



Propuesta Analítica

Respuesta: Proyecto Big Data y Machine Learning de MiBus (Transporte Masivo de Panamá S.A.)

MiBus (TMPSA)

Nelly Ramea

nelly.ramea@mibus.com.pa



Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Perfil de la empresa

Martinexsa es una empresa innovadora de tecnología con más de 35 años en la región. Principalmente nos dedicamos a brindar soluciones tecnológicas para el almacenamiento, protección, procesamiento y análisis de información a nivel empresarial. Contamos con oficinas en Ciudad de México, Guatemala, El Salvador, Honduras, Costa Rica, Panamá, República Dominicana, Colombia, Ecuador y atendemos Perú, Chile y Argentina desde nuestra oficina en Chile.

El departamento de Big Data y Analytics está dedicado a construir soluciones integrales desde la arquitectura de Data Lakes hasta el desarrollo de proyectos de Data Science basados en la predictibilidad, machine learning, deep learning, inteligencia artificial y tiempo real.

Somos un equipo regional de analítica con más de 8 años de experiencia en diseño, desarrollo e implementación de soluciones de Big Data y Analytics basadas en Hadoop y Spark. Somos partner Silver de Cloudera para la región y utilizamos su ecosistema como tecnologías base para la creación de nuestros proyectos. Adicionalmente somos aliados estratégicos de Splunk, Neo4j, Microsoft Azure, Google Cloud y Amazon Web Services.

Estamos formados por científicos de datos, ingenieros de datos, analistas de datos y administradores de sistemas. Nuestro equipo cuenta con certificaciones internacionales de Data Science y Machine Learning, certificaciones de administración y desarrollo de Cloudera, y certificaciones de Spark como herramienta analítica. +70 certificaciones en total.

Nuestros proyectos se centran en la predictibilidad y en tiempo real, lo que hace que nuestros clientes logren un rendimiento financiero basado en sus desafíos comerciales reales. Hemos desarrollado proyectos para la industria de telecomunicaciones, servicios financieros, retail, utilities, manufactura, pharma, salud, agro y gobierno. En total tenemos documentados +65 casos de éxito en toda Latinoamérica.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Beneficios de Martinexsa como Business Partner

- Construimos modelos analíticos predictivos personalizados y hechos a medida basados en necesidades puntuales y retos de negocio.
- Contamos con una fábrica robusta de científicos de datos y profesionales de los datos que poseen amplia experiencia en diversas industrias. Además, tenemos capacidades para manejar proyectos grandes y complejos abordando casos de negocio de importancia empresarial.
- Cubrimos todas las áreas y roles necesarios para desarrollar proyectos de analítica e implementar productos de datos, esto incluye los Data Scientists, Data Engineers, Data Analysts y IT admins.
- Nuestro equipo está certificado en todas las herramientas que utilizamos dentro de nuestro toolbox basado en tecnologías Cloudera, Spark, Splunk, Neo4j, Microsoft Azure, Google Cloud y Amazon Web Services. Extendemos nuestras credenciales con certificaciones internacionales de Ciencia de Datos, Analítica Avanzada, Machine Learning e Inteligencia Artificial.
- Utilizamos metodologías ágiles en nuestros procesos internos de desarrollo y analítica para mostrar resultados entregables en menos de 4 meses. Definimos un ciclo de vida de los proyectos que lleva desde la detección de las oportunidades analíticas hasta la implementación y puesta en producción de los modelos.
- No somos una caja negra de analítica. Todo nuestro trabajo es transparente para nuestros clientes a través de transferencias de conocimiento y sesiones de resultados. El código es propiedad del cliente bajo derechos compartidos.
- A través de Martinexsa Academy proveemos de capacitación y entrenamiento especializado, y con reconocimiento por parte de los fabricantes con el cual apoyamos a nuestros clientes en el desarrollo de sus capacidades y la creación de sus Centros de Excelencia.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Proyecto Data Science MiBus

Dada la necesidad de MiBus de modernizar su arquitectura de datos para poder procesar mayores volúmenes de información y desarrollar modelos analíticos que le permitan ser más ágiles en el negocio, se contempla un proyecto multifacético con el objetivo de guiar al cliente desde el diseño de la hoja de ruta, tanto de soluciones como de la tecnología habilitadora.

En esta **Fase 0** del Proyecto se pretende sentar las bases del desarrollo posterior. Esto se logrará por medio de un trabajo de diseño de solución **Roadmap Analítico** y el desarrollo de un **Data Lake** con todos los componentes necesarios para los casos de uso futuros.

A continuación se describen ambos servicios junto con sus alcances, cronogramas y entregables. Al finalizar esta fase se entregará una propuesta para la Fase 1 en la cual se abordará el desarrollo y puesta en marcha de los modelos analíticos.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Fase 0: Componentes del Proyecto

Roadmap Analítico

En Martinexsa, proponemos una solución integral a través de un diseño de solución especializada, una para negocio y tecnología. Primero, el **Roadmap Analítico**, que tiene como objetivo identificar los desafíos de negocio que pueden ser abordados mediante analítica de datos, priorizar los casos de uso y establecer un roadmap analítico claro. Adicional a esto se tendrá la implementación de una arquitectura sobre Microsoft Azure.

Roadmap Analítico

Implementación Azure

El proyecto de diseño de solución para el cliente se estructurará en cuatro fases principales, con el objetivo de reorganizar y optimizar la gestión y explotación de sus datos para generar valor y monetización, y así moverse con mayor rapidez en el mercado.

