

**Universidade Federal do Amazonas**  
**Projeto de implementação 4 - Banco de dados II**  
**Hilton dos Santos Costa Neto - 21751599**

Consultas em **MongoDB**

*Geração da coleção de dados:*

*criação e povoamento do banco de dados realizada pelo mongodb shell:*

*-uma coleção genérica foi criada para “instanciar” a base de dados Amazon.*

*-“parser.py” gera arquivo json: amazon.json*

- *use Amazon*
- *db.ini.insert \* a coleção genérica*

*-no diretório do arquivo json abrir terminal:*

- *mongoimport --db Amazon --collection produtos Amazon --file amazon.json*

*\*geração do collection produtos Amazon no db criado anteriormente Amazon*

A mesma estrutura dos projetos de implementação foi mantido. As consultas estão comentadas no arquivo.

(a) Dado produto, listar os 5 comentários mais úteis e com maior avaliação e os 5 comentários mais úteis e com menor avaliação.

- 5 comentários mais úteis e com maior avaliação:

| id | rating | helpful |
|----|--------|---------|
| 2  | 5      | 8       |
| 2  | 5      | 8       |
| 2  | 5      | 5       |
| 2  | 5      | 5       |
| 2  | 5      | 4       |

- 5 comentários mais úteis e com menor avaliação:

| id | rating | helpful |
|----|--------|---------|
| 2  | 1      | 9       |
| 2  | 4      | 16      |
| 2  | 4      | 6       |
| 2  | 4      | 5       |
| 2  | 4      | 5       |

(b) Dado um produto, listar os produtos similares com maiores vendas do que ele

| <b>ASIN (0738700797)</b> | <b>title(Candlemas: Feast of Flames)</b>               | <b>salesrank(168596)</b> |
|--------------------------|--|--------------------------|
| 1567184960               | Yule: A Celebration of Light and Warmth                | 103012                   |
| 0738700525               | Midsummer: Magical Celebrations of the Summer Solstice | 159277                   |
| 0738700940               | Lammas   | 58836                    |

(c) Dado um produto, mostrar a evolução diária das médias de avaliação ao longo do intervalo de tempo coberto no arquivo de entrada

| <b>média diária</b> | <b>data</b> |
|---------------------|-------------|
| 5.0                 | 2001-12-16  |
| 4.0                 | 2002-1-7    |
| 5.0                 | 2002-1-24   |
| 5.0                 | 2002-1-28   |
| 4.0                 | 2002-2-6    |
| 4.0                 | 2002-2-14   |
| 4.0                 | 2002-3-23   |
| 5.0                 | 2002-5-23   |
| 5.0                 | 2003-2-25   |
| 5.0                 | 2003-11-25  |
| 1.0                 | 2004-2-11   |
| 5.0                 | 2005-2-7    |

A média apresentou resultado igual tendo em vista que não houve nenhum comentário no mesmo dia. Seria mais interessante agrupar por ano, logo resultados mais interessantes iriam surgir. Devido a forma como foi pensado o parser a data não foi quebrada em em ano mês e dia, sendo apenas uma string. Apesar de ser uma string atende ao que foi pedido.

