Disponibilidad de la información registros de consumo para su monetización de forma segura de la compañía Ola Colombia

Beatriz Amparo Barajas Rodríguez, Elkin Guillermo Gómez Pérez, Jorge Enrique García Olarte

Especialización en Big Data, Fundación Universitaria Compensar

Consultoría fin de Grado

NELSON FELIPE ROSAS JIMENEZ

19 de mayo de 2023

Esta obra está bajo una Licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional.

Disponibilidad de la información registros de consumo para su monetización de forma segura de la compañía Ola Colombia

Beatriz Amparo Barajas Rodríguez, Elkin Guillermo Gómez Pérez, Jorge Enrique García Olarte

Trabajo de grado presentado como requisito parcial para optar al título de:

Especialista en Big Data

Director (a):

M.Sc., Nelson Felipe Rosas Jiménez

Línea de Investigación:

Área Temática

Fundación Universitaria Unipanamericana – Compensar

Facultad de Ingeniería, Especialización en Big Data

Bogotá, Colombia

2023

Resumen

Esta consultoría cuyo objetivo es desarrollar una propuesta técnica orientada al aprovechamiento y explotación de los datos, garantizando la monetización de la información para generar nuevos ingresos a una compañía de telefonía móvil, propone la implementación a mediano plazo de una arquitectura de referencia que se define con zonas y subzonas, y el diseño detallado de una solución tecnológica que se define con los bloques de construcción ABBs y SBBs. Así mismo analiza la relación entre una solución avanzada basada en Cloud Computing y una solución apoyada en el uso de servidores locales, así como sus beneficios y desventajas.

Palabras clave: datos, monetización de datos, disponibilidad de datos, telefonía móvil, bloques de construcción ABBs y SBBs, data lake

Abstract

This consultancy whose objective is to develop a technical proposal oriented to the use and exploitation of data guaranteeing the monetization of information to generate new income to a mobile phone company, proposes the implementation in the medium term of a reference architecture that is defined with zones and subzones, and the detailed design of the technological solution that is defined with the ABBs and SBBs building blocks. It also analyzes the relationship between an advanced solution based on Cloud Computing and a solution supported using local servers, as well as its benefits and disadvantages.

Keywords: data, data monetization, data availability, mobile telephony, ABBs and SBBs building blocks, data lake

Contenido

PÁG.

RESUMEN ..... **V**

INTRODUCCIÓN **11**

ANTECEDENTES Y JUSTIFICACIÓN **12**

NOTA. FUENTE PROPIA 12
FORMULACIÓN DEL PROBLEMA 13
OBJETIVO GENERAL 14
Objetivos Específicos *14*
ALCANCES Y LIMITACIONES 15

1. MARCO TEÓRICO **16**

1.1 FACTORES QUE INFLUYEN EN LA DISPOSICIÓN DE INFORMACIÓN 16
1.2 BIG DATA VALUE CHAIN 17
1.3 MONETIZACIÓN DE DATOS 17

2. DISEÑO METODOLÓGICO **18**

2.1 PROPUESTA TECNOLÓGICA BASADA EN SERVICIOS DE LA NUBE 20
2.2 PROPUESTA TECNOLÓGICA HIBRIDA 28

3. RESULTADOS **33**

3.1 PROPUESTA TECNOLÓGICA BASADA EN SERVICIOS DE LA NUBE 33
3.2 PROPUESTA TECNOLÓGICA HIBRIDA 33

4. DISCUSIÓN **35**

5. CONCLUSIONES Y RECOMENDACIONES **36**

5.1 CONCLUSIONES 36
5.2 RECOMENDACIONES 37

BIBLIOGRAFÍA **39**

Introducción

En la actualidad, las compañías de telefonía móvil junto a las redes sociales pasaron a convertirse en actores clave si en la recolección y almacenamiento de gran cantidad de datos se trata. diferentes datos provenientes del consumo de los usuarios permiten tener acceso a una gran cantidad de información valiosa, especialmente a través de los registros de comportamiento y patrones de uso de sus usuarios.

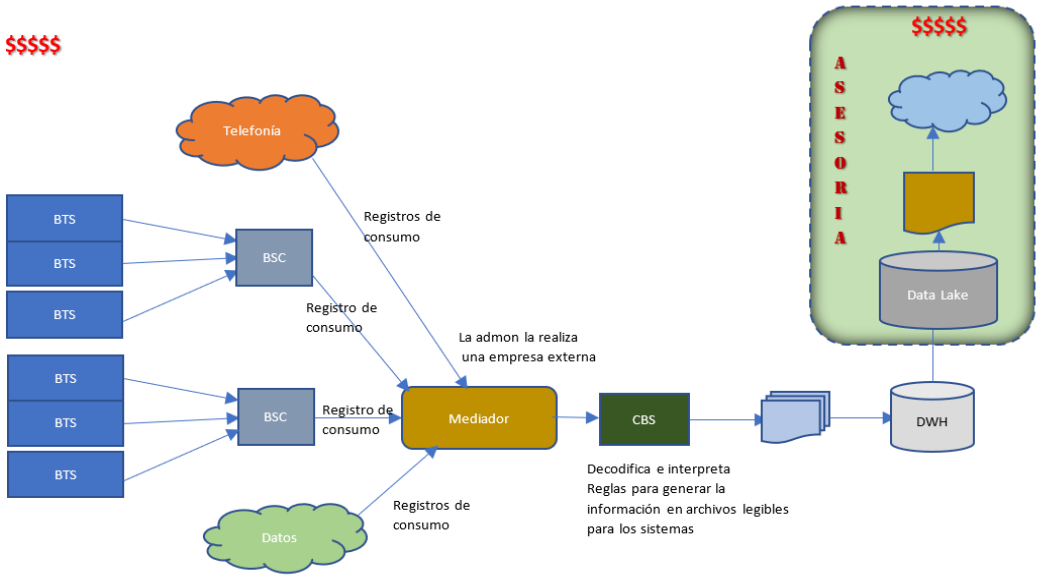
La disponibilidad de la información de los registros de consumo para su monetización de forma segura en las compañías de telefonía móvil es un tema crucial en el ámbito empresarial y tecnológico. Estas compañías generan y recopilan una gran cantidad de datos mediante la prestación de sus diferentes servicios, esto brinda la oportunidad de usar para consumo interno, así como monetizar esa información y utilizarla para mejorar su rentabilidad y eficiencia generando ingresos económicos adicionales a su core business. Sin embargo, es fundamental garantizar la aplicación de la ley de protección de datos y disponibilidad segura de estos registros de consumo para proteger la privacidad de los usuarios y evitar posibles riesgos. Según lo mencionado por (Palmer, Michael, 2006)“los datos son valiosos, pero si no están refinados, en realidad no se pueden usar. El Petróleo debe transformarse en gas, plástico, productos químicos, etc. para crear una entidad valiosa que impulse una actividad rentable; por lo tanto, los datos deben desglosarse y analizarse para que tengan valor”, la monetización de los registros de consumo implica transformar esos datos en valor económico, para hacer esto se basan en procesos de soluciones tecnológicas construidos y desarrollados a partir de la cadena de valor de la Big data en la que mediante una serie de etapas que van desde la minería o generación del dato pasando por la recolección, aplicando una analítica y visualización de los mismos permitan crear ventajas competitivas y futuros modelos de negocio.

Antecedentes y Justificación

Los eventos generados desde los dispositivos celulares de los clientes como los registros de consumo, son anotados y almacenados por medio de las BTS (Base Transceiver Station) las cuales se conectan a una BSC (Base Station Controller) que se encargan de controlar un determinado número de BTSs de un área geográfica específica y a través de esta BSC pasa todo el flujo de comunicaciones, también maneja la configuración de cada una de las BTSs e incluso participa activamente cuando un usuario móvil pasa de una BTS a otra (hand-over). La BSC se conecta al mediador que es una central de conmutación móvil, que es el encargado de tomar decisión sobre el enrutamiento de una llamada y de ejecutar la conmutación correspondiente basándose en la información de la base de datos. El mediador a su vez envía la información o mensajes que se encuentran codificados al sistema CBS el cual se encarga de tomar los datos y la información encriptada dentro del protocolo del estándar GSM, realizando un proceso de decodificación y aplicación de reglas para generar la capa de datos en archivos legibles para los sistemas informáticos implementados para la su cargan en la bodega de datos, DWH, de Oracle para ser dispuestos al interior de la compañía en los procesos de toma de decisiones descrito en la figura 1.