Revisión arquitectura
actual y entrevistas de
negocio

Revisión herramientas y
datos disponibles

Propuesta de
arquitectura y roadmap
analítico

Documentación de
hallazgos y de hoja de
ruta

Adicional se tendrá la implementación de una arquitectura de Microsoft Azure, desplegando los componentes en una modalidad de despliegue mínimo para una vez se comience con fases posteriores donde se realicen diferentes implementaciones, estos componentes se podrán escalar para estos.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica

Objetivos

Los objetivos de la diseño de solución se resumen de la siguiente forma:

1. Diseñar una arquitectura empresarial de administración de datos (EDMS) que permita gobernar los datos como un activo estratégico.
2. Diseñar un único almacén de todos los datos de la empresa, desde el cual se alimenten informes, visualizaciones, análisis de datos.
3. Diseñar una estrategia de datos que tenga capacidad de soportar los requerimientos del negocio en el ámbito analítico y dar continuidad a las operaciones actuales, disminuyendo las brechas que hoy limitan el accionar analítico.
4. Diseñar una arquitectura que brinde mayores capacidades cognitivas/predictivas al interior de las diferentes unidades de negocio de manera transversal.
5. Diseñar una estrategia de datos que permita a los analistas descubrir y reutilizar los activos de datos, así como también que los comprendan e interpreten en el contexto de los requisitos comerciales.
6. Permitir la definición y visualización del ciclo de vida de los datos, describiendo lo que les sucede a los datos a medida que pasan por diversos procesos.
7. Diseñar una arquitectura que garantice la seguridad, integridad y confiabilidad general de los datos que la organización recopila, almacena, crea, recibe o transmite.
8. Definir una estrategia de gobierno del dato de forma transversal para la compañía abarcado el ecosistema completo.
9. Generar un centro de excelencia analítica interno que pueda tomar control y empoderarse de los desarrollos alrededor de los datos.
10. Concientizar y evangelizar de la importancia de la analítica predictiva en los procesos e indicadores de la compañía.
11. Conocer los retos de negocio de la organización que pueden ser abordados con analítica de datos.
12. Permitir la priorización de los casos de uso y con esto poder determinar un roadmap analítico a seguir.



Desarrollo de la diseño de solución

Como resultado del trabajo se obtiene un documento que abarca la propuesta de la arquitectura nueva, reglas y procesos de integración de DWH y el Data Lake, roles y responsabilidades para la implementación y mantenimiento de la solución, una propuesta de configuraciones y optimizaciones posibles para el Data Lake y una visión única de los datos. Como también una construcción de un roadmap analítico basado en los descubrimientos de negocio y disponibilidad de fuentes de datos.

Este documento está construido sobre cuatro grandes pilares:

1. Mejoras en las capacidades Analíticas: La nueva arquitectura EDMS debe permitir tener la cobertura necesaria funcional / técnica.
2. Aprovechamiento de entornos: La nueva arquitectura está pensada en potenciar las capacidades de los entornos Cloud, lo cual facilitará un ecosistema híbrido (on premise + Cloud) junto con facilitar a las unidades operativas la organización acceso a funcionalidades cognitivas para hacer más eficientes sus labores.
3. Formación de un equipo con Habilidades y Capacidades AI: Sugerencias para la formación de un equipo interno el cual posea las competencias y conocimientos para implementar modelos AI.
4. Servicio Cognitivos Predictivos: La nueva arquitectura debe permitir la exploración, implementación, pruebas y puesta en producción de manera recurrente modelos cognitivos para toda la organización (MLOps).

La diseño de solución se basa en estas cuatro fases principales:

1. **Research y Entrevista de Negocio:**

Esta fase inicial consiste en una serie de actividades para obtener una visibilidad completa de la empresa desde perspectivas internas y externas. Se llevarán a cabo entrevistas de negocio para recopilar las estrategias de diversas áreas y comprender los retos y oportunidades existentes. Se realizará una revisión exhaustiva de la arquitectura actual, denominada "radiografía de la empresa", que incluye la evaluación de sistemas operativos, herramientas de almacenamiento (bases de datos, filesystem, etc.), tipos de infraestructura utilizada (bare metal, virtual, contenedores, cloud) y relaciones entre componentes (ingestas ETL/ELT, APIs, entre otros). El resultado será un blueprint de arquitectura que mapeará todos los componentes identificados.

2. **Revisión de Herramientas Disponibles y Exploración de Datos:**

Con el panorama empresarial completamente entendido, se procederá a revisar las herramientas disponibles. Esta etapa busca profundizar en los procesos asociados a cada herramienta, identificando puntos de mejora y deficiencias, y determinando la necesidad



de integrar nuevos componentes. Se generará un listado de hallazgos y recomendaciones para cada herramienta y sus procesos asociados. Además, se incluirá la exploración de datos para conocer la información disponible, sus características (acceso, limpieza, estado, etc.), y realizar un ejercicio de data mining para evaluar la factibilidad de las oportunidades detectadas. Este estudio proporcionará una visión general del funcionamiento de los proyectos de analítica.

3. **Ejecución del Análisis de Factibilidad y Análisis de Situación Actual vs Mejores Prácticas:**
Esta etapa incluye la realización del análisis de factibilidad, un ejercicio de un día en el que se presentarán todos los resultados, análisis y estudios. Se creará una matriz de desarrollo de proyectos, mapeando y priorizando las oportunidades encontradas en las fases anteriores. Paralelamente, los consultores de Martinexsa realizarán un análisis interno de la situación actual (arquitectura y herramientas) en contraste con las mejores prácticas a nivel global. Se identificarán brechas y se propondrá una arquitectura optimizada, considerando hardware, software, seguridad y gobierno. Se generará un nuevo blueprint de arquitectura, comparándolo con el inicial para destacar componentes que requieren atención o modificación, habilitación de módulos, y componentes faltantes para soportar capacidades de Business Intelligence, Analítica Avanzada y Gobierno del Dato.
4. **Generación de Documento Único y Entrega del Roadmap:**
Finalmente, se consolidarán los descubrimientos del diseño de solución en un documento único, que incluirá el roadmap, hallazgos y recomendaciones de negocio y tecnología. Se entregará una hoja de ruta enfocada en tecnología, procesos y personas, detallando los pasos a seguir en el corto, mediano y largo plazo (hasta 5 años) para adoptar una cultura analítica integral.

