

Casos de Estudio Gestión y Arquitectura de Datos

Gestión y Arquitectura de Datos, Universidad de San Andrés

Instrucciones Generales

Los siguientes casos de estudio han sido diseñados para aplicar los conceptos fundamentales de arquitectura y gestión de datos en contextos empresariales reales. Cada caso presenta una situación específica que requiere el diseño de una solución integral de datos.

Metodología de Trabajo

Los estudiantes trabajarán en grupos de 2 a 5 integrantes. Cada grupo será asignado a uno de los casos presentados en este documento. Es fundamental que todos los miembros del grupo participen activamente en el análisis y desarrollo de la solución propuesta. El trabajo debe demostrar comprensión profunda de los conceptos teóricos aplicados a la resolución de problemas prácticos. Se espera que los estudiantes realicen investigación independiente sobre las tecnologías disponibles en el mercado y justifiquen sus decisiones técnicas con base en criterios objetivos.

Investigación de Tecnologías

Una parte fundamental del trabajo consiste en la investigación de soluciones tecnológicas empresariales disponibles en el mercado actual. Los estudiantes deben identificar y comparar al menos tres alternativas diferentes para cada componente principal de su arquitectura propuesta. La investigación debe incluir productos de diferentes proveedores, considerando tanto soluciones comerciales como de código abierto. Es importante evaluar aspectos como funcionalidad, escalabilidad, costo, soporte técnico, y compatibilidad con otros sistemas. Los estudiantes deben consultar documentación oficial, casos de estudio publicados, comparativas técnicas, y cuando sea posible, contactar a representantes comerciales para obtener información detallada sobre las soluciones evaluadas.

Entregables Requeridos

Todos los grupos deben entregar los mismos componentes, independientemente del caso asignado:

- **Arquitectura de Datos:** Diseño conceptual completo de la arquitectura de datos que incluya la justificación detallada de las tecnologías de bases de datos seleccionadas, especificando productos comerciales concretos y sus características técnicas. Se debe proporcionar un análisis comparativo de al menos tres alternativas tecnológicas diferentes para cada componente principal. La propuesta debe incluir diagramas de arquitectura claros y la estrategia de distribución de datos cuando corresponda.
- **Modelo de Datos:** Desarrollo de un modelo de datos detallado que incluya el diseño del diagrama entidad-relación para los componentes que requieran bases de datos relacionales, la especificación de esquemas para bases de datos NoSQL cuando corresponda, y la definición de estructuras de datos para tipos de información específicos del dominio. Se debe incluir la normalización apropiada y la justificación de decisiones de desnormalización cuando sea necesaria.
- **Análisis de Requerimientos:** Documento que demuestre el análisis profundo de los requerimientos del caso, identificando claramente los diferentes tipos de datos a manejar, los patrones de acceso esperados, los volúmenes estimados, y las restricciones técnicas y regulatorias que aplican.
- **Plan de Implementación:** Elaboración de un plan detallado de implementación que contemple las fases de desarrollo del sistema, consideraciones de seguridad y privacidad, estrategias de migración de datos existentes cuando corresponda, y metodologías de validación y testing del sistema propuesto.

Criterios de Evaluación

La evaluación se basará en los siguientes criterios: profundidad del análisis técnico, justificación de decisiones arquitectónicas, calidad de la investigación de tecnologías, viabilidad de la solución propuesta, consideración de aspectos

de seguridad y cumplimiento regulatorio, claridad en la presentación de diagramas y documentación, y capacidad de defenderse técnicamente durante la presentación oral. Se valorará especialmente la capacidad de los estudiantes para relacionar los conceptos teóricos vistos en clase con las decisiones prácticas tomadas en el diseño de la solución.

1. Adven4All: Plataforma de Turismo Inteligente

1.1. Contexto

Adven4All es una startup innovadora con sede en Buenos Aires que ha revolucionado el mercado de turismo y ocio mediante el uso de inteligencia artificial y realidad aumentada para ofrecer experiencias y aventuras personalizadas. La plataforma permite a los usuarios explorar y reservar actividades únicas, desde senderismo y paracaidismo hasta inmersiones culturales y tours gastronómicos, todos adaptados a sus preferencias personales. Tras recibir una inversión de 10 millones de dólares para expandir su operación a Brasil y España, la empresa se encuentra en un momento crítico para revisar y potenciar su arquitectura de datos.

1.2. La Plataforma

Los usuarios de Adven4All interactúan con un sistema de recomendaciones que analiza sus preferencias, historial de navegación, búsquedas, reservas y reseñas para sugerir actividades. La plataforma se integra con redes sociales para enriquecer el perfil del usuario. Una funcionalidad clave es la previsualización de destinos y actividades mediante realidad aumentada, que superpone información contextual, guías virtuales y contenido de otros usuarios en tiempo real. El núcleo operativo es la gestión de reservas, que maneja la selección de experiencias, el procesamiento de pagos, la gestión de reembolsos y la creación de itinerarios personalizados, manteniendo un historial detallado por cliente. Además, la plataforma fomenta una comunidad donde los usuarios comparten reseñas, fotos, videos y participan en foros, con un sistema de gamificación que premia la participación.

1.3. Requerimientos y Desafíos de Datos

La arquitectura debe manejar una gran diversidad de datos. Por el lado estructurado, se encuentra la información de perfiles de usuario, el catálogo de actividades con sus metadatos, los registros de transacciones financieras y los datos de precios y disponibilidad que cambian constantemente. Por el lado no estructurado, se debe gestionar un gran volumen de fotos y videos, textos de reseñas, datos de geolocalización, modelos 3D para realidad aumentada y

logs detallados de interacción del usuario. Se requieren capacidades analíticas para alimentar el motor de recomendaciones, generar reportes de rendimiento operativo, analizar el comportamiento del usuario para optimizar la plataforma y predecir la demanda de actividades.

