

Arquitectura de Datos para E-commerce y Retail Digital

Propuesta Integral de Solución

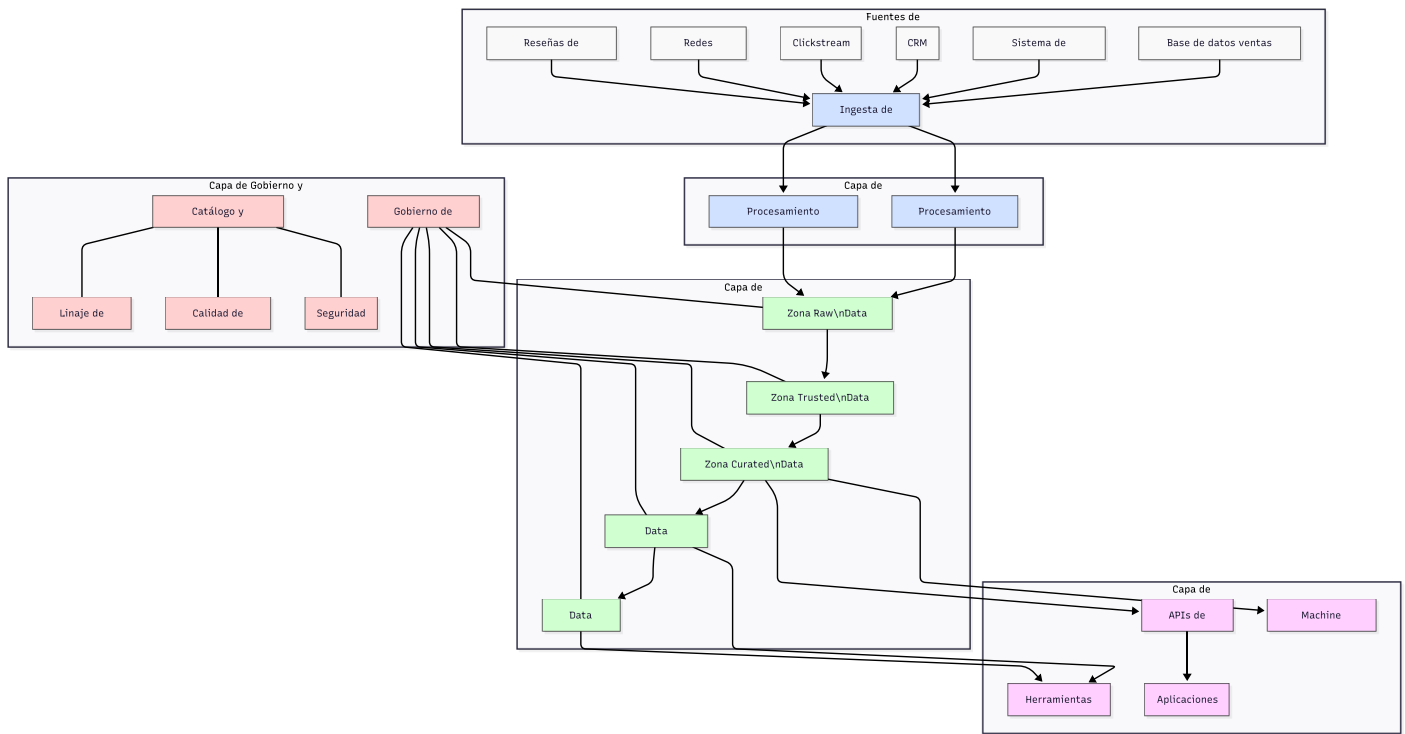
Situación Actual

- Crecimiento acelerado de la empresa
- Gestión de múltiples fuentes de datos heterogéneas
- Necesidad de una arquitectura moderna y escalable
- Requerimiento de calidad y gobernanza de datos
- Requisito de modelo analítico para toma de decisiones