Documentación

- Resumen general de hallazgos
- Roadmap integral de iniciativas con dimensiones de:
 - Datos y tecnología
 - Oportunidades de analítica avanzada
 - Arquitectura tecnológica
 - Aprovechamiento y monetización de los datos
 - Habilitación de capacidades predictivas (casos de uso identificados)
 - Habilitación de un marco para Gobierno del Dato
 - Seguridad de los datos
 - Gestión del cambio y cultura de datos en la organización

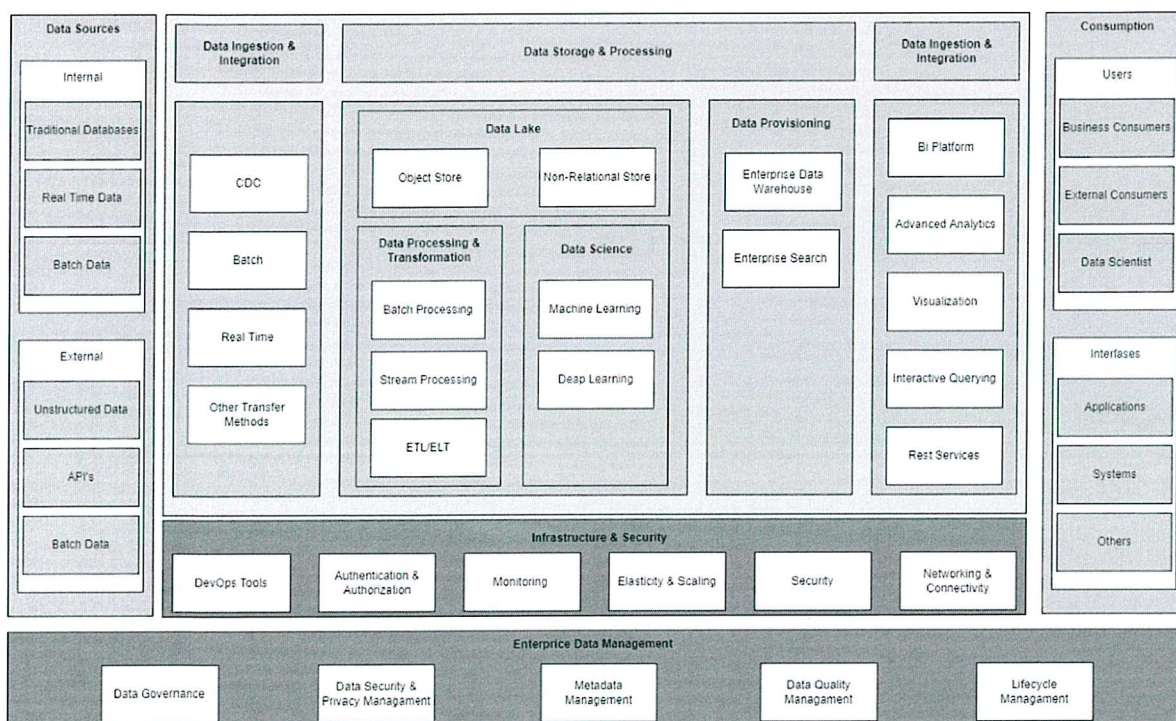




Se realizarán reuniones de entrevista con personal clave del negocio incluyendo un máximo de 3 áreas con un máximo de una entrevista por área.

A continuación, se encuentra una descripción detallada indicando todos los puntos a evaluar para la construcción de una arquitectura optimizada.

La plataforma de Enterprise Data Management Solution (EDMS) debe contar con determinadas características para satisfacer las preocupaciones, requisitos y objetivos comerciales que buscan las organizaciones hoy en día. La arquitectura completa de la solución debe contemplar una serie de capas que ofrecen las capacidades requeridas para soportar los procesos de análisis y reportería de datos de una empresa de telecomunicaciones. En la siguiente imagen se puede apreciar una arquitectura referencial de ejemplo:



Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica

Data Ingestion & Integration

La capa de ingestión e integración tiene la función de conectar, ingerir, transmitir y trasladar todas las fuentes de datos hacia el ecosistema analítico. En función de la necesidad de negocio se determina cuales métodos o patrones de ingestión se adaptan en base a la velocidad, frecuencia y volumen de datos. Esta capa generalmente considera las siguientes capacidades:

- CDC: Capacidad de identificar y capturar solo los datos de producción más recientes y los cambios de metadatos que la fuente ha registrado durante un período de tiempo determinado, generalmente medido en segundos o minutos.
- Batch: Capacidad de agrupar datos en un conjunto y enviarlos periódicamente en lugar de hacerlo de manera individual.
- Real Time: Capacidad que permite la ingestión de datos de origen a medida que se producen o en un período de tiempo muy pequeño después de su producción. En esta categoría se consideran la ingesta del flujo constante y continuo de cualquier formato de datos que se origina en múltiples tipos de fuentes (streaming)
- Other Transfer Methods: Conjunto de tecnologías que puede copiar datos a / desde / dentro de plataformas, que no están considerados en las capacidades antes descritas.

Data Storage & Processing

La capa de almacenamiento y procesamiento es la capa que proporciona datos brutos y preparados para un entorno de análisis robusto e integral. Al mismo tiempo proporciona un potente motor de procesamiento distribuido para el procesamiento de datos. Se compone de cuatro componentes principales:

- Data Lake: es un clúster computacional que aprovecha las capacidades informáticas distribuidas para almacenar y procesar grandes cantidades de datos.
- Data Processing & Transformation: Brinda la capacidad de transformar e integrar datos para admitir varios casos de uso en tiempo real, casi en tiempo real y por lotes. También transforma y limpia los datos para permitir una integración adecuada y estándares de calidad de datos aceptables.
- Data Science: Unifica estadísticas, análisis de datos, aprendizaje automático y sus métodos relacionados con el fin de comprender y analizar fenómenos reales con datos. La ciencia de datos emplea técnicas extraídas del campo de las matemáticas, las estadísticas, las ciencias de la información y la informática.
- Data Provisioning: Es el conjunto de herramientas y técnicas utilizadas para proporcionar datos para el consumo de las aplicaciones, o usuarios, en un formato que puedan consumir





fácilmente. Esta capa contiene vistas y análisis refinados de datos estructurados y no estructurados, en forma de repositorios de datos, que pueden ser buscados por usuarios comerciales y técnicos.

Information Delivery

Esta capa proporciona acceso y los medios para construir, visualizar e interactuar con la información disponible en la arquitectura de datos. Un conjunto de herramientas de BI, así como reportes, dashboards y visualizaciones, junto con herramientas analíticas estándar y avanzadas están disponibles para los consumidores de datos. Las aplicaciones empresariales o de negocios también son herramientas que permiten el acceso de información por medio de APIs.

Infrastructure & Security

Esta capa asigna los componentes de la arquitectura de referencia al hardware y software físicos durante la fase de implementación del proyecto. Las decisiones sobre productos e infraestructura se toman en función de los requisitos del cliente y las capacidades del producto. Esta capa también debe satisfacer varios requisitos de seguridad de datos para garantizar la adherencia y el cumplimiento, proveyendo capacidades de seguridad para Autorización y Autenticación, Seguridad de Red y mapeo basado en roles para el nivel de aplicación.