1.4. Desafíos Regulatorios y de Sociedad

La expansión internacional a Brasil y España impone el desafío de cumplir con múltiples marcos regulatorios de privacidad de datos, como GDPR en España y LGPD en Brasil, además de la ley argentina. La gestión de datos personales y transacciones financieras debe adherirse a las normativas locales de protección al consumidor y bancarias. Es fundamental implementar mecanismos de auditoría y trazabilidad para demostrar el cumplimiento, incluyendo el rastreo del procesamiento de datos personales y la gestión del consentimiento del usuario. Éticamente, la empresa debe considerar el impacto del turismo que promueve en las comunidades locales y la transparencia de sus algoritmos de recomendación para no generar sesgos indeseados.

2. StreamFlix: Plataforma de Streaming de Video

2.1. Contexto

StreamFlix es una plataforma de streaming de video emergente que busca competir con gigantes del mercado a través de una propuesta de valor basada en contenido personalizado por IA, streaming de alta calidad y una experiencia social integrada. Con 2 millones de usuarios y un crecimiento exponencial, la empresa necesita una arquitectura de datos robusta para soportar millones de usuarios concurrentes, análisis de comportamiento, gestión de contenido masivo y funcionalidades sociales.

2.2. La Plataforma

Los usuarios interactúan con un sistema de streaming de video que adapta la calidad de transmisión (de 480p a 4K) al ancho de banda, ofrece subtítulos y permite descargas para visualización offline. El corazón de la experiencia es el motor de recomendaciones, que analiza patrones de visualización, aplica algoritmos de filtrado y considera el contexto (hora, dispositivo, ubicación) para personalizar el contenido sugerido y crear listas dinámicas de tendencias. Las funcionalidades sociales, como "watch parties" sincronizadas con chat en tiempo real, perfiles públicos y sistemas de reseñas, son un diferenciador clave. La gestión de contenido abarca un catálogo detallado de películas y series con metadatos ricos, así como contenido generado por los propios usuarios.

2.3. Requerimientos y Desafíos de Datos

El sistema debe manejar un volumen masivo de datos, incluyendo 50 TB de video nuevo mensualmente, 100 millones de eventos de reproducción diarios y 10 millones de interacciones sociales. Los requerimientos de rendimiento son muy exigentes: la latencia de inicio de reproducción debe ser menor a 2 segundos y las recomendaciones deben generarse en menos de 100 milisegundos. La arquitectura debe ser altamente escalable para soportar picos de tráfico durante estrenos y la expansión internacional. Es necesario gestionar datos transaccionales (suscripciones, pagos), de comportamiento (eventos de reproducción, interacciones) y contenido multimedia masivo (videos, imágenes, subtítulos).

2.4. Desafíos Regulatorios y de Sociedad

La plataforma debe navegar el complejo panorama de los derechos de autor y las licencias de contenido, que varían significativamente por región. Es crucial implementar un sistema de gestión de derechos digitales (DRM) robusto para proteger el contenido. La gestión de datos de usuarios, especialmente sus hábitos de visualización, cae bajo estrictas leyes de privacidad de datos que requieren consentimiento explícito y transparencia. Además, la empresa enfrenta un desafío social relacionado con sus algoritmos de recomendación: debe considerar su responsabilidad en la creación de "burbujas de filtro" decidir activamente si promueve la diversidad de contenidos y voces o si refuerza las tendencias populares.

3. EcoTrack: Sistema IoT de Monitoreo Ambiental

3.1. Contexto

EcoTrack es una startup argentina que desarrolla redes de sensores inteligentes para el monitoreo ambiental en tiempo real. Sus clientes incluyen gobiernos y corporaciones que utilizan los datos para la toma de decisiones sobre sostenibilidad. Tras ganar contratos importantes, la empresa necesita escalar su plataforma para manejar 50,000 sensores en 10 ciudades, procesando 500 millones de mediciones diarias y generando análisis predictivos.

3.2. La Plataforma

La plataforma se centra en una red de sensores IoT que miden calidad del aire (PM2.5, CO2), condiciones meteorológicas, ruido ambiental, calidad del agua (pH, turbidez) y tráfico vehicular mediante cámaras con IA. Utiliza múltiples tecnologías de conectividad como LoRaWAN, 4G/5G y satelital, con procesamiento en el borde (edge computing) para análisis inicial. Los usuarios de la plataforma son diversos: autoridades gubernamentales que monitorean el cumplimiento de normativas a través de dashboards; empresas que controlan sus emisiones y generan reportes de sostenibilidad; e investigadores que acceden a datos históricos para estudios científicos, utilizando herramientas de visualización y exportación de datos.

3.3. Requerimientos y Desafíos de Datos

El principal desafío es el volumen y la velocidad de los datos, con aproximadamente 6,000 mediciones por segundo en momentos pico y una latencia para alertas críticas que debe ser inferior a 30 segundos. Se requiere la retención de datos por 10 años para análisis de tendencias. La calidad de los datos es un problema constante debido a la heterogeneidad de los sensores, fallos de conectividad y la necesidad de limpieza y normalización automática. El sistema debe soportar análisis complejos, como la correlación de múltiples variables, modelos de dispersión de contaminantes y análisis geoespacial avanzado.

3.4. Desafíos Regulatorios y de Sociedad

La plataforma debe establecer una gobernanza de datos clara, definiendo la propiedad de los datos generados, especialmente cuando los sensores están en propiedades privadas. La seguridad de la red de sensores es crítica, ya que representa una infraestructura sensible que podría ser objeto de ataques. A nivel social, existe una tensión entre los beneficios de la transparencia de los datos ambientales abiertos y las preocupaciones por la privacidad, por ejemplo, con el uso de cámaras para el monitoreo de tráfico. La empresa tiene la responsabilidad de garantizar que los datos no se utilicen de forma que puedan generar discriminación o vigilancia indebida.

4. FinTechPay: Plataforma de Pagos Digitales

4.1. Contexto

FinTechPay es una fintech argentina que ofrece una plataforma integral de pagos digitales para el mercado latinoamericano, incluyendo billeteras virtuales, procesamiento de pagos y transferencias internacionales. Con licencia del Banco Central y en plena expansión a Brasil, la empresa procesa 10 millones de transacciones mensuales y experimenta un crecimiento que exige una arquitectura de datos altamente escalable y segura.

