



Proyecto No. 3
Integración y análisis sobre bases de datos
(Grupos de 3 estudiantes)

Modalidad y fecha de entrega:

- El proyecto debe de ser enviado antes de la fecha límite de entrega: viernes 24 de mayo a las 12:00
- No se permitirá la entrega o envío de proyecto más allá de la fecha límite

Objetivo general:

Este proyecto tiene por objetivo implementar un proceso de extracción-transformación-carga (*extraction-transformation-loading*, ETL) para alimentar un datamart a partir de bases de datos transaccionales, con el fin de contar con reportería analítica para la toma de decisiones de manera informada de parte de una gerencia.

El proceso de ETL será trabajado en una herramienta especializada en este tipo de tareas.

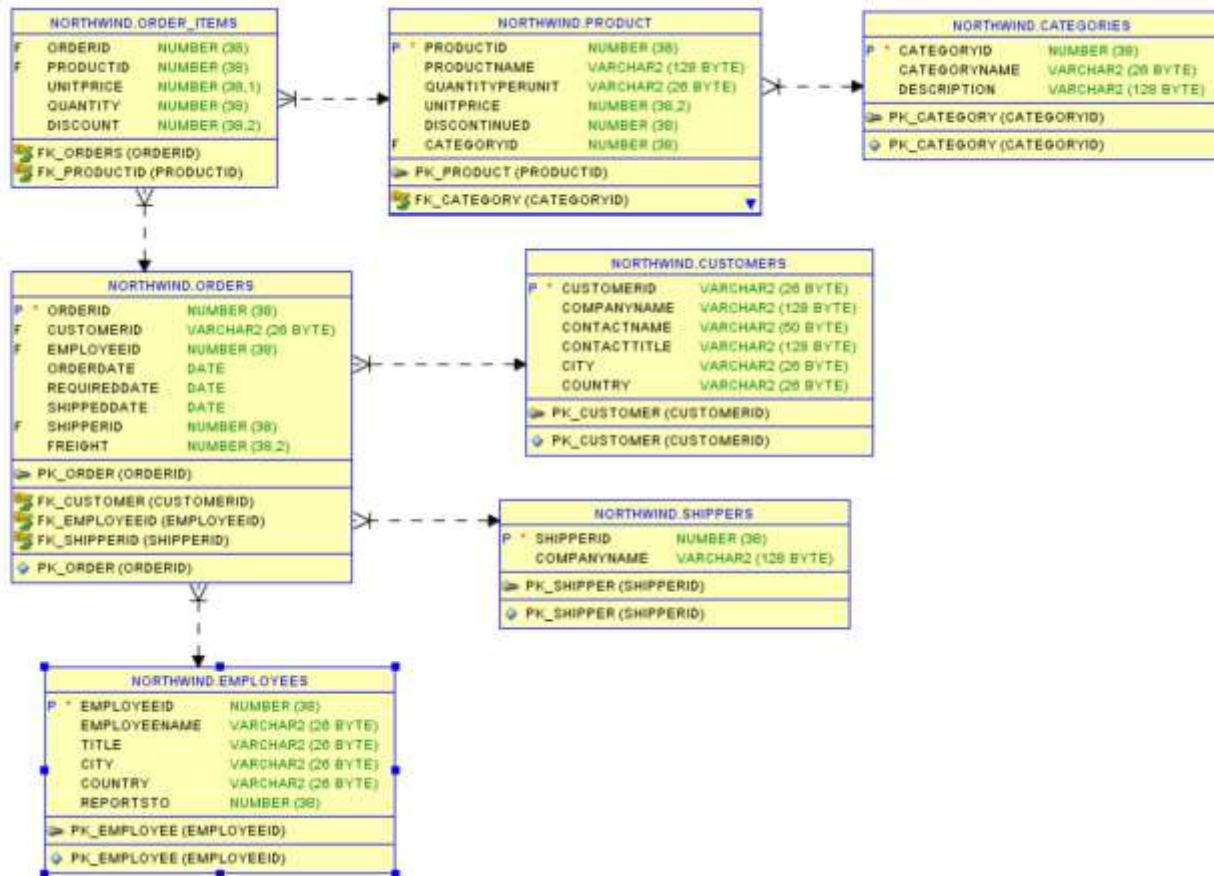
A diferencia de las bases de datos transaccionales (*online-transaction-processing*, OLTP), este datamart está diseñado para realizar análisis (*online-analytical-processing*, OLAP). Se busca también que usted se familiarice con otros motores de bases de datos relacionales, por lo que se usarán motores de bases datos distintos para la base de datos transaccional y la analítica.

