucía planteados en esta consulta preliminar, **¿considera necesario que en el futuro pliego incluyamos tareas / servicios adicionales?**

En este punto nos hemos encontrado con diferentes tipos de respuestas, hacemos un resumen de las diferentes propuestas que nos han hecho las empresas para ser incluidas en los trabajos objeto de la futura contratación de suministros y servicios:

- Análisis de madurez e identificación de necesidades
- Crear un modelo que abarque los bloques de trabajo de gestión de la demanda, integración, arquitectura, modelo de gobierno, gestión de cambio y administración de plataforma alrededor del centro de excelencia del dato como entidad central de la Oficina Digital.
 - Gestión del Programa: Gestión integral del programa bajo metodologías de PMO.
 - Gestión de la Demanda: Gestión de los requerimientos y necesidades del negocio coordinados de manera integral para potenciar la capacidad del centro y los recursos dedicados.
 - Centro de Excelencia: El centro de excelencia de datos incluye bloques de competencias o capacidades:
 - Gobierno del Dato: Equipo dedicado con conocimiento y experiencia en el gobierno del dato (diccionarios de negocio, calidad y datos maestros).
 - Factoría de Analítica de Datos: Equipo dedicado al diseño e implementación de los casos de uso de la analítica de datos.
 - Integración: Capacidades de integración de la plataforma y alineamientos de integración con las aplicaciones departamentales.
 - Arquitectura: Incluye las definiciones y lineamientos de arquitectura de las aplicaciones de procesos, gobierno y analítica de datos.
 - Modelo de Gobierno: Gobierno para asegurar el éxito del programa.

| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 4/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Formación y Gestión del Cambio: El plan de transferencia, apalancándose en la gestión del conocimiento, herramientas de colaboración, mentorías, auditorías, formación, sesiones especiales y herramientas que faciliten el autoaprendizaje.
- Soporte y Administración de la Plataforma: El soporte, actualización y administración de las plataformas incluyendo la monitorización de la seguridad.
- Cómo enfocar la implantación de un Oficina de Gobierno, su misión y buenas prácticas metodológicas.
- Cómo realizar evaluación de la madurez en la nube pública y establecer plan de acción de los aspectos a mejorar.
- Construcción de una comparativa y selección de herramientas (ya existentes y de mercado) de soporte a la función del centro de excelencia de gobierno de los datos.
- Dado el volumen de información que se maneja en documentos y/o formatos de audio y video, sería interesante incluir servicios de procesamiento de lenguaje natural que preparen la información para su posterior almacenamiento y consumo.
- Definición de los roles y responsabilidades tanto del centro de excelencia analítico, de las áreas usuarias de información (Negocio) y las áreas tecnológicas.
- Definición del modelo organizativo en torno al Gobierno del dato.
- Definición e implantación de métricas del despliegue del centro de excelencia de gobierno de los datos y la factoría de analítica de datos de la Junta de Andalucía, de manera complementaria a la medición de los niveles de servicio prestados por la factoría de analítica de datos.
- Definición e implantación de un marketplace de datos y casos de uso generados por la factoría de analítica de datos y puestos a disposición de la Junta de Andalucía y de actores externos en caso necesario.
- Definición e implantación del catálogo de servicios del centro de excelencia de gobierno del dato y de la factoría de analítica de datos de la Junta de Andalucía.
- Dentro de la publicación de APIs, convendría identificar de forma explícita, la integración con fuentes públicas, desde el punto de vista no de consumo de datos, sino desde la posible aportación de datos desde el datalake de la Junta de Andalucía a otros proyectos públicos, a los que se pueda aportar datos.
- Diseño e implantación del modelo de valoración económica y de monetización de los datos y de la analítica para los diferentes casos de uso planteados.
- En la oficina de proyectos (PMO) el soporte legal en materia de seguridad y protección de datos debe estar circunscrito.
- En proyectos de esta índole se suele disponer un servicio transversal de monitorización, gestión y administración de la infraestructura en Cloud, que deberá ser concretado en su nivel de profundidad: L1, L2 o L3 y en la cobertura horaria 8x5, 24x7 u otra.
- Enfoque / roadmap en el que la Junta de Andalucía podrá disponer de una visión del enfoque de la arquitectura funcional, ordenada y sostenible en el tiempo, que debe guiar la implantación de los diferentes casos de uso funcionales particulares a Demanda.
- Enfoque de dimensionamiento progresivo alienado con la visión de Línea Base y visión Bajo demanda.
- Entre las tareas de gobierno del dato se llevan a cabo tareas de mucho esfuerzo y criticidad como son los levantamientos de los EDM. Este tipo de procesos se puede abordar con actividades de minería de procesos que agilizan enormemente el entendimiento del comportamiento de los procesos de analítica.

| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 5/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Es necesario incorporar el análisis legal y jurídico de los diferentes casos de uso, tanto desde un punto de vista de privacidad de los propios datos, necesidades de anonimización o uso final de los mismos, para que se implementen las políticas necesarias, incluso se descarten algunos casos de uso, o se modifique su alcance. Este análisis debe estar acompañado por un estudio de viabilidad de los diferentes casos de uso que se debe incorporar dentro el ciclo metodológico de priorización y valoración del riesgo de cada uno de ellos y así ayudar a la ADA a hacer un uso eficiente de los recursos.
- Establecer KPIs a medir y la prioridad de los mismos.
- Evaluación ética de los algoritmos de inteligencia artificial: como elemento diferenciador, la Unión Europea, aboga por el uso responsable de las tecnologías, y a la adopción de automatismos, promoviendo el uso de tecnologías con responsabilidad. De cara al funcionamiento de este servicio, sería muy interesante que tanto desde el punto de vista del gobierno, como de la factoría y de la oficina de gestión de proyectos, se incluyeran tareas referentes a la ética:
 - Gobierno: definir niveles y valoraciones de ética en la adopción de inteligencia artificial.
 - Oficina de Proyecto: realizar tareas de evaluación de ética, y ver si hay conflictos o cuestiones a considerar
 - Factoría: desarrollar los algoritmos y el procesamiento de datos de acuerdo con criterios éticos definidos."
- Exponer si será necesario ofrecer soporte fuera de horario de oficina o en fin de semana y festivos? Es una variable importante a determinar no reflejada.
- Gestión del cambio y comunicación enfocada a impulsar la iniciativa del centro de excelencia de gobierno de los datos y la factoría de analítica de datos de la Junta de Andalucía.
- Incluir paulatinamente un set de servicios adicionales cuyo objetivo sea la medición y aseguramiento del valor aportado a los distintos departamentos de la Junta de Andalucía a los que se preste servicio.
 - Oficina de medición de valor: Creación de la CVO (Chief Value Officer), cuyo objetivo es la medición continua del valor aportado (ROI, valor al ciudadano, eficiencias, etc.) y la priorización de los casos de uso, fomento de los casos de uso exitosos entre distintas consejerías y departamentos de la Junta.
 - Oficina de compliance: Creación de una oficina que actúe como garante de la seguridad, GDPR y compliance con recursos altamente especializados.
 - Análisis de la organización, modelos operativos y roles y personas relacionadas con los servicios de Data & Analytics dentro de los diversos Organismos de la Junta de Andalucía".
- Incorporar aspectos de ética en el uso de los datos referida a las buenas prácticas en torno a cómo se recogen, comparten y utilizan los datos.
- Orientar el modelo operativo de la factoría analítica de datos al uso de metodologías Agile y a XOps, utilizando como punto de partida las mejores prácticas de DevOps.
- Podría ser interesante plantear un servicio de ‘Observatorio’ de resultados de la implantación de los diferentes casos de uso analíticos.
- Realizar talleres que identifiquen claramente los objetivos de negocio.
- Recomendamos un análisis de la arquitectura actual tanto informacional como operacional ya que puede también condicionar la elección de arquitectura y proveedor.
- Separar Analítica de Datos e Ingeniería de Datos con un equipo de Arquitectura dedicado a Plataformas y Hybrid-Multi-Cloud dentro de una estrategia DataOps.

| | | | |
|--------------|--------------------------------|---|-------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 6/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Servicios de analítica apoyada en servicios GIS que ofrecen unas capacidades de análisis muy demandadas actualmente.
- Tener en cuenta la tipología de iniciativas a desarrollar y sus requerimientos funcionales para la definición de la arquitectura y la selección del proveedor.
- Valorar la posibilidad de, proactivamente a partir del conocimiento cruzado de la información sectorial que se vaya generando en el proyecto, implantar un Servicio de “descubrimiento” de nuevos casos de uso analíticos, desde la PMO y soportado en los analistas de negocio del equipo.