4.2. La Plataforma

Los usuarios interactúan con la plataforma a través de una billetera digital que permite cargar dinero, pagar con QR, realizar transferencias P2P y pagar servicios. Los comercios utilizan un gateway de pagos para procesar transacciones con tarjetas, ofrecer cuotas e integrarse con sus plataformas de e-commerce a través de APIs. La plataforma también facilita transferencias internacionales con tracking en tiempo real y ofrece servicios financieros embebidos, como préstamos express basados en scoring automático y adelanto de cobros para comercios.

4.3. Requerimientos y Desafíos de Datos

Los datos transaccionales deben manejarse con propiedades ACID estrictas para garantizar la consistencia de pagos y balances de cuentas, con logs de auditoría inmutables. El sistema debe almacenar y procesar de forma segura datos sensibles de clientes, incluyendo información personal, documentos de identidad verificados y datos biométricos. También gestiona datos de comercios, como su información fiscal, configuraciones de comisiones y análisis de ventas. La alta disponibilidad del 99.99

4.4. Desafíos Regulatorios y de Sociedad

La plataforma opera en un entorno fuertemente regulado, debiendo cumplir con las normativas del Banco Central, la Ley de Entidades Financieras, las regulaciones de prevención de lavado de dinero (AML/KYC) y la certificación

PCI DSS para el manejo de tarjetas. La seguridad es primordial, requiriendo cifrado punta a punta y autenticación multifactor. Socialmente, la empresa juega un papel en la inclusión financiera, pero también enfrenta desafíos éticos relacionados con sus modelos de scoring de crédito, que deben ser justos y no discriminatorios. La responsabilidad de proteger los datos financieros y personales de millones de usuarios es inmensa.

5. HealthConnect: Plataforma de Salud Digital Integrada

5.1. Contexto

HealthConnect es una plataforma de salud digital que busca unificar el ecosistema de salud argentino, conectando hospitales, clínicas, laboratorios y farmacias. Con el respaldo del Ministerio de Salud, su objetivo es proporcionar a los pacientes acceso a una historia clínica unificada y servicios de salud digitales. Actualmente, maneja 5 millones de historias clínicas y procesa 100,000 consultas mensuales.

5.2. La Plataforma

El eje de la plataforma es la historia clínica unificada, que consolida registros médicos de múltiples instituciones. Los pacientes pueden agendar turnos, recibir resultados de estudios y gestionar tratamientos. Los médicos y profesionales de la salud utilizan la plataforma para acceder al historial del paciente, realizar consultas por telemedicina con video y chat seguro, emitir recetas electrónicas y monitorear a pacientes crónicos a través de dispositivos wearables. La plataforma se integra con sistemas de laboratorios (LIS) y radiología (RIS) para la entrega automática de resultados e incluye un visualizador de imágenes DICOM.

5.3. Requerimientos y Desafíos de Datos

La interoperabilidad es un desafío central, requiriendo la implementación de estándares como HL7 FHIR para integrar sistemas heterogéneos y mapear códigos médicos como CIE-10 y SNOMED. El sistema debe manejar una gran variedad de datos: datos clínicos estructurados (diagnósticos, medicamentos), datos no estructurados (notas médicas, informes), imágenes médicas (DICOM) y datos administrativos (turnos, facturación). La seguridad es crítica, con requerimientos de cifrado de datos en reposo y en tránsito, firma digital de documentos y logs de acceso inmutables.

5.4. Desafíos Regulatorios y de Sociedad

La plataforma está sujeta a la Ley de Protección de Datos Personales (25.326) y a la ley de secreto médico, lo que exige una gestión rigurosa del consentimiento informado para cada uso de los datos. Debe garantizar el derecho al olvido, con las limitaciones que aplican a los datos médicos, y aplicar una anonimización robusta para el uso de datos en investigación. A nivel social, la plataforma enfrenta el desafío de garantizar la equidad en el acceso a la salud digital y evitar la ampliación de la brecha digital. La centralización de la información de salud de una gran parte de la población conlleva una enorme responsabilidad ética y social sobre su custodia y uso correcto.

6. SmartCity Buenos Aires: Plataforma de Ciudad Inteligente

6.1. Contexto

SmartCity Buenos Aires es una iniciativa del Gobierno de la Ciudad para integrar sus sistemas y optimizar los servicios públicos. El proyecto busca consolidar datos de 15 secretarías, 3,000 sensores IoT, 200,000 cámaras y gestionar millones de transacciones de transporte y reclamos ciudadanos para mejorar la calidad de vida de 3 millones de habitantes.

6.2. La Plataforma

La plataforma se compone de varios módulos. El de movilidad y transporte monitorea el tráfico, optimiza semáforos, gestiona el transporte público y el estacionamiento inteligente. El módulo de seguridad ciudadana utiliza video-vigilancia con análisis de video, patrullaje predictivo y alertas tempranas. El de servicios públicos gestiona la recolección de residuos, el alumbrado, la calidad del agua y la respuesta a emergencias. Finalmente, un módulo de participación ciudadana centraliza los reclamos del 147, gestiona el presupuesto participativo y promueve la transparencia a través de datos abiertos y una app móvil.

6.3. Requerimientos y Desafíos de Datos

La plataforma debe integrar sistemas legacy con tecnologías y protocolos heterogéneos. El volumen de datos es masivo, con 10 TB de información nueva generada diariamente, incluyendo datos en tiempo real de sensores, datos históricos de gestión, datos geoespaciales complejos (mapas, imágenes satelitales, modelos 3D) y datos de ciudadanos (reclamos, encuestas). El sistema debe ser capaz de realizar análisis geoespaciales complejos y correlaciones espacio-temporales para la planificación urbana y la gestión de servicios.

