**Professor Eduardo Kugler Viegas**

**Fundamentos de Big Data**

**Prática MapReduce**

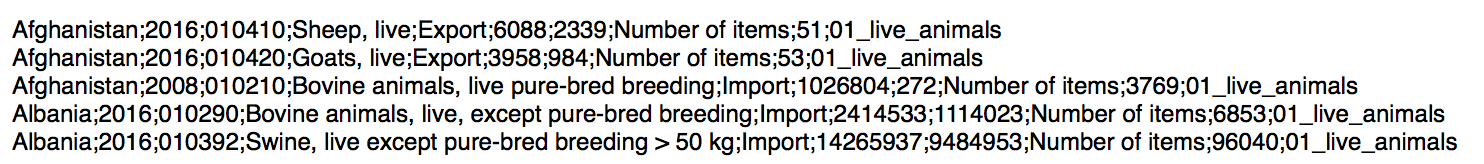
Considerando o caso a seguir, implemente as soluções em Java para extrair o conjunto de informações solicitadas

**Descrição:**

Você foi contratado por uma empresa para efetuar uma análise de dados. Esta empresa possui acesso a uma base de dados com dados sobre as transações comerciais entre países nos últimos 30 anos. Sendo que, para cada transação comercial presente nesta base de dados os seguintes campos são fornecidos:

|  |  |  |
| --- | --- | --- |
| ***#*** | **Campo** | **Descrição** |
| ***0*** | País | País envolvido na transação comercial |
| ***1*** | Ano | Ano em que a transação foi efetuada |
| ***2*** | Código | Código da mercadoria |
| ***3*** | Mercadoria | Descrição da mercadoria |
| ***4*** | Fluxo | Fluxo, e.g. *Exportação* ou *Importação* |
| ***5*** | Valor | Valor em dólares |
| ***6*** | Peso | Peso da mercadoria |
| ***7*** | Unidade | Unidade de medida da mercadoria, e.g. *Quantidade de itens* |
| ***8*** | Quantidade | Quantidade conforme a unidade especificada da mercadoria |
| ***9*** | Categoria | Categoria da mercadoria, e.g. *Produto Animal* |

No total, a base de dados possui mais de 2.8 milhões de transações comerciais. A base de dados foi fornecida no formato CSV, sendo que cada entrada (transação comercial), é representada por uma linha no arquivo. Enquanto cada linha possui os campos listados previamente, estes separados pelo caractere “;”. A imagem a seguir exibe as 5 primeiras transações comerciais presentes na base.



Diante desse contexto, você foi encarregado pelo desenvolvimento de um conjunto de soluções que permitam a extração das seguintes informações sobre a base:

1. País com a maior quantidade de transações comerciais efetuadas;
2. Mercadoria com a maior quantidade de transações comerciais no Brasil (como a base de dados está em inglês utilize Brazil, com Z, utilize a função “contains” do Java);
3. Quantidade de transações financeiras realizadas por ano;
4. Mercadoria com maior quantidade de transações financeiras;
5. Mercadoria com maior quantidade de transações financeiras em 2016;
6. Mercadoria com maior quantidade de transações financeiras em 2016, no Brasil (como a base de dados está em inglês utilize Brazil, com Z); 6 tá Estranha (tudo 2 ou 1), ver com ele.
7. Mercadoria com maior total de peso, de acordo com todas transações comerciais;
8. Mercadoria com maior total de peso, de acordo com todas transações comerciais, separadas de acordo com o ano;
9. Média de peso por mercadoria, separadas de acordo com o ano;
10. Média de peso por mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), separadas de acordo com o ano;
11. Média de peso por mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), em relação ao fluxo, separadas de acordo com o ano;
12. Média de valor por peso, de acordo com a mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), separadas de acordo com o ano;
13. Valor máximo de código;
14. Mercadoria com o maior preço por unidade de peso;
15. Quantidade de transações comerciais de acordo com o fluxo, de acordo com o ano;

Com base no seu conhecimento em MapReduce, para cada uma das soluções requisitadas, forneça:

1. **O código fonte da função de Map**
2. **O código fonte da função de Reduce**

**Inf1)**

**-DESCRIÇÃO:**

País com a maior quantidade de transações comerciais efetuadas;

**-MAP:**

**Tela preta com letras brancas

Descrição gerada automaticamente**

Variáveis:

- Nenhuma.

Info:

Mapeia o País e uma transação: <País, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz a: <País, (Int :: Transações).

**Inf2)**

**-DESCRIÇÃO:**

Mercadoria com a maior quantidade de transações comerciais no Brasil (como a base de dados está em inglês utilize Brazil, com Z, utilize a função “contains” do Java);

**-MAP:**

**Tela preta com letras brancas

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia a mercadoria de País=Brazil e uma transação: <Mercadoria, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela de celular com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz as transações por chave.

**Inf3)**

**-DESCRIÇÃO:**

Quantidade de transações financeiras realizadas por ano;

**-MAP:**

**Tela preta com letras brancas

Descrição gerada automaticamente**

Variáveis:

- ano: ano da ocorrência.