De igual forma, recibimos respuestas indicando que les parecía correcto el planteamiento hecho en la memoria justificativa de la presente consulta.

4. LINEA BASE DE SERVICIOS

Las respuestas recibidas muestran una gran variedad y heterogeneidad en cuanto a los perfiles definidos, así como a su titulación y certificaciones más habituales, porcentaje de dedicación y tarifa hora.

Por ello, se ha decidido tomar la información correspondiente a aquellos perfiles que pueden ser asimilables a los mostrados en las siguientes tablas, registrando en ellas las características de titulación, certificaciones y dedicación más reflejadas entre las distintas respuestas.

Estimación de los perfiles profesionales de la línea base, con indicación de las titulaciones y certificaciones más habituales para dichos perfiles.

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|-----------------------------|--|---|
| Consultor Gobierno del Dato | Titulación universitaria: MECES 3 (Grado + Máster oficial), Licenciatura o Ingeniería superior o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en proyectos de Gobierno de Dato. | Estudios de postgrado (Máster, Diploma universitario, etc.) en Gobierno del Dato (recomendable). Certificación activa en algún tipo de metodologías ágiles aplicables al desarrollo de software (ScrumMaster, Scrum Manager o similar). Certificación activa en alguna herramienta de Gobierno del Dato (DAMA, Data Governance Framework...). |
| Jefe de proyecto | Titulación universitaria: MECES 3 (Grado + Máster oficial), Licenciatura o Ingeniería superior o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 8 años en proyectos de | Certificación activa en dirección de proyectos (PMP o similar). Certificación activa en gestión de servicios (ITIL V3 o superior). Certificaciones básica en servicios en Cloud (Amazon Web Services, Microsoft Azure, Oracle |

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|---|--|---|
| | desarrollo en entornos de analítica de datos (sistemas de inteligencia de negocio y/o analítica avanzada). Experiencia de al menos 1 año en proyectos utilizando servicios o plataformas en la nube (SaaS/PaaS) para el desarrollo de soluciones de inteligencia de negocio (Business Intelligence), soluciones de analítica avanzada (Inteligencia Artificial) o de Big Data. | Cloud, Google Cloud). Certificación activa en algún tipo de metodologías ágiles aplicables al desarrollo de software (ScrumMaster, Scrum Manager o similar). |
| Consultor de negocio especialista en analítica de datos | Titulación universitaria: MECES 2, Diplomatura o Ingeniería técnica o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en proyectos relacionados de analítica de datos (ya sea descriptiva, estadística, predictiva o prescriptiva). Experiencia de 2 años utilizando servicios o plataformas en la nube (SaaS/PaaS). | Certificaciones técnicas en servicios en Cloud en ingeniería/arquitectura de datos (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud). |
| Arquitecto de datos en la nube | Titulación universitaria: MECES 2, Diplomatura o Ingeniería técnica o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en proyectos de desarrollo en entornos de analítica de datos Experiencia de al menos 3 años en proyectos utilizando servicios o plataformas en la nube (SaaS/PaaS) para el desarrollo de soluciones de inteligencia de negocio. | Certificaciones técnicas en servicios en Cloud en ingeniería/arquitectura de datos (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud). Certificación activa en algún tipo de metodologías ágiles aplicables al desarrollo de software (ScrumMaster, Scrum Manager o similar) |
| Ingeniero de pipeline de datos | Grado, Licenciado/a o Ingeniero/a superior o sus equivalencias en | 4 años en proyectos de desarrollo de analítica avanzada. |

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|---|---|---|
| | cualquiera de las áreas de ingeniería, informática o ciencias | 2 años en proyectos utilizando servicios en la nube |
| Ingeniero de plataforma de datos en la nube | Titulación universitaria: MECES 2, Diplomatura o Ingeniería técnica o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en proyectos relacionados de analítica de datos (ya sea descriptiva, estadística, predictiva o prescriptiva). Experiencia de 2 años utilizando servicios o plataformas en la nube (SaaS/PaaS). | Certificaciones técnicas en servicios en Cloud en ingeniería/arquitectura de datos (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud). |
| Analista de datos SQL-BI | Titulación universitaria: MECES 2, Diplomatura o Ingeniería técnica o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 3 años en proyectos de desarrollo de sistemas de inteligencia de negocio (Business Intelligence). Experiencia de al menos 1 año en proyectos utilizando servicios o plataformas en la nube (SaaS/PaaS) para el desarrollo de soluciones de inteligencia de negocio (Business Intelligence), soluciones de analítica avanzada (Inteligencia Artificial) o de Big Data. | Certificación en herramientas y plataformas de inteligencia de negocio (Qlik, Oracle, SAP, Power BI, Pentaho, ...). Certificaciones básicas en servicios en Cloud (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud). |

Estimación del número de recursos y porcentaje de dedicación por perfil profesional de la línea base en función de la experiencia de la empresa en proyectos similares de analítica corporativa.

| Perfil | % dedicación |
|-----------------------------|--------------|
| Consultor Gobierno del Dato | 50% |

| Perfil | % dedicación |
|---|--------------|
| Jefe de proyecto | 100% |
| Consultor de negocio especialista en analítica de datos | 100% |
| Arquitecto de datos en la nube | 100% |
| Ingeniero de pipeline de datos | 100% |
| Ingeniero de plataforma de datos en la nube | 100% |
| Analista de datos SQL-BI | 100% |

Identificación de la tarifa hora por perfil profesional de la línea base.

Para identificar una tarifa hora de referencia por perfil profesional en la línea base, se ha realizado un análisis detallado de todos los datos aportados por las empresas, realizándose una homogeneización de la denominación de los perfiles, una asimilación de perfiles, descartando valores extremos y, finalmente, se ha seleccionado la mediana estadística de los precios obtenidos.

Durante el análisis realizado, se ha llegado a la conclusión de que no es necesario introducir la diferenciación entre perfiles junior y senior, por las propias características recogidas en los perfiles. Por tanto, aunque todos los precios aportados han sido tenidos en cuenta, se han condensado los resultados en una sola tarifa.

Los resultados obtenidos se muestran en la siguiente tabla:

| Perfil | Tarifa hora (IVA no incl.) |
|---|----------------------------|
| Consultor Gobierno del Dato | 62,25 € |
| Jefe de proyecto | 77,25 € |
| Consultor de negocio especialista en analítica de datos | 60,00 € |
| Arquitecto de datos en la nube | 63,58 € |
| Ingeniero de pipeline de datos | 55,00 € |
| Ingeniero de plataforma de datos en la nube | 56,29 € |
| Analista de datos SQL-BI | 52,40 € |

5. SERVICIOS A DEMANDA

Las respuestas recibidas muestran una gran variedad y heterogeneidad en cuanto a los perfiles definidos, así como a su titulación y certificaciones más habituales, porcentaje de dedicación y tarifa hora.