(d) Listar os 10 produtos líderes de venda em cada grupo de produtos.

```
{'_id': {'ASIN': 'B00006JBIX', 'grupo': 'DVD', 'salesrank': '0'}}
{'_id': {'ASIN': 'B000059TOC', 'grupo': 'DVD', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00008DDST', 'grupo': 'DVD', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00005RYLV', 'grupo': 'DVD', 'salesrank': '10002'}}
{'_id': {'ASIN': 'B00006FMAX', 'grupo': 'DVD', 'salesrank': '10004'}}
{'_id': {'ASIN': '0783225539', 'grupo': 'DVD', 'salesrank': '10004'}}
{'_id': {'ASIN': 'B00005QTAU', 'grupo': 'DVD', 'salesrank': '10006'}}
{'_id': {'ASIN': '0783234260', 'grupo': 'DVD', 'salesrank': '10008'}}
{'_id': {'ASIN': 'B00004L8IJ', 'grupo': 'DVD', 'salesrank': '10018'}}
{'_id': {'ASIN': 'B000059TGG', 'grupo': 'DVD', 'salesrank': '10024'}}
```

```
{'_id': {'ASIN': '0310208084', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '0963542613', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '0552139432', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '0945199198', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '9700507734', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00008MNUJ', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '9627762644', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '0879736836', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': '0060537612', 'grupo': 'Book', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00006RGI2', 'grupo': 'Book', 'salesrank': '0'}}
```

```
{'_id': {'ASIN': 'B00005MHUG', 'grupo': 'Music', 'salesrank': '0'}}
{'_id': {'ASIN': 'B0000508ZN', 'grupo': 'Music', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00004XT2L', 'grupo': 'Music', 'salesrank': '0'}}
{'_id': {'ASIN': 'B000065AHM', 'grupo': 'Music', 'salesrank': '0'}}
{'_id': {'ASIN': 'B00000JZZ0', 'grupo': 'Music', 'salesrank': '100001'}}
{'_id': {'ASIN': 'B000000DB6', 'grupo': 'Music', 'salesrank': '100004'}}
{'_id': {'ASIN': 'B00000J913', 'grupo': 'Music', 'salesrank': '100006'}}
{'_id': {'ASIN': 'B000065V9T', 'grupo': 'Music', 'salesrank': '100007'}}
{'_id': {'ASIN': 'B000002KEJ', 'grupo': 'Music', 'salesrank': '10001'}}
{'_id': {'ASIN': 'B0000544DI', 'grupo': 'Music', 'salesrank': '100012'}}
```

```
{'_id': {'ASIN': 'B00005QBZI', 'grupo': 'Video', 'salesrank': '0'}}
{'_id': {'ASIN': '6300215539', 'grupo': 'Video', 'salesrank': '1'}}
{'_id': {'ASIN': 'B00005NNT1', 'grupo': 'Video', 'salesrank': '100'}}
{'_id': {'ASIN': 'B00005A1V9', 'grupo': 'Video', 'salesrank': '1000'}}
{'_id': {'ASIN': '6302593247', 'grupo': 'Video', 'salesrank': '10001'}}
{'_id': {'ASIN': '6304494122', 'grupo': 'Video', 'salesrank': '10002'}}
{'_id': {'ASIN': '0307142612', 'grupo': 'Video', 'salesrank': '10003'}}
{'_id': {'ASIN': 'B00000FDY6', 'grupo': 'Video', 'salesrank': '100048'}}
{'_id': {'ASIN': '0970193203', 'grupo': 'Video', 'salesrank': '100058'}}
{'_id': {'ASIN': 'B000059HI4', 'grupo': 'Video', 'salesrank': '10006'}}
```

```
{ '_id': {'ASIN': 'B000053W7W', 'grupo': 'Toy', 'salesrank': '10732'}}
{'_id': {'ASIN': '1572810939', 'grupo': 'Toy', 'salesrank': '1890'}}
{'_id': {'ASIN': '0963679678', 'grupo': 'Toy', 'salesrank': '2288'}}
{'_id': {'ASIN': 'B000055XW3', 'grupo': 'Toy', 'salesrank': '31296'}}
{'_id': {'ASIN': 'B000059T05', 'grupo': 'Toy', 'salesrank': '4053'}}
{'_id': {'ASIN': '0963679627', 'grupo': 'Toy', 'salesrank': '45241'}}
{'_id': {'ASIN': '0963679600', 'grupo': 'Toy', 'salesrank': '59'}}
{'_id': {'ASIN': 'B000075A80', 'grupo': 'Toy', 'salesrank': '7812'}}
```

É possível encontrar alguns salesrank em posição errada, salesrank que são menores depois de um maior. A consulta fez a ordenação pelo atributo salesrank, porém o parser pegou o salesrank como string e não fez a conversão para inteiro. Devido a muitas das consultas mesclar entre python e MongoDB, foi estabelecido persistir com os valores em salesrank.

Existem outros grupos de produtos no dataset da Amazon, apesar da pouca presença de produtos no dataset, menos que 10: Toy, Video Games, Software, Sports, Baby Products.

(e) Listar os 10 produtos com a maior média de avaliações úteis positivas por produto.

| avgh               | ASIN       |
|--------------------|------------|
| 243.0              | 0967699703 |
| 233.0              | 188553891X |
| 203.0              | 0960316833 |
| 197.0              | 1575025809 |
| 196.0              | 0968519903 |
| 186.46511627906978 | 0140502416 |
| 186.34883720930233 | 0448421658 |
| 186.32558139534885 | 0140951172 |
| 186.30232558139534 | 0670672238 |
| 183.0              | 0811701166 |