Figura 1

Trafico de informacion



Nota. Fuente propia

Formulación del Problema

Falencia en la capacidad del tiempo de respuesta del procesamiento y disposición de los datos capturados correspondientes al tráfico IP generado por la red vía celular de sus usuarios y limitación en la oportunidad de entrega de información a clientes internos y externos según demanda; lo anterior, de un manera detallada con la debida restricción de los campos únicamente requeridos y protegidos según la ley de protección de datos de los usuarios, para su posterior disponibilidad de forma segura en un repositorio previamente definido para sus aliados comerciales según su demanda, lo anterior con fines para estudios analíticos y autorizados por la regulación Colombiana, representado, a su vez, una nueva fuente de ingresos para la compañía al poder monetizar dichas salidas.

Objetivo General

Desarrollar una propuesta técnica que permita garantizar la monetización de la data de la compañía Ola Colombia generando nuevos ingresos y crecimiento empresarial que permita la creación de nuevos servicios, incursión en nuevos negocios y socios como parte de la diversidad y cubrimiento de la gran demanda de mercado existente.

Objetivos Específicos

Plantear un modelo de arquitectura tecnológica que permita alojar y disponer de los datos de acuerdo con los requerimientos funcionales y las condiciones económico-financieras de la compañía.

Presentar un análisis basado en los beneficios y desventajas entre un modelo hibrido de servidor local y un modelo basado en Cloud Computing dirigidos a solucionar la problemática estudiada.

Generar un presupuesto económico que abarque la construcción e implementación de las soluciones tecnológicas ofrecidas.

Alcances y Limitaciones

Con esta consultoría se busca presentar dos propuestas orientadas a la solución de la problemática descrita en la formulación del problema y que se presenta en la compañía Ola Colombia S.A. la cual cuenta con su información relacionada con el tráfico de datos de los usuarios de red móvil (VOZ, SMS y Datos) de su red, que por su volumetría generada a diario, representa una limitante para su ágil procesamiento y generación de resultados para la compañía; sus tiempos de respuesta no cumplen con la demanda ni necesidad del negocio ante el mercado competitivo que representa la competencia actual y las necesidades de sus clientes; por otra parte, incurrir en gastos para la compra de infraestructura como una alternativa inmediata de solución; por lo anterior se hace necesario la búsqueda de una solución que permita probar una infraestructura óptima para la extracción, carga, transformación, anonimizacion, agregación y análisis de datos en tiempos requeridos.

La solución además deberá garantizar, a su vez, la definición de un de esquema de colaboración permanente para la explotación de los datos de Ola Colombia S.A con sus futuros aliados interesados en dicha información cumpliendo con las políticas de protección de datos, su encriptación con una respuesta ágil a la demanda de los aliados, según sus necesidades, de forma automática para finalmente su envió y/o publicación para su consumo como respuestas a los acuerdos comerciales establecidos.

Se evidencian limitaciones técnicas o tecnológicas para la implementación de la propuesta debido a carencia en la arquitectura tecnológica que no esta acorde a los requerimientos, no se cuenta con un recurso de inversión alto y no se tiene una cultura organizacional en adopción de tecnologías actuales.

1. Marco teórico

Disponibilidad de información en una compañía de telecomunicaciones

La disponibilidad de información en todas las organizaciones es un aspecto fundamental pues se convierte en un activo fijo intangible de gran valor. Según el informe del ministerio de las telecomunicaciones (MinTic, 2022) en su boletín trimestral del sector TIC - Cifras cuarto trimestre de 2022, las compañías de telefonía móvil en Colombia al término del cuarto trimestre de 2020 reportaron que Colombia alcanzo 80,8 millones de líneas celulares registradas con un aumento de 5,7 millones más en comparación con las registradas en el trimestre del año inmediatamente anterior. Además, se encontró que existen 156,6 líneas por cada 100 habitantes, hubo un tráfico de telefonía móvil entre las redes de 31.300 millones de minutos. Así mismo como menciona la comisión de regulación de comunicaciones (CRC, 2023) en su informe data flash 2023-005 en Colombia durante el cuarto trimestre del 2022 se generó un tráfico de 910 millones de GB de datos y 40,1 millones de abonados a internet móvil. Lo descrito anteriormente muestra un aumento significativo en el uso y consumo de datos impulsado por la creciente demanda de servicios de comunicación que a su vez es apalancada por el auge en nuevas tecnologías, esto implica la generación constante de grandes volúmenes de datos de manera masiva, diversificada y en constante evolución que puede ser aprovechada y explotada por la compañía de diferentes maneras.

1.1 Factores que influyen en la disposición de información

El término “habeas data” según la ley (1581 de 2012) de protección de datos personales en Colombia hace referencia al derecho fundamental que tienen las personas a que se garantice el control sobre sus datos personales, así como la posibilidad de acceder, suprimir o rectificar su información personal cuando sea requerido, en las compañías de telecomunicaciones este concepto es un tema relevante a la hora de gestionar los datos de sus usuarios, se destaca la importancia del cumplimiento de las medidas y disposiciones legales que realizan estas compañías en materia de protección de datos.

1.2 Big data Value Chain

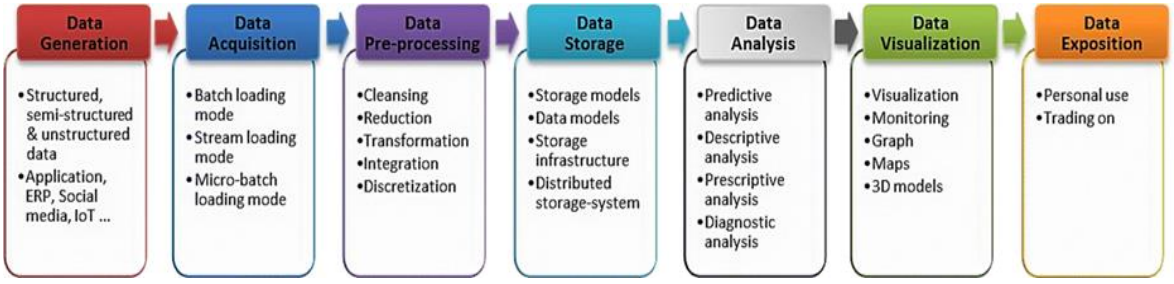
El término Big data (grandes datos) hace relación al almacenamiento de un gran volumen de información variada, así como su adecuada manipulación desde diferentes gestores de bases de datos. El Big data por sí solo no genera valor, recordemos que, si bien los datos tienen un valor intrínseco, por sí mismos no generan ninguna utilidad, por lo que se hace necesario que estos pasen por un proceso de agregación de valor en el cual un dato se convierte en información y este a su vez genera conocimiento y sabiduría. (Faroukhi, 2020) "La cadena de valor de datos es un modelo que ve los datos como materia prima y como un recurso importante en el negocio". En este orden de ideas, el proceso de agregación de valor desde el flujo de datos organizacional esta medido por la Big Data Value Chain que se basa en un conjunto de herramientas (características) para generar procesos claros de explotación de datos para consumo interno o monetizarlos para obtener ingresos de valor no operacionales en las organizaciones. Las herramientas del Big data están definidas por criterios definidos como las 7Vs: volumen, Velocidad, Variedad, Veracidad, Variabilidad, Visualización y Valor. Se entiende por volumen a la gran cantidad. Nos referimos a volumen cuando hablamos de grandes datos, por velocidad a la rapidez de los procesos, por variedad a las diversas fuentes y formatos de origen de los datos, por veracidad a la extracción de datos confiables, por variabilidad al cambio constante del significado de un dato, por visualización al proceso de hacer legible las grandes cantidades de datos y por valor al conocimiento que generan los datos.

1.3 Monetización de datos

Mencionan Faroukhi, (Faroukhi, 2020) "El concepto de monetización tiene como objetivo generar beneficios económicos medibles a partir de datos sin procesar y recursos extraídos." Podemos entonces afirmar que a través de activos intangibles podemos obtener beneficios tangibles siendo este un claro ejemplo de un modelo de negocio de valor económico que a su vez puede ser cuantificable, lo anterior procedente de la explotación y manejo de los datos personales de un cliente para que un tercero pueda procesarlos y usarlos de acuerdo con su necesidad. mediante el uso adecuado de la cadena de valor del Big data que se genera en 7 etapas que son recopiladas y descritas por Faroukhi (Faroukhi, 2020) como "Big Data Value Chain phases" y las podemos observar en la figura 2.

