

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 elaboração dos projetos conceitual e lógico e da implementação de um Banco de Dados contendo dados sobre compras de produtos, além da 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 projeto conceitual do banco de dados usando o Modelo Entidade-Relacionamento (EER), observando as diretrizes da Seção 4, além de um dicionário de dados descrevendo cada tipo de entidade, relacionamento e atributo. Deve ser entregue em um único arquivo PDF nomeado **TP1_1.1_Nome1_Nome2.pdf**
- 2.2 Documentação apresentando um esquema relacional em linguagem SQL criado a partir do EER acima, além de um dicionário de dados descrevendo cada como cada relação e restrição de residencial foi criado a partir dos tipos de entidade e relacionamento. Deve ser entregue em um único arquivo PDF nomeado **TP1_2.1_Nome1_Nome2.pdf**
- 2.3 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 **TP1_3.2_Nome1_Nome2.zip**.
- 2.4 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 **TP1_4.3_Nome1_Nome2.zip**

Os códigos fontes dos Itens 2.3 e 2.4 devem ser desenvolvidos em Python. Os scripts 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 4 arquivos acima deve ser enviado para ronny@icomp.ufam.edu.br até às 23:59 do dia 08/12/2019. Envios recebidos após este horário serão ignorados. No dia 09/12/2019 passarei a lista dos arquivos que recebi para controle.

4. Arquivo de Entrada

O arquivo de entrada de onde serão extraídos os dados de entrada será o “Amazon product co-purchasing network metadata” que faz parte do 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” (que foram adquiridos junto com o produto)
- Informação de categorização do produto – Categorias e subcategorias ao qual o produto pertence
- Comentários sobre os produtos:
 - o 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:

```
Id: 15
ASIN: 1559362022
  title: Wake Up and Smell the Coffee
  group: Book
  salesrank: 518927
  similar: 5 1559360968 1559361247 1559360828 1559361018 0743214552
  categories: 3
    |Books[283155]|Subjects[1000]|Literature & Fiction[17]|Drama[2159]|United States[2160]
    |Books[283155]|Subjects[1000]|Arts & Photography[1]|Performing Arts[521000]|Theater[2154]
reviews: total: 8  downloaded: 8  avg rating: 4
  2002-5-13  cutomer: A2IGOA66Y6O8TQ  rating: 5  votes: 3  helpful: 2
  2002-6-17  cutomer: A2OIN4AUH84KNE  rating: 5  votes: 2  helpful: 1
  2003-1-2   cutomer: A2HN382JNT1CIU  rating: 1  votes: 6  helpful: 1
  2003-6-27  cutomer: A39QMV9ZKRJXO5  rating: 4  votes: 1  helpful: 1
  2004-2-17  cutomer: AUUVMSTQ1TXDI  rating: 1  votes: 2  helpful: 0
  2004-10-13 cutomer: A5XYF0Z3UH4HB  rating: 5  votes: 1  helpful: 1
```

5. **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.

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