



II.2 Esquemas de data warehouse

Extracción de Conocimiento en Bases de Datos

Luis Eduardo Aguilar Sarabia

Abraham Camacho Ríos

Giselle Cantú Chávez

Karla Alejandra De La Cruz Zea

Ricardo Hernández Martínez

IDGS91N

Docente: Luis Enrique Mascote Cano

Introducción

Este documento presenta el diseño del Data Warehouse para *Industrias Nova*, empresa manufacturera de componentes electrónicos con tres plantas. El objetivo es centralizar y estructurar datos de producción, calidad, inventarios y mantenimiento para soportar decisiones estratégicas: optimizar la producción, reducir tiempos de paro y defectos, y gestionar inventarios con precisión.

El alcance incluye la identificación de hechos y dimensiones, la justificación del esquema en estrella, la descripción del flujo de datos desde las fuentes operacionales hasta el DW, el diagrama conceptual (star schema) y el fundamento técnico que sustenta las decisiones de modelado. El diseño prioriza consultas OLAP para KPIs clave: producción por línea, merma por lote, frecuencia de paros y eficiencia por turno.

Identificación de hechos y dimensiones

Hechos principales (fact):

- Fact_Production — medidas: unidades_producidas, unidades_defectuosas, tiempo_paro_min, productividad.
- Fact_Material_Consumption — medidas: cantidad_consumida, desperdicio, merma.
- Fact_Maintenance — medidas: evento_id, tiempo_reparacion_min, costo_mantenimiento.
- Fact_Quality — medidas: numero_pruebas, numero_fallas, tasa_fallas.
(Identificación tomada del caso suministrado.).

Dimensiones clave:

- Dim_Tiempo (día, mes, trimestre, año)
- Dim_Planta (nombre, ubicación)
- Dim_Linea (línea de producción, planta_fk)
- Dim_Turno (nombre, hora_inicio, hora_fin)
- Dim_Producto (sku, familia)
- Dim_Operador (id, nombre)
- Dim_Material (sku_material, unidad)

- Dim_Proveedor (nombre, lead_time)
- Dim_Maquina (tag, modelo)
- Dim_Falla / Dim_PruebaCalidad (catálogos).

Justificación del modelo de esquema seleccionado

Modelo elegido: Esquema en estrella (Star Schema)

Se selecciona este modelo porque ofrece **simplicidad, rapidez de consulta y flexibilidad analítica**, ideales para los objetivos de Industrias Nova: analizar producción, calidad, mantenimiento e inventario desde distintas perspectivas (línea, turno, planta, fecha, producto).

Principales razones:

1. **Simplicidad y claridad:** estructura central con una tabla de hechos conectada a dimensiones descriptivas; facilita la comprensión para usuarios no técnicos y analistas de negocio.
2. **Rendimiento en consultas OLAP:** reduce el número de uniones (joins) al desnormalizar las dimensiones, optimizando cálculos frecuentes como tasas de defectos o productividad.
3. **Flexibilidad y escalabilidad:** permite incorporar nuevas medidas o dimensiones (por ejemplo, costos o energía) sin rediseñar el modelo.
4. **Compatibilidad con herramientas BI:** Power BI o Tableau interpretan de forma nativa el modelo estrella, facilitando reportes y cubos OLAP.
5. **Mantenimiento controlado:** las dimensiones admiten cambios graduales (SCD), preservando el historial y la consistencia de los datos.

Flujo de datos

Fuentes operacionales: incluyen el sistema MES (producción y líneas), ERP (inventarios y materiales), mantenimiento (eventos y fallas), sistema de calidad e incluso sensores IoT. Cada fuente genera datos con diferentes formatos y periodicidad.

Zona de ingestión (landing): recopila los datos crudos tal como se generan, conservando su trazabilidad antes de cualquier cambio.