Objetivos del Proyecto

- Diseñar una arquitectura integral de datos
- Integrar fuentes heterogéneas
- Implementar estrategia de almacenamiento escalable
- Definir plan de calidad de datos
- Desarrollar modelo multidimensional para análisis

Arquitectura Propuesta - Visión General



Fuentes de Datos Identificadas

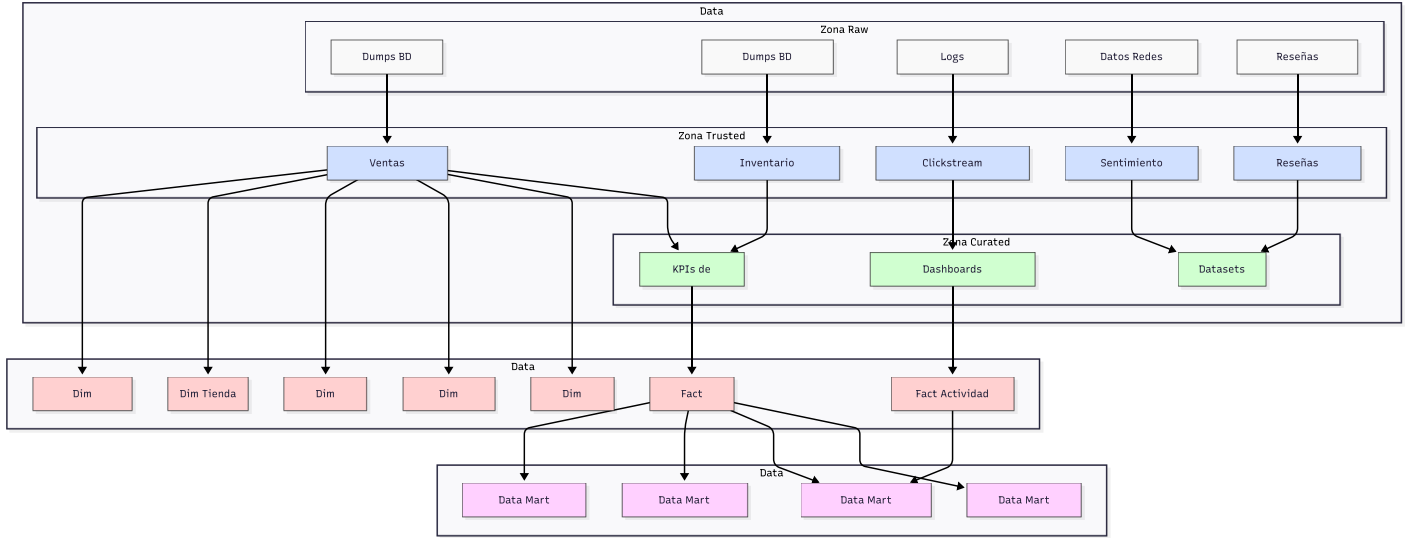
Estructuradas:

- Base de datos transaccional de ventas online (SQL)
- Sistema de inventario (SQL Server)
- CRM (API/SQL)

No Estructuradas:

- Registros de navegación web (JSON/CSV)
- Redes sociales (APIs)
- Comentarios y reseñas de productos (Texto)

Estrategia de Almacenamiento



Zonas de Almacenamiento

Data Lake

- **Zona Raw (Bronze):** Datos originales sin procesar
- **Zona Trusted (Silver):** Datos validados y limpios
- **Zona Curated (Gold):** Datos transformados para casos de uso

