



## **Ingeniería para Big Data**

### **Trabajo obligatorio**

#### **Caso de estudio: “Una comunidad para compartir desafíos deportivos – Data Fitness”**

Matías Bernardo 207870

Florencia Marzano 292054

Martín Martínez 308676

Indice

Introducción ..... 3

Objetivos principales del negocio ..... 3

Requerimientos funcionales ..... 5

Fuentes de datos ..... 6

Arquitectura de referencia..... 7

Tecnologías a utilizar ..... 8

Hoja de ruta..... 10

Diseño de dashboards ..... 11

Conclusión ..... 11

Anexo 1..... 12

Anexo 2..... 12

## Introducción

Data Fitness afronta la necesidad de evolucionar y renovar su aplicación para mantenerse competitiva y brindar una experiencia óptima a sus usuarios.

Este informe aborda el proyecto de rediseño de la aplicación de Data Fitness, una plataforma que surgió con la necesidad de compartir información sobre entrenamientos, competencias y otros datos relevantes para deportistas de diversas disciplinas.

El objetivo principal de este proyecto es modernizar la aplicación, centrándose en el procesamiento de Big Data, consultas en tiempo real, la expansión de las capacidades de vinculación entre usuarios y la introducción de nuevos servicios.

A lo largo de este informe, se explorarán los objetivos de negocio más relevantes, se detallarán los requerimientos funcionales, se analizarán las fuentes de datos involucradas y se establecerá un modelo de arquitectura de referencia. Además, se propondrán las tecnologías a utilizar, el diseño general de resultados y se trazará una hoja de ruta para el desarrollo del proyecto.

## Objetivos principales del negocio

Como primer objetivo, se plantea mejorar la capacidad y rendimiento de la aplicación para el procesamiento y análisis de grandes volúmenes de datos, proporcionando resultados en tiempo real.

Se proyecta una expansión considerable de la base de los usuarios, con la meta de un aumento anual del 200%. Esto implica que, para poder cumplir con este objetivo, se requiere no solo aumentar la capacidad de la aplicación para gestionar un amplio volumen de datos (estructurados y no estructurados), sino que se deben asegurar la escalabilidad y el rendimiento. A su vez, para poder mantener relevancia y ser competitivos en el mercado de aplicaciones deportivas, se debe asegurar que el procesamiento y análisis de los datos sea en tiempo real.

El siguiente objetivo del proyecto consiste en desarrollar e implementar una arquitectura de procesamiento de Big Data, que permita gestionar información estructurada y no estructurada en tiempo real y a gran escala. Dada la necesidad de procesar grandes volúmenes de datos generados por la creciente base de usuarios y la diversidad de dispositivos deportivos, es esencial mejorar la infraestructura tecnológica. La capacidad de procesamiento eficiente respalda la funcionalidad de procesar datos en tiempo real, garantizando una experiencia óptima para los usuarios y preparando la aplicación para un escenario global.

Por último, se busca mejorar la retención de usuarios existentes fomentando la fidelización mediante la entrega de experiencias personalizadas, contenido relevante, servicios exclusivos y creación de comunidades. En lugar de enfocarse únicamente en el crecimiento numérico, con este objetivo se pone énfasis en la calidad de la experiencia de usuario y en mantener a los mismos comprometidos a largo plazo. Se debe buscar incrementar la retención a través de la personalización de servicios. Por ejemplo, ofreciendo planes de entrenamiento y nutrición específicos, y generando espacios que fomenten la creación de una comunidad activa y enriquecedora. Un usuario conforme y comprometido con el servicio no solo contribuirá a la estabilidad de la base de usuarios, sino que también puede convertirse en un defensor de la marca, atrayendo a nuevos usuarios en base a sus propias recomendaciones.

Los stakeholders clave involucrados en el desarrollo e implementación de la nueva solución, desempeñarán roles cruciales en diversas etapas del proyecto. Para ello resulta clave identificarlos.

Como primer parte interesada se pueden identificar a los usuarios actuales y potenciales, pueden ser deportistas, entrenadores, o cualquier persona que utilice los servicios de Data Fitness como usuario final. A su vez, dentro de este grupo se pueden generar comunidades de deportistas y entrenadores profesionales que pueden tener gran influencia en el crecimiento de la plataforma, ya que estos pueden influir en la popularidad y la percepción de la aplicación dentro del ámbito deportivo y pueden actuar como embajadores o críticos de la plataforma.

Dentro de los factores relevantes para este grupo de stakeholders se encuentran: experiencia de usuario, funcionalidades útiles, facilidad de uso, acceso a planes personalizados, interacción social, informes detallados sobre su rendimiento deportivo.