Figura 2.

Big Data Value Chain phases



Nota. extraido de "Big data monetization throughout Big Data Value Chain: a comprehensive review", por Faroukhi, A.Z., El Alaoui, I., Gahi, Y., 2020, J Big Data 7, 3 (2020).

2. Diseño Metodológico

Para la presente consultoría se establece un diseño metodológico divido en 2 propuestas tecnológicas enfocado en ejecución por medio de fases de desarrollo, ambas mencionadas a continuación.

La arquitectura de referencia propuesta define los componentes que harán parte del modelamiento arquitectónico ofrecido como solución tecnológica a la problemática estudiada a lo largo de este documento. Adicionalmente, facilita el desarrollo y el entendimiento de los artefactos de arquitectura y vistas para el diseño. Este modelo de referencia se encuentra alineado con las definiciones de mapa de zonas del Marco de Arquitectura Empresarial (MAE) de MinTIC[1]. La arquitectura de referencia se define con zonas y subzonas, y el diseño detallado de la solución tecnológica se define con los bloques de construcción ABBs y SBBs, estos conceptos se encuentran descritos en la tabla 1.

Tabla 1.

Tabla de jerarquía para el modelamiento arquitectónico

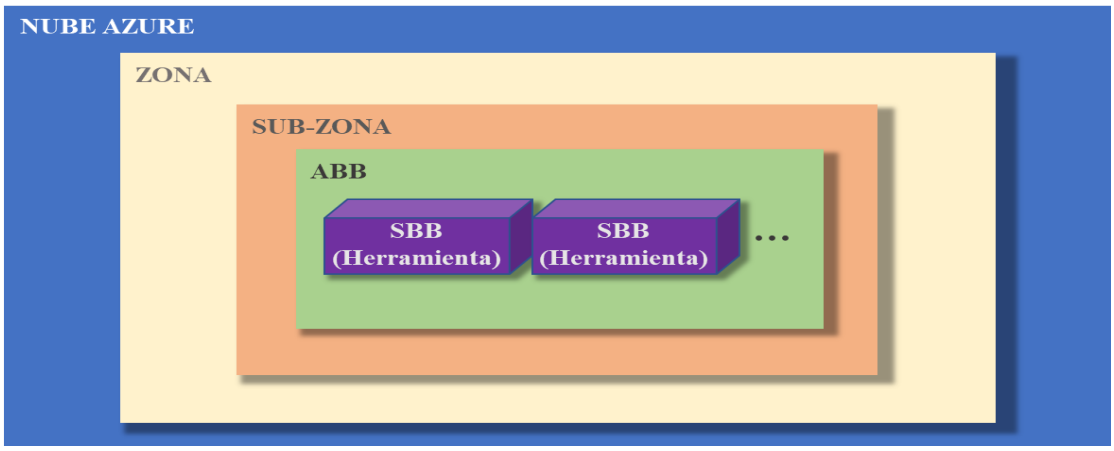
CONCEPTO	*DESCRIPCIÓN*
ZONAS	*Agrupación de alto nivel que contiene de manera agnóstica servicios de gestión de datos y herramientas asociadas a puntos específicos del ciclo de vida de los datos.*
SUB-ZONAS	*Agrupación de nivel intermedio que consolida de manera agnóstica las capacidades de las herramientas, que soportan los requerimientos y servicios de gestión específicos.*
ABBs	*Los bloques de construcción de arquitectura (ABB por sus siglas en inglés) son un componente agnóstico del modelo de la arquitectura estructural que presta y consume servicios de otros bloques y que representa una capacidad de la subzona para la prestación de un servicio tecnológico.*
SBBs	*Los bloques de construcción de solución (SBBs por sus siglas en inglés) son herramientas concretas cuyas funcionalidades cubren las capacidades de los ABBs y suplen los requerimientos de la solución.*

Nota. Fuente propia

Los conceptos anteriores se agrupan en una jerarquía que se puede ver en la figura 3.

Figura 3.

Arquitectura de bloques para el análisis de la propuesta



Nota: Fuente propia

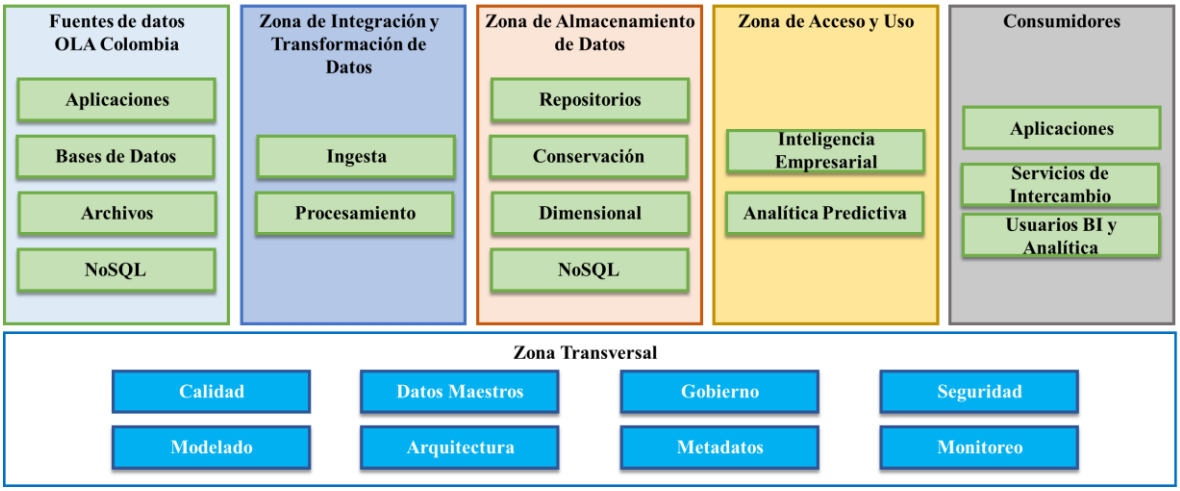
2.1 Propuesta Tecnológica Basada En Servicios De La Nube

Fase 1 – Modelamiento de referencia de la solución

Con base en la arquitectura de bloques anterior, en la siguiente ilustración se despliega el modelo de arquitectura de referencia que se ha diseñado para la propuesta como se visualiza en la figura 4.

Figura 4.

Arquitectura de bloques para el análisis de la propuesta



Nota. Fuente propia

Esta arquitectura de referencia se alinea con el estilo arquitectónico "DataFabric" brindando capacidades para gestionar el ciclo de vida de los datos desde sus fuentes, sus transformaciones, su uso, almacenamiento, apoyo a la gestión hasta los consumidores de datos. A continuación, se describen las zonas definidas:

Fuentes de Datos OLA Colombia: representa las fuentes de datos con las cuales la propuesta de solución en Azure se conectará.

Zona de Integración y Transformación de Datos: para la ingesta y procesamiento de los datos de diferentes fuentes para generar conjuntos de datos procesados para su tránsito, uso y/o almacenamiento.

Zona de almacenamiento de Datos: de gestión completa de repositorios de datos y archivos.

Zona de Acceso y Uso: brinda las estructuras para la explotación de datos y aprovisionamiento de productos de información para los consumidores.

Consumidores: todos aquellos sistemas, organizaciones y/o personas que consumen los productos de información.

