**1. Diagnóstico de la Situación Actual**

La empresa InfoHealth enfrenta problemas significativos de gestión de datos que se derivan de la falta de una arquitectura definida. El crecimiento acelerado ha llevado a una proliferación de sistemas aislados, generando los siguientes desafíos críticos:

* **Silos de Datos:** La información se encuentra fragmentada en múltiples fuentes (historias clínicas, sensores IoT, formularios, hojas de cálculo) que no se comunican entre sí.
* **Duplicación e Inconsistencia:** Sin una fuente única de la verdad, la información de los pacientes se duplica y contiene errores en las diferentes plataformas, lo que afecta la fiabilidad de los datos.
* **Falta de Trazabilidad:** Es imposible seguir el historial completo y unificado de un paciente, combinando sus datos clínicos, administrativos y de monitoreo.
* **Riesgos de Seguridad:** La dispersión de datos sensibles en entornos compartidos y sin un control centralizado aumenta la vulnerabilidad a brechas de seguridad.
* **Ineficiencia Operativa:** Los analistas invierten tiempo valioso en depurar datos en lugar de analizarlos, los directivos desconfían de la precisión de los informes generados y los médicos no pueden acceder a la información clínica con la rapidez y exactitud que necesitan.

**2. Propuesta de Arquitectura y Herramientas**

Se propone una arquitectura de datos moderna de tipo **Lake House en la nube**, diseñada para ser escalable, segura y capaz de manejar la diversidad de datos del sector salud. Esta arquitectura se estructura en las siguientes capas:

|  |  |  |
| --- | --- | --- |
| Capa de Arquitectura | Propósito | Herramientas Sugeridas |
| **1. Fuentes de Datos** | Ingesta de todos los datos generados por la organización, tanto en tiempo real como en lotes. | **AWS Kinesis / Google Pub/Sub** (para datos de sensores IoT) **AWS Glue** (para DB, archivos y APIs) |
| **2. Almacenamiento** | Repositorio centralizado para todos los datos, separando los datos crudos de los procesados. | **Data Lake:** Amazon S3 / Google Cloud Storage **Data Warehouse:** Google BigQuery |
| **3. Procesamiento y Limpieza** | Transformación de los datos crudos en información limpia, estandarizada y lista para el análisis (Proceso ETL/ELT). | **Apache Spark** |
| **4. Acceso y Visualización** | Consumo de los datos procesados a través de herramientas de BI. | **Tableau / Microsoft Power BI / Looker** (para dashboards e informes) |
| **5. Seguridad y Gobernanza** | Capa transversal para garantizar la seguridad, privacidad y calidad de los datos | **AWS Lake Formation** (para gobernanza y catálogo) |

**3. Principios de Diseño y Justificación**

El rediseño de la arquitectura se fundamentará en tres principios clave extraídos del marco de referencia DAMA DMBOK:

1. **El Dato es un Activo Estratégico:** Este principio implica tratar la información con el mismo rigor que cualquier otro activo valioso de la empresa.
2. **Dato Compartido y Consolidado:** Este principio establece que los datos deben ser un recurso compartido entre las diferentes áreas de la organización.
3. **Gobernanza de Datos:** Se establecerá un marco de gobernanza para gestionar la calidad, el ciclo de vida y la seguridad de los datos. Esto es fundamental en el sector salud para cumplir con las normativas de privacidad.

**Justificación de la Arquitectura Propuesta**

La arquitectura de Lake House en la nube es **escalable** porque los servicios cloud como Amazon S3 permiten un crecimiento casi ilimitado del almacenamiento y del cómputo de forma independiente y bajo demanda.

Es **adecuada para el sector salud** porque:

* Maneja la **diversidad de datos** (estructurados, como historiales clínicos, y no estructurados, como datos de sensores).
* Las plataformas cloud ofrecen certificaciones de **cumplimiento y seguridad** robustas, indispensables para proteger la información sensible de los pacientes.
* Habilita **capacidades analíticas avanzadas**, como la aplicación de modelos de ML sobre los datos del Data Lake para análisis predictivo, posicionando a InfoHealth a la vanguardia tecnológica.

**4. Conclusiones**

La falta de una arquitectura de datos definida es el principal obstáculo que enfrenta InfoHealth para capitalizar el valor de su información. La implementación de una arquitectura Lake House en la nube no solo resolverá los problemas actuales de duplicación, seguridad y falta de trazabilidad, sino que también proporcionará una plataforma escalable y segura que soportará el crecimiento futuro de la organización.