- **Authentication & Authorization:** Capacidad de realizar autenticación utilizando el directorio activo para autenticar usuarios, cuentas de servicio y hosts para acceder a la infraestructura y se realiza para remediar el acceso no autorizado a la plataforma de datos y análisis. El control de acceso basado en roles proporciona acceso granular a los componentes de la plataforma.
- **Monitoring:** Corresponde a la capacidad de proporcionar un marco de trabajo / herramientas de operaciones integradas para monitorear y administrar plataformas analíticas y de datos modernas.
- **Elasticity & Scaling:** Capacidad para aumentar / disminuir las características de cómputo y/o almacenamiento de manera automática basado en las cargas de trabajo. Esta es una característica de la infraestructura de la nube.
- **Security:** Corresponde a las herramientas utilizadas para complementar una plataforma de datos y análisis en términos de seguridad. También considera los aspectos de autorización basada en la propiedad de los datos, proporcionando a los administradores acceso granular y grueso.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>





- **Network & Connectivity:** Considera la utilización de redes de alta velocidad para recopilar y procesar datos en la plataforma de datos y análisis. Las opciones incluyen ancho de banda, resistencia y otras características de QoS.

Enterprise Data Management

Esta capa se refiere al desarrollo y la ejecución de arquitecturas, políticas, prácticas y procedimientos que gestionan adecuadamente las necesidades del ciclo de vida completo de los datos de una empresa. El framework de EDM consta de cinco capacidades.

- **Data Governance:** Conjunto de tecnologías habilitadoras para garantizar la propiedad de los datos de la empresa y que estén administrados de manera precisa y consistente.
- **Data Security & Privacy Management:** Esta capacidad controla y supervisa el acceso a las aplicaciones y los datos contenidos en la arquitectura. Proporciona protección de ciertos datos en función de políticas, redacción / cifrado de datos, tokenización, autenticación y capacidades de autorización.
- **Metadata Management:** Corresponde a la capacidad que habilita políticas y proporciona acceso a los datos. El sistema debe mantener metadatos de negocios, técnicos y operativos integrales.
- **Data Quality Management:** Capacidad que combina procesos y conjuntos de herramientas que proporcionan un enfoque estructurado para resolver errores e inconsistencias en los datos.
- **Lifecycle Management:** Corresponde a la capacidad que gestiona el flujo de información y los datos asociados desde la creación y el almacenamiento.

Martinexsa LATAM | Data Team

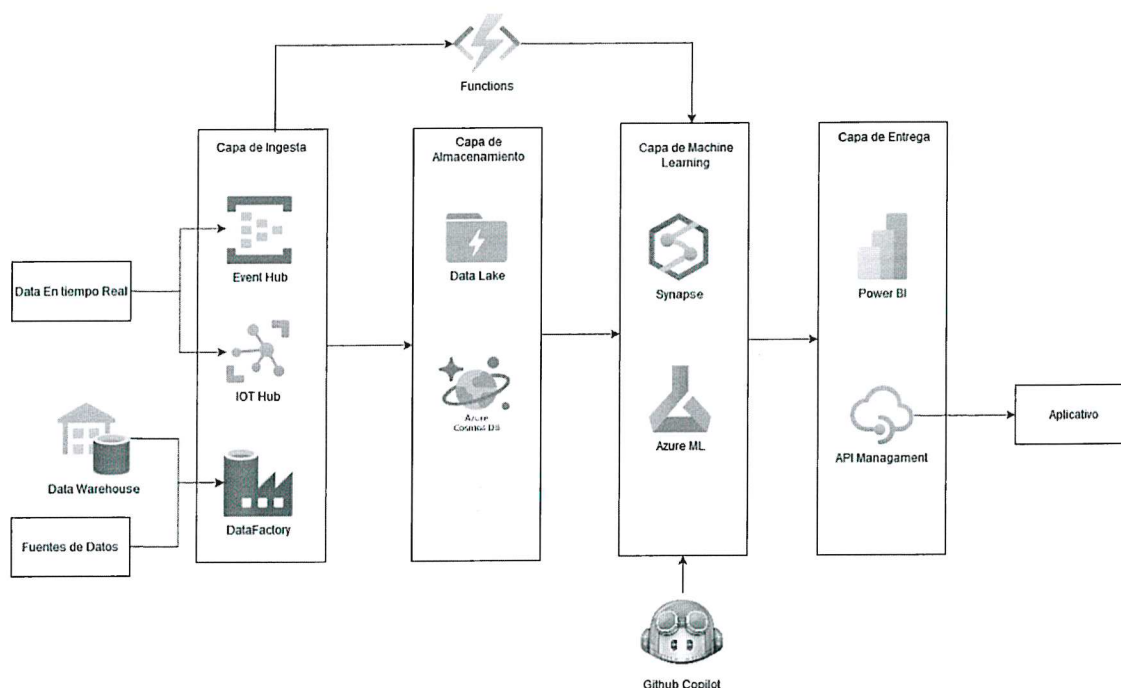
bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Data Lake: Microsoft Azure Deploy & Configuration

Diseño conceptual



Como parte de la arquitectura que propone para este proyecto se consideran diferentes componentes dentro del *Stack* tecnológico de **Microsoft Azure**, la cual se compone de la siguiente manera:

- **Event Hub:** Para la ingesta en tiempo real de eventos. Este es un servicio de *streaming* de eventos y que permitirá recibir datos continuamente desde múltiples fuentes.
- **Data Factory:** Para el manejo de flujos de información entre las diferentes fuentes de información
- **Azure IoT Hub:** Es el centro de control para dispositivos IoT. Permite conectar, monitorear y controlar una gran cantidad de dispositivos de forma segura.
- **DataLake:** para el almacenamiento de información
- **Azure Cosmos DB:** Una base de datos súper flexible y escalable, ideal para almacenar grandes cantidades de datos de diversas fuentes, como aplicaciones móviles, juegos o IoT.
- **Synapse:** Para la utilización de Apache Spark, lenguajes tanto Python como R, generación de diferentes casos de uso de Machine Learning



- **Azure Machine Learning:** Este es el servicio para crear y desplegar modelos de inteligencia artificial. Esta herramienta utiliza modelos predefinidos y low code.
- **Functions:** Servicio que se activará como un *trigger* cada vez que llegue un evento a Event Hubs. Esta función ejecutará el script de Python que procesará los datos en tiempo real.
- **API Management:** para exponer los resultados de los diferentes casos de uso como un *endpoint* de API. Se puede configurar límites de acceso y manejar el tráfico de las consultas.
- **Power BI:** La herramienta de visualización y análisis de datos.
- **GitHub Copilot:** Es un copiloto en la programación. Este código mientras se escribe, pudiendo simplificar el trabajo de generación de nuevos modelos.