Otros actores clave son quienes conforman el equipo de desarrollo y tecnología, los cuales pueden ser programadores, científicos de datos o personas que trabajan en la implementación de la arquitectura de Big Data. Tienen como finalidad implementar de forma eficiente la arquitectura de Big Data, desarrollar y realizar el mantenimiento de la aplicación y lograr la integración efectiva con diversos dispositivos deportivos.

La directiva y gerencia, junto a los inversionistas de Data Fitness son quienes están a cargo de la toma de decisiones, planeamiento estratégico y financiamiento de la organización. Este grupo tiene como objetivo lograr el crecimiento de usuarios, posicionarse de forma competitiva en el mercado, formar alianzas estratégicas y buscar nuevas estrategias de inversión a largo plazo.

Se identifican a su vez, las empresas proveedoras de dispositivos deportivos, que pueden ser tanto compañías que produzcan smartwatches (amazfit, apple watch, etc), como también distintas marcas que se encarguen de desarrollar y comercializar smartphones. El foco de interés de este grupo se centra en la integración efectiva de sus dispositivos con la nueva aplicación, lo que permitirá mantener una relación comercial continua. Pueden beneficiarse del acceso a datos de usuarios para mejorar sus productos.

Las empresas asociadas y patrocinadores conforman parte de los involucrados en la nueva solución. La colaboración efectiva entre Data Fitness y las empresas asociadas se traduce en brindar servicios personalizados a los usuarios, acceso a datos agregados para análisis de mercado, participación en estrategias comerciales conjuntas, crecimiento de la red de empresas asociadas. A su vez, el éxito de la aplicación puede atraer a nuevas empresas y marcas interesadas en asociarse para realizar promociones cruzadas, publicidad, o para ofrecer productos y servicios complementarios a los usuarios de la aplicación.

Los departamentos de marketing y comunicaciones tienen como objetivo principal la búsqueda de estrategias de promoción de la nueva aplicación, lograr una comunicación efectiva de las características y beneficios que se ofrecen, atraer a nuevos usuarios y fidelizar a los ya existentes.

Las entidades regulatorias y el departamento legal buscan cumplir con las regulaciones de privacidad de datos, seguridad de la información, transparencia en las prácticas comerciales y gestión ética de datos de usuarios. Serán los responsables de que la aplicación pueda lanzarse de forma pública al mercado por lo que también se identifican como actores clave.

Por último, los proveedores de servicios financieros son las entidades que brindan las herramientas para que los usuarios finales puedan realizar pagos dentro de la plataforma.

Dentro de los factores de interés más relevantes para este grupo se encuentran la integración efectiva de servicios financieros en la aplicación, oportunidades de financiamiento para productos deportivos y la colaboración en estrategias de venta.

## Requerimientos funcionales

Luego de establecer los objetivos estratégicos e identificar los stakeholders clave, se abordan los requisitos funcionales que representan el conjunto de capacidades esenciales para cumplir con las expectativas de los usuarios finales, y garantizar la viabilidad y el rendimiento de la plataforma en un mercado dinámico y competitivo.

La buena gestión de usuarios es el primer requisito que se identifica. La aplicación debe permitir a los usuarios registrarse, suscribirse en modalidades básicas, mensuales o anuales, facturar y mantener un registro de pagos y de uso de las diferentes funcionalidades. Además, se podría contemplar la futura integración de servicios de venta de productos deportivos con opciones de financiamiento y entrega a domicilio, lo cual podría promover a la generación de alianzas con diferentes marcas. Para ello resulta necesario asegurar la escalabilidad, ya que a medida que aumente el número de usuarios y por ende la cantidad de datos, el sistema deberá escalar de manera eficiente, asegurando un buen manejo de la carga de datos. Por otro lado, resulta fundamental garantizar la seguridad de los datos personales de los usuarios.

Más allá de la experiencia del usuario, se coloca el foco en la integración con dispositivos deportivos. La aplicación debe ser compatible con la mayor cantidad de dispositivos deportivos existentes y ser fácilmente adaptable a nuevos dispositivos que puedan surgir en el futuro, para esto es necesario que exista una buena interoperabilidad, lo cual implica una buena capacidad para trabajar con diferentes protocolos y formatos de datos.

El siguiente requisito funcional se centra en el acceso multiplataforma. Los usuarios deben poder acceder a la aplicación desde diferentes dispositivos, tanto para ingresar información como para realizar consultas, así sean de datos en real time o en batch. La aplicación debe garantizar una experiencia optimizada en diferentes tipos de dispositivos, así como celulares (tanto con *Android* o *ios*), tablets y computadoras. Por lo tanto, se debe asegurar la usabilidad de la plataforma. La experiencia de uso del usuario es sumamente importante, la app debe ser consistente y estar optimizada en los dispositivos y plataformas en las que puede ser utilizada. A su vez, se debe asegurar la disponibilidad de la app, dado que el usuario debe poder utilizar la aplicación en todo momento.