Data Warehouse

- Modelo dimensional optimizado para análisis

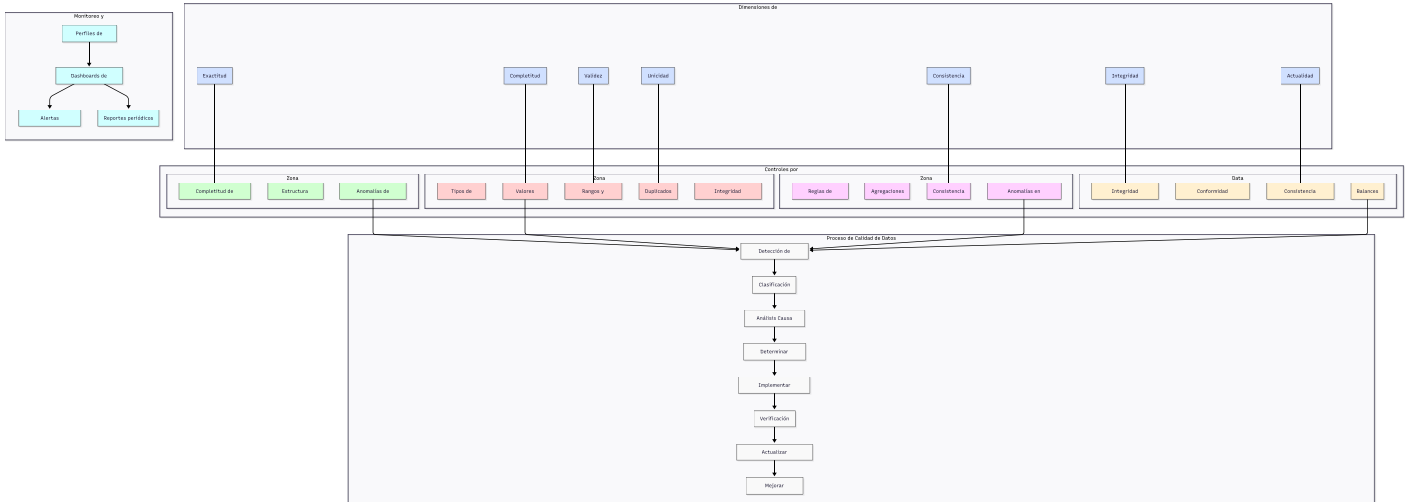
Data Marts

- Subconjuntos departamentales (Ventas, Marketing, Logística, Finanzas)

Tecnologías Recomendadas

Componente	Tecnología AWS	Alternativa
Data Lake	S3	MinIO/HDFS
Procesamiento	EMR, Lambda, Glue	Spark, Databricks
Data Warehouse	Redshift	Snowflake
Orquestación	Step Functions, Airflow	Airflow, Dagster
Visualización	QuickSight	Tableau, Power BI