Info:

Mapeia o ano e uma transação: <Ano, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz as transações por chave.

**Inf4)**

**-DESCRIÇÃO:**

Mercadoria com maior quantidade de transações financeiras;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- Nenhuma.

Info:

Mapeia a mercadoria e uma transação: <Mercadoria, (Int :: Transação -> 1)>.

**-REDUCE:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- qt: quantidade de transações.

Info:

Reduz as transações por chave.

**Inf5)**

**-DESCRIÇÃO:**

Mercadoria com maior quantidade de transações financeiras em 2016;

**-MAP:**

**Tela preta com letras brancas

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia a mercadoria e uma transação para ano=2016: <Mercadoria, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela preta com letras brancas

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz as transações por chave.

**Inf6)**

**-DESCRIÇÃO:**

Mercadoria com maior quantidade de transações financeiras em 2016, no Brasil (como a base de dados está em inglês utilize Brazil, com Z);

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia a mercadoria de País=Brazil e uma transação para ano=2016: <Mercadoria, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz as transações por chave.

**Inf7)**

**-DESCRIÇÃO:**

Mercadoria com maior total de peso, de acordo com todas transações comerciais;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- p: peso.

Info:

Mapeia a mercadoria e o peso: <Mercadoria, (Long :: Peso)>.

**-REDUCE:**

Tela de celular

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz os pesos à soma deles por chave.

**Inf8)**

**-DESCRIÇÃO:**

Mercadoria com maior total de peso, de acordo com todas transações comerciais, separadas de acordo com o ano;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- p: peso.

Info:

Mapeia a mercadoria e peso com o ano da ocorrência: <Mercadoria+Ano, (Int :: Peso)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- qt: quantidade de transações.

Info:

Reduz os pesos à soma por chave.

**Inf9)**

**-DESCRIÇÃO:**

Média de peso por mercadoria, separadas de acordo com o ano;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- p: peso.

Info:

Mapeia a mercadoria e peso com o ano da ocorrência: <Mercadoria+Ano, (Int :: Peso)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- avg: soma dos elementos.

- qntd: quantidade de elementos.

Info:

Reduz à média os valores por chave.

**Inf10)**

**-DESCRIÇÃO:**

Média de peso por mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), separadas de acordo com o ano;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia a mercadoria e peso com o ano da ocorrência: <Mercadoria+Ano, (Int :: Peso)>.

**-REDUCE:**

Tela preta com letras brancas

Descrição gerada automaticamente

Variáveis:

- avg: soma dos elementos.

- qntd: quantidade de elementos.

Info:

Reduz à média os valores por chave.

**Inf11)**

**-DESCRIÇÃO:**

Média de peso por mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), em relação ao fluxo, separadas de acordo com o ano;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- p: peso.

Info:

Mapeia País+Fluxo+Ano e o peso da transação: <País+Fluxo+Ano, Float :: Peso>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- avg: soma dos elementos.

- qntd: quantidade de elementos.

Info:

Reduz à média os valores por chave.

**Inf12)**

**-DESCRIÇÃO:**

Média de valor por peso, de acordo com a mercadoria comercializadas no Brasil (como a base de dados está em inglês utilize Brazil, com Z), separadas de acordo com o ano;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- price: preço.

Info:

Mapeia a mercadoria e o ano da ocorrência para país=Brazil: <Mercadoria+Ano, (Int :: Preço)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- avg: soma dos elementos.

- qntd: quantidade de elementos.

Info:

Reduz à média os valores por chave.

**Inf13)**

**-DESCRIÇÃO:**

Valor máximo de código;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia o código e seu valor: <Código, (Float :: Preço)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- ff: valor base de comparação.

Info:

Reduz à máxima os valores por chave.

**Inf14)**

**-DESCRIÇÃO:**

Mercadoria com o maior preço por unidade de peso;

**-MAP:**

**Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

- price: preço.

Info:

Mapeia a mercadoria e preço por unidade de peso: <Mercadoria, (Float :: Preço)>.

**-REDUCE:**

Tela de computador com texto preto sobre fundo branco

Descrição gerada automaticamente

Variáveis:

- dd: valor base de comparação.

Info:

Reduz à máxima os valores por chave.

**Inf15)**

**-DESCRIÇÃO:**

Quantidade de transações comerciais de acordo com o fluxo, de acordo com o ano;

**-MAP:**

**Tela preta com letras brancas

Descrição gerada automaticamente**

Variáveis:

- b: vetor com as informações das colunas do arquivo .csv.

Info:

Mapeia país, fluxo e ano com uma transação: <País+Fluxo+Ano, (Int :: Transação -> 1)>.

**-REDUCE:**

Tela preta com letras brancas

Descrição gerada automaticamente

Variáveis:

- qt: quantidade.

Info:

Reduz o número de transações por chave: <País+Fluxo+Ano, (Int :: Transações)>.