Asimismo, la aplicación debe ser capaz de registrar y procesar información en tiempo real (procesamiento de datos en stream) de cada actividad en el modelo de datos correspondiente. Lo cual implica gestionar miles de registros simultáneos y permitir a los usuarios enviar actividades no registradas en modo batch en caso de problemas de conectividad. Por ejemplo, en casos donde las actividades deportivas se realicen en lugares donde la conexión pueda ser mala o inexistente. Lo cual implica que debe asegurarse el rendimiento, permitiendo que Data Fitness pueda procesar grandes volúmenes de datos en tiempo real y de forma eficiente. Por otro lado, deberá tener tolerancia a fallos, permitiendo a los usuarios enviar registros de actividades no registradas en modo batch cuando lo requiera.

El usuario debe poder establecer sus gustos y preferencias por tipo de actividad, así también como sus datos personales, para que luego al integrarlos con los datos específicos de cada actividad realizada se pueda construir el perfil completo del deportista. Para lograrlo, la aplicación debe ser capaz de recopilar y almacenar esta información eficientemente. Teniendo

en cuenta esto es importante que se cumpla con excelentes estándares de privacidad, de manera de garantizar la seguridad de la información personal de los usuarios.

Abordando aspectos de consulta y análisis de actividades, los usuarios deben poder realizar consultas de actividades en forma individual, analizarlas por períodos determinados y/o agruparlas según criterios específicos (por ejemplo, por tipo de actividad). A su vez, los entrenadores deben poder visualizar los registros de los usuarios que conformen equipo de entrenamiento. Finalmente, la aplicación debe enviar periódicamente a sus usuarios resúmenes de su actividad deportiva, mostrando mejores registros y promedios para identificar áreas de mejora. Por lo tanto, debe asegurar la velocidad y la fiabilidad de los datos, para poder ofrecerle al cliente una mejor experiencia de uso.

Es importante que los usuarios puedan realizar la evaluación de su rendimiento para enfrentar competencias específicas en función de su rendimiento previo. Esto, entre otros beneficios, puede permitir a los usuarios establecer planes de entrenamiento y/o nutrición personalizados para mejorar sus habilidades y prepararse para desafíos específicos. Para esto se deberá asegurar la precisión de los datos, para que la evolución del rendimiento sea realmente representativa.

Los usuarios deben tener la posibilidad de vincularse como si se tratase de una red social de deportistas, generando recomendaciones de interacción por actividades realizadas, e incentivando la competencia a través de resultados. También deberán poder formar grupos por equipos de entrenamiento y asociar al entrenador. Además, se debe permitir la creación de álbumes de fotos, así como también el realizar comentarios y dejar likes sobre todo el contenido social de la plataforma. Para esto es necesario garantizar la seguridad de las interacciones sociales y una conectividad eficiente y rápida, permitiendo la interacción entre usuarios en tiempo real.

Por último, la aplicación debe ser capaz de generar reportes consolidados de actividades deportivas a nivel regional, de ciudad o país. Estos reportes deben incluir información sobre las actividades más ejecutadas, horarios, tipo de terreno, y otros datos relevantes para usuarios y socios comerciales. Por lo tanto, se debe asegurar la fiabilidad y precisión de los datos para la confección de los reportes.

## Fuentes de datos

Para el análisis de la arquitectura de la nueva aplicación para Data Fitness, se requiere explorar las fuentes de datos que conformarán su funcionamiento. Estas fuentes no solo actúan como puntos de origen de información, sino también como apoyo a decisiones estratégicas. Desde los datos de dispositivos utilizados para actividades deportivas, hasta los datos personales de los usuarios, los datos de pagos y facturación y los datos de redes sociales y de interacción entre usuarios, cada fuente de datos, una vez integrada y procesada de manera eficaz, contribuirá a la creación de una plataforma dinámica y adaptable. (Ver anexo 1)

La fuente de datos de dispositivos para actividades deportivas se centra en la información relacionada con los dispositivos utilizados para registrar actividades deportivas. Incluye diferentes tipos de dispositivo, como por ejemplo smartwatches, smartbands, smartphones, etc. Es importante que el equipo de desarrollo tenga en cuenta las diferentes características técnicas de estos dispositivos para asegurar un correcto funcionamiento de la aplicación en cada uno de los mismos, de manera de establecer una buena comunicación con la aplicación. A su vez, este tipo de dispositivos cuenta con información geoespacial, como la localización de las actividades