Por ello, se ha decidido tomar la información correspondiente a aquellos perfiles que pueden ser asimilables a los mostrados en las siguientes tablas, registrando en ellas las características de titulación, certificaciones y dedicación más reflejadas entre las distintas respuestas.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 10/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

Estimación de los perfiles profesionales a demanda, con indicación de las titulaciones y certificaciones más habituales para dichos perfiles.

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|---|--|--|
| Gestor de proyectos | Titulación universitaria: MECES 3 (Grado + Máster oficial), Licenciatura o Ingeniería superior o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en Gestión de Proyectos y Oficinas de Gestión de Proyectos. | Certificación activa en dirección de proyectos (PMP o similar). Certificación activa en gestión de servicios (ITIL V3 o superior). |
| Formador especialista en analítica de datos | Ingeniería Técnica, Diplomatura o Grado, o cualquier titulación superior a esta | |
| Consultor de cumplimiento legal | Titulación universitaria: Licenciado o Ingeniero superior o sus equivalencias en cualquiera de las áreas de ingeniería, informática, ciencias o derecho y humanidades | |
| Científico de datos | Titulación universitaria: MECES 3 (Grado + Máster oficial), Licenciatura o Ingeniería superior o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia de al menos 5 años en proyectos relacionados de analítica de datos (ya sea descriptiva, estadística, predictiva o prescriptiva) Experiencia de al menos 3 años en proyectos con implementación de algoritmos y desarrollo integral de modelos analíticos (Machine Learning/Deep Learning) Experiencia de 2 años utilizando servicios o | Estudios de postgrado (Máster, Diploma universitario, etc) en Big Data y/o Data Science (recomendable) Certificación en herramientas o soluciones de analítica avanzada en Cloud (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud, IBM Cloud) Certificación activa en algún tipo de metodologías ágiles aplicables al desarrollo de software (ScrumMaster, Scrum Manager o similar) |

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|--|---|--|
| | plataformas en la nube (SaaS/PaaS) | |
| Desarrollador | Titulación universitaria: MECES 2, Grado, Diplomatura o Ingeniería Técnico en cualquiera de las áreas de ingeniería, informática o ciencias. Experiencia profesional contrastable de al menos cinco años como analista y programador con lenguajes como Java, HTML, CSS, JavaScript, ECMAScript. Y desarrollo de APIs y Web services | Certificaciones básicas en servicios en Cloud (Amazon Web Services, Microsoft Azure, Oracle Cloud, Google Cloud). Certificación activa en algún tipo de metodologías ágiles aplicables al desarrollo de software (ScrumMaster, Scrum Manager o similar). |
| Diseñador gráfico | Titulación media | Diseño Gráfico o equivalente |
| Anotador y depurador de datos | - Experiencia y habilidades en áreas como programación, matemáticas e informática - Titulación universitaria: Grado, Licenciado o ingeniero superior o todas sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias | - CCA Spark and Hadoop Developer - AWS Certified Data Analytics – Specialty (DAS-C01) - Microsoft Certified: Data Analyst Associate - Google Professional Data Engineer |
| Consultor de soporte del proveedor de servicios en la nube | - Titulación universitaria: Grado, Licenciado o Ingeniero superior o sus equivalencias en cualquiera de las áreas de ingeniería, informática o ciencias - Profundo conocimiento de las nuevas tecnologías, entiende la relación entre ellas y sabe cómo se pueden integrar y combinar para resolver eficazmente cualquier problema relacionado con los datos | - AWS Certified Solutions Architect – Associate (SAAC02) - AWS Solution Architect Profesional - Microsoft Certified: Azure Solutions Architect Expert - AWS Technical Professional - AWS DevDay Online - IBM Certified Solution Architect - Cloud Pak for Data V2.5 - Google |

| Perfil | Titulación Requerida | Certificaciones Requeridas |
|--------|----------------------|------------------------------|
| | | Professional Cloud Architect |

Estimación del número anual de horas por perfil profesional a demanda en función de la experiencia de la empresa en proyectos similares de analítica corporativa.

| Perfil | Numero anual de horas |
|--|-----------------------|
| Gestor de proyectos | 3520 |
| Formador especialista en analítica de datos | 3520 |
| Consultor de cumplimiento legal | 3520 |
| Científico de datos | 3520 |
| Desarrollador | 3520 |
| Diseñador gráfico | 3520 |
| Anotador y depurador de datos | 3520 |
| Consultor de soporte del proveedor de servicios en la nube | 1760 |

Identificación de la tarifa hora por perfil profesional a demanda.

Para identificar una tarifa hora de referencia por perfil profesional a demanda, se ha realizado un análisis detallado de todos los datos aportados por las empresas, realizándose una homogeneización de la denominación de los perfiles, una asimilación de perfiles, descartando valores extremos y, finalmente, se ha seleccionado la mediana estadística de los precios obtenidos. Los resultados inicialmente obtenidos por perfil a demanda, en promedio, han sido un 3% superiores a los recogidos en los mismos perfiles de la línea base, existiendo un caso en el que el perfil a demanda es un 13% superior a su equivalente en la línea base, y otro caso en el que el perfil a demanda es un 9% inferior al equivalente en la línea base. Finalmente, se ha considerado que la diferencia encontrada no es significativa y se ha tomado una tarifa unificada para los perfiles de la línea base y a demanda.

Por otro lado, al igual que en el caso de la línea base, se ha llegado a la conclusión de que no es necesario introducir la diferenciación entre perfiles junior y senior, por las propias características recogidas en los perfiles. Por tanto, aunque todos los precios aportados han sido tenidos en cuenta, se han condensado los resultados en una sola tarifa.

Los resultados obtenidos se muestran en la siguiente tabla:

| Perfil | Tarifa hora (IVA no incl.) |
|---|----------------------------|
| Gestor de proyectos | 56,25 € |
| Formador especialista en analítica de datos | 47,10 € |
| Consultor de cumplimiento legal | 63,80 € |

| Perfil | Tarifa hora (IVA no incl.) |
|--|----------------------------|
| Científico de datos | 60,00 € |
| Desarrollador | 51,05 € |
| Diseñador gráfico | 47,74 € |
| Anotador y depurador de datos | 24,10 € |
| Consultor de soporte del proveedor de servicios en la nube | 226,00 € |

6. SUMINISTRO DE SERVICIOS EN NUBE

En función de su experiencia en otros proyectos de este ámbito y alcance, ¿qué enfoque tecnológico en nube considera que es el más acertado teniendo en cuenta que es importante no tener una excesiva dependencia de una nube concreta pero también es importante acelerar los proyectos y reducir la gestión haciendo uso de servicios gestionados?

Las respuestas recibidas muestran una gran variedad y heterogeneidad en cuanto al enfoque y alcance considerado. Este es un resumen y compendio de las opiniones expresadas por las empresas participantes en la consulta preliminar:

En general un enfoque Híbrido combinando piezas en entornos on-premise con otros IaaS/PaaS/SaaS en público y Multicloud con servicios de diferentes proveedores de nube pública suele ser el más propuesto, lo que implica las siguientes ventajas:

- Permite seleccionar el mejor servicio entre todos los proveedores para cada servicio de la solución.
- Deberá permitir una futura migración a otro proveedor cloud, los posibles condicionantes derivados del uso de algún servicio particular deberán minimizarse, por tanto toda funcionalidad debería ser empaquetada (Kubernetes, Docker, etc...) para facilitar el traspaso entre diferentes infraestructuras.
- PaaS ofrece la ventaja de poder migrar a otros entornos fácilmente, aunque suele ser algo más caro tanto el mantenimiento como el desarrollo de esa solución con perfiles más especializados.
- Proporciona capacidad de negociación con diferentes proveedores.
- La escalabilidad, en el caso del IaaS se verá afectada por el procesamiento horizontal del software utilizado. Siendo los modelos PaaS y SaaS los que aseguren una escalabilidad óptima.
- Previene excesivas dependencias tecnológicas.
- Combina la rapidez de adaptación de entornos IaaS con la potencia y capacidades de entornos PaaS.
- Elimina posibles problemas de latencias con la capacidad de on-premise.
- SaaS ofrece las mayores ventajas en cuanto a precio y velocidad de desarrollo, pero lógicamente te atas mucho más a la plataforma en la que trabajas.
- Facilita la selección de personal experto, al ampliar el número de posibles tecnologías.
- El enfoque que realmente permitirá explotar al máximo las capacidades cloud, siempre que la plataforma lo permita, es sobre un modelo de servicios PaaS o SaaS, ya que es con estos modelos cuando realmente se le saca partido a la reducción de costes y donde la disminución de la carga en la gestión de recursos es importante. En la mayoría de los casos no implica que exista una dependencia del proveedor de cloud, ya que en la mayoría se basan en software Open Source, lo que permitiría migrar el código de un proveedor a otro sin un elevado coste.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 14/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Es también más adecuada la aproximación PaaS para explotar la modularidad y facilidad de integración entre servicios independizando el ciclo de vida de cada componente dentro de la solución global adoptada.
- Por último, los perfiles técnicos necesarios en una arquitectura compuesta con servicios PaaS pueden centrarse completamente en el objetivo funcional y por tanto en los resultados finales para los usuarios ya que tareas eminentemente técnicas de infraestructura no deben ser abordadas.
- Habitualmente se busca un enfoque de compromiso entre IaaS vs PaaS, en función de la potencial necesidad de interoperabilidad de servicios de plataforma y coste de desarrollo/mantenimiento de funcionalidades existentes en modo PaaS. Lo más común es usar PaaS en aquellos servicios con un beneficio relevante y cuya portabilidad a otro proveedor sea acotable y de coste predecible.
- PaaS permite aprovechar mucho más las ventajas que proporciona una nube pública, ya que reduce en gran medida la gestión y acelera en gran medida los proyectos al permitir hacer uso de servicios ya existentes sin necesidad de implementarlos, pero puede provocar dependencia de dicho proveedor y de sus soluciones exclusivas.

Sin embargo, este enfoque presenta también una serie de retos a considerar:

- El modo de contratación puede ser complejo al combinar múltiples proveedores con diferentes condiciones.
- La arquitectura tecnológica es más compleja.
- Precisa de una mayor capa de gestión (aunque esta puede ser proporcionada por proveedores específicos).
- El plazo hasta disponer de toda la solución completa es más largo.
- La opción IaaS, aunque lógicamente es más fácil moverte a otras Cloud, los costes son mucho más altos y apenas ofrece las ventajas que tiene la nube.

Un enfoque tecnológico basado en Cloud debe basarse en los siguientes pilares:

- Debe soportar cargas de trabajo con un rendimiento al menos como el esperado del mejor CPD On-premise. Para ello debe estar diseñada aprovechando las características de alto rendimiento de la infraestructura de nube como pueden ser las redes de baja latencia, clustering de máquinas con acceso RDMA entre ellas, computación y almacenamiento flexible, etc. así como dar respuesta a otros requerimientos como sistemas de recuperación frente a desastres, gobernabilidad y control preciso del consumo de recursos.
- Debe permitir cargas de trabajo tanto tradicionales como las nuevas aplicaciones nativas Cloud con facilidades para la migración de entornos actuales de forma eficaz que puedan aprovecharse de toda la nueva infraestructura y que la integración entre ambos mundos sea posible y eficaz.
- Debe ir un paso más allá en los componentes de arquitectura basándose en servicios suministrados como Plataforma, servicios Autónomos y servicios manejados. El planteamiento de sistemas de DR debe ser transparente, flexible en implementación (Dominios de Disponibilidad, Dominios de Fallo o Regiones), flexible de consumo de recursos (bajo mínimos para réplica, consumo completo sólo si se utiliza)
- Debe contemplar una solución viable y eficaz para entornos híbridos de forma que, si no todo puede ser subido a la nube, sea la nube la que llegue al CPD con los mismos criterios expuestos.
- La seguridad debe ser la base, no una opción ni un coste aparte.

La Agencia Digital de Andalucía se plantea hacer del dato un elemento clave en la toma de decisiones, de forma que estas sean más eficientes y productivas. Para obtener todo el valor de múltiples fuentes de datos cada vez con mayor volumen, es necesaria una solución que proporcione agilidad, escalabilidad y eficiencia sin olvidar valores básicos en toda infraestructura de IT como son la seguridad y la disponibilidad.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 15/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

La solución adoptada debe incluir un modelo de gobernanza centralizado, en el que se definan el catálogo de servicios para la preparación, análisis y visualización de los datos, automatizando todas las tareas posibles, teniendo así, procedimientos reutilizables y eficientes.

Al iniciar un proceso de migración – implantación en el Cloud lo más importante a tener en cuenta es la capacidad que tenga el entorno o entornos seleccionados para poder albergar toda la funcionalidad necesaria que dé respuesta a las necesidades funcionales de la Junta de Andalucía.

Respecto al enfoque de nube pública frente a nube privada, para una arquitectura de datos moderna es la de un lago de datos en el que diferentes fuentes de datos coexisten en un único entorno, son gobernadas conjuntamente y pueden ser accedidas por diferentes unidades de negocio en función de sus necesidades. se propone un despliegue de la solución en nube pública con una conexión hacia los CPD de la Junta de Andalucía. En aquellos casos en los que el dato pueda residir de forma permanente en la nube, se alojarán en el lago de datos y si tuviesen que persistir en los CPD, se replicarán bajo demanda hacia la nube.

Se deberá evaluar un despliegue de la solución en nube pública con una conexión hacia los CPD (Centro de Procesamiento de Datos) de la Junta de Andalucía. En aquellos casos en los que el dato pueda residir de forma permanente en la nube, se alojarán en el lago de datos y si tuviesen que persistir en los CPD, se replicarán bajo demanda hacia la nube.

Los clientes que usan múltiples proveedores de cloud pública tienden a tener la mayoría de las cargas de trabajo en un proveedor y las cargas de trabajo restantes (bien por su especificad o por mantener alternativas) en un proveedor secundario o incluso divididas entre múltiples proveedores complementarios.

Los servicios de Cloud Pública son muy similares en todos los fabricantes, si bien cada uno puede disponer de algunos aspectos diferenciadores.

Una arquitectura de datos moderna debe permitir:

- Construir de forma rápida un lago de datos a escala: Un emplazamiento del dato que permita el acceso desde múltiples herramientas utilizando formatos estándar y que sea altamente escalable, disponible, seguro y flexible.
- Servicios de datos para cada uso: Una única herramienta no sirve para tratar todos los tipos de datos. Es necesario contar con un conjunto de servicios que permitan procesar datos tanto en tiempo real como en lotes, datos estructurados, semiestructurados o no estructurados y todo ello con la escala, flexibilidad y eficiencia en costes que demanda la misión de la Agencia Digital de Andalucía.
- Movimiento ilimitado de datos: Los datos pueden estar alojados en diferentes sistemas. La solución debe ser capaz de mover el dato entre servicios y fuentes de datos sin limitaciones. Para facilitar esta tarea de movimientos entre diferentes repositorios de datos, es necesario contar con servicios de integración de datos que proporcionen a desarrolladores, analistas y científicos de datos la capacidad de crear flujos de datos que transformen, normalicen y cataloguen las fuentes de datos.
- Gobernanza unificada: Una de las piezas más importantes es la capacidad de autorizar, gestionar y auditar el acceso a los diferentes repositorios de datos. Sin las herramientas adecuadas, este proceso puede ser complejo, largo y propenso a errores.
- Rendimiento y eficiencia en costes: El rendimiento a escala de la plataforma junto con la capacidad de ofrecer un modelo de costes sostenible.

