Universidad Mayor de San Andrés
Facultad de Ciencias Puras y Naturales

Informe DW con metodología Hefesto - Parte Teórica

DAT 251 - Base de datos 3

Docente:

Lic. Celia Elena Tarquino Peralta

Estudiantes:

- Henrry Yonathan Condori Casa
- Ricardo Andrés Beizaga Marquez
- Muñoz Callisaya Gabriel Marcelo

Fecha de entrega:

22 de marzo del 2025

La Paz - Bolivia

Índice del Proyecto

1. Introducción

1.1. Objetivo del Proyecto

El objetivo principal de este proyecto es diseñar e implementar un Data Warehouse (DW) para analizar la producción y el desperdicio en PIL Andina S.A., una empresa real en Bolivia dedicada a la producción de productos lácteos, como leche pasteurizada, yogures y quesos. Este DW tiene como propósito identificar procesos ineficientes en la elaboración de estos productos, reducir costos asociados al desperdicio de materias primas como leche cruda o envases, y proporcionar a la gerencia información clave para la toma de decisiones estratégicas. Con ello, se busca mejorar la competitividad en el mercado boliviano y optimizar la rentabilidad de PIL Andina S.A., alineando los procesos productivos con estándares de calidad y eficiencia.

1.2. Alcance y Limitaciones

El alcance del proyecto abarca un análisis detallado de los volúmenes de producción y desperdicio generados por cada línea de producción en PIL Andina S.A., considerando factores como el procesamiento de leche y derivados. Incluye la identificación de operarios, turnos y máquinas que registran mayores niveles de desperdicio, el cálculo del costo total del desperdicio según el tipo de producto lácteo (e.g., leche líquida o yogur) o material (e.g., envases desechados), y el monitoreo de la eficiencia de producción a lo largo del tiempo para detectar tendencias estacionales o anomalías. Las limitaciones del proyecto radican en que el análisis se centra exclusivamente en datos relacionados con la producción y el desperdicio, excluyendo áreas como ventas, distribución o logística. Además, depende de la calidad y disponibilidad de la información en los sistemas transaccionales (OLTP) de PIL Andina S.A., lo que podría limitar la profundidad del análisis si los registros son incompletos o inconsistentes.

1.3. Metodología HEFESTO aplicada

La metodología HEFESTO se emplea para estructurar el desarrollo del Data Warehouse de manera sistemática y alineada con las necesidades específicas del negocio de PIL Andina S.A. Esta metodología se organiza en fases clave: análisis de requerimientos, donde se identifican las necesidades de información; diseño conceptual, que establece una visión inicial del DW; diseño lógico y físico, que detalla la estructura técnica; implementación, que pone en marcha el sistema; y validación, que asegura su correcto funcionamiento. Este enfoque garantiza una transición fluida desde la identificación de problemas operativos, como el desperdicio en la producción de lácteos, hasta la entrega de una solución analítica robusta y práctica, adaptada al contexto industrial de la empresa.

1.4. Herramientas y Tecnologías utilizadas

Para llevar a cabo este proyecto, se seleccionaron herramientas específicas que aseguran un desarrollo eficiente y una presentación clara de los resultados. **Pentaho Data Integration (PDI)** se utilizará para los procesos de extracción, transformación y carga (ETL) de datos, permitiendo integrar información de diversas

fuentes de PIL Andina S.A. **Power BI** se empleará para la creación de dashboards y visualizaciones interactivas, facilitando el análisis de desperdicio y producción por parte de la gerencia. **PostgreSQL** servirá como base de datos relacional para almacenar el DW, ofreciendo robustez y escalabilidad. Finalmente, **Typst** se usa para la redacción y presentación de este informe teórico, asegurando un formato profesional y legible.

2. Análisis y Definición de Requerimientos

2.1. Identificación de fuentes de datos

Las fuentes de datos principales para el DW provienen de los sistemas operativos de PIL Andina S.A. El **Sistema ERP de PIL Andina S.A.** registra datos detallados de producción (e.g., litros de leche procesados por turno), inventario (e.g., stock de materia prima) y costos de materiales (e.g., precio por litro de leche cruda o envases). El **Sistema de control de calidad** contiene información crítica sobre desperdicio y scrap generado durante la producción de lácteos, como leche derramada o productos rechazados por defectos. El **Sistema de recursos humanos** proporciona datos sobre operarios, turnos y horarios, esenciales para correlacionar el desempeño humano con los niveles de desperdicio en las líneas de producción.