Esta arquitectura de datos está diseñada para procesar datos en **tiempo real**, así como para conectarse a las bases de datos actuales en premisas (Data Warehouse actual), ingestar data de forma automatizada, y generar los diferentes casos de uso que TMPSA tiene contemplados actualmente. De la misma manera, esta arquitectura permitirá utilizar aplicativos como **Power BI** utilizando la información del Data Lake y de los diferentes modelos analíticos que se tendrán del Roadmap Analítico.

Dentro de los beneficios de la nube y la arquitectura propuesta se encuentra:

- Escalamiento tanto vertical como horizontal, tanto en storage como procesamiento para datos estructurados y no estructurados.
- Simplificación de procesos de ETL y ELT, tanto para batch como en tiempo real.
- Procesamiento de soluciones de ciencia de datos y Machine Learning
- Administración detallada de uso de cada servicio.

Dentro del servicio de *Deploy & Implementation* se realizará la configuración de los diferentes servicios de manera de que queden listos para que, una vez se tenga los casos de uso, se puedan subir recursos de los mismos y poder ejecutar sobre esta arquitectura y minimizar el downtime entre proyectos. De igual forma se desarrollará una (1) conexión a una base de datos On Premise.





A continuación se presenta un escenario de uso de créditos mínimo para la implementación de la arquitectura y un caso de ciencia de datos.

Créditos mensuales

Nombre de servicio	Horas al mes o métrica	Créditos Mensuales
Event Hub	730 horas/mes	11.09
IOT Hub	Dispositivos ilimitados, 400,000 mensajes/día	10.00
Data Factory	730 horas/mes	786.96
Data Lake Storage	10TB Storage Hot	209.58
Cosmos DB	1000 RU/s por 730 horas/mes	132.60
Synapse Analytics	730 horas/mes	1,297.65
Machine Learning	730 horas/mes	167.17
Functions	4,000,000 ejecuciones/mes	0.60
Api Management	730 horas/mes	48.03
Power BI	1 nodo por 360 horas/mes	362.92
Github Copilot*	5 licencias	50.00
TOTAL		3,076.60

*Github Copilot tiene opción de comprar licenciamiento por 1 año con un descuento añadido.

Para esta Fase 0 se pretende aprovisionar los créditos que nos servirán para la configuración de la arquitectura y desarrollos futuros, para después complementar en las fases posteriores con los créditos que hagan falta para cumplir con un año completo de desarrollos y modelos en producción.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>

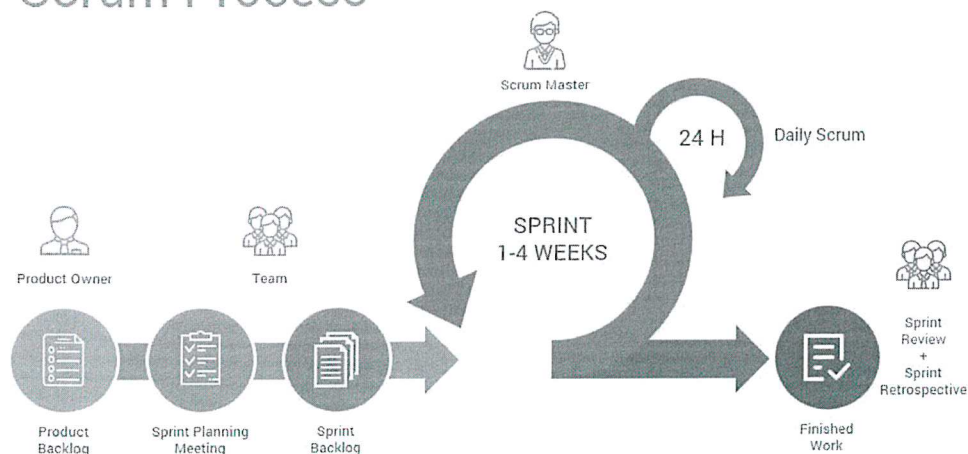


Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica

Consideraciones del proyecto

Metodologías ágiles

Scrum Process



Scrum es un marco colaborativo de trabajo entre equipos que se basa en un conjunto de bien definidas prácticas y roles que deben ser involucradas durante el proceso de desarrollo. Es una metodología flexible que se basa en los 12 principios ágiles y se maneja en un contexto donde miembros del equipo están enterados y en constante comunicación.'

Scrum se ejecuta temporalmente en pequeños y constantes periodos llamados sprints los cuales tienen una duración de 1 a 4 semanas, estos al terminan mejoran el valor percibido por el cliente. Se ejecutan Daily Scrums para que aumente la visibilidad de que hace cada miembro del equipo. Mediante esta metodología se quiere dar entregas continuas para dar un valor añadido al desarrollo que se esté generando.



Recursos

- **Delivery Manager**
 - Más de 5 años trabajando en proyectos de analítica de datos y ciencia de datos, con proyectos exitosos en toda Latinoamérica. Responsable de conducir la entrega del proyecto desde la parte técnica.
- **Project Manager**
 - Especialista en metodologías ágiles con maestría en administración de proyectos. Responsable de la gestión del proyecto y punto de contacto para el cliente.
 - Funciones principales:
 - Dar guía para realizar la implementación e integración de las metodologías ágiles.
 - Acompañar a los equipos en el uso de las herramientas ágiles y software de soporte.
 - Fomentar y motivar el compromiso de los involucrados y líderes para adoptar la Cultura Agile.
 - Planificar la implantación de Scrum.
 - Ayudar a entender que interacciones con el equipo aportan valor y cuáles no.
 - Ayudar al equipo a crear productos de valor.
 - Eliminar cualquier impedimento del equipo.
 - Facilitar los eventos Scrum.
 - Junto con el equipo Scrum, actualizar el burn down chart.
- **Líder de ingeniería de datos/ Consultor principal**
 - Más de 5 años trabajando en proyectos de ingeniería de datos, desarrollando integraciones tecnológicas para las compañías más importantes de Guatemala y Centroamérica. Conocimiento de amplia variedad de tecnologías de datos. Responsable de coordinar el desarrollo del proyecto desde la perspectiva tecnológica.
- **Ingeniero de datos/Científico de datos**
 - Recursos del equipo asignados a cada tarea que tienen experiencia en el desarrollo de proyectos de datos. Se involucran en función de las fases del proyecto.
 - Funciones principales:
 - Manejo y ejecución de procesos Machine Learning
 - Desarrollar patrones, métodos o algoritmos predictivos o de comportamiento para el análisis de información.
 - Desarrollar modelos de aprendizaje automático.



