



Especificação de Projeto Integrador (PI)

Módulo 3

Peso na nota final de cada disciplina no módulo: 40%

Entrega via Moodle: até 11/12/2022

Este trabalho tem como objetivo realizar um trabalho prático que aplique os conceitos das disciplinas do Módulo 3: O trabalho prático pode ter como tema um problema/aplicação oriunda de sua empresa ou um tema fictício a escolha dos alunos envolvidos. O tema escolhido para desenvolvimento deve cumprir com os requisitos dispostos neste documento.

O trabalho pode ser feito em grupo de até 4 alunos. Entretanto, cada aluno do grupo deve fazer uma gravação breve sobre o trabalho, mostrando um documento oficial no início do vídeo para identificação. A solução deve ser proposta exclusivamente pelo grupo com base nos conhecimentos adquiridos ao longo das aulas e pesquisas feitas nas disciplinas.

Requisitos

Descrição Geral. Expandir o Data Warehouse (DW) desenvolvido com PostgreSQL no PI do Módulo 2 para recuperar imagens pelo seu conteúdo (ou seja, **vetores de características**) bem como realizar **consultas** e **produzir gráficos** para dashboard utilizando **Python**. Além disso, efetuar a **especificação** de como implantar esse mesmo DW em um NoSQL de sua escolha (**não é necessária a implementação do NoSQL, porém é imprescindível a descrição, no relatório, de quais adaptações seriam necessárias**). Para tanto, seu Projeto Integrador (PI) deve atender os requisitos conforme descrito ao longo deste documento. O PI forma portanto um trabalho completo que corresponde a uma única avaliação transversal. O tema do data warehouse é de livre escolha do grupo.

Arquivos solicitados para entrega. Na entrega do PI, o grupo enviará um arquivo compactado contendo um relatório em pdf (que inclui os nomes dos membros) em conjunto com os demais arquivos solicitados abaixo com o conteúdo definido por cada disciplina (**entrega do grupo**). Cada membro do grupo deve produzir também um vídeo curto (entre 3 a 5 minutos) descrevendo o que faz sua aplicação (por exemplo, mostrando sua execução). No início do vídeo, é requerido mostrar um documento oficial com foto para comprovar sua identidade (por exemplo, RG). **Somente um**

membro do grupo pode fazer a entrega do grupo, entretanto, todos os membros do grupo devem entregar o vídeo de explicação. Compacte todos esses arquivos e envie o arquivo resultante no Moodle, conforme instruções na disciplina geral da Coordenação do Curso.

Entrega do grupo:

- I. Relatório em pdf;
- II. Banco de dados (arquivo de backup gerado pelo PostgreSQL ou script de criação);
- III. Código fonte da implementação (e.g. arquivo .py **ou** se usar Jupyter Lab/Notebook incluir tanto o .ipynb como o .pdf do notebook com todas as células executadas).

Entrega individual:

- I. Vídeo.

Requisitos da disciplina: Recuperação de Informação Baseada em Conteúdo

Descrição. Considerando o data warehouse construído, o grupo deve:

- Descrever como armazenar as imagens (ou outros dados complexos) no data warehouse projetado e implementado
- **Armazenar os vetores de características no data warehouse mantido no PostgreSQL**
- Criar consultas por similaridade range query e kNN (aceita-se o valor fixo para k, ex: os 5 mais similares) sobre o data warehouse que mantém os vetores de características

Observação: utilizar os vetores de características disponíveis na Internet. Exemplos:

<https://www.kaggle.com/datasets>

<http://archive.ics.uci.edu/ml/index.php>

<https://registry.opendata.aws>

<https://toolbox.google.com/datasetsearch>

Arquivos solicitados para entrega. O relatório (Item I) deve conter os vetores de características que são armazenados no data warehouse, além das respostas das consultas range query e kNN escritas em SQL, utilizando os mecanismos que o professor irá descrever durante as aulas.

Requisitos da disciplina: Linguagens de Programação para Ciência de Dados

Descrição. O grupo deverá utilizar a linguagem de programação Python para realizar a implementação de consultas ao data warehouse e gerar gráficos que possam ser incorporados a um dashboard.

Arquivos solicitados para entrega. O relatório (item I) deve conter os gráficos gerados. Também será avaliado o código fonte (Item III).

Requisitos da disciplina: Introdução ao Big Data

Descrição. Considerando o data warehouse construído (conforme especificado nos requisitos anteriores), **porém sem levar em conta o armazenamento de vetores de características**, o grupo deve discutir como ele poderia ser implantado em um NoSQL. O relatório (Item I) deve descrever essa transformação baseando-se no seguinte guia:

- Qual é o NoSQL escolhido e por que da sua escolha?
- Qual é o modelo de dados do NoSQL escolhido?
- Por que seria interessante migrar do PostgreSQL ao NoSQL escolhido? detalhar a motivação.
- Como implantar as tabelas de dimensão e de fato no NoSQL escolhido?
- Qual seria a sintaxe das consultas analíticas no NoSQL escolhido? apresente exemplos.

Observação: Não há a necessidade de implementar efetivamente o NoSQL e as consultas podem ser apresentadas em linhas gerais (mas utilizando a sintaxe do NoSQL escolhido).

Arquivos solicitados para entrega. O relatório (Item I) deve conter o consenso do grupo em relação ao plano de migração para o NoSQL escolhido, esclarecendo, no mínimo, as questões apontadas anteriormente (utilize gráficos, desenhos, tabelas, ..., se necessário).

Material para Entregar

Arquivo compactado. Deve-se preparar o arquivo compactado contendo todos os arquivos mencionados em cada requisito descrito anteriormente. **Somente um membro do grupo pode fazer a entrega deste arquivo compactado, entretanto, todos os membros do grupo devem entregar o vídeo de explicação (de 3 a 5 minutos, e mostrando um documento oficial para fins de identificação).**

Critério de Correção

Critério de avaliação do trabalho. A combinação dos requisitos forma um sistema completo. A nota final do projeto integrador fará parte da composição da nota final de cada disciplina do módulo.

Bom Trabalho!