

Análisis de Caso – Arquitectura de Datos

1. Diagnóstico de la situación actual

La empresa InfoHealth presenta múltiples desafíos asociados a la ausencia de una arquitectura de datos formal:

- Fragmentación y duplicación de datos por múltiples fuentes no integradas (HCE, sensores, formularios, hojas de cálculo).
- Falta de trazabilidad sobre el origen y la calidad de los datos.
- Bajos niveles de confianza en los reportes generados.
- Procesos manuales y lentos para análisis y preparación de datos.
- Problemas de seguridad y privacidad, especialmente en información sensible como historiales clínicos.
- Ausencia de gobernanza de datos clara y procesos de control.

2. Propuesta de arquitectura de datos

Capa	Descripción propuesta	Ejemplos de tecnologías
Fuentes de datos	Historias clínicas electrónicas, sensores IoT, formularios online, planillas y correos institucionales.	HL7/FHIR, formularios Google Forms, dispositivos IoT.
Almacenamiento	Almacenamiento híbrido para datos estructurados y no estructurados.	Amazon S3, Amazon RDS (PostgreSQL), MongoDB.
Procesamiento y limpieza	ETL para depuración y transformación de datos. Limpieza de duplicados y estandarización de formatos.	AWS Glue, Apache NiFi, Talend.
Acceso y visualización	Dashboards, reportes y consultas según perfiles (médicos, analistas, directivos).	Power BI, Amazon QuickSight, Tableau, Looker Studio.
Seguridad y gobernanza	Control de accesos, encriptación, catalogación de datos y políticas de privacidad.	AWS Lake Formation, Apache Atlas, IAM, GDPR compliant.

3. Herramientas propuestas (2 por capa)

Capa	Herramienta 1	Herramienta 2
Fuentes de datos	HL7/FHIR	IoT + MQTT / HTTP APIs
Almacenamiento	Amazon S3	MongoDB
Procesamiento	AWS Glue	Apache NiFi
Acceso y visualización	Power BI	Amazon QuickSight
Seguridad y gobernanza	AWS IAM	Apache Atlas

4. Principios clave de arquitectura aplicados (según DAMA-DMBOK)

1. Principio de Integridad de los Datos: Evitar duplicaciones y asegurar consistencia entre fuentes.
2. Gobierno de Datos Efectivo: Definición de roles, responsabilidades y políticas claras.
3. Seguridad y Privacidad desde el diseño: Encriptación, autenticación y trazabilidad por defecto.

5. Justificación de escalabilidad y adecuación al sector salud

- Permite crecer horizontalmente agregando nuevas fuentes como dispositivos móviles o apps médicas.
- Soporta tanto datos estructurados como no estructurados, ideal para combinar historias clínicas con datos IoT.
- Segmentación de acceso para cumplir con normativas del sector (GDPR, HIPAA).
- Mejora la eficiencia del análisis de datos clínicos en tiempo real y retroalimentación a los profesionales.

6. Esquema visual / Tabla resumen

Capa	Elemento principal	Herramienta sugerida
Fuentes de datos	HCE, sensores IoT, formularios	HL7/FHIR, HTTP APIs, MQTT (IoT)
Almacenamiento	S3 para documentos, RDS para registros	Amazon S3, PostgreSQL, MongoDB
Procesamiento	Limpieza y ETL	AWS Glue, Apache NiFi
Acceso y visualización	Paneles de consulta y reportes	Power BI, QuickSight
Seguridad y gobernanza	Accesos, catalogación, trazabilidad	AW IAM, Apache Atlas