Finalmente se ejercitará también la presentación de resultados en herramientas de inteligencia de negocios proponiendo incluso una comparación entre diferentes herramientas disponibles en el mercado.

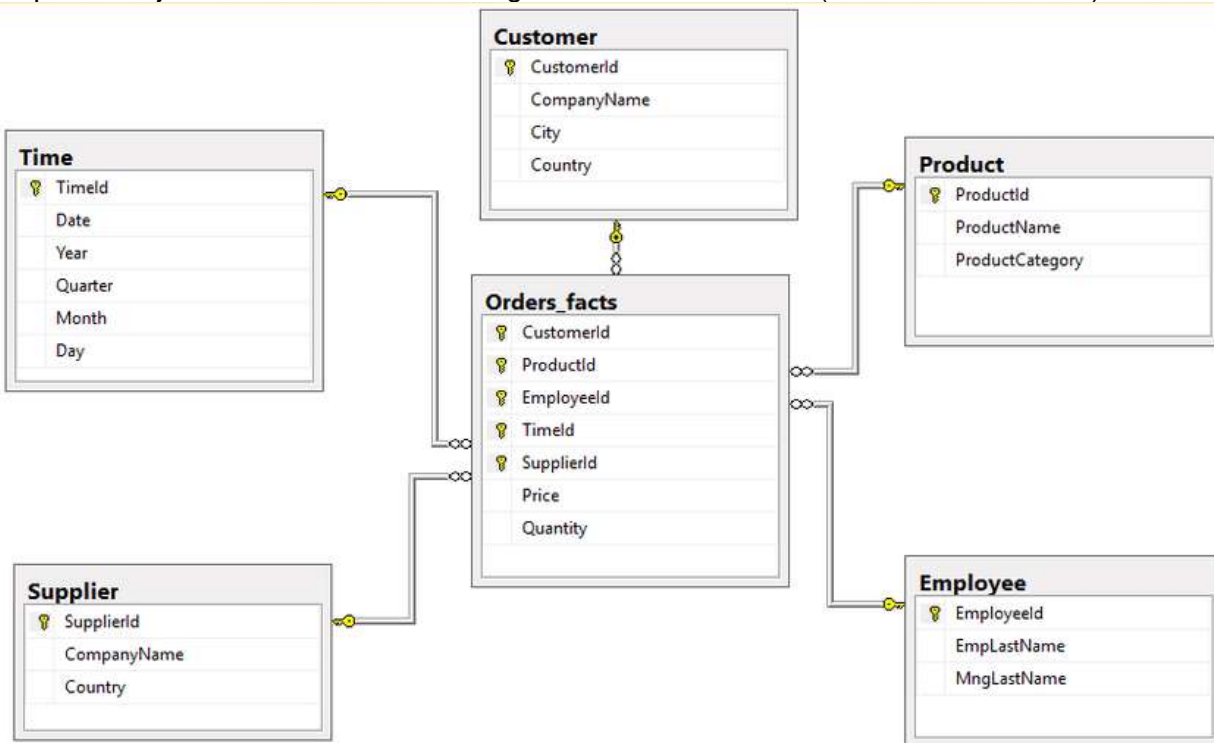
Descripción general del proyecto:

Para este proyecto se le proporcionará un dataset de órdenes de compra de nombre Northwind en formato CSV que representará el sistema transaccional del cual se originará el modelo estrella a formar en su datamart, que servirá de base para construir reportería analítica.