o información regional. Por otro lado, los usuarios al loguearse a la aplicación deberán completar datos de usuario que contienen información personal como su nombre, edad, género, preferencias deportivas, país, etc. A su vez, quedará registrada la fecha de suscripción y de sus entrenamientos. Estos datos podrían complementarse con información de salud relevante como por ejemplo condiciones médicas que sean de interés para la actividad física. Para los usuarios que opten por la suscripción anual en lugar de la suscripción básica, se contará con datos de pago y facturación realizados, como el detalle de preferencia de pago, elección de la tarjeta, si se realiza en cuotas o en un único pago, si opta por la opción de renovación y cobro automático y el detalle del plan de preferencia. Por último, se encuentran los tipos de planes de suscripción. Esta fuente de datos se relaciona con la actividad de los usuarios en redes sociales, incluye conexiones entre usuarios, actividades compartidas, comentarios y mensajes. Estos aspectos facilitan y fomentan la conexión entre los miembros de la comunidad, y además promueven una cultura de apoyo mutuo entre pares.

Para la nueva aplicación de Data Fitness resulta fundamental la capacidad de extraer, procesar y transformar datos provenientes de diversas fuentes. En este contexto, cada fuente de datos resulta esencial para la toma de decisiones. Para ello, a continuación, se explorarán los atributos más relevantes de cada fuente.

Para la fuente de datos de dispositivos para actividades deportivas se consideran los siguientes atributos: ID de usuario (clave primaria), tipo de dispositivo (smartwatch, smartband, smartphone, etc.), Características técnicas del dispositivo, Información geoespacial (ubicación de actividades, datos regionales), fecha de la actividad, sistema operativo.

Para los datos de usuarios es necesario contar con: ID de usuario (clave primaria), nombre, edad, género, preferencias deportivas, país, etc.; e-mail, fecha de suscripción y entrenamientos registrados, datos de salud relevantes, detalles de pagos (si aplica).

Para la base de datos de pagos y facturación se destacan los siguientes atributos: ID de usuario (clave foránea), ID de transacción (clave primaria), detalle de preferencia de pago, elección de tarjeta, opciones de pago (cuotas, único pago), opción de renovación y cobro automático, detalle del plan de preferencia.

Para obtener datos sociales y de interacción entre usuarios son necesarios: ID de usuario (clave primaria), conexiones entre usuarios (amistades, seguidores), actividades compartidas, comentarios y mensajes.

La tabla de datos de usuarios se relaciona con datos de dispositivos para actividades deportivas a través del ID de usuario. Los datos de pagos y facturación están asociados a la tabla de datos de usuarios mediante el ID de usuario. La interacción entre usuarios se conecta a la tabla de datos de usuarios a través del dato de ID de usuario. Este modelo de datos asegura una gestión efectiva de la información, permitiendo la adaptabilidad a diferentes dispositivos, registros detallados de usuarios y sus interacciones sociales, así como un seguimiento preciso de pagos y suscripciones.

## Arquitectura de referencia

El diseño de procesamiento de Big Data que se utiliza en la nueva aplicación de Data fitness está basado en la arquitectura Lambda, la cual utiliza en simultáneo flujos de datos en streaming y batch. Esta arquitectura permite reducir las latencias, escalar la solución y aumentar la tolerancia a fallos. Para esto, se identifican 4 capas distintas de procesamiento:

La primer capa es la de ingesta o de consumo de datos, en esta etapa se obtienen los datos desde las fuentes, para este caso los datos provienen de dispositivos como smartwatches, smartbands, smartphones, etc mediante los cuales se generan archivos en formato json o csv, e inmediatamente envía los datos a la capa de Streaming. En el caso de la ingesta en Batch, la herramienta de ingesta almacena los datos acumulados en un almacenamiento persistente para luego procesarlos y enviarlos a la capa de batch.

La capa batch es de almacenamiento distribuido, con tolerancia a fallos y replicación. La cual tiene todos los datasets disponibles en el sistema de forma tal, que puede ejecutar procesos de consolidación y actualizar las vistas. Dado que se espera un incremento significativo en el número de usuarios y la cantidad de datos generados por las actividades deportivas, la capa de procesamiento en batch permite manejar grandes volúmenes de información de manera eficiente. Esto es importante para garantizar un rendimiento óptimo y evitar problemas de escalabilidad.

