2023-2024



Projeto de processamento estruturado de infomração

Trabalho Prático – época normal

**Trabalho elaborado por:**

**Grupo 19**

8220169 – César Ricardo Barbosa Castelo

8220337 – Hugo Ricardo Almeida Guimarães

8220307 – Pedro Marcelo Santos Pinho

Índice

[Índice de Figuras 2](#_Toc154400533)

[Chave de Siglas 3](#_Toc154400534)

[Introdução 4](#_Toc154400535)

[1. Mongo DB 5](#_Toc154400536)

[a. Estrutura da base de dados 5](#_Toc154400537)

[a. Clientes 5](#_Toc154400538)

[b. Produtos 5](#_Toc154400539)

[c. Vendas 5](#_Toc154400540)

[d. Devoluções 5](#_Toc154400541)

[b. Importação dos dados fornecidos 5](#_Toc154400542)

[c. Integração do BaseX com o MongoDB 5](#_Toc154400543)

[2. Organização do XML 6](#_Toc154400544)

[a. Estrutura do XML 6](#_Toc154400545)

[b. Regras do XML (XML Schema) 6](#_Toc154400546)

[Conclusão 7](#_Toc154400547)

# Índice de Figuras

# Chave de Siglas

|  |  |
| --- | --- |
| XML | Extensible Markup Language |
| XSD | XML Schema Definition |
| CSV | Comma Separated Values |
| JSON | Java Script Object Notation |

# Introdução

Este trabalho foi realizado para o âmbito da disciplina de *Processamento Estruturado de Informação*, que funcionará como integrador dos conhecimentos adquiridos no decorrer das aulas.

Este trabalho consiste na construção de um vocabulário XML para suportar o envio de relatórios de vendas e devoluções de cada parceiro da empresa *Phone for You*, uma empresa que comercializa smartphones através de várias lojas. Assim cada parceiro terá de implementar nos seus sistemas informáticos, um módulo que suporta a geração de documentos XML de vendas e devoluções de acordo com vocabulário estabelecido.

Para além da construção de um vocabulário XML para suportar o envio de relatórios de vendas e devoluções, também será preciso fazer uma transformação dos dados já existentes de vários ficheiros CSV para uma base de dados orientada a documentos (MongoDB), para além disso, também será necessário fazer uma API para aceder aos dados, e construir os relatórios para depois retornar ao utilizador.

# Mongo DB

## Estrutura da base de dados

Para a realização deste trabalho, foi-nos enviado vários ficheiros, em formato CSV, que representam um subconjunto de informação tipicamente armazenada por um dos parceiros da *Phone for You*, com o objetivo de fazer a importação deles para o Mongo DB. No entanto, como a informação de ficheiros estava muito fragmentada, uma importação direta era inviável, já que as consultas iriam demorar imenso tempo. Então para resolver esse problema, fez-se uma transformação dos dados quando estes foram importados. Ficando os dados divididos segundo as seguintes coleções:

### Clientes

[imagem do json]

Para poder importar informação dos clientes, foi preciso juntar vários ficheiros em uma única coleção, sendo esses ficheiros: “*address.csv”*, “city.csv” e “country.csv”.

Como cliente tem uma morada, e cada morada tem que ter registado uma cidade, e uma cidade está registada a um país, decidiu-se juntar a informação dessas tabelas em uma única coleção.

### Produtos

[imagem do json]

Para os produtos aconteceu algo semelhante à coleção dos clientes, no entanto em vez de se juntar a morada, juntou-se o produto com as categorias, visto que um produto tem uma ou mias categorias, e uma categoria pode ter uma ou mais subcategorias. Juntando assim a informação dos ficheiros *“Product.csv”*, *“sub\_category\_product.csv”*, *“sub\_category.csv”* e *“category.csv”*.

### Vendas

[imagem do json]

Para poder criar a coleção das vendas, foi preciso unir os ficheiros: *“sales\_header.csv”*, *“sales\_lines.csv”*, pois, uma “sales\_header” pode ter uma ou mais “sales\_lines”. Para além de unir esses dois ficheiros, também se colocou alguma informação do cliente que comprou o produto, e dos produtos comprados, assim evita-se ter de percorrer várias coleções para quando se for criar relatórios, fazendo com a aplicação tenha um desempenho muito melhor.

### Devoluções

[imagem do json]

Nas devoluções, juntou-se um pouco de toda a informação das outras coleções criadas anteriormente, pois, uma devolução é efetuada por um cliente, sobre um produto em específico, e esse produto está associado a uma venda a partir do código da sua fatura (invoice\_id).

## Importação dos dados fornecidos

## Integração do BaseX com o MongoDB

# Organização do XML

## Estrutura do XML

## Regras do XML (XML Schema)

# Conclusão

A realização deste trabalho foi bem concebida tendo em conta o que nos foi proposto e também as nossas próprias exigências para o trabalho, entre elas, a constante busca pela estrutura da base de dados, que conseguisse gerar relatórios da forma mais eficiente possível.

Ao longo da realização do projeto deparamo-nos com vários desafios, como por exemplo, a parte da criação da migração dos dados dos ficheiros CSV para o mongoDB, e a passagem dos relatórios de JSON para XML a partir do baseX, que de acordo com todos os elementos do grupo foi o que deu mais trabalho. Mas com isto concluímos que este projeto foi muito importante para o nosso desenvolvimento, pois conseguimos colocar em prática aquilo que foi lecionado durante as aulas, assim, sedimentando os nossos conhecimentos.