

- La **CLAVE** de la dimensión **TIEMPO** es un **número entero compuesto por el año, el mes y el día** (NO AUTO INCREMENTAL) del formato AAAAMMDD -> convertir la fecha en formato *date* a un *integer* en el formato AAAAMMDD
- ¿Por qué no puede existir una cardinalidad de muchos a muchos entre una tabla FACT y una tabla de DIMENSIONES?
 - Porque multiplicaría exponencialmente los datos entre una tabla y la otra, generando un **Cross join** que en el que por cada FACT, las medidas identificadas no tendrían un número real, tendrían un número exponencial
- Los cálculos se hacen entre la intermedia y la fact
- Si el negocio lo requiere, arrastro los id de los empleados, pero si no no. En el DW voy a tener datos de varios sistemas fuente, pueden haber id repetidos.

Se pueden cargar los ID de los empleados a modo de dato, y utilizar una key unica (por ejemplo, autoincremental) para relacionarla con la table de hecho

- LAS MEDIDAS NO TIENEN QUE ESTAR EN LAS DIMENSIONES, SOLO EN LA FACT
- AUTO INCREMENTAL DE LA PK EN LA FACT NO VA LA MAYORÍA DE LAS VECES
- diferencias controlflow y dataflow!!
 - Control Flow (job) es la parte que organiza y controla qué tareas (task) se ejecutan y en qué orden.
 - Data Flow (task) es la parte que se encarga de mover y transformar datos de un lugar a otro.
- Definir trabajo (job) y transformaciones
- Stored procedure
 - Conjunto de instrucciones SQL precompiladas y almacenadas en una bbdd. Se utilizan para tareas repetitivas o periódicas.
- Integration service que es, para que lo usamos, ejemplos, comparaciones
 - Es una herramienta de Microsoft diseñada para realizar tareas de integración de datos y soluciones de business intelligence. SSIS permite la migración, transformación y limpieza de datos, así como la automatización de flujos de trabajo y procesos
- “de pentaho no hay ninguna pregunta, si puede haber algo de integration, algo muy conceptual” qué es pentaho y para qué puede servir?
-
- Dashboards, que son ? ¿para que se usan ? Ejemplos
 - Un dashboard es, en síntesis, un “informe”, una herramienta visual que presenta información, métricas, etc, de forma resumida.
- Bases de datos SQL vs NOSQL
 - NoSQL no utiliza solo el lenguaje SQL para sus consultas, incluso evitan utilizarlo o solo se usa de apoyo, no tiene estructuras fijas como tablas, no suelen permitir operaciones JOIN.
- que es ETL
 - ETL es el proceso de extraccion, transformacion y carga previo al DW, en este proceso se cargan los datos de diversas fuentes, pasan por un proceso de transformacion y limpieza de datos y por último se vuelcan al DW.

Existen varias metodologías de extracción, entre las más comunes tenemos full extract (donde se extraen todos los datos) e incremental extract (donde solo se procesa lo modificado o agregado)

- que es Big Data/ diferencias entre Big Data y BI
 - Se puede definir como información de mucho volumen procesada a gran velocidad y muy variada, que facilita la obtención de conocimiento y toma de decisiones.
 - BI suele definirse como la transformación de los datos de la compañía en conocimiento. Por un lado se refiere al tipo de análisis de datos. Por otro hace referencia al conjunto de tecnologías que soportan este tipo de análisis

Pelado Rules

escribir más de dos renglones y poner ejemplos de todo.