

1. Apresentação

Objetivo deste trabalho prático é projetar e implementar um banco de dados sobre produtos vendidos em uma loja de comércio eletrônico, incluindo avaliações e comentários de usuários sobre estes produtos. O trabalho consiste na criação de um Banco de Dados Relacional contendo dados sobre compras de produtos e elaboração de um *Dashboard*, um painel para monitoramento dos dados de compra, gerando uma série de relatórios e gráficos. Os dados para o banco de dados serão fornecidos de um arquivo de entrada que será indicado aos alunos.

2. O que entregar

- 2.1 Documentação apresentando um diagrama correspondendo ao esquema do Banco de Dados relacional, observando as diretrizes da Seção 4, além de um dicionário de dados descrevendo cada relação, atributo, restrição de integridade referencial ou de outro tipo que fizer parte do esquema do banco de dados. Deve ser entregue em um único arquivo PDF nomeado **TP2_2.1_Nome1_Nome2.pdf**
- 2.2 Código fonte dos programas desenvolvidos para extração de dados do arquivo de entrada, criação do esquema do banco de dados, e povoamento das relações com estes dados. Deve ser entregue em um único arquivo ZIP nomeado **TP2_2.2_Nome1_Nome2.zip**.
- 2.3 Código fonte do Dashboard implementando no mínimo as consultas descritas na Seção 5. Embora o uso de gráficos para apresentação dos resultados das consultas não seja obrigatório, sua utilização é bastante encorajada e contará para uma boa avaliação do trabalho. Deve ser entregue em um único arquivo ZIP nomeado **TP2_2.3_Nome1_Nome2.zip**

Os códigos fontes dos itens 2.2 e 2.3 devem ser desenvolvidos em C++. Devem ser entregues com os respectivos scripts de geração do executável em ambiente Linux de versão recente. Os códigos deverão estar prontos para ser compilados e executados sem erros. Os executáveis devem receber como parâmetros de entrada os dados de conexão a um banco de dados *PostgreSQL*, ou seja, endereço do servidor; usuário e senha; e nome do banco de dados. Os alunos podem considerar um ambiente onde estão instalados drivers JDBC e/ou ODBC para conexão ao servidor.

3. Como entregar

Cada um dos 3 arquivos acima deve ser copiado em uma pasta Google Drive que ficará aberta até às 23:59 do dia 02/10/2017.

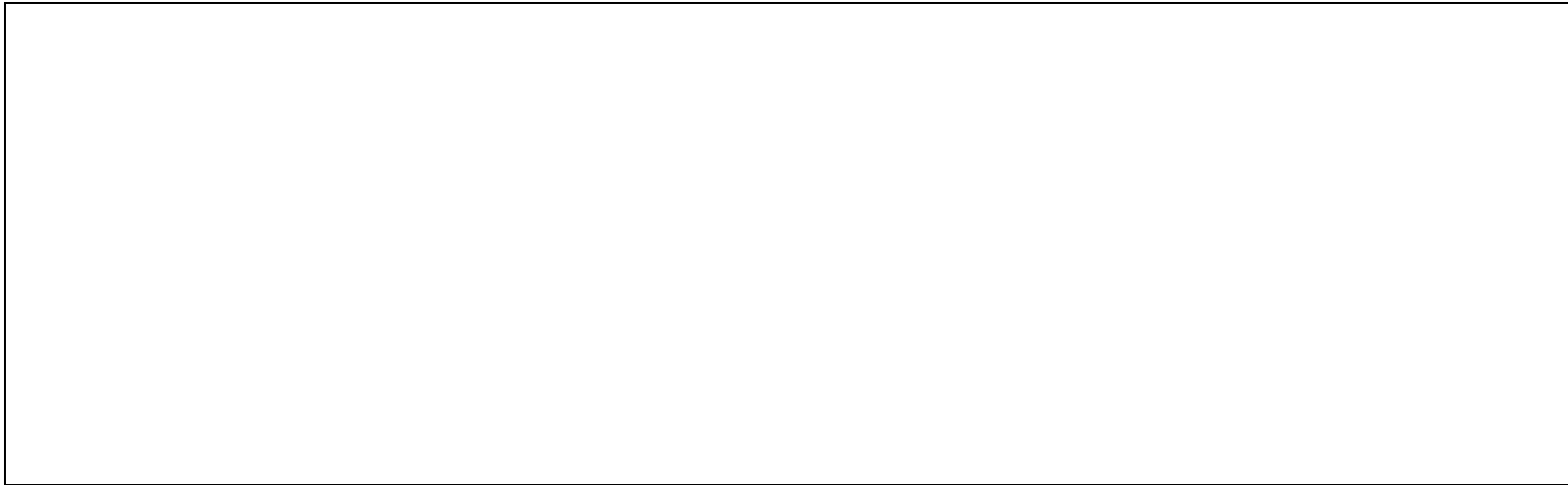
4. Arquivo de Entrada

Amazon product co-

Stanford Network Analysis Project (SNAP). Os dados foram coletados em 2006 do site Amazon e informações de produto e comentários de clientes sobre 548.552 produtos diferentes (livros, CDs de música, DVDs e fitas de vídeo VHS). Para cada produto, a seguinte informação está disponível:

- Título
- Posição no ranking de vendas (Salesrank)
- Lista de produtos ``similares
- Informação de categorização do produto Categorias e subcategorias ao qual o produto pertence
- Comentários sobre os produtos:
 Informação data, id do cliente, classificação, número de votos, o número de pessoas que acharam a avaliação útil

Um trecho do arquivo é apresentado abaixo:



5. **Sobre o Esquema do Banco de Dados:** O esquema de bancos de dados a ser desenvolvido deverá seguir o modelo relacional. Seu desenvolvimento deverá seguir a técnica ascendente (bottom-up) de projeto de bancos de dados relacionais e deve necessariamente observar as regras de uma das formas normais de alto nível tais como a **Forma Normal de Boyce-Codd**, **Terceira Forma Normal** ou **Quarta Forma Normal**. O estudo dessa técnica de projeto e das formas normais faz parte do trabalho. É sugerido que para este estudo os alunos consultem as referências [1] ou [2], ou outra com material semelhante.

6. Dashboard: O dashboard a ser implementado deve dar suporte a pelo menos as seguintes consultas, as quais devem todas ser implementadas com consultas em linguagem SQL.

- (a) Dado produto, listar os 5 comentários mais úteis e com maior avaliação e os 5 comentários mais úteis e com menor avaliação
- (b) Dado um produto, listar os produtos similares com maiores vendas do que ele
- (c) Dado um produto, mostrar a evolução diária das médias de avaliação ao longo do intervalo de tempo coberto no arquivo de entrada
- (d) Listar os 10 produtos líderes de venda em cada grupo de produtos
- (e) Listar os 10 produtos com a maior média de avaliações úteis positivas por produto
- (f) Listar a 5 categorias de produto com a maior média de avaliações úteis positivas por produto
- (g) Listar os 10 clientes que mais fizeram comentários por grupo de produto

7. Referências

- [1]** GARCIA-MOLINA Hector, ULLMAN, Jeffrey D., WIDOM, Jennifer. Database Systems: The Complete Book. 2ª ed. Prentice Hall, 2008. Seções 3.1, 3.3, 3.5 e 3.6
- [2]** ELMASRI, Ramez; NAVATHE, Shamkant B. Fundamentals of Database Systems, 6th edition. Addison Wesley, 2010. Capítulos 10 e 11