- Desarrollo de reportes y dashboards.
 - Proporcionar coaching para promover las mejores prácticas para perfeccionar la infraestructura de aprendizaje automático.
 - Acompañamiento y coaching en la construcción, prueba y mantenimiento de arquitecturas de datos.
 - Desarrollo de algoritmos, esquemas de datos y soluciones.
 - Experto en lenguajes de programación como Python y R.
 - Identificar maneras de mejorar la confiabilidad, eficiencia y calidad de los datos.
 - Preparar datos para modelos predictivos y prescriptivos.
 - Desarrollar modelos para descubrir tareas que puedan automatizarse.
 - Manipular grandes conjuntos de datos para abordar problemas del negocio.
 - Buscar nuevas fuentes de información para incorporarla en las plataformas de datos utilizadas por la empresa.
- Arquitecto de datos
 - Especialista en arquitecturas de datos. Responsable de hacer el dimensionamiento y arquitectura de las soluciones planteadas por el equipo.
 - Funciones Principales:
 - Participar en el entendimiento de las necesidades, con el fin de estructurar la mejor solución técnica para viabilizar y satisfacer los objetivos del cliente.
 - Participar activamente en la definición y el diseño de las diferentes soluciones que se requieran, identificando la necesidad real de los usuarios.
 - Dar lineamientos a nivel de arquitectura para el desarrollo de las soluciones para entornos On Premise, Cloud e híbridos, entre otras.
 - Participar en la toma de decisiones adecuadas para lograr una arquitectura que garantice un mejor desempeño, flexibilidad, mantenibilidad, robustez y reusabilidad de las soluciones.



Cronograma Roadmap Analítico

Mes	1				2			
Tarea	1	2	3	4	1	2	3	4
Entrevistas con áreas de negocio y tecnología								
Revisión exhaustiva de herramientas								
Definición de ecosistema actual de tecnología y datos								
Definición de oportunidades de negocio								
Definición de la arquitectura recomendada								
Recomendaciones de analítica avanzada								
Documentación de actividades y generación de road map								

*Los cronogramas presentados pueden estar sujetos a cambios, los finales serán entregados luego de una sesión inicial de proyecto.

**El cronograma también depende del tiempo necesario para llevar a cabo las entrevistas con las personas clave de las áreas de negocio indicadas, su disponibilidad y la cantidad de roles a entrevistar.

Cronograma Deploy & Implementation Data Lake Azure

Semana	1					2				
Actividades	1	2	3	4	5	1	2	3	4	5
Levantamiento de Prerequisitos										
Levantamiento de Arquitectura y configuraciones										
Transferencia de conocimiento										

*Los cronogramas presentados pueden estar sujetos a cambios, los finales serán entregados luego de una sesión inicial de proyecto.

**Este cronograma es asumiendo una arquitectura mínima, se requerirán trabajos para aumentar capacidad de las configuraciones por cada proyecto.



Entregables

Como resultado de la implementación se pueden esperar los siguientes puntos:

Roadmap Analítico

- Diseño de solución para la creación de una estrategia de implementación de nuevas tecnologías, arquitecturas nuevas y procesos de integración.
 - Resumen general de los hallazgos.
 - Documento que resume todo el trabajo realizado
 - Estado actual (Resumen, hallazgos y recomendaciones acerca de lo encontrado)
 - Proyecto e iniciativas
 - Retos actuales
 - Proyectos para solución de retos actuales
 - Priorización de proyectos
 - Desarrollo de capacidades analíticas
 - Dentro de este rubro se hacen recomendaciones acerca de las capacidades necesarias de perfiles de ciencia de datos.
 - Asimismo, se contempla el incluir diferentes modalidades de contratación (contratación directa, outsourcing, horas, etc.).
 - Arquitectura propuesta
 - Blueprint de arquitectura
 - Arquitectura alternativa
 - Conclusiones de los escenarios
 - Beneficios
 - Hoja de ruta para ejecución de arquitectura propuesta
 - Mapa de ruta y estrategia de implementación
 - Planes de inversión y recursos necesarios
 - Gobierno y Seguridad del dato
 - Estado actual de gobierno del dato
 - Data Ownership
 - Estrategia del gobierno del dato
 - Medidas de seguridad y cumplimiento
 - Cultura de datos
 - Cultura actual
 - Hallazgos
 - Recomendaciones

Data Lake Azure

- Implementación de Arquitectura Microsoft Azure
 - Incluye la conexión a una (1) fuente de datos On Premise
- Créditos para arquitectura mínima (4 meses).
- Transferencia de conocimiento e insights sobre el proyecto.





Propuesta económica

Propuesta Económica servicios

Contenido	Precio
Licenciamiento Azure Servicios de aprovisionamiento de 12,307 créditos de Microsoft Azure, estimados para el desarrollo propuesto y 4 meses de modelo analítico en producción. Incluye despliegue e implementación de Arquitectura Azure según descrito en el alcance, identificación de retos más importantes del negocio y diseño del roadmap analítico	USD \$49,307.00

*Precio no incluye IVA, ITBMS ni ningún impuesto asociado.