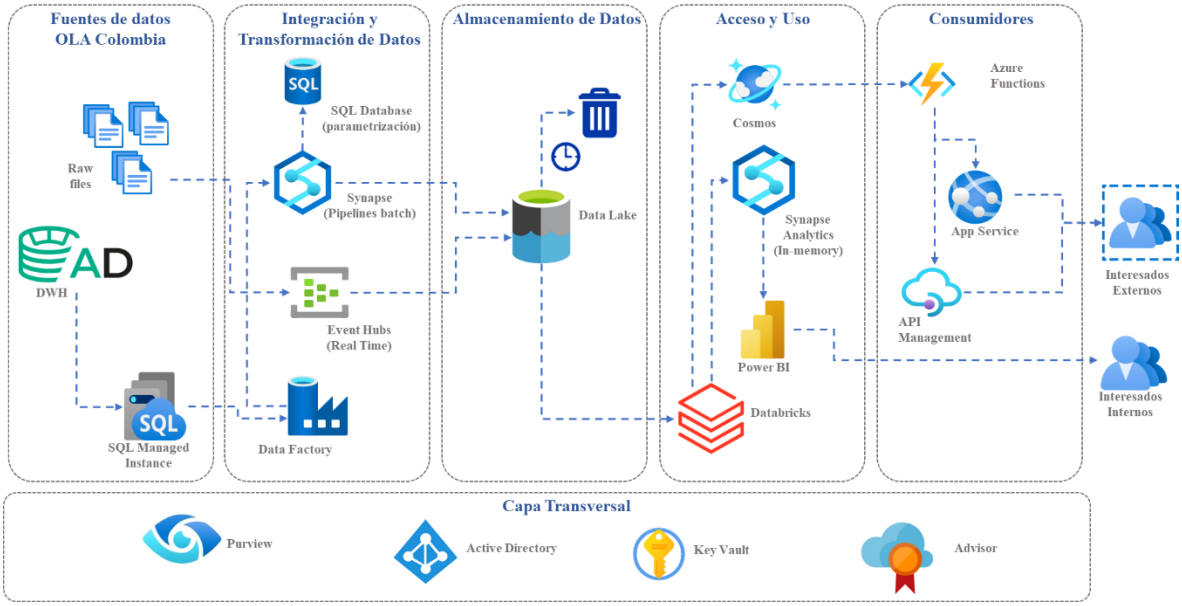
Zona transversal: agrupación de servicios utilitarios de interés común: gestión de auditoría, logs, reportes, gestión holística de la solución a través de herramientas de gobierno y seguridad de datos y de la arquitectura, donde se determinan los patrones, roles, flujos y estructura arquitectónica de los componentes.

Fase 2 – Modelamiento arquitectónico para la solución

A continuación, se presenta el diseño contemplado para la solución y su subsecuente activación de herramientas como se muestra en la figura 5.

Figura 5.

Arquitectura en bloques de herramientas dispuestas para la solución



Nota. Fuente propia

El diseño está pensado para ser racional con el uso de los recursos. Contempla un modelo de crecimiento orgánico y extensible tanto a nivel de capacidad de las herramientas como de la activación de estas. Gracias al modelo de herramientas de Azure es posible incrementar las capacidades de las herramientas en procesamiento, niveles de servicio, almacenamiento y tráfico, lo que flexibiliza el modelo de crecimiento según la demanda.

A continuación, se presenta una tabla que expone el catálogo de bloques de solución, se relaciona: bloque de arquitectura, herramienta o bloque de solución, descripción y principales funcionalidades, se describe en la siguiente tabla.

Tabla 2.

Bloque de arquitectura, herramienta o bloque de solución, descripción y principales funcionalidades

Bloque de Arquitectura / ABB	**Herramienta / SBB**	**Descripción del SBB**	**Funcionalidades principales**
Fuentes de datos OLA Colombia	SQL Managed Instance	SSB que se encarga de la gestión y administración de base de datos SQL, soporta la ingesta continua de datos, la actualización continua de información, de igual forma realiza de forma automática la optimización y gestión de carga de trabajo para consultas repetitivas. SQL Managed Instance es un servicio permite la integración de redes virtuales ofreciendo una capa de seguridad interna y perimetral para el acceso a los datos.	Gestión y administración de base de datos SQL. Parcheo automático de software. Monitoreo y métricas de forma incorporada.
Integración y Transformación de Datos	Data Factory, Azure Synapse Analytics	SBB que presta los servicios de migración de datos de distintas fuentes mediante pipeline, integration runtime y copy data.	Migración de datos. Integración con distintas fuentes de datos.
	SQL Database	SSB que se encarga de la gestión y administración de base de datos con el fin de soportar la ingesta continua de datos, actualización continua de información, realiza de forma automática la optimización y gestión de carga de trabajo para consultas repetitivas.	Gestión y administración de base de datos. Copias de seguridad de forma automática. Escalabilidad y seguridad
	Event Hubs	SBB que presta el servicio de procesamiento de eventos en tiempo real, permite a las aplicaciones procesar grandes cantidades de datos de eventos en tiempo real.	Admite el procesamiento de eventos a gran escala. Detección de anomalías, la monitorización de eventos de

Bloque de Arquitectura / ABB	Herramienta / SBB	Descripción del SBB	Funcionalidades principales
			seguridad, y el análisis de datos en tiempo real.
Almacenamiento de Datos	Azure DataLake	SBB que presta el servicio asociado con permitir que los datos sean limpiados, transformados, fusionados, mejorados o agregados y purgados que propicien la conservación y disposición final de la información procesada y almacenada en el repositorio.	Almacenamiento central de los datos en el repositorio. Dispondrá de espacio para carga inicial desde el Datawarehouse Onpremises para habilitar su procesamiento masivo, el espacio se liberará una vez los datos estén en un estado Gold y luego habilitará almacenamiento basado en deltas o en tiempo real.
Acceso y uso	Azure Synapse Analytics	SBB que se encarga del análisis de macrodatos y BigData mediante el movimiento, control y validación de datos bulk/batch y la transformación de los datos mediante la gestión de capas de almacenamiento: Bronce (datos sin procesar de forma nativa en un Data Lake), Silver (datos procesados y limpios almacenados en un Data Lake) y Gold (datos finales que se presentaran al consumidor), que provienen de distintas fuentes de datos estructurados y no estructurados.	Análisis de Big Data. Data Warehousing ETL. Bulk copy program utility (bcp). SqlBulkCopy API. Soporte a Pipelines. Telemetría de datos mediante Data Explorer.
	Azure Databricks	SBB que se encarga del procesamiento y transformación de grandes cantidades de datos a través de modelos de aprendizaje automático. Permite ingesta de datos a través de conectores aprovisionados por Azure.	Aprendizaje Automático (ML). Análisis de Big Data. Data Warehousing ETL.
	Azure Cosmos DB	Base de datos NoSQL que permite manejar datos a gran escala basado en formato JSON de forma distribuida geográficamente y multimodal a través de diferentes APIs de acceso.	Gestión e implementación de base de datos NoSQL.
	Power BI	SBB que presta los servicios de integración de distintas fuentes de datos para presentar un análisis y creación de cuadros de mando mediante transformación de datos para la toma de decisiones empresariales.	Visualización y creación de cuadros de mando. Análisis de datos de diversas fuentes.
Consumidores	Azure Functions	SBB que presta los servicios de implementación de código fuente a través del modelo Serverless (ejecución sin servidor) mediante múltiples lenguajes de programación, permitiendo el desarrollo a	Ejecución y desarrollo de microservicios y APIs sin servidor.

Bloque de Arquitectura / ABB	Herramienta / SBB	Descripción del SBB	Funcionalidades principales
		nivel local o en el portal de Azure para desplegar microservicios. Azure function se desarrolla mediante un desencadenador o trigger lo cual inicia el flujo de ejecución y permite integración con recursos de Azure mediante bindings (enlaces) de datos de entrada, salida o ambos.	
	App Service	SBB que presta los servicios de alojamiento de aplicaciones web en diferentes lenguajes de programación.	Hospedaje de aplicaciones web. API Rest. Back-Ends. Descarga FTP/SFTP de archivos CSV o demás formatos planos.
	API Management	SBB que presta los servicios de desarrollo, administración y publicación de APIs propias o de terceros que conectan los datos con las aplicaciones de forma escalable y segura mediante un único punto de entrada a través de API Gateway.	Distribución, implementación y monitoreo de APIs.
Transversal	Key Vaults	SBB que se encarga de encriptar y mantener de manera estricta y controlada el acceso a llaves, secretos, claves de API y certificados. Permite la integración de componentes el cual realizan cadenas de conexión de forma segura sin guardar credenciales en archivos de configuración.	Bóveda centralizada y encriptada de llaves, secretos, claves de API y certificados SSL/TLS. Integración de componentes mediante identidades administradas.
	Advisor	SBB que presta los servicios de monitoreo, análisis y supervisión de recursos de Azure y On-Premise. Permite la correlación de métricas y diagnóstico de anomalías asociados a recursos, creación de tableros de visualización personalizados, operaciones automáticas y supervisión de aplicaciones y componentes en tiempo real.	Monitoreo de Logs, métricas y diagnósticos de aplicaciones y componentes de Azure.
	Azure Active Directory	SBB que se encarga de administrar usuarios, contraseñas, grupos, permisos y accesos de forma centralizada a componentes propios de Azure, Microsoft 365 o recursos externos SaaS. Permite integrarse con aplicaciones mediante Single Sign-On y generar auditoria e información.	Servicio de directorio activo. Gobernabilidad de usuarios, grupos y aplicaciones. Aplicación de políticas y directivas de grupo. Integración con múltiples factores de autenticación. Gestión de licenciamiento.