6.4. Desafíos Regulatorios y de Sociedad

La iniciativa debe establecer un marco de gobernanza de datos públicos claro. El uso de tecnologías como el reconocimiento facial en el espacio público plantea importantes debates sobre privacidad vs. seguridad, que deben ser

abordados con una regulación específica y transparente. Existe el riesgo de que la plataforma amplíe la brecha digital, dejando atrás a los ciudadanos con menor acceso o habilidades tecnológicas. Además, la centralización de los datos operativos y personales de toda una ciudad crea un objetivo de alto valor para ciberataques, lo que exige los más altos estándares de ciberseguridad y resiliencia.

7. AgriTech Solutions: Plataforma de Agricultura de Precisión

7.1. Contexto

AgriTech Solutions es una empresa tecnológica que ofrece una plataforma de agricultura de precisión a grandes explotaciones agrícolas en Argentina. Su objetivo es optimizar el rendimiento de los cultivos y promover la sostenibilidad mediante el uso intensivo de datos. Con una cobertura de 2 millones de hectáreas, la empresa necesita una arquitectura de datos robusta para manejar la información de sensores, drones e imágenes satelitales.

7.2. La Plataforma

Los agricultores y agrónomos utilizan la plataforma para tomar decisiones. Un módulo de monitoreo de cultivos procesa imágenes satelitales y de drones para generar mapas de vigor (NDVI) que identifican áreas con estrés hídrico o plagas. Otro módulo se integra con la telemetría de la maquinaria agrícola (tractores, cosechadoras) para registrar datos de ubicación y operación, permitiendo crear prescripciones para la aplicación variable de insumos. Un tercer módulo de riego inteligente utiliza datos de sensores de humedad del suelo y pronósticos meteorológicos para automatizar y optimizar el uso del agua.

7.3. Requerimientos y Desafíos de Datos

El sistema debe manejar un gran volumen y variedad de datos: imágenes satelitales multiespectrales de gran tamaño, nubes de puntos 3D de drones, series temporales de sensores de suelo y estaciones meteorológicas, y datos de telemetría de maquinaria. La integración y alineación espacio-temporal de estas fuentes de datos heterogéneas es un desafío clave para poder correlacionar, por ejemplo, el rendimiento de la cosecha con la aplicación de fertilizantes. El procesamiento de modelos agronómicos y las consultas geo-espaciales requieren una capacidad de cómputo intensiva.

7.4. Desafíos Regulatorios y de Sociedad

La plataforma gestiona datos que son considerados un activo valioso para los productores. La propiedad y la portabilidad de los datos agrícolas son temas de debate en la industria, y la plataforma debe tener políticas claras al respecto. El uso de datos para optimizar la aplicación de agroquímicos tiene un impacto ambiental positivo, pero también conlleva la responsabilidad de garantizar que las recomendaciones cumplan con las regulaciones ambientales locales. Además, se debe considerar la seguridad de la plataforma para evitar que la manipulación de datos pueda llevar a decisiones agronómicas erróneas con un impacto económico significativo para el productor.

8. RetailLink: Plataforma de Cadena de Suministro

8.1. Contexto

RetailLink gestiona la cadena de suministro de una gran cadena de supermercados en Argentina, conectando 500 tiendas, 20 centros de distribución y más de 3,000 proveedores. El objetivo es optimizar el inventario y reducir costos logísticos. El desafío es modernizar su arquitectura para pasar de procesos batch nocturnos a una operación en tiempo real.

8.2. La Plataforma

Los gerentes de tienda, planificadores de demanda y personal de logística son los usuarios principales. El sistema de gestión de inventario se actualiza en tiempo real con cada venta registrada en los puntos de venta (POS) de las tiendas y genera órdenes de reposición automáticas. Un módulo de logística planifica las rutas de la flota de camiones, optimizando la carga y proporcionando seguimiento por GPS. Los proveedores interactúan a través de un portal de colaboración donde reciben órdenes de compra y confirman despachos, con integración automatizada vía EDI o APIs.

8.3. Requerimientos y Desafíos de Datos

La consistencia de los datos de inventario es crítica y requiere transacciones ACID para evitar quiebres de stock o sobreventas. La plataforma debe procesar un alto volumen de transacciones concurrentes con baja latencia, especialmente en picos de venta. La integración con la red heterogénea de proveedores, cada uno con sus propios sistemas y formatos de datos, exige una capa de integración de datos muy flexible. El modelo de datos debe ser lo suficientemente rico para capturar las complejas relaciones entre productos, proveedores, órdenes, envíos y facturas.

8.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con las regulaciones de la cadena de frío para productos perecederos, garantizando y documentando que se mantengan las

temperaturas adecuadas durante el transporte y almacenamiento. La gestión de datos compartidos con miles de proveedores plantea cuestiones de confidencialidad comercial y seguridad de la información. Socialmente, la eficiencia de la plataforma puede tener un impacto en la reducción del desperdicio de alimentos, un objetivo importante. Sin embargo, la optimización extrema de la cadena de suministro también puede aumentar la presión sobre los proveedores más pequeños.

9. EduSphere: Plataforma de Aprendizaje Online (LMS)

9.1. Contexto

EduSphere es un LMS utilizado por universidades de América Latina para ofrecer educación virtual a 150,000 estudiantes. El desafío es evolucionar su arquitectura de datos para permitir una personalización del aprendizaje a gran escala, analizando las interacciones de los estudiantes para adaptar los contenidos y detectar alumnos en riesgo.

9.2. La Plataforma

Los docentes utilizan la plataforma para crear cursos, subir materiales (videos, PDF, contenido SCORM/xAPI) y diseñar evaluaciones. Los estudiantes acceden a estos materiales, participan en foros de discusión, entregan tareas, rinden exámenes y asisten a clases en vivo por videoconferencia. La plataforma registra cada interacción: tiempo en cada recurso, visualizaciones de video, intentos en cuestionarios y mensajes en foros, con el fin de construir un perfil detallado del proceso de aprendizaje de cada alumno.

