

# Proyecto Data Alive: Visión Estratégica de la Plataforma

Autor: Emmanuel Eduardo Pérez Cabrera

Presencia Profesional: [datalquemy.github.io](https://datalquemy.github.io)

LinkedIn: [Emmanuel Perez](#)

Email: emmanuel.eduardo@datalquemy.com

## Recursos del Proyecto:

- Seguimiento en Tiempo Real (Trello): [ProyectoDatAlive](#)
- Repositorio de Código y Documentación (GitHub): [DatAlive](#)

## 1. Visión Ejecutiva: Una Plataforma Cognitiva, No Solo un Pipeline

*"Data Alive no es un ETL ni un dashboard. Es una plataforma cognitiva, multi-dominio y parametrizable."*

El **Proyecto Data Alive** ha evolucionado para convertirse en una **plataforma de datos agnóstica y adaptable**, diseñada para entregar a cada cliente una solución a medida sin reinventar el núcleo tecnológico. Su misión es transformar datos de cualquier industria en insights accionables, con la capacidad única de **explicar el porqué de los fenómenos** gracias a su capa cognitiva.

La arquitectura se basa en un **Núcleo (Core)** puro y agnóstico que gestiona la ingesta, transformación y orquestación, sobre el cual se implementan **Paquetes de Dominio (Domain Packs)** que contienen la lógica de negocio específica para cada industria (ej. petróleo, agro, retail).

## 2. Hoja de Ruta Estratégica: De la Fábrica al Producto

El proyecto se gestiona a través de un plan maestro que asegura una construcción robusta y escalable, enfocada en crear primero la "fábrica" y luego el primer "producto".

### Pilar I: Cimientos de la Plataforma (Fases 1-5)

Esta etapa se enfoca en construir la infraestructura base segura y gobernada en la nube (AWS y Azure). Se establecen los cimientos de seguridad, redes, cuentas y el Data Lake, preparando el terreno para el desarrollo del núcleo de la plataforma.

### Pilar II: Construcción del Núcleo Agnóstico (Fases 6-11)

Aquí se construye el **"Core Engine"** de Data Alive. Se desarrollan los componentes reutilizables para la ingesta (batch y streaming con Kafka), procesamiento (Spark), orquestación (Airflow), calidad de datos (Great Expectations) y la capa de consumo analítico, todo de forma parametrizable y sin lógica de negocio específica.

### Pilar III: Implementación del Primer Dominio (Fases 12-19)

Una vez que el núcleo está listo, se demuestra su poder implementando el primer **"Domain**

**Pack"** para la industria del **Petróleo y Gas**. En esta etapa se construyen los dashboards de producción diaria, se desarrollan los modelos de Machine Learning para pronósticos y se implementa la capa **RAG** para explicar las variaciones en la producción.

#### **Pilar IV: Operación y Escalabilidad (Fases 20-21)**

La etapa final se centra en las pruebas integrales, la validación del primer "Domain Pack" y la creación de la documentación y los manuales de operación. Esto deja la plataforma lista no solo para operar, sino para empezar a desarrollar nuevos "Domain Packs" para otras industrias.

### **3. Gestión de Proyecto y Transparencia Estratégica**

El proyecto se gestiona con una **metodología ágil (Kanban)**, demostrando una sólida capacidad de liderazgo y ejecución. La planificación estratégica se refleja en la detallada hoja de ruta, asegurando que cada paso tenga un propósito claro.

Para fomentar una total transparencia, el progreso puede ser consultado en tiempo real a través del **tablero público en Trello**, y toda la documentación y el código fuente están disponibles en el **repositorio público de GitHub**. Esta gestión abierta posiciona al proyecto como un portafolio vivo de las mejores prácticas en arquitectura de datos y gestión de productos tecnológicos.

**"From BI to AI to DA. It's not just AI... it's Data Alive."**