Bloque de Arquitectura / ABB	Herramienta / SBB	Descripción del SBB	Funcionalidades principales
			Gestión de DNS administrado. Administración de usuarios y dispositivos mediante integración con Microsoft Intune.
	Purview	SBB gestor de metadatos crea, clasifica y gestiona automáticamente los metadatos desde las fuentes de datos de OLA y el repositorio en Azure. La información de metadatos fluye desde los bloques ABB hacia la gestión de metadatos por solicitud del gestor a través de la suscripción de la información de cada bloque, lo cual, es implementado por los SBB (bloques de solución) o herramientas materializadas.	Capacidades para la gestión como: actualización, descubrimiento, etiquetado, modificación, descripción e integración, Posee integración con fuentes de metadatos propias de la solución y aquellas que están fuera de esta, esto con el fin de registrar información sobre integraciones, transformaciones, calidad, registro y descripción de los activos de los datos.

Nota. Fuente propia

Fase 3 – Costos, calculo económico de la solución

Se presenta el cálculo de costos referente a tiempos de ejecución y recursos tecnológicos para los cuales se hizo uso de la calculadora de Azure, posterior se registran y obtienen los valores desagregados de cada elemento de la solución tecnológica basada en la nube, en la fase 3 podemos observar 3 tablas distribuidas en costos de recursos tecnológicos, recursos de programación y cronograma de ejecución

Tabla 3.

Cálculo de costos calculadora Azure

categoria de servicio	tipo servicio	Region	descripcion	costo mes estimado	costo upfront estimado
Bases de datos	Azure SQL Database	E|ast US	Single Database, vCore, General Purpose, Provisioned, Standard-series (Gen 5), Locally Redundant, 1 - 12 vCore Database(s) x 1 Month, 3000 GB Storage, RA-GRS Backup Storage Redundancy, 1500 GB Point-In-Time Restore, 0 x 1000 GB Long Term Retention	$ 2.957,62	$ 0,00
Análisis	Event Hubs	East US	Basic tier: 10 Throughput unit(s) x 1 Month, 2 million Ingress events	$ 109,56	$ 0,00

Análisis	Azure Data Factory	East US	Azure Data Factory V2 Type, Data Pipeline Service Type, Azure Integration Runtime: 10 Activity Run(s), 100 Data movement unit(s), 20 Pipeline activities, 0 Pipeline activities – External, Data Flow: 1 x 8 General Purpose vCores x 1 Month, 0 x 8 Memory Optimized vCores x 1 Month, Data Factory Operations: 1 x 50,000 Read/Write operation(s), 1 x 50,000 Monitoring operation(s)	$ 1.631,63	$ 0,00
Bases de datos	Azure Synapse Analytics	East US	Tier: Compute Optimized Gen2, Dedicated SQL Pools: DWU 100 x 1 Month, 3 TB of storage with Geo-redundant disaster recovery; Serverless SQL Pools: 3 TB of data queried; Apache Spark Pools: Memory optimized: (1 Instance(s), Small-sparkpool x 1 Month); Este de EE. UU. Region, 3 TB of data collected per day, 7 days of Hot Cache, 30 days of total retention, 7 times estimated data compression, 1 Month of 2 x Large (16 vCores) Engine Instances, 1 Month of 2 x 4 vCores Data Management Instances; Data Integration: Data flows (8 x 1 Basic vCores, 8 x 1 Standard vCores)	$ 11.939,01	$ 0,00
Análisis	Azure Data Lake Storage Gen1	East US 2	Pay-as-you-go: 3 TB Storage, 10,000 Read Transactions, 10,000 Write Transactions	$ 659,81	$ 0,00
Bases de datos	Azure Cosmos DB	East US	Azure Cosmos DB for NoSQL (formerly Core), Standard provisioned throughput (manual), Always-free quantity disabled, Single Region Write (Single-Master) - Este de EE. UU. (Write Region), 400 RU/s x 1 Month, 3,000 GB transactional storage, Analytical storage disabled, 2 copies of periodic backup storage, Dedicated Gateway not enabled	$ 773,36	$ 0,00
Análisis	Azure Databricks	West US	All-Purpose Compute Workload, Premium Tier, 1 D8AV4 (8 vCPU(s), 32 GB RAM) x 1 Month, Pay as you go, 1.5 DBU x 1 Month	$ 929,29	$ 0,00
Análisis	Power BI Embedded	East US 2	1 node(s) x 1 Month, Node type: A1, 1 Virtual Core(s), 3GB RAM, 1-300 Peak renders/hour	$ 735,91	$ 0,00
Compute	Azure Functions	West US	Consumption tier, Pay as you go, 1024 MB memory, 10,000 milliseconds execution time, 3,000,000 executions/mo	$ 474,00	$ 0,00
Compute	App Service	West US	Basic Tier; 1 B1 (1 Core(s), 1.75 GB RAM, 10 GB Storage) x 1 Month; Windows OS; 0 SNI SSL Connections; 0 IP SSL Connections	$ 54,75	$ 0,00
Web	API Management	East US	Developer tier, 1 unit(s) x 1 Month; Self-hosted Gateways, 3 Gateways x 1 Month, 1 x 1 Month x 5 workspaces	$ 48,03	$ 0,00

Análisis	Microsoft Purview	East US	Elastic Data Map: 1 Capacity Unit Month, 1 Month, Automated Scanning and Classification: 1 Total scan duration in minute x 32 Total vCores across scans (For other data sources), Other features: 1 Resources Set Month, Microsoft Purview Data Catalog: C0 Service	$ 453,67	$ 0,00
Identidad	Azure Active Directory External Identities	West US	Premium P1 tier: 50,000 monthly active user(s), 0 SMS/Phone Events	$ 0,00	$ 0,00
Seguridad	Key Vault	East US	Vault: 50,000 operations, 50,000 advanced operations, 50 renewals, 50 protected keys, 0 advanced protected keys; Managed HSM Pools: 0 Standard B1 HSM Pool(s) x 1 Month	$ 200,90	$ 0,00
Administración y Gobernanza	Azure Advisor		There are no charges to use Azure Advisor.	$ 0,00	$ 0,00
Support		**Support**	0	$ 0,00	
		Total		**$ 20.967,54**	**$ 0,00**

Nota. Fuente propia, valores en $US

Tabla 4.

Estimación horaria fases de desarrollo

Alcance	Estimación Horaria	Valor Hora 2023	Valor Total
Fase 1	138	90.803	**$ 12.530.814**
Fase 2	340	90.803	**$ 30.873.020**
Total			**$ 43.403.834**

Nota. Fuente propia, valores en COP$.

Tabla 5.

Cronograma ejecución fases de desarrollo, tiempo esperado

Desarrollo	tiempo estimado	sem 1	sem 2	sem 3	sem 4	sem 5	sem 6	sem 7	sem 8	sem 9	sem 10	sem 11	sem 12
fase 1	4 semanas												
fase 2	8 semanas												