9.3. Requerimientos y Desafíos de Datos

La arquitectura debe manejar una gran diversidad de datos: información estructurada de estudiantes y calificaciones, contenido multimedia masivo y un enorme volumen de datos de eventos de interacción semi-estructurados. La entrega eficiente de contenido de video requiere una solución de almacenamiento robusta integrada con una CDN. El desafío más complejo es capturar e ingerir el flujo de eventos de interacción en tiempo real y modelarlos de una forma que permita análisis complejos sobre los patrones de aprendizaje y comportamiento.

9.4. Desafíos Regulatorios y de Sociedad

La plataforma maneja datos personales y académicos de estudiantes, que están protegidos por leyes de privacidad. Es crucial gestionar el consentimiento para el uso de datos de interacción con fines de análisis. Éticamente, el uso de algoritmos para detectar estudiantes "en riesgo" debe ser transparente y evitar

sesgos que puedan perjudicar a ciertos grupos de estudiantes. Existe una responsabilidad social en garantizar que la personalización del aprendizaje no cree "burbujas educativas" que la plataforma sea accesible para estudiantes con discapacidades, cumpliendo con las normativas de accesibilidad web.

10. GameStorm: Backend para Videojuego Multijugador Masivo

10.1. Contexto

GameStorm está desarrollando "Chronicles of Aethel", un MMORPG que espera soportar a cientos de miles de jugadores concurrentes. El reto es diseñar un backend escalable, de baja latencia y robusto frente a trampas y ataques.

10.2. La Plataforma

El mundo del juego es gestionado por un "Servicio de Mundo" que sincroniza el estado de monstruos, objetos y misiones para miles de jugadores con latencia mínima. El "Servicio de Jugador" gestiona de forma persistente la información de cada personaje (nivel, inventario, etc.), asegurando que cada cambio se guarde de forma rápida y consistente. La experiencia del jugador se complementa con servicios de chat en tiempo real, matchmaking para formar grupos, una tienda virtual para microtransacciones y un sistema anti-trampas que analiza patrones de comportamiento.

10.3. Requerimientos y Desafíos de Datos

La latencia es crítica. La base de datos del "Servicio de Mundo" debe manejar una tasa de lecturas y escrituras extremadamente alta con respuestas de milisegundos, lo que sugiere el uso de soluciones in-memory. La consistencia de los datos del jugador es fundamental; las transacciones deben ser atómicas y duraderas para no perder progreso. La arquitectura debe ser políglota para manejar diferentes tipos de datos: datos de personaje estructurados y consistentes, logs de chat y eventos de juego semi-estructurados de alto volumen (ideales para bases de datos documentales o de series de tiempo), y la red social de amigos y clanes (ideal para una base de datos de grafos).

10.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con regulaciones sobre economías virtuales y microtransacciones, especialmente en lo que respecta a la protección del consumidor y la prevención del juego en menores. La gestión de la comunidad

de jugadores es un desafío social importante, requiriendo políticas y herramientas para combatir el comportamiento tóxico, el acoso y las trampas. La empresa tiene la responsabilidad de crear un entorno de juego justo y seguro. Además, la recopilación de datos de comportamiento de los jugadores debe ser transparente y respetar su privacidad.

11. EnergyGrid: Gestión de Red Eléctrica Inteligente

11.1. Contexto

EnergyGrid, una distribuidora eléctrica provincial, está modernizando su red a una Smart Grid mediante el despliegue de millones de medidores inteligentes y sensores. El desafío es construir una plataforma de datos para analizar la información masiva generada, con el fin de mejorar la eficiencia, detectar fallos proactivamente y ofrecer nuevos servicios.

11.2. La Plataforma

El sistema de medición inteligente (AMI) ingiere datos de consumo eléctrico de millones de medidores cada 15 minutos. Un sistema de gestión de la red (DMS) utiliza datos de sensores en transformadores y líneas para monitorear la salud de la red en tiempo real, detectando anomalías para realizar mantenimiento predictivo. Los clientes finales interactúan con la plataforma a través de un portal donde pueden visualizar su consumo en detalle, recibir recomendaciones de ahorro y adherirse a programas de tarifas dinámicas.

11.3. Requerimientos y Desafíos de Datos

El principal desafío técnico es el manejo de datos de series temporales a gran escala, procesando miles de millones de registros diarios. Esto exige una base de datos especializada en series temporales. La fiabilidad y disponibilidad de la plataforma son críticas, ya que se utiliza para operaciones en tiempo real, requiriendo una arquitectura redundante y con recuperación ante desastres. El sistema debe combinar el procesamiento de streaming para alertas en tiempo real con un data warehouse o data lake para el análisis de patrones históricos de consumo y su correlación con datos operativos de la red.

11.4. Desafíos Regulatorios y de Sociedad

La plataforma gestiona datos de consumo eléctrico, que son considerados datos personales y están protegidos por ley. La ciberseguridad es una prioridad nacional, ya que la red eléctrica es una infraestructura crítica. La implementación de tarifas inteligentes y programas de gestión de la demanda plantea

desafíos de equidad social, ya que podrían beneficiar más a los consumidores con mayor capacidad de adaptar su consumo, potencialmente perjudicando a los hogares de menores ingresos. La empresa debe asegurar que la transición a una red inteligente sea justa y no cree nuevas formas de exclusión.

12. AutoDrive: Plataforma de Datos para Vehículos Autónomos

12.1. Contexto

AutoDrive desarrolla un sistema de conducción autónoma de Nivel 4. Para entrenar sus algoritmos de IA, opera una flota de 100 vehículos de prueba que generan varios terabytes de datos de sensores por día cada uno. El desafío es construir una plataforma para gestionar este volumen masivo de información.

12.2. La Plataforma

Cada vehículo está equipado con un sistema de adquisición que captura y sincroniza datos de cámaras de alta resolución, LiDAR, RADAR, IMU y GPS, almacenándolos temporalmente a bordo. En la base, un sistema de ingesta transfiere estos datos a un data lake central. Los ingenieros y científicos de datos utilizan un catálogo de datos para buscar y seleccionar conjuntos de datos específicos para sus experimentos (ej: recorridos nocturnos con lluvia") y acceden a entornos de cómputo para entrenar modelos de machine learning a gran escala.