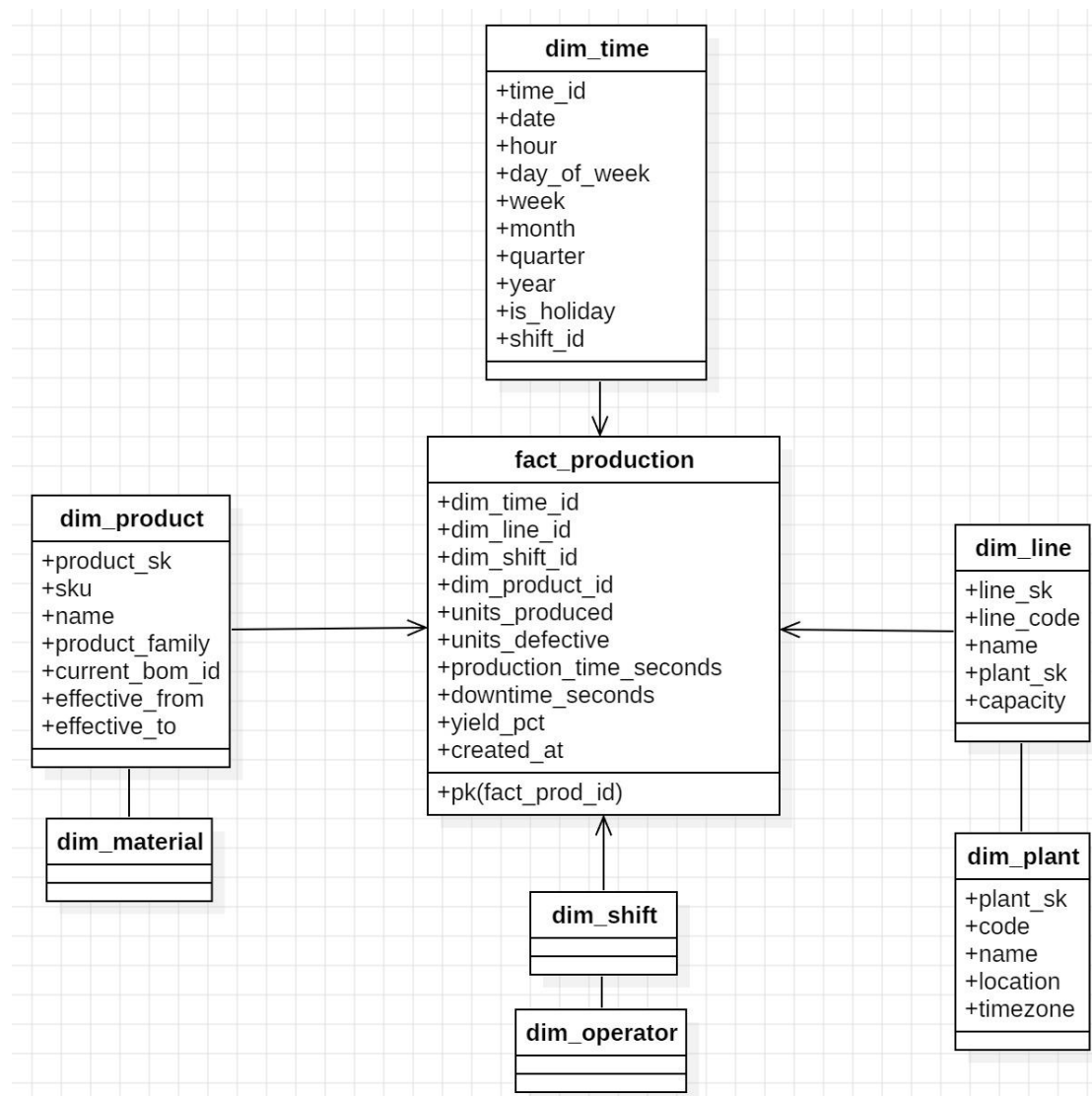
Staging o preparación: se limpian, validan y estandarizan los datos, corrigiendo unidades, formatos y duplicados para garantizar consistencia.

Transformación (ETL): los datos se enriquecen con cálculos clave (por ejemplo, productividad, tasa de defectos, merma) y se vinculan con sus dimensiones (tiempo, turno, línea, producto, etc.).

Carga al DW: las dimensiones se actualizan (según tipo de cambio) y los hechos se cargan de forma incremental, dejando el modelo listo para consultas.

Consumo analítico: los usuarios acceden a los datos mediante herramientas BI (como Power BI), generando reportes y dashboards sobre producción, calidad, mantenimiento e inventarios.