2.2. Requerimientos del negocio

Mediante entrevistas con gerentes de producción, responsables de planta y operarios de PIL Andina S.A., se identificaron requerimientos clave para optimizar la producción de productos lácteos. Estos incluyen cuantificar volúmenes de producción y desperdicio por línea en un período determinado (e.g., diario o mensual), determinar qué operarios, turnos o máquinas generan mayores niveles de desperdicio (como fallos en pasteurizadoras), calcular el costo total del desperdicio y su distribución por tipo de producto (e.g., yogur vs. queso) o material (e.g., envases plásticos), y evaluar la eficiencia de producción a lo largo del tiempo para detectar tendencias, como incrementos en desperdicio durante turnos nocturnos o picos estacionales.

2.3. Definición de KPIs y métricas clave

Los KPIs y métricas definidos reflejan las necesidades operativas de PIL Andina S.A. en la producción de lácteos: **Cantidad Producida** mide las unidades o lotes fabricados en un período (e.g., litros de leche procesada). **Cantidad de Desperdicio** cuantifica el volumen o peso de material perdido (e.g., litros de leche derramada). **Costo de Desperdicio** calcula el valor monetario del material perdido, usando la fórmula

$$\text{Costo\_Desperdicio} = \text{Cantidad\_Desperdicio} \times \text{Costo\_Unitario}$$

(e.g., costo de leche cruda por litro). **Porcentaje de Desperdicio** evalúa la relación

$$\text{Porcentaje\_Desperdicio} = \frac{\text{Cantidad\_Desperdicio}}{\text{Cantidad\_Producida}} \times 100$$

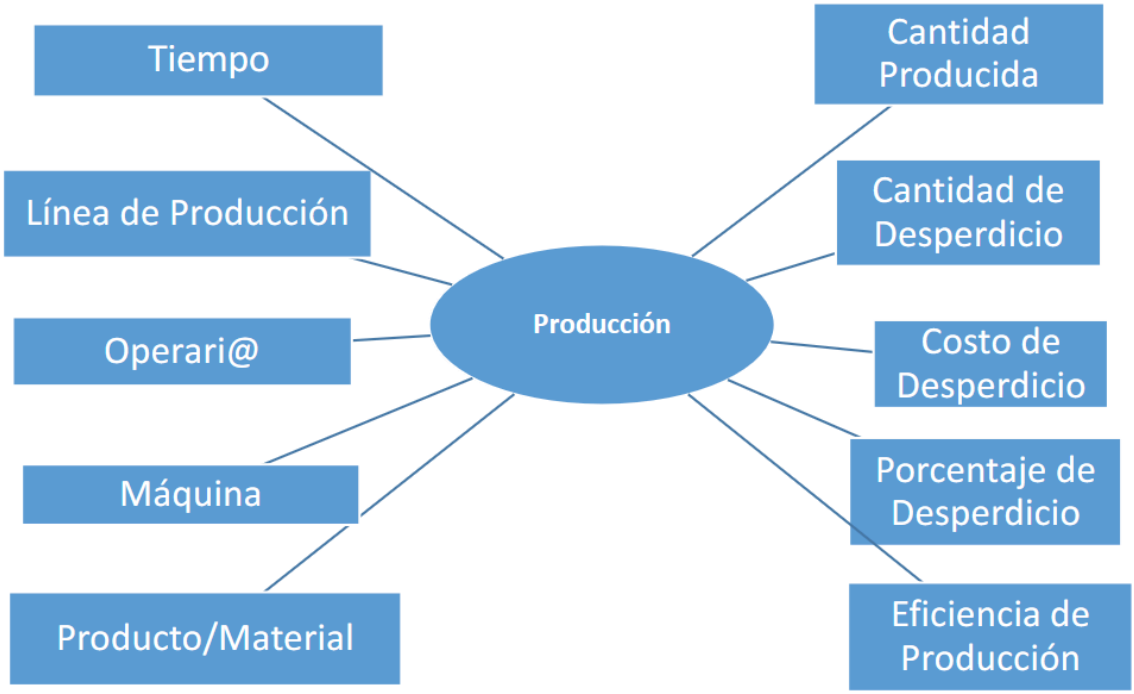
, indicando eficiencia relativa. **Eficiencia de Producción** mide

