

Proyecto Data Alive: Visión y Hoja de Ruta Estratégica

Autor: Emmanuel Eduardo Pérez Cabrera

Presencia Profesional: [DatAlive](#)

LinkedIn: [Emmanuel Perez](#)

Email: emmanuel.eduardo@dataquemy.com

Recursos del Proyecto:

- **Seguimiento en Tiempo Real (Trello):** [DA Trello](#)
- **Repositorio de Código y Documentación (GitHub):** [Dataquemy/DA](#)

1. Visión Ejecutiva: De los Datos Inertes a la Inteligencia Viva

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

El **Proyecto Data Alive** es una iniciativa de arquitectura de datos de nueva generación diseñada para transformar la forma en que interactuamos con la información. El objetivo es evolucionar más allá de la inteligencia de negocios tradicional (BI) y la inteligencia artificial (AI) para llegar a un nuevo paradigma: **Data Alive (DA)**, un ecosistema donde los datos viven, respiran y conversan con nosotros.

La misión es construir una plataforma analítica multinube (AWS y Azure) capaz de:

- **Reportar** la operación diaria de producción con la máxima claridad.
- **Predecir** el rendimiento futuro mediante modelos de Machine Learning.
- **Explicar** en lenguaje natural las causas de cualquier variación o anomalía, transformando los dashboards de simples visualizadores a puntos de partida para el diálogo estratégico.

Este proyecto representa la construcción de un "sistema nervioso digital" que permite a los líderes empresariales no solo ver datos, sino comprender su historia y dialogar con ellos para tomar decisiones más rápidas e inteligentes.

2. Hoja de Ruta Estratégica: Un Enfoque por Fases

El proyecto está estructurado en una hoja de ruta de 21 fases, organizadas en cuatro pilares estratégicos que aseguran una construcción robusta, segura y escalable.

Pilar I: Cimientos y Gobierno (Fases 1-4)

Antes de construir, aseguramos el terreno. Esta etapa inicial se enfoca en establecer las bases de seguridad, la estructura de cuentas en la nube, la arquitectura de red y los mecanismos de control de costos. Es la fase más crítica para garantizar la integridad y el gobierno de todo el ecosistema.

Pilar II: El Ecosistema de Datos (Fases 5-12)

Donde los datos crudos se convierten en información. Este pilar cubre la construcción del Data Lake (AWS S3 y Azure ADLS), la definición de catálogos y formatos de datos, la

implementación de pipelines de ingesta (tanto en batch como en tiempo real con Kafka), el procesamiento distribuido con Spark (EMR y Databricks) y la orquestación automatizada de flujos de trabajo.

Pilar III: Inteligencia y Automatización (Fases 13-19)

Donde la información se convierte en inteligencia. Aquí es donde el proyecto cobra vida. Se desarrollan y despliegan modelos de Machine Learning (MLOps), se implementa la capa de **Generación Aumentada por Recuperación (RAG)** para permitir las explicaciones en lenguaje natural, se exponen los insights a través de APIs (FastAPI) y se integra todo con motores analíticos de alto rendimiento como Snowflake.

Pilar IV: Operación y Entrega (Fases 20-21)

Asegurando la calidad y la sostenibilidad. La etapa final se centra en las pruebas integrales de todo el sistema, la planificación del paso a producción (*cutover*), y la creación de la documentación y los manuales de operación (*runbooks*) que no solo servirán de guía, sino que también alimentarán al propio sistema RAG para hacerlo aún más inteligente.

3. Gestión de Proyecto y Transparencia Estratégica

Más allá de la excelencia técnica, el **Proyecto Data Alive** se distingue por una gestión de proyecto profesional y transparente, demostrando una sólida capacidad de liderazgo y ejecución. La iniciativa se gestiona bajo una **metodología ágil (Kanban)**, lo que permite una entrega de valor incremental y una gran adaptabilidad ante los desafíos.

La planificación estratégica se refleja en la detallada hoja de ruta de 21 fases, asegurando que cada paso de la construcción tenga un propósito claro y contribuya al objetivo final. Este enfoque metódico garantiza una ejecución ordenada y minimiza los riesgos.

Para fomentar una total transparencia y colaboración, el progreso del proyecto puede ser consultado en tiempo real a través del **tablero público en Trello**, y todo el código fuente y la documentación están disponibles en el **repositorio público de GitHub**. Esta gestión abierta no solo asegura la calidad, sino que también posiciona al proyecto como un portafolio vivo de las mejores prácticas en arquitectura de datos y gestión de proyectos en la nube.