



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. Os dados para o banco de dados serão fornecidos de um arquivo de entrada que será indicado aos alunos.

O trabalho deverá ser desenvolvido em **trios**.

2. Ambiente de Desenvolvimento

Os códigos fontes devem ser desenvolvidos em linguagem Python e o SGBD relacional usado deverá ser o PostgreSQL. Os scripts Python devem fazer acesso direto ao SGDB usando comandos SQL, sem camadas de software intermediário¹.

3. O Que Entregar

3.1 Documentação apresentando um diagrama correspondendo ao esquema do Banco de Dados Relacional, observando as diretrizes da **Seção 6**, 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 tp1_3.1.pdf**

3.2 Um script python para extração dos dados do arquivo de entrada, criação do esquema do banco de dados, e povoamento das relações com estes dados. O script deverá ser entregue em um único arquivo chamado **tp1_3.2.py**

3.3 Um script python para as consultas do Dashboard descritas na Seção 7. O script deverá ser entregue em um único arquivo chamado **tp1_3.3.py**

Atenção: os scripts deverão estar prontos para serem executados sem erros.

¹ Em particular, deve ser evitado o uso dos chamados "frameworks" de mapeamento objeto-relacional (ORM) como Django e outros.

4. Como entregar

A entrega será feita em **um container do Docker** que deverá incluir, além da documentação e dos scripts desenvolvidos, todo o ambiente operacional necessário para o desenvolvimento e execução do trabalho, incluindo o SGBD PostgreSQL, o interpretador Python necessários para a execução dos scripts e toda e qualquer biblioteca necessária para execução do trabalho sem erros.

O container deverá ser nomeado "tp1_<nome1_nome2_nome3>" onde <nome1_nome2_nome3> deve ser substituído pelos nomes dos integrantes do trio que desenvolveu o trabalho. O container deverá ser copiado para [este endereço](#) no Google Drive até as 23:59 do dia **27/06/2022**. Para controle, no dia seguinte, passarei a lista dos containers que recebi.

5. 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.com e contém informações sobre produtos 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: 2
|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
```

6. **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.
7. **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 um 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
- [3] PostgreSQL Python <https://www.postgresqltutorial.com/postgresql-python/>