12.3. Requerimientos y Desafíos de Datos

El volumen de datos es el desafío más extremo, con cientos de terabytes de datos nuevos diariamente, lo que requiere una solución de almacenamiento de objetos en la nube, escalable y rentable. La plataforma debe manejar una gran variedad de datos complejos: video, nubes de puntos 3D, series temporales, etc. Un catálogo de metadatos robusto es esencial para describir los datos, incluyendo la calibración de los sensores y las condiciones de cada recorrido. La gestión del ciclo de vida de los datos, desde los datos crudos hasta las versiones anotadas y curadas, es crucial para controlar los costos y rastrear el linaje de los datos.

12.4. Desafíos Regulatorios y de Sociedad

El desarrollo de vehículos autónomos está sujeto a un marco regulatorio en evolución. La plataforma debe ser capaz de proporcionar datos y análisis para certificar la seguridad del sistema ante las autoridades. La recopilación

masiva de datos del entorno público (imágenes de calles, otros vehículos, peatones) plantea serias cuestiones de privacidad que deben ser gestionadas mediante técnicas de anonimización. A nivel social, existe un debate sobre la ética de los algoritmos de decisión en caso de accidente inevitable y sobre el impacto que los vehículos autónomos tendrán en el empleo y en el diseño de las ciudades.

13. BioGenomics: Plataforma de Investigación Bioinformática

13.1. Contexto

BioGenomics es una empresa de biotecnología que investiga nuevos tratamientos contra el cáncer analizando datos genómicos y clínicos. Utiliza secuenciadores de ADN de última generación y colabora con hospitales. El desafío es crear una plataforma unificada y segura para que sus científicos puedan realizar sus investigaciones.

13.2. La Plataforma

El corazón de la plataforma es un repositorio que almacena datos genómicos (archivos FASTQ, BAM, VCF), que son masivos en volumen. En paralelo, un registro de datos clínicos almacena de forma segura y anonimizada la información de los pacientes (diagnósticos, tratamientos). Los científicos utilizan un entorno de análisis interactivo que les da acceso controlado a ambos tipos de datos y a herramientas bioinformáticas para ejecutar análisis complejos, como estudios de asociación de genoma completo (GWAS).

13.3. Requerimientos y Desafíos de Datos

La integración de datos genómicos (masivos, semi-estructurados) y clínicos (provenientes de sistemas heterogéneos) es un desafío complejo que requiere un modelo de datos flexible y procesos de armonización. El rendimiento de los análisis bioinformáticos, que son computacionalmente muy intensivos, exige una arquitectura de almacenamiento y cómputo distribuido de alto rendimiento. Es crucial implementar un pipeline de datos que gestione de forma segura y eficiente la carga, procesamiento y análisis de esta información.

13.4. Desafíos Regulatorios y de Sociedad

El cumplimiento regulatorio es primordial. La plataforma debe adherirse estrictamente a leyes como la de Protección de Datos Personales y normativas internacionales como HIPAA, implementando controles de acceso, cifrado y auditoría. La gestión del consentimiento del paciente para el uso de sus datos

en investigación es un pilar fundamental. Éticamente, la plataforma debe asegurar que los resultados de la investigación no se utilicen para crear formas de discriminación genética. Existe una responsabilidad social de asegurar que los avances logrados contribuyan de manera equitativa al bien común.

14. MediaPulse: Plataforma de Análisis de Medios y Redes Sociales

14.1. Contexto

MediaPulse es una agencia de relaciones públicas que ofrece a sus clientes servicios de monitoreo de su presencia en medios y redes sociales. Necesita una plataforma que rastree menciones en tiempo real a través de miles de fuentes online y las analice para medir la reputación de sus clientes.

14.2. La Plataforma

La plataforma se basa en un sistema de ingesta de datos que se conecta a las APIs de redes sociales y a los feeds de sitios de noticias y blogs. Un motor de procesamiento de lenguaje natural (NLP) analiza en tiempo real cada texto ingerido para identificar menciones de clientes, extraer entidades, determinar el sentimiento y clasificarlo por tema. Los clientes finales utilizan un dashboard interactivo para visualizar un feed de menciones, gráficos de evolución de sentimiento, nubes de palabras y análisis de redes de influencia, entendiendo quién habla de ellos y con qué alcance.

14.3. Requerimientos y Desafíos de Datos

La velocidad y el volumen de la ingesta de datos son el principal desafío, requiriendo una arquitectura de streaming capaz de procesar millones de documentos por hora. El manejo de datos no estructurados (texto libre) es central, necesitando un motor de búsqueda o base de datos documental que permita una indexación y consulta eficiente. Para analizar las redes de influencia y la propagación de información, es necesario modelar las relaciones entre usuarios y temas, lo que sugiere el uso de una base de datos de grafos en una arquitectura políglota.

14.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con los términos de servicio de las APIs de las redes sociales y respetar las leyes de derechos de autor sobre el contenido de los medios de comunicación. El análisis de sentimiento automatizado plantea desafíos éticos, ya que la ironía y el sarcasmo pueden ser malinterpretados,

llevando a conclusiones erróneas sobre la opinión pública. Socialmente, la plataforma puede ser utilizada para identificar y contrarrestar campañas de desinformación, pero también podría ser usada para fines de manipulación. La empresa tiene la responsabilidad de establecer políticas de uso ético de su tecnología.

15. LegalMind: Plataforma de Investigación Legal con IA

15.1. Contexto

LegalMind es una startup que crea una plataforma de investigación para abogados en Argentina, utilizando IA para acelerar la búsqueda de jurisprudencia y el análisis de documentos. El desafío es construir un sistema que pueda procesar un cuerpo masivo de documentos legales no estructurados de forma segura y confidencial.

15.2. La Plataforma

La plataforma se centra en un repositorio de documentos legales, tanto públicos (leyes, fallos) como privados (documentos de casos de clientes). La herramienta principal para los abogados es un motor de búsqueda semántica que entiende el significado de las consultas legales más allá de las palabras clave. Los usuarios también pueden subir documentos, como un contrato, y la plataforma utiliza IA para identificar cláusulas clave, señalar riesgos o analizar un escrito judicial para sugerir jurisprudencia de apoyo.