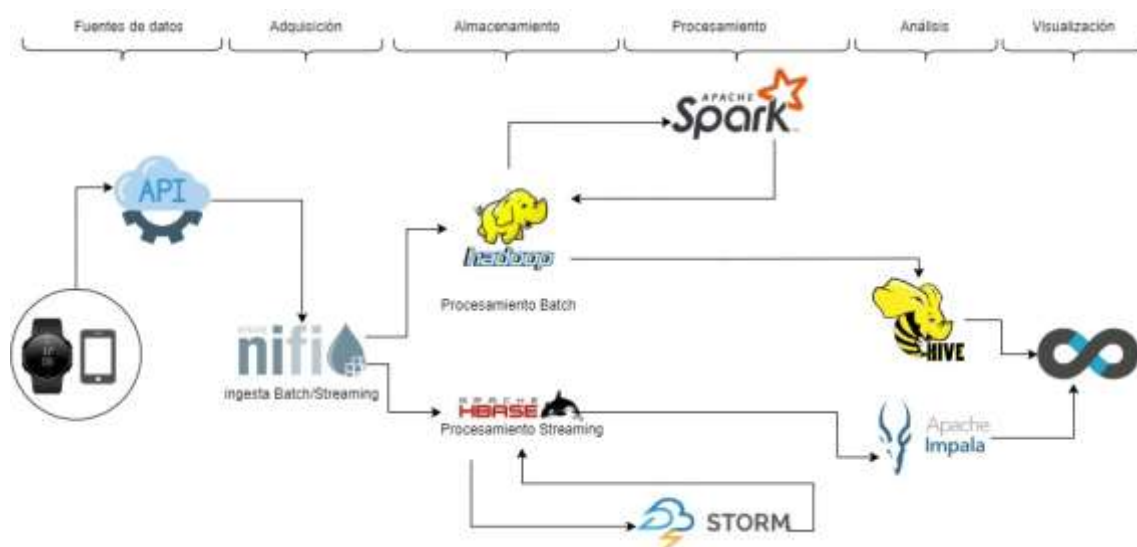
Aplicado al caso, se incluyen en la capa batch la información correspondiente a los pagos realizados por los usuarios, actualizando los datos cada cierta cantidad de tiempo dependiendo del volumen diario de nuevas transacciones realizadas.

La capa de streaming o speed layer permite procesar y analizar los datos de actividades deportivas en tiempo real a medida que son generados por los dispositivos de los usuarios. Esto garantiza que los usuarios obtengan información instantánea sobre sus actividades, lo cual es clave para mejorar la experiencia del usuario y brindar retroalimentación inmediata.

Con la capa de streaming, la generación de reportes y resúmenes de actividades puede realizarse de manera inmediata. Esto significa que los usuarios podrán recibir informes periódicos sobre su rendimiento en tiempo real, lo que brinda una experiencia más dinámica y atractiva.

La capa de servicio o serving layer es la encargada de combinar los resultados de los pipelines batch y de streaming para tener los resultados unificados ya que las actualizaciones de streaming se producen en tiempo real y las actualizaciones de la capa batch de manera periódica. La capa de servicio facilita un acceso eficiente a datos consolidados almacenados en la capa de almacenamiento (Hadoop). Los usuarios pueden realizar consultas y acceder a informes sin comprometer la velocidad de respuesta, lo que garantiza una buena experiencia de usuario.

## Tecnologías a utilizar





La generación de datos comienza en dispositivos inteligentes (smartphones, smartbands, smartwatches, etc), donde se recopila la información relacionada con las actividades deportivas de los usuarios. Estos datos se transfieren de manera segura y eficiente a través de una API hacia NiFi.

La API actúa como un puente de conexión, permitiendo la transmisión fluida de los datos; Apache NiFi toma los datos que fueron recopilados desde los dispositivos y fueron transferidos hacia la API. A su vez, los datos se procesan y se almacenan en Hadoop (ingesta en batch) y Apache HBase (ingesta en streaming). NiFi proporciona una transferencia segura y eficiente de datos entre sistemas y a su vez, facilita la integración de datos.

Hadoop, cumple el rol de almacenar los datos que se generan en ingesta batch, esta elección se da porque proporciona almacenamiento distribuido y escalable, lo cual es necesario para manejar grandes volúmenes de datos, y además permite gestionar grandes volúmenes de información estructurada y no estructurada.

Apache HBase almacena los datos generados en ingesta por streaming, dentro de los beneficios que posee esta herramienta, se destaca la posibilidad de acceder rápidamente a los datos a través de claves primarias, esto es muy importante ya que permite realizar efectivamente consultas y análisis en escalas de tiempo reducidas.