Diagrama Conceptual



Fundamento técnico

1. Granularidad y consistencia

Cada tabla de hechos tiene un grano definido, lo que garantiza que las métricas sean coherentes entre sí:

- *Fact_Production*: una fila por **lote × línea × turno × fecha**.
- *Fact_Material_Consumption*: por **lote × material × fecha**.
- *Fact_Maintenance*: por **evento de mantenimiento**.
- *Fact_Quality*: por **inspección × lote × fecha**.

Definir el grano evita duplicidades y asegura comparaciones válidas entre dimensiones.

2. Claves sustitutas y relaciones

Todas las dimensiones usan claves sustitutas (surrogate keys) de tipo entero para conectar con las tablas de hechos mediante llaves foráneas (FK).

Esto desacopla el DW de los códigos de origen (ERP, MES), permitiendo mantener integridad referencial y flexibilidad frente a cambios en los sistemas operativos.

3. Mantenimiento del historial (SCD)

Algunas dimensiones se actualizan mediante Slowly Changing Dimensions (SCD):

- Tipo 2 (mantiene historial) en *Producto*, *Proveedor* y *Máquina*.
- Tipo 1 (sobrescribe) en *Planta* y *Línea*, dado que cambian con poca frecuencia. Así se conserva el historial de atributos relevantes sin duplicar información innecesaria.

4. Rendimiento y almacenamiento

Las tablas de hechos se particionan por fecha (*date_key*) para optimizar consultas por periodos de tiempo y mejorar la carga incremental.

Además, se crean índices compuestos en columnas de filtrado frecuente (fecha, planta, línea, turno), lo que acelera las consultas de BI.

5. Calidad y trazabilidad de datos

Durante el proceso ETL se aplican validaciones para evitar errores:

- Comparación entre cantidades producidas y consumidas.

- Revisión de rangos válidos (tiempos negativos, valores nulos).
- Registro de auditoría con número de registros cargados y fecha del proceso. Esto garantiza datos confiables y reproducibles en los reportes.

6. Escalabilidad y uso analítico

El modelo soporta crecimiento en volumen de datos (nuevos lotes o periodos) sin afectar la estructura existente.

Además, su diseño en estrella facilita la creación de cubos OLAP y dashboards interactivos, permitiendo análisis de productividad, defectos y mantenimiento con tiempos de respuesta adecuados.

Conclusión

El esquema en estrella propuesto consolida información crítica de MES, ERP, mantenimiento y calidad en un repositorio analítico consistente y eficiente. Al definir granos claros en las tablas de hechos y dimensiones gestionadas con claves sustitutas y SCD cuando corresponde, el DW garantiza trazabilidad y comparabilidad entre indicadores (tasa de defectos, productividad, merma y tiempo de reparación).

Con particionamiento por fecha, índices adecuados y cálculos clave realizados en staging, el modelo ofrece rendimiento para dashboards y análisis ad-hoc, facilitando la detección de cuellos de botella y la toma de decisiones operativas. Se recomienda implementar pipelines ETL/ELT con monitoreo de calidad de datos y comenzar con cargas incrementales para validar métricas antes de ampliar agregaciones y cubos OLAP.

Referencias

What is a Data Warehouse? | Google Cloud. (s. f.). Google Cloud.

<https://cloud.google.com/learn/what-is-a-data-warehouse/>

Denglishbi. (s. f.). *Understand star schema and the importance for Power BI - Power BI.*

Microsoft Learn. <https://learn.microsoft.com/en-us/power-bi/guidance/star-schema/>

Data, S. (s. f.). *Discover best practices your organization should implement to get the most out of your data warehouse and facilitate data analytics.* Stitch Data.

<https://www.stitchdata.com/resources/data-warehouse-best-practices/>

ETL Tutorial: Get Started with ETL | Panoply Guide. (s. f.-b). Panoply.

<https://panoply.io/data-warehouse-guide/etl-tutorial/>

GeeksforGeeks. (2024, 6 agosto). *Slowly changing dimensions*. GeeksforGeeks.
<https://www.geeksforgeeks.org/software-testing/slowly-changing-dimensions/>