Nota. Fuente propia

2.2 Propuesta Tecnológica Hibrida

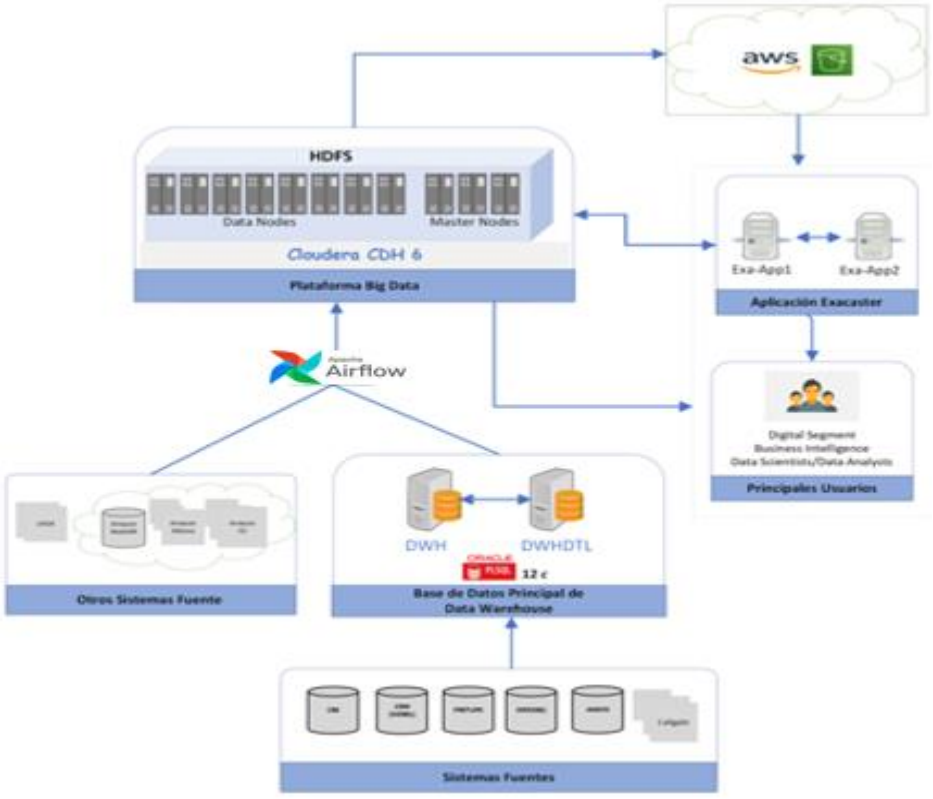
Fase 1 – Modelamiento de referencia de la solución

Apalancada en la resiente infraestructura implementada por la compañía y siendo pioneros en la implementación de este tipo de solución por la volumetría manejada, su procesamiento y demás procesos implicados para la disponibilizacion de la Data representa un reto para la calibración de dicha infraestructura, tiempos de procesamiento y sus resultados; permitirán visualizar las mejoras más significativas de cara a la problemática presentada en el proceso actual.

A continuación, como se muestra en la figura 6, la Arquitectura implementada actualmente en la compañía Ola Colombia y sobre la cual se implementaría esta propuesta tecnológica.

Figura 6.

Infraestructura de Big Data

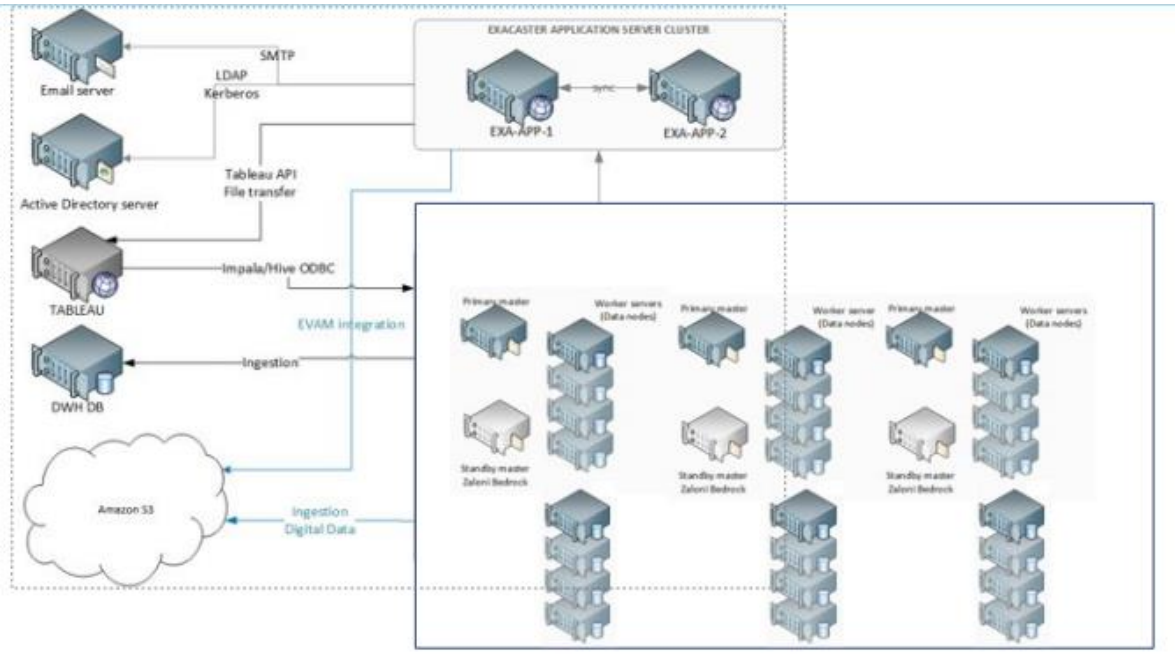


Nota. Fuente propia

En la siguiente figura02 se puede visualizar la conformación de la plataforma a nivel de equipos.

Figura 7.

Infraestructura Cloudera de Big Data



Nota. Fuente propia

A nivel de Software la plataforma Big Data cuenta con las siguientes opciones:

- Arena (orquestador de flujos)
- Airflow (orquestador de Dags)

Cloudera (integración de la plataforma); dentro de Cloudera los siguientes ejecutores:

- HIVE
- Impala
- Spark (YARN)

Scripts

- Python
- Spark
- Scala
- Json

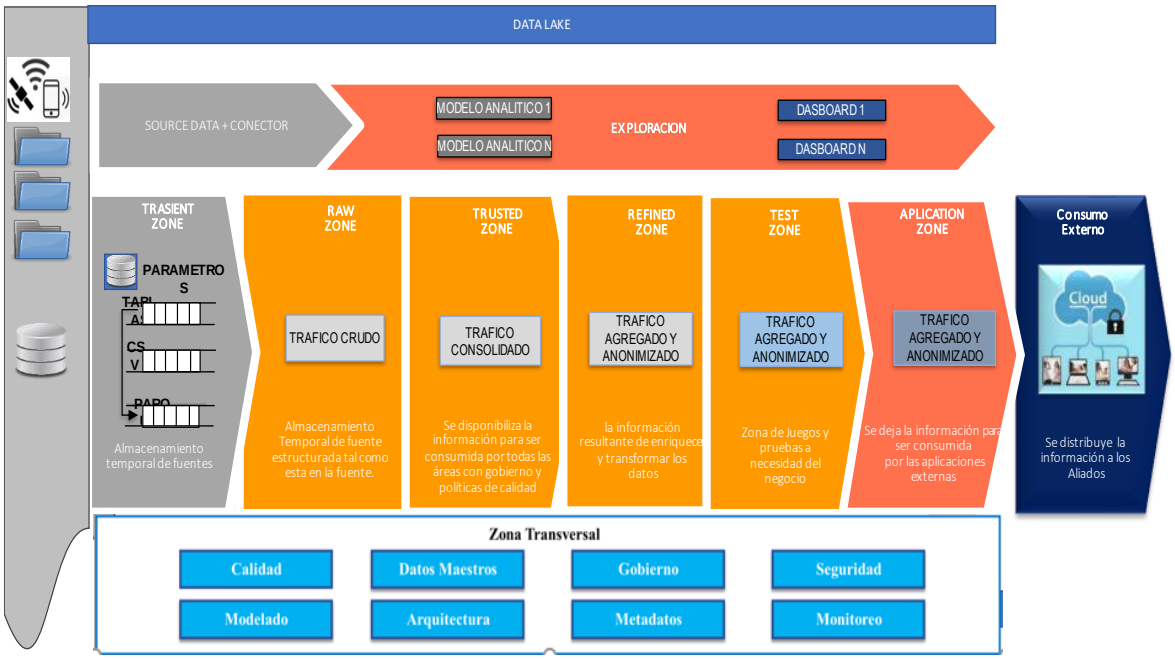
El gran diferenciador de esta propuesta es la utilización de la Infraestructura actual mitigando los costos de inversión los cuales ya estarían cubiertos en la implementación realizada por TI y en funcionamiento.

Fase 2 – Modelamiento arquitectónico para la solución

La solución se adaptaría al modelo preestablecido para la plataforma y se aplicarían los procesos de seguridad, protección de datos, herramientas ponibles para los procesos de carga, almacenamiento, limpieza, procesamiento, visualización y publicación disponibles en el ecosistema como lo muestra la figura 8.

Figura 8.