Plan de Calidad de Datos



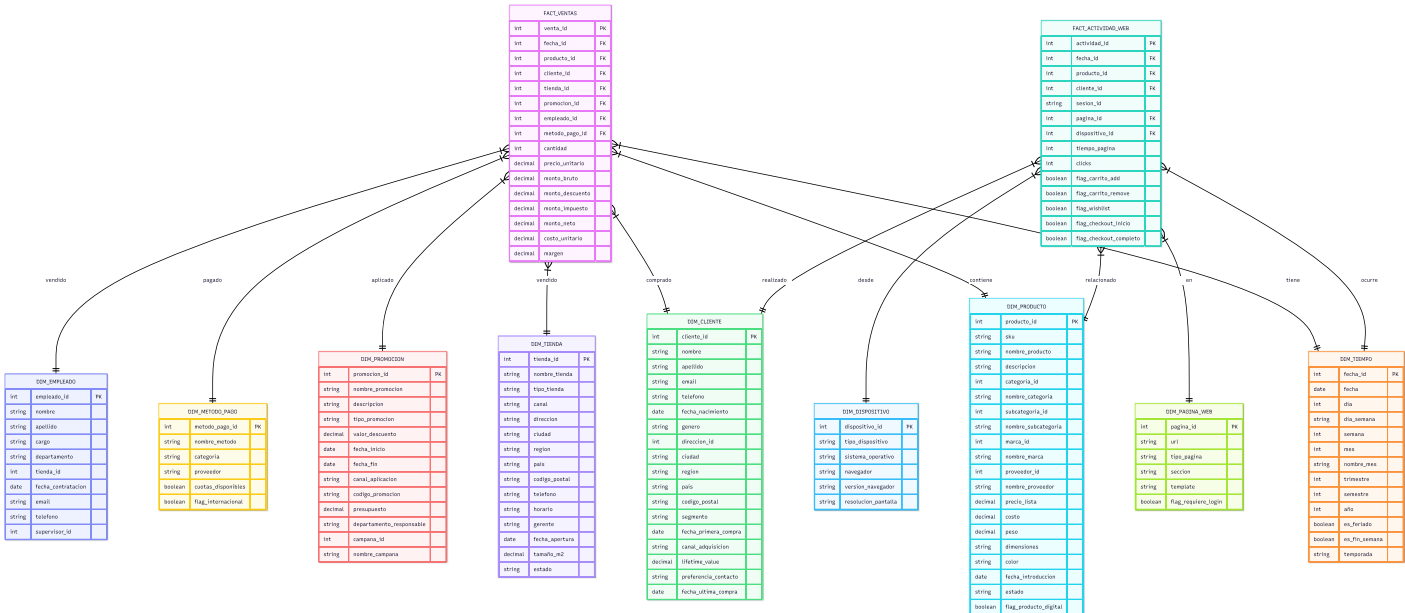
Dimensiones de Calidad

- **Exactitud:** Precisión de valores (crítica para precios, inventario)
- **Compleitud:** Presencia de todos los valores requeridos
- **Consistencia:** Coherencia entre datos relacionados
- **Actualidad:** Datos actualizados oportunamente
- **Unicidad:** Ausencia de duplicados
- **Validez:** Conformidad con reglas de negocio
- **Integridad:** Mantenimiento de relaciones entre datos

Proceso de Monitoreo y Remediación

1. **Detección:** Identificación automática o manual
2. **Clasificación:** Por severidad e impacto
3. **Investigación:** Análisis de causa raíz
4. **Corrección:** Implementación de solución
5. **Verificación:** Validación de la corrección
6. **Prevención:** Mejora de controles

Modelo Multidimensional



Tablas de Hechos y Dimensiones

Hechos:

- **FACT_VENTAS:** Transacciones de venta
- **FACT_ACTIVIDAD_WEB:** Comportamiento en sitio

Dimensiones Principales:

- Tiempo, Producto, Cliente, Tienda
- Promoción, Método de Pago, Empleado
- Página Web, Dispositivo

Beneficios del Esquema Estrella

- **Simplicidad de consultas:** Reducción de joins
- **Rendimiento optimizado:** Para consultas analíticas
- **Facilidad de comprensión:** Estructura intuitiva

- **Compatibilidad:** Mejor soporte en herramientas BI

Casos de Uso Analíticos

1. Análisis de ventas multidimensional

- Ventas por categoría, región y período

2. Análisis de conversión

- Tasa de conversión por dispositivo y canal

3. Efectividad de promociones

- Impacto de promociones en ventas por categoría

Implementación Progresiva

Fase 1 (0-3 meses)

- Infraestructura básica
- Ingesta de fuentes críticas
- Controles de calidad básicos

Fase 2 (3-6 meses)

- Data Warehouse y modelo dimensional
- Procesos ETL/ELT completos
- Controles de calidad extendidos

Fase 3 (6-12 meses)

- Data Marts departamentales
- Capacidades avanzadas de análisis
- Automatización completa

Conclusiones

- Arquitectura escalable y flexible para crecimiento futuro
- Enfoque integral de calidad y gobernanza
- Modelo dimensional optimizado para análisis de negocio
- Solución adaptada a las necesidades específicas de e-commerce
- Base sólida para análisis avanzados y machine learning

Recomendaciones

- Establecer comité de gobierno de datos
- Priorizar casos de uso de alto impacto
- Invertir en capacitación del personal
- Revisar periódicamente la arquitectura
- Desarrollar métricas de éxito para medir el valor generado

¡Gracias!

Propuesta desarrollada por: [Tu Nombre] Ingeniero de Datos