Al lago de datos se le proporcionan herramientas de transferencia o replicación de datos para que estos puedan ser explotados en nube sin impactar al uso primario que de estos se hace actualmente. Diseminar el lago de datos en más de un proveedor de nube pública conlleva crear diferentes silos de

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 16/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

datos, dificultando la explotación transversal de los mismos y, sobre todo, la gobernanza única de la plataforma.

Un enfoque Unicloud es mucho más rápido de implementar, precisa mucho menos gestión, pero a su vez tiene inconvenientes:

- Introduce dependencias de un proveedor que puede dificultar o impedir la portabilidad a otros proveedores.
- Reduce las posibilidades de negociación económica.
- Se encuentra limitado a los servicios que presta dicho proveedor, así como al número de expertos en la misma.

El enfoque Unicloud/MultiCloud, suele ser un proceso evolutivo. Debido a que la transición al cloud requiere de nuevas capacidades y formas diferentes de operar respecto a los entornos on-premise, se suele optar por un enfoque single cloud hasta que se tiene la madurez operativa en Cloud para poder abordar el multicloud. Frecuentemente la aproximación de la mayoría de los clientes suele ser tener un cloud vendor de referencia y otro para casos adicionales. Algunos clientes con una madurez muy alta en entornos cloud van a estrategias best-of-breed, buscando el mejor servicio de cada cloud vendor. Esto conlleva complejidad adicional porque el ciclo de vida del dato y su gobierno se vuelve mucho más complejo, pudiendo tener datos replicados en diversas clouds, con las dificultades de consistencia que conlleva.

Es muy difícil, si no imposible, tener una estrategia Unicloud, y menos cuando se intenta dar un servicio a muchos clientes de diferentes tipologías con un alto componente de soluciones on-premise.

Es importante recordar que la versión y funcionalidad de la Base de Datos es la misma tanto para OnPremise como para Cloud, aunque mientras que en OnPremise tendremos que instalar, administrar, mantener, etc. en Cloud serán servicios manejados. Ciertamente, un paso siguiente son los servicios autónomos que sólo están disponibles en Cloud (Pública o At Customer) y componentes como Golden Gate o Data Integrator, funcionalmente son similares pero la implementación Cloud es un servicio manejado simplificado frente a OnPremise.

La escalabilidad, flexibilidad y la cantidad de capacidades innovadoras que ofrecen los proveedores de cloud pública se ha ampliado considerablemente a medida que dicho ecosistema ha madurado. Para aprovechar en mayor medida estas capacidades es un mayor número de casos de uso, se recomienda adoptar arquitecturas multicloud.

Se recomienda la implementación de soluciones que permitan una gestión de todos los entornos de almacenamiento desde una única consola central.

Los distintos proveedores cloud aportan y sugieren el uso de sus servicios nativos, esto aporta una mayor velocidad de adopción del cloud e incluso beneficios económicos y simplicidad de operación, y gestión frente a un posible vendor lock in. Es necesario realizar un análisis de la situación actual y las expectativas antes de optar por una decisión.

Una propuesta es Nube Pública Única por entidad u organismo, pero con un equipo de trabajo y Gobierno Multi-Cloud que tenga las capacidades técnicas para pivotar en la nube más adecuada para cada entidad. El equipo de trabajo deberá realizar un análisis técnico-económico de la nube adecuada, proponiendo y justificando la creación de lago de datos, migraciones o replicaciones de éstos si el proyecto lo requiere. La clave es tener un equipo multidisciplinar coordinado que tenga la capacidad de aprovecharse de lo mejor de cada nube en cada momento. En cualquier caso, también sería interesante contar con un integrador que tenga una visión completa e integral del mundo de la nube pública.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 17/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

Existen tres áreas principales a considerar en una adopción multicloud:

- Estrategia multicloud. Las decisiones estratégicas que debe tomarse al adoptar el multicloud son muy diversas. De hecho, el multicloud no debe considerarse la estrategia de la organización en sí misma. En su lugar, la estrategia multicloud debe ser parte de la estrategia general Cloud de la organización.
- Gestión multicloud. Una vez que se ha decidido adoptar una estrategia multicloud, se debe realizar el arduo trabajo de identificar cómo escalar y administrar adecuadamente los entornos multicloud, contemplando: gobierno, arquitectura, distribución de servicios, operaciones.
- Arquitectura multicloud. En una arquitectura multicloud, las cargas de trabajo abarcan múltiples proveedores. Algunas cargas de trabajo pueden ser una combinación de diferentes tipos de proveedores y proveedores de servicios. Por ejemplo, puede haber IaaS y PaaS integrados (IaaS+PaaS) de un solo proveedor, junto con otros elementos de PaaS de múltiples proveedores, como PaaS de integración (iPaaS), componentes verticales específicos o servicios de API, como un servicio de voz. API. Algunas cargas de trabajo pueden ser redundantes e implementarse en diferentes proveedores de cloud en diferentes momentos; por ejemplo, una aplicación de trabajo Batch podría implementarse en el proveedor de cloud menos costoso.

Lo recomendable sería aplicar soluciones multicloud con el objetivo de aprovechar las ventajas y funcionalidades que ofrecen los distintos proveedores cloud y obtener la máxima eficiencia. Al mismo tiempo que para ahorrar costes o acceder a mejores precios, mover las cargas de trabajo entre distintas nubes, potenciar el cumplimiento y evitar el bloqueo de un único proveedor. Cada proveedor de nube posee capacidades únicas, por lo que resulta beneficioso combinar nubes cuando se intenta maximizar el rendimiento. Es muy importante asegurar la conectividad adecuada entre los distintos proveedores cloud involucrados.

Mantener una estrategia multicloud implica mayor complejidad e introduce complejos desafíos de integración y operaciones. Es necesario realizar una planificación cuidadosa, disponer de un mayor conjunto de conocimientos, y gozar una madurez operativa y una excelencia en desarrollo e integración de servicios.

Una de las problemáticas más comunes que existen cuando se plantean proyectos de despliegue basados en entornos multicloud es el conocer si realmente va a ser fácil implementar esa capacidad de poder decidir dónde, cómo y cuándo correr nuestras cargas de trabajo.

Las soluciones de almacenamiento de datos han de contar con herramientas de migración de datos totalmente integradas. Así mismo, estas soluciones han de permitir la migración de datos desde los entornos on-premise hacia el cloud público, y viceversa, para así posibilitar una experiencia real de entorno multicloud.

Una arquitectura de datos moderna es la de un lago de datos federado, entendiendo por tal aquel en el que diferentes fuentes de datos coexisten en un único entorno, son gobernadas conjuntamente y pueden ser accedidas por diferentes unidades de negocio en función de sus necesidades. Al lago de datos se le proporcionan herramientas de transferencia o replicación de datos para que estos puedan ser explotados en nube sin impactar al uso primario que de estos se hace actualmente. Esto requerirá mecanismos de conexión entre la nube y la red corporativa de la Junta de Andalucía que garantice esta transferencia de forma ágil y segura.

Para un despliegue completamente nuevo, lo recomendable es llevar a cabo un proyecto en el cual se separe la gestión de la infraestructura de la gestión de las cargas de trabajo. De esta manera, nuestros clientes pueden seguir una estrategia multi-cloud, aprovechando las ventajas que les ofrezcan en cada momento las nubes públicas. El objetivo es variabilizar al máximo todo lo concerniente a plataforma y gestión de cargas, pagando por consumo.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 18/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

Multi-Cloud puede parecer muy atractivo, pero diseminar el lago de datos en más de un proveedor de nube pública conlleva crear diferentes silos de datos, dificultando la explotación transversal de los mismos y, sobre todo, la gobernanza única de la plataforma, ya que cada proveedor trabaja con un paradigma diferente y herramientas dispares. También resulta atractivo poder disfrutar de las mejores herramientas de todos los proveedores para cada caso de uso, sin ataduras a ninguno de ellos y obteniendo las mejores ofertas en un modelo de competencia.