Ecosistema



Nota. Fuente propia

A continuación, realizamos una breve explicación de las diferentes zonas del modelo:

TRASIENT ZONE: Almacenamiento temporal de fuentes; la zona de aterrizaje transitoria es un almacenamiento basado en archivos donde los datos están organizados por sistemas de origen

RAW ZONE: Almacenamiento Temporal de fuente estructurada tal como está en la fuente, se implementa como un almacenamiento basado en archivos (Hadoop–HDFS). Incluye

un área de "Almacenamiento de datos sin procesar" para conservar los datos de origen para el uso activo y el archivo

TRUSTED ZONE: Se dispone la información para ser consumida por todas las áreas con gobierno y políticas de calidad; integrar varios datos y aplicar transformaciones comunes en los datos sin procesar en una estructura limpia y estandarizada que está optimizada para los consumidores de datos.

REFINED ZONE: la información resultante de enriquecer y transformar los datos; proporciona procesos para la mejora, el aumento, la clasificación y la estandarización de datos. Incluye procesos para el procesamiento de reglas comerciales automatizadas y procesos para derivar o agregar nuevos atributos a los registros existentes de fuentes internas y externas.

TEST ZONE: Zona de Juegos y pruebas a necesidad del negocio.

APLICATION ZONE: Se deja la información para ser consumida por las aplicaciones externas; se rige por un proceso de descubrimiento que se implementa internamente como funcionalidad de búsqueda, localización y recuperación a través de herramientas como Elasticsearch o Solr/Lucene. Un descubrimiento es posible gracias a la gran cantidad de metadatos que se han recopilado en las zonas anteriores.

CONSUMO EXTERNO: Se distribuye la información a los aliados; es la puerta de entrada principal para los usuarios externos al lago de datos. La clave para implementar un nivel de consumo funcional es la cantidad y la calidad de los metadatos que habríamos recopilado en las zonas anteriores y la forma inteligente en que expondremos estos metadatos para la búsqueda y recuperación de datos

ZONA TRANSVERSAL: La capa de seguridad y gobierno de datos corrige la responsabilidad de regular el derecho de acceso a los datos y los derechos de definición y modificación de datos. Esta capa se asegura de que haya un proceso bien documentado para el cambio y el control de acceso de todos los artefactos de datos. El mecanismo de gobierno supervisa los métodos para la creación, el uso y el seguimiento del linaje de datos en varios niveles del lago de datos para que pueda combinarse con las reglas de seguridad.

Fase 3 – Costos, calculo económico de la solución

Como se mención anteriormente esta solución tecnológica utilizara la infraestructura actual por lo cual sus costos iniciales serán mínimos y por ende ofrecerá con su entrega de valor ingresos para la compañía de forma inmediata, ver tabla 6.

Tabla 6.

Infraestructura actual

Categoria	Servicio	Costo
Infraestrututa	Equipos - Cloudera	$0,0
Base datos	Oracle	$0,0
Software	Licencias	$0,0
Mantenimiento	Validacion de recursos y funcionamiento	$0,0
Operación	Gestion de escalamientos y validacion de procesos	$0,0
Total		$0,0

Nota. Fuente propia

Tabla 7.

Estimación horaria y costos fases de desarrollo

Alcance	Estimación horaria	valor hora 2023	valor total
fase 1	270	85.000	$ 22.950.000
fase 2	405	85.000	$ 34.425.000
Total			**$ 57.375.000**

Nota. Fuente propia

Tabla 8.

Cronograma ejecución fases de desarrollo, tiempo esperado

Desarrollo	T Estimado	SP01			SP02			SP03		
		Sem1	Sem2	Sem3	Sem4	Sem5	Sem6	Sem7	Sem8	Sem9
Fase1	3 semanas									
Fase2	6 semanas									

Nota. Fuente propia

3. Resultados

3.1 Propuesta Tecnológica Basada En Servicios De La Nube

El modelo propuesto integra los datos en un Datalake, los dispone en las tres capas de procesamiento. (Microsoft, 2023) Bronze (Raw), Silver (Trusted) y Gold (Refined) que mediante la ingesta ágil con la implementación de procesos ELT y flujos de procesamiento por lotes provenientes desde distintos orígenes, conforma una plataforma de datos unificada y mediante la aplicación de reglas de negocio y de seguridad con procesos de enmascaramiento, habilita los datos para que puedan ser usados por las unidades organizacionales internas para análisis, informes de BI, ciencia de datos, inteligencia artificial y aprendizaje automático, y también para las organizaciones externas con quien se tenga convenio.

Ola Colombia presenta un grado importante de resistencia de adopción de la tecnología de computación en la nube, no se desprende del uso de hardware y software propio por cuanto valora en demasía su experiencia en los dominios de accesibilidad, seguridad, almacenamiento de la información y control económico de sus activos. Pero da un paso muy importante hacia el uso de la nube al iniciar un proyecto como este donde se abre la posibilidad de mejorar sus ingresos mediante la comercialización de los datos que gestiona de las llamadas por celular.

3.2 Propuesta Tecnológica Hibrida

La utilización de Big data para la recopilación de grandes cantidades de datos y la búsqueda de tendencias dentro de los datos permite que la empresa se mueva mucho más rápidamente, sin problemas y de manera eficiente. También le permite eliminar las áreas problemáticas antes de que los problemas acaben con sus beneficios o su reputación.

Con respecto al cloud computing, se observa que ninguno de los sistemas está actualmente alojado en la nube y que el equipo no posee conjuntos de habilidades relacionadas con la tecnología. Estos hechos, junto con las preocupaciones sobre la privacidad de los datos, llevan a esta propuesta a que la solución sea creada en Big Data local; pero se deja la opción de alojamiento basado en la nube porque se especula que el paso a seguir por parte de la compañía en futuro es iniciar con la migración de software como servicio alojado en la nube.

Esta propuesta híbrida utiliza gran porcentaje de su infraestructura tecnológica en tierra más su experiencia relacionada aunada con la experiencia y conocimiento que el proveedor de

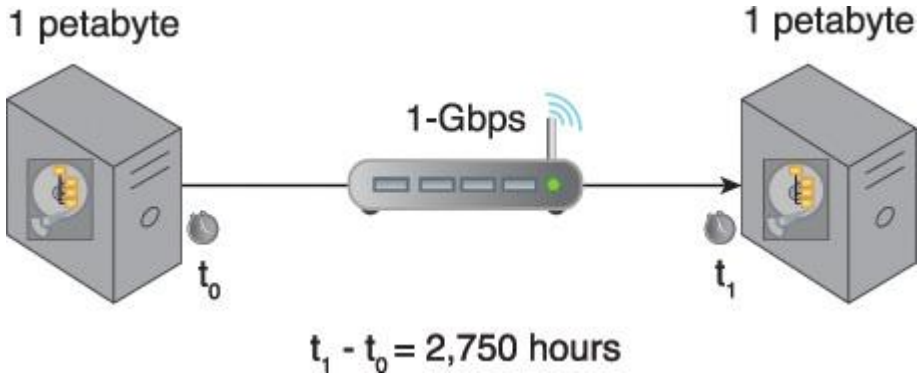
nube ofrece, en el marco de un contrato solido que cuente con un soporte al cliente adecuado y acuerdos de nivel de servicio (SLA) garantizados, el inicio progresivo de la adopción de Cloud Computing será exitoso generando óptimas operaciones en las cadenas de valor, reduciendo los reprocesos, garantizando la seguridad y comprometiéndose en la competitividad de economía de los datos.

Desafío de Rendimiento

Debido a los volúmenes de datos que a futuro algunas soluciones de Big Data deban procesar, el rendimiento suele ser una preocupación. Por ejemplo, grandes conjuntos de datos junto con algoritmos de búsqueda complejos pueden generar tiempos de consulta prolongados. Otro desafío de rendimiento está relacionado con el ancho de banda de la red. Con el aumento de los volúmenes de datos, el tiempo para transferir una unidad de datos puede exceder el tiempo real de procesamiento de datos, como se muestra en la Figura 9.

Figura 9.

Transferencia de datos



Nota. Fuente propia

La transferencia de 1 PB de datos a través de una conexión LAN de 1 Gigabit con un rendimiento del 80 % llevará aproximadamente 2,750 horas.