Prerrequisitos y consideraciones generales

- Acceso a los datos que fueran a ser necesarios para el proyecto.
- Consideración del trabajo 100% remoto
- Se debe tener un punto de contacto para consulta de dudas y presentación de resultados.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Servicios fuera de alcance

- Cualquier servicio, tarea o actividad distinto de lo mencionado específicamente en este documento.
- Configuración o instalación de componentes de Hardware o Software no mencionados en este documento.
- Servicios administrados de infraestructura o plataforma.
- Modificaciones del código fuente de la plataforma.
- Horas para soporte de solución de problemas.
- Resolución de problemas con Hardware, redes, sistemas operativos u otro tipo de infraestructura.
- Gobierno de la información y seguridad.
- Integración con herramientas y sistemas terceros no especificados en el alcance.
- No se encuentra dentro de alcance ningún requerimiento que se encuentre fuera del documento presente. Cualquier requerimiento incluido en otro documento deberá ser previamente evaluado por Martinexsa.

Lo expuesto en este documento sirve como referencia del entregable final y desarrollo a realizar en el proyecto. El detalle específico de ejecución, así como también el alcance final del proyecto, debe ser definido en conjunto con el equipo técnico y la versión final debe ser aceptada por el cliente. Cualquier requerimiento adicional a dicha aceptación, se informará del tiempo y costos que conlleva el cambio.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Términos y condiciones

La presente propuesta incluye únicamente lo definido en cada una de las secciones descritas en el documento y cualquier requisito que no se encuentre explícitamente listado queda fuera del alcance de la misma. La propuesta está sujeta a los términos y condiciones descritos a continuación. (1) Los precios se encuentran expresados en USD (dólares de Estados Unidos) a menos que el cuadro de propuesta económica indique otra moneda, como quetzales. (2) Si se requiere aplicar tipos de cambio de dólares hacia alguna moneda local será Martinexsa quien defina la tasa, en el caso de quetzales será al menos de 8 a 1. (3) Los precios no incluyen IVA, Impuesto de Venta ni ningún otro impuesto asociado a la legislación del país, a menos que el cuadro de propuesta económica indique lo contrario. (4) Los precios y descuentos otorgados son exclusivos de la presente propuesta y tienen efecto únicamente en este documento. No se pueden utilizar ni tomar de referencia para futuras negociaciones. (5) El precio final de la propuesta puede variar dependiendo de cambios en el alcance y/o cambios en los requerimientos. (6) No se pueden separar los ítems de la propuesta económica, el precio de la misma viene definido por la contratación de todas las partes que la conforman. Si se requiere contratar ítems separados se debe solicitar una propuesta económica para cada uno. (7) Ningún entregable de Martinexsa puede ser reproducido o comercializado por el cliente, esto incluye código fuente de modelos analíticos. Tampoco pueden ser compartidos con personas externas a la organización del cliente. (8) Martinexsa se reserva el derecho de solicitar al cliente cartas de referencia de los proyectos trabajados una vez hayan sido finalizados. (9) Todas las condiciones de pago se encuentran definidas en los siguientes incisos, a menos que se haya pactado un plan de pagos con el cliente y se vea reflejado en la propuesta económica. (10) La vigencia de esta propuesta es 15 días hábiles posterior a la fecha de emisión de la misma, fecha para la cual deberá contarse con la orden de compra, o contrato firmado en su defecto. En caso de expirada, se debe solicitar la propuesta nuevamente y esto puede impactar en los precios y descuentos reflejados.

Para la contratación de licencia y soporte de plataformas de datos Cloudera (CDP, CDH, CDF, CDSW), los términos de soporte son directos con el fabricante y bajo las condiciones que estipula en su sitio web. La modalidad y versión a contratar están descritas en la sección de propuesta económica y no incluyen servicios administrados sobre las plataformas a menos que se encuentren cotizados ítems específicos para esto. En plazo de contratación es regularmente por un año con renovación anual y las condiciones de pago son 100% por anticipado a la generación de la licencia. Si la propuesta incluye varios años se deben pagar todos los años por anticipado a menos que la propuesta indique un plan de pagos distinto. La vigencia de la licencia empieza a correr desde la fecha indicada en la propuesta. Si se van a sincronizar distintos paquetes de suscripciones debe venir indicado en la propuesta económica un ítem de co-término que hace referencia a los meses adicionales para emparejar el licenciamiento.

Para los servicios de instalación, configuración e implementación de plataformas Cloudera (CDP, CDH, CDF, CDSW) se requiere que el cliente proporcione todos los requisitos previos a comenzar la instalación. Además, el cliente deberá proveer conexión de red entre todos los nodos del clúster, proveer conexión remota al equipo de Martinexsa y acceso a internet temporal desde el clúster para hacer la descarga de paquetes. Estos servicios son individuales para cada una de las plataformas y ambientes, por tanto, la propuesta económica debe tener detallado el alcance de la plataforma que será implementada. El stack de herramientas Cloudera de gobierno y seguridad son ítems aparte en la propuesta económica y no se incluyen por defecto. La implementación se da por finalizada 24 horas después de enviado el estado de salud de la instalación. Las condiciones de pago para estos servicios son 100% por anticipado, a menos que se indique un plan de pago distinto en la sección de condiciones de pago.

Para la contratación de proyectos de Analítica, Ciencia de Datos o Inteligencia Artificial; se define un calendario provisional para los tiempos de entrega que puede variar dependiendo de la definición final del proyecto. El costo de los proyectos son estimados a partir de las horas-hombre dedicadas al diseño, desarrollo, implementación y puesta en producción. El proyecto comienza con un kick-off oficial en conjunto con el cliente. Se solicita al cliente el acceso ilimitado a las fuentes de información de relevancia para el proyecto, guardando la confidencialidad respectiva. Se solicita también un punto de contacto de parte del cliente, dedicado al proyecto y a gestionar los requerimientos internos del cliente que Martinexsa solicite como consecuencia del trabajo a realizar. Martinexsa asignará un PM para que pueda gestionar el proyecto y la comunicación con el cliente. Cualquier retraso en las actividades asignadas al cliente implicarán un retraso en el proyecto y será notificado por el PM de Martinexsa. Las fuentes de datos a utilizar en el proyecto se limitan a las descritas en la propuesta, a menos que Martinexsa decida ampliar el alcance. Ampliar el alcance en cuanto a fuentes de información puede incurrir en costos adicionales al proyecto. Para estos proyectos Martinexsa no es responsable de ningún componente tecnológico de hardware, software o suscripciones vigentes necesarios para ejecutar el proyecto, entendiendo que los mismos son insumos provistos por el cliente y asumiendo que se encuentran configurados bajo las mejores prácticas. Así mismo, que los recursos tecnológicos sean los suficientes, según solicitud de Martinexsa, para la puesta en producción del proyecto. El cierre del proyecto se hará 24 horas después de la presentación final de resultados, una vez que el proyecto haya quedado implementado y puesto en producción de acuerdo a lo definido en la contratación. El soporte y mantenimiento del modelo, posterior a la implementación, queda del lado del cliente; a menos que sea solicitado un ítem de "soporte y mantenimiento anual" y este se encuentren incluido de forma individual dentro de la propuesta económica. Las condiciones de pago para estos proyectos son 100% por anticipado, a menos que se indique un plan de pago distinto en la sección de condiciones de pago.