El enfoque tecnológico pasa por poder adoptar mecanismos que permitan un entorno de Cloud Híbrido y Multicloud. Esta estrategia engloba el poder mantener relaciones entre las diferentes nubes, tanto públicas como privadas, sin olvidarnos nunca de las posibilidades de los diferentes proveedores de servicio locales o nacionales (aquí el concepto de multicloud), y así como el seguir interactuando con los actuales entornos on-premise, o centros de datos locales, que hacen uso las corporaciones (concepto de hibridación del cloud integrándolo con el mundo on-premise).

Los modelos híbridos multicloud son una opción en la que se podrá disfrutar de los beneficios de los diferentes proveedores de cloud según el caso de uso, siempre que se mantenga un modelo de gobernanza único y centralizado.

Un enfoque multi-cloud híbrido permite tener un crecimiento de datos escalable a la vez que nos permite automatizar todas las tareas para que de manera sencilla los datos sean procesados de inicio a fin. Permite mejorar los rendimientos (escalabilidad e interoperabilidad) ya que la combinación de diferentes plataformas en la nube permite a las organizaciones centrarse en las características exactas que creen que beneficiarán a sus operaciones sin obligarlas a pagar por características que no tendrán impacto. Otro punto importante es la flexibilidad ya que permite que las aplicaciones y soluciones diseñadas sobre la plataforma permitan ser migradas directamente a otra cloud sin necesidad de procesos de adaptación o reprogramación por estar atado a un solo proveedor. Adicionalmente tiene la ventaja de poder optimizar el gasto. Como ejemplo, una forma concreta de ahorrar dinero con un enfoque multi-cloud híbrido es a través de los costes de ancho de banda. Por último, otro beneficio de esta estrategia multi-cloud es la reducción de riesgos, ya que este modelo de nube múltiple disminuye las posibilidades de vulnerabilidad de la ciberseguridad con su estrategia de diversificación, al distribuir los productos, servicios e infraestructura en diferentes proveedores. Todo esto hace que los servicios sean más resilientes y tengan un menor tiempo de recuperación, en caso de invasión.

Una solución multi-cloud híbrido permite poder tener datos distribuidos (on-premise y on cloud) y poder hacer procesamiento distribuido de estos datos sin necesidad de moverlos entre diferentes plataformas evitando problemas de latencia de datos, costes no planificados por transferencias de datos.

Este tipo de arquitecturas permite montar lagos de datos federados en los que diferentes fuentes de datos coexisten en un único entorno, son gobernadas conjuntamente y pueden ser accedidas por diferentes unidades de negocio en función de sus necesidades independientemente del enfoque tecnológico de las fuentes de información, e incluso si disponen de nubes privadas o públicas particulares, disponiendo de herramientas de transferencia o replicación de datos para que los datos de origen puedan ser explotados en nube sin impactar al uso primario que de estos se hace actualmente.

Estos son los escenarios en los que la nube híbrida mejor permite desarrollar todo el potencial de aunar los mundos del cloud público y privado.

- Cargas de trabajo dinámicas o que cambian frecuentemente.
- Separación de las cargas de trabajo más críticas o sensibles de la que son menos.
- Procesamiento de Big Data.
 - Analítica de datos
 - Los backups de datos, archivados y repositorios múltiples de datos (lago de datos)

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 19/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Transición a la nube de forma gradual.
- Necesidades de capacidad de procesamiento temporal.
- Flexibilidad para el futuro.
- Kubernetes como estándar de facto.

En los proyectos BigData, uno de los hándicaps para abordar una estrategia multicloud, es dónde reside el lago de datos. Aunque la mayor parte de las tecnologías BigData desacoplan el procesamiento del almacenamiento, estos deben estar en la misma localización para que el procesamiento sea óptimo. Un acceso al dato desde distintos proveedores de Cloud implica:

1. Replicación de datos entre los distintos clouds, con los costes que derivan de ello, ya que los proveedores de Cloud no cobran por subir los datos, pero si por extraerlos.
2. Duplicidad de ingesta, ya sea porque el dato se envía a los n proveedores de Cloud o porque se envía a uno y se replica desde este.
3. Desincronización de los datos.
4. Modelo de gobierno más complejo.
5. Mayor dispersión de los datos.
6. Cada proveedor trabaja con un paradigma diferente y herramientas dispares.

En el ámbito particular de Big Data e Inteligencia Artificial, la tendencia en el mercado es la utilización de tecnologías basadas en microservicios y contenedores, que aportan la flexibilidad, independencia y portabilidad que requiere una estrategia multicloud.

Todos los proveedores de Cloud disponen de herramientas no code / low code lo que permite acelerar los desarrollos, minimizar el testing y eliminar el mantenimiento de los desarrollos ya que esto se descarga en el proveedor de Cloud. Para minimizar el riesgo de comprometerse con un único proveedor lo ideal es apoyarse en la mayor medida posible en paquetes de software Open Source.

Es importante tener en cuenta a la hora de evitar dependencias excesivas de proveedores, reducir las necesidades de gestión y facilitar la portabilidad:

- Intentar hacer uso de servicios en nube comunes o que sean compatibles entre diferentes proveedores de nube pública (por ejemplo, mismos servicios de bases de datos analíticos).
- Emplear capas de aislamiento que faciliten la migración sin impacto entre proveedores: por ejemplo, capas de orquestación de contenedores y servicios homologados en diferentes proveedores.
- En caso de interesar el uso de servicios que son específicos de un proveedor y no compatibles con otros, se recomienda construir piezas de arquitectura que aislen el servicio del resto del sistema, facilitando la portabilidad de aquellos componentes no exclusivos.
- Seleccionar proveedores o servicios de estos en los cuales exista un número suficiente de profesionales formados (tanto en el ámbito de desarrollo y arquitectura como de operaciones).
- Elegir servicios con la suficiente implantación en el mercado nacional e internacional que garantice la viabilidad de la solución adoptada.

La propuesta se basaría en un lago de datos ubicado en una Región dentro de la Unión Europea utilizando servicios gestionados basados en estándares de código abierto que liberarán al personal de la Junta de Andalucía de las tareas de mantenimiento de la plataforma y les permitirá centrarse en la innovación y la extracción de valor del dato.

La estrategia de Lake House permitirá contar con un repositorio de datos común para cada entidad sobre el que alojar múltiples fuentes de datos de diferentes entidades, con un modelo de gobernanza único y con acceso al catálogo de servicios de preparación, análisis y visualización que ya están utilizando decenas de miles de clientes por todo el mundo.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 20/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

Los componentes necesarios para una Agencia Digital capaz de afrontar tales retos deben responder efectivamente tanto al procesamiento por lotes (BATCH) como en tiempo real (STREAMING), incluida la ingesta de datos, el almacenamiento, las soluciones y componentes de BI automatizados para el procesamiento analítico avanzado.

Para lograr estos requisitos, normalmente se implementan arquitecturas Lambda en la nube. La arquitectura Lambda es una arquitectura de procesamiento de datos diseñada para gestionar cantidades masivas de datos aprovechando tanto los métodos de procesamiento por lotes y como los métodos de procesamiento en tiempo real. Este tipo de arquitecturas permiten el procesamiento/consulta de datos complejos desde la capa Batch y la consulta y el procesamiento en tiempo real desde la capa Streaming.

Se recomienda que estos recursos se implementen en una plataforma desatendida, como un servicio (PaaS – Platform as a Service), que reduce el esfuerzo de administración de TI y simplifica el mantenimiento de la plataforma, compartiendo responsabilidades con el proveedor de servicios en la nube seleccionado.

En función de su experiencia en otros proyectos de este ámbito, ¿qué requisitos funcionales deberían exigirse a la nube pública en la que se lleven a cabo los trabajos para poder llevar a cabo los trabajos con garantías de éxito?