4. Discusión

Esta consultoría plantea dos propuestas tecnológicas, las cuales son presentadas para dar una solución completa a la necesidad de monetizar la información de registro de llamadas telefónicas móviles o de tecnología celular que gestiona la empresa Ola Colombia, la solución tecnológica propuesta basada en la nube representa una mejora significativa en comparación a la propuesta tecnológica hibrida que decanta el uso de servidores en nube solo para disposición final y consta de servidores locales para el almacenamiento y manejo de la información, si bien en la decisión final intervienen factores estructurales y económicos revisados en el capítulo de resultados, se opta por la alternativa más económica donde se hace un gran porcentaje de reutilización del sistema informático Onpremises que ya está provisto en la actualidad. Cabe anotar que esta propuesta de consultoría es la primera incursión de la empresa en el uso de servicios de nube buscando tener mayor confianza en este tipo de computación de acuerdo con los resultados que arroje.

Existe cierto grado de resistencia de adopción de este tipo de tecnología tanto a nivel informático como económico debido a la credibilidad de los mecanismos de respaldo y garantía que proveen computación en la nube con respecto a la gobernabilidad de los datos y su seguridad, y la difícil tarea de llevar un correcto control y disposición de los activos de hardware y software que ya se poseen para no generar pérdidas económicas al momento de que estos elementos pasen a un segundo plano en la producción o se deban dejar de utilizar.

Pero lo anterior no es una barrera para la implementación a mediano plazo de la primera alternativa ya que representa el inicio de una adopción casi completa de la tecnología que ofrece la nube, solo que Ola Colombia requiere tener mayor certeza y con resultados tangibles de su viabilidad, basándose en las estimaciones y transacciones económicas registradas en cada aspecto de la segunda alternativa y que en combinación con la adquisición progresiva de experiencia tecnológica por parte de los ingenieros que hacen parte de planta de personal, la puedan poner en marcha junto con otro conjunto de iniciativas que se habilitarían al contar con ese Cloud Computing. La alternativa seleccionada pondrá en manifiesto los beneficios y bondades no solo para la solución propuesta sino para otros diversos sistemas de información que explotaría más efectivamente los datos de telefonía celular en conjunto con otros datos de diferente tipo y de diversas fuentes que se gestionan al interior y que también podrían recibirse de empresas asociadas, impulsando la transición a la primera alternativa.

En el apartado de costos que tenemos en el capítulo de desarrollo metodológico de la presente consultoría, encontramos los valores relacionados con la implementación, soporte y desarrollo de ambas propuestas tecnológicas. Se resaltan las siguientes diferencias entre ambas por su marcada relevancia a la hora de decidir cual implementar, la propuesta tecnológica basada en la nube tiene un costo inicial de implementación de $43.403.834 COP y un valor mensual por soporte de $93.232.166 COP. La propuesta hibrida tiene un costo inicial de $57.375.000 COP y no requiere de un valor mensual de soporte. Si bien la diferencia en costos inicial es del 24%, esta diferencia pasa a segundo plano si tenemos presente que en un año el gasto de soporte de una propuesta tecnológica es de $1.118.785.992 COP y la de la otra es $0 COP.

5. Conclusiones y recomendaciones

5.1 Conclusiones

Con la propuesta tecnológica seleccionada en esta consultoría suministra la facilidad de almacenar, procesar y dejar disponible cualquier tipo de datos y archivos sin preocuparse por la compatibilidad ya que la garantiza al permitir formatos multivariados, asegura los datos con sistemas modernos de almacenamiento, su resguardo, durabilidad, seguridad y disponibilidad por medio de dispositivos móviles y de ordenadores conectados a internet. Al optimizar la disponibilidad de los datos se reduce el tiempo de su comercialización, mejora la experiencia de los clientes y partners, se generan nuevos flujos de ingresos llevando estos datos al mercado en forma de productos y servicios de datos a la medida, puestos a disposición de los interesados y cualquier otra organización que sea parte de la economía de los datos. (PowerData, 2023) Es una solución que garantiza un mayor provecho de los datos mediante una racional inversión en tecnología y adecuada reutilización de la infraestructura actual.

La configuración de los datos en repositorios o buckets de Amazon Simple Storage Service S3, permite que sean consumidos por los clientes mediante aplicaciones nativas en la nube y aplicaciones móviles. Esta plataforma de almacenamiento posee clases de almacenamiento rentables y funcionalidades fáciles de administración, generando optimización de costos, organización de los datos y uso de controles de acceso acordes a los requisitos empresariales. Esta disposición de tecnología impulsa a Ola Colombia a compartir los datos a nivel local y/o global, a identificar nuevas oportunidades de mercado, a racionalizar

sus operaciones y a desarrollar relaciones de datos con socios del ecosistema empresarial. También a nivel de sus procesos internos, la explotación de estos datos y en correlación con los propios de cada área, habilitan el análisis histórico, predictivo y prescriptivo, generando ventajas competitivas.

Esta consultoría genera un planteamiento económico en torno a una propuesta tecnológica basada en servicios de la nube y una propuesta tecnológica hibrida, ambas dirigidas a la compañía Ola Colombia, en el capítulo de desarrollo metodológico se explica y hace referencia a cada propuesta y sus fases de aplicación así como los costos relacionados estimados que sean requeridos para el desarrollo y la implementación tecnológica en un tiempo estimado mínimo de 9 semanas y máximo de 12 semanas, los valores entre las propuestas tecnológicas presentan 2 diferencias importantes, la propuesta tecnológica basada en servicios de la nube tiene una inversión inicial menor a su contraparte pero esta pierde peso si tenemos en cuenta que se requiere de un gasto mensual de soporte mientras que la propuesta tecnológica hibrida solo requiere una inversión inicial y se apoya en infraestructura ya existente en la compañía. La propuesta seleccionada se encuentra dentro del rango de costo promedio de mercado y ofrece una solución definitiva para la disponibilidad y monetización de los datos, generando ingresos superiores con relacion a la inversión tecnológica.

5.2 Recomendaciones

Con base en los resultados obtenidos en el uso de la solución seleccionada, la segunda, se opte por la utilización total, parcial o mejorada que ofrece la primera propuesta, de acuerdo con la velocidad que cambia esta tecnología. El Cloud Computing es una realidad y los sistemas modernos se están rediseñando para la transición a este tipo de tecnologías y de acuerdo con su aceptación empresarial global se evidencia que el uso de este tipo de arquitecturas ya no hay vuelta atrás.

Se deja claro que una vez se adopta Cloud Computing también se puede y de acuerdo con el escenario organizacional específico, integrar con los activos IT Onpremises, en data centers propios, pero siempre buscando el equilibrio en la solución, de manera prudente y lógica.

Ola Colombia debería continuar con el uso de la computación en la nube tanto para expandirse en su ecosistema empresarial, mejorar la experiencia del cliente con relación a sus

productos y servicios, y al mismo tiempo lograría reducir gastos en algunos sistemas antiguos, legados y difíciles de mantener por tecnología, conocimientos, logística, etc.

Bibliografía

CRC. (16 de 05 de 2023). *Postdata mas alla de los datos*. Obtenido de https://www.postdata.gov.co/dataflash/data-flash-2023-005-internet-movil

Curry, E. (2016). The Big Data Value Chain: Definitions, Concepts, and Theoretical Approaches. *New Horizons for a Data-Driven Economy*, 29 - 37.

Faroukhi, A. E. (08 de 06 de 2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal Big Data*. Obtenido de journal Big Data: https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0281-5

Microsoft. (2023). *Azure*. Obtenido de Canalizaciones de ingesta, ETL y procesamiento de flujos con Azure Databricks y Delta Lake: https://learn.microsoft.com/es-es/azure/architecture/solution-ideas/articles/ingest-etl-stream-with-adb

MinTic. (12 de 2022). *Boletín trimestral del sector TIC - Cifras cuarto trimestre de 2022*. Obtenido de https://colombiatic.mintic.gov.co/679/w3-article-275427.html

packt. (2023). *Data Lake Architecture*. Obtenido de Packt: https://subscription.packtpub.com/book/big-data-and-business-intelligence/9781785888083/1/ch01lvl1sec14/data-lake-architecture

Palmer, Michael. (2006). *red-forbes-los-datos-son-el-nuevo-petroleo*. Obtenido de https://www.forbes.com.mx/red-forbes-los-datos-son-el-nuevo-petroleo/

PowerData. (2023). *Cómo ganar en la economía de datos actual*. Obtenido de https://landings.powerdata.es/lp-guia-como-ganar-en-la-economia-de-datos-actual

SIC.GOV.CO. (2023). *superintendencia de industria y comercio*. Obtenido de GOV.CO: https://www.sic.gov.co/manejo-de-informacion-personal