Para la contratación de entrenamientos y capacitaciones. La modalidad que se cotiza por defecto es impartida por consultores certificados de Martinexsa. Esa modalidad es en español y puede ser presencial o remota, dependiendo de lo detallado en la propuesta económica. El contenido y los horarios son pactados en conjunto con el cliente para un máximo de 10 personas, a menos que la propuesta económica contemple una cantidad distinta de participantes. Las propuestas de entrenamiento no incluyen ningún costo de T&L (pasaje, hospedaje, movilización y viáticos) a menos que la propuesta económica incluya un ítem separado con esta información. La logística de ubicación, recursos tecnológicos, parqueo y demás corre por parte del cliente, a menos que la propuesta económica incluya un ítem específico para estos temas. Si se requiriera, se puede solicitar incluir en la cotización la sala de capacitaciones de Martinexsa, la cual se encuentra disponible bajo demanda con los recursos necesarios para impartir estos entrenamientos. Las capacitaciones impartidas por Martinexsa incluyen un diploma de participación a todos los asistentes al final del curso y con esto se da por finalizado el servicio. Para los cursos impartidos directamente por Cloudera; las condiciones, horarios y contenido están definidos en la página del fabricante y vienen descritos en la sección de propuesta económica en caso haya sido requerida su cotización. En este caso toda la logística corre por parte del cliente y la gestión del instructor y fechas la hace el cliente directamente con Cloudera. Ni los entrenamientos de Cloudera ni los de

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Martinexsa incluyen ningún voucher para tomar Evaluación de certificación, a menos que se incluya una línea específica para esto en un ítem de la propuesta económica. Las condiciones de pago de todos los entrenamientos son 100% por anticipado, a menos que se indique un plan de pago distinto en la sección de condiciones de pago.

En caso de contratar horas de servicios profesionales para desarrollo o soporte sobre plataformas Cloudera o proyectos de Ciencia de Datos, el cobro se hace por cantidad de horas basadas en un esquema escalonado. Las actividades que incluyen y las que no incluyen estas horas se encuentran definidas en el documento de propuesta. Los horarios de atención de estas horas son bajo un esquema 8x5 en horario laboral, a menos que la propuesta económica indique lo contrario. Al no ser horas críticas se deben solicitar con al menos 48 horas hábiles de anticipación para poder ser atendidas, a menos que la propuesta económica indique un escenario distinto. Si las horas contratadas son consumidas en su totalidad se debe contactar a su asesor para cotizar horas adicionales previo a atender nuevas solicitudes. Las horas contratadas vencen luego de un año desde la fecha de su contratación y ya no pueden ser utilizadas, a menos que la propuesta económica indique lo contrario. Las condiciones de pago para estas horas adicionales son 100% por anticipado, a menos que se indique un plan de pago distinto en la sección de condiciones de pago.

Para los proyectos que son en modalidad "Analytics as a service", la vigencia del contrato debe ser al menos de 1 año, a menos que se especifiquen tiempos menores en la propuesta económica. Esto implica 12 pagos mensuales del servicio. Las renovaciones también se hacen de forma anual, a menos que se pacte un esquema distinto de renovación. El servicio incluye la utilización a estilo "caja negra" de plataforma analítica en la nube y el costo de desarrollo, implementación, soporte y mantenimiento de modelos analíticos en dicha plataforma. La distribución del costo para cada ítem se hará según lo descrito en la propuesta económica. Si se contratan horas individuales en este esquema, el uso y distribución de las mismas es independiente a la modalidad de pago. Esto indica que las horas se pueden acabar en un plazo menor a la cantidad de meses pactados a pagar en el contrato. Las renovaciones anuales se hacen de forma obligatoria sobre la plataforma analítica en la nube. Los pagos deben comenzar desde el primer mes de desarrollo, a menos que se haga un plan de pagos distinto. Existe una limitación en cuanto a la cantidad de información que se puede alojar en este servicio, esta limitación estará definida en la propuesta y exceder este límite puede incurrir en costos adicionales. Una vez finalizado un contrato anual el cliente puede optar por bajar la información a una arquitectura propia, pero debe incurrir en los costos de movimiento de información. Ahora bien, el modelo es propiedad compartida del cliente y puede bajarlo sin ningún costo. Si requiriera implementarlo localmente se debe cotizar por aparte un ítem referente a migración de modelo analítico y este debe reflejarse en la sección de propuesta económica. Las condiciones de pago para este servicio son de forma mensual con la fecha exacta a pactar con el cliente. Cualquier modelo o desarrollo sobre la plataforma dejará de funcionar si el pago de la plataforma y soporte dejan de estar vigentes.

Para otros assessments puntuales, como Data Strategy o Big Data Discovery Workshop, todas las secciones, contenido y condiciones vienen descritas en la propuesta. Hay actividades que se hacen de forma presencial y otras de forma remota, la distribución y modalidad de las actividades se describe también en la propuesta. La consultoría se da por finalizada una vez entregado el documento con los resultados o realizada la presentación, si es que aplica. Las condiciones de pago para estas consultorías son 100% por anticipado, a menos que se indique un plan de pago distinto en la sección de condiciones de pago.

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica



Aceptación de la propuesta

Firma:

Sello de la empresa:

Nombre de la persona que autoriza:

Fecha:

Esperando que el presente documento sea de su entera satisfacción, quedo a su disposición para ampliar cualquier información,

Atentamente,

Francesco Petrone
Account manager
Martinexsa Panamá
fpetrone@martinexsa.com
+507 6528-0882

Martinexsa LATAM | Data Team

bigdata@martinexsa.com

<https://www.martinexsa.com/bigdata/>



Guatemala | México | El Salvador | Honduras | Costa Rica | Panamá | República Dominicana | Colombia | Ecuador | Chile | Perú | Argentina | Jamaica