15.3. Requerimientos y Desafíos de Datos

El desafío principal es el manejo de documentos no estructurados (texto). Se necesita una base de datos documental o un motor de búsqueda que soporte consultas de texto completo y búsqueda semántica por vectores (vector search). El procesamiento de NLP para generar los ".embeddings"(representaciones vectoriales del texto) de millones de documentos es computacionalmente intensivo y requiere un pipeline de datos robusto. La plataforma debe ser capaz de indexar estos vectores y realizar búsquedas de similitud de manera eficiente.

15.4. Desafíos Regulatorios y de Sociedad

La seguridad y la confidencialidad de los datos de los clientes son de máxima importancia, y están protegidas por el secreto profesional. La arquitectura debe garantizar un aislamiento total entre los datos de los diferentes estudios jurídicos, con cifrado y permisos granulares. Éticamente, el uso de IA

en el ámbito legal plantea preguntas sobre el sesgo en los algoritmos. Si los datos de entrenamiento (fallos históricos) reflejan sesgos sociales, la IA podría perpetuarlos. La plataforma tiene la responsabilidad social de ser una herramienta que promueva la justicia y no una que refuerce las desigualdades existentes.

16. AeroConnect: Plataforma de Operaciones y Mantenimiento Aeronáutico

16.1. Contexto

AeroConnect proporciona una plataforma de gestión de operaciones y mantenimiento para aerolíneas comerciales. Su principal cliente, una aerolínea regional, necesita integrar información de sistemas diversos y críticos para la seguridad, cumpliendo con las estrictas regulaciones de la autoridad aeronáutica.

16.2. La Plataforma

El personal de mantenimiento utiliza el módulo de "Salud de la Flota" para analizar datos de los sensores de las aeronaves (motores, hidráulica) con el fin de aplicar modelos de mantenimiento predictivo. El sistema de "Gestión de Mantenimiento"^{es} utilizado por técnicos e inspectores para planificar y registrar todas las tareas, manteniendo un historial auditable de cada componente. El personal de operaciones de vuelo usa un módulo que integra la planificación de vuelos y la asignación de tripulaciones para tener una visión unificada de la operación en tiempo real y gestionar retrasos de forma eficiente.

16.3. Requerimientos y Desafíos de Datos

La integración de sistemas heterogéneos (aviónica, sistemas legacy, etc.) es un desafío clave que requiere una arquitectura de microservicios. La plataforma debe manejar un gran volumen de datos de series temporales de alta frecuencia provenientes de los sensores de los aviones. La integridad y la auditabilidad de los datos son absolutamente críticas; cada registro de mantenimiento debe ser almacenado de forma segura e inmutable. La base de datos que soporta estos registros debe garantizar consistencia ACID y tener un sistema de versionado robusto para cumplir con las regulaciones.

16.4. Desafíos Regulatorios y de Sociedad

La plataforma está sujeta a las regulaciones de la autoridad aeronáutica (ANAC en Argentina), que dicta requisitos estrictos sobre la trazabilidad y

el registro de las tareas de mantenimiento. La seguridad de la plataforma es un asunto de seguridad nacional, ya que un fallo o un ciberataque podrían tener consecuencias catastróficas. Socialmente, la eficiencia y seguridad mejoradas por la plataforma contribuyen directamente a la seguridad de los pasajeros. Sin embargo, la creciente automatización y el mantenimiento predictivo también pueden transformar los roles y requerimientos de habilidades del personal técnico aeronáutico.

17. Construct360: Plataforma de Gestión para la Construcción

17.1. Contexto

Construct360 es una plataforma de software para la gestión integral de grandes proyectos de construcción. Busca centralizar toda la información del proyecto para mejorar la colaboración entre arquitectos, ingenieros, contratistas y clientes. El desafío es manejar tipos de información extremadamente diversos, desde modelos 3D a presupuestos, manteniendo una única "fuente de la verdad".

17.2. La Plataforma

Arquitectos e ingenieros utilizan el "Common Data Environment" (CDE) para gestionar los modelos de Building Information Modeling (BIM), visualizarlos y detectar colisiones. Los gerentes de proyecto usan el módulo de gestión para seguir el cronograma, asignar tareas y controlar el presupuesto, vinculando las tareas con los objetos del modelo BIM para una visualización 4D. Los equipos en el sitio de obra usan tablets para acceder a los últimos planos, reportar problemas (RFIs), registrar el progreso diario con fotos y realizar listas de verificación de seguridad y calidad.

17.3. Requerimientos y Desafíos de Datos

El manejo de los complejos y pesados archivos de modelos BIM es un desafío singular, que requiere la capacidad de extraer e indexar sus datos para realizar consultas. La plataforma debe gestionar una extrema variedad de datos: archivos de diseño, documentos de texto, hojas de cálculo, cronogramas, imágenes y videos. El versionado y la gestión de cambios son críticos, ya que los diseños evolucionan constantemente. El sistema debe asegurar que todos trabajen con la última versión de la información y mantener un registro auditable de cada cambio.

17.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con los códigos y normativas de construcción locales e internacionales. La gestión de contratos y órdenes de cambio a través

de la plataforma tiene implicaciones legales, por lo que la auditabilidad y la validez de los registros digitales son fundamentales. Socialmente, la plataforma puede mejorar la seguridad en las obras al facilitar la comunicación y el cumplimiento de los protocolos. Sin embargo, también puede crear una brecha digital entre las grandes empresas constructoras que pueden permitirse estas tecnologías y los contratistas más pequeños.

18. InsureRight: Plataforma de Seguros Basada en Datos

18.1. Contexto

InsureRight es una compañía de seguros en proceso de transformación digital. Busca crear una plataforma de datos centralizada para optimizar la tarificación de pólizas, la gestión de siniestros y la detección de fraudes, integrando datos de sistemas legacy con nuevas fuentes de datos de IoT.