El modelo que se proporcionará está superficialmente basado en el siguiente modelo transaccional:



Un primer objetivo es transformarlo al siguiente modelo estrella (cubo de información):



Consideraciones:

- 1) La dimensión de tiempo en su modelo estrella debe ser generada de manera programática, no obtenida del sistema transaccional.
- 2) El procedimiento a realizar debe de ser el siguiente:
 - a. Instalar una versión express de Oracle Database. Crear allí el usuario que contendrá el transaccional de Northwind. Debe cargarlo basado en los archivos .csv que se le proporciona como base y además crearle los respectivos constraints – llaves primarias y extranjeras.
 - b. En instalación de postgresql, crear allí la base de datos para el datamart y construir allí adentro su modelo estrella.
 - c. Utilizar la herramienta [Knlme](#) para configurar el proceso de ETL, el cual transformará el modelo relacional que se encuentra en Oracle, y lo convertirá en modelo estrella en el manejador de base de datos PostgreSQL. Esto deberá poder hacerse desde dos instancias distintas de Oracle.



- d. Tome en cuenta que debe de incluir las jerarquías respectivas en las correspondientes dimensiones, lo que implica una de-normalización.
 - e. Tome en cuenta que el modelo relacional proporcionado no es ideal. Este modelo relacional es una aproximación, pero para poder construir los queries que darán origen a las dimensiones y la tabla de hechos será necesario escudriñar y estudiar las tablas físicas que están en Oracle.
 - f. Recuerde que un buen modelo estrella asegura buen rendimiento en la reportería analítica independientemente de la herramienta de inteligencia de negocios que utilice.
 - g. Puede utilizar índices, triggers, procedimientos almacenados, vistas e incluso vistas materializadas si considera necesario.
- 3) Una vez tenga construido su datamart en la base de datos Oracle debe de construir en herramientas de inteligencia de negocios la reportería analítica correspondiente para que la gerencia pueda realizar sus análisis. Por tanto:
- a. Debe dejar el rol programador y asumir el de analista. Entre las recomendaciones que debe tomar en cuenta es que para hacer tableros (*dashboards*) y reportería analítica debe considerar cómo funciona el negocio, por lo que los tableros y reportes a construir deben de ser coherentes.
 - b. Se sugiere trabajar reportes que inicialmente representen elementos bastante generales, y que luego se vaya profundizando en gráficas más específicas, y que permitan ir contando la historia de lo que desea plantear (*storytelling*).
 - c. Es necesario que para esta fase construya los mismos reportes y tableros en dos herramientas distintas de inteligencia de negocios, como por ejemplo PowerBI / Tableau / Microstrategy / Qlik, etc.
 - d. El objetivo de utilizar estas dos herramientas es:
 - i. Poder observar cómo un modelo correcto permite intercambiar la herramienta de visualización sin necesidad de modificaciones de fondo.
 - ii. Poder comparar las herramientas en cuanto a los siguientes criterios:

Datamart	¿Qué diferencias observa con respecto al modelo transaccional?
Business intelligence / Herramienta / análisis	¿Resultó factible replicar los tableros y funcionalidades de una herramienta a otra?
Facilidad	¿Qué conclusión tiene con respecto a utilizar las dos herramientas desde el punto de vista de facilidad?
Licenciamiento	Investigar entre precios de licencias entre las dos herramientas
Tipos de usuarios	¿Las herramientas cuentan con usuarios solo de lectura y uso, versus usuarios de construcción de tableros y reportes?



	¿Cuentan con servicio en servidor en la nube u <i>on-premise</i> (local) para publicar los dashboards, y que se tenga reportería analítica corporativa?
Facilidad de soporte técnico y capacitación	Que tal fácil es obtener soporte técnico por bugs y capacitación de las herramientas

Conclusión: ¿Cuál de las dos herramientas de inteligencia de negocios recomendaría utilizar?

Especificación de tecnología:

- Sistema manejador de base de datos: PostgreSQL y Oracle
- Lenguaje de programación a discreción

Temas a reforzar:

- Diseño de bases de datos
- Índices
- Vistas
- Triggers y procedimientos almacenados
- Construcción de modelo estrella para Datamarts
- Uso de herramientas de inteligencia de negocios

Documentos a entregar:

1. ETL construido con Knime
2. Backup del modelo estrella sobre Oracle
3. Informe de visualizaciones desarrolladas en herramienta de inteligencia de negocios
4. Informe y presentación de las conclusiones a las que llegaron

VI. Evaluación

- Construcción del datamart: 25%
- Implementación de reportes y uso de índices: 25%
- Herramientas de inteligencia de negocios: 25%
- Calidad de la presentación: 25%
 - **Total: 100%**

Total: 20 puntos netos