Este es un resumen de requisitos funcionales que planteados por las empresas:

- Acceso a la innovación.
- Acceso a otras soluciones basadas en la nube.
- Acceso a proveedores, socios o entidades de las AA.PP.
- Acceso a una amplia gama de capacidades
- Acceso a una gama más amplia de geografías.
- Almacenamiento altamente disponible que aloje los datos en al menos dos zonas de disponibilidad, entendiéndolas como diferentes dominios de fallo.
- Almacenamiento altamente disponible que aloje los datos en al menos dos zonas de disponibilidad, entendiéndolas como diferentes dominios de fallo.
- Aprovechando las ventajas del cloud computing
 - Ahorro de Costes: Pago por uso.
 - Velocidad y agilidad: En unos pocos clics y en cuestión de minutos, tener recursos disponibles y desplegados.
 - Alta disponibilidad: Garantizar la disponibilidad de los recursos incluso ante fallos de dependencias.
 - Seguridad: Los proveedores de cloud invierten grandes cantidades de recursos, económicos y humanos, en mejorar y mantener actualizadas las medidas de seguridad.
 - Escalabilidad y flexibilidad: Aumentar o disminuir la capacidad de los recursos en cuestión de minutos.
 - Recuperación de desastres: Además de que los entornos cloud se encuentran replicados en diferentes data centers dentro de una misma región, existe la posibilidad de tener copias en otras regiones, lo que permitiría garantizar la seguridad de los datos en caso de un desastre regional importante.
 - Entorno siempre actualizado: Las plataformas son actualizadas continuamente garantizando tener siempre la última o la mejor versión de cada servicio.
- Capacidades de encriptación de los datos en tránsito y almacenados en el proveedor.
- Capacidades y conocimientos del Cloud.
- Características que garantizarían el éxito en un proyecto de datos de gran escala:

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 21/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Permitir la ingesta de datos de cualquier tipo de origen.
- Poder realizar ingesta de datos en modo batch o en realtime (streaming).
- Costes adaptados al consumo de datos ingestados en cada momento.
- Almacenar y transportar los datos de forma segura.
- Cumplir con las exigencias normativas de seguridad más altas a nivel global como locales.
- Capacidad de auditar y monitorizar que está sucediendo con los datos en cada momento.
- Catálogo de datos centralizado y una herramienta única de gobernanza que permita definir usuarios del Data Lake y los permisos con los que estos cuentan para acceder a cada dataset con la máxima granularidad posible (permitiendo filtrado a nivel de columna y de fila en el caso de datos en formato tabular).
- Catálogo de datos centralizado y una herramienta única de gobernanza que permita definir usuarios del Data Lake y los permisos con los que estos cuentan para acceder a cada dataset con la máxima granularidad posible (permitiendo filtrado a nivel de columna y de fila en el caso de datos en formato tabular).
- Complejidad de herramientas.
- Conceptualización del sistema
- Consideraciones sobre la situación geográfica:
 - Localización con los mismos requisitos normativos y regulatorios.
 - Cercanía física que evite problemas de latencias
- Considerar el coste de esta transmisión de datos.
- Créditos del fabricante para la gestión de pilotos y testeo de servicios
- Cuando la solución se apoye en un servicio gestionado y no en IaaS, éste debe utilizar estándares de código abierto (Apache Hadoop, Apache Spark, HBase, Apache Presto, etc.) que permitan aprovechar código y conocimiento existente en la Junta de Andalucía y una salida sencilla del servicio.
- Cuando la solución se apoye en un servicio gestionado y no en IaaS, éste debe utilizar estándares de código abierto (Apache Hadoop, Apache Spark, HBase, Apache Presto, etc.) que permitan aprovechar código y conocimiento existente en la Junta de Andalucía y una salida sencilla del servicio.
- Debe contener técnicas de analítica avanzada habituales:
 - Data Mining y Text Mining.
 - Aprendizaje automático (Machine Learning).
 - Aprendizaje profundo (Deep Learning).
 - Patrones y tendencias.
 - Visualización y simulación.
 - Análisis semántico.
 - Estadística multivariada.
 - Análisis de grafos.
 - Clustering y algoritmos de clasificación.
 - Real time analytics
- Dentro de las capacidades cognitivas que debe tener una plataforma cloud, se recomienda contar con:
 - Data Science Workbench o Analytics Workbench -> que es una herramienta que permite acelerar su sistema de análisis de data science y machine learning y generará un entorno colaborativo y de autoservicio que facilitará la exploración, la visualización y el modelaje de datos.
 - Reconocimiento y detección de imágenes
 - Procesamiento del lenguaje natural
- Desarrollo de la Plataforma

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 22/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Desarrollo de sistema de explotación IA
- Descuentos y niveles de compromiso.
- Disponer de los servicios homologados acordes a los objetivos del proyecto, entre otros:
 - Almacén de datos en la nube
 - Conectores con múltiples fuentes de datos
 - Servicios analíticos.
 - Capacidades de visualización
 - Servicios ETL
 - Servicios para las funcionalidades del Gobierno del dato.
 - Servicios que ayuden a garantizar la Calidad del Dato
- Disponibilidad de herramientas de búsqueda, visualización y análisis sobre texto y datos no estructurados.
- Disponibilidad de herramientas de ingesta, limpieza, preparación y normalización de datos.
- Disponibilidad de herramientas de migración de bases de datos.
- Disponibilidad de herramientas de tratamiento de datos de streaming.
- Disponibilidad de herramientas que faciliten la creación de lagos de datos.
- Disponibilidad de servicios de análisis de flujos de datos y videos en tiempo real.
- Disponibilidad de servicios de Apache Hadoop y Apache Kafka.
- Disponibilidad de servicios serverless, entendiendo por estos aquellos en los que la infraestructura crece y decrece en función de la demanda de forma transparente al usuario, minimizando costes y asegurando escalabilidad. Estos servicios deben ser altamente disponibles por diseño.
- Disponibilidad de servicios serverless, entendiendo por estos aquellos en los que la infraestructura crece y decrece en función de la demanda de forma transparente al usuario, minimizando costes y asegurando escalabilidad. Estos servicios deben ser altamente disponibles por diseño.
- El entorno Cloud debe tener capacidad para conectarse con otras aplicaciones o entornos a partir de los siguientes métodos:
 - API REST
 - SOA
 - SDK
 - GATEWAY
 - Directamente a entornos On-Premise.
- En cuanto a escalabilidad al menor coste posible.
 - Datalakes escalables, en formatos de almacenamiento de objetos o de ficheros.
 - Servicios de análisis que permitan el rendimiento necesario al menor coste posible.
- Es importante que el proveedor de Cloud disponga de servicios específicos para cada una de las capas necesarias para desplegar una solución end to end.
- Es importante que la nube pública cuente con un catálogo de servicios propios o de terceros (en su “Marketplace”) que faciliten:
 - Un despliegue ágil y flexible de los mismos. Los servicios que vayamos a necesitar de las nubes públicas deben tener APIs y/o herramientas que faciliten la automatización de su gestión (despliegue, repliegue, etc) a nivel de infraestructura.
 - Pago por uso real. Si hay que recurrir a herramientas o productos cuyo coste es fijo o por periodos fijos de tiempo estaremos incurriendo en costes innecesarios y que no estarán en línea con la filosofía de los servicios en la nube. Es importante que todos los componentes que se van a necesitar en momentos puntuales se puedan pagar acorde al uso real que hagamos de los mismos.
- Estrategia de Gobernanza de datos
- Estrategia de salida de la nube.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 23/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Existencia de suficientes profesionales formados en diferentes proveedores de servicios profesionales.
- Es imprescindible, desde el punto de vista de la gestión y la monitorización del dato, dotar de herramientas que permitan gestionar desarrollos de infraestructura híbrida (on-prem y cloud) y multicloud (múltiples nubes), las cuales permitan:
 - Supervisar su infraestructura
 - Optimizar el gasto
 - Controlar y proteger los datos
- Herramienta de monitorización de servicios
- Herramienta de tarificación de servicios
- Integración de fuentes de datos
- Implementación de las características más importantes del LakeHouse que son:
 - Soporte a transacciones (ACID).
 - Integridad del dato.
 - Almacenamiento desacoplado.
 - Almacenamiento Open Source.
 - Soporte de datos semi estructurados.
- Los requisitos funcionales vendrán definidos por los casos de uso de negocio a implementar.
- Manejo de conceptos como DataFabric, DataMesh o propiamente LakeHouse para diferenciar las acciones que se toman en las diferentes fases, así como para diferenciar su implementación.
- Máxima integración de los servicios de analítica de datos y los de Inteligencia Artificial y Machine Learning, de forma que estos últimos formen parte del Data Lake para mejorar la analítica predictiva.
- Máxima integración de los servicios de analítica de datos y los de Inteligencia Artificial y Machine Learning, de forma que estos últimos formen parte del Data Lake para mejorar la analítica predictiva.
- Mecanismos de protección del dato, con cifrado en tránsito y en destino (at-rest) así como herramientas que permitan identificar la presencia de datos de carácter personal y la auditoría de acceso a la información.
- Mecanismos de protección del dato, con cifrado en tránsito y en destino (at-rest) así como herramientas que permitan identificar la presencia de datos de carácter personal y la auditoría de acceso a la información.
- Para gestión de almacenamiento (Data Lake):
 - Ingesta y adquisición de datos: Este componente aborda la capacidad de la ingesta de datos en el ecosistema de la plataforma. Proporciona flexibilidad en poder disponer diferentes fuentes de datos, tipos y métodos de ingestión para cumplir con los requisitos de abastecimiento integrados y escalables de datos multi-estructurados.
 - Almacenamiento y gestión de datos: Este componente aborda la capacidad de almacenar y administrar datos en la plataforma en las capas sin procesar (Raw Layer) y curadas (Curated Layer). Soporta tenencia múltiple (multi-clientes), proporciona opciones de compresión y serialización. Contiene la capacidad de disponer datos heterogéneos de herramientas de ingestión en tiempo real y por ingestión mediante procesos batch.
 - Procesamiento de datos: Este componente aborda la capacidad de procesamiento y la necesidad de un sistema de tenencia múltiple en la plataforma. Proporciona un marco flexible para sistema de tenencia múltiple que admite aplicaciones en tiempo real y aplicaciones por batch para diferentes fuentes y soporta CDC y otros algoritmos de procesamiento paralelo específicos de aplicaciones.
 - Aprovechamiento de datos: Este componente aborda la capacidad de acceso a datos y aprovisionamiento desde la plataforma. Proporciona mecanismos de acceso