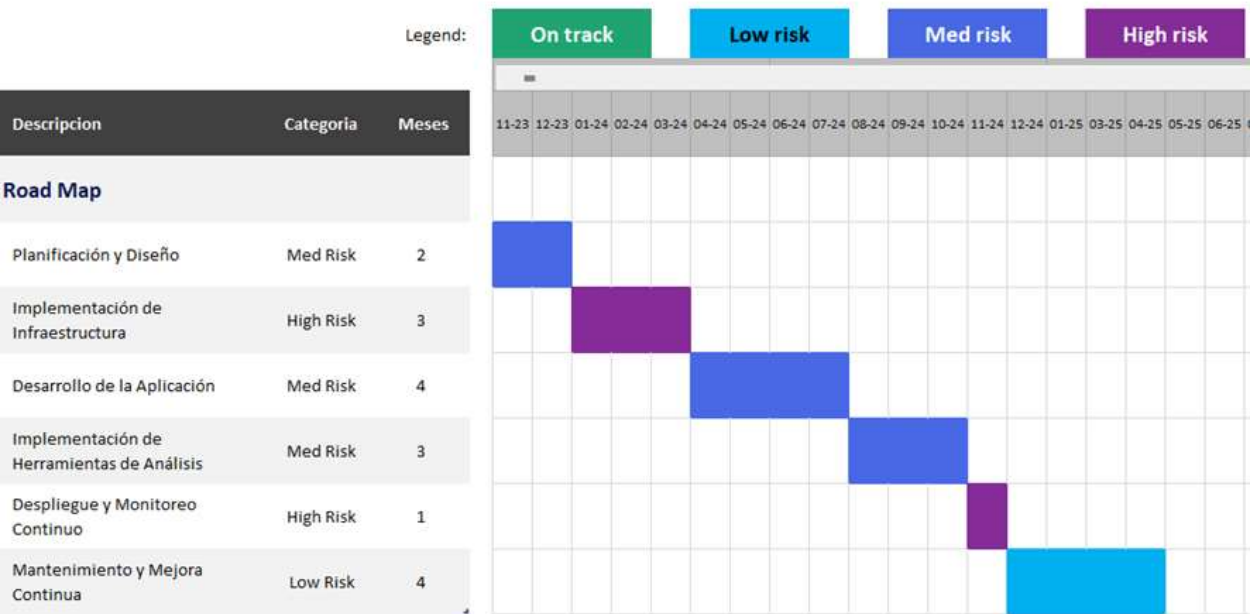
Para el procesamiento de datos en tiempo real (streaming) se utiliza Apache Storm debido a su capacidad de procesamiento rápido y eficiente. Ya que es de gran utilidad para refinar los datos de las actividades deportivas, asegurando un procesamiento eficaz para la ingesta en HBase. En cuanto a los datos almacenados en batch, se utiliza Spark para realizar el refinamiento de los mismos.

Luego, para el caso de batch se integra Hive a modo de proporcionar una interfaz gráfica de usuario que facilite la ejecución de consultas y el análisis interactivo de datos. Lo cual es necesario para que los usuarios puedan realizar consultas respecto a las actividades realizadas y generar informes personalizados. Para el caso de streaming, se integra Apache Impala, ya que ofrece buen rendimiento para consultas en tiempo real en grandes conjuntos de datos, dado que está optimizado para minimizar la latencia y responder de manera rápida a las consultas.

Por último, Superset se integra para la creación de dashboards interactivos, proporciona visualizaciones claras y personalizadas. Cumple con el requisito de generar tableros de gestión a medida del usuario que se pueden visualizar en diferentes dispositivos.

Este modelo de arquitectura se selecciona para garantizar la escalabilidad, el rendimiento y la capacidad de procesamiento necesario para cumplir con los requisitos específicos necesarios para asegurar eficiencia en el modelo de negocio. Además, la elección de herramientas como NiFi, Hadoop, HBase, Apache Storm Spark, Hive, Apache Impala y Superset se basa en su comprobada eficacia en entornos de Big Data.

## Hoja de ruta



Fase 1: Planificación y Diseño. Tiene una duración de 2 meses. El objetivo es analizar los requisitos del proyecto con el equipo de desarrollo y los interesados, definir la arquitectura del sistema, seleccionando tecnologías y herramientas específicas necesarias, realizar la planificación del proyecto para establecer un cronograma detallado para el desarrollo, teniendo en cuenta hitos importantes y asignación de tareas con fechas de inicio y finalización. Y por último, configurar los entornos de desarrollo, prueba y producción, incluyendo bases de datos, sistemas de almacenamiento y servicios en la nube.

Fase 2: Implementación de Infraestructura, la cual conlleva 3 meses y en la que se realizará el desarrollo de las bases de datos, y su correcta integración con los dispositivos que generan la información y las APIs que se encargan de transferirlas a Nifi. Además se debe realizar la correcta configuración de las conexiones que se dan entre Nifi y Hadoop (para procesamiento de datos en batch) y de Nifi y Spark (para procesamiento de datos en streaming). Finalmente se desarrollan los procesos en Spark para el refinamiento de datos y posterior almacenamiento de las nuevas bases de datos en Hadoop.