$$\text{Eficiencia} = \frac{\text{Cantidad\_Producida} - \text{Cantidad\_Desperdicio}}{\text{Cantidad\_Producida}} \times 100$$

, reflejando el rendimiento óptimo de las líneas.

2.4. Modelado conceptual del Data Warehouse

El modelo conceptual establece las bases del DW para PIL Andina S.A., definiendo dimensiones y hechos clave adaptados a su producción de lácteos. Las **dimensiones** incluyen: **Tiempo** (día, semana, mes, trimestre, año, para análisis temporal), **Línea de Producción** (e.g., línea de pasteurización), **Operario** (personal involucrado), **Máquina** (e.g., envasadoras), y **Producto/Material** (e.g., leche líquida o envases). Los **hechos** comprenden: **Cantidad Producida** (litros o unidades), **Cantidad de Desperdicio** (litros o kilos desperdiciados), y **Costo de Desperdicio** (valor en bolivianos). Este modelo se representa gráficamente en una estructura que vincula producción y desperdicio, como se muestra en:



3. Diseño del Data Warehouse

3.1. Diseño de la arquitectura del DW

La arquitectura del DW sigue un enfoque de bus de datos, diseñado para integrar eficientemente las tablas de hechos y dimensiones en una base de datos PostgreSQL. Este esquema centralizado optimiza las consultas analíticas requeridas por PIL Andina S.A., como el análisis de desperdicio por línea o producto lácteo, y permite una escalabilidad futura para incorporar datos adicionales (e.g., nuevos productos o plantas). La elección de PostgreSQL asegura robustez y soporte para grandes volúmenes de datos generados en la producción diaria de lácteos.

3.2. Modelado dimensional (Star Schema / Snowflake)

Se seleccionó un esquema en estrella (Star Schema) por su simplicidad y alto rendimiento en consultas multidimensionales, ideal para las necesidades analíticas de PIL Andina S.A. Las dimensiones se mantienen denormalizadas para minimizar uniones complejas, facilitando reportes rápidos en Power BI sobre desperdicio por turno o máquina. Este modelo contrasta con el Snowflake Schema, descartado por su mayor complejidad, que no se justifica frente a la estructura operativa relativamente directa de la producción de lácteos.

3.3. Diseño de tablas de hechos y dimensiones

El diseño incluye una **Tabla de Hechos** llamada `produccion_hechos`, con campos como claves foráneas a dimensiones, `cantidad_producida` (e.g., litros de leche), `cantidad_desperdicio` (e.g., litros perdidos), y `costo_desperdicio` (en bolivianos). Las **Tablas de Dimensiones** son: `dim_tiempo` (fecha, mes, trimestre, año), `dim_linea_produccion` (ID, nombre, e.g., «Línea Yogur»), `dim_operario` (ID, nombre, turno), `dim_maquina` (ID, nombre, tipo, e.g., «Pasteurizadora»), y `dim_producto` (ID, nombre, categoría, e.g., «Leche Entera»). Esta estructura soporta análisis específicos de PIL Andina S.A.

3.4. Definición de procesos ETL

Los procesos ETL están diseñados para integrar datos desde las fuentes OLTP de PIL Andina S.A. de manera eficiente. La **extracción** se realizará desde sistemas como el ERP y control de calidad, capturando registros de producción y desperdicio diarios. La **transformación** incluye limpieza de datos (e.g., corrección de valores nulos en litros producidos), estandarización de formatos y cálculo de métricas derivadas como el Porcentaje de Desperdicio. La **carga** se hará en las tablas del DW en PostgreSQL, asegurando integridad referencial y consistencia para análisis futuros.