**Diagnóstico del problema**

La organización enfrenta:

* **Demoras en reportes** → falta de automatización y centralización.
* **Errores por duplicación** → ausencia de control de calidad y reconciliación de datos.
* **Decisiones basadas en datos inconsistentes** → falta de trazabilidad y gobierno de datos.
* **Bases de datos aisladas** en ventas, marketing y logística → silos de información.

**Propuesta de integración**

**Tipo de integración recomendada:**

**Integración híbrida:**

* Batch para sistemas históricos (ventas y logística).
* Streaming para eventos en tiempo real (marketing digital, campañas).
* APIs para conectar sistemas externos (CRM, e-commerce).

**Tareas y herramientas sugeridas**:

|  |  |  |
| --- | --- | --- |
| Tarea | Herramienta | Justificación |
| Orquestación de flujos | Apache NiFi | Visual, flexible, soporta múltiples fuentes. |
| ETL batch | Talend / AWS Glue | Transformación y limpieza de datos. |
| Ingesta en tiempo real | Apache Kafka | Manejo de eventos y datos en tiempo real. |
| Repositorio central | Amazon S3 / Data Lake | Almacenamiento escalable y económico. |
| Virtualización | Denodo / Dremio | Acceso unificado sin mover datos. |
| Calidad de datos | Great Expectations / Talend Data Quality | Validación, duplicados, reglas de negocio. |
| Seguridad | IAM (AWS) / Roles RBAC | Control de accesos y auditoría. |

**Flujo de integración propuesto**

1. **Extracción** desde bases aisladas (ventas, marketing, logística).
2. **Transformación** y limpieza con Talend o Glue.
3. **Carga** en un Data Lake (S3).
4. **Virtualización** con Denodo para acceso unificado.
5. **Visualización** en dashboards (Power BI, Tableau).

**Beneficios esperados**

* Reducción de tiempos en generación de reportes.
* Eliminación de duplicados y errores.
* Mejora en la calidad y consistencia de los datos.
* Toma de decisiones más confiable y ágil.
* Mayor trazabilidad y seguridad.