

La arquitectura propuesta de la solución de inteligencia de negocios se diseñó para garantizar la integración, procesamiento y análisis eficiente de datos provenientes de fuentes heterogéneas relacionadas con la salud y el ambiente. En la capa de fuentes de datos, se integran los registros del IDEAM–SISAIRE (concentraciones de contaminantes atmosféricos), la Secretaría Distrital de Salud – RIPS (morbilidad respiratoria en población infantil) y el DANE (proyecciones poblacionales por localidad y grupo etario).

Posteriormente, en la capa ETL, se implementa un flujo de transformación híbrido: primero mediante Python (pandas) para limpiar, estandarizar y consolidar las fuentes, y luego mediante Google Cloud Data Fusion para orquestar el proceso en la nube. Los datos transformados se almacenan en BigQuery, sirviendo como repositorio analítico central en la infraestructura de Google Cloud Platform (GCP), que cumple un rol equivalente a la capa de almacenamiento en AWS en otras arquitecturas de referencia.

Finalmente, la capa de presentación se desarrolla en Power BI, donde se construyeron tres tableros de control —descriptivo, inferencial y clínico— que permiten explorar relaciones entre contaminación y morbilidad desde perspectivas estadísticas, espaciales y temporales. Esta arquitectura garantiza trazabilidad, escalabilidad y soporte analítico para la toma de decisiones en salud pública basada en datos.