Fase 3: Desarrollo de la Aplicación. Se realiza en 4 meses. Implica el desarrollo de la Interfaz de Usuario, priorizando la autenticación y el ingreso de datos. Se implementarán las funcionalidades sociales, como conexiones entre usuarios, actividades compartidas, comentarios, likes, etc. Se realizará la integración de pagos y facturación, integrando servicios de pago, como Stripe o PayPal, para gestionar suscripciones y facturación.

Fase 4: Implementación de Herramientas de Análisis, la cual tiene una duración de 3 meses. Implica la integración de Hive y Superset para lo que será necesario configurar y desplegar Hive para consultas interactivas y Superset para la creación de dashboards. Seguido de una fase de pruebas y optimización, realizar pruebas de todas las funcionalidades, de manera de poder identificar y corregir posibles problemas de rendimiento.

Fase 5: Despliegue y Monitoreo Continuo, que tiene 1 mes de duración. Implica el despliegue a Producción, asegurando una transición sin problemas, y el monitoreo continuo para lo que es

necesario implementar diferentes herramientas para supervisar el rendimiento y la disponibilidad de la aplicación en tiempo real, evitando así caídas prolongadas del servicio.

Fase 6: Mantenimiento y Mejora Continua, la cual inicia después del despliegue. Y conlleva tareas como recolectar comentarios de usuarios y partes interesadas para identificar áreas de mejora. Realizar actualizaciones y parches de manera regular para solucionar problemas identificados y mantener seguridad en los datos. Y por último, evaluar la posibilidad de expandir características y funcionalidades en función de la retroalimentación del usuario y las tendencias del mercado.

## Diseño de dashboards

Mediante la integración con Superset podemos visualizar los datos trabajados con las distintas aplicaciones. Esto permite visualizar insights y detectar desvíos del negocio para poder tomar acciones de mejora. Por ejemplo, se puede visualizar la cantidad de usuarios nuevos que se registran por mes, las actividades preferidas por los usuarios y los ingresos generados por cada plan diferente. (Ver anexo 2)

## Conclusión

En este informe, se generó un proyecto de rediseño para la aplicación de Data Fitness, con el objetivo principal de modernizar la plataforma y proporcionar una experiencia óptima a sus usuarios. A lo largo del proceso, se establecen objetivos clave que abarcan desde la mejora del rendimiento y la capacidad de procesamiento hasta la expansión de la base de usuarios y la fidelización a través de experiencias personalizadas.

Los requerimientos funcionales que se detallan en el informe aportan un marco sólido para el desarrollo de la aplicación, ya que aseguran una gestión efectiva de usuarios, una integración perfecta con dispositivos deportivos, acceso multiplataforma y capacidades avanzadas de consulta y análisis.

Se han explorado las fuentes de datos, desde dispositivos deportivos hasta interacciones sociales entre usuarios, destacando la importancia de cada una para la toma de decisiones estratégicas y el desarrollo de una plataforma dinámica y adaptable.

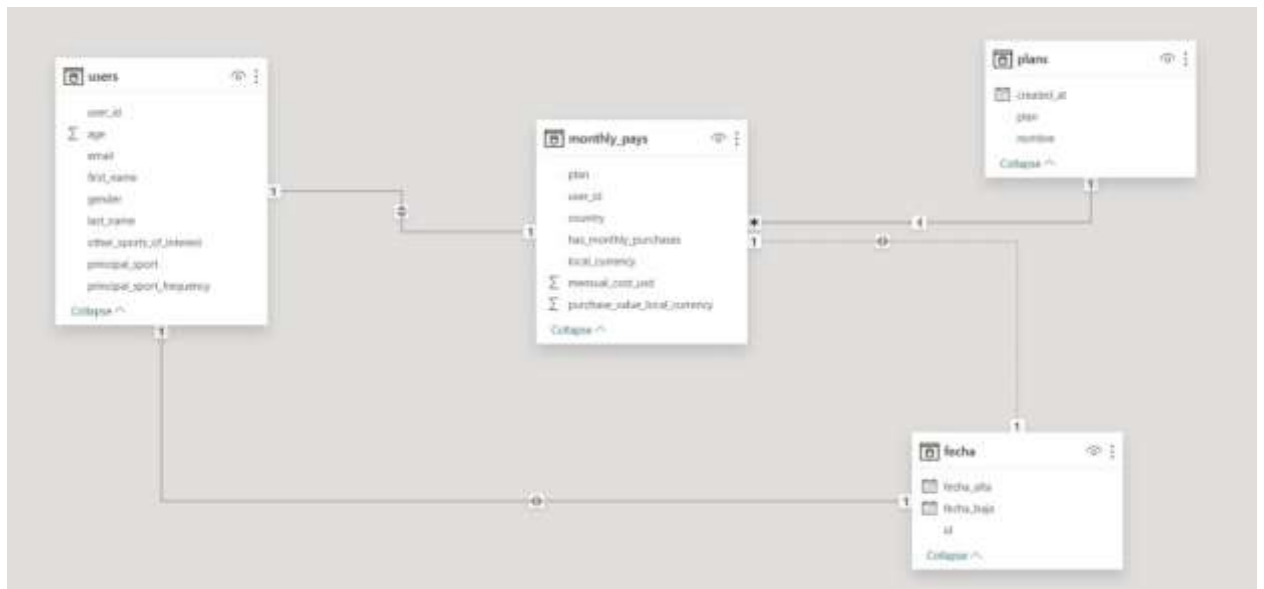
El modelo de arquitectura que se propone, engloba tecnologías como NiFi, Hadoop, Spark, Hive y Superset, y proporciona una infraestructura sólida para el manejo eficiente de grandes volúmenes de datos. Esta elección se basa en la comprobada eficacia de estas herramientas en entornos de Big Data.

