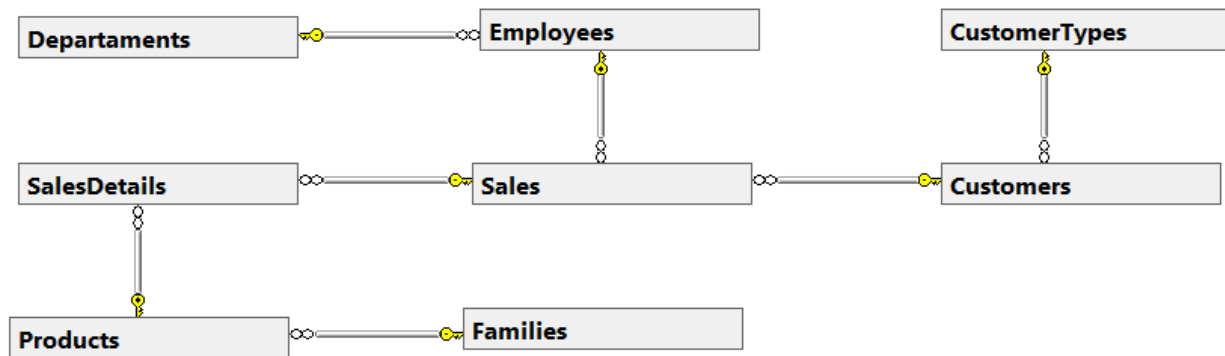


Descrição do problema

O presente trabalho tem por objetivo o desenvolvimento de um data mart, a partir dos dados de vendas (faturação), efetuadas aos clientes de uma empresa (*Steinway*) que se dedica à comercialização de equipamentos/instrumentos musicais. A *Steinway* possui um sistema operacional que suporta o registo das vendas realizadas. O modelo de dados relacional inerente a esse sistema operacional encontra-se apresentado na figura seguinte.



O sistema operacional possui capacidades muito limitadas ao nível da realização de análises de dados às vendas efetuadas aos clientes. Na sequência desta limitação, e também para permitir análises de cariz histórico, pretende-se que **desenvolva um data mart** que permita a realização de **análises variadas às vendas efetuadas aos clientes**.

Requisitos do trabalho

Nos pontos seguintes enumeram-se os requisitos que devem ser respeitados no trabalho a efetuar:

- Desenvolver um processo de análise dimensional, a fim de definir e criar o esquema conceptual para o *data mart*, com base na informação disponibilizada e seguindo a metodologia de *Kimball*. Todas as opções tomadas na criação do modelo dimensional devem ser devidamente justificadas no relatório a produzir. O *data mart* a desenvolver deverá ser concebido de modo a permitir a realização de consultas/análises **aos dados das vendas no nível de granularidade mais elementar**, de modo a possibilitar a maior flexibilidade possível para a realização futura de análises de dados.

- Proceder à extração, transformação, limpeza e carregamento dos dados no *data mart*, por intermédio de um *Integration Services Project* do *Visual Studio* e dos componentes mais adequados que este disponibiliza para as referidas tarefas. Mais especificamente, pretende-se que:
 - Todo o processo de extração, transformação, limpeza, integração e carregamento dos dados seja feito do modo mais eficiente possível. Por questões de confidencialidade relacionadas com o negócio, apenas foi disponibilizada para a construção do *data mart* uma amostra dos dados que se encontram no sistema operacional. No entanto, o volume de registos é bastante elevado.
 - O processo de transformação e limpeza filtre todos os registos cujos dados apresentem problemas de qualidade. Em particular, os dados podem estar afetados por valores em falta em atributos supostamente de preenchimento obrigatório e valores que violam o domínio do atributo. Assim, qualquer registo afetado por problemas de qualidade, não deve ser carregado no *data mart*, sendo armazenado em tabelas especialmente criadas para o efeito na *Staging Area*. A correção destes problemas e posterior carregamento desses registos para o *data mart* fica excluída do âmbito deste trabalho.
 - O carregamento dos dados no *data mart* (tabelas de dimensões e tabela de factos) deve ser concebido de modo a poder ser executado de forma incremental, isto é, em cada execução só serem carregados os dados novos e atualizados os que já existem, caso tenham sofrido alterações.
 - O *Integration Services Project* seja desenvolvido tendo em conta que será posteriormente instalado na pasta *C:\Temp*. Como tal, assumindo que o projeto do Visual Studio é criado com a designação de *TrabPrat*, este deve ser armazenado no interior de *C:\Temp*, o que dará origem ao seguinte path/caminho de pastas: *C:\Temp\TrabPrat*.
- Elaborar um relatório no qual constem todos os elementos relevantes para a avaliação do trabalho, como: arquitetura do *data mart*; modelo dimensional; estruturas de dados criadas na *staging area*; mapeamento de dados entre o sistemas fonte e o *data mart*; processos de extração, transformação, limpeza, integração e carregamento de dados efetuados; *scripts* SQL criados; justificação das opções tomadas; melhoramentos possíveis; etc.

Prazo, formas de realização e entrega do trabalho:

- A realização do trabalho é feita em **grupos de dois ou três alunos**.
- **O Trabalho** (relatório final + *Visual Studio Integration Services Project*) tem de ser **submetido no Moodle** até às **23h59** do dia **11 de janeiro de 2026**.
- O trabalho deve ser submetido sob a forma de um **único ficheiro ZIP**, com o seguinte nome: **XXXXXXX_YYYYYY_ZZZZZZ.zip**, em que XXXXXX, YYYYYY e ZZZZZZ representam os **números dos alunos** que constituem o grupo.
- A discussão do trabalho será efetuada com **os elementos do grupo obrigatoriamente presentes**, em data e hora a combinar oportunamente.

O **incumprimento do prazo de entrega** implica uma **penalização na nota de 20% por cada dia de atraso**.