18.2. La Plataforma

Los actuarios y científicos de datos utilizan un motor de "Dynamic Pricing" que calcula primas personalizadas usando modelos de machine learning, los cuales pueden incorporar datos de telemática sobre los hábitos de conducción. Los clientes interactúan con la plataforma a través de una app móvil para reportar siniestros, subiendo fotos y documentos. Un sistema de "Gestión Inteligente de Siniestros" utiliza IA para analizar esta información, estimar costos y acelerar el proceso. Los investigadores de fraudes usan un módulo que analiza continuamente los datos en busca de patrones sospechosos y asigna un puntaje de riesgo a cada reclamo.

18.3. Requerimientos y Desafíos de Datos

El principal desafío es la integración de datos: combinar datos estructurados de sistemas core legacy, datos no estructurados como fotos y texto, y datos de streaming de alta velocidad de dispositivos IoT. Esto requiere una arquitectura de data lake o data lakehouse. La calidad de los datos es fundamental, por lo que la plataforma debe tener un componente robusto para limpiar y estandarizar la información. Para el análisis de fraudes, es necesario modelar redes de colaboración complejas, una tarea ideal para una base de datos de grafos.

18.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con las regulaciones de la Superintendencia de Seguros de la Nación. El uso de datos personales, como los de telemática, para calcular precios, plantea importantes cuestiones de privacidad y exige un

consentimiento explícito del cliente. Éticamente, los modelos de tarificación y detección de fraude deben ser auditados para evitar sesgos discriminatorios. A nivel social, si bien la tarificación personalizada puede parecer más justa, también podría llevar a que personas con perfiles de mayor riesgo (a menudo correlacionados con factores socioeconómicos) no puedan acceder a un seguro asequible.

19. MusicVerse: Plataforma de Streaming y Gestión de Regalías

19.1. Contexto

MusicVerse es una startup que lanza una plataforma de streaming de música con un modelo de regalías transparente y justo para los artistas como diferenciador. El desafío es diseñar una arquitectura que soporte un streaming de alta calidad y, a la vez, gestione el complejo cálculo y distribución de regalías.

19.2. La Plataforma

Los usuarios finales utilizan la plataforma para escuchar música, recibiendo recomendaciones personalizadas. El contenido de audio se distribuye a través de una CDN para garantizar baja latencia. El "Sistema de Ingesta y Procesamiento de Reproducciones" es el núcleo para los artistas y sellos discográficos, ya que captura un evento detallado por cada reproducción. Un "Motor de Cálculo de Regalías" procesa estos miles de millones de eventos mensualmente, aplicando las reglas de contratos complejos para calcular con precisión los micropagos correspondientes a cada titular de derechos.

19.3. Requerimientos y Desafíos de Datos

La arquitectura debe soportar dos cargas de trabajo muy diferentes: una analítica masiva para procesar los eventos de reproducción (ideal para un data lake/warehouse) y una transaccional para gestionar el catálogo y los contratos (requiere una base de datos relacional ACID). El volumen de eventos de reproducción es enorme y su ingesta debe ser en tiempo real y sin pérdidas, ya que cada evento tiene un valor monetario. El modelo de datos para los derechos musicales es extremadamente complejo, con una sola canción pudiendo tener decenas de titulares de derechos con condiciones variables, lo que exige un esquema relacional muy preciso.

19.4. Desafíos Regulatorios y de Sociedad

La plataforma debe navegar el complejo mundo de las leyes de derechos de autor internacionales y los acuerdos con las sociedades de gestión colectiva.

La promesa de transparencia en las regalías es un desafío en sí mismo, requiriendo que los reportes generados sean claros y auditables por los artistas. Socialmente, la plataforma tiene la oportunidad de empoderar a artistas independientes, pero también enfrenta la presión de un mercado dominado por grandes sellos. La forma en que su algoritmo de recomendación promueva a artistas emergentes versus a los ya establecidos tendrá un impacto cultural significativo.

20. PharmaChain: Trazabilidad de la Cadena de Suministro Farmacéutica

20.1. Contexto

PharmaChain es un consorcio de laboratorios y empresas de logística que construye una plataforma de trazabilidad de medicamentos en Argentina. Su objetivo es combatir la falsificación y garantizar la integridad de los productos, especialmente los sensibles a la temperatura, utilizando blockchain e IoT.

20.2. La Plataforma

El sistema es utilizado por personal de laboratorios, logística y farmacias. Cada caja de medicamento tiene un código único serializado. El "Sistema de Trazabilidad" registra cada evento en la vida de la caja (producción, transporte, almacenamiento) en una red de blockchain, creando un registro inmutable. Para productos sensibles, sensores IoT de un solo uso monitorean la temperatura y la humedad durante el transporte, transmitiendo sus datos a la plataforma. Un "Sistema de Alertas" monitorea los datos de la blockchain y los sensores para detectar anomalías y generar alertas inmediatas.

20.3. Requerimientos y Desafíos de Datos

El principal desafío es la integración de la tecnología blockchain, decidiendo qué datos almacenar on-chain (inmutables pero lentos y caros) y cuáles off-chain. La plataforma debe ingerir y procesar datos de streaming de millones de sensores IoT, lo que requiere una arquitectura escalable para series temporales. Un reto clave es la capacidad de correlacionar los datos transaccionales y discretos de la blockchain con los datos continuos de los sensores IoT y los datos maestros de productos y participantes de la cadena.

20.4. Desafíos Regulatorios y de Sociedad

La plataforma debe cumplir con las regulaciones de trazabilidad de medicamentos de la ANMAT. La validez legal de los registros en blockchain como prueba en disputas comerciales o regulatorias es un campo en evolución. La

colaboración entre competidores (los laboratorios del consorcio) en una plataforma compartida plantea desafíos de gobernanza y confidencialidad de datos comerciales. Socialmente, la plataforma tiene un impacto directo en la salud pública al reducir el riesgo de medicamentos falsificados o en mal estado. Sin embargo, el costo de la tecnología podría ser una barrera para los laboratorios más pequeños.