La hoja de ruta establecida, desde la planificación y diseño hasta el mantenimiento y mejora continua, brinda una estructura clara para el desarrollo del proyecto, garantiza una implementación eficiente y un monitoreo continuo del rendimiento de la aplicación.

Este proceso de rediseño no solo busca mantener a Data Fitness competitiva en el mercado de aplicaciones deportivas, sino que también aspira a elevar la experiencia de usuario, fomentar la expansión y retención de usuarios, y establecer alianzas estratégicas para un crecimiento sostenible y a largo plazo.

## Anexo 1

Bases de datos y relación entre las mismas:

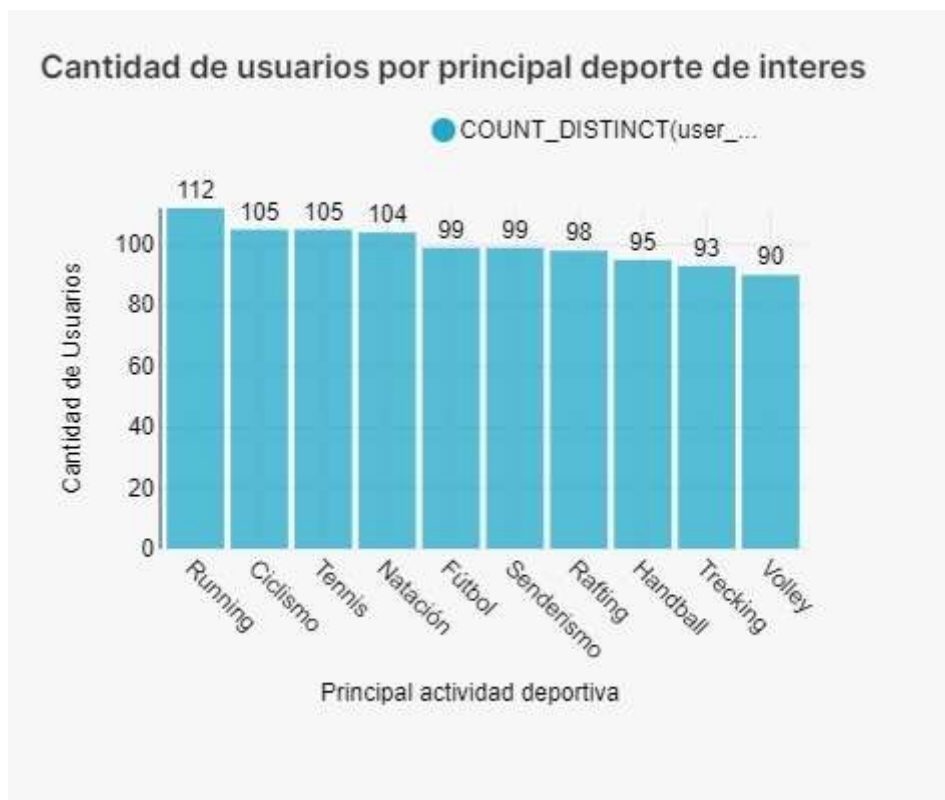


## Anexo 2

Diseño de dashboards



Al ver como se distribuyen los ingresos por plan en el último mes, se destaca que la distribución de los usuarios entre los distintos planes no tiene una gran variación.



Al observar el gráfico de barras que representa la distribución de usuarios entre las principales actividades deportivas, se puede notar que no hay una variación significativa, aunque los deportes más practicados son Running, Ciclismo y Tenis.



Este dashboard muestra un crecimiento marginalmente decreciente, ya que es notoria una tendencia a la baja en cuanto a la adquisición de nuevos usuarios por mes. Esto demuestra una maduración en cuanto a la etapa del negocio, en el que debemos retener los usuarios y realizar nuevas estrategias para incrementar la obtención de nuevos clientes.