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 24/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

de datos consistentes para sistemas externos, usuarios y técnicos a través de las interfaces de Hadoop y procesos personalizados de extracción de datos.

- Gestión de metadatos: Este componente aborda la capacidad de gestión integrada de metadatos en toda la plataforma para datos técnicos, operativos y de negocio, así como el linaje de datos (trazabilidad de los datos) para abordar las necesidades de los usuarios finales, el personal técnico y auditores.
 - Administración y monitorización: Este componente aborda la capacidad para la administración de la plataforma y las actividades de mantenimiento del mismo. Proporciona un marco, un proceso y un conjunto de herramientas necesarias para soportar y monitorear las aplicaciones, las ejecuciones y los servicios que se ejecutan dentro de la plataforma.
 - Gestión de seguridad: Este componente aborda la capacidad de los requisitos de seguridad para la plataforma. Define un marco para la autenticación y autorización dentro de la plataforma y un marco de integración para aplicaciones con múltiple tendencia de los sistemas empresariales para autenticación y autorización.
 - Gobierno: Este componente aborda la capacidad para la evaluación de la calidad de los datos y la elaboración de perfilado de datos (Data Profiling) como parte del flujo de trabajo de procesamiento automatizado, así como el servicio para los usuarios de la plataforma para proporcionar transparencia en la calidad de los datos.
- Plataforma DataOps que debe tener una serie de capacidades que permitan al ADA manejar los datos allá donde tengan más sentido, facilitar sacarles todo su partido, alimentar todos los casos de uso minimizando el tiempo en ponerlos en producción, reducir el coste total de los datos del ADA, mantener los datos seguros y gestionados
 - Plataforma escalable tanto en capacidad de almacenamiento como en cómputo.
 - Plataforma escalable tanto en capacidad de almacenamiento como en cómputo.
 - Pool de servicios suficientemente amplio
 - Posibilidades de aplicar las políticas corporativas de la Junta de Andalucía: Identificación, Autenticación, Auditorías, etc.
 - Proporcionar capacidades de migración a otras plataformas o de usar capas de aislamiento (por ejemplo, plataformas de orquestación de contenedores)
 - Relaciones de proveedores.
 - Servicios que cumplan tanto la Regulación de Protección de Datos de la Unión Europea (GDPR) como el Esquema Nacional de Seguridad, modalidad Alta.
 - Servicios que cumplan tanto la Regulación de Protección de Datos de la Unión Europea (GDPR) como el Esquema Nacional de Seguridad, modalidad Alta.
 - Si bien ciertos tipos de datos estructurados podrían necesitar sus propios repositorios, los servicios deberán ser capaces de acceder al Data Lake para explotar los datos sin necesidad de copias intermedias.
 - Si bien ciertos tipos de datos estructurados podrían necesitar sus propios repositorios, los servicios deberán ser capaces de acceder al Data Lake para explotar los datos sin necesidad de copias intermedias.
 - Solución o servicios asociados implantados en un número suficiente de clientes que garanticen la viabilidad y evolución de los servicios seleccionados.
 - Soportes, correctivos y mantenimiento evolutivo
 - Trabajo de integración adicional.
 - Tener distintos tipos de entorno, dependiendo de la fase de desarrollo en la que nos encontremos:
 - Desarrollo: Entorno en el que se producen todos los futuros entregables.
 - Pre-Producción: Entorno lo más similar posible al de producción en el que se prueba cualquier nuevo entregable antes de subirlo a producción.

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 25/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |

- Producción: Entorno al que accede el usuario final y donde se encuentra alojada la última versión válida puesta a disposición.
- Velocidad líneas de comunicación
- Se recomienda tener funcionalidades de:
 - Devops
 - Dataops (Data Operations, Data Managements, Data Distributions, Data infraestructure, Data security)

En función de su experiencia en otros proyectos con alcance similar a este, ¿a cuánto considera que puede llegar a ascender el presupuesto anual dedicado a costear los servicios en la nube (computación, servicios de analítica de datos gestionados, almacenamiento, transferencia de datos de entrada/salida, productos adquiridos en el market, etc)?

Los valores aportados por las empresas participantes en la consulta preliminar indican que para un periodo de 3 años estos costes estarían entre un 4,7% y un 23,8% del coste total del del caso de uso. El coste total del caso de uso es la suma del coste de servicios en la nube y del coste de los servicios profesionales necesarios para llevar a cabo el desarrollo e implantación del caso de uso (excluyendo Centro de excelencia de Gobierno del dato y PMO).

EL TÉCNICO RESPONSABLE DEL INFORME

Fdo: Víctor Manuel Iglesias Palomo

| | | | |
|--------------|--------------------------------|---|--------------|
| FIRMADO POR | VICTOR MANUEL IGLESIAS PALOMO | 28/06/2022 | PÁGINA 26/26 |
| VERIFICACIÓN | Pk2jm6VER5ETD447VTVEC5MFRK89KS | https://ws050.juntadeandalucia.es/verificarFirma | |